Л. ДЕВРОЙ, Л. ДЬЁРФИ

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ПЛОТНОСТИ

L_1

ИЗДАТЕЛЬСТВО "МИР"

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ПЛОТНОСТИ

Wiley Series in Probability and Mathematical Statistics

NONPARAMETRIC DENSITY ESTIMATION

THE L_1 VIEW

LUC DEVROYE
McGill University
Montreal, Canada

LÁSZLÓ GYÖRFI
Hungarian Academy of Sciences
Budapest, Hungary

John Wiley & Sons
New York · Chichester · Brisbane · Toronto · Singapore

1985

Л. ДЕВРОЙ, Л. ДЬЁРФИ

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ПЛОТНОСТИ

L_1-ПОДХОД

Перевод с английского
А. Б. ЦЫБАКОВА

под редакцией
М. Б. МАЛЮТОВА

МОСКВА «МИР» 1988

ББК 22.172
Д25
УДК 519.22

Деврой Л., Дьёрфи Л.

Д 25 Непараметрическое оценивание плотности. L_1-подход: Пер. с англ. — М.: Мир, 1988. — 408 с.

ISBN 5-03-000475-0

Книга известных специалистов (Канада, ВНР), представляющая собой первое достаточно полное изложение теории непараметрического оценивания плотности распределения — важного раздела современной математической статистики. Описаны основные типы оценок плотности, приведены результаты, относящиеся к их состоятельности и скорости сходимости. Основное внимание уделяется исследованию сходимости оценок в естественной для плотностей распределения метрике L_1. Русское издание дополнено новым материалом.

Для научных работников, имеющих дело с обработкой статистических данных, для студентов, специализирующихся по математической статистике.

Д $\frac{1702060000—146}{041\ (01)—88}$ 27—88, ч. 1 ББК 22.172

Редакция литературы по математическим наукам

ISBN 5-03-000475-0 (русск.)

ISBN 0-471-81646-9 (англ.)

© 1985 by John Wiley & Sons, Inc. All Rights Reserved. Authorized translation from English language edition published by John Wiley & Sons, Inc.

© перевод на русский язык, с дополнениями, «Мир», 1988

Предисловие редактора перевода

Оценивание распределения повторных изменений и, в частности, оценивание плотности распределения (существование которой обычно естественно предположить) — центральная задача математической статистики. Гистограммная оценка плотности издавна является общепринятым элементом описательной статистики. Теоретические результаты о состоятельности и скорости сходимости непараметрических оценок плотности, развивающих гистограммные, начали появляться с конца пятидесятых годов на основе теоретико-информационных результатов в теории приближений. Оказалось, что наилучшая скорость сходимости таких оценок — более медленная, чем для параметрических моделей, в зависимости от априорного класса плотностей, которому принадлежит искомая. В настоящее время для некоторых метрик (например, L_2) наилучшая скорость сходимости найдена с точностью до эквивалентности.

В книге Л. Девроя и Л. Дьёрфи основное внимание уделено разработке красивой теории, изучающей состоятельность оценок f_n плотности f и скорость сходимости L_1-расстояния

$$J_n = \int |f_n - f| \, dx.$$

Полученные фундаментальные результаты затем эффективно применяются: к исследованию поведения широкого класса непараметрических оценок, охватывающего основные используемые на практике приемы; к исследованию корректности применения оценок плотности при приближенном статистическом моделировании искомого распределения (весьма популярном сейчас в связи с развитием метода *бутстрэп*); к задачам дискриминации популяций и классификации измерений.

Уже эти приложения оправдывают усилия, которые понадобятся читателю для овладения содержательной, но сложной теорией поведения величины J_n, впервые подробно представленной в этой книге. Список книг, посвященных более узким классам оценок плотности или практическим аспектам их применения, приведен в литературе к гл. 1 и комментарию к ней.

В комментариях к пяти главам книги переводчик А. Б. Цыбаков сообщил сведения о полезной дополнительной литературе

на ту же тему (в частности, советских авторов, мало цитируемых в книге), о дополнительных результатах авторов, полученных после выхода книги. Авторы любезно прислали также список замеченных ими ошибок. Ряд неточностей был устранен А. Б. Цыбаковым, деятельность которого намного превышала обычную работу переводчика.

Перевод книги снабжен двумя дополнениями, написанными нашими известными специалистами Н. Н. Ченцовым и В. Н. Вапником. В первом из них, в частности, объяснена важность подхода L_1, изучаемого в книге, с точки зрения общей теории статистических решений и геометрии распределений. Во втором дополнении дана исчерпывающая теория состоятельности общих оценок плотности и разработаны общие методы построения классов таких оценок, основанные на принципе упорядоченной минимизации риска и общей теории некорректных задач.

Таким образом, в книге впервые дается изложение принципов, лежащих в основе одной из центральных проблем математической статистики, и углубленный обзор методов, применяемых при ее решении. Книга послужит ценным пособием для студентов, аспирантов и специалистов в области математической статистики и будет способствовать росту числа грамотных применений этих методов.

М. Б. Малютов

Предисловие к русскому изданию

Мы очень благодарны издательству «Мир» за публикацию нашей книги в переводе на русский язык, благодаря чему более широкий круг представителей сильной советской школы по теории вероятностей и математической статистике сможет познакомиться с книгой.

Настоящее издание на русском языке нечто большее, чем просто пассивный перевод, поскольку в нем содержатся важные редакционные замечания А. Б. Цыбакова и М. Б. Малютова, а также превосходные статьи Н. Н. Ченцова и В. Н. Вапника, поясняющие уместность L_1-подхода к оцениванию плотности.

В нашей книге приводятся аргументы в пользу исследования оценок плотности в метрике L_1. Затем мы проводим само исследование, обнаруживая ряд новых заманчивых свойств оценок плотности. Хотя многие задачи при этом остаются нерешенными, мы надеемся заразить своим энтузиазмом по поводу L_1-теории оценивания плотности других ученых и полагаем, что издание на русском языке будет способствовать развитию советских исследований в этой области

Люк Деврой
Монреаль, март 1987 г.
Ласло Дьёрфи
Будапешт, март 1987 г.

Посвящается Би и Кати

Предисловие

В книге рассматривается сходимость в метрике L_1 оценок плотности распределения, основанных на выборке из независимых одинаково распределенных случайных векторов со значениями в R^d. Мы предприняли попытку разработать хорошую L_1-теорию, так как более полно изученная L_2-теория приводит к ряду аномальных эффектов и неправильных представлений. Было бы неверно утверждать, что в книге содержится исчерпывающее описание всех известных результатов в метрике L_1. Скорее, это изложение совокупности взглядов, в котором упор делается на результаты, позволяющие лучше понять оценивание плотности. Умышленно ограничиваясь L_1-теорией, мы, разумеется, не могли не упустить некоторые интересные и зачастую глубокие исследования по непараметрическому оцениванию плотности.

Хочется надеятся, что некоторые разделы книги увлекут читателя, и все-таки в основном она должна казаться довольно скучной (всем, кроме редких любителей технических деталей). Так что книга не рекомендуется ни для использования в качестве шпаргалки, ни для изучения во время телевизионных рекламных пауз. Надо было пожертвовать хоть чем-то ради краткости, общности и оптимальности — трех целей, которые мы перед собой поставили. Например, неглубокие результаты, перегруженные избыточными условиями, имеют, как правило, простые и короткие доказательства. Иногда необходимо пройти длинными окольными путями, чтобы обобщить эти результаты и избавиться от таких удобных условий. По пути часто наталкиваешься на достойные упоминания изящные побочные результаты, и вся работа превращается, прежде чем успеешь это осознать, в почти непроходимые джунгли из технических деталей. Создание книги сопровождалось вспышками возбуждения и энтузиазма: нас охватывало волнение всякий раз, когда одному из нас удавалось заполнить пробел или построить мост между какими-нибудь понятиями, а энтузиазм вызывали такие простые вещи, как красивые неравенства. Именно волнение и поспешность ответственны за ошибки, которые читатель, быть может, обнаружит.

В выборе тем и математического аппарата на нас оказали влияние ранние работы по непараметрическому оцениванию плот-

ности (Парзен, Розенблат), современная французская школа (Жеффруа, Боск, Дехёвельс, Абу-Жауд, Бретаньоль, Юбер, Бирже) и некоторые относительно надавние публикации на близкие темы (Гиман, Стил, Стоун). Мы весьма признательны Терри Вагнеру, Шандору Чиби, Кларку Пенроду, Полю Дехёвельсу, Адаму Кшижаку, Питеру Холлу и Годфриду Туссену за прямую помощь, выразившуюся в форме обсуждений и советов. Мы также благодарны канадскому фонду NSERC за щедрую финансовую поддержку, а Макгиллскому университету за то, что он не сделал из этой субсидии никаких вычетов. Наконец, мы хотели бы поблагодарить всех коллег и друзей, которые, беседуя с нами, способствовали, сами того не осознавая, лучшему пониманию вопросов оценивания плотности. Это — Ален Берлине, Люсьен Бирже, Денис Боск, Жан Бретаньоль, Пэт Брокетт, Жерар Коломб, Том Ковер, Бен Фокс, Стьюарт Гиман, Пит Грёнебом, Вилфрид Гроссман, Антонио Гвалтиеротти, Катрин Юбер, Жан-Пьер Лекутр, Фред Мэчелл, Мэнни Парзен, Георг Плюг, Пал Ревес, Дэвид Скотт, Майк Стил, Джим Томсон и Вольфганг Вертц.

Люк Деврой,
Ласло Дьёрфи

Монреаль, Канада,
Будапешт, Венгрия
Октябрь, 1984 г.

Глава 1

Введение

В обширной литературе по непараметрическому оцениванию плотности [1]) каждая книга неизбежно охватывает лишь ограниченный круг вопросов. Не является исключением и данная книга. При выборе подхода мы руководствовались общими принципами, в частности в отличие от других работ рассматривали все плотности как элементы пространства L_1, а не L_2 или L_∞. Мы также не рассматривали оценки, не являющиеся плотностями, считая, что плотности следует оценивать плотностями. Так как пространство L_1 — естественное пространство для плотностей, то использование его глубинных свойств приводит к очень хорошей теории, не обремененной ненужными условиями. Мы попытаемся сформулировать все теоремы в наиболее общей и простейшей форме.

В этой книге рассматривается следующая задача. Пусть *данные* $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные векторы со значениями в R^d, имеющие общую плотность распределения f. *Оценка плотности* есть последовательность $f_1, f_2, \ldots$, где $f_n(x) = f_n(x; X_1, \ldots, X_n)$ является действительнозначной борелевской функцией своих аргументов при каждом n и плотностью на R^d при фиксированных $n, X_1, \ldots, X_n$.

Выбор L_1-расстояния $J_n = \int |f_n - f|$ мотивируется тем, что оно инвариантно относительно монотонных преобразований координатных осей и всегда определено. Пусть, например, X и Y — случайные векторы с плотностями f и g соответственно. Применим к X и Y преобразование $T: R^d \to R^d$, достаточно богатое в том смысле, что $\{T^{-1}B \mid B \in \mathcal{B}\} = \mathcal{B}$, где $\mathcal{B}$ — класс всех борелевских множеств на R^d (это влечет за собой, что преобразование является взаимно однозначным). Независимо от вида T имеем

$$\int |f - g| = \int |f^* - g^*|, \tag{1}$$

---

[1]) Здесь и далее термин *плотность* означает *плотность распределения* случайной величины (вектора). — *Прим. перев.*

где f^* и g^* — плотности распределения $T(X)$ и $T(Y)$. В частности, при $d = 1$ расстояние J_n инвариантно относительно непрерывных строго монотонных преобразований. Свойство (1) вытекает из следующего результата.

Теорема 1 (Scheffé (1947)). *Для любых плотностей f и g на R^d имеем*

$$\int |f - g| = 2 \sup_{B \in \mathscr{B}} \left| \int_B f - \int_B g \right|. \tag{2}$$

Доказательство. Пусть $B = \{f > g\}$ и $A \in \mathscr{B}$. Так как $\int (f - g) = 0$, $\int |f - g| = 2 \int_B (f - g)$, то левая часть (2) меньше или равна правой. Кроме того,

$$\left| \int_A f - \int_A g \right| = \left| \int_{A \cap B} (f - g) + \int_{A \cap B^c} (f - g) \right| \leqslant$$

$$\leqslant \max \left(\int_{A \cap B} (f - g), \int_{A \cap B^c} (g - f) \right) \leqslant$$

$$\leqslant \max \left(\int_B (f - g), \int_{B^c} (g - f) \right) = \frac{1}{2} \int |f - g|, \quad \forall A \in \mathscr{B},$$

где через B^c обозначается дополнение множества B. Соотношение (2) доказано.

Теперь легко получить свойство инвариантности (1):

$$\int |f^* - g^*| = 2 \sup_B \left| \int_B f^* - \int_B g^* \right| = 2 \sup_B \left| \int_{T^{-1}B} f - \int_{T^{-1}B} g \right| =$$

$$= 2 \sup_B \left| \int_B f - \int_B g \right| = \int |f - g|.$$

Иными словами, при $d = 1$ можно получить наглядное представление о величине J_n, оценивая площадь между графиками плотностей f и g, нарисованных в любом выбранном нами линейном или нелинейном масштабе, или даже между графиками соответствующих преобразованных плотностей на $[0, 1]$ при преобразованиях $T\colon R \to [0, 1]$. Теорема 1 также устанавливает связь L_1-расстояния между f и g с максимальным расхождением вероятностей любых борелевских множеств, вычисленных с помощью плотностей f и g соответственно.

Рассмотрим теперь L_p-расстояние $\left(\int |f - g|^p \right)^{1/p}$. Заменим X и Y на aX и aY, где $a \neq 0$ — масштабный множитель и размер-

ность X и Y равна 1. Тогда плотность распределения случайной величины aX равна $f^*(x) = (1/a) f(x/a)$. Следовательно,

$$\left(\int |f^* - g^*|^p\right)^{1/p} = a^{(1-p)/p} \left(\int |f - g|^p\right)^{1/p}. \tag{3}$$

За исключением случая $p = 1$, все L_p-расстояния зависят от используемого масштаба. Для их сравнения друг с другом нет такой универсальной основы, как соотношение (2). Например, сравнение величин $\int (f_n - f)^2$ и $\int (f_n - g)^2$, где f_n — некоторая оценка, используемая как для f, так и для g, не имеет смысла ввиду (3). В то же время в силу теоремы Шеффе [1]) величины $\int |f_n - f|$ и $\int |f_n - g|$ вполне сравнимы. При этом имеет смысл утверждение типа: «оценка f_n лучше для данной плотности f, чем для плотности g».

Читатель без труда может доказать следующее утверждение: для любых f и любых $p > 1$ существуют последовательности плотностей f_n и g_n, такие, что $\int |f_n - f| \downarrow 0$, $\int |f_n - f|^p \uparrow \infty$, $\int |g_n - f| = c > 0$ и $\int |g_n - f|^p \downarrow 0$. Отсюда видно, что не существует простых соотношений или неравенств между L_1- и L_p-расстояниями.

Все оценки плотности основаны на теореме Лебега о плотностях:

$$\lim_{h \downarrow 0} \int_{S_{xh}} \frac{f(y)\, dy}{\lambda(S_{xh})} = f(x) \tag{4}$$

для почти всех x, где S_{xh} — замкнутый шар радиуса h с центром в точке x и λ — мера Лебега. Выражение под знаком предела в левой части (4) равно $P(X_1 \in S_{xh})/\lambda(S_{xh})$, так что его можно аппроксимировать величиной

$$f_n(x) = \sum_{i=1}^{n} \frac{I_{[X_i \in S_{xh}]}}{n\lambda(S_{xh})}, \tag{5}$$

где I — индикаторная функция. Оценка (5) была предложена Розенблатом в 1956 г.[1] Чтобы выражение под знаком предела в (4) хорошо аппроксимировало $f(x)$, необходима малость h. Однако при малых h дисперсия величины (5) возрастает, так как уменьшается среднее число точек, попадающих в S_{xh}. При вы-

[1]) Теорема 1 данной главы. — *Прим. перев.*

боре h необходимо уравновесить эти два эффекта, в связи с чем возникают интересные теоретические задачи.

В гл. 2 представлены общие аппроксимационные теоремы типа (4). В гл. 3, 5 и 6 параллельно рассматриваются две оценки — *ядерная* и *гистограммная*. В частности, приводятся условия на h, необходимые и достаточные для всех типов сходимости величины J_n (гл. 3), результаты о скорости сходимости средней ошибки $E(J_n)$, среди которых выделяется универсальная нижняя граница для $\liminf_{n\to\infty} n^{2/5}E(J_n)$ при любых ядерных оценках и любых плотностях f (гл. 5), и теоремы о сходимости ядерных оценок, в которых h выбирается как функция от данных (гл. 6).

В гл. 4 показано, что при любых оценках плотности можно добиться стремления величины $E(J_n)$ к 0 с любой заданной скоростью, если только выбирать f из соответствующих классов плотностей, таких, как класс всех бесконечно дифференцируемых плотностей или класс всех плотностей с носителем $[0, 1]$, ограниченных величиной 2. Таким образом, ни для какой сколь угодно сложной оценки величина $E(J_n)$ не может убывать с некоторой заданной скоростью при любых f. Для изучения скоростей сходимости величины $E(J_n)$ следует наложить некоторые условия на f, причем из результатов гл. 4 ясно, что не достаточно только условий на хвосты или только условий гладкости. В остальных главах поясняются эти основные теоретические положения.

Главы 7—12 можно читать в любом порядке; их уровни сложности различны и зависят от круга читателей, которому каждая из них адресована. В гл. 9 обсуждается преобразованная ядерная оценка. В гл. 12 рассматривается несколько оценок, связанных с разложениями в ортогональные ряды. Другие оценки описаны в гл. 7. Главы 8, 10 и 11 в большей степени ориентированы на приложения. Так, в гл. 8 затрагивается проблема использования f_n при численном моделировании.

В гл. 10 показано, что любая оценка плотности имеет аналог в задаче классификации и что вероятность ошибки классификации тесно связана с величиной J_n. Наконец, в гл. 11 среди прочего рассмотрены некоторые приложения к теории обнаружения.

Далеко не все направления удалось осветить в книге и многие вопросы так и остались без ответа. Наиболее существенным недостатком является отсутствие теории асимптотического распределения величины J_n, закона повторного логарифма для J_n, результатов о скоростях сходимости $E(J_n)$ в многомерном случае, методов оценивания J_n и доверительных интервалов для J_n.

Каждая глава снабжена собственным списком литературы. Дополнительные ссылки о других свойствах рассмотренных здесь оценок (например, об их поведении в случае, когда $X_1, \ldots$

..., X_n зависимы, об их свойствах в метриках L_p при $1 < p \leqslant \infty$, о законах повторного логарифма для них) или о других оценках можно найти в монографиях Wertz (1978), Tapia, Thompson (1978), Надарая (1983), Prakasa Rao (1983) и в обзорах и библиографиях Cover (1972), Fryer (1977), Földes, Révész (1974), Leonard (1978), Révész (1972), Tarter, Kronmal (1976), Wegman (1972a, 1972b), Wertz, Schneider (1979) и Bean, Tsokos (1980).

В каждой главе формулы нумеруются (1), (2), (3) и т. д. а теоремы имеют номера 1, 2, 3 и т. д. Если в некоторой главе есть ссылка на теорему 3, то имеется в виду теорема 3 той же главы. В противном случае добавляется номер главы, например теорема 2.3. Главы связаны друг с другом следующим образом: гл. 2 необходима для чтения гл. 3, 4 и 6; гл. 3 содержит предварительные сведения для гл. 7 и 10; гл. 4 нужна для чтения гл. 5; в свою очередь гл. 5 содержит предварительные сведения для гл. 8 и 9.

ЛИТЕРАТУРА

Bean S. J., Tsokos C. P.
(1980) Developments in nonparametric density estimation, International Statistical Review, 48, pp. 267—287.
Cover T. M.
(1972) A hierarchy of probability density function estimates, in Frontiers of Pattern Recognition, Academic Press, New York, pp. 83—98.
Földes A., Révész P.
(1974) A general method for density estimation, Studia Scientiarium Mathematicarum Hungarica, 9, pp. 81—92.
Fryer M. J.
(1977) A review of some nonparametric methods of density estimation, Journal of the Institute of Mathematics and Applications, 20, pp. 335—354.
Leonard T.
(1978) Density estimation, stochastic processes and prior information, Journal of the Royal Statistical Society, B40, pp. 113—146.
Prakasa Rao B. L. S.
(1983) Nonparametric Functional Estimation, Academic Press, New York.
Révész P.
(1972) On empirical density function, Periodica Mathematica Hungarica, 2, pp. 85—110.
Rosenblatt M.
(1956) Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, pp. 832—835.
Scheffé H.
(1947) A useful convergence theorem for probability distributions, Annals of Mathematical Statistics, 18, pp. 434—458.
Tapia R. A., Thompson J. R.
(1978) Nonparametric Probability Density Estimation, The Johns Hopkins University Press, Baltimore.
Tarter M. E., Kronmal R. A.
1976) An introduction to the implementation and theory of nonparametric density estimation, The American Statistician, 30, pp. 105—112.
Wegman E. J.
(1972a) Nonparametric probability estimation. I. A summary of available methods, Technometrics, 14, pp. 533—546.

(1972b) Nonparametric probability density estimation. II. A comparison of density estimation methods, Journal of Statictical Computation and Simulation, 1, pp. 225—245.

Wertz W.

(1978) Statistical Density Estimation. A Survey, Vandenhoeck and Ruprecht, Gottingen, Applied Statistics and Econometrics Series, 13.

Wertz W., Schneider B.

(1979) Statistical density estimation: A bibliography, International Statistical Review, 47, pp. 155—175.

Надарая Э. А.

(1983) Непараметрическое оценивание плотности вероятности и кривой регрессии. — Тбилиси: Изд-во Тбилисского гос. ун-та.

КОММЕНТАРИЙ ПЕРЕВОДЧИКА

В перечне нерешенных проблем здесь указана задача оценивания J_n и построения доверительных интервалов для J_n. Недавно достаточно полное решение этой задачи найдено Девроем (см. Devroye (1987*) и комментарий к гл. 3). Ответ на другой вопрос — об оптимальных скоростях сходимости для $E(J_n)$ в многомерном случае — можно извлечь, например, из общих результатов Birgé (1983*), конкретизируя их для определенных классов плотностей.

Дополнительные сведения об оценках плотности можно найти в книгах Тарасенко (1976*) и Silverman (1986*). В последней много внимания уделяется прикладным аспектам.

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Тарасенко Ф. П.

(1976*) Непараметрическая статистика. — Томск: Изд-во Томского гос. ун-та.

Birgé L.

(1983*) Approximation dans les espaces métriques et théorie de l'estimation//Zeitschrift für Wahrscheinlichkeitstheor. verw. Geb., 65, p. 181—237.

Devroye L.

(1987*) A course in density estimation. Boston e. a.: Birhäuser. Silverman B. W.

(1986*) Density estimation for statistics and data analysis. London: Chapman and Hall.

Глава 2

Дифференцирование интегралов

Наиболее важным инструментом в нашем исследовании является теорема Лебега о плотностях (1.4). Все аналогичные ей результаты собраны в данной главе. За доказательствами и блестящими обсуждениями мы отсылаем читателя к гл. 7 и 9 книги Wheeden, Zygmund (1977) и гл. 1—3 книги Guzman (1975). См. также Shapiro (1969), Stein (1970), Hayes, Pauc (1970) и Guzman (1981). В этой главе λ — мера Лебега на R^d, K — борелевская функция на R^d, f — плотность на R^d, h — положительное число, $K_h(x) = (1/h^d) K(x/h)$ и «$*$» — оператор свертки; например, если $K \in L_1(\lambda)$, то

$$f * K(x) = \int f(y) K(x-y)\, dy = \int K(y) f(x-y)\, dy.$$

Теорема 1. *Для любых функций* $f, g \in L_1(\lambda)$ *имеем* $\int |f*g| \leqslant \int |f| \int |g|$ (*неравенство Юнга*). *Для любых* $f, K \in L_1(\lambda)$, *таких, что* $\int K = 1$, *имеем*

$$\lim_{h \downarrow 0} \int |f * K_h - f| = 0.$$

Доказательство. Первое неравенство получается с помощью перемены порядка интегрирования (допустимой в силу неотрицательности подынтегральных функций):

$$\int \left| \int f(y) g(x-y)\, dy \right| dx \leqslant \int\int |f(y)|\, |g(x-y)|\, dy\, dx =$$

$$= \int |f(y)| \int |g(x-y)|\, dx\, dy = \int |g| \int |f|.$$

Доказательство второго утверждения теоремы 1 проведем сначала для непрерывных f с компактным носителем. Пусть $\omega(t)$ — модуль непрерывности f, $\omega(t) = \sup_{\|x-y\| \leqslant t} |f(x) - f(y)|$, и пусть M — некоторое большое число. Представим K в виде $K = K' + K''$, где $K' = K I_{[\|x\| \leqslant M]}$, $K'' = K I_{[\|x\| > M]}$. Ясно, что

$$\int |f * K_h - f| \leqslant \int \left| f * K'_h - f \left(\int K'_h \right) \right| + \int |f * K''_h| + \int f \int |K''_h|. \quad (1)$$

Последние два слагаемых в (1) не превосходят величины $2\int |K_h''| = 2\int |K''|$, которую можно сделать сколь угодно малой за счет выбора M. Первое слагаемое в правой части (1) есть $o(1)$, поскольку оно равно

$$\int_A \left| f * K_h' - f\left(\int K_h'\right)\right| \leqslant \iint_A |f(x-y) - f(x)|\,|K_h'(y)|\,dy\,dx \leqslant$$

$$\leqslant \omega(Mh)\iint_A |K_h'(y)|\,dy\,dx \leqslant$$

$$\leqslant \omega(Mh)\,\lambda(A)\int |K| = o(1),$$

где A — некоторый большой компакт. Для любых f и любых непрерывных g с компактным носителем имеем

$$\int |f * K_h - f| \leqslant \int |f-g| * |K_h| + \int |f-g| + \int |g * K_h - g| \leqslant$$

$$\leqslant \left(\int |K| + 1\right)\int |f-g| + o(1).$$

Это выражение можно сделать сколь угодно малым за счет выбора g.

Теорема 2 (теорема Лебега о плотностях). *Пусть $\mathscr{B}$ — класс всех борелевских множеств на R^d, обладающих следующим свойством*

$$\sup_{B\in\mathscr{B}} \frac{\lambda\,(\text{наименьший куб с центром в } 0, \text{ содержащий } B)}{\lambda(B)} < \infty.$$

Тогда для любой последовательности множеств B_k из $\mathscr{B}$, такой, что $\lambda(B_k) \to 0$, имеем

$$\lim_{k\to\infty} \int_{x+B_k} \frac{|f(y) - f(x)|}{\lambda(B_k)}\,dy = 0 \tag{2}$$

при почти всех x, и потому

$$\lim_{k\to\infty} \int_{x+B_k} \frac{f(y)\,dy}{\lambda(B_k)} = f(x) \tag{3}$$

при почти всех x. Точки, в которых справедливы соотношения (2) и (3), называются точками Лебега функции f. Множество точек Лебега зависит только от f.

Доказательство теоремы 2 см. в книге Wheeden, Zygmund (1977, с. 108—109). Заметим, что в качестве $\mathscr{B}$ можно взять мно-

ЭБИБЛИОТЕКА

жество всех шаров с центром в начале координат (тогда получается классический вариант теоремы Лебега о плотностях) или класс всех множеств вида aA, где $a > 0$ и A — фиксированный компакт в R^d, но нельзя взять класс всех прямоугольников, содержащих начало координат.

Теорема 3. *Пусть* $K \in L_1(\lambda)$ *и* $\int K = 1$. *Предположим, что* K *имеет интегрируемую радиальную мажоранту* $\psi \in L_1(\lambda)$ $(\psi(x) = \sup_{\|y\| \geqslant \|x\|} |K(y)|)$. *Тогда*

$$f * K_h \to f$$

при $h \downarrow 0$ *для почти всех* x.

Теорема 3 принадлежит Stein (1970, с. 77—78 по русскому изданию). В ней достаточно, например, предполагать, что функция K ограниченна, принадлежит $L_1(\lambda)$, $\int K = 1$ и $K(x) \leqslant a/\|x\|^{d+\varepsilon}$ при некоторых $\varepsilon > 0$, $a > 0$. В такой форме теорему 3 можно найти в работе Wheeden, Zygmund (1977, с. 152—153). Конечно, в случае ограниченных K с компактным носителем теорема 3 легко следует из теоремы 2.

Обратной к теореме 1 является

Теорема 4. *Пусть* K *— плотность на* R^d. *Тогда* $\int |f * K_h - f| > 0$ *при любых* $h > 0$. *Если* $h = h_n$ *— некоторая числовая последовательность, то из соотношения* $\lim_{n\to\infty} \int |f * K_h - f| = 0$ *следует, что* $h \to 0$.

Доказательство. Пусть φ и ψ — характеристические функции плотностей f и K соответственно. Тогда характеристическая функция $f * K_h$ равна $\psi(ht)\,\varphi(t)$, $t \in R^d$. Очевидно, из равенства $\int |f * K_h - f| = 0$ следует, что $f = f * K_h$ для почти всех x и, значит, $\varphi(t) = \varphi(t)\,\psi(ht)$ для всех $t \in R^d$. При $\varphi(t) \neq 0$, т. е. по крайней мере в некоторой окрестности нуля, имеем $\psi(ht) = 1$. Но так как $h \neq 0$, отсюда следует, что ψ не может быть характеристической функцией плотности распределения на R^d. Полученное противоречие доказывает первое утверждение теоремы 4.

Для доказательства второго утверждения теоремы 4 предположим сначала, что $\lim h = \infty$. В силу леммы Фату соотношение $\int |f * K_h - f| \to 0$ влечет, что $\liminf |f * K_h - f| = 0$ для почти всех x. Но так как $f * K_h \to 0$ для почти всех x, то обязательно $f = 0$ для почти всех x, что невозможно. Предположим теперь,

что $\lim h = c \in (0, \infty)$. Ясно, что $\int |f * K_h - f| \geqslant \int |f * K_c - f| - \int |f * K_c - f * K_h|$. Для завершения доказательства достаточно показать, что $\int |f * K_c - f * K_h| \to 0$, и, учитывая первое утверждение теоремы, прийти к противоречию. Пусть K' — некоторая непрерывная функция с компактным носителем. В силу теоремы 1 и теоремы Лебега о мажорируемой сходимости

$$\int |f * K_c - f * K_h| \leqslant \int |K_c - K_h| \leqslant$$

$$\leqslant \int |K_c - K'_c| + \int |K'_c - K'_h| + \int |K'_h - K_h| =$$

$$= 2 \int |K - K'| + \int |K'_c - K'_h| = 2 \int |K - K'| + o(1). \quad (4)$$

Последнее выражение в цепочке неравенств (4) можно сделать малым, выбирая K' близким к K в $L_1(\lambda)$.

Для изучения гистограммных оценок нам понадобятся некоторые теоремы о сходимости мартингалов. Рассмотрим последовательность разбиений $\mathscr{P}_n = \{A_{nj},\ j = 1,\ 2, \ldots\}$, $n \geqslant 1$, где $\lambda(A_{nj}) \in (0, \infty)$ при любых n, j и все A_{nj} — борелевские множества в R^d. Эта последовательность называется *вложенной*, если при любом n разбиение $\mathscr{P}_{n+1}$ является измельчением разбиения $\mathscr{P}_n$. Последовательность разбиений называется *кубической*, если существуют положительные постоянные $a_1, \ldots, a_d$ и последовательность положительных чисел $h = h_n$, такие, что каждое из множеств A_{nj} имеет вид $\prod_{i=1}^{d} [a_i k_i h,\ a_i (k_i + 1) h)$, где $k_1, \ldots, k_d$ — целые числа. Пусть в дальнейшем $\mathscr{B}_n = \sigma(\mathscr{P}_n)$ есть σ-алгебра, порожденная разбиением $\mathscr{P}_n$, $\mathscr{B}'_n = \sigma\left(\bigcup_{m \geqslant n} \mathscr{P}_m\right)$ и $\mathscr{B}$ — класс всех борелевских множеств на R^d. Везде будет предполагаться, что

$$\mathscr{B} = \bigcap_{n=1}^{\infty} \mathscr{B}'_n. \quad (5)$$

Условие (5) означает, что последовательность разбиений должна быть достаточно богатой.

Рассмотрим функции

$$g_n(x) = \int_{A_{nj}} f/\lambda(A_{nj}), \quad x \in A_{nj}. \quad (6)$$

Для последовательностей разбиений, удовлетворяющих (5), Abou-Jaoude (1976) доказал следующий сильный аналог теорем 1 и 4.

Теорема 5 (Abou-Jaoude (1976)). *Сходимость*

$$\int |g_n - f| \to 0$$

при любых плотностях f имеет место тогда и только тогда, когда для любого множества $A \in \mathcal{B}$, такого, что $0 < \lambda(A) < \infty$, и любого $\varepsilon > 0$ существует n_0, такое, что при каждом $n \geqslant n_0$ найдется множество A_n из $\mathcal{B}_n$, удовлетворяющее условию $\lambda(A \Delta A_n) < \varepsilon$ (здесь Δ — символ симметрической разности).

Теорема 5 доказана в работе Abou-Jaoude (1976, с. 216—219).

Теорема 6. *В случае кубической последовательности разбиений условие* (5) *и сходимость $\int |g_n - f| \to 0$ при любых плотностях f имеют место тогда и только тогда, когда* $\lim h = 0$.

Доказательство. Прежде всего ясно, что условие $\lim h = 0$ необходимо и достаточно для (5). В частности, достаточность следует из того, что в пересечении $\bigcap_{n=1}^{\infty} \mathcal{B}_n$ содержатся все множества вида $\prod_{i=1}^{d} (-\infty, x_i)$ при любых $x = (x_1, \ldots, x_d) \in R^d$ и что эти множества порождают борелевскую σ-алгебру.

Итак, остается просто проверить условия теоремы 5. Так как λ — регулярная мера на R^d (т. е. все борелевские множества являются пределами стягивающихся последовательностей открытых множеств), то будем рассматривать только ограниченные открытые множества O. Но каждое множество O является счетным объединением прямоугольников. Следовательно, для любого $\varepsilon > 0$ существует конечное множество прямоугольников $R_1, \ldots, R_N$, такое, что $\lambda\left(O - \bigcup_{i=1}^{N} R_i\right) < \varepsilon$. Таким образом, достаточно установить, что условия теоремы 5 выполнены для конечного числа прямоугольников, т. е. по существу, для одного прямоугольника. Но для одного прямоугольника эти условия удовлетворяются тривиальным образом.

Отметим здесь, что для кубических последовательностей разбиений, вложенных или нет, из теоремы 2 легко следует, что $g_n \to f$ для почти всех x.

Из предыдущего видно, что в теоремах о поточечной сходимости обычно требуется больше условий, чем в теоремах об интегральной сходимости. Пример тому дает сравнение теорем 1 и 3. Это вызвано тем, что поточечная сходимость является строго более сильным свойством, чем сходимость в L_1.

Теорема 7 (Scheffé (1947)). *Пусть f_n — последовательность плотностей на R^d, сходящаяся почти всюду к плотности f. Тогда* $\int |f_n - f| \to 0$.

Доказательство. Из теоремы 1.1 и теоремы Лебега о мажорируемой сходимости следует, что $\int |f_n - f| = 2 \int\limits_{f>f_n} (f - f_n) \to 0$.

Теорема 8 (Glick (1974)). *Пусть f_n — оценка плотности на R^d и f — плотность на R^d. Если $f_n \to f$ по вероятности при $n \to \infty$ для почти всех x, то $\int |f_n - f| \to 0$ по вероятности (и, следовательно, в среднем) при $n \to \infty$. Если $f_n \to f$ почти наверное при $n \to \infty$ для почти всех x, то $\int |f_n - f| \to 0$ почти наверное при $n \to \infty$.*

Доказательство. Обозначим $(\cdot)_+$ положительную часть функции. По условию имеем $(f - f_n)_+ \to 0$ по вероятности при почти всех x. Так как $(f - f_n)_+ \leqslant f$, то в силу теоремы Лебега о мажорируемой сходимости $E((f - f_n)_+) \to 0$ при почти всех x. Применяя эту теорему еще раз, получим $E\left(\int |f_n - f|\right) = E\left(2 \int (f - f_n)_+\right) = 2 \int E((f - f_n)_+) \to 0$.

При доказательстве второго утверждения теоремы обозначим через $(\Omega, \mathcal{F}, P)$ вероятностное пространство последовательностей $X_1, X_2, \ldots$ и через ω — соответствующие элементарные исходы. В силу теоремы Фубини соотношение

$$P(\omega: f_n(x) \not\to f(x)) = 0$$

выполнено для почти всех x по мере λ тогда и только тогда, когда множество $\{(\omega, x): f_n(x) \not\to f(x)\}$ имеет $P \times \lambda$-меру нуль, и тогда и только тогда, когда

$$\lambda(x: f_n(x) \not\to f(x)) = 0$$

для почти всех ω по мере P. Пусть Ω' — множество исходов ω, для которых выполнено последнее равенство. По теореме Лебега о мажорируемой сходимости $\int |f_n - f| \to 0$ для всех $\omega \in \Omega'$. Так как $P(\Omega') = 1$, то теорема доказана.

ЛИТЕРАТУРА

Abou-Jaoude S.
(1976) Conditions nécessaires et suffisantes de convergence L_1 en probabilité de l'histogramme pour une densité, Annales de l'Institut Henri Poincaré, 12, pp. 213—231.

de Guzman M.
(1975) Differentiation of Integrals in R^n, Lecture Notes in Mathematics # 481, Springer-Verlag, Berlin. (Имеется русский перевод: Гусман М. Дифференцирование интегралов в R^n. — М.: Мир, 1978.)
(1981) Real Variable Methods in Fourier Analysis, North-Holland, Amsterdam.
Devroye L.
(1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, Annals of Statistics, 11, pp. 896—904.
Glick N.
(1974) Consistency conditions for probability estimators and integrals of density estimators, Utilitas Mathematica, 6, pp. 61—74.
Hayes C. A., Pauc C. Y.
(1970) Derivation and Martingales, Springer-Verlag, New York.
Neveu J.
(1975) Discrete-Parameter Martingales, North-Holland, Amsterdam.
Scheffé H.
(1947) A useful convergence theorem for probability distributions, Annals of Mathematical Statistics, 18, pp. 434—458.
Shapiro H. S.
(1969) Smoothing and Approximation of Functions, Van Nostrand Reinhold, New York.
Stein E. M.
(1970) Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, New Jersey. (Имеется русский перевод: Стейн И. Сингулярные интегралы и дифференциальные свойства функций. — М.: Мир, 1973.)
Wheeden R. L., Zygmund A.
(1977) Measure and Integral, Marcel Dekker, New York.

Глава 3

Состоятельность

1. Ядерная оценка

Ядерная оценка (Parzen (1962), Rosenblatt (1956), Cacoullos (1966)) определяется равенством

$$f_n(x) = (nh^d)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

где $h = h_n$ — последовательность положительных чисел и K — борелевская функция (ядро), удовлетворяющая условиям $K \geqslant 0$, $\int K = 1$. Основной результат этого параграфа состоит в том, что для ядерной оценки все типы сходимости J_n к 0 эквивалентны. В теореме 1 ниже утверждается, что либо $J_n \to 0$ вполне [1]) при любых f, либо J_n не стремится к 0 по вероятности ни при каких f. Промежуточного варианта нет. Слабый аналог теоремы 1 для гистограммных оценок приведен в § 3. Теорема 1 впервые была опубликована Devroye (1983), но некоторые ее ключевые идеи восходят к Abou-Jaoude (1977).

Теорема 1. *Пусть K — неотрицательная борелевская функция на R^d, такая, что $\int K = 1$. Тогда следующие утверждения эквивалентны:*

(i) $J_n \to 0$ *по вероятности при* $n \to \infty$ *для некоторой плотности* f;
(ii) $J_n \to 0$ *по вероятности при* $n \to \infty$ *для любых* f;
(iii) $J_n \to 0$ *почти наверное при* $n \to \infty$ *для любых* f;
(iv) $J_n \to 0$ *экспоненциально при* $n \to \infty$ *(т. е. для любого* $\varepsilon > 0$ *существуют* r, $n_0 > 0$, *такие, что* $P(J_n \geqslant \varepsilon) \leqslant e^{-rn}$, $n \geqslant n_0$*) для любых* f;

(v) $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh^d = \infty$.

---

[1]) То есть $\sum_{n=1}^{\infty} P(J_n > \varepsilon) < \infty$ при любых $\varepsilon > 0$. — *Прим. перев.*

В (iv) число r можно выбрать не зависящим от f. Кроме того, (iv) следует из (v), если K — просто абсолютно интегрируемая функция, такая, что $\int K = 1$.

Замечание 1. Мы покажем, что из (v) следует неравенство $P(J_n \geqslant \varepsilon) \leqslant e^{-rn\varepsilon^2}$ при любых $\varepsilon \in (0, 1)$ и любых $n \geqslant n_0$, где n_0 зависит от f и ε. Для заданной плотности f существуют функции $h_0(\varepsilon)$ и $c_0(\varepsilon)$, такие, что эта экспоненциальная граница верна при $(c_0(\varepsilon)/n)^{1/d} < h < h_0(\varepsilon)$. Таким образом, при заданном ε экспоненциальное неравенство остается в силе, даже если величина h является постоянной.

2. Доказательство теоремы 1

Попытаемся выделить ключевые факты, необходимые для доказательства теоремы 1. Сформулируем их в виде ряда лемм, представляющих самостоятельный интерес. Нам также понадобятся теоремы 2.1, 2.2 и 2.4. Включение (i) $\Rightarrow$ (v) будет установлено в лемме 3, включение (v) $\Rightarrow$ (iv) — в лемме 2. Так как, очевидно, (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i), тем самым будет доказана теорема 1.

В этом параграфе будет использоваться обозначение

$$g_h(x) = E(f_n(x)) = \int h^{-d} K\left(\frac{x-y}{h}\right) f(y)\, dy. \tag{1}$$

Лемма 1 (неравенство для полиномиального распределения). *Пусть* $(X_1, \ldots, X_k)$ — *случайный вектор, имеющий полиномиальное распределение с параметрами* $(n, p_1, \ldots, p_k)$. *Для* $\varepsilon \in (0, 1)$ *и любых* k, *таких, что* $k/n \leqslant \varepsilon^2/20$, *имеем*

$$P\left(\sum_{i=1}^{k} |X_i - E(X_i)| \geqslant n\varepsilon\right) \leqslant 3e^{-n\varepsilon^2/25}.$$

Доказательство основано на пуассонизации. Пусть N — пуассоновская с параметром n случайная величина и $U_1, U_2, \ldots$ — не зависящая от N последовательность независимых случайных величин со значениями $1, \ldots, k$, распределенных по закону: $P(U_1 = i) = p_i$, $1 \leqslant i \leqslant k$. Пусть X_i — число появлений значения i среди $U_1, \ldots, U_n$ и X_i' — число появлений значения i среди $U_1, \ldots, U_N$. Ясно, что $X_1', \ldots, X_k'$ — независимые пуассоновские случайные величины со средними $np_1, \ldots, np_k$ и что $X_1, \ldots, X_k$ — случайный вектор, имеющий полиномиальное распределение с параметрами $(n, p_1, \ldots, p_k)$. Имеем

$$\sum_{i=1}^{k} \frac{1}{n} |X_i - np_i| \leqslant \sum_{i=1}^{k} \frac{1}{n} |X_i - X_i'| + \sum_{i=1}^{k} \frac{1}{n} |X_i' - np_i|. \tag{2}$$

Далее, для пуассоновской с параметром λ случайной величины U имеем $E\,(e^{t\,|\,U-\lambda\,|}) \leqslant E\,(e^{t\,(U-\lambda)} + e^{t\,(\lambda-U)}) = e^{\lambda\,(e^t-1)-t\lambda} + e^{\lambda\,(e^{-t}-1)+t\lambda} \leqslant 2e^{\lambda\,(e^t-1-t)}$ при $t > 0$, так как $e^{-t} + t \leqslant e^t - t$. Таким образом,

$$P(|\,U-\lambda\,| \geqslant \lambda\varepsilon) \leqslant E\,(e^{t\,|\,U-\lambda\,|-t\lambda\varepsilon}) \leqslant 2e^{-\lambda t\varepsilon}e^{\lambda\,(e^t-1-t)} =$$
$$= 2e^{\lambda\,(\varepsilon-(1+\varepsilon)\ln(1+\varepsilon))} \leqslant 2e^{-\lambda\varepsilon^2/2\,(1+\varepsilon)} \leqslant 2e^{-\lambda\varepsilon^2/4}, \quad (3)$$

где мы положили $t = \ln(1+\varepsilon)$. Повторяя предыдущее рассуждение, получим

$$P\left(\sum_{i=1}^{k}\frac{1}{n}\,|\,X_i - np_i\,| \geqslant \varepsilon\right) \leqslant$$

$$\leqslant P(|\,N-n\,| \geqslant 2\varepsilon n/5) + P\left(\sum_{i=1}^{k}|\,X_i' - np_i\,| \geqslant 3\varepsilon n/5\right) \leqslant$$

$$\leqslant 2e^{-n\,(2\varepsilon/5)^2/4} + e^{-tn\,(3\varepsilon/5)}\prod_{i=1}^{k}\left(2e^{np_i\,(e^t-1-t)}\right) \leqslant \qquad \text{(в силу (3))}$$

$$\leqslant 2e^{-n\varepsilon^2/25} + 2^k e^{n\,(e^t-1-t-3\varepsilon t/5)} \leqslant 2e^{-n\varepsilon^2/25} + e^{k-n\,(3\varepsilon/5)^2/4} \leqslant$$
$$\text{(при } t = \ln(1+3\varepsilon/5))$$
$$\leqslant 3e^{-n\varepsilon^2/25} \qquad \text{(если } k \leqslant n\varepsilon^2/20). \quad (4)$$

Замечание. После того как книга была сдана в печать, мы заметили, что

$$\sum_{i=1}^{k}\left|\frac{X_i}{n} - E\left(\frac{X_i}{n}\right)\right| = 2\sup_{A}\left|\sum_{i\in A}\left(\frac{X_i}{n} - E\left(\frac{X_i}{n}\right)\right)\right|,$$

где A — множество всех 2^k подмножеств набора $\{1, \ldots, k\}$. Следовательно, применяя непосредственно неравенство Бонферрони и неравенство Хёфдинга (Hoeffding (1963)), получим, что верхнюю границу леммы 1 можно заменить на $2^{k+1}e^{-n\varepsilon^2/2}$ при любых k, n и любых $\varepsilon > 0$. Похожее неравенство для полиномиального распределения доказано несколько иным способом в работе Bretagnolle, Huber (1978).

Лемма 2. *Для любой плотности f на R^d и любой абсолютно интегрируемой функции K, такой, что $\int K(x)\,dx = 1$, утверждение* (iv) *имеет место, если*

$$\lim_{n\to\infty} h = 0, \quad \lim_{n\to\infty} nh^d = \infty.$$

Доказательство. Пусть g_h — функция, определенная в (1). В силу теоремы 2 достаточно показать, что интеграл $\int |f_n(x) - g_h(x)|\,dx \to 0$ экспоненциально. Заметим, что

$$f_n(x) = h^{-d} \int K\left(\frac{x-y}{h}\right) \mu_n(dy),$$

где μ_n — эмпирическая вероятностная мера для $X_1, \ldots, X_n$. Для заданного $\varepsilon > 0$ найдем конечные постоянные M, L, N, $a_1, \ldots, a_N$ и непересекающиеся конечные прямоугольники $A_1, \ldots, A_N$ из R^d, такие, что функция

$$K^*(x) = \sum_{i=1}^{N} a_i I_{A_i}(x)$$

удовлетворяет следующим условиям: $|K^*| \leqslant M$, $K^* = 0$ вне $[-L, L]^d$ и $\int |K(x) - K^*(x)|\,dx < \varepsilon$. Определим g_h^* и f_n^* так же, как g_h и f_n, подставляя K^* вместо K. Тогда

$$\begin{aligned}
\int |f_n(x) - g_h(x)|\,dx &\leqslant \int |f_n(x) - f_n^*(x)|\,dx + \\
&+ \int |f_n^*(x) - g_h^*(x)|\,dx + \int |g_h^*(x) - g_h(x)|\,dx \leqslant \\
&\leqslant \int h^{-d} \int \left| K^*\left(\frac{x-y}{h}\right) - K\left(\frac{x-y}{h}\right)\right| f(y)\,dy\,dx + \\
&+ \int h^{-d} \int \left| K^*\left(\frac{x-y}{h}\right) - K\left(\frac{x-y}{h}\right)\right| \mu_n(dy)\,dx + \\
&+ \int |f_n^*(x) - g_h^*(x)|\,dx \leqslant \\
&\leqslant 2\varepsilon + \int |f_n^*(x) - g_h^*(x)|\,dx,
\end{aligned}$$

где дважды произведена замена порядка интегрирования. Но

$$\begin{aligned}
\int |f_n^*(x) - g_h^*(x)|\,dx &\leqslant \\
&\leqslant \sum_{i=1}^{N} |a_i| \int \left| h^{-d} \int_{x+hA_i} f(y)\,dy - h^{-d} \int_{x+hA_i} \mu_n(dy)\right| dx \leqslant \\
&\leqslant Mh^{-d} \sum_{i=1}^{N} \int |\mu(x + hA_i) - \mu_n(x + hA_i)|\,dx,
\end{aligned}$$

где μ — вероятностная мера, соответствующая f. Лемма 2 будет доказана, если показать, что

$$h^{-d} \int |\mu(x + hA) - \mu_n(x + hA)|\,dx \to 0$$

экспоненциально при $n \to \infty$ для любых конечных прямоугольников A из R^d. Выберем произвольный прямоугольник A и произвольное $\varepsilon > 0$. Рассмотрим разбиение пространства R^d на множества B, представляющие собой d-кратные произведения интервалов вида $[(i-1)\,h/N,\ ih/N)$, где i — целое и N — новая постоянная, которая будет выбрана далее. Обозначим это разбиение через Ψ. Пусть

$$A = \prod_{i=1}^{d} [x_i,\ x_i + a_i), \quad \min_i a_i \geqslant 2/N$$

и

$$A^* = \prod_{i=1}^{d} [x_i + 1/N,\quad x_i + a_i - 1/N).$$

Положим

$$C_x = x + hA - \bigcup_{\substack{B \in \Psi \\ B \subseteq x+hA}} B \subseteq x + h(A - A^*) = C_x^*.$$

Ясно, что

$$\int |\mu(x + hA) - \mu_n(x + hA)|\, dx \leqslant$$

$$\leqslant \int \sum_{\substack{B \in \Psi, \\ B \subseteq x+hA}} |\mu(B) - \mu_n(B)|\, dx + \int (\mu(C_x^*) + \mu_n(C_x^*))\, dx. \qquad (5)$$

Последнее слагаемое в (5) равно

$$2\lambda(h(A - A^*)) = 2h^d \lambda(A - A^*) =$$

$$= 2h^d \left(\prod_{i=1}^{d} a_i - \prod_{i=1}^{d} (a_i - 2/N) \right) =$$

$$= 2h^d \lambda(A) \left(1 - \prod_{i=1}^{d} \left(1 - \frac{2}{Na_i} \right) \right) \leqslant$$

$$\leqslant 4h^d \lambda(A) \sum_{i=1}^{d} a_i^{-1}/N \leqslant \varepsilon h^d$$

при некотором выборе N. Здесь мы воспользовались тем, что для любого множества C и любой вероятностной меры ν на борелев-

ских множествах из R^d выполнено равенство $\int \nu(x+hC)\,dx = \lambda(hC)$. Первое слагаемое в (5) ограничено сверху величиной

$$\sum_{\substack{B\in\Psi,\\ B\cap S_{0R}\neq\varnothing}} |\mu_n(B)-\mu(B)| \int\limits_{B\subseteq x+hA} dx + \int\limits_{B\subseteq x+hA} dx\,(\mu_n(S_{0R}^c)-\mu(S_{0R}^c)+2\mu(S_{0R}^c)), \quad (6)$$

где $R>0$ — произвольная конечная постоянная. Здесь символом $(\cdot)^c$ обозначается дополнение множества. Ясно, что $h^{-d}\int\limits_{B\subseteq x+hA} dx \leqslant \lambda(A)$, и можно выбрать R так, что $\mu(S_{0R}^c)<\varepsilon$. Кроме того,

$$P(\mu_n(S_{0R}^c)-\mu(S_{0R}^c)>\varepsilon) \leqslant e^{-2n\varepsilon^2}$$

в силу неравенства Хёфдинга для биномиальных случайных величин (Hoeffding (1963)). Наконец, поскольку существует не более чем $(2RN/h+2)^d = o(n)$ множеств $B\in\Psi$, таких, что $B\cap S_{0R}\neq\varnothing$, то в силу леммы 3

$$P\left(\sum_{\substack{B\in\Psi,\\ B\cap S_{0R}\neq\varnothing}} |\mu_n(B)-\mu(B)|>\varepsilon\right) \leqslant 3e^{-n\varepsilon^2/25}$$

при всех достаточно больших n. Теперь соберем вместе полученные неравенства. Лемма 2 доказана.

Лемма 3. *Пусть K и f — плотности на R^d. Если $J_n\to 0$ по вероятности при $n\to\infty$, то $\lim\limits_{n\to\infty} h=0$ и $\lim\limits_{n\to\infty} nh^d=\infty$.*

Доказательство. Так как $J_n\leqslant 2$ при любых n, то $J_n\to 0$ по вероятности тогда и только тогда, когда $\lim\limits_{n\to\infty} J(F_n)=0$. Пусть g_h — функция, определенная в (1). Тогда

$$E(J_n)=E\left(\int |f_n(x)-f(x)|\,dx\right)\geqslant$$

$$\geqslant \int |E(f_n(x))-f(x)|\,dx = \int |g_h(x)-f(x)|\,dx.$$

Используя теорему 2.4, заключаем, что $\lim\limits_{n\to\infty} h=0$. Далее в доказательстве считается, что это условие выполнено. Докажем второе утверждение леммы. Заметим, что $E\left(\int |f_n(x)-g_h(x)|\,dx\right)\to$ $\to 0$ в силу теоремы 2.1. Пусть M — некоторое большое число и

$K^*(x) = K(x) I_{[K(x) \leqslant M]}$. Определим f_n^* и g_h^* так же, как f_n и g_h, подставляя K^* вместо K. В силу теоремы 2.1

$$\int |f_n(x) - g_h(x)|\,dx \geqslant \int |f_n^*(x) - g_h^*(x)|\,dx -$$

$$- \int |f_n(x) - f_n^*(x)|\,dx - \int |g_h(x) - g_h^*(x)|\,dx =$$

$$= \int |f_n^*(x) - g_h^*(x)|\,dx - 2\int |K(x) - K^*(x)|\,dx. \quad (7)$$

Введем следующие дополнительные обозначения: L — еще одно большое число, A — событие, состоящее в том, что ни одно из значений X_i, $1 \leqslant i \leqslant n$, не принадлежит $S_{x,\,hL}$, $K' = K^* I_{S_{0L}}$, $K'' = K^* - K'$, и f_n' и f_n'' — функции, которые определяются так же, как f_n, при замене K на K' и на K''. Ясно, что

$$\int E(|f_n^*(x) - g_h^*(x)|)\,dx \geqslant \int E(|f_n^*(x) - g_h^*(x)|\, I_A)\,dx \geqslant$$

$$\geqslant \int g_h^*(x) P(A)\,dx - \int E(f_n(x) I_A)\,dx = U_n - V_n. \quad (8)$$

Нам понадобятся следующие факты, являющиеся следствиями теорем 2.2 и 2.3: $g_h^*(x) \to f(x) \int K^*(x)\,dx$ при почти всех x и ограниченных K^* с компактным носителем; $\mu(S_{y+hz,\,hL})/\lambda(S_{y+hz,\,hL}) \to f(y)$ при всех $z \in R^d$ и почти всех $y \in R^d$. Предположим, что $\lim\limits_{n\to\infty} nh^d = s \in [0, \infty)$ и C — объем шара S_{01}. По лемме Фату

$$\liminf_{n\to\infty} U_n \geqslant \int \liminf_{n\to\infty} g_h'(x) \liminf_{n\to\infty} P(A)\,dx =$$

$$= \int f(x) \liminf_{n\to\infty} (1 - \mu(S_{x,\,hL}))^n\,dx \int K'(z)\,dz \geqslant$$

$$\geqslant \int f(x) \exp\left(-\limsup_{n\to\infty} \left(\frac{n\mu(S_{x,\,hL})}{1 - \mu(S_{x,\,hL})}\right)\right) dx \int K'(z)\,dz =$$

$$= \int f(x) \exp\left(-sCL^d f(x)\right) dx \int\limits_{S_{0L}} K^*(z)\,dz. \quad (9)$$

Кроме того,

$$V_n \leqslant \int E\left(\frac{1}{n}\sum_{i=1}^{n} h^{-d} K''\left(\frac{x - X_i}{h}\right) I_A\right) dx =$$

$$= \int\int h^{-d} K''\left(\frac{x-y}{h}\right) I_{y \notin S_{x,\,hL}} f(y)\,dy\,(1 - \mu(S_{x,\,hL}))^{n-1}\,dx =$$

$$= \int f(y) \int\limits_{x \notin S_{y,hL}} h^{-d} K''((x-y)/h)(1-\mu(S_{x,hL}))^{n-1}\, dx\, dy \leqslant$$

$$\leqslant \int f(y) \int\limits_{z \notin S_{0L}} K''(z) \exp(-(n-1)\mu(S_{x+hz,hL}))\, dz\, dy. \qquad (10)$$

Подынтегральное выражение во внутреннем интеграле в (10) ограничено интегрируемой функцией K''. Таким образом, в силу теоремы Лебега о мажорируемой сходимости и сделанного выше замечания получим

$$\limsup_{n\to\infty} V_n \leqslant \int f(y) \int\limits_{z \notin S_{0L}} K^*(z) \exp(-sCL^d f(y))\, dz\, dy =$$

$$= \int f(y) \exp(-sCL^d f(y))\, dy \int\limits_{z \notin S_{0L}} K^*(z)\, dz. \qquad (11)$$

Неравенства (7), (8), (9) и (11) вместе дают

$$\liminf_{n\to\infty} \int E(|f_n(x) - g_h(x)|)\, dx + 2\int |K(x) - K^*(x)|\, dx \geqslant$$

$$\geqslant \int f(x) \exp(-sCL^d f(x))\, dx \left(2 \int\limits_{S_{0L}} K^*(z)\, dz - 1\right). \qquad (12)$$

Так как M — произвольное число, то

$$\liminf_{n\to\infty} \int E(|f_n(x) - g_h(x)|)\, dx \geqslant \int f e^{-sCL^d f} \left(2 \int\limits_{S_{0L}} K - 1\right).$$

Выберем теперь в качестве L достаточно большое конечное число, такое, что $\int\limits_{S_{0L}} K > 1/2$. Тогда, для того чтобы правая часть последнего неравенства равнялась 0, необходимо $s = \infty$. Это противоречит предположению. Следовательно, никакая подпоследовательность последовательности nh^d не может стремиться к конечному пределу, и потому $\lim\limits_{n\to\infty} nh^d = \infty$.

3. Гистограммная оценка

Гистограммная оценка определяется последовательностью разбиений $\mathscr{P}_n = \{A_{nj},\ j = 1, 2, \ldots\}$, $n \geqslant 1$, где все A_{nj} — борелевские множества конечной меры Лебега. Будем считать, что последовательность разбиений столь богата, что

$$\mathscr{B} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{m=n}^{\infty} \mathscr{P}_m\right), \qquad (13)$$

где $\mathscr{B}$ — класс всех борелевских множеств и символ σ используется для обозначения σ-алгебры, порожденной классом множеств. Гистограммная оценка определяется соотношением

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{I_{[X_i \in A_{nj}]}}{\lambda(A_{nj})}, \quad x \in A_{nj},$$

и ее среднее значение равно

$$g_n(x) = E(f_n(x)) = \int_{A_{nj}} f/\lambda(A_{nj}), \quad x \in A_{nj}.$$

Abou-Jaoude (1976 a, c) доказал следующее утверждение.

Теорема 2. *Предположим, что последовательность разбиений $\mathscr{P}_n$ удовлетворяет условию* (13). *Тогда следующие утверждения эквивалентны*:

(i) $J_n \to 0$ *по вероятности при* $n \to \infty$ *для любых* f;

(ii) $J_n \to 0$ *почти наверное при* $n \to \infty$ *для любых* f;

(iii) $J_n \to 0$ *экспоненциально при* $n \to \infty$ (*см. теорему* 1) *для любых* f (*как и в теореме* 1, *показатель экспоненты можно выбрать не зависящим от* f *и от разбиения*).

(iv) *Для любого множества* $A \in \mathscr{B}$, *такого, что* $0 < \lambda(A) < \infty$, *и любого* $\varepsilon > 0$ *существует* n_0, *такое, что при любом* $n \geqslant n_0$ *найдется множество* $A_n \in \sigma(\mathscr{P}_n)$, *удовлетворяющее условию*

$$\lambda(A \Delta A_n) < \varepsilon; \tag{14}$$

и

$$\sup_{\substack{M>0, \\ \text{все} \\ \text{множества } C \\ \text{конечной} \\ \text{меры Лебега}}} \limsup_{n\to\infty} \lambda\left(\bigcup_{j\,:\,\lambda(A_{nj}\cap C)\leqslant M/n} A_{nj} \cap C\right) = 0. \tag{15}$$

Наше доказательство этой теоремы отличается от доказательства Abou-Jaoude только некоторыми деталями. Например, мы сократили его, используя мощную лемму 1. Условия (14) и (15) в ряде случаев легко проверяются. Рассмотрим, например, *кубическую гистограммную оценку*, для которой каждое из множеств A_{nj} имеет вид $\prod_{i=1}^{d} [a_i k_i h,\ a_i(k_i + 1)h)$, где k_i — целые числа, h — параметр сглаживания, такой же, как в ядерной оценке, и a_i — положительные постоянные. В теореме 2.5 было показано, что для этой оценки (14) имеет место тогда и только тогда, когда

$$\lim_{n\to\infty} h = 0.$$

Кроме того, легко видеть, что (15) имеет место тогда и только тогда, когда

$$\lim_{n\to\infty} nh^d = \infty.$$

Отметим, что в другой статье Abou-Jaoude (1976b) приведены похожие необходимые и достаточные условия слабой состоятельности в L_1 гистограммных оценок на R^1, для которых разбиения зависят от порядковых статистик выборки $X_1, \ldots, X_n$. По этому поводу см. теорему 7.3.

4. Доказательство теоремы 2

Всегда имеет место неравенство $E\left(\int |f_n - f|\right) \geqslant \int |g_n - f|$. Кроме того, в силу ограниченности J_n сходимость J_n к 0 в среднем эквивалентна сходимости по вероятности. Следовательно, по теореме 2.5 получим (i) $\Rightarrow$ (14). Так как, очевидно, (iii) $\Rightarrow$ $\Rightarrow$ (ii) $\Rightarrow$ (i), то остается лишь доказать включения (iv) $\Rightarrow$ (iii) и (i) $\Rightarrow$ (15). Это будет сделано в двух отдельных леммах.

Лемма 4. (iv) $\Rightarrow$ (iii).

Доказательство. Мы знаем, что $\int |g_n - f| \to 0$ (теорема 2.5). Таким образом, достаточно показать, что $\int |f_n - g_n| \to 0$ экспоненциально. Пусть μ_n — эмпирическая мера для $X_1, \ldots, X_n$ и μ — вероятностная мера на борелевских множествах из R^d, определяемая плотностью f. Имеем

$$\int |f_n - g_n| = \sum_j |\mu_n(A_{nj}) - \mu(A_{nj})|.$$

Разобьем множество натуральных чисел на две части — множество H_n, содержащее все целые j, для которых $\lambda(A_{nj}C) > M/n$, и его дополнение H_n^c. Имеем

$$\begin{aligned}\int |f_n - g_n| &\leqslant \sum_{j\in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + \\ &\quad + \sum_{j\in H_n^c} (\mu_n(A_{nj}) + \mu(A_{nj})) \leqslant \\ &\leqslant \sum_{j\in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + 2\mu(A_n) + \\ &\quad + |\mu_n(A_n) - \mu(A_n)|, \qquad (16)\end{aligned}$$

где $A_n = \bigcup_{j\in H_n^c} A_{nj}$.

Так как $\lambda(A_{nj} \cap C) > M/n$ при $j \in H_n$, то множество H_n содержит не более $1 + n\lambda(C)/M$ элементов. Кроме того, случайный вектор $\{n\mu_n(A_n);\ n\mu_n(A_{nj}C),\ j \in H_n\}$ имеет полиномиальное распределение. Следовательно, если

$$\lambda(C)/M + 2/n \leqslant \varepsilon^2/20, \tag{17}$$

то в силу леммы 1

$$P\left(\sum_{j \in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + |\mu_n(A_n) - \mu(A_n)| > \varepsilon\right) \leqslant$$

$$\leqslant 3\exp(-n\varepsilon^2/25).$$

Величину $\mu(A_n)$ можно сделать сколь угодно малой за счет выбора C. В самом деле,

$$\mu(A_n) = \mu(A_n \cap C) + \mu(A_n \cap C^c) \leqslant o(1) + \mu(C^c), \tag{18}$$

где $o(1)$ возникает в силу соотношения (15) (в котором утверждается, что $\lambda(A_n \cap C) = o(1)$) и того, что мера μ абсолютно непрерывна относительно λ.

Таким образом, для заданного $\varepsilon > 0$ выберем C так, что (18) $\leqslant \varepsilon + o(1)$, а затем выберем M так, чтобы (17) имело место при всех достаточно больших n. Подстановка всех полученных неравенств в (16) дает

$$P\left(\int |f_n - g_n| > 4\varepsilon\right) \leqslant 3\exp(-n\varepsilon^2/25)$$

при всех достаточно больших n. Лемма 4 доказана.

Лемма 5. (i) $\Rightarrow$ (15).

Доказательство. Сохраним обозначения леммы 4. В частности, $M > 0$ — постоянная, C — множество конечной меры Лебега. Предположим, что $\lambda(C) > 0$, и обозначим $f = I_C/\lambda(C)$ и

$$Z_n = \sum_j \lambda(A_{nj} \cap C) \frac{I_{[X_i \notin A_{nj} \cap C,\ i=1,\ \ldots,\ n]}}{\lambda(C)}.$$

Имеем

$$\int |f_n - g_n| = \sum_j |\mu_n(A_{nj}) - \mu(A_{nj})| =$$

$$= \sum_j |\mu_n(A_{nj}) - \lambda(A_{nj} \cap C)/\lambda(C)| \geqslant Z_n.$$

Так как $0 \leqslant Z_n \leqslant 1$, то из (i) следует, что $E(Z_n) \to 0$. Далее,

$$E(Z_n) = \sum_j \frac{\lambda(A_{nj} \cap C)}{\lambda(C)} \left(1 - \frac{\lambda(A_{nj} \cap C)}{\lambda(C)}\right)^n \geqslant$$

$$\geqslant \sum_{j \in H_n^c} \frac{\lambda(A_{nj} \cap C)}{\lambda(C)} \left(1 - \frac{M}{n\lambda(C)}\right)^n \geqslant$$

$$\geqslant \frac{\lambda(A_n \cap C)}{\lambda(C)} \exp\left(-\frac{M/\lambda(C)}{1 - M/n\lambda(C)}\right) \sim$$

$$\sim \frac{\lambda(A_n \cap C)}{\lambda(C)} \exp\left(-\frac{M}{\lambda(C)}\right).$$

Но отсюда следует, что $\lim\limits_{n\to\infty} \lambda(A_n \cap C) = 0$ для любого множества C, такого, что $\lambda(C) > 0$. Кроме того, если $\lambda(C) = 0$, то ясно, что $\lambda(A_n \cap C) = 0$ при любых n. Таким образом, условие (15) выполнено.

5. Относительная устойчивость

При сравнении различных оценок плотности работать с величиной J_n неудобно. Можно было бы использовать квантили или моменты J_n, мы же воспользуемся величиной $E(J_n)$. Однако такой выбор был бы неудачным, если бы величина J_n не была близка к $E(J_n)$ в некотором смысле. В самом деле, желательно, чтобы оценки были *относительно устойчивыми* (по вероятности, почти наверное), т. е. чтобы выполнялось соотношение

$$\frac{J_n}{E(J_n)} \to 1 \qquad (19)$$

(по вероятности, почти наверное). Заметим, что обычно последовательность случайных величин J_n называется относительно устойчивой, если существует последовательность действительных чисел a_n, такая, что J_n/a_n стремится к единице в некотором стохастическом смысле. Наше определение несколько отличается от этого, поскольку мы дополнительно полагаем $a_n = E(J_n)$. Тем не менее доказать соотношение (19) фактически так же трудно, как установить предельное распределение J_n. К счастью, гораздо легче доказать, что *вариация* оценки $\int |f_n - E(f_n)|$ относительно устойчива. В силу лемм 6 и 7 это приводит к утверждениям относительно J_n, близким к соотношению (19).

Лемма 6. *Для любой плотности f на R^d и любой оценки плотности f_n имеем*

$$\max\left(\int |f - E(f_n)|, \frac{1}{2} E\left(\int |f - E(f_n)|\right)\right) \leqslant E\left(\int |f_n - f|\right).$$

Доказательство. По неравенству Йенсена $E\left(\int |f_n - f|\right) \geqslant \int |E(f_n) - f|$. Кроме того, по неравенству треугольника

$\int |f_n - f| \geqslant \int |f_n - E(f_n)| - \int |E(f_n) - f|$. Соединяя первое неравенство с усредненным вторым, получим требуемый результат.

Лемма 7. *Если вариация оценки плотности относительно устойчива по вероятности, т. е.*

$$\frac{\int |f_n - E(f_n)|}{E\left(\int |f_n - E(f_n)|\right)} \to 1$$

по вероятности, то $P(J_n/E(J_n) \notin (0, 3+\varepsilon)) \to 0$ *при* $n \to \infty$ *для любых* $\varepsilon > 0$. *Если вариация относительно устойчива почти наверное, то* $P(J_n/E(J_n) \notin (0, 3+\varepsilon)$ *бесконечно часто*$) = 0$ *для любых* $\varepsilon > 0$.

Доказательство. Применяя лемму 6 и неравенство треугольника, имеем

$$\frac{J_n}{E(J_n)} \leqslant \frac{\int |f_n - E(f_n)| + \int |f - E(f_n)|}{\max\left(\int |f - E(f_n)|,\ \frac{1}{2} E\left(\int |f_n - E(f_n)|\right)\right)} \leqslant$$

$$\leqslant 2 \frac{\int |f_n - E(f_n)|}{E\left(\int |f_n - E(f_n)|\right)} + 1.$$

Грубо говоря, если вариация оценки относительно устойчива, то величина $J_n/E(J_n)$ с большой вероятностью остается в интервале $[0, 3]$. Это указывает на то, что величина $E(J_n)$ является довольно хорошей мерой сравнения оценок плотности. (Именно таким сравнением мы будем заниматься в гл. 4, 5, 7, 8 и 9).

Оставшаяся часть этого параграфа следует работе Abou-Jaoude (1977), который показал, что вариации гистограммной оценки и ядерной оценки с равномерным ядром $K(x) = I_{[-1/2,\ 1/2]^d}$ относительно устойчивы по вероятности *для любых* f. Чтобы доказать это, понадобятся некоторые неравенства для биномиального распределения.

Лемма 8 (неравенство для абсолютного уклонения биномиальной случайной величины). *Пусть* X *— случайная величина, имеющая биномиальное распределение с параметрами* (n, p), *где* $p \leqslant 1/2$. *Тогда*

$$E\left(\left(p - \frac{X}{n}\right)_+\right) \geqslant \begin{cases} p/e^2 & \text{при } p < 1/n, \\ c\sqrt{p/n} & \text{при } p \geqslant 1/n, \end{cases}$$

где $c = (\sqrt{4\pi e^{13/6}})^{-1}$ — *универсальная постоянная. Кроме того,*

$$E\left(\left(p - \frac{X}{n}\right)_+\right) \leqslant \sqrt{\frac{p}{n}}.$$

Доказательство. Пусть $m = [np]$ (наибольшее целое, не превосходящее np). С помощью элементарных вычислений получим

$$E((np - X)_+) = \sum_{i=0}^{m} (np - i)\binom{n}{i} p^i (1-p)^{n-i} =$$

$$= np\binom{n-1}{m} p^m (1-p)^{n-m}$$

при $n \geqslant 2$. Если $p < 1/n$, так что $m = 0$, то эта величина равна $np\,(1-p)^n \geqslant npe^{-np/(1-p)} \geqslant npe^{-2}$. Если $p \geqslant 1/n$, так что $m > 0$, то по формуле Стирлинга получим

$$E((np - X)_+) = (2\pi)^{-1/2} p \sqrt{\frac{n-m}{m/n}} \left(\frac{np}{m}\right)^m \left(\frac{n(1-p)}{n-m}\right)^{n-m} g(n, m),$$

где $g(n, m) = \exp(u/12n - v/12m - w/12(n-m)) \geqslant \exp(-1/6)$ (здесь u, v, w — числа, принадлежащие отрезку $[0, 1]$). Кроме того, $n - m \geqslant n/2$ и $m/n \leqslant p$. Более того, так как $m = np - z$, $z \in [0, 1]$, то

$$\left(\frac{np}{m}\right)^m \left(\frac{n(1-p)}{n-m}\right)^{n-m} = \left(1 - \frac{z}{np}\right)^{-(np-z)} \left(1 + \frac{z}{n(1-p)}\right)^{-(n-np+z)} \geqslant$$

$$\geqslant \exp\left(+z - \frac{z^2}{np} - z - \frac{z^2}{n-np}\right) \geqslant$$

$$\geqslant \exp\left(-\frac{1}{np} - \frac{1}{n-np}\right) \geqslant \exp(-2).$$

Объединяя эти оценки, получим искомый результат для $n \geqslant 2$.

Для $n = 1$ заметим, что $E((p - X/n)_+) = p(1-p)$ и, таким образом, искомое неравенство выполняется для всех n. Верхнюю границу легко получить, используя неравенство Коши—Шварца и замечая, что $E((p - X/n)^2) = p(1-p)/n \leqslant p/n$.

Лемма 9. (Geffroy; см. Abou-Jaoude (1977, с. 52—53)). *Пусть* p_1, p_2, p_3 — *вероятностный вектор и* X_1, X_2, X_3 — *случайный вектор, имеющий полиномиальное распределение с параметрами* (n, p_1, p_2, p_3). *Тогда*

$$E\left(\left(p_1 - \frac{X_1}{n}\right)_+ \left(p_2 - \frac{X_2}{n}\right)_+\right) \leqslant E\left(\left(p_1 - \frac{X_1}{n}\right)_+\right) E\left(\left(p_2 - \frac{X_2}{n}\right)_+\right).$$

Доказательство. Пусть выполнено следующее условие:

Функция $E\left(\left(p_1 - \frac{X_1}{n}\right)_+ \middle| X_2 = n_2\right)$ монотонно возрастает и выпукла по n_2. (20)

Тогда

$$E\left(\left(p_1-\frac{X_1}{n}\right)_+\right)=\sum_{n_2=0}^{n}P(X_2=n_2)\,E\left(\left(p_1-\frac{X_1}{n}\right)_+\Big|\,X_2=n_2\right)\geqslant$$

$$\geqslant E\left(\left(p_1-\frac{X_1}{n}\right)_+\Big|\,X_2=[np_2]\right)\geqslant \text{ (по неравенству Йенсена)}$$

$$\geqslant E\left(\left(p_1-\frac{X_1}{n}\right)_+\Big|\,X_2=n_2\right)$$

при любых $n_2\leqslant np_2$. Таким образом,

$$E\left(\left(p_1-\frac{X_1}{n}\right)_+\left(p_2-\frac{X_2}{n}\right)_+\right)=\sum_{n_2=0}^{[np_2]}(p_2-n_2/n)\,P(X_2=n_2)\times$$

$$\times E\left(\left(p_1-\frac{X_1}{n}\right)_+\Big|\,X_2=n_2\right)\leqslant E\left(\left(p_1-\frac{X_1}{n}\right)_+\right)E\left(\left(p_2-\frac{X_2}{n}\right)_+\right),$$

что и требовалось доказать.

Теперь докажем (20). Необходимо показать, что функция $\psi(m)=\varphi(m+1)-\varphi(m)$ положительна и монотонно возрастает по m, где φ — функция, определенная в (20). Пусть Y_m — случайная величина X_1 при фиксированном $X_2=m$. Очевидно, справедливо следующее разложение: $Y_m=Y_{m+1}+Z$, где Z — бернуллиева случайная величина с параметром $p_1/(1-p_2)$, не зависящая от Y_{m+1}. Таким образом,

$$\psi(m)=E\left(\left(p_1-\frac{Y_{m+1}}{n}\right)_+-\left(p_1-\frac{Y_{m+1}}{n}-Z\right)_+\right)=E(U).$$

Но случайная величина U принимает следующие значения:

$$U=\begin{cases}0, & \text{если } Z=0 \text{ или если } Z=1,\ p_1-Y_{m+1}/n\leqslant 0,\\ p_1-Y_{m+1}/n, & \text{если } Z=1,\ 0<p_1-Y_{m+1}/n\leqslant 1/n,\\ 1/n, & \text{если } Z=1,\ 1/n<p_1-Y_{m+1}/n.\end{cases}$$

Если $z=np_1-[np_1]$, то

$$\psi(m)=P(Z=1)\left(\frac{z}{n}P(Y_{m+1}=[np_1])+\frac{1}{n}P(Y_{m+1}\leqslant np_1-1)\right)=$$

$$=\frac{p_1}{n(1-p_2)}(zP(Y_{m+1}\leqslant np_1)+(1-z)P(Y_{m+1}\leqslant np_1-1)).$$

Эта величина положительна. Заметим также, что в силу указанного выше разложения она возрастает с ростом m. Лемма 9 доказана.

Лемма 10. *Пусть Z_1, Z_2, ..., Z_n — последовательность неотрицательных случайных величин, такая, что $E(Z_n)\neq 0$ и*

$E(Z_n^2) < \infty$ при любых n. Тогда если

$$\lim_{n\to\infty} E(Z_n^2)/(E(Z_n))^2 = 1,$$

то $Z_n/E(Z_n) \to 1$ по вероятности.

Доказательство. По неравенству Чебышева

$$P\left(\left|\frac{Z_n}{E(Z_n)} - 1\right| > \varepsilon\right) \leqslant \frac{\mathrm{Var}(Z_n)}{\varepsilon^2 E^2(Z_n)} = o(1)$$

при любых $\varepsilon \geqslant 0$.

Теорема (Abou-Jaoude (1977)). *Предположим, что для гистограммной оценки из § 3 найдется постоянная $\eta > 0$, такая, что*

$$A_n(\varepsilon) = \sum_{\mu(A_{nj}) \leqslant \varepsilon} \mu(A_{nj}) \geqslant \eta, \quad n \geqslant n_0,$$

при любых $\varepsilon > 0$ и некотором n_0. (Это условие выполнено для кубической гистограммной оценки при $h \to 0$). Тогда вариация гистограммной оценки относительно устойчива по вероятности:

$$\frac{\int |f_n - E(f_n)|}{E\left(\int |f_n - E(f_n)|\right)} \to 1$$

по вероятности.

Доказательство. Заметим, что

$$\int |f_n - E(f_n)| = 2Z_n,$$

где

$$Z_n = \sum_j (\mu(A_{nj}) - \mu_n(A_{nj}))_+.$$

Ввиду леммы 10 и очевидного неравенства $E^2(Z_n) \leqslant E(Z_n^2)$ достаточно показать лишь, что

$$\limsup_{n\to\infty} E(Z_n^2)/E^2(Z_n) \leqslant 1. \tag{21}$$

Используя лемму 9, получим

$$E(Z_n^2) = \sum_j E\left((\mu(A_{nj}) - \mu_n(A_{nj}))_+^2\right)$$

$$+ \sum_{i\neq j} E\left((\mu(A_{nj}) - \mu_n(A_{nj}))_+ (\mu(A_{ni}) - \mu_n(A_{ni}))_+\right) \leqslant$$

$$\leqslant \frac{1}{n} \sum_j \mu(A_{nj})(1 - \mu(A_{nj}))$$

$$+\sum_{i\neq j} E\left((\mu(A_{nj})-\mu_n(A_{nj}))_+\right) E\left((\mu(A_{ni})-\mu_n(A_{ni}))_+\right) \leqslant$$

$$\leqslant \frac{1}{n} + E^2(Z_n). \tag{22}$$

Но (21) следует из (22) и соотношения $\sqrt{n}E(Z_n) \to \infty$, которое сейчас будет доказано. По лемме 8

$$\sqrt{n}\,E(Z_n) \geqslant c \sum_{\varepsilon \geqslant \mu(A_{nj}) \geqslant 1/n} \sqrt{\mu(A_{nj})} + e^{-2}\sqrt{n} \sum_{\mu(A_{nj})<1/n} \mu(A_{nj}) \geqslant$$

$$\geqslant \eta \min\left(c/\sqrt{\varepsilon},\ \sqrt{n}/e^2\right), \quad n \geqslant n_0,$$

где c — постоянная из леммы 8. Теперь теорема 3 следует из произвольности ε.

Замечание. Мы знаем, что (в силу абсолютной непрерывности μ относительно меры Лебега) $\sup \mu(A_{nj}) \to 0$ для кубической гистограммной оценки с параметром сглаживания $h = h_n \to 0$, так что

$$\lim_{n\to\infty} A_n(\varepsilon) = 1$$

при любых $\varepsilon > 0$. Таким образом, выполнено условие теоремы 3.

Теорема 4 (Abou-Jaoude (1977)). *Рассмотрим ядерную оценку с ядром* $K(x) = I_{[-1/2,\ 1/2]^d}$ *и параметром сглаживания* $h \to 0$. *Тогда ее вариация относительно устойчива по вероятности.*

Доказательство. Обозначим $[-1/2, 1/2]^d$ через C, $\mu(x+hC)$ — через $p(x)$ и $\mu_n(x+hC)$ — через $p_n(x)$. Напомним, что $\sup_x p(x) \to 0$ при $n \to \infty$ в силу абсолютной непрерывности μ относительно меры Лебега. Рассуждая так же, как в доказательстве теоремы 3, замечаем, что вариация равна $2Z_n/h^d$, где

$$Z_n = \int (p - p_n)_+.$$

Снова достаточно установить (21). Пусть теперь D — множество всех пар x, y из R^{2d}, для которых L_∞-расстояние $\|x-y\|$ не превосходит h, и пусть D^c — дополнение D. Заметим, что

$$E(Z_n^2) = \int\int E\left((p(x)-p_n(x))_+(p(y)-p_n(y))_+\right) dx\,dy$$

$$\leqslant \int_D E\left((p(x)-p_n(x))_+(p(y)-p_n(y))_+\right) dx\,dy + E^2(Z_n) =$$

$$= a_n + E^2(Z_n),$$

где на D^c применена лемма 9. Для завершения доказательства достаточно показать, что $E(Z_n)/\sqrt{a_n} \to \infty$. Чтобы сделать это, следует получить хорошие верхние границы для a_n.

Рассмотрим теперь множества $A = (x + hC) - (y + hC)$, $A' = (x + hC) \cap (y + hC)$ и $A'' = (y + hC) - (x + hC)$ при фиксированных x, y. Тогда

$$(p(x) - p_n(x))_+ (p(y) - p_n(y))_+ =$$
$$= (\mu(A + A') - \mu_n(A + A'))_+ (\mu(A' + A'') - \mu_n(A' + A''))_+ \leqslant$$
$$\leqslant ((\mu(A) - \mu_n(A))_+ + (\mu(A') - \mu_n(A'))_+) \times$$
$$\times ((\mu(A') - \mu_n(A'))_+ + (\mu(A'') - \mu_n(A''))_+) \leqslant$$
$$\leqslant (\mu(A) - \mu_n(A))_+^2 + 2(\mu(A') - \mu_n(A'))_+^2 + (\mu(A'') - \mu_n(A''))_+^2$$

и математическое ожидание этого выражения не превосходит

$$\frac{1}{n}(\mu(A) + 2\mu(A') + \mu(A'')) \leqslant \frac{1}{n}(\mu(x + hC) + \mu(y + hC)).$$

Таким образом,

$$a_n \leqslant \int_D \frac{1}{n}(\mu(x + hC) + \mu(y + hC))\, dx\, dy \leqslant$$
$$\leqslant \frac{2}{n}(2h)^d \int \mu(x + hC)\, dx = \frac{2}{n} 2^d h^{2d}.$$

Но если $\sup p(x) \leqslant 1/2$, то по лемме 8

$$E(Z_n) \geqslant \int_{1/2 \geqslant p(x) > 1/n} \frac{c}{\sqrt{n}} \frac{p(x)}{\sqrt{\sup p(x)}}\, dx + \int_{p(x) \leqslant 1/n} e^{-2} p(x)\, dx \geqslant$$
$$\geqslant \frac{h^d}{\sqrt{n}} \min\left(\frac{c}{\sqrt{\sup p(x)}},\ e^{-2}\sqrt{n}\right) \int \frac{p(x)}{h^d}\, dx.$$

Теорема 4 следует теперь из того факта, что $\sup p(x) \to 0$ и $\int (p(x)/h^d)\, dx = 1$.

Замечание. Распространение этого результата на случай общего K мы оставляем в качестве упражнения. Если $f \in L_2(R)$, то интегральная квадратичная ошибка относительно устойчива, т. е. $\int (f_n - f)^2 / E\left(\int (f_n - f)^2\right) \to 1$ по вероятности, по крайней мере если h, f и K удовлетворяют некоторым условиям регулярности (Hall (1982)). В важной статье Hall (1984) недавно получено асимптотическое распределение величины $\int (f_n - f)^2$ в случае, когда f имеет две ограниченные равномерно непрерывные производные на R^d и K — ограниченная плотность, соответствующая случайному вектору с нулевым средним и единичной матрицей ковариаций.

Методы, использованные в доказательствах теорем 1 и 2, позволяют получить полезные результаты об устойчивости почти наверное. Ниже это показано для кубической гистограммной оценки.

Теорема 5. *Пусть f_n — кубическая гистограммная оценка на R^d, в которой используются положительные постоянные a_i, $1 \leqslant i \leqslant d$, и параметр сглаживания h (обозначения из § 3), причем $\lim\limits_{n\to\infty} h = 0$, $\lim\limits_{n\to\infty} nh^d = \infty$. Пусть f — произвольная плотность на $[0, 1]^d$ и c — постоянная из леммы 8. Тогда для любого $\varepsilon \in (0, 1)$ существует $n_0 > 0$, такое, что*

$$P\left(\frac{J_n}{E(J_n)} \geqslant 1 + \frac{\sqrt{20}}{c\int\sqrt{f}\,(1-\varepsilon)}\right) \leqslant$$

$$\leqslant 3\exp\left(-\frac{4}{5}\left(\prod_{i=1}^{d} a_i h^d (1-\varepsilon)\right)^{-1}\right)$$

при $n \geqslant n_0$. Если, кроме того, $\lim\limits_{n\to\infty} h^d \log n = 0$, то

$$\limsup_{n\to\infty} \frac{J_n}{E(J_n)} \leqslant 1 + \frac{\sqrt{20}}{c\int\sqrt{f}}$$

почти наверное.

Доказательство. Длины сторон прямоугольников A_{nj}, $j = 1, \ldots,$ определяющих кубическую гистограммную оценку, равны ha_i, $i = 1, \ldots, d$. Обозначим через N число прямоугольников A_{nj}, для которых $\mu(A_{nj}) > 0$. Ясно, что

$$N \leqslant \prod_{i=1}^{d}\left(2 + \frac{1}{ha_i}\right).$$

Определим постоянную $b = 1 + \sqrt{20}\,/\left(c\int\sqrt{f}\,(1-\varepsilon)\right)$. Теперь, используя обозначения § 3 и 4 и леммы 6 и 8, получим

$$P(J_n/E(J_n) \geqslant b) \leqslant$$

$$\leqslant P\left(1 + 2\int |f_n - E(f_n)| \Big/ E\left(\int |f_n - E(f_n)|\right) \geqslant b\right) =$$

$$= P\left(\sum_j |\mu_n(A_{nj}) - \mu(A_{nj})| \geqslant \frac{b-1}{2}\sum_j E\left(|\mu_n(A_{nj}) - \mu(A_{nj})|\right)\right) =$$

$$= P\left(\sum_j |\mu_n(A_{nj}) - \mu(A_{nj})| \geqslant (b-1)\sum_j E((\mu(A_{nj}) - \mu_n(A_{nj}))_+)\right) \leqslant$$

$$\leqslant P\left(\sum_j |\mu_n(A_{nj}) - \mu(A_{nj})| \geqslant (b-1) \times \right.$$

$$\left. \times \sum_j \min\left(e^{-2}\mu(A_{nj});\ c\sqrt{\mu(A_{nj})/n}\right)\right).$$

Но

$$\sum_j \min\left(e^{-2}\mu(A_{nj}),\ c\sqrt{\mu(A_{nj})/n}\right) \geqslant$$

$$\geqslant \sum_j c\sqrt{\mu(A_{nj})/n} - \sum_{j:\,\mu(A_{nj})<c^2e^4/n} c\sqrt{\mu(A_{nj})/n} \geqslant$$

$$\geqslant \sum_j c \int_{A_{nj}} \sqrt{f}\Big/\sqrt{n\lambda(A_{nj})} - Nc^2e^2/n =$$

$$= c\int \sqrt{f}\Big/\sqrt{\prod a_i\,(nh^d)} - O(1/(nh^d)).$$

Поскольку $nh^d \to \infty$, то, таким образом,

$$P(J_n/E(J_n) \geqslant b) \leqslant P\left(\sum_j |\mu_n(A_{nj}) - \mu(A_{nj})| \geqslant \delta\right)$$

при достаточно больших n, где

$$\delta = \left(\frac{20}{(\prod a_i)\,nh^d\,(1-\varepsilon)}\right)^{1/2}.$$

Далее, заметим, что $N/n \leqslant \delta^2/20$ при достаточно больших n (так как $N/n \sim (nh^d \prod a_i)^{-1}$). Таким образом, в силу леммы 1

$$P(J_n/E(J_n) \geqslant b) \leqslant 3\exp(-n\delta^2/25),$$

что и требовалось доказать. Последнее утверждение теоремы 5 следует из этого неравенства и леммы Бореля—Кантелли.

ЛИТЕРАТУРА

Abou-Jaoude S.
(1976a) Sur une condition nécessaire et suffisante de L_1-convergence presque complète de l'estimateur de la partition fixe pour une densité, Comptes Rendus de l'Academie des Sciences de Paris Série A, 283, pp. 1107—1110.
(1976b) Sur la convergence L_1 et L_∞ de l'estimateur de la partition aléatoire pour une densité, Annales de l'Institut Henri Poincaré, 12, pp. 299—317.
(1976c) Conditions nécessaires et suffisantes de convergence L_1 en probabilité de l'histogramme pour une densité, Annales de l'Institut Henri Poincaré, 12, pp. 213—231.
(1977) La convergence L_1 et L_∞ de certains estimateurs d'une densité de probabilité. Thèse de Doctorat d'État, Université Paris VI, Paris.
Bretagnolle J., Huber C.
(1978) Lois empiriques et distance de Prokhorov, in Seminaire de Probabilites XII, Springer Notes in Mathematics, vol. 649, pp. 332—341.
Cacoullos T.
(1966) Estimation of a multivariate density, Annals of the Institute of Statistical Mathematics, 18, pp. 178—179.
Devroye L.
(1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, Annals of Statistics, 11, pp. 896—904.

Hall P.
(1982) Limit theorems for stochastis measurs of the accuracy of density estimators, Stochastic Processes and Applications, 13, pp. 11–25.
(1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, Journal of Multivariate Analysis, 14, pp. 1—16.
Hoeffding W.
(1963) Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association, 58, pp. 13—30.
Parzen E.
(1962) On estimation of a probability density function and mode, Annals of Mathematical Statistics, 33, pp. 1065—1076.
Rosenblatt M.
(1956) Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, pp. 832—837.

КОММЕНТАРИЙ ПЕРЕВОДЧИКА

Исследованию предельного распределения интегрального риска $\int (f_n - f)^2$, о котором идет речь в замечании после теоремы 4, посвящено много работ, начиная с Bickel, Rosenblatt (1973*). Более подробно об этом см. книгу Надарая (1983*).

Понятие относительной устойчивости введено Abou-Jaoude (1977). Относительная устойчивость ядерной оценки плотности недавно доказана в работе Devroye (1986a*), где из экспоненциальных границ для величины $\sup_{h>0,\, f} P(|J_n - E(J_n)| > \varepsilon)$ извлечен следующий результат:

Теорема 1. *Пусть f — произвольная плотность на R^1, f_n — ее ядерная оценка с неотрицательным ограниченным ядром K, имеющим компактный носитель. Тогда если оценка f_n состоятельна (т. е. $E(J_n) \to 0$), то она относительно устойчива почти наверное.*

Деврой (Devroye (1986a*)) получил аналогичные утверждения для оценок плотности на R^d и для случая, когда параметр сглаживания h зависит от выборки $X_1, \ldots, X_n$. При исследовании относительной устойчивости по вероятности можно использовать следующий изящный результат, основанный на неравенстве Эфрона—Стейна (Efron, Stein (1981*)):

Теорема 2 (Devroye (1986b*)). *Пусть f — произвольная плотность на R^d. Если f_n — ядерная оценка с произвольным параметром сглаживания h и произвольным ядром K, таким, что $\int K = 1$, то*

$$\operatorname{Var}\left(\int |f_n - f|\right) \leqslant 4\left(\int |K|\right)^2 / n.$$

Если f_n — гистограммная оценка с произвольным параметром сглаживания h, то

$$\operatorname{Var}\left(\int |f_n - f|\right) \leqslant 4/n.$$

Теорема 2 и частный случай теоремы 1 приведены в книге Devroye (1987*) (см. список дополнительной литературы к гл. 1).

Из теоремы 2 и неравенства Чебышева следует, что если

(A) $$\lim_{n\to\infty} \sqrt{n}\, E\left(\int |f_n - f|\right) = \infty,$$

то $J_n/E(J_n) \xrightarrow{p} 1$, $n \to \infty$, т. е. оценка f_n относительно устойчива по вероятности. Условие (A) выполняется в типичных для непараметрического оценивания случаях (см. теоремы 2 и 5 гл. 5), однако при некоторых сочетаниях f, K и h оно нарушается (теорема 17 гл. 5). Существует гипотеза (Devroye (1986b*)), что $\sqrt{n}$-состоятельные оценки плотности не являются относительно устойчивыми.

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Надарая Э. А.
(1983*) Непараметрическое оценивание плотности вероятностей и кривой регрессии. — Тбилиси: Изд-во Тбилисск. гос. ун-та.

Bickel P., Rosenblatt M.
(1973*) On some global measures of the deviations of density function estimates// Ann. Statist., v. 1, n. 6, p. 1071—1095.

Devroye L.
(1986a*) The kernel estimate is relatively stable. Techn. Report, McGill Univ.
(1986b*) An application of the Efron—Stein inequality in density estimation. Techn. Report, McGill Univ.

Efron B., Stein C.
(1981*) The jackknife estimate of variance//Ann. Statist., v. 9, n. 3, p. 586—596.

Глава 4

Нижние границы для скоростей сходимости

1. Введение

В этой главе мы попытаемся получить нижние границы, содержащие общую информацию о возможных скоростях сходимости для $E\left(\int |f_n - f|\right)$ при любых оценках плотности. Здесь возникают две задачи:

(i) Получить нижние границы для

$$\sup_{f \in \mathcal{F}} E\left(\int |f_n - f|\right),$$

где $\mathcal{F}$ — соответствующим образом ограниченный класс плотностей. Такие границы называются *равномерными нижними границами.*

(ii) Получить нижние границы для

$$\sup_{f \in \mathcal{F}} \limsup_{n \to \infty} a_n^{-1} E\left(\int |f_n - f|\right),$$

где a_n — некоторая последовательность положительных чисел.

Таким образом, в (ii) ставится задача отыскания наихудших скоростей сходимости для отдельных плотностей f из $\mathcal{F}$ [1]). Иногда мы будем называть величину $\inf_{f_n} \sup_{f \in \mathcal{F}} E\left(\int |f_n - f|\right)$ (которая зависит только от n и $\mathcal{F}$) *минимаксным риском*, а нижние и верхние границы для нее — *минимаксными нижними границами* и *минимаксными верхними границами.*

Рассмотрим следующие классы плотностей на R^1:

G — класс всех плотностей, равных нулю вне $[0, 1]$ и ограниченных числом 2;

G_∞ — класс всех плотностей, равных нулю вне $[0, 1]$, ограниченных числом $2 + \delta$ (при некотором $\delta > 0$) и бесконечное число раз непрерывно дифференцируемых на $[0, 1)$;

---

[1]) Границы для величины, указанной в (ii), под знаком sup далее называются границами для фиксированной плотности f (в оригинале — individual bounds). — *Прим. перев.*

$H(g)$—класс всех плотностей вида $\sum_{i=1}^{\infty} p_i g(x+x_i)$, где g — произвольная заданная плотность, носитель которой содержится в $[0, 1]$, $(p_1, p_2, \ldots)$ — вероятностный вектор [1]) и x_i — возрастающая последовательность действительных чисел, такая, что $x_{i+1} - x_i > 1$;

U — класс всех монотонных плотностей на $[0, \infty)$, имеющих максимум в точке 0;

U_∞ — класс всех симметричных бесконечное число раз непрерывно дифференцируемых унимодальных плотностей с центром в точке 0.

Теорема 1. *Пусть f_n — произвольная оценка плотности. Тогда*

(i) $$\inf_n \sup_{f \in \mathcal{F}} E\left(\int |f_n - f|\right) \geqslant c,$$

где $c = 1$ при $\mathcal{F} = G, G_\infty, H(g)$ и $c = 1/8$ при $\mathcal{F} = U, U_\infty$.

(ii) *Пусть $\{a_n\}$ — последовательность положительных чисел, стремящаяся к 0. Тогда для всех классов $\mathcal{F}$, упомянутых в* (i).

$$\sup_{f \in \mathcal{F}} \limsup_{n \to \infty} \frac{1}{a_n} E\left(\int |f_n - f|\right) = \infty.$$

Теорема 1 называется *теоремой о медленной сходимости*. Так как она охватывает классы U_∞ и $H(g)$, то для исследования скоростей сходимости произвольных оценок плотности, очевидно, недостаточно условий непрерывности на $\mathcal{F}$. Например, если $g(x) = c \exp(-1/x(1-x))$, $0 \leqslant x \leqslant 1$, то любая плотность f из $H(g)$ бесконечное число раз непрерывно дифференцируема. Более того, так как класс G тоже включен в теорему 1, то одного условия на хвосты или условия ограниченности также недостаточно. Таким образом, для получения содержательных равномерных нижних границ и нижних границ для фиксированной плотности f, по-видимому, необходимо сочетание условий непрерывности и условий на хвосты. Но даже при этом следует проявлять осторожность, имея в виду класс G_∞: недифференцируемость f даже в одной точке достаточна для того, чтобы получить результат о медленной сходимости.

Заметим, что в утверждении (ii) теоремы 1 выбирается *одна* плотность f из $\mathcal{F}$; она обычно зависит от последовательностей f_n и a_n, но, будучи выбранной, остается одной и той же при всех n. Утверждение (ii) показывает, что на рассматриваемых здесь клас-

---

[1]) То есть $p_i \geqslant 0$, $i = 1, 2, \ldots$ и $\sum_{i=1}^{\infty} p_i = 1$. — *Прим. перев.*

сах $\mathcal{F}$ достигается любая скорость сходимости. Отметим, что в утверждении (ii) можно заменить lim sup на lim inf (Birgé (1983b)).

Теорема 1 (i) не вполне удовлетворительна, ибо возможно, что значения $E\left(\int |f_n - f|\right)$ велики для тех отдельных плотностей f внутри классов $\mathcal{F}$, для которых велики значения некоторого критерия, измеряющего, насколько плотность f не гладкая и насколько длинны ее хвосты. Одним из таких критериев, который еще раз появится естественным образом в гл. 5, является

$$D_s(f) = \left(\int |f^{(s)}|\right)^{1/2s+1} \left(\int \sqrt{f}\right)^{2s/(2s+1)},$$

где $f^{(s)}$ есть s-я производная f. Отметим здесь, что $\int \sqrt{f}$ — мера того, насколько тяжелы хвосты f. Число s может принимать любое целое положительное значение. Теперь мы увидим, что теорема 1 (i) в сильной степени обусловлена присутствием внутри каждого класса $\mathcal{F}$ плотностей с большими значениями $D_s(f)$.

Теорема 2. *Пусть f_n — произвольная оценка плотности и g — произвольная плотность на* [0, 1], *имеющая непрерывную s-ю производную $g^{(s)}$. Тогда*

$$\liminf_{n\to\infty} n^{s/(2s+1)} \sup_{f \in H(g)} \frac{E\left(\int |f_n - f|\right)}{D_s(f)} \geqslant \frac{(s/e\,(2s+1))^{s/(2s+1)}}{D_s(g)},$$

при любых $s \geqslant 1$.

Заметим здесь, что $\inf_g D_s(g) = C(s) > 0$, для $s = 1, 2$ (см. гл. 5), так что нижняя граница равна

$$\frac{(s/e\,(2s+1))^{s/(2s+1)}}{C(s)}, \quad s = 1, 2,$$

при условии, что верхняя грань по всем f из $H(g)$ заменена на верхнюю грань по всем g на [0, 1] и всем f из $H(g)$.

Ясно, что в отличие от теоремы 1 верхнюю грань в теореме 2 нельзя приблизить плотностями, для которых $D_s(f) = \infty$. Таким образом, в теореме 2 речь идет о наихудших f, которые в некотором смысле достаточно хороши (так как $D_s(f) < \infty$). Теорема 2 также указывает на важность нормирующего множителя.

Чтобы проиллюстрировать теорему 2 в случае $s = 2$, мы немного забежим вперед и аннонсируем следующий результат гл. 5 относительно ядерных оценок f_n: при всех f имеем

$$\liminf_{n\to\infty} n^{2/5} \frac{E\left(\int |f_n - f|\right)}{D_2(f)} \geqslant c > 0,$$

где c — универсальная постоянная, а $D_2(f)$ — величина, определяемая так же, как и выше, для дважды непрерывно дифференци-

руемых f (и иначе — для других f). В этом утверждении содержится информация, которую нельзя получить из теоремы 2, поскольку это результат для фиксированной плотности f (а не равномерный). Однако из него следует, что при $s \geqslant 3$ нижняя граница теоремы 2 не достигается на ядерной оценке. При $s \geqslant 3$ нужны либо другие оценки, либо существенно модифицированная ядерная оценка. В § 5.9 и 7.6 будет показано, что достаточно допустить, чтобы ядро K принимало отрицательные значения.

Внимательный анализ доказательства теоремы 2 показывает, что $D_s(f_n^*) \to \infty$ для плотностей f_n^* из класса $H(g)$, для которых велико значение $E\left(\int |f_n - f|\right) / D_s(f)$. Мы можем теперь еще немного продвинуть технику получения границ, рассматривая классы хороших плотностей, таких, как

$F_{s,r}$ — класс всех плотностей на $[0, 1]$, имеющих $s-1$ абсолютно непрерывных производных, s-ю производную $f^{(s)}$ и таких, что $D_s(f) \leqslant r$,

или

$F_{s,\infty}$ — класс всех плотностей на $[0, 1]$, имеющих $s-1$ абсолютно непрерывных производных, s-ю производную $f^{(s)}$ и таких, что $D_s(f) < \infty$.

Эти классы не связаны операцией включения ни с какими из рассмотренных до сих пор классов. Еще одна теорема гласит:

Теорема 3 (Bretagnolle, Huber (1979)) [1]). *Существует число $r^* > 0$, зависящее лишь от s, такое, что для любой оценки плотности f_n*

$$\liminf_{n\to\infty} n^{s/(2s+1)} \sup_{f\in F_{s,r}} E\left(\int |f_n - f|\right) \geqslant (2e)^{-4}\left(\frac{r}{r^*} - 1\right)$$

при любых $r > r^*$ и

$$\liminf_{n\to\infty} n^{s/(2s+1)} \sup_{f\in F_{s,\infty}} E\left(\int |f_n - f|\right) = \infty.$$

Теорема 3 сильнее, чем теорема 2, в том смысле, что в ней верхние грани берутся по классам плотностей $F_{s,r}$ с равномерно ограниченными значениями $D_s(f)$. По этой причине рассуждения в ней более тонкие и изощренные. Заметим, что скорость $n^{-s/(2s+1)}$

[1]) В английском оригинале доказательства этой теоремы имеется пробел. Значение r^* в доказательстве, помещенном в переводе, не вычисляется. — *Прим. перев.*

приближается к $n^{-1/2}$ при $s \to \infty$. Хотя при $s \geqslant 3$ эта скорость не достигается на ядерной оценке ни для каких f, она достигается на других оценках, во всяком случае, в смысле границ для фиксированных плотностей f из $F_{s,\infty}$, имеющих компактный носитель (см., например, оценку Бартлета, описанную в § 7.6 и § 5.9).

Все рассмотренные до сих пор классы все же довольно обширны. Ясно, что дальнейшее существенное сокращение размеров классов приведет к уменьшению нижних границ. Если мы сократим классы так сильно, что останется семейство с одним-единственным параметром θ, то нижние границы должны выполняться для всех параметрических оценок плотностей из данного семейства. Нижние границы, полученные для таких семейств, обычно не достигаются на общих оценках плотности, рассматриваемых в данной книге. Однако они достигаются на некоторых специальных параметрических оценках плотности. В качестве примера рассмотрим следующее простое семейство плотностей:

$\Pi(g)$ — класс всех плотностей вида $f(x) = pg(x) + (1-p)g(x+2)$, где g — произвольная плотность, носитель которой содержится в $[0, 1]$, и $p \in [0, 1]$ — параметр смеси, не известной пользователю.

Заметим, что все f из класса $\Pi(g)$ имеют компактный носитель и бесконечное число раз непрерывно дифференцируемы, если g бесконечное число раз непрерывно дифференцируема. Мы докажем следующую теорему.

Теорема 4. *Пусть f_n — некоторая оценка и g — произвольная плотность, носитель которой содержится в $[0, 1]$. Тогда*

(i) *при любых $n \geqslant 4$ имеем*

$$\sup_{f \in \Pi(g)} \sqrt{n}\, E\left(\int |f_n - f|\right) \geqslant 0{,}030153\ldots$$

$$\left(\textit{в действительности } \sup_{f \in \Pi(g)} E\left(\int |f_n - f|\right) \geqslant \right.$$

$$\left. \geqslant (0{,}0849856\ldots + o(1))/\sqrt{n}\right);$$

(ii) $$\sup_{f \in \Pi(g)} \limsup_{n \to \infty} \sqrt{n}\, E\left(\int |f_n - f|\right) \geqslant 0{,}0424928\ldots.$$

Таким образом, можно сделать вывод, что при не совсем тривиальном выборе класса $\mathcal{F}$ наилучшая из возможных скоростей сходимости в L_1 равна $1/\sqrt{n}$. Заметим, например, что если $\mathcal{F}$ содержит только конечное число элементов $g_1, \ldots, g_N$ и по определению $f_n = g_i$, где номер i вычисляется по данным $X_1, \ldots, X_n$,

то $E\left(\int |f_n - f|\right) \leqslant 2P(g_i \neq f)$. Последняя величина стремится к 0 экспоненциально с ростом n, если g_i выбирается по принципу максимума правдоподобия (см. § 11.9, посвященный обнаружению). Таким образом, для всех f из этого конечного класса $\mathscr{F}$ имеется экспоненциально убывающая *верхняя* граница!

Пользователи обычно жалуются, что большинство статистических теорий имеет асимптотическую природу. Что же можно сделать при малых выборках? Существует ли «оптимальность при малых выборках»? Один способ получения хороших оценок при малых n состоит в том, чтобы подходящим образом ограничить множество рассматриваемых плотностей и решить следующую задачу о минимаксе: для каких f_n величина $\sup_{f \in \mathscr{F}} E\left(\int |f_n - f|\right)$ минимальна? (Это минимальное значение $m(n, \mathscr{F})$ есть функция только от n и $\mathscr{F}$.) Здесь можно было бы рассмотреть классы всех монотонных плотностей на $[0, \infty)$, всех симметричных унимодальных плотностей, всех C-липшицевых плотностей, всех логарифмически вогнутых плотностей с модой в точке 0, всех плотностей с возрастающей интенсивностью отказов [1]) на $[0, \infty)$ и так далее. В частности, если задано только X_1 (*задача с одним наблюдением*), то что представляет собой f_1 для каких-либо из этих классов? Интересно, конечно, получить хорошие границы для $m(n, \mathscr{F})$ при любых n. Асимптотически хорошие границы можно получить, как правило, методами, развитыми в этом и в следующем параграфах, но ввиду сказанного они представляют меньший практический интерес.

2. Лемма Ассуа

Идеи, на которых основаны результаты § 1, можно разбить на три группы: во-первых, нельзя оценить плотность f на данном интервале, если в этот интервал не попадает ни одно из наблюдений X_i (теоремы 1 и 2), во-вторых, можно получать нижние границы теоретико-информационными методами (теорема 3), в-третьих, можно использовать такие неравенства, как неравенство Крамера—Рао, и свойства достаточных статистик (теорема 4). В этом параграфе мы хотели бы привлечь внимание к мощной и простой технике, изложенной в работах Assouad (1983) и Birgé (1980, 1983, 1986), позволяющей заново вывести некоторые результаты § 1 и получить ряд новых нижних границ для важных классов плотностей. Бирже использовал понятия ε-энтропии и ε-емкости,

---

[1]) Интенсивностью отказов плотности f на $[0, \infty)$ называется функция $\lambda(x) = f(x)/(1 - F(x))$, где F — функция распределения, соответствующая f. Это понятие используется в теории надежности. — *Прим. перев.*

введенные Колмогоровым и Тихомировым (1959), что позволило ему не только получить равномерные нижние, но и верхние границы для минимаксного риска

$$\inf_{f_n} \sup_{f \in \mathcal{F}} E\left(\int |f_n - f|\right).$$

Он успешно ответил на вопрос о получении для некоторых верхних и нижних границ, имеющих одинаковую зависимость от n (но с различными коэффициентами). Вторая часть работы Бирже здесь не рассматривается, поскольку важные верхние границы будут получены в гл. 5.

Ключевым результатом является мощная лемма, принадлежащая Ассуа (Assouad (1983)) (теорема 5 ниже). Она приводится в виде, более удобном для применения в книге.

Теорема 5. (Assouad, 1983) (общая формулировка). *Пусть* $r \geqslant 1$ *— целое число и* $b = (b_1, \ldots, b_r) \in \{-1, 1\}^r$ *— параметр семейства плотностей* $f(b, \cdot)$, *содержащего* 2^r *элементов. Пусть параметры* b_{i+} *и* b_{i-} *определены равенствами*

$$b_{i+} = (b_1, b_2, \ldots, b_{i-1}, +1, b_{i+1}, \ldots, b_r),$$
$$b_{i-} = (b_1, b_2, \ldots, b_{i-1}, -1, b_{i+1}, \ldots, b_r).$$

Если существует такое разбиение $A_0, A_1, \ldots, A_r$ *пространства* R^d, *что для любых* b *и любых* $1 \leqslant i \leqslant r$ *выполнены неравенства*

$$\int_{A_i} |f(b_{i+}, \cdot) - f(b_{i-}, \cdot)| \geqslant \alpha > 0$$

и

$$\int \sqrt{f(b_{i+}, \cdot) f(b_{i-}, \cdot)} \geqslant \beta > 0,$$

то для любой оценки плотности f_n

$$\sup_b E\left(\int |f_n - f|\right) \geqslant \begin{cases} (r\alpha/2)\left(1 - \sqrt{2 - 2\beta^n}\right), \\ (r\alpha/4)\,\beta^{2n}. \end{cases}$$

В терминах величины $\gamma = 1 - \beta$ *нижние границы можно заменить на* $(r\alpha/2)\left(1 - \sqrt{2n\gamma}\right)$ *и* $(r\alpha/4)\,(1 - \gamma)^{2n}$ *соответственно.*

(Частная формулировка) (Birgé, 1986). *Пусть* $r \geqslant 1$ *— целое число, пусть* $A = [0, l)$, $l \leqslant 1/r$, *— интервал, на котором определена измеримая функция* g, *удовлетворяющая условиям*

$$|g| \leqslant 1, \quad \int_A g = 0,$$

пусть $g = 0$ *вне* A, *и пусть* $y_1, \ldots, y_r$ *— действительные числа, такие, что множества* $A + y_i$ *не пересекаются. Пусть* f_0 *— плот-*

ность на R, принимающая значение 1 на множестве $\bigcup_i A + y_i$. Пусть F — класс 2^r плотностей, параметризованный с помощью

$$b = (b_1, \ldots, b_r) \in \{-1, 1\}^r,$$

элементы которого $f(b, \cdot)$ определяются следующим образом:

$$f(b, x) = \begin{cases} f_0(x), & x \notin \bigcup_i A + y_i, \\ f_0(x) + b_i g(x - y_i), & x \in A + y_i. \end{cases}$$

Тогда

$$\sup_{f \in F} E\left(\int |f_n - f|\right) \geqslant r \int_A |g| \left(1 - \sqrt{2n \int_A g^2}\right) \geqslant \frac{r}{2} \int_A |g|$$

для любой оценки плотности f_n, если $n \int_A g^2 \leqslant 1/8$.

Если задано некоторое общее семейство $\mathscr{F}$, то сперва следует найти r, l, g, f_0 и y_1, ..., y_r, такие, что семейство F из теоремы 5 полностью содержится в $\mathscr{F}$. Тогда любая нижняя граница, полученная для F, необходимо является нижней границей для $\mathscr{F}$, которую и требовалось найти. Покажем теперь, как действует этот подход для некоторых важных классов.

Определим *класс Липшица* $W(s, \alpha, C)$ как класс всех плотностей f, сосредоточенных на отрезке $[0, 1]$, имеющих $(s-1)$ абсолютно непрерывных производных и удовлетворяющих условию

$$|f^{(s)}(x) - f^{(s)}(y)| \leqslant C|x - y|^\alpha, \quad x, y \in R,$$

где $s \geqslant 0$ — целое, C — положительное число и $\alpha \in (0, 1]$. Говорят, что функция g является *C-липшицевой*, если

$$|g(x) - g(y)| \leqslant C|x - y|, \quad x, y \in R.$$

В анализе, проведенном в § 1, центральным понятием были функционалы $D_s(f)$. Здесь же мы начнем с небольшого и приятного класса — класса Липшица $W(s, \alpha, C)$, не заботясь по началу о величине $D_s(f)$. В гл. 5 будет показано, что соответствующим образом обобщенный функционал, близкий к $D_{s+1}(f)$, равномерно ограничен на $W(s, 1, C)$.

Наиболее важными классами Липшица являются $W(1, 1, C)$ и $W(0, 1, C)$. Последний из них есть класс всех *C-липшицевых* плотностей, сосредоточенных на $[0, 1]$. Очевидно, класс $W(0, 1, C)$ пуст при $C < 4$ и содержит только один элемент (равнобедренную треугольную плотность на $[0, 1]$ [1]) при $C = 4$.

[1]) Равнобедренной треугольной плотностью на $[0, 1]$ здесь и далее называется плотность $f(x) = 4x$, $0 \leqslant x \leqslant 1/2$, $f(x) = 4(1 - x)$, $1/2 < x \leqslant 1$. — *Прим. перев.*

Заметим, что точно так же любой из классов $W(s, \alpha, C)$ не пуст, если C больше, чем некоторая постоянная C_0, и пуст при всех $C < C_0$.

Теорема 6. *Пусть f_n — произвольная оценка плотности. Для любого целого $s \geqslant 0$ и любого $\alpha \in (0, 1]$ существуют положительные постоянные c_1, c_2, c_3, γ_1 и γ_2, зависящие только от s и α, такие, что*

$$\sup_{f \in W(s, \alpha, C)} E\left(\int |f_n - f|\right) \geqslant$$

$$\geqslant \begin{cases} (c_3 + o(1))\, C^{1/(2(s+\alpha)+1)} n^{-(s+\alpha)/(2(s+\alpha)+1)}, \\ \gamma_1 \frac{C}{4}\left((16\gamma_2 n C^2)^{1/(2(s+\alpha)+1)} + 4\right)^{-(s+\alpha)}, \quad n \geqslant C c_2, \end{cases}$$

для всех $C \geqslant c_1$. Если $W^*(s, \alpha) = \bigcup\limits_{C>0} W(s, \alpha, C)$, то

$$\liminf_{n \to \infty} \sup_{f \in W^*(s, \alpha)} E\left(\int |f_n - f|\right) n^{(s+\alpha)/(1+2(s+\alpha))} = \infty.$$

Постоянные c_2 и c_3 можно вычислить следующим образом:

$$\gamma_0 = \left[(s+\alpha)^{s+1}\, 2^{1-\alpha} \exp\left(\frac{s+\alpha}{2} \exp\left(\frac{s+1}{s+\alpha}\right)\right)\right]^{-1},$$

$$\gamma_1 = \gamma_0 \Gamma^2(s+\alpha+1)\, \Gamma^{-1}(2s+2\alpha+2),$$

$$\gamma_2 = \gamma_0^2 \Gamma^2(2s+2\alpha+1)\, \Gamma^{-1}(4s+4\alpha+2),$$

$$\gamma_3 = \gamma_0 / 4^{s+\alpha},$$

$$c_2 = \gamma_3^{(2(s+\alpha)+1)/(s+\alpha)} (16\gamma_2)^{-1},$$

$$c_3 = \frac{1}{4} \gamma_1 (16\gamma_2)^{-(s+\alpha)/(1+2(s+\alpha))}.$$

Теорема 6 хороша тем, что она дает непрерывный спектр полиномиальных скоростей сходимости. К сожалению, постоянные c_i в теореме не оптимальны, так что стоит приложить дополнительные усилия в надежде получить полезные нижние границы для $W(1, 1, C)$ и $W(0, 1, C)$. Это сделано в следующей теореме.

Теорема 7. *Пусть f_n — произвольная оценка плотности. Тогда*

$$\sup_{f \in W(0, 1, C)} E\left(\int |f_n - f|\right) \geqslant \frac{21}{160}\left(\frac{12C}{25n}\right)^{1/3}, \quad n \geqslant \max\left(10, \frac{3C}{50}\right),$$

при любом $C \geqslant 72$ и

$$\sup_{f \in W(1, 1, C)} E\left(\int |f_n - f|\right) \geqslant \begin{cases} \left(\frac{1}{32}\left(\frac{30}{23}\right)^{2/5} C^{1/5} + o(1)\right) n^{-2/5}, \\ \frac{C}{32}\left(\left(\frac{23nC^2}{30}\right)^{1/5} + 8\right)^{-2}, \quad n \geqslant \frac{15C}{368}, \end{cases}$$

при любом $C \geqslant 288$.

В гл. 5 будет показано, что эти скорости достигаются на ядерной оценке и только для $W(0, 1, C)$ — на гистограммной оценке. Если нижние границы для C представляются не реалистичными, то пользователь может без большого труда понизить их за счет увеличения коэффициентов при $n^{-1/3}$ и $n^{-2/5}$ в теореме 7.

Покажем теперь, что из теоремы 5 можно также получить теоремы о быстрой и о медленной сходимости в духе теорем 1 и 4.

Теорема 8. *Пусть f_n — произвольная оценка плотности, $r \geqslant 1$ — фиксированное число и g — фиксированная измеримая функция на $[0, 1/r)$, такая, что $|g| \leqslant 1$, $\int_0^{1/r} g = 0$. Пусть $Q_r(g)$ — класс всех плотностей следующего вида:*

$$f(x) = 1 + \varepsilon b_i g\left(x - \frac{i}{r}\right), \quad \frac{i}{r} \leqslant x < \frac{i+1}{r}, \quad i = 0, 1, \ldots, r-1,$$

где $\varepsilon \in [0, 1]$ и числа $b_i \in \{-1, 1\}$, $1 \leqslant i \leqslant r$. Тогда

$$\sup_{f \in Q_r(g)} E\left(\int |f_n - f|\right) \geqslant \frac{r \int |g|}{\sqrt{32n \int g^2}}, \quad n \geqslant \left(8 \int g^2\right)^{-1}.$$

В частности, если $g = 1$ на $[0, 1/(2r))$ и $g = -1$ на $[1/(2r), 1/r)$, то

$$\sup_{f \in Q_r(g)} E\left(\int |f_n - f|\right) \geqslant \sqrt{r/32n}, \quad n \geqslant r/8,$$

и

$$\sup_{f \in \bigcup_r Q_r(g)} E\left(\int |f_n - f|\right) \geqslant 1/2$$

при любых n.

Несмотря на свою простоту, теорема 8 содержит полезную информацию. Равномерная нижняя граница для $Q_1(g)$ имеет порядок $1/\sqrt{n}$. Эта нижняя граница применима к задаче *параметрического* оценивания, так как она остается в силе, даже если функция g задана и только ε и b_i неизвестны. Таким образом, ситуация снова сравнима с тем, что было в теореме 4. Неудивительно, что нижняя граница возрастает с ростом r и что для $\bigcup_r Q_r(g)$ фактически получается результат, аналогичный теореме 1. В этом результате нет ничего удивительного, поскольку $\bigcup_r Q_r(g)$ — густо населенный подкласс класса G всех плотностей на $[0, 1]$, ограниченных числом 2.

В этом параграфе были представлены только равномерные результаты. Индивидуальные нижние границы, такие, как в теореме 1 (ii), тоже можно получить из теоремы 5 путем сложных построений (см., например, Birgé (1986)).

Закончим этот параграф равномерной нижней границей для класса $M_B = \{$все монотонно невозрастающие плотности на $[0, 1]$, такие, что $f(0) \leqslant B\}$. Ясно, что этот класс пуст, если $B < 1$.

Теорема 9. *Для любой оценки плотности и любого* $B \geqslant 2$ *имеем*

$$\sup_{f \in M_B} E\left(\int |f_n - f|\right) \geqslant \frac{1}{16\left(3 + 2\,(n/4)^{1/3}\right)} \sim \frac{1}{32}\left(\frac{4}{n}\right)^{1/3}.$$

Ясно, что такая же нижняя граница верна для всех симметричных унимодальных плотностей на $[-1, 1]$, таких, что $f(0) \leqslant B/2$, и для всех унимодальных плотностей на $[-1, 1]$ с модой $m \in [-1, 1]$, таких, что $f(m) \leqslant B/2$. Полученная здесь нижняя граница — далеко не наилучшая, хотя бы потому, что она не возрастает с ростом B. Однако доказательство теоремы 9 удивительно просто и по крайней мере степень n ($n^{-1/3}$) является правильной, так как равномерные верхние границы для средней L_1-ошибки ядерной и гистограммной оценок убывают как $n^{-1/3}$ (это будет показано в гл. 5). Урок, который отсюда следует, состоит в том, что если предполагать только условия монотонности и компактности, то ни от какой оценки плотности нельзя ожидать очень хорошего поведения, и что, по-видимому, не очень выгодно строить специальные оценки для M_B, возможно, не состоятельные вне M_B, поскольку они не могут быть существенно лучше ядерной оценки.

Важное замечание. Получение равномерных нижних и верхних границ для $E\left(\int |f_n - f|\right)$ на неоднородных или очень широких классах $\mathcal{F}$ является рискованным и часто бессмысленным упражнением. Оно очень напоминает определение максимального числа червяков в яблоке из кучи, куда наряду с только что собранными яблоками подброшена пара старых. С большой вероятностью результат определяется старыми яблоками и дает мало информации о только что собранных. Например, этот эффект возникает с классами Липшица $W(s, \alpha, C)$, для которых верхняя и нижняя границы возрастают с ростом C (при фиксированных s и α). Из теорем 2 и 3 и из верхних границ, которые будут получены в гл. 5, видно, что величина $D_s(f)$ является очень хорошей мерой «трудностей», доставляемых плотностью f. Но так как имеется лишь весьма слабая связь между C и $D_s(f)$, то теорема 6, например, содержит очень мало информации о подавляющем большин-

стве плотностей из класса $W(s, \alpha, C)$. В связи с этим более естественными и реалистичными представляются классы $F_{s,r}$, которые далее называются *классами Бретаньоля—Юбер*. Добавим, что в отличие от $W(s, \alpha, C)$ классы $F_{s,r}$ замкнуты относительно перенормировок [1]).

3. Некоторые исторические замечания

Нижние границы, собранные в § 1, имеют некоторые аналоги в L_p. По очевидным причинам подробное представление нижних границ в L_p здесь неуместно. Однако интересно напомнить некоторые исторические вехи, которые позволят лучше понять различие между нормой L_1 и нормами L_p при $p \neq 1$.

В § 1 были рассмотрены результаты о медленной сходимости (теорема 1), нижние границы с промежуточными скоростями сходимости (теоремы 2 и 3) и слабые нижние границы (теорема 4). Например, теорема 1 позволяет сделать важное замечание относительно результатов о сходимости в L_2 ядерной оценки,

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \quad x \in R,$$

где K — ограниченная симметричная плотность. Розенблат (Rosenblatt (1971)) показал, что

$$E\left(\int (f_n - f)^2\right) \sim \frac{\alpha}{nh_n} + \frac{\beta}{4} h_n^4,$$

если $h_n \to 0$ и $nh_n \to \infty$ при $n \to \infty$ и f — ограниченная плотность, имеющая две непрерывные производные, и $\int (f'')^2 < \infty$. Постоянные равны

$$\alpha = \int K^2, \quad \beta = \left(\int x^2 K(x)\,dx\right)^2 \int (f'')^2.$$

Выбирая $h_n = (\alpha/\beta n)^{1/5}$, получим оптимальную скорость сходимости в L_2:

$$E\left(\int (f_n - f)^2\right) \sim \frac{5}{4} \frac{\alpha^{4/5}\beta^{1/5}}{n^{4/5}}$$

(см. также Надарая (1974)). В то же время для некоторой плотности f из данного класса имеем

$$E\left(\int |f_n - f|\right) \geqslant \frac{1}{\log\log\log\log n}$$

[1]) То есть плотности вида $f_1(x) = a^{-1} f(x/a)$, $0 < a < 1$, принадлежат $F_{s,r}$, если $f \in F_{s,r}$. — *Прим. перев.*

бесконечно часто. Действительно, достаточно в теореме 1 взять класс $H(g)$, для которого $g(x) = \text{const} \cdot \exp(-1/x(1-x))$ на $[0, 1]$, и заметить, что все f из $H(g)$ ограниченны и удовлетворяют условию $\int (f'')^2 = \sum_{i=1}^{\infty} p_i^2 \int (g'')^2 \leqslant \int (g'')^2 < \infty$. Другими словами, классический результат Розенблата и бо́льшая часть других результатов о скорости сходимости в L_2 содержит мало информации о близости f_n к f. При выборе параметра h_n эти результаты следует использовать чрезвычайно осторожно. Такое несоответствие между хорошими скоростями сходимости в L_2 и плохими скоростями в L_1 вызвано тем, что в L_2 менее существенны хвосты распределений (и участки, где значения f малы). Заметим, однако, что при дополнительных условиях на хвосты для класса плотностей, рассмотренного Розенблатом, достигается оптимальная в L_1 скорость сходимости ядерной оценки $n^{-2/5}$ (см. гл. 5).

Укажем без доказательства аналог теоремы 1 для L_p.

Теорема 10 (Devroye (1983)). *Пусть f_n — произвольная оценка плотности, $p \geqslant 1$ — фиксированное действительное число и $f \in L_p$. Тогда*

$$\text{(i)} \qquad \inf_n \sup_{f \in \mathcal{F}} \frac{E\left(\int |f_n - f|^p\right)}{\int f^p} \geqslant \frac{1}{2^{p-1}},$$

где $\mathcal{F} = H(g)$ при произвольном g или $\mathcal{F} = G$.

(ii) *Пусть $\{a_n\}$ — последовательность положительных чисел, стремящаяся к 0. Тогда*

$$\sup_{f \in G} \limsup_{n \to \infty} a_n^{-1} \frac{E\left(\int |f_n - f|^p\right)}{\int f^p} = \infty.$$

Имеется также ряд аналогов нижних границ с промежуточными скоростями сходимости. Эти аналоги будут приведены здесь без доказательств или только с краткими набросками доказательств. Читателю не следует забывать, что легко сбиться с курса в широком море результатов, представленных в литературе, — их столько же, сколько нормирующих множителей, норм и классов плотностей $\mathcal{F}$. Ниже дается определение обобщенного множителя $D_s(f)$ для L_p:

$$D_{sp}(f) = \left(\int |f^{(s)}|^p\right)^{(1/(2s+1)} \left(\int f^{p/2}\right)^{2s/(2s+1)}.$$

Следующий результат (Breatagnolle, Huber (1979)), имеющий место при $d = 1$, $s \geqslant 1$ и $p > 1$, служит дополнением к равномерным нижним границам теорем 1 и 3:

$$\liminf_{n\to\infty} n^{sp/(2s+1)} \sup_{f\in\mathcal{F}} E\left(\int |f_n - f|^p\right) \geqslant rC_{sp} > 0,$$

где C_{sp} — постоянная, зависящая только от s и p, $r > 0$ — действительное число и $\mathcal{F}$ — класс всех плотностей, имеющих $s-1$ абсолютно непрерывных производных, таких, что $f^{(s)} \in L_p$, $f \in L_p$ и $D_{sp}(f) \leqslant r$. Заметим, что при $p \geqslant 2$ из условия $D_{sp}(f) \leqslant r$ не следуют ограничения на хвосты f, так что данный результат не противоречит результату Розенблата о скорости сходимости ядерных оценок в L_2.

Подчеркнем еще раз важность нормировок. Например, при $p \geqslant 2$ можно найти плотность g, такую, что $H(g) \subseteq \mathcal{F}$. Это следует из того, что

$$D_{sp}(f)^{2s+1} = \left(\sum p_i^p\right)\left(\sum p_i^{p/2}\right)^{2s} D_{sp}(g)^{2s+1} \leqslant D_{sp}(g)^{2s+1}$$

при $f \in H(g)$. Таким образом, в силу теоремы 9 имеем $\sup_{f\in\mathcal{F}} E\left(\int |f_n - f|^p\right) \Big/ \int f^p \geqslant 1/2^{p-1}$ при любых n и любых $p \geqslant 2$.

Будем теперь постепенно уменьшать размеры классов $\mathcal{F}$. Прежде всего снова рассмотрим $H(g)$. Расчет, проведенный ниже в § 5, показывает, что

$$\begin{aligned}
\sup_{f\in H(g)} E\left(\int |f_n - f|^p\right) \geqslant \\
\geqslant \int g^p 2^{-(p-1)} \sup_{\substack{\text{все вероятностные} \\ \text{векторы } p_1, p_2, \ldots}} \sum_{i=1}^{\infty} p_i^p (1-p_i)^n \geqslant \\
\geqslant \int g^p \left(\frac{p-1}{2(1+n)}\right)^{p-1} \left(1 - \frac{p-1}{1+n}\right)^n \geqslant \\
\geqslant \int g^p \left(\frac{p-1}{2(n+1)} \exp\left(-\frac{n}{2+n-p}\right)\right)^{p-1} \sim \\
\sim \int g^p \frac{((p-1)/2)^{p-1}}{n^{p-1}}
\end{aligned}$$

при $n \to \infty$. При $p = 1$ эта нижняя граница равна 1 для любых n. При $p = 2$ она пропорциональна $1/n$ независимо от выбора g. В самом деле, так как $\int g^p \geqslant 1$ при $p \geqslant 1$, то

$$\sup_{f\in H(g)} E\left(\int (f_n - f)^2\right) \geqslant \frac{1}{2(n+1)} \exp\left(-\frac{n}{1+n}\right)$$

при любых n и g. Однако при другой нормировке получается

Теорема 11. *Пусть g — бесконечное число раз непрерывно дифференцируемая плотность с носителем* $[0, 1]$, *и пусть* $d = 1$. *Тогда*

$$\liminf_{n\to\infty} / n^{ps/(2s+1)} \sup_{f\in H(g)} \frac{E\left(\int |f_n - f|^p\right)}{D_{sp}(f)} \geqslant \frac{1}{2^{p-1}} \frac{(ps/(e(2s+1)))^{ps/(2s+1)}}{D_{sp}(g)}$$

для любых p, $s \geqslant 1$ *и любых оценок плотности* f_n.

Как показано выше, наилучшая скорость сходимости в L_2, достижимая равномерно на любом классе $H(g)$, равна $1/n$. Бойд и Стил (Boyd, Steele (1978)) получили несколько более сильное утверждение для небольшого класса BS всех нормальных с параметрами $(0, \sigma^2)$ плотностей.

Теорема 12 (Boyd, Steele (1978)). *Для любой оценки плотности* f_n *существует плотность* $f \in BS$, *такая, что*

$$\limsup_{n\to\infty} nE\left(\int (f_n - f)^2\right) \geqslant c(f) > 0,$$

где $c(f)$ — *постоянная, зависящая только от* f.

Результат Бойда и Стила нельзя улучшить при оценивании нормальных плотностей, так как если, например, оценкой f_n для f является нормальная с параметрами $(\hat{\mu}, \hat{\sigma}^2)$ плотность, где $\hat{\mu}$ и $\hat{\sigma}^2$ — обычные выборочные оценки для μ и σ^2, то

$$\lim_{n\to\infty} P\left(\int (f_n - f)^2 < \frac{x}{16\sqrt{\pi n}\sigma}\right) = F(x),$$

где F — функция распределения величины $4V + 3U$ и V, U — независимые случайные величины, имеющие распределение хи-квадрат с одной степенью свободы (см. работу Мания (1969), который также получил сходный результат при $d > 1$). Таким образом, скорость, предсказанная Бойдом и Стилом, достигается. Теорему 12 следует рассматривать как аналог теоремы 4 для L_2. Предположительно существует универсальная постоянная c, такая, что

$$\sup_{f\in BS} \limsup_{n\to\infty} \sqrt{n} E\left(\int |f_n - f|\right) \geqslant c > 0.$$

Кифер (Kiefer (1982)) приводит обзор литературы по нижним границам скорости сходимости оценок плотности; до настоящего времени основное внимание уделялось скоростям поточечной сходимости (см. Farrell (1967, 1972), Wahba (1975), Stone (1980), Ибрагимов, Хасьминский (1979)). Глобальные скорости сходимости (скорости сходимости в L_p) рассматривались Самаровым (1976) и Bretagnolle, Huber (1979). По поводу границ в L_∞ см.,

например, Ибрагимов, Хасьминский (1979) и Stone (1983). Скорость поточечной сходимости ядерной оценки подробно исследовали Wahba (1975), Rosenblatt (1971) и многие другие. Скорость сходимости ядерной оценки в L_2 была найдена Rosenblatt (1971), Надарая (1974), Deheuvels (1977a), Bretagnolle,Huber (1979) и другими авторами. Для случая $d > 1$ укажем работу Deheuvels (1977b).

4. Доказательство теоремы 1

Классы G и H (g). Начнем с определения двух семейств плотностей. Семейство 1 параметризовано действительным числом $b \in [0, 1]$ и вероятностным вектором $(p_1, p_2, \ldots)$. Указывая явно только зависимость от b, запишем

$$f(b, x) = \sum_{i=1}^{\infty} p_i g(x - 2i - b_i),$$

где g — плотность с носителем, содержащимся в $[0, 1]$, и b_i есть i-й бит в двоичной записи $0, b_1 b_2 b_3 \ldots$ числа b. При каждом b функция $f(b, x)$ является плотностью по x.

Семейство 2 параметризовано тем же b и тем же вероятностным вектором $(p_1, p_2 \ldots)$. Отрезок $[0, 1]$ разбивается на множества A_i, A_i', такие, что $\int_{A_i} dx = \int_{A_i'} dx = p_i/2$, и затем определяется

$$f(b, x) = 2 \sum_{i=1}^{\infty} I_{b_i A_i + (1-b_i) A_i'}(x).$$

В качестве A_i и A_i' всегда можно выбрать смежные интервалы, такие, что $A_i' = A_i + p_i/2$. Ясно, что семейство 2 является подклассом G, а семейство 1 — подклассом $H(g)$.

Доказательство основано на следующей схеме вложения. Пусть B — случайная величина, равномерно распределенная на $[0, 1]$, и пусть $X_1^*, \ldots, X_n^*$ — независимые случайные величины, не зависящие от B, с общей плотностью $f(0, \cdot)$. Определим множества $C_i = [2i, 2i + 2)$ для семейства 1 и $C_i = A_i \cup A_i'$ для семейства 2. Ясно, что в обоих случаях $P(X_1^* \in C_i) = p_i$. Обозначим B_i i-й бит в двоичной записи B. Определим теперь случайные величины $X_1, \ldots, X_n$ соотношениями

$$X_j = X_j^* + B_i \quad \text{при} \quad X_j^* \in C_i$$

для семейства 1;

$$X_j = X_j^* + \frac{B_i p_i}{2} \quad \text{при} \quad X_j^* \in C_i$$

для семейства 2.

Кроме того, будут использоваться случайные величины $N_i = \sum_{j=1}^{n} I_{[X_j \in C_i]} = \sum_{j=1}^{n} I_{[X_j^* \in C_i]}$, для которых $N = (N_1, N_2, \ldots)$ по построению *не зависит от* B. Наконец, пусть B' и B'' — случайные величины, отличающиеся от B только в i-м двоичном знаке, причем $B_i' = 0$, $B_i'' = 1$.

Используются два основных приема. Первый прием основан на том, что верхняя грань больше, чем среднее значение:

$$\sup_b E\left(\int |f_n - f(b, \cdot)|\right) \geqslant E\left(\int |f_n - f(B, \cdot)| = \right.$$

$$= E\left(\sum_{i=1}^{\infty} \int_{C_i} |f_n - f(B, \cdot)|\right). \qquad (1)$$

Второй прием состоит в том, что с использованием условной независимости B_i и $X_1, \ldots, X_n$ на множестве $[N_i = 0]$ из (1) исключаются f_n. Имеем

$$E\left(I_{[N_i=0]} \int_{C_i} |f_n - f(B, \cdot)| \,\middle|\, X_1, \ldots, X_n\right) =$$

$$= E\left(I_{[N_i=0]} I_{[B_i=0]} \int_{C_i} |f_n - f(B', \cdot)| + \right.$$

$$\left. + I_{[N_i=0]} I_{[B_i=1]} \int_{C_i} |f_n - f(B'', \cdot)| \,\middle|\, X_1, \cdots, X_n\right) =$$

$$= \frac{1}{2} I_{[N_i=0]} E\left(\int_{C_i} |f_n - f(B', \cdot)| + \right.$$

$$\left. + \int_{C_i} |f_n - f(B'', \cdot)| \,\middle|\, X_1, \ldots, X_n\right) \geqslant$$

$$\geqslant I_{[N_i=0]} \frac{1}{2} E\left(\int_{C_i} |f(B', \cdot) - f(B'', \cdot)|\right) = p_i I_{[N_i=0]} \qquad (2)$$

для обоих семейств. Объединяя теперь (1) и (2), получим

$$\sup_b E\left(\int |f_n - f(b, \cdot)|\right) \geqslant \sum_{i=1}^{\infty} p_i P(N_i = 0) = \sum_{i=1}^{\infty} p_i (1 - p_i)^n. \qquad (3)$$

Неравенство (3) оказывается достаточно сильным, чтобы доказать утверждение (i) для семейств G и H (g), так как имеется еще свобода в выборе $(p_1, p_2, \ldots)$. Пусть, например, $p_i = 1/M$, $1 \leqslant i \leqslant M$ и $p_i = 0$, $i > M$. Тогда правая часть (3) равна $(1 - 1/M)^n$ и

верхняя грань этого выражения по M равна 1. Доказательство для классов G_∞, U и U_∞ пока отложим.

Для доказательства части (ii) необходима

Лемма 1. *Если* $0 < a_n \leqslant 1/8$ *при всех* n *и* $\lim\limits_{n\to\infty} a_n = 0$, *то существует вероятностный вектор* $(p_1, p_2, \ldots)$, *такой, что*

$$\sum_{i=1}^{\infty} p_i (1 - p_i)^n \geqslant a_n$$

при любых n.

Доказательство. Сперва построим последовательность

$$a_n' = \max_{m \geqslant n} a_m + \frac{1}{4(n+1)}.$$

Эта последовательность такова, что $a_n' \geqslant a_n$ при всех n и $1/4 \geqslant a_1' \geqslant a_2' \geqslant \ldots \geqslant a_n' \downarrow 0$. Следовательно, найдутся целые $1 = k_1 < k_2 < \ldots$ и положительные числа p_i, такие, что $p_1 = 1 - 2a_1'$ и

$$p_i \leqslant (2n)^{-1}, \qquad \sum_{i=k_{n-1}+1}^{k_n} p_i = 2(a_{n-1}' - a_n')$$

при $n \geqslant 2$, $k_{n-1} < i \leqslant k_n$. Заметим, что, в частности,

$$\sum_{i=1}^{\infty} p_i = p_1 + \sum_{n=2}^{\infty} 2(a_{n-1}' - a_n') = 1 - 2a_1' + 2a_1' = 1.$$

Кроме того,

$$\sum_{i=1}^{\infty} p_i (1 - p_i)^n \geqslant \left(1 - \frac{1}{2n}\right)^n \sum_{p_i \leqslant 1/2n} p_i \geqslant$$

$$\geqslant \frac{1}{2} \sum_{p_i \leqslant 1/2n} p_j \geqslant \frac{1}{2} \sum_{i=k_{n-1}+1}^{\infty} p_i =$$

$$= \frac{1}{2} \sum_{i=n}^{\infty} 2(a_{i-1}' - a_i') = a_{n-1}' \geqslant a_n' \geqslant a_n$$

при $n \geqslant 2$. При $n = 1$ имеем $p_1(1 - p_1) = 2a_1'(1 - 2a_1') \geqslant a_1' \geqslant a_1$, так что лемма доказана.

Доказательство утверждения (ii). Воспользуемся здесь прежним вложением, но распространим его на бесконечные последо-

вательности $X_1^*, X_2^*, \ldots$, и $X_1, X_2, \ldots$. Нам понадобятся величины

$$J_n(b) = E\left(\int |f_n - f(b, \cdot)|\right) \text{ и } \bar{J}_n(b) = \sup_{m>n} J_m(b)/a_m.$$

Пусть $D_n = \{b: \bar{J}_n(b) > 1\}$. Поскольку D_n монотонно убывает, существует предельное множество D и, следовательно,

$$\int_{D_n \cap [0,1]} dx \downarrow \int_{D \cap [0,1]} dx = \lambda(D) \ (\lambda \text{ — мера Лебега}).$$

Пусть D_n^c — дополнение D_n.

Множество D_n введено для того, чтобы можно было применить лемму Фату:

$$\sup_b \limsup_{n\to\infty} \frac{J_n(b)}{a_n} \geqslant \sup_b \limsup_{n\to\infty} \left(\frac{J_n(b)}{a_n}\right) I_{D_n^c}(b) \geqslant$$

$$\geqslant E\left(\limsup_{n\to\infty} \left(\frac{J_n(B)}{a_n}\right) I_{D_n^c}(B)\right) \geqslant \limsup_{n\to\infty} E\left(\left(\frac{J_n(B)}{a_n}\right) I_{D_n^c}(B)\right). \quad (4)$$

Для членов последовательности, стоящей в (4), воспользуемся соответствующим образом видоизмененными неравенствами (1) и (2). Полагая $X = (X_1, X_2, \ldots)$, изменим левую часть (2), придав ей вид

$$E\left(I_{[N_i=0]} a_n^{-1} I_{D_n^c}(B) \int_{C_i} |f_n - f(B, \cdot)| \,\middle|\, X\right) \geqslant$$

$$\geqslant I_{[N_i=0]}(2a_n)^{-1} E\left(I_{[\bar{J}_n(B')\leqslant 1]} \int_{C_i} |f_n - f(B', \cdot)| + \right.$$

$$\left. + I_{[\bar{J}_n(B'')\leqslant 1]} \int_{C_i} |f_n - f(B'', \cdot)| \,\middle|\, X\right) \geqslant$$

$$\geqslant I_{[N_i=0]}(2a_n)^{-1} E\left(I_{[\max(\bar{J}_n(B'), \bar{J}_n(B''))\leqslant 1]} \int_{C_i} |f(B', \cdot) - \right.$$

$$\left. - f(B'', \cdot)| \,\middle|\, X\right) \geqslant$$

$$\geqslant I_{[N_i=0]} \frac{p_i}{a_n} E\left(I_{[\max(\bar{J}_n(B'), \bar{J}_n(B''))\leqslant 1]} \,\middle|\, X\right),$$

где математическое ожидание последнего выражения равно

$$p_i a_n^{-1} P(N_i = 0, \max(\bar{J}_n(B'), \bar{J}_n(B'')) \leqslant 1) \geqslant$$

$$\geqslant p_i a_n^{-1} (P(N_i = 0) - P(N_i = 0, \bar{J}_n(B') > 1) -$$

$$- P(N_i = 0, \bar{J}_n(B'') > 1)) \geqslant$$

$$\geqslant p_i a_n^{-1}(P(N_i=0)-2P(N_i=0,\ \bar{J}_n(B)>1)-$$
$$-2P(N_i=0,\ \bar{J}_n(B)>1))=$$
$$=p_i a_n^{-1}(P(N_i=0)-4P(N_i=0,\ B\in D_n)). \qquad (5)$$

Пусть $A_n=[k_{n-1},\ \infty)$, где k_n определено в доказательстве леммы 1. Определим также

$$Z_n=\sum_{i\in A_n} p_i I_{[N_i=0]}.$$

Выше было показано, что

$$E\left(\frac{J_n(B)}{a_n}I_{D_n^c}(B)\right)\geqslant E\left(\frac{Z_n}{a_n}\right)-4E\left(\frac{Z_n}{a_n}I_{D_n}(B)\right)\geqslant$$
$$\geqslant E\left(\frac{Z_n}{a_n}\right)-4\left(E\left(\frac{Z_n^2}{a_n^2}\right)P(B\in D_n)\right)^{1/2}$$

(по неравенству Шварца).

Выберем теперь вероятностный вектор $(p_1,\ p_2,\ \ldots)$, предписываемый конструкцией леммы 1. Тогда

$$E\left(\frac{Z_n}{a_n}\right)=a_n^{-1}\sum_{i\geqslant k_{n-1}} p_i(1-p_i)^n\to 0.$$

Кроме того,

$$E(Z_n^2)=\sum_{i\in A_n} p_i^2(1-p_i)^n+\sum_{i\neq j,\ i,\ j\in A_n} p_i p_j(1-p_i-p_j)^n\leqslant$$
$$\leqslant 2\sum_{i\in A_n} p_i^2(1-p_i)^{2n}+\sum_{i\neq j,\ i,\ j\in A_n} p_i(1-p_i)^n p_j(1-p_j)^n\leqslant$$
$$\leqslant 2E^2(Z_n),$$

где используется тот факт, что $(1-p_i)^n\geqslant 1/2$ на A_n и что $1-p_i-p_j\leqslant(1-p_i)(1-p_j)$ при любых $i,\ j$. Объединяя последние два факта с (5) и (4) и учитывая соотношение $\lim\limits_{n\to\infty} P(B\in D_n)=\lambda(D)$, заключаем, что

$$\sup_b \limsup_{n\to\infty}\frac{J_n(b)}{a_n}\geqslant(1-4\sqrt{2\lambda(D)})\limsup_{n\to\infty} E\left(\frac{Z_n}{a_n}\right).$$

В силу определения D всегда

$$\sup_b \limsup_{n\to\infty}\frac{J_n(b)}{a_n}\geqslant\lambda(D).$$

Так как при $\lambda(D)>0$ обязательно $\limsup\limits_{n\to\infty} J_n(b)/a_n>0$ для некоторого b и при $\lambda(D)=0$ то же самое выполнено в силу ра-

нее полученного неравенства, то можно заключить, что существует некоторое b, такое, что

$$\limsup_{n\to\infty} \frac{J_n(b)}{a_n} > 0.$$

Но так как при выборе вероятностного вектора $(p_1, p_2, \ldots)$ всегда можно заменить a_n на $\sqrt{a_n}$, то из этого результата следует утверждение (ii) для семейств G и $H(g)$.

Класс G_∞. Изменим слегка определения семейства 2. Рассмотрим плотность $g(x) = C\exp(-1/x(1-x))$ на $[0, 1]$. Пусть, далее, символами $C_1, C_2, \ldots$ обозначаются интервалы $[0, p_1]$, $[p_1, p_1+p_2]$, ... Разобьем каждый из интервалов на два интервала одинаковой длины и обозначим левый интервал A_i и правый интервал A_i'. Далее, положим

$$f(b, x) = \sum_{i=1}^{\infty} 2\left(b_i g_i + (1-b_i) g_i'\right),$$

где g_i — функция $g(x/(p_i/2))$, сдвинутая вправо на $p_1 + \ldots + p_{i-1}$, и g_i' есть функция g_i, сдвинутая на $p_i/2$ вправо. Таким образом, g_i равна нулю вне A_i и g_i' равна нулю вне A_i'. Более того, семейство 2 принадлежит G_∞. Чтобы убедиться в этом, достаточно заметить, что максимальное значение f, т. е. удвоенное максимальное значение g, равно $2C\exp(-4)$. Теперь

$$\int_0^1 \exp\left(-\frac{1}{x(1-x)}\right) = 2\int_0^{1/2} \exp\left(-\frac{4}{1-4y^2}\right) dy =$$

$$= 2e^{-4}\int_0^{1/2} \exp\left(-\frac{16y^2}{1-4y^2}\right) dy \geqslant 2e^{-4}\int_0^{1/\sqrt{20}} (1-20y^2)\, dy,$$

где было использовано неравенство $e^{-u} \geqslant 1-u$ и тот факт, что $1-4y^2 \leqslant 1$. Но данная нижняя граница равна $4e^{-4}/(3\sqrt{20})$. Следовательно, $2Ce^{-4} = 3\sqrt{20}/2 = 3\sqrt{5}$. Легко видеть, что для любого $\delta > 0$ из определения класса G_∞ можно выбрать подходящую функцию g, постоянную на $[\varepsilon, 1-\varepsilon]$ и очень гладкую на $[0, \varepsilon]$ и $[1-\varepsilon, 1]$. Нетрудно установить, что доказательство теоремы 1 переносится на такое семейство *без изменений*.

Класс U. Рассмотрим подкласс класса U, параметризованный, как и выше, с помощью b и $(p_1, p_2, \ldots)$. Величины p_i не возрастают с ростом i, и их сумма равна $1/2$. Прежде всего определим интервалы, вводя узловые точки

$$x_0 = 0, \quad x_1 = 1, \quad x_{i+1} - x_i = 2^i p_i, \quad i \geqslant 1.$$

Плотность $f(b, \cdot)$ равна нулю всюду, кроме интервалов $[x_i, x_{i+1})$, на которых она постоянна. При $x \in [x_i, x_{i+1})$ она определяется следующим образом:

$$f(x) = \begin{cases} \frac{1}{2} + \sum\limits_{\substack{i \geqslant 1, \\ b_i = 1}} \frac{p_i}{2}, & \text{если } i = 0, \\ 2^{-i}, & \text{если } i \neq 0, \quad b_i = 0, \\ 2^{-(i+1)}, & \text{если } i \neq 0, \quad b_i = 1. \end{cases}$$

Заметим, что f унимодальна на $[0, \infty]$ и что она является плотностью, поскольку

$$\int f = (x_1 - x_0)\left(\frac{1}{2} + \sum_{\substack{i \geqslant 1, \\ b_i = 1}} \frac{p_i}{2}\right) + \sum_{\substack{i \geqslant 1, \\ b_i = 1}} (x_{i+1} - x_i) 2^{-i-1} + \\ + \sum_{\substack{i \geqslant 1, \\ b_i = 0}} (x_{i+1} - x_i) 2^{-i} =$$

$$= \frac{1}{2} + \sum_{\substack{i \geqslant 1, \\ b_i = 1}} \frac{p_i}{2} + \sum_{\substack{i \geqslant 1, \\ b_i = 1}} \frac{p_i}{2} + \sum_{\substack{i \geqslant 1, \\ b_i = 0}} p_i = \frac{1}{2} + \sum_{i=1}^{\infty} p_i = 1.$$

Несколько сложнее теперь оказывается вложение. Начнем с независимых случайных величин B, X_1^*, ..., X_n^*, Y_1, ..., Y_n, где величина B такая же, как раньше, X_i^* имеют плотность $f(0, \cdot)$ (напомним, что $f(0, x) = 1/2$ на $[x_0, x_1)$ и $f(0, x) = 2^{-i}$ на $[x_i, x_{i+1})$, $i \geqslant 1$) и Y_i — бернуллиевы случайные величины с параметром $1/2$. Определим выборку X_1, ..., X_n из $f(b, \cdot)$ следующим образом:

$$X_j = \begin{cases} X_j^*, & \text{если } X_j^* \in [x_0, x_1) \\ & \text{или если } X_j^* \in [x_i, x_{i+1}), \quad i \geqslant 1, \text{ и } b_i = 0, \\ & \text{или если } X_j^* \in [x_i, x_{i+1}), \quad i \geqslant 1, \text{ и } b_i = 1, \\ & Y_j = 0, \\ \dfrac{X_j^* - x_i}{x_{i+1} - x_i}, & \text{если } X_j^* \in [x_i, x_{i+1}), \quad i \geqslant 1, \ b_i = 1, \\ & Y_j = 1. \end{cases}$$

Таким образом, в последнем случае X_j^* заменяется на случайную величину, равномерно распределенную на $[x_0, x_1)$. Легко удостовериться, что X_j имеют плотность $f(b, \cdot)$.

Проверьте, что (1) остается в силе при замене C_i на $[x_i, x_{i+1})$. Далее, так как при $N_i = 0$ $(i \geqslant 1)$, где N_i — число X_j в интервале $[x_i, x_{i+1})$, величины B_i и $X_1, \ldots, X_n$ условно независимы, то снова можно вывести (2) при всех $i \geqslant 1$ со следующими изменениями: математическое ожидание брать при фиксированных $X_1^*, Y_1, \ldots, X_n^*, Y_n$ и в последней строке заменить p_i на $p_i/4$. Таким образом, можно заключить, что справедливо соотношение (3) с правой частью, деленной на 4.

Эта нижняя граница зависит от n и от вектора величин p_i. Как показано выше, можно найти невозрастающую последовательность величин p_i, сумма элементов которой равна 1/2, такую, что $\sum_{i=1}^{\infty} (p_i (1 - p_i)^n / a_n)/4 \to \infty$ при $n \to \infty$. Кроме того, супремум рассматриваемой нижней границы по всем таким векторам величин p_i равен 1/8. Это завершает доказательство утверждения (i) для данного класса плотностей. Доказательство утверждения (ii) также не представляет значительной трудности.

Класс U_∞. Рассмотрим семейство кусочно-прямоугольных унимодальных плотностей, построенное для U, и построим новое семейство, содержащееся в U_∞ и состоящее из сходных кусочно-«почти прямоугольных» унимодальных плотностей. Это можно сделать, определяя $f(b, \cdot)$ на каждом из интервалов $[x_i, x_{i+1})$ так же, как и прежде, за исключением окрестностей точек x_i и x_{i+1}, где делается поправка с целью получения непрерывной функции. Так как U можно рассматривать в качестве предельного случая для U_∞, то нетрудно видеть, что все предыдущие результаты остаются здесь в силе.

5. Доказательства теорем 2 и 11

Заметим, что теорема 2 есть не что иное, как частный случай теоремы 11 (при $p = 1$). Рассмотрим снова семейство 1 из доказательства теоремы 1. Нетрудно проверить, что

$$\int f^p(b, \cdot) = \left(\sum p_i^p\right) \int g^p$$

и

$$D_{sp}(f) = D_{sp}(g) \left(\sum p_i^p\right)^{1/(2s+1)} \left(\sum p_i^{p/2}\right)^{2s/(2s+1)}.$$

Теперь снова проделаем выкладки (1) и (2) вплоть до предпоследнего выражения в (2), где следует заменить множитель 1/2 на 2^{-p}. Интеграл $\int_{C_i} |f(B', \cdot) - f(B'', \cdot)|^p$ равен $2p_i^p \int g^p$, откуда

$$\sup_b E\left(\int |f_n - f(b, \cdot)|^p\right) \geqslant 2^{-(p-1)} \int g^p \sum p_i^p (1 - p_i)^n.$$

Положим теперь $p_i = 1/M$ при $i = 1, 2, \ldots, M$ и $p_i = 0$ при других i. Тогда

$$\sup_b \frac{E\left(\int |f_n - f(b, \cdot)|^p\right)}{D_{sp}(f)} \geqslant$$

$$\geqslant \frac{\int g^p}{D_{sp}(g)} 2^{-(p-1)} \left(\frac{1}{M}\right)^{ps/(2s+1)} \left(1 - \frac{1}{M}\right)^n.$$

Положим $M \sim n/(ps/(2s+1))$, заметим, что $\int g^p \geqslant 1$ и устремим n к ∞.

6. Доказательство теоремы 3

Так же как в доказательстве теоремы 1 (i), воспользуемся методом рандомизации (см. (1)). Построим семейство плотностей, строго вложенное в $F_{s,t}$ и группирующееся вокруг некоторой центральной плотности g. Эта центральная плотность в свою очередь равна свертке $g_0 * g_1$, где g_0 — плотность с носителем, содержащимся в $[-1/4, 1/4]$, имеющая непрерывную $(s-1)$-ю производную, и g_1 — равномерная плотность на $[-1/2, 1/2]$. Таким образом,

$$(g_0 * g_1(x))^{(s)} = \int_{x-1/2}^{x+1/2} g_0^{(s)}(y)\, dy = g_0^{(s-1)}(x + 1/2) - g_0^{(s-1)}(x - 1/2)$$

и

$$\int |(g_0 * g_1)^{(s)}| = \int |g_0^{(s-1)}(x + 1/2) - g_0^{(s-1)}(x - 1/2)|\, dx =$$

$$= 2 \int |g_0^{(s-1)}(x)|\, dx.$$

Чтобы произвести рандомизацию, рассмотрим случайную величину B, равномерно распределенную на $[0, 1]$ и имеющую двоичную запись $B = 0, B_1 B_2 \ldots$, в которой все двоичные единицы равны $+1$ или -1 (т. е. все символы 0 заменены на -1). Отдельные реализации случайной величины B будут обозначаться через b. Также потребуются последовательности действительных чисел h_j, u_j, $j \geqslant 1$, такие, что $6\sum h_j \leqslant 1$, $0 \leqslant u_j \leqslant 1$. Найдем теперь действительные числа x_j, $j \geqslant 1$, для которых множества $A_j = [x_j - 3h_j/2, x_j + 3h_j/2]$ не пересекаются между собой и содержатся в $\{x: g(x) = 1\}$. (Это можно сделать, поскольку $6\sum h_j \leqslant 1$.)

При фиксированном $b \in [0, 1)$ определим плотность $f(b, x)$ из нашего семейства следующим образом:

$$f(b, x) = g(x)\left(1 + b_j u_j \left(g\left(\frac{x - x_j}{h_j} + \frac{3}{4}\right) - g\left(\frac{x - x_j}{h_j} - \frac{3}{4}\right)\right)\right), \quad x \in A_j,$$

и $f(b, x) = g(x)$ при x, не принадлежащих никакому из множеств A_j. Легко проверить, что $f \geqslant 0$, поскольку $|b_j u_j| \leqslant 1$ при любых j и $g \leqslant 1$. Кроме того, в силу определения g интеграл от f равен интегралу от g.

Ясно, что

$$\sup_{0 \leqslant b < 1} E\left(\int |f_n(x) - f(b, x)|\right) \geqslant \int_0^1 E\left(\int |f_n(x) - f(b, x)|\right) db \geqslant$$

$$\geqslant \sum_j \int_0^1 E\left(\int_{A_j} |f_n(x) - f(b, x)|\right) db, \qquad (6)$$

где зависимость f_n от выборки с плотностью $f(b, \cdot)$ и интегрирование по dx не указываются. Рассмотрим теперь какое-либо из слагаемых в (6) и фиксируем последовательность $b_1, b_2, \ldots, b_{j-1}, b_{j+1}, \ldots$. Пусть b^+, b^- — числа, двоичной записью которых является данная последовательность с добавлением $b_j = +1$ и $b_j = -1$ соответственно [1]). Тогда j-е слагаемое в (6), в котором оставлено только усреднение в j-м двоичном знаке b_j и усреднение E, равно [2])

$$\frac{1}{2}\left(E\left(\int_{A_j} |f_n(x) - f(b^+, x)|\right) + E\left(\int_{A_j} |f_n(x) - f(b^-, x)|\right)\right).$$

Обозначим $y = (z_1, \ldots, z_n) \in R^n$, $u^+ = u^+(y) = \int_{A_j} |f_n(x; y) - f(b^+, x)|$,

$u^- = \int_{A_j} |f_n(x; y) - f(b^-, x)|$, $f^+(y) = \prod_{i=1}^{n} f(b^+, z_i)$ и $f^-(y) =$

---

[1]) Точнее $b^+ = 0, b_1 \ldots b_{j-1} 1 b_{j+1} \ldots$ и $b^- = 0, b_1 \ldots b_{j-1} - 1 b_{j+1} \ldots$ в двоичной записи. — *Прим. перев.*

[2]) Здесь учтено, что j-й двоичный знак B_j равномерно распределенной случайной величины B принимает значения $+1$ и -1 с равными вероятностями. — *Прим. перев.*

$= \prod_{i=1}^{n} f(b^-, z_i)$. Тогда последнее выражение ограничено снизу величиной

$$\frac{1}{2}\left(\int u^+(y) f^+(y)\,dy + \int u^-(y) f^-(y)\,dy\right) \geqslant$$

$$\geqslant \frac{1}{2}\int (u^+(y) + u^-(y)) \min(f^+(y),\ f^-(y))\,dy \geqslant$$

$$\geqslant \frac{1}{2}\, u \int \min(f^+,\ f^-) \geqslant \frac{1}{4}\, u \exp\left(-\int f^+ \log\left(\frac{f^+}{f^-}\right)\right), \qquad (7)$$

где $u = \int\limits_{A_j} |f(b^+, x) - f(b^-, x)|$ не зависит от b_i, $i \neq j$. Последнее неравенство в (7) доказано в теореме 8.2. Показатель степени в (7) равен $n \int f(b^+, x) \log (f(b^-, x)/f(b^+, x))$. Кроме того, так как

$$\left|(1+z)\log\left(\frac{1-z}{1+z}\right) - 2z\right| \leqslant \frac{2z^2}{1-|z|},$$

то, записывая f на A_j в виде $g(1 + b_j g_1)$, получим

$$\int f^+ \log\left(\frac{f^-}{f^+}\right) = n \int\limits_{A_j} g(1+g_1) \log\left(\frac{1-g_1}{1+g_1}\right) \geqslant$$

$$\geqslant \int\limits_{A_j} 2gg_1 - \int\limits_{A_j} \frac{2gg_1^2}{1-|g_1|} \geqslant -\frac{4nu_j^2 h_j}{1-u_j}.$$

Учитывая, что $u = 2\int |g_1| \geqslant 2u_j h_j$, объединим все предыдущие границы в одну и заметим, что (6) не меньше, чем

$$\sum_j \frac{1}{2}\, u_j h_j \exp\left(-\frac{4nu_j^2 h_j}{1-u_j}\right). \qquad (8)$$

Положим $u_j = u$, $h_j = h$, $1 \leqslant j \leqslant N$, так что $6hN = 1$, $u/h^s = a > 0$. При таком выборе параметров можно вычислить $D_s(f)$ при каждом b. В частности, всегда

$$\sqrt{1-u}\int \sqrt{g} \leqslant \int \sqrt{f} \leqslant \sqrt{1+u}\int \sqrt{g},$$

$$\int |f^{(s)}| = (1 + 2Nuh^{1-s}) \int |g^{(s)}| = \left(1 + \frac{a}{3}\right) \int |g^{(s)}|$$

и, таким образом,

$$D_s(f) \to D_s(g) \left(1 + \frac{a}{3}\right)^{1/(2s+1)},$$

поскольку полагаем $N \to \infty$, $u \to 0$. Минимум нижней границы (8), которую можно переписать в виде $(u/12)\exp(-4nu^2 \times \times (u/a)^{1/s}/(1-u))$, достигается приблизительно при $u = ah^s$, $h^{2s+1} = (s/(2s+1))/4na^2$, и соответствующее значение этой нижней границы равно

$$\frac{1}{12}\left(\frac{s}{4ne(2s+1)}\right)^{s/(2s+1)} a^{1/(2s+1)} \geqslant \frac{1}{12} e^{-4} a^{1/(2s+1)} n^{-s/(2s+1)}.$$

Чтобы все плотности $f(b, \cdot)$ при достаточно больших n принадлежали $F_{s,r}$, должно выполняться неравенство $(1 + a/3)^{1/(2s+1)} \times \times D_s(g) \leqslant r$. Поэтому выберем $a = 3((r/D_s(g))^{2s+1} - 1)$ (эта величина, конечно, положительна при $r > D_s(g)$). Тогда в силу неравенств $3^{1/(2s+1)}/12 \geqslant 2^{-4}$ и $((r/D_s(g))^{2s+1} - 1)^{1/(2s+1)} \geqslant r/D_s(g) - 1$ нижняя граница равна

$$(2e)^{-4}\frac{r/D_s(g) - 1}{n^{s/(2s+1)}}.$$

Эта граница справедлива при всех $r > D_s(g)$ и всех достаточно больших n[1]).

7. Доказательство теоремы 4

Пусть $\tilde{f}_n$ — некоторая оценка плотности. Тогда оценка

$$f_n(x) = \tilde{p}g(x) + (1 - \tilde{p})g(x+2),$$

где $\tilde{p} = \int\limits_{[0,1]} \tilde{f}_n$, лучше, чем $\tilde{f}_n$, поскольку

$$\int |f_n - f| = \int\limits_{[0,1]} g|\tilde{p} - p| + \int\limits_{[0,1]} g|(1-\tilde{p}) - (1-p)| =$$

$$= \left|\int\limits_{[0,1]} \tilde{f}_n - f\right| + \left|\int\limits_{([0,1])^c} \tilde{f}_n - f\right| \leqslant \int |\tilde{f}_n - f|.$$

Иными словами, при выводе нижних границ можно, не нарушая общности, предполагать, что $f_n(x) = \tilde{p}g(x) + (1 - \tilde{p})g(x+2)$, где $\tilde{p} = \tilde{p}(X_1, \ldots, X_n)$ — борелевская функция своих аргументов, принимающая значения из $[0, 1]$. Положим теперь $N = \sum\limits_{i=1}^{n} I_{[X_i \in [0,1]]}$. Тогда

$$E\left(\int |f_n - f|\right) = 2E(|\tilde{p} - p|) \geqslant 2E(|E(\tilde{p}|N) - p|) =$$

$$= 2E(|\psi(N) - p|),$$

[1]) Ясно, что существует функция $g = g_0 * g_1$, такая, что $D_s(g) \leqslant r^*$, где $0 < r^* < \infty$ — некоторая постоянная. — *Прим. ред.*

где ψ — некоторая борелевская функция со значениями из $[0, 1]$. Здесь мы воспользовались условным вариантом неравенства Йенсена и тем, что N — достаточная статистика для p.

Обозначим E_p математическое ожидание по распределению вектора $(X_1, \ldots, X_n)$ в случае, когда параметр смеси для f равен p. Пусть N_p и N_q — случайные величины, имеющие биномиальные распределения с параметрами (n, p) и (n, q) соответственно, где $0 \leqslant p \leqslant q \leqslant 1$. Основным неравенством, на которое опирается дальнейшее доказательство, является

$$\frac{1}{2}\left(E_p\left(\int |f_n - f|\right) + E_q\left(\int |f_n - f|\right)\right) \geqslant |p - q| P(N_q \leqslant np).$$

Это неравенство легко доказать следующим образом:

$$\frac{1}{2}\left(E_p\left(\int |f_n - f|\right) + E_q\left(\int |f_n - f|\right)\right) \geqslant$$

$$\geqslant E(|\psi(N_p) - p|) + E(|\psi(N_q) - q|) \geqslant$$

$$\geqslant \sum_{j \leqslant np} (|\psi(j) - p| P(N_p = j) + |\psi(j) - q| P(N_q = j)) \geqslant$$

$$\geqslant \sum_{j \leqslant np} (|\psi(j) - p| + |\psi(j) - q|) P(N_q = j) \geqslant$$

$$\geqslant |p - q| P(N_q \leqslant np).$$

Часть (i) теоремы 4 получается непосредственно из нашего основного неравенства с помощью рандомизации. Пусть случайная величина p распределена на $[0, 1]$ с вероятностной мерой μ. Тогда

$$\sup_p E_p\left(\int |f_n - f|\right) \geqslant \int E_p\left(\int |f_n - f|\right) \mu(dp).$$

Пусть половина массы меры μ сосредоточена в точке $a = 1/2$ и половина — в точке $b = 1/2 + c/\sqrt{n}$, где c — некоторая постоянная. В силу основного неравенства замечаем, что

$$\sup_p E_p\left(\int |f_n - f|\right)\sqrt{n} \geqslant cP(N_b \leqslant na) \sim c\Phi(-2c),$$

где Φ — функция нормального распределения с параметрами $(0, 1)$. Последнее соотношение следует из центральной предельной теоремы. Максимальное значение нижней границы $c\Phi(-2c)$ равно 0,0849856 ... ; оно достигается при $2c = 0{,}7517915241 \ldots$. Если требуется граница при каком-либо значении n, то можно поступить различными способами; например, предположим, что

$c = 19/100$, $n \geqslant 4$ (такой выбор нужен для дальнейшего). Заметим, что $n(1-b) \leqslant [na] \leqslant nb$. Slud (1977) показал, что

$$cP(N_b \leqslant na) \geqslant c\Phi\left(\frac{[na]-nb}{\sqrt{nb(1-b)}}\right).$$

При четных n это выражение больше или равно

$$c\Phi(-c/\sqrt{(1/4)-c/2\sqrt{n}}) \geqslant c\Phi(-c/\sqrt{(1-c)/4}) = c\Phi(-2c/\sqrt{1-c}).$$

При нечетных n нижняя граница дается величиной $c\Phi((-2c - 1/\sqrt{n})/\sqrt{1-c}) \geqslant c\Phi((-2c-1/2)/\sqrt{1-c})$. Нижняя граница при нечетных n меньше, чем при четных, и равна $(19/100) \times \Phi(-44/45) \geqslant (19/100)\Phi(-1) = 0{,}030153\ldots$. Утверждение (i) доказано.

Для доказательства утверждения (ii) обозначим $J_n(p) = E_p\left(\int |f_n - f|\right)\sqrt{n}$ и предположим, что p — случайная величина с равномерной вероятностной мерой μ на $[1/2-\varepsilon, 1/2+\varepsilon]$, где $\varepsilon > 0$ — некоторое малое число. Кроме того, обозначим $q = c/\sqrt{n}$ и выберем c так же, как в доказательстве утверждения (i), т. е. положим $2c = 0{,}7517915241\ldots$. Пусть $a > 0$ — произвольная постоянная. По лемме Фату

$$\begin{aligned}
\sup_p \limsup_{n\to\infty} J_n(p) &\geqslant E\left(\limsup_{n\to\infty} J_n(p)\right) \geqslant \\
&\geqslant E\left(\limsup_{n\to\infty} \min(J_n(p), a)\right) \geqslant \\
&\geqslant \limsup_{n\to\infty} E(\min(J_n(p), a)) = \\
&= \limsup_{n\to\infty} (E(\min(J_n(p), a)) + E(\min(J_n(p\oplus q), a)))/2 \geqslant \\
&\geqslant \limsup_{n\to\infty} E(\min((J_n(p) + J_n(p\oplus q))/2, a))/2 \geqslant \\
&\geqslant \limsup_{n\to\infty} E(\min(cP(N_{p+q} \leqslant np \mid p), a) I_{p+q\leqslant 1/2+\varepsilon})/2,
\end{aligned}$$

где величина $p \oplus q$ по определению равна $p+q$, если эта сумма меньше, чем $1/2+\varepsilon$, и $p+q+2k\varepsilon$ при некоторых других целых k, таких, что $p+q+2k\varepsilon$ принадлежит отрезку $[1/2-\varepsilon, 1/2+\varepsilon]$. Символом N_{p+q} обозначается случайная величина, имеющая биномиальное распределение с параметрами $(n, p+q)$. Если положить $a = c$, то последний член рассматриваемой цепочки неравенств равен

$$\limsup_{n\to\infty} E(cP(N_{p+q} \leqslant np \mid p) I_{p+q\leqslant 1/2+\varepsilon})/2.$$

Теперь в силу центральной предельной теоремы $cP(N_{p+q} \leqslant np \mid p) \to c\Phi(-c/\sqrt{p(1-p)})$ при фиксированном p, таком,

что $p \leqslant 1/2 + \varepsilon$. Таким образом, теорема о мажорируемой сходимости влечет за собой, что рассматриваемый верхний предел не меньше, чем $cE\left(\Phi\left(-c/\sqrt{p(1-p)}\right)\right)/2$. Так как $|p - 1/2| \leqslant \varepsilon$ и ε можно выбирать, то этот предел можно сделать сколь угодно близким к $c\Phi(-2c)/2 = 0{,}0424928\ldots$. Теорема 4 доказана.

8. Доказательства теорем 5—9

Доказательство теоремы 5. Снова будем действовать путем рандомизации, используя равномерное распределение на множестве всех возможных векторов $b = (b_1, \ldots, b_r)$. Обозначим через $\sum_b$ сумму по всем таким векторам, через b_{i+} и b_{i-} — r-векторы, совпадающие с b во всех компонентах, кроме, быть может, i-й компоненты, которая для b_{i+} всегда равна $+1$, а для b_{i-} всегда равна -1. Обозначим через $\mathbf{X}_n$ вектор $(X_1, \ldots, X_n) \in R^n$ и через $\mathbf{x}_n$ — вектор $(x_1, \ldots, x_n) \in R^n$. Наконец, все произведения $\prod$ будем брать по $j = 1, \ldots, n$, например $\prod f(b, x_j)$. Оценку плотности $f_n(x)$, $x \in R$, будет более удобно записывать в виде $f_n(x, \mathbf{X}_n)$, указывая явно зависимость от данных — выборки $\mathbf{X}_n$, состоящей из n независимых одинаково распределенных случайных величин с плотностью $f(b, x)$. Имеем

$$\sup_b E\left(\int |f_n - f(b, \cdot)| \geqslant\right.$$

$$\geqslant 2^{-r} \sum_b \int\int |f_n(x, \mathbf{x}_n) - f(b, x)|\, dx \cdot \prod f(b, x_j)\, d\mathbf{x}_n \geqslant$$

$$\geqslant 2^{-r} \sum_b \int \sum_{i=1}^{r} \int_{A_i} |f_n(x, \mathbf{x}_n) - f(b, x)|\, dx \cdot \prod f(b, x_j)\, d\mathbf{x}_n =$$

$$= 2^{-r} \sum_b \int \sum_{i=1}^{r} \frac{1}{2} \left(\int_{A_i} |f_n(x, \mathbf{x}_n) - f(b_{i+}, x)|\, dx \cdot \prod f(b_{i+}, x_j) + \right.$$

$$\left. + \int_{A_i} |f_n(x, \mathbf{x}_n) - f(b_{i-}, x)|\, dx \cdot \prod f(b_{i-}, x_j) \right) d\mathbf{x}_n \geqslant$$

$$\geqslant 2^{-r} \sum_b \int \sum_{i=1}^{r} \frac{\alpha}{2} \min\left(\prod f(b_{i+}, x_j), \prod f(b_{i-}, x_j)\right) d\mathbf{x}_n \geqslant$$

$$\geqslant \frac{r\alpha}{2} \inf_{b,\, i} \int \min\left(\prod f(b_{i+}, x_j), \prod f(b_{i-}, x_j)\right) d\mathbf{x}_n.$$

Далее, известно, что если f и g — плотности, то $\int \min(f, g) \geqslant \geqslant \left(\int \sqrt{fg}\right)^2 / 2$ (теорема 8.5) и $\int \min(f, g) = 1 - \int |f - g|/2 \geqslant \geqslant \left(\int (\sqrt{f} - \sqrt{g})^2\right)^{1/2}$ (теорема 8.4). Но

$$\int \sqrt{\prod f(b_{i+}, x_j) \prod f(b_{i-}, x_j)}\, d\mathbf{x}_n = \prod \int \sqrt{f(b_{i+}, x_j) f(b_{i-}, x_j)}\, dx_j \geqslant \beta^n,$$

что доказывает первую часть теоремы 5.

Для доказательства второй части заметим, что $\alpha = 2\int\limits_A |g|$ и $\gamma \leqslant \int\limits_A g^2$. Это следует из того, что

$$1 - \int \sqrt{f(b_{i+}, x) f(b_{i-}, x)} = \int\limits_{A+y_i} \left(f_0 - \sqrt{f_0 + g}\sqrt{f_0 - g}\right) =$$

$$= \int\limits_{A+y_i} \left(1 - \sqrt{1 - g^2}\right) \leqslant \int\limits_A g^2$$

при любых b и i.

Замечание. Первая часть доказательства теоремы 5 в основных чертах совпадает с цепочкой неравенств (6)—(7) в доказательстве теоремы 3 (Bretagnolle, Huber (1979)).

Доказательство теоремы 6. Прежде всего выберем неотрицательную функцию g_0, содержащуюся в $W(s, \alpha, C)$ и такую, что

$$\int\limits_0^1 |g_0| \geqslant \gamma_1 C, \quad \int\limits_0^1 g_0^2 \leqslant \gamma_2 C^2, \quad \sup g_0 \leqslant \gamma_3 C.$$

Затем так же, как в теореме 5, построим семейство функций следующим образом. Возьмем постоянную $a_0 > 1$. Положим $A = [0, 1/a_0 r)$ и

$$g(x) = \begin{cases} \dfrac{g_0(2a_0 r x)}{(2a_0 r)^{s+\alpha}}, & 0 \leqslant x \leqslant \dfrac{1}{2a_0 r}, \\ \dfrac{-g_0(2a_0 r(x - 1/(2a_0 r)))}{(2a_0 r)^{s+\alpha}}, & \dfrac{1}{2a_0 r} \leqslant x < \dfrac{1}{a_0 r}. \end{cases}$$

Пусть $y_i = (1 - 1/a_0)/2 + i/a_0 r$, $i = 0, 1, \ldots, r-1$. Тогда $\bigcup\limits_i A + y_i = [(1 - 1/a_0)/2, (1 + 1/a_0)/2]$, т. е. имеется разбиение последнего интервала на r равных частей длины $1/a_0 r$ каждая. Функция f_0 из теоремы 5 равна 1 на этом большом центральном интервале, а вне его ее следует аккуратно доопределить. В част-

ности, нужно обеспечить выполнение следующих условий: $f_0 = 0$ вне $[0, 1]$, $\int_0^1 f_0 = 1$, $f_0 \geqslant 0$, $f_0^{(i)}(0) = 0$ при $i = 0, 1, \ldots, s-1$, $f_0((1 - 1/a_0)/2) = 1$ и $f_0^{(i)}((1 - 1/a_0)/2) = 0$, $i = 1, \ldots, s-1$. Кроме того, функция f_0 должна принадлежать $W(s, \alpha, C)$ и должна быть симметричным образом доопределена на $[(1 + 1/a_0)/2, 1]$. Это всегда можно сделать, выбирая C просто выше некоторого порога $c_1(s, \alpha, a_0)$. Далее будем полагать $a_0 = 2$ (для этого нет никаких особых оснований, кроме удобства), что позволяет ввести условие $C \geqslant c_1(s, \alpha)$. Давайте быстро проверим, что семейство из теоремы 5 действительно содержится в $W(s, \alpha, C)$. Прежде всего

$$g^{(s)}(x) = \frac{g_0^{(s)}(2a_0 rx)}{(2a_0 r)^{\alpha}}, \quad 0 \leqslant x \leqslant \frac{1}{2a_0 r},$$

так что

$$|g^{(s)}(x) - g^{(s)}(y)| \leqslant (2a_0 r)^{-\alpha} C (2a_0 r)^{\alpha} |x - y|^{\alpha} = C|x - y|^{\alpha}.$$

Кроме того,

$$\int_A |g| = 2\int_0^1 \frac{|g_0|}{(2a_0 r)^{s+\alpha+1}} \geqslant \frac{2\gamma_1 C}{(2a_0 r)^{s+\alpha+1}},$$

$$\int_A g^2 = 2\int_0^1 \frac{g_0^2}{(2a_0 r)^{2(s+\alpha)+1}} \leqslant \frac{2\gamma_2 C^2}{(2a_0 r)^{2(s+\alpha)+1}}$$

и

$$\sup_A |g| = \sup_{[0,1]} \frac{|g_0|}{(2a_0 r)^{s+\alpha}} \leqslant \frac{\gamma_3 C}{(2a_0 r)^{s+\alpha}}.$$

Нижняя граница из теоремы 5 принимает теперь вид

$$r\int_A |g|/2 \geqslant \gamma_1 C r (2a_0 r)^{-s-\alpha-1} = \gamma_1 C (2a_0)^{-s-\alpha-1} r^{-(s+\alpha)}$$

при условии, что

$$n\int_A g^2 \leqslant \frac{2n\gamma_2 C^2}{(2a_0 r)^{2(s+\alpha)+1}} \leqslant \frac{1}{8}$$

и

$$\frac{\gamma_3 C}{(2a_0 r)^{s+\alpha}} \leqslant 1.$$

Первое из этих дополнительных условий используется для определения r. Наилучший результат получается, если положить

$$r = \lceil (2a_0)^{-1} (16\gamma_2 n C^2)^{1/(2(s+\alpha)+1)} \rceil,$$

где $\lceil x \rceil$ — ближайшее целое, большее или равное x. Тогда второе условие становится нижней границей для n. Оно выполняется, например, если

$$(16\gamma_2 nC^2)^{(s+\alpha)/(2(s+\alpha)+1)} \geqslant \gamma_3 C,$$

что эквивалентно условию $n \geqslant Cc_2$. Подставляя $a_0 = 2$ и полученное значение r в нижнюю границу, получим асимптотический результат, указанный в теореме 6. Используя тот факт, что $\lceil y \rceil$ не больше, чем $y + 1$, получим нижнюю границу

$$\frac{\gamma_1 C/4}{((16\gamma_2 nC^2)^{1/(2s+2\alpha+1)} + 4)^{s+\alpha}},$$

которая верна при всех $n \geqslant Cc_2$.

Остается указать исходную функцию g_0. Мы умышленно рассмотрим не оптимальную, но удобную функцию g_0, а именно

$$g_0(x) = C\gamma_0 (x(1-x))^{s+\alpha}.$$

Без труда убеждаемся, что $\int_0^1 |g_0| = C\gamma_1$, $\int_0^1 g_0^2 = C^2\gamma_2$ и $\sup g_0 = C\gamma_3$. Таким образом, теорема 6 будет доказана, если показать, что g_0 принадлежит классу $W(s, \alpha, C)$. Используя обозначение

$$\binom{u}{j} = \begin{cases} 1 \text{ при } j = 0, \\ \dfrac{u(u-1)\ldots(u-j+1)}{j!} \quad \text{при } j > 0, \ u \in R, \end{cases}$$

и теорему о биномиальном разложении, убеждаемся, что

$$\frac{g_0(x)}{C\gamma_0} = \sum_{j=0}^{\infty} (-1)^j \binom{s+\alpha}{j} x^{s+\alpha+j}.$$

Функция $g_0(x)/(C\gamma_0)$ имеет s-ю производную $\sum\limits_{j=0}^{\infty} (-1)^j h_j(x)$, где

$$h_j(x) = \binom{s+\alpha}{j} (s+\alpha+j)\ldots(1+\alpha+j) x^{\alpha+j}.$$

Теперь при $0 \leqslant x \leqslant y \leqslant 1/2$ имеем

$$|h_j(x) - h_j(y)| = \binom{s+\alpha}{j} (s+\alpha+j)\ldots(1+\alpha+j) |x^{\alpha+j} - y^{\alpha+j}|$$

и

$$|x^{\alpha+j} - y^{\alpha+j}| \leqslant \begin{cases} |y-x|^{\alpha}, & j = 0, \\ (\alpha+j) y^{\alpha+j-1} |y-x|, & j > 0. \end{cases}$$

Объединяя все эти оценки в общую верхнюю границу, получим

$$|g_0^{(s)}(x) - g_0^{(s)}(y)|/C\gamma_0 \leqslant (s+\alpha)\ldots(1+\alpha)|y-x|^\alpha +$$

$$+ \sum_{j=1}^{\infty} \binom{s+\alpha}{j} (s+\alpha+j)\ldots(\alpha+j)\, y^{\alpha+j-1}|y-x| \leqslant$$

$$\leqslant |y-x|^\alpha \left((s+\alpha)\ldots(1+\alpha) + \sum_{j=1}^{\infty} \frac{(s+\alpha+j)^{s+1}(s+\alpha)^j}{2^{\alpha+j-1} j!} \right) \leqslant$$

$$\leqslant |y-x|^\alpha \left((s+\alpha)\ldots(1+\alpha) + \right.$$

$$\left. + \sum_{j=1}^{\infty} (s+\alpha)^{s+1} 2^{1-\alpha} \frac{((s+\alpha)\, e^{(s+1)/(s+\alpha)}/2)^j}{j!} \right) \leqslant$$

$$\leqslant |y-x|^\alpha (s+\alpha)^{s+1} 2^{1-\alpha} \exp\left(\frac{s+\alpha}{2} \exp\left(\frac{s+1}{s+\alpha}\right)\right),$$

где мы воспользовались неравенством $(1+u) \leqslant e^u$, $u \in R$. Но последнее выражение в цепочке неравенств равно $|y-x|^\alpha/\gamma_0$, что и требовалось доказать.

Доказательство теоремы 7. В первой части, относящейся к классу $W(0, 1, C)$, будет показано, как можно оптимизировать неравенство из теоремы 5 с целью получения хороших границ. Поначалу будем следовать доказательству теоремы 6, используя функцию g_0 на $[0, 1]$ вида

$$g_0(x) = \begin{cases} Cx, & 0 \leqslant x \leqslant 1/2, \\ C(1-x), & 1/2 \leqslant x \leqslant 1. \end{cases}$$

Очевидно, $\int_0^1 |g_0| = C/4$, $\int_0^1 g_0^2 = C^2/12$ и $\sup g_0 = C/2$. Таким образом, постоянные γ_1, γ_2 и γ_3 из теоремы 6 формально можно заменить на 1/4, 1/12 и 1/2. Определим функцию g через g_0 так же, как в теореме 6. Будем считать, что и величины $a_0 > 1$, $r \geqslant 1$ имеют такой же смысл, как в теореме 6. Тогда по теореме 5

$$\sup_{f \in W(0,1,C)} E\left(\int |f_n - f|\right) \geqslant r \int_A |g| \left(1 - \sqrt{2n \int_A g^2}\right) =$$

$$= \frac{\gamma_1 C}{a_0} \left(\frac{1}{y} - \frac{L}{y^{5/2}}\right)$$

при достаточно больших C, где $L = \sqrt{4\gamma_2 C^2 n}$ и $y = 2a_0 r$. Эта граница справедлива при $y \geqslant \gamma_3 C$. Если рассматривать данную нижнюю границу как функцию только от y, то она достигает максимума при $y = (5L/2)^{2/3}$, принимая значение

$$\frac{1}{a_0}\gamma_1 C \cdot \frac{3}{5}\left(\frac{2}{5L}\right)^{2/3} = \frac{3}{20a_0}\left(\frac{12C}{25n}\right)^{1/3}.$$

Небольшое неудобство доставляет только множитель a_0, который необходимо выбрать так, чтобы решение r уравнения $2a_0 r = y$ было целым числом.

Пусть b_0 — некоторая другая постоянная, большая единицы, и пусть $r = \lceil y/2b_0 \rceil$. Тогда

$$a_0 = y/(2\lceil y/2b_0 \rceil) \in [b_0,\ b_0/(1 - 2b_0/y)] \subseteq [b_0,\ b_0/(1 - u_0)]$$

при $y \geqslant 2b_0/u_0$. Наша нижняя граница теперь равна

$$\frac{3}{20}\,\frac{1-u_0}{b_0}\left(\frac{12C}{25n}\right)^{1/3}$$

при условиях

(i) $y \geqslant \gamma_3 C$ (это эквивалентно тому, что $n \geqslant 3C/50$),
(ii) $y \geqslant 2b_0/u_0$.

Кроме того, построив C-липшицеву функцию f_0 на $[0,\ (1 - 1/a_0)/2]$, получим, так же как в теоремах 5 и 6, нижнюю границу для C. Рассмотрим функцию, состоящую из двух линейных кусков с изломом в точке $(1 - 1/a_0)/4$ (где значение этой функции равно 3/2) и концами в точках 0 и $(1 - 1/a_0)/2$ (где значения функции, очевидно, равны 0 и 1). Легко проверить, что

$$\int_0^{(1/2)\,(1-1/a_0)} f_0 = \frac{1}{2}\left(1 - \frac{1}{a_0}\right)$$

и что функция f_0 является C-липшицевой при $(1 - 1/a_0)\,C/4 \geqslant 3/2$, т. е. при $C \geqslant 6/(1 - 1/a_0)$. Последнее условие всегда выполнено, если

(iii) $C \geqslant 6/(1 - 1/b_0)$.

Положим $b_0 = 12/11$, $u_0 = 1/22$. Из условия (iii) следует, что $C \geqslant 72$. Условие (ii) влечет за собой, что $n \geqslant 12 \cdot 48^3/(25C^2)$. Если $C \geqslant 72$, то это неравенство выполнено при $n \geqslant 10$. Подстановка значений u_0 и b_0 в нижнюю границу приводит к уменьшению коэффициента 3/20 до 21/160. Этим завершается доказательство первой части теоремы 7.

Прежде чем приступить к доказательству второй части, отметим, что мы нашли компромисс между нижней границей, малой в асимптотике, и нижней границей, пригодной при небольших

значениях n. Заинтересованные читатели могут теперь без особого дополнительного труда получить свои собственные границы для интересующего их диапазона значений. Например, можно уменьшить нижнюю границу для C за счет введения «уровня постоянства» функции f_0, значительно превосходящего 1. Это может оказаться полезным, хотя и потребует изменений во всех теоремах, включая теорему 5. Для $W(1, 1, C)$ можно решить задачу, подобную небольшой оптимизационной задаче, решенной при выводе границы для $W(0, 1, C)$, но мы поступим здесь иначе. Вместо этого воспользуемся прямым методом, в котором значения $a_0 r$ определяются из уравнения $n\int_A g^2 = 1/8$ и подставляются в выражение $r\int_A |g|/2$.

Для класса $W(1, 1, C)$ начнем с определения функции g_0 на $[0, 1]$, имеющей вид

$$g_0(x) = \begin{cases} Cx^2/2, & 0 \leqslant x \leqslant 1/4, \\ (C/8 - C(x - 1/2)^2)/2, & 1/4 \leqslant x \leqslant 3/4, \\ C(1 - x)^2/2, & 3/4 \leqslant x \leqslant 1. \end{cases}$$

Ясно, что функция $g_0^{(1)}$ является C-липшицевой и что g_0 удовлетворяет всем требованиям, которые предъявлялись к ней в доказательстве теоремы 6. Далее, $\int |g_0| = C/32$, $\int g_0^2 = 23C/(60 \cdot 256)$, $\sup g_0 = C/16$, так что в теореме 6 можно положить $\gamma_1 = 1/32$, $\gamma_2 = 23/(256 \cdot 60)$ и $\gamma_3 = 1/16$.

Используя это, так же как в теореме 6, вычислим постоянные

$$c_3 = (30/23)^{2/5}/32, \quad c_2 = 15/368.$$

Таким образом, эта часть теоремы 7 следует непосредственно из теоремы 6. Для удобства пользователя вычислим постоянную c_1 в явном виде. Выше было принято $a_0 = 2$. Необходимо соединить точки $(0, 0)$ и $((1 - 1/a_0)/2, 1) = (1/4, 1)$ с помощью функции f_0 из $W(1, 1, C)$ со средним значением 1 на рассматриваемом интервале. Утверждается, что это можно сделать при $C \geqslant 288$. Сначала определим на базовом отрезке длины $2a$ следующую функцию:

$$Cx^2/2, \quad 0 \leqslant x \leqslant a,$$
$$Ca^2 - C(x - 2a)^2/2, \quad a \leqslant x \leqslant 2a.$$

Эта функция имеет C-липшицеву производную на отрезке $[0, 2a]$ и нулевые производные на обоих его концах. Площадь под этой функцией (интеграл от 0 до $2a$) равна Ca^3, максимальное значение функции достигается в точке $2a$ и равно Ca^2. Если взять такую же функцию, но перевернутую, и с постоянной D вместо C, где $D \leqslant C$, и приклеить ее справа к первой функции, выбирая для нового

куска определенную длину b, то получим искомую функцию f_0 при условии, что разрешимы следующие уравнения:

(i) $2a + 2b = 1/4$,
(ii) $Ca^3 + Db^3 + 2b = 2a + 2b$ (условие на среднее значение),
(iii) $Ca^2 - Db^2 = 1$ (правильное значение на конце),

где $D \leqslant C$, $a \geqslant 0$, $b \geqslant 0$. Решение a равно

$$\frac{4}{C}\left(1 + \sqrt{1 + \frac{C}{16}}\right).$$

Эта величина не превосходит 1/8 (и, таким образом, $b \geqslant 0$) при $C \geqslant 128$. Наконец, $D = 64 \cdot 8a/(1 - 8a)^2 \leqslant C$ тогда и только тогда, когда $C/64 \geqslant \sqrt{1 + C/16}$, а последнее условие выполнено при $C \geqslant 288$. Теорема 7 доказана.

Доказательство теоремы 8. Из теоремы 5 следует, что уже для подсемейства класса $Q_r(g)$ с фиксированным значением ε имеется нижняя граница $(r/2)\,\varepsilon \int |g|$ при условии, что $n\varepsilon^2 \int g^2 \leqslant 1/8$. Полагая теперь $\varepsilon = \int (8n\, g^2)^{-1/2} \leqslant 1$, получим первое неравенство. Вторая граница следует из того, что при данном выборе g имеем $\int |g| = \int g^2 = 1/r$. Третья граница следует из второй при $r = 8n$. Заметим, что первое неравенство также можно было получить, используя более сильную половину теоремы 5 и проводя оптимизацию так же, как в доказательстве теоремы 7.

Доказательство теоремы 9. Применим теорему 5 в общей формулировке. Семейство $f(b, \cdot)$ построим с помощью центральной функции f_0, такой, что $f_0(x) = B(1 - 2i\delta)$, $(i - 1)/rB \leqslant x \leqslant i/rB$, $i = 1, \ldots, r$, и $f_0(x) = 0$ в остальной части отрезка $[0, 1]$. Здесь $\delta = 1/2(2r + 1)$. Площадь под этой функцией, т. е. интеграл от нее в пределах от 0 до $1/B$, равна

$$\sum_{i=1}^{r} (B - 2i\delta)(rB)^{-1} = 1 - \delta(r + 1)/B = 1 - (r + 1)/(2B(2r + 1)).$$

Возьмем теперь некоторое малое возмущение, такое, как функция $g_0(x)$, равная δB на $[0, 1/2rB)$ и равная $-\delta B$ на $[1/2rB, 1/rB)$. Интеграл от этой функции в пределах от 0 до $1/rB$ равен 0. Так как она меняется от δB до $-\delta B$, то можно складывать ее кусочно с f_0, не нарушая монотонности; определим $f(b, \cdot)$ на $[0, 1/B)$ следующим образом:

$$f(b, x) = \begin{cases} f_0(x) + g\left(x + \dfrac{i-1}{rB}\right) & \text{на } \left[\dfrac{i-1}{rB}, \dfrac{i}{rB}\right) \quad \text{при } b_i = 1, \\ f_0(0) & \text{на } \left[\dfrac{i-1}{rB}, \dfrac{i}{rB}\right) \quad \text{при } b_i = -1. \end{cases}$$

Таким образом, при любом b функция $f(b,\cdot)$ монотонна на $[0, 1/B)$ и $f(0) \leqslant B$. Теперь распределим остаток вероятностной массы равномерно на $[1/B, 1)$. Значение плотности на этом равномерном участке равно

$$\frac{\delta(r+1)/B}{1-1/B} = \frac{\delta(r+1)}{B-1}.$$

На $[0, 1/B)$ значения $f(b, \cdot)$ больше или равны $B(1 - (2r+1) \times \times \delta) = B/2$. Чтобы имела место монотонность, должно выполняться неравенство $2\delta(r+1) \leqslant B(B-1)$, или $(r+1)/(2r+1) \leqslant B(B-1)$. Оно выполняется при любом целом r, если $B \geqslant 2$. Теперь построение закончено. Требуется только вычислить α и γ из теоремы 5. В качестве α можно взять $\int_0^{1/rB} |g| = \delta/r = 1/(2r(2r+1))$. Кроме того, можно положить $\gamma = \delta^2/(2r(1-2r\delta))$, что явствует из следующего неравенства, справедливого при любых i и b:

$$1 - \int \sqrt{f(b_{i+},x) f(b_{i-},x)} = \int_{(i-1)/rB}^{i/rB} (f_0 - \sqrt{f_0(f_0+g)}) \leqslant$$

$$\leqslant \frac{B(1-2i\delta)}{2rB}\left(1 - \sqrt{1 + \frac{\delta B}{B(1-2i\delta)}} + 1 - \sqrt{1 - \frac{\delta B}{B(1-2i\delta)}}\right) \leqslant$$

$$\leqslant \frac{1-2i\delta}{2r}\left(\frac{\delta B}{B(1-2i\delta)}\right)^2 =$$

(в силу неравенства $\sqrt{1+u} + \sqrt{1-u} \geqslant 2 - u^2$, $|u| \leqslant 1$)

$$= \frac{\delta^2}{2r(1-2i\delta)} \leqslant \frac{\delta^2}{2r(1-2r\delta)}.$$

Нижняя граница равна

$$\frac{r\alpha}{4}(1-\gamma)^{2n} = \frac{\delta}{4}\left(1 - \frac{\delta^2}{2r(1-r/(2r+1))}\right)^{2n} \geqslant$$

$$\geqslant \frac{\delta}{4}\left(1 - \frac{1}{r}\delta^2\right)^{2n} \geqslant \frac{\delta}{4}\left(1 - \frac{2n\delta^2}{r}\right).$$

Если $r = \lceil (n/4)^{1/3} \rceil$, то ясно, что $r/\delta^2 = 4r(2r+1)^2 \geqslant 16 \times \times ((n/4)^{1/3})^3 = 4n$, так что нижняя граница не меньше, чем

$$\frac{\delta}{8} = \frac{1}{16(2r+1)} \geqslant \frac{1}{16(2(n/4)^{1/3}+3)},$$

и теорема 9 доказана.

ЛИТЕРАТУРА

Assouad P.
(1983) Deux remarques sur l'estimation, Comptes Rendus de l'Académie des Sciences de Paris, 296, pp. 1021—1024.
Birgé L.
(1980) Thèse, 3e partie, Université de Paris VII, Paris, France, 1980.
(1983) Approximation dans les espaces métriques et théorie de l'estimation, Zeit-

schrift für Wahrscheinlichkeitsteorie und verwandte Gebiete, 65, pp. 181—237.
(1986) On estimating a density using Hellinger distance and some other strange facts, Probability Theory and Related Fields, 71, pp. 271—291.
Boyd D. W., Steele J. M.
(1978) Lower bounds for nonparametric density estimation rates, Annals of Statistics, 6, pp. 932—934.
Bretagnolle J., Huber C.
(1979) Estimation des densités: risque minimax, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 47, pp. 119—137.
Deheuvels P.
(1977a) Estimation non paramétrique de la densité par histogrammes généralisés, Revue de Statistique Appliquée, 25, pp. 5—42.
(1977b) Estimation non paramétrique de la densité par histogrammes généralisés, Publications de l'Institut de Statistique de l'Université de Paris, 22, pp. 1—23.
Devroye L.
(1983) On arbitrarily slow rates of global convergence in density estimation, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 62, pp. 475—483.
Farrel R. H.
(1967) On the lack of a uniformly consistent sequence of estimators of a density function in certain cases, Annals of Mathematical Statistics, 38, pp. 471—474.
(1972) On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, Annals of Mathematical Statistics, 43, pp. 170—180.
Kiefer J.
(1982) Optimum rates for non-parametric density and regression estimates, under order restrictions, in Statistics and Probability: Essays in Honor of C. R. Rao, G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh Eds., North-Holland, Amsterdam, pp. 419—428.
Rosenblatt M.
(1971) Curve estimates, Annals of Mathematical Statistics, 42, pp. 1815—1842.
Slud E. V.
(1977) Distribution inequalities for the binomial law, Annals of Probability, 5, pp. 404—412.
Stone C. J.
(1980) Optimal rates of convergence for nonparametric estimators, Annals of Statistics, 8, pp. 1348—1360.
(1983) Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives, Department of Statistics, University of California, Berkeley, preprint.
Wahba G.
(1975) Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation, Annals of Statistics, 3, pp. 15—29.
Ибрагимов И. А., Хасьминский Р. З.
(1979) Асимптотическая теория оценивания. — М.: Наука.
Колмогоров А. Н., Тихомиров В. Н.
(1959) ε-энтропия и ε-емкость множеств в функциональных пространствах// Успехи матем. наук. — Т. 14. — 2. — С. 3 — 86.
Мания Г. М.
(1969) Квадратическая погрешность оценки плотности многомерного нормального распределения по данным выборки//Теория вероятн. и ее примен. — Т. 14. — 1, С. 151—155.
Надарая Э. А.
(1974) Об интегральной среднеквадратичной ошибке некоторых непараметрических оценок плотности вероятностей//Теория вероятн. и ее примен. — Т. 19. — 1. — С. 131—140.
Самаров А. М.
(1976) О минимаксной границе риска непараметрических оценок плотности// Проблемы передачи информации. — Т. 12 3. — С. 108—111.

КОММЕНТАРИЙ ПЕРЕВОДЧИКА

Первые из известных нижних границ для минимаксных рисков оценок плотности принадлежат Ченцову (1972*) и Farrell (1972). Нижние границы Ченцова построены для L_2-рисков оценок плотностей на абстрактных пространствах и выражены через поперечники классов плотностей. Ченцов (1972*) также показал, что эти нижние границы достигаются по порядку величины на проекционных оценках. Нижние границы Farrell (1972) относятся к поточечным рискам оценок. Правильные по порядку величины нижние границы в метрике L_∞ для классов Гёльдера найдены Хасьминским (1978*), а в метриках L_p, $1 \leqslant p < \infty$, для классов $F_{s,r}$ — Bretagnolle, Huber (1979). Наиболее общим является последующий результат Ибрагимова, Хасьминского (1980*), которые получили нижние и верхние границы для рисков оценок плотности на R^d, $d \geqslant 1$, в метриках L_p, $1 \leqslant p \leqslant \infty$, для широкого класса функций потерь. Границы Ибрагимова, Хасьминского (1980*) охватывают классы Гёльдера, а также классы аналитических плотностей. Нижние и верхние границы для поточечных рисков оценок плотностей и производных плотностей на R^d получены в работе Stone (1980).

Невозможность построения оценок, равномерно состоятельных в L_1 на классах гладких плотностей (теорема 1 (i)), была впервые замечена Ибрагимовым, Хасьминским (1980*). Они показали, что теорема 1 (i) имеет место для класса $\mathcal{F}$ всех плотностей f на R^d, удовлетворяющих условию Гёльдера

$$\int_{R^d} |f(x+h) - f(x)|\, dx \leqslant L |h|^\beta, \quad h \in R^d,$$

с заданными $L > 0$, $0 < \beta \leqslant 1$.

Первое утверждение теоремы 5 (названное в тексте леммой Ассуа) следует из Bretagnolle, Huber (1979, с. 122). Лемма Ассуа (Assouad (1983)) — более общий результат, позволяющий рассматривать также другие метрики и функции потерь. С ее помощью, однако, не удается получить правильные по порядку величины нижние границы в метрике L_∞. Единственным способом получения таких границ пока остается метод Ибрагимова, Хасьминского (1979), основанный на неравенстве Фано (1965*, с. 220).

Можно выделить три общих приема получения непараметрических нижних границ: 1) лемма Ассуа, т. е. сведéние к большому числу задач различения двух гипотез (Bretagnolle, Huber (1979), Assouad (1983)), 2) переход к некоторому «плохому» параметрическому подсемейству класса $\mathcal{F}$ и применение известных информационных нижних границ для этого подсемейства (Ченцов (1972*), Ибрагимов, Хасьминский (1979, с. 319)), 3) сведéние к задаче различения большого числа гипотез с применением неравенства Фано (Хасьминский (1978*), Ибрагимов, Хасьминский (1979, гл. VII)). Последний подход, по-видимому, имеет более широкую область действия, хотя иногда дает более грубые границы.

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Ибрагимов И. А., Хасьминский Р. З.
(1980*) Об оценке плотности распределения//Исследования по матем. статистике IV. Записки научн. семинаров ЛОМИ. — Т. 98. — Л.: Наука. — С. 61—85.

Фано Р.
(1965*) Передача информации. Статистическая теория связи. — М.: Мир.

Хасьминский Р. З.
(1978*) О границе снизу рисков непараметрических оценок плотности в равномерной метрике//Теория вероятн. и ее примен. — Т. 23. — 4. — С. 824—828.

Ченцов Н. Н.
(1972*) Статистические решающие правила и оптимальные выводы. — М.: Наука.

Глава 5

Скорости сходимости в L_1

1. Введение

Следует ожидать, что при конечном наборе данных любая заданная оценка плотности имеет ограниченные возможности даже для самых хороших плотностей f. Эта ограниченность возможностей оценок в определенной степени обнаруживается в нижних границах из гл. 4. В данной главе мы хотим получить очень точную информацию относительно $E(J_n)$ для конкретных оценок плотности, таких, как ядерная и гистограммная оценки. Нас будут в особенности интересовать асимптотические выражения для ошибки $E(J_n)$ этих оценок при $d = 1$.

Рассматриваемая здесь ядерная оценка имеет вид

$$f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

(Parzen (1962), Rosenblatt (1956)), где $h = h_n$ — заданная последовательность положительных чисел и K — заданная плотность (ядро), удовлетворяющая условию

$K(x) = K(-x)$ при любых x, функция K ограниченна и имеет компактный носитель. (2)

Кроме того, ввиду теоремы 3.1 не будут рассматриваться последовательности h, для которых не выполнены условия

$$\lim_{n\to\infty} h = 0, \quad \lim_{n\to\infty} nh = \infty. \tag{3}$$

В данной главе будут исследованы скорости сходимости величины $E(J_n)$ при фиксированной плотности f [1]). Будет показано, что значения этих скоростей тесно связаны с величинами

$$A(K) = \left(\int K^2\right)^{2/5} \left(\int x^2 K\right)^{1/5}$$

и

$$B(f) = \left(\frac{1}{2}\left(\int \sqrt{f}\right)^4 \int |f''|\right)^{1/5}.$$

---

[1]) В оригинале individual rates of convergence. — *Прим. перев.*

Для упрощения записи будем обозначать величины $\sqrt{\int K^2}$ и $\int x^2 K$ символами α и β соответственно. Таким образом, $A(K) = (\alpha^4\beta)^{1/5}$. Величина $B(f)$ будет рассматриваться для любых плотностей f, принадлежащих классу $\mathcal{F}$ функций, таких, что

(i) f абсолютно непрерывна и имеет п. в. производную f',
(ii) f' абсолютно непрерывна и имеет п. в. производную f'',
(iii) f'' непрерывна и ограниченна.

Но так как желательно еще иметь информацию о плотностях f, не принадлежащих $\mathcal{F}$, то определение $B(f)$ надо обобщить. Определим для *любых* f величину

$$B^*(f) = \left(\frac{1}{2}\left(\int \sqrt{f}\right)^4 \sup_{h>0} \int |(f * \varphi_h)''|\right)^{1/5},$$

где $*$ — оператор свертки, φ — плотность с компактным носителем, имеющая четыре непрерывные ограниченные производные, $\varphi \in \mathcal{F}$, $\varphi'' \in \mathcal{F}$ и $\varphi_h(x) = (1/h)\,\varphi(x/h)$. В лемме 5 будет доказано, что величина $B^*(f)$ не зависит от выбора φ и что для любых f из класса $\mathcal{F}$ рассмотренные два определения совпадают, т. е. $B^*(f) = B(f)$.

Наконец, для описания точного асимптотического поведения величины $E(J_n)$ требуется ввести функцию

$$\psi(u) = \sqrt{2/\pi}\left(u\int_0^u e^{-x^2/2}\,dx + e^{-u^2/2}\right), \quad u \geqslant 0.$$

По-видимому, полезно получить представление о том, как ведет себя функция ψ. В силу формулы Миллза

$$\int_u^\infty e^{-x^2/2}dx \leqslant (1/u)\,e^{-u^2/2}$$

имеем $\psi(u) \geqslant u$. Заметим также, что $\psi(u) \geqslant \sqrt{2/\pi}$. Таким образом,

$$\max\left(u, \sqrt{2/\pi}\right) \leqslant \psi(u) \leqslant u + \sqrt{2/\pi},$$

$$\psi'(u) = \sqrt{2/\pi}\int_0^u e^{-x^2/2}\,dx \geqslant 0 \quad \text{(т. е. } \psi \text{ монотонно не убывает)},$$

$$\psi''(u) \geqslant 0 \quad \text{(т. е. } \psi \text{ выпукла)},$$

$$\lim_{u \downarrow 0} \psi(u) = \sqrt{2/\pi}.$$

Теорема 1. *Для любых плотностей f из класса $\mathcal{F}$, имеющих компактный носитель, ошибка ядерной оценки, определенной в* (1)—(3), *удовлетворяет соотношению*

$$E(J_n) = J(n, h) + o(h^2 + (nh)^{-1/2}),$$

где

$$J(n, h) = \int \frac{\alpha \sqrt{f}}{\sqrt{nh}} \psi\left(\sqrt{nh^5}\, \frac{\beta |f''|}{2\alpha \sqrt{f}}\right).$$

Кроме того,

$$J(n, h) \leqslant \sqrt{\frac{2}{\pi}}\, \frac{\alpha \int \sqrt{f}}{\sqrt{nh}} + \frac{\beta}{2} h^2 \int |f''|.$$

Если f имеет компактный носитель, то

$$E(J_n) \leqslant \sqrt{\frac{2}{\pi}}\, \frac{\alpha \int \sqrt{f}}{\sqrt{nh}} + \frac{\beta}{2} h^2 \sup_{a>0} \int |(f * \varphi_a)''| + o((nh)^{-1/2}),$$

где функция φ — такая же, как в определении величины $B^*(f)$. В частности,

$$\limsup_{n\to\infty} \inf_{h>0} n^{2/5} E(J_n) \leqslant C^* A(K)\, B^*(f),$$

где

$$C^* = 5(8\pi)^{-2/5} = 1{,}3768102 \ldots .$$

В случае когда плотность f имеет компактный носитель и $B^*(f) < \infty$, данная верхняя граница не превышается при следующем выборе параметра h:

$$h = \left[\frac{\alpha}{2\beta} \frac{\int \sqrt{f}}{\sup\limits_{a>0} \int |(f * \varphi_a)''|} \sqrt{\frac{2}{\pi}}\right]^{2/5} n^{-1/5}.$$

В верхней границе для $J(n, h)$ мы обнаруживаем составляющую, которая соответствует смещению (второе слагаемое), и дисперсионную составляющую (первое слагаемое). Утверждение теоремы 1 состоит в том, что для плотностей с конечным значением $B^*(f)$ величина $E(J_n)$ убывает со скоростью $n^{-2/5}$, если h выбрано пропорциональным $n^{-1/5}$. Отметим, что при слабо отличающихся условиях регулярности (задающих класс, не вложенный в $\mathcal{F}$) Розенблат (Rosenblatt (1979)) получил сходную верхнюю границу с несколько большей постоянной $5/2^{9/5} = 1{,}435872 \ldots$. В его доказательстве используется (хотя и без явной формулировки) неоптимальное неравенство $\psi(u) \leqslant 1 + u$.

Доказательство теоремы 1 и различных ее следствий будет отложено и проведено в других параграфах, с тем чтобы здесь можно было продолжить изложение основных результатов. В до-

полнение к верхней границе теоремы 1 предлагается следующая нижняя граница.

Теорема 2. *Для любых плотностей f ядерная оценка, определенная в* (1) *и* (2), *удовлетворяет соотношению*

$$\lim_{n\to\infty} \inf_h n^{2/5} E(J_n) \geqslant C A(K) B^*(f) \geqslant C C_1 A(K) \geqslant C C_1 C_2 = C_3,$$

где

$$C = \inf_u \frac{\psi(u)}{u^{1/5}} = 1{,}028493\ldots$$

— *универсальная постоянная,*

$$C_1 = \inf_f B^*(f) = \left(\frac{2^9}{3^4}\right)^{1/5} = 1{,}4459624\ldots$$

и

$$C_2 = \inf_{\substack{\text{четные} \\ \text{плотности} \\ K}} A(K) = \left(\frac{9}{125}\right)^{1/5} = 0{,}59083538\ldots.$$

В теореме 2 дается, таким образом, универсальная нижняя граница, которую ввиду теоремы 1 нельзя очень сильно улучшить: действительно, постоянная C^* в верхней границе примерно лишь на 35 % превосходит постоянную C в нижней границе. Нижняя граница теоремы 2 в определенном смысле более информативна, чем границы из гл. 4, — она относится ко всем плотностям, а не только к «худшим» плотностям из некоторых классов. Еще более важно, что эта нижняя граница помогает принять решение о том, достаточен ли имеющийся объем выборки n для достижения требуемой точности в L_1.

Данная нижняя граница остается в силе, если h — случайная величина, не зависящая от $X_1, \ldots, X_n$. Это может быть полезно в случае, когда h оценивается адаптивно по независимой выборке, объем которой обычно много меньше n.

Существование универсальной нижней границы $C_3/n^{2/5}$ является следствием замечания об универсальной ограниченности снизу величины $B^*(f)$ постоянной C_1. Доказательство этого утверждения содержится в теореме 3. Следует упомянуть, что точная нижняя грань всех величин $B^*(f)$ достигается на равнобедренной треугольной плотности. Следовательно, с помощью ядерной оценки легче всего оценивать равнобедренную треугольную плотность. На этом замечании основано исследование преобразованной ядерной оценки в гл. 9. Заинтересовавшийся читатель мог бы сейчас непосредственно перейти к гл. 9, не потеряв нити изложения.

Минимальное значение постоянной $A(K)$ равно C_2, и оно достигается на ядре Епанечникова

$$K(x) = 3(1 - x^2)/4, \quad |x| \leqslant 1$$

(см. Bartlett (1963), Епанечников (1969), а также Tapia, Thompson (1978)). Краткое доказательство этого факта приведено в лемме 18. Строго говоря, эту функцию следовало бы называть не ядром Епанечникова, как это делается обычно в литературе, а ядром Бартлетта. Известно, что значения $A(K)$ для наиболее естественных четных плотностей K очень близки к минимальному значению C_2 (см. Rosenblatt (1971) или Deheuvels (1977)).

В связи с важностью коэффициента $B^*(f)$ его изучению будет посвящен целый параграф. Еще в одном параграфе будут приведены основные леммы и доказательства. Затем аналогичным образом будет исследована гистограммная оценка, скорость сходимости которой, как мы увидим, равна $n^{-1/3}$, а не $n^{-2/5}$.

В § 9 рассмотрены равномерные верхние границы для средней ошибки в L_1, т. е. границы, которые можно применять при любом значении n. Верхние границы для $E(J_n)$ такого же типа, как в теореме 1, даны в § 10 применительно к плотностям f с некомпактным носителем. Наконец, в § 11 показано, что для некоторых гладких, но, как правило, имеющих длинные хвосты плотностей и специальным образом построенной ядерной оценки достигается неравенство $E(J_n) \leqslant c/\sqrt{n}$. За это заплачено дорогой ценой: оценка несостоятельна для подавляющего большинства плотностей, в том числе для плотностей с компактным носителем.

2. Коэффициент $B^*(f)$

Из теорем 1 и 2 следует вывод, что коэффициент $B^*(f)$ является мерой сложности оценивания f с помощью обычной ядерной оценки плотности. В этом параграфе будут установлены различные свойства величины $B^*(f)$ и ее составляющих. Грубо говоря, имеется составляющая $\int \sqrt{f}$ величины $B^*(f)$, которая служит мерой того, насколько тяжелы хвосты плотности f, и составляющая $\sup_{h>0} \int |(f * \varphi_h)''|$ — мера того, насколько сильно плотность f осциллирует. Рассмотрим эти составляющие раздельно, начав с $\int \sqrt{f}$.

Требуется пояснить утверждение о том, что из малости интеграла $\int \sqrt{f}$ следует малость хвостов плотности f, и обратно. Для этого приведем сначала одно обобщение неравенства Карлсона

(Carlson (1934)) (см. Beckenbach, Bellman (1965, с. 242 по русскому изданию)).

Лемма 1. *Для любой случайной величины X с плотностью f на R имеем*

(i) $\int \sqrt{f} \leqslant \sqrt{2\pi}\,(\mathrm{Var}\,(X))^{1/4}$,

(ii) $\int \sqrt{f} \leqslant \inf\limits_{a \in R} C_\varepsilon (E(|X-a|^{1+\varepsilon}))^{1/2\,(1+\varepsilon)}$,

где $\varepsilon > 0$ — произвольная постоянная и $C_\varepsilon = (8\pi \sin^{-1}(\pi/(1+\varepsilon))\,\varepsilon^{-\varepsilon/(1+\varepsilon)})^{1/2}$ — величина, зависящая только от ε.

Доказательство. Неравенство Карлсона для неотрицательных функций g на $[0, \infty)$ имеет вид

$$\int_0^\infty g \leqslant \sqrt{\pi}\left(\int_0^\infty g^2\right)^{1/4}\left(\int_0^\infty x^2 g^2\right)^{1/4}.$$

Положим теперь $g = \sqrt{f}$, и пусть $p = \int\limits_0^\infty f$, $a = \int\limits_0^\infty x^2 f \Big/ \int\limits_{-\infty}^{\infty} x^2 f$. Ясно, что тогда

$$\int \sqrt{f} \leqslant \sqrt{\pi}\left(\int x^2 f\right)^{1/4}(p^{1/4}a^{1/4} + (1-p)^{1/4}(1-a)^{1/4}).$$

При фиксированном $a \in (0, 1)$ последний множитель достигает максимума, если $p = b/(1+b)$, где $b = (a/(1-a))^{1/3}$, причем его максимальное значение $(a^{1/3} + (1-a)^{1/3})^{3/4}$, соответствующее этому p, в свою очередь никогда не превосходит $\sqrt{2}$. Таким образом, имеем

$$\int \sqrt{f} \leqslant \sqrt{2\pi}\,(E(X^2))^{1/4}.$$

Но так как всегда можно сдвинуть X на некоторую постоянную, скажем $E(X)$, то отсюда непосредственно следует (i).

Второе неравенство будет получено независимым образом. При $\varepsilon = 1$ его правая часть лишь вдвое больше, чем в неравенстве (i). Если $b, c > 0$ — постоянные, то всегда

$$\int \sqrt{f} \leqslant c\int (1 + b|x|^{1+\varepsilon}) f + \int\limits_{c(1+b|x|^{1+\varepsilon}) f \leqslant \sqrt{f}} \sqrt{f} \leqslant$$

$$\leqslant c + bcE(|X|^{1+\varepsilon}) + 2\int_0^\infty \frac{1}{c(1+b|x|^{1+\varepsilon})}\,dx =$$

$$= c + bcE(|X|^{1+\varepsilon}) + \frac{1}{c}\cdot\frac{1}{b^{1/(1+\varepsilon)}}\cdot\frac{2\pi/(1+\varepsilon)}{\sin(\pi/(1+\varepsilon))}.$$

Если A, B — неотрицательные постоянные, то при $c = \sqrt{B/A}$ функция $cA + B/c$ достигает минимального значения, равного $2\sqrt{AB}$. Отсюда следует, что интеграл $\int \sqrt{f}$ ограничен величиной

$$2\left((1 + bE(|X|^{1+\varepsilon}))\, b^{-1/(1+\varepsilon)}\, \frac{2\pi}{1+\varepsilon}\, \frac{1}{\sin(\pi/(1+\varepsilon))}\right)^{1/2}.$$

Нетрудно проверить, что при $b = (\varepsilon E(|X|^{1+\varepsilon}))^{-1}$ эта величина достигает как функция от b своего минимального значения, равного верхней границе в (ii), где $a = 0$. Нижняя грань по всем $a \in R$ добавлена в силу того, что всегда можно сдвинуть X на произвольную величину. Лемма 1 доказана.

Заметим, что доказанное очень близко к утверждению о том, что из неравенства $E(|X|) < \infty$ вытекает конечность интеграла $\int \sqrt{f}$. В силу следующего примера легко видеть, что это утверждение верно не всегда. Пусть f монотонно не возрастает на $[0, \infty)$, так что $f(x) \sim (x \log x)^{-2}$ при $x \to \infty$. Ясно, что $\int \sqrt{f} = \infty$, но $E(|X|) < \infty$.

С другой стороны, возможно, что $E(|X|) = \infty$ и $\int \sqrt{f} < \infty$. Пусть, например, f — индикаторная функция множества A. Ясно, что $\int \sqrt{f} = 1$. Но если выбрать $A = \bigcup_i [x_i,\ x_i + 2^{-i}]$, где $\sum x_i 2^{-i} = \infty$, то $E(|X|) = \infty$. Таким образом, маленькие хвосты обеспечивают малость величины $\int \sqrt{f}$, тогда как обратное верно разве что при особых условиях. Например, имеются обратные неравенства, такие, как

$$E(|X|) \leqslant \sup(|x|\sqrt{f}) \int \sqrt{f}.$$

Для плотностей с правильно меняющимися хвостами порядка r, т. е. таких, что $\lim_{x\to\infty} f(tx)/f(x) = t^r$ при любых $t > 0$ и аналогичное соотношение имеет место в пределе при $x \to -\infty$, значения $\int \sqrt{f}$ конечны при $r < -2$. В частности, $\int \sqrt{f} < \infty$ для всех плотностей с экспоненциально убывающим хвостом или хвостами, но $\int \sqrt{f} = \infty$ для плотности распределения Коши. Если $r > -2$, то $\int \sqrt{f} = \infty$.

Рассмотрим теперь осцилляционный коэффициент [1]). Дока-

---

[1]) Этими словами здесь и далее называется величина $\sup_{h>0} \int |(f * \varphi_h)'|$. — *Прим. перев.*

жем сначала, что при $f \in \mathcal{F}$ он представляет собой не что иное, как $\int |f''|$.

Лемма 2. *$B^*(f) = B(f)$ при любых $f \in \mathcal{F}$.*

Доказательство. Вспоминая теорему 2.1, получим

$$\lim_{h \downarrow 0} \int |(f * \varphi_h)'' - f''| = \lim_{h \downarrow 0} \int |f'' * \varphi_h - f''| = 0$$

в случае, когда $\int |f''| < \infty$. Для таких f имеем $\lim\limits_{h \downarrow 0} \int |(f * \varphi_h)''| = \int |f''|$. Если $\int |f''| = \infty$, то, прибегая к помощи леммы Фату, получим

$$\liminf_{h \downarrow 0} \int |(f * \varphi_h)''| \geqslant \int \liminf_{h \downarrow 0} |f'' * \varphi_h| = \int |f''|$$

в силу непрерывности f''. Таким образом, $B^*(f) \geqslant B(f)$ для любых f из класса $\mathcal{F}$.

Более того,

$$\int |(f * \varphi_h)''| = \int |f'' * \varphi_h| \leqslant \int |f''|,$$

так что $B^*(f) \leqslant B(f)$. Лемма доказана.

Проведем теперь подготовку к доказательству того, что величина $B^*(f)$ не зависит от выбора φ. Наши усилия не пропадут даром, поскольку некоторые из лемм, доказываемых в ходе этой подготовки, окажутся весьма полезными в дальнейшем. Леммы 3 и 4 частично пересекаются с леммой 22, но пока они достаточны для наших целей.

Лемма 3. *Для любой плотности K, удовлетворяющей условию* (2), *и любой плотности $f \in \mathcal{F}$ имеем*

$$f * K_h - f = h^2 \tilde{K}_h * f'',$$

где $\tilde{K}$ — некоторая неотрицательная симметричная функция с компактным носителем, интеграл от которой равен $\beta/2$, т. е.

$$\int \tilde{K} = \frac{1}{2} x^2 K = \frac{\beta}{2}.$$

Доказательство. Рассмотрим следующее разложение в ряд Тейлора в окрестности фиксированной точки x:

$$f(y) = f(x) + (y - x) f'(x) + \int_x^y (y - z) f''(z)\, dz.$$

В силу симметричности K имеем

$$f * K_h - f = \int \frac{1}{h} K\left(\frac{x-y}{h}\right) \int_x^y (y-z) f''(z)\, dz\, dy =$$

$$= \int \frac{1}{h} K\left(\frac{x-y}{h}\right) \int [I_{y>x} I_{x\leqslant z<y} (y-z) f''(z)\, dz +$$

$$+ I_{y\leqslant x} I_{x>z>y} (-(y-z)) f''(z)\, dz]\, dy =$$

$$= \int \frac{1}{h} K\left(\frac{x-y}{h}\right) \int [I_{y>x} I_{x\leqslant z\leqslant y} + I_{y\leqslant x} I_{x>z>y}] \times$$

$$\times |y-z| f''(z)\, dz\, dy =$$

$$= h^2 \int \frac{1}{h} \tilde{K}\left(\frac{x-z}{h}\right) f''(z)\, dz,$$

где

$$h^2 \cdot \frac{1}{h} \tilde{K}\left(\frac{x-z}{h}\right) =$$

$$= \int \frac{1}{h} K\left(\frac{x-y}{h}\right) [I_{y>x} I_{x\leqslant z\leqslant y} + I_{y\leqslant x} I_{x>z>y}]\, |y-z|\, dy =$$

$$= \int K(u) [I_{x-hu>x} I_{x\leqslant z\leqslant x-hu} + I_{x-hu\leqslant x} I_{x>z>x-hu}]\, |x-hu-z|\, du =$$

$$= h^2 \left(\frac{1}{h} \int K(u) [I_{u\leqslant 0} I_{0\leqslant (z-x)/h\leqslant -u} + I_{u>0} I_{0>(z-x)/h>-u}] \left|u + \frac{z-x}{h}\right| du\right).$$

Легко проверить, что если функция K ограниченна, симметрична и имеет компактный носитель, то такими же свойствами обладает и $\tilde{K}$. Кроме того,

$$\int \tilde{K} = \int K(u) [I_{u\leqslant 0} I_{0\leqslant -x\leqslant -u} + I_{u>0} I_{0>-x>-u}]\, |u-x|\, du\, dx =$$

$$= \int K(u) \left(I_{u\leqslant 0} \int_u^0 |u-x|\, dx + I_{u>0} \int_0^u |u-x|\, dx \right) du =$$

$$= \frac{1}{2} \int u^2 K(u)\, du = \beta/2,$$

что и требовалось доказать.

Лемма 4. *Пусть плотность K удовлетворяет условию* (2), *и пусть $\varphi \in \mathcal{F}$ — четырежды непрерывно дифференцируемая плотность с компактным носителем, такая, что $\varphi'' \in \mathcal{F}$. Тогда справедливы следующие неравенства*:

(i) $\int |f * K_h - f| \leqslant h^2 (\beta/2) \int |f''|$ *при любых $f \in \mathcal{F}$ и любых $h > 0$;*

(ii) $\int |f*K_h - f| \leqslant h^2(\beta/2) \liminf_{a \downarrow 0} \int |(f*\varphi_a)''|$ *при любых* f *и любых* $h \geqslant 0$;

(iii) $\int |f*K_h - f| \geqslant h^2\left[(\beta/2)\int |(f*\varphi_a)''| - h^2\beta_1 \int |(f*\varphi_a)''''|\right]$ *при любых* f *и любых* h, $a > 0$. (*Здесь* β_1 — *неотрицательная постоянная, зависящая только от* K.)

Доказательство. В силу леммы 3 интеграл в (i) равен

$$h^2 \int |\tilde{K}_h * f''| \leqslant h^2 \int \tilde{K} \int |f''| = h^2 \frac{\beta}{2} \int |f''|,$$

где мы воспользовались теоремой 2.1. Это показывает, что выполнено (i).

Неравенство (ii) следует непосредственно из (i) и теоремы 2.1. В самом деле, заметим, что

$$\int |f*K_h - f| \leqslant \int |f*K_h - f*\varphi_a*K| + \int |f - f*\varphi_a| +$$

$$+ \int |f*\varphi_a*K_h - f*\varphi_a| \leqslant 2\int |f - f*\varphi_a| + h^2 \frac{\beta}{2} \int |(f*\varphi_a)''|,$$

и устремим a к 0.

Для доказательства неравенства (iii) рассмотрим сначала четырежды непрерывно дифференцируемые плотности f из класса $\mathscr{F}$, такие, что $f'' \in \mathscr{F}$ и $\int |f''| < \infty$. В силу леммы 3

$$|f*K_h - f| = |h^2\left(\int \tilde{K}_h(x-z)f''(x)\,dz - \right.$$

$$\left. - \int \tilde{K}_h(x-z)(f''(x) - f''(z))\,dz\right)| \geqslant$$

$$\geqslant h^2\left(\frac{\beta}{2}|f''(x)| - \left|\int \tilde{K}_h(x-z)(f''(x) - f''(z))\,dz\right|\right).$$

К последнему слагаемому снова применима лемма 3 при замене f и K на f'' и $\tilde{K}$. Таким образом, существует неотрицательная постоянная β_1, зависящая только от K, такая, что

$$\int |f*K_h - f| \geqslant h^2\left(\frac{\beta}{2}\int |f''| - h^2\beta_1 \int |f''''|\right).$$

В случае произвольной плотности f применим только что полученное неравенство к функции $f * \varphi_a$ при любом $a > 0$ и, объединяя его с соотношением

$$\int |f*K_h - f| \geqslant \int |(f*\varphi_a)*K_h - f*\varphi_a|,$$

получим (iii).

Основной результат, относящийся к осцилляционному коэффициенту, заключен в следующей лемме.

Лемма 5. *Пусть f — произвольная плотность, K — плотность, удовлетворяющая условию* (2), *и φ — четырежды непрерывно дифференцируемая плотность с компактным носителем, такая, что $\varphi'' \in \mathcal{F}$. Тогда следующие величины равны друг другу и не зависят от K и φ:*

$$\lim_{h \downarrow 0} \frac{\int |f * K_h - f|}{h^2 \beta/2} = \lim_{a \downarrow 0} \inf \int |(f * \varphi_a)''| = \sup_{a>0} \int |(f * \varphi_a)''|.$$

Доказательство. В силу леммы 4 (ii) первая величина не превосходит

$$\lim_{a \downarrow 0} \inf \int |(f * \varphi_a)''|.$$

В силу леммы 4 (iii) она больше или равна, чем $\sup_{a>0} \int |(f * \varphi_a)''|$. Следовательно, нижний предел совпадает с верхней гранью, что влечет за собой равенство трех величин из леммы 5.

Таким образом, установлена прочная связь между смещением ядерной оценки и осцилляционным коэффициентом. Покажем теперь, что величина $B^*(f)$ ограничена снизу универсальной постоянной. Мы будем продвигаться вперед очень медленно, чтобы читатель смог понять существо рассуждений, используемых при выводе таких нижних границ. Сначала покажем, что $B^*(f) \geqslant 1$ при любых f из класса $\mathcal{F}$. Затем покажем, что эта нижняя граница верна при любых f и, наконец, что ее можно улучшить до $(2^9/81)^{1/5}$, причем последнее значение достигается на равнобедренной треугольной плотности.

Лемма 6. $B^*(f) \geqslant 1$ *для любых f из класса $\mathcal{F}$.*

Доказательство. Можно считать, что $\int |f''| < \infty$. Так как функция f' абсолютно непрерывна, то

$$f'(y) - f'(x) = \int_x^y f''(z)\, dz$$

при любых $x < y$. Так как функция f'' ограниченна, то f' удовлетворяет условию Липшица. Кроме того, $f'(x)$ стремится к 0 при $|x| \to \infty$ в силу того, что функция f'' абсолютно интегрируема.

Таким образом, обозначая символами $(\ )_+$ и $(\ )_-$ положительную и отрицательную части функции, получим

$$\int_{-\infty}^{\infty} (f''(y))_- dy \leqslant f'(x) \leqslant \int_{-\infty}^{\infty} (f''(y))_+ dy$$

при любых x и

$$\int_{-\infty}^{\infty} (f''(y))_+ dy + \int_{-\infty}^{\infty} (f''(y))_- dy = 0,$$

так что

$$\sup |f'(x)| \leqslant \frac{1}{2} \int_{-\infty}^{\infty} |f''(y)|\, dy.$$

Но, кроме того, имеем $1 = \int f \leqslant \sqrt{\sup f} \int \sqrt{f}$. Соединяя эти неравенства, получим $B(f)^5 \geqslant \sup |f'(x)|/\sup f^2(x)$. В силу геометрических соображений ясно, что

$$1 = \int f(x)\, dx \geqslant \int (\sup f(x) - |y| \sup |f'(x)|)_+ dy = \frac{\sup f^2(x)}{\sup |f'(x)|}.$$

Лемма 6 доказана.

Лемма 7. *$B^*(f) \geqslant 1$ при любых f.*

Доказательство. В этом доказательстве используются следующие утверждения.

A. $\int \sqrt{f} = \int \sqrt{f} * \varphi_a$ (где φ и a такие, как в определении величины $B^*(f)$).

B. $\int \sqrt{f} * \varphi_a \cdot \sup(\sqrt{f} * \varphi_a) \geqslant \int (\sqrt{f} * \varphi_a)^2$.

C. $\liminf\limits_{a \downarrow 0} \int (\sqrt{f} * \varphi_a)^2 \geqslant \int \liminf\limits_{a \downarrow 0} (\sqrt{f} * \varphi_a)^2$ (по лемме Фату)

$= \int (\sqrt{f})^2 =$ (так как $\sqrt{f} * \varphi_a \to \sqrt{f}$ при почти всех x; см. теорему 2.3)

$= 1$.

D. $f * \varphi_a \in \mathscr{F}$.

E. $\sup |(f * \varphi_a)'| \leqslant \int |(f * \varphi_a)''|/2$ (см. доказательство леммы 6).

F. $1 \geqslant \sup (f * \varphi_a)^2 / \sup |(f * \varphi_a)'|$ (см. доказательство леммы 6).

G. $(\sqrt{f} * \varphi_a)^2 \leqslant f * \varphi_a$ (по неравенству Йенсена).

Таким образом, при фиксированном a имеем

$$B^*(f)^5 \geqslant \left(\int (\sqrt{f} * \varphi_a)^2\right)^4 \frac{\sup (f * \varphi_a)^2}{\sup (\sqrt{f} * \varphi_a)^4} \geqslant \quad \text{(в силу A, B, E, F)}$$

$$\geqslant \left(\int (\sqrt{f} * \varphi_a)^2\right)^4 \quad \text{(в силу G)}.$$

Перейдем к нижнему пределу при $a \downarrow 0$ и применим утверждение C.

Теорема 3. *$B^*(f) \geqslant (2^9/81)^{1/5}$ при любых f. Эта нижняя граница достигается на равнобедренной треугольной плотности.*

Доказательство. Предположим сначала, что $f \in \mathscr{F}$ и $\int |f''| < \infty$. Пусть функции K и φ те же, что в (2) и в определении $B^*(f)$. Имеем $f'(x) \to 0$ при $|x| \to \infty$. Таким образом, используя представление $f'' = f''_+ + f''_-$ функции f'' в виде суммы ее положительной и отрицательной частей, получим

$$\int |f''| = 2\int f''_+ = -2\int f''_- \geqslant 2(\sup f' - \inf f').$$

Следовательно, $(B^*(f))^5 \geqslant \left(\int \sqrt{f}\right)^4 (\sup f' - \inf f')$. Введем класс $\mathscr{G}$, состоящий из всех плотностей, таких, что

$$-D(y-x) \leqslant f(y) - f(x) \leqslant C(y-x),$$

при любых $x < y$; класс $\mathscr{G}$ включает $\mathscr{F}$. Здесь C и D — положительные постоянные. Среди таких плотностей минимум величины $\int \sqrt{f}$ достигается на треугольной плотности с высотой b и основанием, которое разбивается модой на отрезки длины c и d, где $b/c = C$, $b/d = D$ и $bc + bd = 2$. Таким образом, $c^2 = 2/(C + C^2/D)$, $d^2 = 2/(D + D^2/C)$. Для плотностей f из класса $\mathscr{G}$ имеем $\sup f' \leqslant C$ и $\inf f' \geqslant -D$. Следовательно, $(B^*(f))^5 \geqslant \inf\limits_{C,\,D} \inf\limits_{g \in \mathscr{G}} (C+D)\left(\int \sqrt{g}\right)^4$. Так как величина $(C + D)\left(\int \sqrt{g}\right)^4$ инвариантна относительно масштаба, то можно положить $C = 1$. Для упомянутой выше треугольной плотности имеем

$$\left(\int \sqrt{g}\right)^4 (1+D) \geqslant (2/3)^4 \left(c^{3/2} + \sqrt{D}\, d^{3/2}\right)^4 (1+D) =$$

$$= \left(\frac{2}{3}\right)^4 (1+D)\left(\left(\frac{8D^3}{(1+D)^3}\right)^{1/4} + \left(\frac{8D^2}{(D+D^2)^3}\right)^{1/4}\right)^4 =$$

$$= (2/3)^4\, 8\,(1+D)^2/D \geqslant 2\,(4/3)^4,$$

замечая, что величина $(1 + D)^2/D$ минимальна при $D = 1$. Таким образом, $\inf\limits_{f \in \mathscr{F}} (B^*(f))^5 \geqslant 2\,(4/3)^4$.

Пусть f — плотность с компактным носителем и K — плотность с компактным носителем, принадлежащая классу $\mathscr{F}$. Обо-

значим функцию $f * K_{1/n}$ через f_n. Покажем, что $B^*(f_n) \leqslant$ $\leqslant B^*(f)(1 + o(1))$. Так как каждая из функций f_n имеет компактный носитель и принадлежит классу $\mathscr{F}$, то тем самым требуемый результат будет получен для всех плотностей с компактным носителем. Прежде всего в силу леммы Фату и теоремы 2.3

$$\liminf_{n\to\infty} \int \sqrt{f * K_{1/n}} \geqslant \int \sqrt{\liminf_{n\to\infty} f * K_{1/n}} = \int \sqrt{f}.$$

Кроме того, для достаточно большого компактного множества T, содержащего носитель f, и достаточно больших n имеем

$$\int \sqrt{f_n} \leqslant \int \sqrt{|f_n - f|} + \int \sqrt{f} \leqslant$$

$$\leqslant \sqrt{\lambda(T)} \sqrt{\int |f_n - f|} + \int \sqrt{f} = o(1) + \int \sqrt{f},$$

где мы воспользовались теоремой 2.1. Далее, в силу того что для любых плотностей φ из класса $\mathscr{F}$, имеющих компактный носитель, и для любых $h > 0$ выполнены соотношения

$$\int |(f * K_{1/n} * \varphi_h)''| = \int |(f * \varphi_h)' * K_{1/n}| \leqslant \int |(f * \varphi_h)''|,$$

получим, что $B^*(f_n) \leqslant (1 + o(1))\, B^*(f)$.

Чтобы обобщить этот результат, полученный для всех плотностей с компактным носителем, на всевозможные плотности, нужны лишь обычные аналитические построения. Например, можно аппроксимировать f последовательностью функций g_t, таких, что значение g_t на $[-t, t]$ равно 1, вне $[-t-1, t+1]$ равно 0, а на промежуточных интервалах функция g_t является непрерывной и гладкой. (Заметим, что метод свертки применять нельзя, поскольку существуют плотности, для которых $\int \sqrt{f} < \infty$, но $\int \sqrt{f * K_h} = \infty$ при любых плотностях K и любых $h > 0$.)

3. Доказательства теорем 1 и 2

В начале этого параграфа поясним, каким образом функция ψ проникла в выражение для $J(n, h)$ из теоремы 1. С этой целью приведем отдельно две важные леммы. В оставшейся части доказательства теоремы 1 остановимся на некоторых результатах о смещении и дисперсии ядерной оценки.

Лемма 8. *Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины. Пусть $E(X_1) = 0$, $E(X_1^2) = \sigma^2 > 0$ и $\rho = E(|X_1|^3) < \infty$. Тогда*

$$\sup_{a \in R} \left| E\left(\left| (\sigma\sqrt{n})^{-1} \sum_{i=1}^{n} X_i - a \right| \right) - E(|N - a|) \right| \leqslant \frac{c\rho\sigma^{-3}}{\sqrt{n}},$$

где c — универсальная положительная постоянная и N — нормальная с параметрами (0, 1) *случайная величина. Заметим, что*

$$E(|N-a|) = |a|P(|N| \leqslant |a|) + \sqrt{2/\pi}e^{-a^2/2} = \psi(|a|).$$

Доказательство. Пусть F_n — функция распределения случайной величины $X = (\sigma\sqrt{n})^{-1}\sum_{i=1}^{n} X_i$, и пусть Φ — функция распределения N. Ясно, что

$$E(|X-a|) = \int_0^\infty P(|X-a|>t)\,dt = \int_0^\infty (1 - F_n(a+t) + F_n(a-t))\,dt$$

и аналогичное равенство имеет место для N и Φ. Абсолютная величина разности между правыми частями этих двух равенств не превосходит

$$\int_0^\infty |\Phi(a+t) - F_n(a+t)|\,dt + \int_0^\infty |\Phi(a-t) - F_n(a-t)|\,dt =$$

$$= \int_{-\infty}^{\infty} |\Phi(t) - F_n(t)|\,dt.$$

В силу хорошо известных неравномерных оценок в центральной предельной теореме типа Берри — Эссеена (Петров (1972, теорема 14, с. 115))

$$|\Phi(t) - F_n(t)| \leqslant \frac{c\rho\sigma^{-3}}{(1+|t|^3)\sqrt{n}},$$

где c — некоторая универсальная постоянная. Так как функция $(1+|t|^3)^{-1}$ интегрируема, то получим требуемый результат.

Чтобы получить выражение для $E(|N-a|)$, заметим, что при $a > 0$

$$E(|N-a|) = E(|N|) + E(|N-a| - |N|) =$$
$$= E(|N|) + aP(N<0) + E((a-2N)I_{[0<N<a]}) - aP(N>a) =$$
$$= E(|N|) + a - 2E(NI_{[0<N<a]}) - 2aP(N>a) =$$
$$= \sqrt{2/\pi} + a - aP(|N|>a) - 2\int_0^a \frac{te^{-t^2/2}}{\sqrt{2\pi}}\,dt =$$
$$= \sqrt{2/\pi} + aP(|N|<a) - 2(1-e^{-a^2/2})/\sqrt{2\pi},$$

что и требовалось доказать.

Далее в этом параграфе через T обозначается произвольный интервал, через $[-r, r]$ — носитель K, через K^* — верхняя граница для K, через T^* — множество $\{x: |x-y| \leqslant hr$ при

некотором $y \in T\}$. Таким образом, T^* зависит от h. Кроме того, пусть c — постоянная из леммы 8, $B_n(x) = E(f_n(x)) - f(x)$ — *смещение* в точке x, $V_n(x) = f_n(x) - E(f_n(x))$ — *вариация* в точке x, $\sigma_n^2(x) = E(V_n^2(x))$ — *дисперсия* в точке x.

Лемма 9.

$$\left| E(|f_n(x) - f(x)|) - \sigma_n(x)\psi\left(\frac{|B_n(x)|}{\sigma_n(x)}\right)\right| \leqslant \frac{cK^*}{nh}$$

для любых плотностей K, удовлетворяющих условию (2).

Доказательство. Применим лемму 8 к случайным величинам

$$Y_i = \frac{1}{h}K\left(\frac{X_i - x}{h}\right) - E\left(\frac{1}{h}K\left(\frac{X_i - x}{h}\right)\right),$$

полагая $a = B_n(x)/\sigma_n(x)$. Получим, что величина ошибки из леммы 8 равна

$$c\frac{E(|Y_1|^3)}{nE(Y_1^2)} \leqslant \frac{cK^*}{nh}.$$

Лемма 10. *Пусть T — ограниченный интервал. Тогда для любых $h > 0$ и любых плотностей K имеем*

$$-\sqrt{h}/\alpha - \sqrt{\int |f * K_h^+ - f|\lambda(T)} \leqslant (\sqrt{nh}/\alpha)\int_T \sigma_n - \int_T \sqrt{f} \leqslant$$

$$\leqslant \sqrt{\int |f * K_h^+ - f|\lambda(T)},$$

где $K^+ = K^2 \Big/ \int K^2$. Кроме того, $\int_T |\sigma_n\sqrt{nh}/\alpha - \sqrt{f}| = o(1)$ при $h \downarrow 0$.

Доказательство. Прежде всего заметим, что на ограниченных множествах всегда $\int_T \sqrt{f} < \infty$. Имеем $\sigma_n^2(x) = a(x) + b(x)$, где

$$a(x) = \frac{\alpha^2 f(x)}{nh}$$

и

$$b(x) = \frac{(f * K_h^+ - f)\alpha^2}{nh} - \frac{(f * K_h)^2}{n}.$$

Ясно, что $a(x) \geqslant 0$. Следовательно, $\sqrt{a(x) + b(x)} \leqslant \sqrt{a(x)} + \sqrt{b_+(x)}$ и $\sqrt{a(x) + b(x)} \geqslant \sqrt{a(x)} - \sqrt{|b(x)|}$. Интегрируя по T и применяя неравенство Коши—Шварца, получим

$$\int_T \sigma_n \leqslant (\alpha/\sqrt{nh})\left(\int_T \sqrt{f} + \int_T \sqrt{|f * K_h^+ - f|}\right) \leqslant$$

$$\leqslant (\alpha/\sqrt{nh})\left(\int_T \sqrt{f} + \sqrt{\int |f * K_h^+ - f|\lambda(T)}\right)$$

и

$$\int_T \sigma_n \geqslant (\alpha/\sqrt{nh})\left(\int_T \sqrt{f} - \sqrt{\int |f * K_h^+ - f|\,\lambda(T)} - \right.$$
$$\left. - (\sqrt{h}/\alpha)\int f * K_h\right).$$

Отсюда легко следует первое утверждение леммы 10. Второе утверждение вытекает из соотношения $\int_T \sqrt{|b(x)|} = o(1)$. Но это соотношение является следствием того, что $h = o(1)$, $\lambda(T) < \infty$ и $\int |f * K_h^+ - f| = o(1)$ (см. теорему 2.1).

Лемма 11. *Для любых $f \in \mathcal{F}$, любых K, удовлетворяющих условию* (2), *и любых ограниченных интервалов T имеем*

$$\int_T \left||B_n| - \beta h^2 |f''|/2\right| = o(h^2)$$

при $h \downarrow 0$. Это отношение остается в силе при замене T на R, если плотность f имеет компактный носитель, $f \in \mathcal{F}$ и ядро K удовлетворяет условию (2).

Доказательство. Пусть $\tilde{K}$ — функция из леммы 3. Тогда

$$\int_T \left|\,|B_n| - \frac{\beta}{2}h^2|f''|\,\right| = \frac{\beta}{2}h^2 \int_T \left|\,\left|\frac{2}{\beta}\tilde{K}_h * f''\right| - |f''|\,\right| \leqslant$$
$$\leqslant \frac{\beta}{2}h^2 \int_T \left|\frac{2}{\beta}\tilde{K}_h * f'' - f''\right| = o(h^2)$$

по теореме Лебега о мажорируемой сходимости. Здесь использовано то, что $(2/\beta)\,\tilde{K}_h * f'' \to f''$ при любых x (теорема 2.3), функция $|f''|$ ограниченна и $|(2/\beta)\,\tilde{K}_h * f''| \leqslant \sup_x |f''(x)|$.

Чтобы приступить к доказательству теоремы 1, потребуется еще одна техническая лемма.

Лемма 12. *Если u, v, w, z — неотрицательные числа, то*

$$\left|u\psi\left(\frac{v}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| \leqslant |v - z| + \sqrt{\frac{2}{\pi}}\,|u - w|.$$

Доказательство. Сначала проверяем, что $0 \leqslant \psi'(u) \leqslant 1$ при любых $u \geqslant 0$ и $|(u\psi(v/u))'| \leqslant \sqrt{2/\pi}$ при любых $v \geqslant 0$. Таким образом,

$$\left|u\psi\left(\frac{v}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| \leqslant \left|u\psi\left(\frac{v}{u}\right) - u\psi\left(\frac{z}{u}\right)\right| +$$
$$+ \left|u\psi\left(\frac{z}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| \leqslant |v - z| + \sqrt{\frac{2}{\pi}}\,|u - w|.$$

Доказательство теоремы 1. Теорема 1 состоит из нескольких частей. Сначала предположим, что $f \in \mathcal{F}$ и что функция f имеет компактный носитель, заключенный в ограниченном интервале T. Пусть a — достаточно большое число, такое, что $K_h(u) = 0$ при любых $h \leqslant \max\limits_n h_n$ и любых $|u| > a$. Выберем интервал T достаточно большим для того, чтобы для любого x из носителя функции f отрезок $[x - a, x + a]$ содержался в T.

Воспользуемся сначала неравенством леммы 12 при

$$u = \sigma_n(x), \quad v = |B_n(x)|,$$

$$w = \alpha\sqrt{f(x)}/\sqrt{nh}, \quad z = \beta h^2 |f''(x)|/2.$$

По лемме 11

$$\int_T |v - z| = o(h^2),$$

и по лемме 10

$$\int_T |u - w| = o((nh)^{-1/2}).$$

Подстановка этих соотношений в неравенство леммы 9 дает

$$\int_T E(|f_n - f|) - J(n, h) \Big| \leqslant$$

$$\leqslant \left| \int_T E(|f_n - f|) - \int \sigma_n \psi\left(\frac{|B_n|}{\sigma_n}\right) \right| + \left| \int \sigma_n \psi\left(\frac{|B_n|}{\sigma_n}\right) - J(n, h) \right| \leqslant$$

$$\leqslant \frac{cK^*}{nh} \lambda(T) + o(h^2) + o((nh)^{-1/2}),$$

где мы воспользовались тем, что $J(n, h) = \int\limits_T w\psi(z/w)$.

Неравенство для $J(n, h)$ следует, из того, что $\psi(u) \leqslant u + \sqrt{2/\pi}$. Действительно,

$$J(n, h) = \int_T w\psi\left(\frac{z}{w}\right) \leqslant \int_T z + \sqrt{\frac{2}{\pi}} \int_T w.$$

Рассмотрим теперь всевозможные плотности f с компактным носителем и обозначим через L величину $\sup\limits_{a>0} \int |(f * \varphi_a)''|$, участвующую в определении $B^*(f)$. Снова в силу леммы 9 и неравенства $\psi(u) \leqslant u + \sqrt{2/\pi}$ получим

$$\int_T E(|f_n - f|) \leqslant \int_T \left(\sqrt{\frac{2}{\pi}}\sigma_n + |B_n|\right) + \frac{cK^*}{nh} \lambda(T),$$

и в силу лемм 4, 5 и 10 эта величина в свою очередь ограничена сверху выражением

$$\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{nh}}\int\sqrt{f}+\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{nh}}\sqrt{\int|f*K_h^+-f|\lambda(T)}+$$

$$+\frac{\beta}{2}h^2L+\frac{cK^*}{nh}\lambda(T),$$

где K^+ — плотность, определенная в лемме 10. Второе слагаемое здесь равно $o((nh)^{-1/2})$ при $h=o(1)$ (теорема 2.1). Последнее слагаемое равно $o((nh)^{-1/2})$ при $nh\to\infty$. Тем самым доказана первая верхняя граница для общих плотностей f. Если выбрать значение h, указанное в формулировке теоремы (т. е. значение, доставляющее минимум главному члену в верхней границе), то

$$\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{nh}}\int\sqrt{f}+\frac{\beta}{2}h^2L=C^*A(K)\frac{B^*(f)}{n^{2/5}},$$

и теорема 1 доказана.

Доказательство теоремы 2. Имеем

$$\inf_h E(J_n)\geqslant\min\left(\inf_{h\sqrt{n}\leqslant 1}E(J_n),\ \inf_{h\sqrt{n}>1}E(J_n)\right). \tag{4}$$

Рассмотрим последовательность h, такую, что $E(J_n)\sim\inf_h E(J_n)$. Так как условие (3) достаточно для сходимости $E(J_n)\to 0$ (теорема 3.1), то ясно, что $E(J_n)\to 0$ при любых f. Но в силу неравенства $E(J_n)\geqslant\int|f*K_h-f|$ необходимо, чтобы $h\to 0$ (теорема 2.4). Рассмотрим теперь по отдельности две нижние грани в (4).

Прежде всего, если последовательность h такова, что $h\geqslant 1/\sqrt{n}$ при любых n и $E(J_n)\sim\inf_{h\sqrt{n}\geqslant 1}E(J_n)$, то, как указано выше, $h\to 0$. Кроме того, $nh\to\infty$ и в действительности $nh/n^{2/5}\to\infty$. Пусть теперь T — ограниченный интервал и $a>0$ — произвольная постоянная. При таких h нижняя граница для $E(J_n)$ такова:

$$\int_T E(|f_n-f|)\geqslant\int_T\sigma_n\psi\left(\int_T|B_n|\Big/\int_T\sigma_n\right)-cK^*\lambda(T)/(nh)\geqslant$$

(в силу леммы 9, выпуклости ψ и неравенства Йенсена)

$$\geqslant C\left(\int_T\sigma_n\right)^{4/5}\left(\int_T|B_n|\right)^{1/5}-o(n^{-2/5})\geqslant\quad\text{(по определению } C)$$

$$\geqslant Cn^{-2/5}\left(\alpha\int_T\sqrt{f}\right)^{4/5}\left((\beta/2)\int_T|(f*\varphi_a)''|\right)^{1/5}(1+o(1))\sim$$

(в силу лемм 4 и 10)

$$\sim n^{-2/5} CA(K) \left[\left(\int_T \sqrt{f} \right)^4 \int_T |(f * \varphi_a)''|/2 \right]^{1/5}$$

(по определению $A(K)$). (5)

Далее, пусть последовательность h такова, что $h \leqslant 1/\sqrt{n}$ при любых n и $E(J_n) \sim \inf_{h\sqrt{n} \leqslant 1} E(J_n)$. В силу теоремы 3.1 известно, что $nh \to \infty$. Кроме того,

$$E(J_n) \geqslant E\left(\int |f_n - f * K_h| \right) / 2$$

в силу леммы 3.6. По лемме Фату

$$\liminf_{n\to\infty} n^{2/5} E(J_n) \geqslant \int \liminf_{n\to\infty} n^{2/5} E(|f_n - f * K_h|)/2,$$

и правая часть этого неравенства равна ∞, если $\liminf_{n\to\infty} n^{2/5} \times$ $\times E(|f_n - f * K_h|) = \infty$ при почти всех x, таких, что $f(x) > 0$. Чтобы показать это, применим центральную предельную теорему Берри—Эссеена, которая использовалась в лемме 8. Пусть $\sigma_n^2(x) = \operatorname{Var}(K_h(X_1 - x))$ и M — произвольно большое положительное число. Пусть Z — нормальная случайная величина с параметрами (0, 1). Тогда

$$\begin{aligned} n^{2/5} E(|f_n - f * K_h|) &\geqslant \\ &\geqslant MP(|f_n - f * K_h| \geqslant M/n^{2/5}) = \\ &= MP(|f_n - f * K_h| \sqrt{n}/\sigma_n \geqslant M\sqrt{n}/(\sigma_n n^{2/5})) \geqslant \\ &\geqslant M\left(P(|Z| \geqslant Mn^{1/10}/\sigma_n) - 2c\sigma_n^{-3} n^{-1/2} E(|K_h(X_1 - x) - \right. \\ &\qquad \left. - E(K_h(X_1 - x))|^3)\right). \end{aligned}$$

Из доказательств леммы 10 и теоремы 2.4 легко извлечь, что $\sigma_n^2(x) \sim \alpha^2 f(x)/h$ для почти всех x при $h \to 0$. Кроме того, в силу c_r-неравенства [1]) и теоремы 2.3

$$\begin{aligned} E(|K_h(X_1 - x) - E(K_h(X_1 - x))|^3) &\leqslant \\ &\leqslant 4E\left(h^{-3} K^3\left(\frac{X_1 - x}{h}\right)\right) + 4(E(K_h(X_1 - x)))^3 = \\ &= 4h^{-2} f * (K^3)_h + 4(f * K_h)^3 \sim \\ &\sim 4h^{-2} f(x) \int K^3 + 4f(x)^3 \sim 4h^{-2} f(x) \int K^3 \end{aligned}$$

---

[1]) Имеется в виду неравенство $(a+b)^r \leqslant 2^{r-1}(a^r + b^r)$, $a, b > 0$, $r \geqslant 1$. — *Прим. перев.*

для почти всех x. Так как $n^{1/10}/\sigma_n(x) \sim n^{1/10}\sqrt{h}/\alpha\sqrt{f(x)} \leqslant \leqslant 1/(n^{3/20}\alpha\sqrt{f(x)}) \to 0$ при почти всех x, таких, что $f(x) > 0$, то

$$n^{2/5}E(|f_n - f * K_h|) \geqslant$$

$$\geqslant M\left(1 - (2c + o(1))\frac{4\int K^3\sqrt{f(x)}}{\sqrt{nh}\alpha^3}\right) = M(1 - o(1))$$

при почти всех x, таких, что $f(x) > 0$. Так как M произвольно, то тем самым показано, что

$$\liminf_{n\to\infty} \inf_{h\sqrt{n}\leqslant 1} n^{2/5}E(J_n) = \infty.$$

Последнее вместе с (5), определением $B^*(f)$ и теоремой о монотонной сходимости дает

$$\liminf_{n\to\infty}\inf_h n^{2/5}E(J_n) \geqslant \sup_{\substack{a>0,\\ \text{ограниченные } T}} CA(K)\left(\int_T \sqrt{f}\right)^{4/5} \frac{\left(\int_T |(f * \varphi_a)''|\right)^{1/5}}{2^{1/5}} =$$

$$= CA(K)B^*(f).$$

Теорема 2 доказана.

4. Гистограммная оценка

Исследование гистограммной оценки на R можно провести так же, как исследование ядерной оценки. Рассмотрим только простейшую гистограммную оценку, которая определяется разбиениями вида

$$\mathscr{P}_n = \{[kh, (k+1)h), k - \text{целое}\}, \tag{6}$$

где $h = h_n$ — последовательность положительных чисел. Фридман и Диаконис (Freedman, Diaconis (1981)) полностью разработали L_2-теорию этой оценки для плотностей из класса $\mathscr{F}_2$, т. е. функций, удовлетворяющих условиям

(i) $f \in L_2$, f абсолютно непрерывна и имеет п. в. производную f',

(ii) $0 < \int (f')^2 < \infty$.

В частности, в работах Scott (1979) и Freedman, Diaconis (1981) доказана

Теорема 4. *При $f \in \mathscr{F}_2$ гистограммная оценка, определяемая разбиениями* (6), *удовлетворяет соотношению*

$$\inf_h E\left(\int (f_n - f)^2\right) \sim \frac{\left(36\int (f')^2\right)^{1/3}/4}{n^{2/3}},$$

и указанная в этом соотношении скорость сходимости достигается при $h \sim \left(6 \Big/ \left(n \int (f')^2\right)\right)^{1/3}$

Теорема 4 формулируется здесь без доказательства, только для последующих ссылок. Заметим, что при самых гладких плотностях f указанная оптимальная скорость сходимости в L_2 равна $n^{-2/3}$ в отличие от скорости $n^{-4/5}$ для ядерной оценки. Поэтому мы полагаем, что при гладких f ядерная оценка будет вести себя в L_2 лучше, чем гистограммная, по крайней мере при больших n. Покажем в этом параграфе, что то же самое остается в силе для $E(J_n)$. Конечно, такая иерархия оценок не распространяется на случай произвольных плотностей f. Например, если f — равномерная плотность на $[0, 1]$ и $h = 1$, то $E(J_n) = 0$ при любых n, тогда как для ядерной оценки

$$\liminf_{n\to\infty} n^{2/5} E(J_n) = \infty$$

независимо от выбора h (теорема 2), поскольку $\sup_{a>0} \int |(f * \varphi_a)''| = \infty$ и, следовательно, $B^*(f) = \infty$.

В этом параграфе через $\mathcal{F}$ обозначим класс функций, удовлетворяющих условиям

(i) f абсолютно непрерывна и имеет п. в. производную f',

(ii) f' ограниченна и непрерывна. (Заметим, что $\int |f'| < \infty$.)

Теорема 5. *При любых* $f \in \mathcal{F}$ *ошибка гистограммной оценки, определяемой разбиениями* (6), *удовлетворяет следующей нижней границе*

$$\liminf_{n\to\infty} \inf_h n^{1/3} E(J_n) \geqslant C_H B_H(f) \geqslant C_H,$$

где

$$C_H = \inf_u \psi(u)/(2u)^{1/3} = 0{,}880261\ldots$$

и

$$B_H(f) = \left(\left(\int \sqrt{f}\right)^2 \int |f'| \Big/ 2\right)^{1/3}$$

(*последняя величина* $\geqslant 1$ *при любых* f *из класса* $\mathcal{F}$).

В этом параграфе и в § 5 нам потребуется еще одна функция r_n, определяемая следующим образом:

$$r_n(x) = \frac{x - a}{b - a}$$

при любых $x \in A_{nj} = [a, b) = [jh, (j+1)h)$. Это — пилообразная функция на действительной оси со значениями между 0 и 1. Нам также понадобится функция

$$z_n(x) = (1 - 2r_n(x)) f'(x),$$

которая имеет смысл при любых f из класса $\mathscr{F}$. Функция z_n колеблется между $-f'$ и $+f'$.

Теорема 6 (точное асимптотическое поведение величины $E(J_n)$). *Пусть плотность $f \in \mathscr{F}$ имеет компактный носитель. Ошибка гистограммной оценки, определяемой разбиениями* (6) *и условием* (3), *равна*

$$E(J_n) = J(n, h) + o\left(h + \frac{1}{\sqrt{nh}}\right),$$

где

$$J(n, h) = \int \sqrt{\frac{f}{nh}}\,\psi\left(\frac{h}{2}\,|z_n|\,\sqrt{\frac{nh}{f}}\right).$$

Кроме того,

$$\limsup_{n\to\infty} \inf_h n^{1/3} E(J_n) \leqslant C_H^* B_H(f),$$

где $C_H^* = (27/4\pi)^{1/3} = 1{,}290381\ \ldots$.

Таким образом, на $\mathscr{F}$ величина $E(J_n)$ зажата между двумя близкими друг к другу границами $C_H B_H(f)/n^{1/3}$ и $C_H^* B_H(f)/n^{1/3}$, где $C_H^*/C_H = 1{,}46590\ \ldots$. Это показывает, что коэффициент $B_H(f)$ в определенной степени является мерой сложности оценивания f с помощью гистограммной оценки. Интересно, что $B_H(f) \not\equiv B^*(f)$. Иными словами, плотности, которые легко оценивать с помощью ядерной оценки, могут оказаться сложными для гистограммной оценки, и наоборот. Например, $B^*(f) = \infty$ для равномерной на $[0, 1]$ плотности и вместе с тем ее можно аппроксимировать последовательностью из $\mathscr{F}$ со значениями B_H, стремящимися к 1, т. е. к минимальному из возможных значений B_H. Таким образом, с помощью гистограммной оценки легче всего оценивать равномерную плотность.

Второе следствие теорем 5 и 6 состоит в том, что средняя ошибка в L_1 для гистограммной оценки изменяется со скоростью, в лучшем случае равной $n^{-1/3}$, что хуже, чем для ядерной оценки, имеющей скорость сходимости $n^{-2/5}$. При конечных n ответ на вопрос о том, лучше ли ядерная оценка, в большой степени зависит от значений $B_H(f)$ и $B^*(f)$.

Для гистограммной оценки нет свободных от распределения нижних границ, таких, как границы для ядерной оценки из теоремы 2. По-видимому, для их получения следует рассматривать критерии вида

$$\inf_h \sup_{x_0} E(J_n),$$

где $\mathscr{P}_n$ заменяется на $\mathscr{P}_n(x_0) = \{[kh + x_0,\ (k+1)h + x_0),\ k$ — целое$\}$. Здесь это не делается, так как уже установлено, что на большом подклассе класса $\mathscr{F}$ гистограммная оценка, несомненно, не является асимптотически оптимальной.

5. Доказательства теорем 5 и 6

В этом параграфе μ_n обозначает эмпирическую меру для $X_1, \ldots, X_n$, и μ — меру, соответствующую f. Пусть $A_{nj} = [jh, (j+1)h)$,

$$f_n(x) = \frac{\mu_n(A_{nj})}{h}, \quad x \in A_{nj},$$

и

$$g_n(x) = E(f_n(x)) = \frac{\mu(A_{nj})}{h}, \quad x \in A_{nj}.$$

На протяжении этого параграфа через T обозначается ограниченный интервал вида $[-t, t]$.

Лемма 13. *Если* $\int |g_n - f| \to 0$ *для некоторой плотности* $f \in \mathscr{F}$, *то* $\lim\limits_{n\to\infty} h = 0$.

Доказательство. Заметим, что лемма 13 не является следствием теоремы 2.6, так как ее условие выполнено не для всех f. Предположим сначала, что $\lim\limits_{n\to\infty} h = \infty$. Тогда $g_n \to 0$ при любых x и $\liminf\limits_{n\to\infty} \int |g_n - f| \geqslant \int \liminf\limits_{n\to\infty} |g_n - f| = 1$, что противоречит условию. Далее, предположим, что $\lim\limits_{n\to\infty} h = c > 0$. При $h = c$ из условия $\int |g_n - f| = 0$ следует, что $f' = 0$ п. в. Но в силу абсолютной непрерывности f последнее влечет за собой, что $f = 0$ п. в. Таким образом, заключаем, что $\int |g_n - f| > 0$ при $h = c > 0$. Будем теперь явно указывать зависимость от h, введя символ g_h. Ясно, что будет снова получено противоречие, если удастся показать, что из сходимости $h \to c$ следует $\int |g_h - g_c| \to 0$. Этим противоречием будет доказана лемма 13.

Чтобы получить это противоречие, достаточно показать, что $g_h \to g_c$ при почти всех x, и применить теорему 2.8. Предположим, не нарушая общности, что $x > 0$. Далее, $x \in [jh, (j+1)h) \cap [kc, (k+1)c)$ при некоторых целых j, k. При h, достаточно близких к c, имеем $j = k$. Пусть A — пересечение указанных двух интервалов и B — их симметрическая разность. Имеем

$$|g_h - g_c| \leqslant \mu(A)\left|\frac{1}{h} - \frac{1}{c}\right| + \mu(B)\left(\frac{1}{h} + \frac{1}{c}\right) \leqslant$$

$$\leqslant \frac{|h-c|}{hc} + o(1) = o(1).$$

Лемма 13 доказана.

Лемма 14. *Если случайная величина X имеет биномиальное распределение с параметрами (n, p), то*

$$\sup_i P(X = i) \leqslant c/\sqrt{np(1-p)},$$

где c — некоторая универсальная постоянная.

Доказательство. Воспользуемся тем, что $k! = (k+1)^k \times \times \sqrt{2\pi(k+1)}\, e^{-(k+1)} \exp(\theta/12(k+1))$ при некотором $0 < \theta < 1$ (см., например, Whittaker, Watson (1963)). Тогда, в силу того что все величины $P(X = i)$ меньше, чем k-я, где $k = [(n+1)p]$ (см. Feller (1968), с. 167 по русскому изданию), получим

$$P(X = i) \leqslant \left(1 + \frac{1}{n}\right)^n \frac{e^{1+1/24}\sqrt{n+1}}{\sqrt{2\pi(k+1)(n-k+1)}} \leqslant$$

$$\leqslant \frac{e^{2+1/24}\sqrt{2n}}{\sqrt{2\pi(n+1)p(n+1)(1-p)}} \leqslant \frac{c}{\sqrt{np(1-p)}};$$

где мы воспользовались неравенством $(1+u) \leqslant e^u$.

Лемма 15.

$$\inf_{f \in \mathcal{F}} B_H(f) = 1.$$

Доказательство. Известно, что $\int \sqrt{f} \geqslant \int f / \sqrt{\sup f} = 1/\sqrt{\sup f}$. Так как функция f абсолютно непрерывна, то $\int |f'| \geqslant 2\sup f$. Соединяя эти неравенства, получим

$$\left(\int \sqrt{f}\right)^2 \left(\int |f'|\right) / 2 \geqslant 1.$$

Для доказательства того, что эта нижняя граница достигается, построим последовательность плотностей $f = g * \varphi_h$ из $\mathcal{F}$ и устремим h к 0. Здесь φ — нормальная плотность с параметрами (0,1) и g — равномерная плотность на $[0,1]$.

Доказательство того, что $\int \sqrt{g * \varphi_h} \to \int \sqrt{g} = 1$, является простым упражнением из анализа.

Более того, так как свертка двух симметричных унимодальных распределений унимодальна (Feller (1971), с. 197 по русскому изданию), то каждая плотность f унимодальна. Следовательно,

$$\int |(g * \varphi_h)'| = 2\sup(g * \varphi_h) \to 2\sup g = 2$$

где $h \to 0$. Это показывает, что для рассматриваемой последовательности $\limsup B_H(f) \leqslant 1$.

Сохраним обозначения § 2 и 3. В частности, $\sigma_n^2(x) = \operatorname{var}(f_n(x)) = \operatorname{var}(\mu_n(A_{nj})/h)$, $x \in A_{nj}$ и $B_n(x) = g_n(x) - f(x)$. Кроме того,

$$E(J_n) = \sum_j E(J_n(A_{nj})),$$

где $J_n(A) = \int\limits_A |f_n - f|$. По аналогии с леммой 9 при любых j будем иметь

$$\left| E(J_n(A_{nj})) - \int\limits_{A_{nj}} \sigma_n(x)\psi\left(\frac{|B_n(x)|}{\sigma_n(x)}\right)dx \right| \leqslant \frac{c}{n}, \qquad (7)$$

где для величины ошибки из леммы 8 использована оценка

$$c\int\limits_{A_{nj}} \frac{E(|I_{A_{nj}}(X_1) - \mu(A_{nj})|^3/h^3)}{nE(|I_{A_{nj}}(X_1) - \mu(A_{nj})|^2/h^2)}dx \leqslant \frac{c}{nh}\int\limits_{A_{nj}} dx = \frac{c}{n}.$$

Кроме того, по аналогии с (5) имеем

$$E(J_n(T)) \geqslant \int\limits_T \sigma_n\psi\left(\int\limits_T |B_n| \Big/ \int\limits_T \sigma_n\right) - (c/n)(\lambda(T)/h + 2) \geqslant$$

$$\geqslant C_H\left(\int\limits_T \sigma_n\right)^{2/3}\left(2\int\limits_T |B_n|\right)^{1/3} - (c/n)(\lambda(T)/h + 2), \qquad (8)$$

поскольку $C_H = \inf\ \psi(u)/(2u)^{1/3}$.

Лемма 16. *Предположим, что* $\lim\limits_{n\to\infty} h = 0$. *Тогда* $\int\limits_T |\sigma_n\sqrt{nh} - \sqrt{f}| = o(1)$.

Доказательство. Пусть $x \in A_{nj}$. Тогда $\sigma_n^2(x) = \mu(A_{nj})(1 - \mu(A_{nj}))/(nh^2)$ и, следовательно,

$$|\sigma_n\sqrt{nh} - \sqrt{f}| = |\sqrt{\mu(A_{nj})(1 - \mu(A_{nj}))/h} - \sqrt{f}| \leqslant$$

$$\leqslant \mu(A_{nj})/\sqrt{h} + |\sqrt{\mu(A_{nj})/h} - \sqrt{f}| =$$

$$= g_n\sqrt{h} + |\sqrt{g_n} - \sqrt{f}| \leqslant g_n\sqrt{h} + \sqrt{|g_n - f|}.$$

Взяв теперь интегралы от левой и правой частей этого неравенства и воспользовавшись неравенством Коши—Шварца и теоремой 2.6, получим

$$\int\limits_T |\sigma_n\sqrt{nh} - \sqrt{f}| \leqslant \sqrt{h} + \int\limits_T \sqrt{|g_n - f|} \leqslant$$

$$\leqslant \sqrt{h} + \sqrt{\int\limits_T |g_n - f|\lambda(T)} = o(1).$$

Лемма 17. *Предположим, что* $\lim\limits_{n\to\infty} h = 0$ *и что плотность* $f \in \mathcal{F}$ *имеет компактный носитель. Тогда*

$$q_n(x) = o(h)$$

при любых x *и*

$$\int q_n = o(h),$$

где $q_n(x) = |B_n(x) - (h/2)\, z_n(x)|$. *При любых* $f \in \mathcal{F}$ *имеем*

$$\int_T q_n = o(h)$$

и

$$\int_T |B_n| \sim \frac{h}{2} \int_T |z_n| \sim \frac{h}{4} \int_T |f'|.$$

Доказательство. Используя разложение в ряд Тейлора с остаточным членом, получим

$$f(y) = f(x) + (y-x)f'(x) + (y-x)(f'(\xi) - f'(x)), \quad x \leqslant \xi \leqslant y.$$

Таким образом, при $x \in A_{nj}$ имеем

$$g_n(x) = \frac{1}{h} \int_{A_{nj}} f = f(x) + \frac{1}{h} \int_{A_{nj}} (y-x) f'(x)\, dy +$$

$$+ \frac{1}{h} \int_{A_{nj}} (y-x)(f'(\xi) - f'(x))\, dy = f(x) + \frac{h}{2} z_n(x) + b(x, h),$$

где функция $|b(x, h)| \leqslant \sup\limits_{|y-x| \leqslant h} |f'(y) - f'(x)|\, h/2$ не превосходит некоторой интегрируемой функции, умноженной на h (в силу того что f имеет компактный носитель и $\sup |f'| < \infty$), и $b(x, h)/h \to 0$ при любых x (в силу того что f' непрерывна). Таким образом, $\int q_n \leqslant \int |b(x, h)| = o(h)$ и $q_n(x) = o(h)$ при любых x. Если f имеет некомпактный носитель, то остается в силе соотношение $\int\limits_T q_n = o(h)$.

Таким образом, $\int\limits_T |B_n| \sim (h/2) \int\limits_T |z_n|$ при любых $f \in \mathcal{F}$. Доказательство леммы будет завершено, если показать, что $\int\limits_T |z_n| \sim \int\limits_T |f'|/2$ при $f \in \mathcal{F}$. Пусть теперь N_n — набор индексов j, для которых $A_{nj} \subseteq T$. Очевидно,

$$\sum_{j \notin N_n} \int_{T \cap A_{nj}} |f'|\, |1 - 2r_n| \leqslant 2h \sup |f'| = o(1).$$

Так как функция f' непрерывна на T, то она равномерно непрерывна на T. Именно, для фиксированного $\varepsilon > 0$ имеем $|f'(x) - f'(y)| < \varepsilon$ при любых $x, y, \in T$, таких, что $|x - y| \leqslant h$, где h — произвольное, достаточно малое число. Используя это, получим следующую цепочку неравенств:

$$\sum_{i \in N_n} \int_{T \cap A_{nj}} |f'||1 - 2r_n| \leqslant \sum_{i \in N_n} \sup_{T \cap A_{nj}} |f'| \int_{A_{nj}} |1 - 2r_n| =$$

$$= \sum_{i \in N_n} \sup_{T \cap A_{nj}} |f'| \int_0^h |1 - 2x/h| \, dx =$$

$$= \sum_{i \in N_n} \sup_{T \cap A_{nj}} |f'| h/2 \leqslant$$

$$\leqslant \sum_{i \in N_n} \int_{T \cap A_{nj}} (|f'| + \varepsilon)/2 + O(h) =$$

$$= \int_T |f'|/2 + \int_T \varepsilon/2 + O(h).$$

Точно так же получается нижняя граница $\int_T |f'|/2 - \int_T \varepsilon/2 - O(h)$. Лемма доказана.

Доказательство теоремы 5. Используя неравенство

$$\inf_h E(J_n) \geqslant \min\left(\inf_{h\sqrt{n} > 1} E(J_n), \inf_{h\sqrt{n} \leqslant 1} E(J_n)\right),$$

разобьем доказательство на две части. Пусть сначала последовательность h такова, что $E(J_n) \sim \inf_{h\sqrt{n} \geqslant 1} E(J_n)$. Известно, что $h \to 0$ (это имеет место в силу того, что $E(J_n) \geqslant \int |g_n - f|$ и по теореме 2.6 и лемме 13 последняя величина стремится к 0 при $f \in \mathscr{F}$ тогда и только тогда, когда $h \to 0$). Кроме того, $1/nh = o(n^{-1/3})$. Таким образом, объединяя (8) и леммы 16 и 17, получим

$$\inf_{h\sqrt{n} > 1} E(J_n) \sim E(J_n) \geqslant E(J_n(T)) \geqslant$$

$$\geqslant \int_T \sigma_n \psi\left(\int_T |B_n| \Big/ \int_T \sigma_n\right) - (c/n)(\lambda(T)/n + 2) \geqslant$$

$$\geqslant C_H \left(\int_T \sqrt{f}\right)^{2/3} \left(\int_T |f'|\right)^{1/3} (2n)^{-1/3} (1 + o(1))$$

для любых $T = [-t, t]$. Устремляя теперь T к R, заключаем, что

$$\liminf_{n\to\infty} \inf_{h\sqrt{n}\leqslant 1} E(J_n)\, n^{1/3} \geqslant C_H B_H(f). \tag{9}$$

Далее, рассмотрим последовательность, для которой $n^{1/3}E(J_n) \sim \inf_{h\sqrt{n}\leqslant 1} n^{1/3}E(J_n)$. По лемме Фату

$$n^{1/3}E(J_n) \geqslant \int \liminf_{n\to\infty} n^{1/3}E(|f_n - g_n|)/2.$$

Пусть $x \in A_{nj}$ фиксировано и Z — случайная величина, имеющая биномиальное распределение с параметрами $(n, \mu(A_{nj}))$. По лемме 14

$$\begin{aligned} n^{1/3}E(|f_n - g_n|) &= n^{1/3}(nh)^{-1}\, E(|Z - E(Z)|) \geqslant \\ &\geqslant MP(|Z - E(Z)| \geqslant Mhn/n^{1/3}) \geqslant \\ &\geqslant M\left(1 - \frac{2Mhn}{n^{1/3}} \frac{c}{\sqrt{n\mu(A_{nj})(1-\mu(A_{nj}))}}\right) \end{aligned} \tag{10}$$

при любом $M > 0$, где c — постоянная из леммы 14. В силу теоремы 2.2 имеем $\mu(A_{nj})h \to f(x)$ при почти всех x, так как $h \to 0$. Кроме того, $\mu(A_{nj}) \to 0$ при почти всех x. Таким образом, (10) $\sim M\left(1 - 2Mc\sqrt{hn^{1/3}/f(x)}\right) \sim M$ при почти всех x, таких что $f(x) > 0$. Так как M — произвольное число, то

$$\liminf_{n\to\infty} \inf_{h\sqrt{n}\leqslant 1} n^{1/3}E(J_n) = \infty. \tag{11}$$

Объединяя теперь (9) и (11) и применяя лемму 15, получим утверждение теоремы 5.

Доказательство теоремы 6. Рассмотрим сначала неравенство (7), считая, что T — множество вида $[-t, t]$, содержащее носитель f. Тогда

$$\left|E(J_n) - \int \sigma_n \psi(|B_n|/\sigma_n)\right| \leqslant \frac{c}{n}\left(\frac{\lambda(T)+2h}{h} + 2\right) = o\left(\frac{1}{nh}\right). \tag{12}$$

Пусть q_n обозначает то же, что в лемме 17, и через p_n обозначается величина $|\sigma_n - \sqrt{f/nh}|$. В лемме 16 было доказано, что $\int p_n = o(1/\sqrt{nh})$. Рассуждая так же, как в доказательстве теоремы 1, получим

$$\begin{aligned} &\left|\sigma_n\psi\left(\frac{|B_n|}{\sigma_n}\right) - \sqrt{\frac{f}{nh}}\,\psi\left(\frac{h}{2}|z_n|\sqrt{\frac{nh}{f}}\right)\right| \leqslant \\ &\leqslant \sigma_n\left|\psi\left(\frac{h}{2}\frac{|z_n|}{\sigma_n}\right) - \psi\left(\frac{|B_n|}{\sigma_n}\right)\right| + \\ &+ \left|\sigma_n\psi\left(\frac{h}{2}\frac{|z_n|}{\sigma_n}\right) - \sqrt{\frac{f}{nh}}\,\psi\left(\frac{h}{2}|z_n|\sqrt{\frac{hn}{f}}\right)\right| \leqslant q_n + \sqrt{\frac{2}{\pi}}\,p_n. \end{aligned} \tag{13}$$

Объединенные вместе соотношения (12), (13) и леммы 16 и 17 показывают, что $E(J_n) - J(n, h) = o(h + 1/\sqrt{nh})$.

Для доказательства второго утверждения теоремы 6 воспользуемся для $J(n, h)$ неравенством $\psi(u) \leqslant \sqrt{2/\pi} + u$. Получим

$$J(n, h) \leqslant \sqrt{\frac{2}{\pi}} \int \sqrt{\frac{f}{nh}} + \frac{h}{2} \int |z_n| \sim \sqrt{\frac{2}{\pi}} \int \sqrt{\frac{f}{nh}} + \frac{h}{4} \int |f'|.$$

Затем, выбирая значение h, минимизирующее эту верхнюю границу, т. е.

$$h = \left(\frac{8}{\pi} \left(\frac{\int \sqrt{f}}{\int |f'|} \right)^2 \frac{1}{n} \right)^{1/3},$$

получим требуемый результат.

6. Выбор параметра сглаживания

В теоремах 1 и 6 были получены точные асимптотические выражения для $E(J_n)$. К сожалению, последовательности h, минимизирующие главные члены этих асимптотических выражений, трудно выписать в явном виде. Для такой минимизации более подходящими являются верхние границы теорем 1 и 6. В случае ядерной оценки при выборе ядра Епанечникова $3(1 - x^2)/4$, $|x| \leqslant 1$ (для которого $\int K^2 = 3/5$ и $\int x^2 K = 1/5$) получим

$$h = \left[\sqrt{\frac{15}{2\pi}} \frac{\int \sqrt{f}}{\int |f''|} \right]^{2/5} n^{-1/5} \qquad (14)$$

при f, принадлежащих классу $\mathcal{F}_K$ плотностей, для которых справедлива теорема 1. Если f принадлежит классу $\mathcal{F}_H$ плотностей, для которых верна теорема 6, то соответствующее значение h для гистограммной оценки равно

$$h = \left[\sqrt{\frac{8}{\pi}} \frac{\int \sqrt{f}}{\int |f'|} \right]^{2/3} n^{-1/3}. \qquad (15)$$

Обратим внимание на то, что формулы (14) и (15) верны только при $f \in \mathcal{F}_K$ (или $f \in \mathcal{F}_H$) и, следовательно, только для f с конечной величиной $\int \sqrt{f}$ и $\int |f''|$ (или $\int |f'|$). Если один из этих интегралов бесконечен, то получается странный результат — оптимальное значение h приблизительно равно 0 или ∞. Это противоречие вызвано, разумеется, тем, что если один из участвующих в формуле членов равен бесконечности или нулю (см., например, верхнюю границу для $J(n, h)$ в доказательстве тео-

ремы 6), то формальная минимизация неправомерна. Так как трудно проверить все указанные условия на f, то любая попытка определить h путем аппроксимации соотношений (14) или (15) вызовет у пользователей серьезные опасения.

Выбор параметра h рассматривается также в § 5.8—5.10 и в гл. 6 и 9. В частности, в гл. 6 и 9 приводятся очень общие теоремы состоятельности для оценок плотности с зависящими от данных параметрами сглаживания. Здесь же мы хотели бы указать лишь определенные черты *параметрического метода* определения h.

Предположим, что плотность f близка к некоторому элементу параметрического семейства f_θ, $\theta \in R^k$ или принадлежит этому семейству. В силу предположения для этого семейства величины (14) и (15) известны в явном виде и равны $c_K(\theta)/n^{1/5}$ и $c_H(\theta)/n^{1/3}$. Построим обычным образом по набору данных оценку $\hat{\theta}$ параметра θ и воспользуемся величинами $c_K(\hat{\theta})/n^{1/5}$ и $c_H(\hat{\theta})/n^{1/3}$ вместо (14) и (15). Желательно использовать робастные оценки $\hat{\theta}$, если это в принципе возможно. Этот метод, называемый параметрическим, был предложен для выбора оптимальных в L_2 значений параметра h ядерной оценки в работах Deheuvels (1977) и Deheuvels, Hominal (1980). Похожие рассмотрения в случае гистограммной оценки можно найти у Scott (1979) и Freedman, Diaconis (1981). В частности, для плотностей, удовлетворяющих условию Розенблата (§ 4.3), оптимальное в L_2 значение параметра h ядерной оценки равно

$$h = \left(\frac{15}{n \int (f'')^2} \right)^{1/5} \tag{16}$$

(в случае ядра Епанечникова), а для плотностей из теоремы 4 оптимальное в L_2 значение параметра h гистограммной оценки равно

$$h = \left(\frac{6}{n \int (f')^2} \right)^{1/3}. \tag{17}$$

Выражения (16), (17) имеют очень мало сходства с (14), (15), за исключением зависимости от n. Любопытно, что даже для самых простых плотностей, таких, как нормальная плотность с параметрами (0, 1), эти величины различны. Для такой плотности величины (16) и (17), например, равны

$$h = \left(\frac{40\sqrt{\pi}}{n} \right)^{1/5} = 2{,}345 \ldots n^{-1/5}$$

и

$$h = \left(\frac{12\sqrt{\pi}}{n} \right)^{1/3} = 2{,}7706 \ldots n^{-1/3}$$

соответственно. Вместе с тем величины (14) и (15) равны

$$h = \left(\frac{15e\sqrt{2\pi}}{8n}\right)^{1/5} = 1{,}6644\ldots n^{-1/5}$$

и

$$h = \left(4\frac{\sqrt{8\pi}}{n}\right)^{1/3} = 2{,}7168\ldots n^{-1/3}$$

соответственно. (Легко проверить, что формулы (14), (15) остаются в силе для нормальной плотности [1].) Можно, разумеется, построить более экстремальные примеры. Так, если приближать плотность распределения Коши с помощью семейства t-плотностей Стьюдента, то коэффициент при n в (16) стремится к некоторой постоянной, тогда как коэффициент при n в (14) стремится к ∞. Это вызвано тем, что L_1-теория более чувствительна к тяжелым хвостам распределений. В табл. 1 приведены значения $\int \sqrt{f}$, $\int |f'|$, $\int |f''|$, $B^*(f) = \left[\left(\int \sqrt{f}\right)^4 \int |f''|/2\right]^{1/5}$, $B_H(f) = \left[\left(\int \sqrt{f}\right)^2 \int |f'|/2\right]^{1/3}$, $c_K(\theta)$ и $c_H(\theta)$ для различных семейств распределений. Если $f \notin \mathscr{F}_H$, то вместо $\int |f'|$ стоит предел величин $\int |f'|$ для последовательности плотностей из $\mathscr{F}_H$, стремящейся к f. Если $f \notin \mathscr{F}_K$, то вместо величины $\int |f''|$ используется ее обобщение $\sup_{a>0} \int |(f * \varphi_a)''|$.

В рассматриваемых семействах (и, следовательно, в выражениях для $c_K(\theta)$ и $c_H(\theta)$) нет необходимости указывать параметры сдвига и масштаба, поскольку

$$c_K(\mu, \sigma, \gamma) = \sigma c_K(0, 1, \gamma)$$

и

$$c_H(\mu, \sigma, \gamma) = \sigma c_H(0, 1, \gamma),$$

где μ — параметр сдвига, σ — параметр масштаба и γ — набор параметров формы, возможно пустой. Вычисление величин из табл. 1 упрощается тем, что $\int |f'| = 2 \sup f$ для унимодальных плотностей f из $\mathscr{F}_H$ и что $\int |f''| = 4 \sup |f'|$ для симметричных $f \in \mathscr{F}_K$ с унимодальной на $[0, \infty)$ производной f'.

Ни одна из плотностей в табл. 1 не имеет параметра формы. Так как параметр масштаба можно выделить в виде множителя,

---

[1]) То есть утверждается, что неравенства из теорем 1, 6 верны в случае нормальной плотности f (ср. с § 10 данной главы, где даются аналогичные неравенства, но с другими, более грубыми, числовыми коэффициентами). — *Прим. перев.*

Таблица 1

Плотность	$\int \sqrt{f}$	$\int \lvert f' \rvert$	$\int \lvert f'' \rvert$	$B^*(f)$	$B_H(f)$	$c_K(\theta)$	$c_H(\theta)$
Равномерная на [0, 1]	1	2[1]	∞[2]	∞	1 (минимальное значение)	0	$(2/\pi)^{1/3}$
Равнобедренная треугольная на [0, 1]	$(2^{3/2})/3$	4[1]	16[2]	$(2^9/3^4)^{1/5}$ (минимальное значение)	$(16/9)^{1/3}$	$(5/192\pi)^{1/5}$	$(4/9\pi)^{1/3}$
Нормальная с параметрами (0, 1)	$(8\pi)^{1/4}$	$\sqrt{2/\pi}$	$\sqrt{8/\pi e}$	$(128\pi/e)^{1/10}$	$2^{1/3}$	$(225\pi e^2/32)^{1/10}$	$(4\sqrt{8\pi})^{1/3}$
Лапласа $(e^{-\lvert x \rvert}/2)$	$4/\sqrt{2}$	1[1]	2[2]	$64^{1/5}$	$4^{1/3}$	$(15/\pi)^{1/5}$	$(64/\pi)^{1/3}$
Экспоненциальная $(e^{-x},\ x > 0)$	2	2[1]	∞[2]	∞	$4^{1/3}$	0	$(8/\pi)^{1/3}$
Коши $(\pi(1 + x^2))^{-1}$	∞	$2/\pi$	$9/2\pi\sqrt{3}$	∞	∞	∞	∞
$1/2(2\sqrt{x}),\ 0 \leqslant x \leqslant 1$	$4/3\sqrt{2}$	∞[1]	∞[2]	∞	∞	0	0
Стьюдента $t3$, $\left(\frac{2/\pi\sqrt{3}}{(1 + x^2/3)^2}\right)$	$\sqrt{2\pi}3^{1/4}$	$4/(\pi\sqrt{3})$	$500/81\pi\sqrt{5}$	$(2000\pi/54\sqrt{5})^{1/5}$	$4^{1/3}$	$(81\pi 3^{3/4}/100)^{2/5}$	$(\pi 3^{3/4})^{2/3}$

[1]) Предел последовательности плотностей из $\mathscr{F}_H$.

[2]) Обобщенное определение $\int \lvert f'' \rvert$.

то в табл. 1 содержится полная информация относительно $c_K(\theta)$ и $c_H(\theta)$. Например, если плотность f близка к нормальной, то для гистограммной оценки можно было бы взять

$$h = (\sqrt{8\pi})^{1/3}\hat{\sigma}n^{-1/3},$$

где $\hat{\sigma}$ — оценка для σ, построенная по данным в предположении, что они получены из плотности нормального распределения с параметрами (μ, σ^2).

В общем случае можно строить робастные оценки для σ следующим образом. Выберем два числа, $p, q \in (0, 1)$ и предположим, что известны квантили x_1 и x_2, соответствующие значениям p и q функции распределения с параметрами $\mu = 0$, $\sigma^2 = 1$. Порядковые статистики $X_{(np)}$ и $X_{(nq)}$ набора $X_1, \ldots, X_n$ можно, таким образом, рассматривать как оценки для $\mu + \sigma x_1$ и $\mu + \sigma x_2$ соответственно. Следовательно, можно оценивать σ величиной $(X_{(nq)} - X_{(np)})/(x_2 - x_1)$. Оценки такого типа обычно более предпочтительны, чем оценки, основанные на усреднении. Например, если в плотность распределения Коши из табл. 1 введен параметр масштаба, то при $p = 1/4$, $q = 3/4$ для него получается оценка $(X_{(3n/4)} - X_{(n/4)})/2$ (в случае распределения Коши усреднение было бы абсурдно). Имеется обширная литература со сложным теоретическим обоснованием по робастным оценкам параметра масштаба плотности нормального распределения. Одним из примеров подробно изученных оценок параметра σ плотности нормального (μ, σ^2)-распределения является

$$\hat{\sigma} = 0{,}1174\,(X_{(0{,}9765n)} - X_{(0{,}0235n)} + 2\,(X_{(0{,}8721n)} - X_{(0{,}1279n)}))$$

(Kulldorf (1963, 1964)).

В табл. 1 заключена и другая полезная информация. В § 4 и 5 было установлено, что величины $B^*(f)$ и $B_H(f)$ являются мерами сложности оценивания f с помощью ядерного и гистограммного методов соответственно. Мы знаем, что посредством гистограммной оценки легче всего оценивать равномерную на $[0, 1]$ плотность f (B_H достигает своего минимального значения 1). Таким образом, для гистограммной оценки не является препятствием то, что плотность f разрывна, если она ограниченна. Для неограниченных плотностей, таких, как $1/2\sqrt{x}$, имеем $B_H(f) = \infty$ в силу того, что $\int |f'| = \infty$. Конечно, ни гистограммная, ни ядерная оценки не могут очень хорошо работать для плотностей с тяжелыми хвостами (см., например, плотность распределения Коши, для которой $B^*(f) = B_H(f) = \infty$). Укажем также на различие между экспоненциальной и двойной экспоненциальной плотностями в случае ядерной оценки, которое возникает из-за того, что у двойной экспоненциальной плотности нет разрыва в нуле. Точно так же $B^*(f) = \infty$ для равномерной на $[0, 1]$

плотности. Наилучшей плотностью для ядерной оценки является равнобедренная треугольная.

Таблица 1 слегка дезориентирует. Она приводит к мысли о том, что, скажем, для равномерной на [0, 1] плотности при предлагаемых методах выбора h лучше использовать гистограммную оценку. Это неправильно — таблица просто *показывает*, что оптимальная скорость сходимости величины $E(J_n)$, равная $n^{-2/5}$, не достигается на ядерной оценке. Хорошим упражнением будет доказательство того, что для ядерной оценки и равномерной на [0, 1] плотности наилучшая скорость сходимости величины $E(J_n)$ равна $n^{-1/3}$, и что эта скорость достигается, если h меняется как $c/n^{1/3}$ (см. § 7 ниже).

Все плотности из табл. 1 — унимодальные. Интуитивно ясно, что мультимодальные плотности оценивать сложнее. Изучим то, как влияет наличие нескольких максимумов на величины B^* и B_H. С этой целью рассмотрим основную плотность f с носителем [0, 1] и определим следующую мультимодальную плотность:

$$g(x) = \frac{1}{N} \sum_{i=1}^{N} f(x - 2i).$$

Так как $\int \sqrt{g} = \sqrt{N} \int \sqrt{f}$, $\int |g'| = \int |f'|$ и $\int |g''| = \int |f''|$, то

$$B^*(g) = N^{2/5} B^*(f)$$

и

$$B_H(g) = N^{1/3} B_H(f).$$

Присутствие нескольких максимумов, по-видимому, хуже влияет на ядерную оценку, чем на гистограммную. Параметрический метод, который в основных чертах описан выше, следует применять для мультимодальных плотностей с исключительной осторожностью. Хороший подход состоит в том, чтобы изолировать различные максимумы и разбить задачу оценивания на несколько простых подзадач с унимодальными плотностями.

Выше было показано, что если плотность f ведет себя достаточно хорошо, то оптимальное значение h для ядерной оценки имеет вид: $h = cn^{-1/5}$. Важно понять, насколько ухудшается качество оценки, т. е. величина $E(J_n)$, в случае когда значение c не оптимально. Изучение этого обычно называют *анализом чувствительности*.

Аналитически слишком трудно исследовать точное выражение для $J(n, h)$, приведенное в теореме 1; вместе с тем верхняя граница для $J(n, h)$ равна $n^{-2/5}(ac^{-1/2} + bc^2)$, где

$$a = \sqrt{\frac{2}{\pi}}\alpha \int \sqrt{f}, \quad b = \frac{\beta}{2} \sup_{u>0} \int |(f * \varphi_u)''|.$$

Легко проверить, что эта верхняя граница как функция от c достигает минимума при $c = (a/4b)^{2/5}$, когда ее значение равно произведению $n^{-2/5}$ на $C^*A(K)B^*(f)$. Обозначим это оптимальное c через c_0. Если истинное c равно rc_0, то верхняя граница для $J(n, h)$ равна произведению $n^{-2/5}$ на $B^*(f)A(K)C^*$ и на $H(r)$, где $H(r) = r^2/5 + 4/5\sqrt{r}$. При $c = c_0/r$ дополнительный множитель в верхней границе для $J(n, h)$ равен $G(r) = 1/5r^2 + 4\sqrt{r}/5$. Обе функции, $H(r)$ и $G(r)$, разумеется, достигают минимума в точке $r = 1$, где их значение равно 1. Покажем теперь, что $G(r) < H(r)$ при любых $r > 1$, т. е. что лучше занизить, чем завысить c_0 в r раз.

Доказательство неравенства $G(r) < H(r)$. Требуется доказать, что $4r^{3/2} + r^4 > 4r^{5/2} + 1$ при $r > 1$. Положим $r = 1 + u$ и заметим, что из усеченного разложения в ряд Тейлора следует

$$\begin{aligned} 4r^{3/2} + r^4 &\geqslant 4(1 + 3u/2) + (1 + 6u + 10u^2 + 6u^3 + u^4) = \\ &= 5 + 12u + 10u^2 + 6u^3 + u^4 > \\ &> 5 + 10u + 15u^2/2 + 15u^3/2 = \\ &= 1 + 4\left(1 + \frac{5u}{2} + \left(\frac{5}{2}\cdot\frac{3}{2}\right)\frac{u^2}{2!} + \left(\frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\right)\frac{u^3}{3!}\right) \geqslant \\ &\geqslant 1 + 4(1 + u)^{5/2} = 1 + 4r^{5/2}. \end{aligned}$$

Наконец, заметим, что отношение $H(r)/G(r)$ изменяется следующим образом: 1 ($r = 1$), 1,038 ... ($r = 1{,}21$), 1,156 ... ($r = 2$), 2,232 ... ($r = 4$) и 6,787 ... ($r = 9$).

Закончим этот параграф одним интересным замечанием относительно оптимальных значений (14) и (15) параметра h в случае, когда вместо $\int |f''|$ и $\int |f'|$ используются обобщенные определения этих величин. Замечание состоит в том, что при *любых* f имеют место неравенства

$$h \leqslant \left(\frac{98415\pi^4}{65536}\right)^{1/5} \sigma n^{-1/5} = 6{,}7726100\ldots\,\sigma n^{-1/5} \qquad (18)$$

для ядерной оценки и

$$h \leqslant (16\pi^2)^{1/3}\,\sigma n^{-1/3} = 5{,}40513538\ldots\,\sigma n^{-1/3} \qquad (19)$$

для гистограммной оценки, где $\sigma = \sqrt{\mathrm{Var}\,(X)}$ — стандартное уклонение f. Чтобы убедиться в этом, воспользуемся соотношениями (14) и (15) вместе с теоремой 5.3 ($B^*(f) \geqslant (2^9/3^4)^{1/5}$), теоремой 5.9 ($B_H^*(f) \geqslant 1$, см. § 8 ниже) и леммой 5.1 $\left(\left(\int\sqrt{f}\right)^2 \Big/ \sqrt{\mathrm{Var}\,(X)} \leqslant 2\pi\right)$. Если заменить σ хорошей оценкой, построенной по выборке, то данные неравенства можно использовать для получения очень грубых, но полезных верхних границ для h при *полном* отсутствии информации относительно f.

7. Равномерная плотность

Равномерная на [0, 1] плотность f заслуживает отдельного рассмотрения, поскольку из того, что она разрывна, следует, что $\liminf_{n\to\infty} h^{2/5} E(J_n) = \infty$ для ядерной оценки. В этом параграфе будет показано, что при подходящем выборе h величина $E(J_n)$ убывает как $n^{-1/3}$. Материал этого параграфа можно перенести на плотности с более ярко выраженными разрывами, такие, как бета-плотности

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}, \quad 0 < x < 1, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

где можно положить $a = b \in (0, 1]$. Предлагается показать, что оптимальная скорость сходимости $E(J_n)$ к 0 равна $n^{-\varepsilon}$, где ε зависит только от a и может быть выбрано сколь угодно близким к 0.

В ходе проводимого анализа следует упомянуть несколько интересных побочных результатов.

Лемма 18. *При любых* $p > 0$ *имеем*

$$\inf_K \left(\int K^2\right)^p \int |x|^p K \geqslant \left(\frac{p+1}{2p+1}\right)^p \cdot \frac{1}{2p+1},$$

где нижняя грань берется по всем плотностям K *на* R^1. *Нижняя грань достигается при*

$$K(x) = \frac{p+1}{2p}(1 - |x|^p), \quad |x| \leqslant 1.$$

При $p = 2$ этот результат принадлежит Епанечникову (1969) и Bartlett (1963).

Доказательство. Для плотности K, определенной в утверждении леммы 18, имеем $\int K = 1$, $\int |x|^p K = 1/(2p+1)$ и $\int K^2 = (p+1)/(2p+1)$. Так как величина $\left(\int K^2\right)^p \int |x|^p K$ масштабно-инвариантна, то достаточно взять нижнюю грань по всем плотностям K, таким, что $\int |x|^p K = 1/(2p+1)$. Эта нормировка предполагается выполненной для всех рассматриваемых здесь плотностей. Любую плотность g можно записать в виде

$g = K + g^*$, где $\int g^* = 0$ и $\int |x|^p g^* = 0$ (поскольку $\int |x|^p g = \int |x|^p K$). Таким образом,

$$\int g^2 = \int K^2 + \int (g^*)^2 + 2\int_{-1}^{1} \frac{p+1}{2p}(1-|x|^p)g^* =$$

$$= \int K^2 + \int (g^*)^2 + \frac{p+1}{p}\int_{[-1,\,1]^c} (|x|^p - 1)g^* \geqslant \int K^2 + \int (g^*)^2$$

в силу того, что $g^* \geqslant 0$ на $[-1,1]^c$ (иначе функция g не была бы плотностью) и $|x|^p \geqslant 1$ на $[-1,1]^c$. Правая часть этого неравенства минимальна, если $g^* = 0$ почти всюду. Лемма 18 доказана.

Лемма 19. *Пусть* $B_n = E(f_n) - f$. *Тогда для ядерной оценки* (1) *при любых* $K \in L_1$ *имеем*

$$E\left(\int |f_n - f|\right) \geqslant \int |B_n| \sim 2h\int |x| K(x)\,dx, \quad h \to 0.$$

Доказательство. Первое неравенство следует из границы Йенсена. Для доказательства второго соотношения заметим, что величина $E(f_n)$ не превосходит единицы на $[0, 1]$, так что $\int |B_n| = 2\int\limits_{[0,\,1]^c} E(f_n)$.

Но если плотности K соответствует функция распределения F и Y — случайная величина с плотностью K, то

$$\int\limits_{[0,\,1]^c} E(f_n) = \int\limits_{[0,\,1]^c} \left(F\left(\frac{1-x}{h}\right) - F\left(-\frac{x}{h}\right)\right)dx =$$

$$= \int_{-\infty}^{0} (-P(hY > 1-x) + P(hY > -x))\,dx +$$

$$+ \int_{1}^{\infty} (P(hY < 1-x) - P(hY < -x))\,dx =$$

$$= \int_{0}^{\infty} (P(hY < -x) + P(hY > x))\,dx +$$

$$+ \int_{1}^{\infty} (-P(hY > x) - P(hY < -x))\,dx =$$

$$= hE(|Y|) + \tilde{O}(h)$$

при $E(|Y|) < \infty$ и $h \to 0$.

Лемма 20. *Если* $h \to 0$, $nh \to \infty$ *и* K — *ограниченная плотность с компактным носителем, то* $E(J_n) \geqslant (\alpha + o(1))\sqrt{2/\pi nh}$, *где* $\alpha = \sqrt{\int K^2}$.

Доказательство. Из леммы 9 и неравенства $\psi(u) \geqslant \sqrt{2/\pi}$ следует

$$E(J_n) \geqslant \int_T \sqrt{\frac{2}{\pi}}\,\sigma_n - \frac{cK^*\lambda(T^*)}{nh},$$

где $T = [0, 1]$. Кроме того, $\lambda(T^*) \to 1$, так как $h \to 0$. В силу леммы 10

$$\int_T \sigma_n \sim \frac{\alpha}{\sqrt{nh}} \int \sqrt{f} = \frac{\alpha}{\sqrt{nh}}.$$

Лемма 20 доказана.

Теорема 7. *Если* K — *ограниченная плотность с компактным носителем, то для ядерной оценки* (1), *примененной к равномерной на* $[0, 1]$ *плотности* f, *имеем*

$$\liminf_{n\to\infty} \inf_{h>0} n^{1/3} E(J_n) \geqslant \left(\frac{4}{\pi}\int K^2 \int |x|\,K\right)^{1/3} \geqslant \left(\frac{8}{9\pi}\right)^{1/3}.$$

Доказательство. Воспроизводя доказательство теоремы 2, получим

$$\inf_h E(J_n) \geqslant \max_{t>0} \min\left(2t\int |x|\,K, \sqrt{\frac{2\alpha}{\pi nt}}\right)(1 + o(1)),$$

где мы применили леммы 19 и 20. Данный максимум достигается при $t = \left(\alpha^2 \Big/ \left(2\pi n\left(\int |x|\,K\right)^2\right)\right)^{1/3}$. Подстановка этого значения t в максимизируемую функцию дает средний член в цепочке неравенств теоремы 7. Применяя лемму 18 при $p = 1$, получим нижнюю границу, не зависящую от K.

Теорема 8. *Если* K — *ограниченная плотность с компактным носителем, то для ядерной оценки* (1), *примененной к равномерной на* $[0, 1]$ *плотности* f, *выполнено неравенство*

$$\limsup_{n\to\infty} \inf_{h>0} n^{1/3} E(J_n) \leqslant \left(\left(\frac{8}{\pi}\right)^{1/3} + \left(\frac{1}{\pi}\right)^{1/3}\right)\left(\int K^2 \int |x|\,K\right)^{1/3}.$$

Доказательство. Пусть $T = [0, 1]$, и пусть все символы, используемые без определения, обозначают то же, что в § 2 и 3. Из доказательства теоремы 1 следует

$$E(J_n(T)) = \int_T \sigma_n \psi\left(\frac{|B_n|}{\sigma_n}\right) + O\left(\frac{1}{nh}\right).$$

В силу леммы 10

$$\int_T \sigma_n = (\alpha + o(1)) \int \frac{\sqrt{f}}{\sqrt{nh}} = \frac{\alpha + o(1)}{\sqrt{nh}}.$$

Далее, в силу неравенства $\psi(u) \leqslant u + \sqrt{2/\pi}$ и леммы 19

$$E(J_n) = E(J_n(T^c)) + E(J_n(T)) = E\left(\int_{T^c} f_n\right) + E(J_n(T)) \leqslant$$

$$\leqslant \int_{T^c} E(f_n) + \sqrt{\frac{2}{\pi}} \int_T \sigma_n + \int_T |B_n| =$$

$$= 2h \int |x| K + o(h) + \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{nh}} + o\left(\frac{1}{\sqrt{nh}}\right).$$

Главные члены этой верхней границы достигают минимума при $h^{3/2} = (\alpha \sqrt{2/\pi}) / \left(4 \int |x| K \sqrt{n}\right)$. Получаем

$$n^{1/3} E(J_n) \leqslant \left(\frac{2}{\pi} \int |x| K \int K^2\right)^{1/3} (4^{1/3} + 2^{-1/3} + o(1)),$$

что и требовалось доказать.

Здесь не исследовалось точное асимптотическое поведение оптимальной ядерной оценки. Тем не менее с помощью незамысловатых рассуждений были получены верхняя граница (теорема 8) и нижняя граница (теорема 7), отношение которых равно

$$2^{1/3} + 2^{-2/3} = 1{,}8898816\ldots.$$

Границы верны при любых ограниченных K с компактным носителем. Значение h, при котором достигается верхняя граница теоремы 8, равно

$$h = \left(\frac{\int K^2}{\left(\int |x| K\right)^2 8\pi n}\right)^{1/3}.$$

Таким образом, если имеющейся информации недостаточно для исключения возможности того, что f — равномерная плотность на $[0, 1]$, то опасно позволять параметру h изменяться по закону $c^* n^{-1/5}$. В самом деле, при этом из леммы 19 явствует, что для равномерной плотности

$$E(J_n) \geqslant \left(\int |x| K + o(1)\right) 2c^* n^{-1/5},$$

т. е. скорость сходимости заведомо ниже, чем оптимальная скорость, указанная в теоремах 7 и 8.

В заключение заметим, что для равномерной плотности f оптимальным является не ядро Епанечникова (оно оптимально

для узкого класса $\mathscr{F}$ из теоремы 1), а равнобедренное треугольное ядро $1 - |x|$, $|x| \leqslant 1$. Это следует из теорем 7 и 8 и леммы 18. Для остальных членов семейства бета-распределений получаются другие оптимальные ядра, форма которых зависит от характера имеющегося разрыва.

8. Минимаксная стратегия выбора параметра сглаживания

Возможны случаи, когда гладкость плотности f вызывает сомнения, например, если есть подозрение, что f имеет разрыв. Для таких случаев теорема 1 не дает никаких указаний по выбору параметра h ядерной оценки. Действительно, если f — равномерная плотность на $[0, 1]$, то, как показано в § 7, следует опасаться выбора h в виде произведения некоторой постоянной на $n^{-1/5}$. Можно действовать осторожно, увеличивая класс плотностей, находя на нем верхнюю границу для $E(J_n)$ и получая h в результате минимизации данной верхней границы. В этом состоит минимаксная смешанная стратегия. Стратегия такого типа для ядерной оценки в случае критерия L_2 разработана в статье Wertz (1972).

Рассмотрим сначала класс плотностей $\mathscr{F}$ из § 4, т. е. класс всех абсолютно непрерывных плотностей f, имеющих ограниченную и непрерывную п. в. производную f' и, следовательно, таких, что $\int |f'| < \infty$. Для плотностей из этого класса был определен коэффициент

$$B_H(f) = \left(\left(\int \sqrt{f}\right)^2 \int |f'|/2\right)^{1/3}.$$

Несколько обобщая определение этого коэффициента, положим

$$B_H^*(f) = \left(\left(\int \sqrt{f}\right)^2 \sup_{a>0} \int |(f * \varphi_a)'|/2\right)^{1/3},$$

где $\varphi \in \mathscr{F}$ — некоторая непрерывно-дифференцируемая плотность с компактным носителем.

Основное утверждение этого параграфа состоит в том, что если выбрать h в виде произведения некоторой постоянной на $n^{-1/3}$, то $E(J_n) = o(n^{-1/3})$ для любых плотностей с компактным носителем, для которых величина $B_H^*(f)$ конечна. Этим условиям удовлетворяют все абсолютно непрерывные плотности с компактным носителем и даже плотности с простыми разрывами, такие, как равномерная плотность на $[0, 1]$. Установим ряд свойств, аналогичных тем, которые найдены в § 2 и 3. В некоторых леммах дадим лишь наброски доказательств.

Лемма 21. *Пусть f и K — произвольные плотности на R. Тогда*

$$\int |f * K_h - f| \leqslant h\gamma \liminf_{a \downarrow 0} \int |(f * \varphi_a)'|$$

при любых $h > 0$, где $\gamma = \int |x| K$.

Доказательство. Пусть сначала $f \in \mathscr{F}$. Так как

$$f(y) - f(x) = \int_x^y f'(z)\, dz,$$

то

$$\int |f * K_h - f| = \int \left| \int \tilde{K}\left(\frac{x-y}{h}\right) f'(y)\, dy \right| dx,$$

где

$$\tilde{K}(x) = \begin{cases} \int_x^\infty K(z)\, dz, & x \geqslant 0, \\ \int_{-\infty}^x K(z)\, dz, & x < 0. \end{cases}$$

Отсюда следует, что

$$\int |f * K_h - f| \leqslant h \int |f'| \int |\tilde{K}| = h \int f |f'| \gamma.$$

Обобщение на случай произвольной плотности f проводится так же, как в лемме 4 (ii); оно основано на том, что функция $f * \varphi_a$ принадлежит классу $\mathscr{F}$ при любых f и любых $a > 0$.

Теорема 9. *$B_H^*(f) = B_H(f)$ при любых $f \in \mathscr{F}$. При любых f имеет место граница $B_H^*(f) \geqslant 1$, которая достигается на равномерной плотности на $[0, 1]$.*

Доказательство. Доказательство первого утверждения не представляет большого труда (см., например, аналогичное доказательство в лемме 2), и оно опускается. Второе утверждение частично, т. е. для всех $f \in \mathscr{F}$, доказано в лемме 15. Покажем здесь, что оно справедливо при любых f. Используя неравенства

$$\int \sqrt{f} \geqslant 1/\sqrt{\sup f} \quad \text{и} \quad \int |(f * \varphi_a)'| \geqslant 2 \sup (f * \varphi_a)$$

(см. лемму 15), получим

$$B_H^*(f)^3 \geqslant \sup (f * \varphi_a)/\sup f$$

при любых $a > 0$. Но так как $f * \varphi_a \to f$ для почти всех x при $a \downarrow 0$ (теорема 2.3), то ясно, что $\sup_{a>0} \sup (f * \varphi_a) \geqslant \sup f$ (между

прочим, всегда $\sup(f * \varphi_a) \leqslant \sup f$) и, таким образом, $B_H^*(f)^3 \geqslant \geqslant 1$.

Теорема 10. *Пусть K — ограниченная плотность с компактным носителем, и пусть h удовлетворяет условию* (3). *Тогда для ядерной оценки* (1) *и любой плотности f с компактным носителем*

$$E(J_n) \leqslant \sqrt{\frac{2}{\pi}} \frac{\alpha \int \sqrt{f}}{\sqrt{nh}} + h\gamma \sup_{a>0} \int |(f * \varphi_a)'| + o((nh)^{-1/2}).$$

Более того,

$$\limsup_{n\to\infty} \inf_h n^{1/3} E(J_n) \leqslant C_1^* A_1(K) B_H^*(f),$$

где

$$C_1^* = 3/\pi^{1/3} = 2{,}0483522\ldots$$

и

$$A_1(K) = (\alpha^2\gamma)^{1/3}.$$

Если величина $B_H^*(f)$ конечна, то верхняя граница $C_1^* A_1(K) B_H^*(f)$ не нарушается при следующем выборе параметра h:

$$h = \left[\sqrt{\frac{2}{\pi}} \frac{\alpha \int \sqrt{f}}{2\gamma \sup_{a>0} \int |(f * \varphi_a)'|}\right]^{2/3} n^{-1/3}.$$

Доказательство. Так же как в доказательстве теоремы 1, на ограниченных интервалах T имеем

$$\int_T E(|f_n - f|) \leqslant \int_T \left(\sqrt{\frac{2}{\pi}}\sigma_n + |B_n|\right) + \frac{cK^*}{nh}\lambda(T) \leqslant$$

$$\leqslant \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{nh}} \int \sqrt{f} + h\gamma \sup_{a>0} \int |(f * \varphi_a)'| + o((nh)^{-1/2}),$$

где мы воспользовались леммами 10 и 21. Если K имеет компактный носитель, то, выбирая интервал T достаточно большим, получим, что $E(J_n) = E(J_n(T))$ при любых n, откуда следует искомый результат.

Главные члены верхней границы имеют вид $uh^{-1/2} + vh$. Это выражение как функция от h достигает при $h = (u/2v)^{2/3}$ минимума, равного $(u^2v)^{1/3}\, 3/4^{1/3}$. Но последнее выражение можно записать в виде $C^* A_1^*(K) B_H^*(f)$, и теорема доказана.

Здесь надо сделать несколько замечаний. Во-первых, в силу леммы 18 величина $A_1(K)$ больше или равна $(2/9)^{1/3}$, и это минимальное значение достигается на равнобедренной треугольной плотности $1 - |x|$ на $[-1, 1]$. Если подставить это значение

A_1 (K) в верхнюю границу, то она станет равной $(6/\pi)^{1/3}B_H^*$ $(f) = 1{,}240701\ldots B_H^*$ (f). Данный результат лучше, чем верхняя граница теоремы 6 для гистограммной оценки, и он очень близок к нижней границе теоремы 5 для той же оценки. Если используется оптимальное ядро K, то величина h, указанная в теореме 10, равна

$$h = \left(\frac{3}{\pi}\right)^{1/3}\left(\frac{\int \sqrt{f}}{\sup\limits_{a>0}\int |(f * \varphi_a)'|}\right)^{2/3} n^{-1/3}$$

(достаточно подставить $\alpha = \sqrt{2/3}$ и $\gamma = 1/3$ в формулу для h).

Наконец, так как верхняя граница теоремы 10 достигает минимума на равномерной на $[0, 1]$ плотности (см. теорему 9), то перед построением ядерной оценки пользователю выгодно преобразовать данные, стремясь сделать их как можно «более равномерными».

9. Классы Липшица, классы Бретаньоля—Юбер и равномерные верхние границы

Так же как в § 4.2, обозначим $W(s, \alpha, C)$ *класс Липшица* с параметрами s, α, C, т. е. класс всех плотностей на $[0, 1]$, имеющих $(s-1)$ абсолютно непрерывных производных и таких, что

$$|f^{(s)}(x) - f^{(s)}(y)| \leqslant C|x-y|^{\alpha}$$

при любых x, $y \in R$. Здесь $\alpha \in (0, 1]$, s — неотрицательное целое число и $C > 0$. Обозначим $F_{s,r}$ *класс Бретаньоля — Юбер* с параметрами s и r (s — положительное целое число и $r > 0$), т. е. класс всех плотностей на $[0, 1]$, для которых $D_s^*(f) \leqslant r$, где величина $D_s^*(f)$ определяется следующим образом:

$$D_s^*(f) = \left(\left(\int \sqrt{f}\right)^{2s} \sup_{a>0} \int |(f*\varphi_a)^{(s)}|\right)^{1/(2s+1)}.$$

Здесь φ — четная ограниченная плотность с компактным носителем, монотонная на $[0, \infty)$ и имеющая s абсолютно непрерывных производных. Заметим, что данное определение несколько отличается от определений аналогичных величин в гл. 4 и § 5.1—5.8.

В теореме 4.6 было показано, что при достаточно больших C для любых оценок плотности f_n выполнено соотношение

$$\sup_{f \in W(s,\alpha,C)} E\left(\int |f_n - f|\right) \geqslant$$

$$\geqslant (c_3 + o(1))\, C^{1/(2(s+\alpha)+1)} n^{-(s+\alpha)/(2(s+\alpha)+1)}$$

(в обозначениях теоремы 4.6). Напомним, что $c_3 = c_3(s, \alpha) > 0$. Аналогичная минимаксная нижняя граница для $F_{s,r}$, равная

$c_4(r)\, n^{-s/(2s+1)}$, получена в теореме 4.3 при достаточно больших r.

Наша цель сейчас — показать, что если несколько обобщить определение ядерной оценки, то, с точностью до коэффициента пропорциональности, *не зависящего от* n, C и r, на этой оценке будут достигаться указанные границы. Отметим, что этот факт очень важен. На других оценках можно достичь в лучшем случае лишь уменьшения коэффициента пропорциональности в границе сверху для минимаксного L_1-риска. Поэтому мы будем серьезно рассматривать другие оценки только для специальных классов плотностей (таких, как множество монотонных плотностей на $[0, 1]$ и т. д.).

Для классов Липшица рассмотрим только случай, когда $\alpha = 1$, поскольку случай $0 < \alpha < 1$, конечно, менее интересен. Однако исследование этого случая после знакомства с общим анализом при $\alpha = 1$ является хорошим упражнением. Отметим еще, что будет избран краткий способ доказательства: верхние границы будут получены очень просто, путем весьма грубого раздельного рассмотрения членов, соответствующих смещению и дисперсии. Это приведет лишь к незначительному ухудшению коэффициентов пропорциональности. При изложении будем частично следовать работе Bretagnolle, Huber (1979).

Лемма 22 (равномерные границы для смещения). *Пусть* K — *измеримая симметричная функция, удовлетворяющая условиям*

$$\int K = 1, \quad \int x^i K = 0, \quad i = 1, \ldots, s-1,$$

$$\int |x^s|\,|K| < \infty,$$

и пусть L — *ядро, ассоциированное с* K, *т. е.*

$$L(x) = (-1)^s \int_x^\infty \frac{(y-x)^{s-1}}{(s-1)!} K(y)\, dy, \quad x > 0,$$

$$L(-x) = (-1)^s L(x), \quad x < 0,$$

Тогда $\int |L| < \infty$. *Если* $s = 1$, $K \geqslant 0$, *то* $\int |L| = \int |x|\, K$, *и если* $s = 2$, $K \geqslant 0$, *то* $\int |L| = \int x^2 K/2$. *Если функция* f *имеет* $(s-1)$ *абсолютно непрерывных производных, то*

$$\int |f * K_h - f| \leqslant h^s \int |L| \int |f^{(s)}|.$$

При любых f *имеем*

$$\int |f * K_h - f| \leqslant h^s \int |L| \liminf_{a \downarrow 0} \int |(f * \varphi_a)^{(s)}|.$$

При $f \in W(s-1, 1, C)$, $s \geqslant 1$, *последняя верхняя граница не превосходит*

$$Ch^s \int |L|.$$

Доказательство.

$$\int |L| = 2\int_0^\infty \left| \int_x^\infty \frac{(y-x)^{s-1}}{(s-1)!} K(y)\, dy \right| dx \leqslant$$

$$\leqslant 2\int_0^\infty \left(\int_0^y \frac{(y-x)^{s-1}}{(s-1)!}\, dx \right) |K(y)|\, dy = 2\int_0^\infty \frac{|y|^s}{s!} |K(y)|\, dy < \infty.$$

При $s = 1$, $s = 2$ и $K \geqslant 0$ здесь везде выполнены равенства, что позволяет получить в явном виде простые выражения для $\int |L|$.

Если функция f имеет $(s-1)$ абсолютно непрерывных производных, то, используя разложение в ряд Тейлора, получим

$$f(x+y) - f(x) = \sum_{j=1}^{s-1} \frac{y^j}{j!} f^{(j)}(x) + \int_x^{x+y} \frac{(x+y-u)^{s-1}}{(s-1)!} f^{(s)}(u)\, du.$$

Пусть через L_h обозначается $(1/h)\, L(x/h)$ и функция $(L)_h$ определяется так же, как L, с заменой K на K_h. Тогда

$$f * K_h - f = f^{(s)} * (L)_h = h^s f^{(s)} * L_h.$$

Таким образом,

$$\int |f * K_h - f| \leqslant h^s \int |L_h| \int |f^{(s)}| = h^s \int |L| \int |f^{(s)}|.$$

Известно, что при фиксированном $h > 0$ и любых f имеем

$$|f * K_h - f| = \liminf_{a \downarrow 0} |(f * K_h - f) * \varphi_a| =$$

$$= \liminf_{a \downarrow 0} |f * \varphi_a * K_h - f * \varphi_a|$$

для почти всех x. Следовательно, при том же h получим

$$\int |f * K_h - f| \leqslant \liminf_{a \downarrow 0} \int |f * \varphi_a * K_h - f * \varphi_a| \leqslant$$

$$\leqslant \liminf_{a \downarrow 0} h^s \int |(f * \varphi_a)^{(s)}| \int |L|,$$

если воспользоваться леммой Фату. При любых f из класса $W(s-1, 1, C)$ интеграл в последнем выражении ограничен сверху величиной C в силу неравенства $\int |f^{(s-1)} * \varphi_a'| \leqslant C$. Для доказательства этого неравенства заметим, что функция

$f^{(s-1)}$ является C-липшицевой. Но для любой C-липшицевой функции g на R имеем

$$\left|\int g(y)\,\varphi_a'(x-y)\,dy\right| = \left|\int (g(y)-g(x))\,\varphi_a'(x-y)\,dy\right| \leqslant$$
$$\leqslant \int |g(y)-g(x)|\,|\varphi_a'(x-y)|\,dy \leqslant$$
$$\leqslant C\int |x-y|\,|\varphi_a'(x-y)|\,dy =$$
$$= C\int |z|\,|\varphi'(z)|\,dz =$$
$$= -2\int_0^\infty Cz\varphi'(z)\,dz = 2C\int_0^\infty \varphi(z)\,dz = C.$$

Так как $\liminf\limits_{a\downarrow 0} \int\limits_{[0,1]^c} |g * \varphi_a'| = 0$ для любых C-липшицевых функций g, сосредоточенных на $[0,1]$, то лемма 22 доказана. (Чтобы убедиться в этом, заметим, что функция $g * \varphi_a'$ ограничена по абсолютной величине постоянной C и равна нулю вне $[-aM, 1+aM]$, где M — число, зависящее от носителя φ.)

Лемма 23 (равномерные границы для вариации). *Пусть K — ядро на R, равное нулю вне отрезка $[-1,1]$ и удовлетворяющее условиям леммы 22. Положим $C_s = \liminf\limits_{a\downarrow 0} \int |(f * \varphi_a)^{(s)}|$ и $\alpha = \sqrt{\int K^2} < \infty$. Тогда, если плотность f сосредоточена на $[0,1]$,*

$$E\left(\int |f_n - f * K_h|\right) \leqslant (nh)^{-1/2}\left(\alpha\int\sqrt{f} + \sqrt{C_1(h+2h^2)\int |x|\,K^2/2}\right).$$

Доказательство. Применяя к $E(|f_n - f * K_h|)$ неравенство Коши — Шварца, получим

$$E\left(\int |f_n - f * K_h|\right) \leqslant (nh)^{-1/2}\int\sqrt{f * K_h^2},$$

где $K_h^2(x) = (1/h)\,K^2(x/h)$. Если обозначить $K^+ = K^2 \Big/ \int K^2$, то эту верхнюю границу можно переписать в виде

$$(nh)^{-1/2}\,\alpha\int\sqrt{f * K_h^+} \leqslant (nh)^{-1/2}\,\alpha\left(\int\sqrt{f} + \int\sqrt{|f - f * K_h^+|}\right) \leqslant$$
$$\leqslant (nh)^{-1/2}\,\alpha\left(\int\sqrt{f} + \sqrt{\int_{-h}^{1+h} dx\int |f - f * K_h^+|}\right)$$

(по неравенству Коши)

$$\leqslant (nh)^{-1/2}\alpha\left(\int\sqrt{f}+\sqrt{(1+2h)\,hC_1\int|x|\,K^+/2}\right)$$

(по лемме 22),

что и требовалось доказать.

Теорема 11 (минимаксные верхние границы). *Пусть функции K и L — такие же, как в леммах 22 и 23, и пусть α и C_s — величины, определенные в лемме 23. Тогда для любых плотностей f, сосредоточенных на $[0, 1]$, имеем*

$$E\left(\int|f_n-f|\right)\leqslant h^sC_s\int|L|+$$

$$+(nh)^{-1/2}\left(\alpha\int\sqrt{f}+\sqrt{C_1(h+2h^2)\int|x|\,K^2/2}\right),$$

где f_n — ядерная оценка с ядром K. В частности, для $W(s-1, 1, C)$ при любых $s\geqslant 1$ имеет место следующая минимаксная верхняя граница:

$$\inf_{g_n}\sup_{f\in W(s-1,\,1,\,C)}E\left(\int|g_n-f|\right)\leqslant$$

$$\leqslant\inf_{h>0,\ K}\sup_{f\in W(s-1,\,1,\,C)}E\left(\int|f_n-f|\right)\leqslant$$

$$\leqslant\inf_K\frac{2s+1}{2s}\left(2sC\int|L|\,\alpha^{2s}n^{-s}\right)^{1/(2s+1)}(1+o(1)),$$

где через g_n обозначается произвольная оценка плотности. Нижние грани здесь берутся по всем K, удовлетворяющим условиям лемм 22 и 23. Последнее неравенство можно получить, например, если выбрать

$$h=\left(\frac{\alpha}{2sC\int|L|}\right)^{2/(2s+1)}n^{-1/(2s+1)}.$$

В частности, при $s=1$ верхняя граница равна произведению $(1+o(1))$ на

$$\inf_{K>0}\frac{3}{2}\left(2C\int|x|\,K\alpha^2n^{-1}\right)^{1/3}=\left(\frac{3C}{2n}\right)^{1/3}.$$

Последняя величина получается при выборе треугольного ядра K (для которого $\alpha^2=2/3$, $\int|x|\,K=1/3$). В случае $s=2$ нижняя грань верхней границы, равная

$$(1+o(1))\ (225C/512)^{1/5}n^{-2/5},$$

достигается на ядре Епанечникова $3(1-x^2)_+/4$ (для которого $\int x^2K=1/5$, и $\int K^2=3/5$).

Доказательство. Первое неравенство следует из лемм 22 и 23. Для доказательства второй цепочки неравенств требуются три факта. Во-первых, для любых плотностей f, сосредоточенных на $[0, 1]$, имеем $C_1 \leqslant C_2 \leqslant C_3 \leqslant \ldots \leqslant C_s$. Кроме того, $C_s \leqslant C$ для любой плотности $f \in W$ $(s - 1, 1, C)$ (см. доказательство леммы 22). Во-вторых, если плотность f равна нулю вне $[0, 1]$, то $\int \sqrt{f} \leqslant 1$. В-третьих, функция $h^s C \int |L| + \alpha/\sqrt{nh}$ достигает минимума при значении h, указанном в теореме. Формальной подстановкой этого значения h получим требуемую верхнюю границу. Оставшаяся часть доказательства очевидна.

Минимаксная нижняя граница теоремы 4.6 для W $(s - 1, 1, C)$ и минимаксная верхняя граница теоремы 11 зависят от n и C одинаковым образом. Они отличаются друг от друга только коэффициентами пропорциональности, которые в свою очередь зависят только от s. Полезно понять, каково расхождение между этими границами для наиболее важных классов W $(0, 1, C)$ и W $(1, 1, C)$. Напомним, что для W $(0, 1, C)$ коэффициент при $(C/n)^{1/3}$ в теореме 4.7 равен $(21/160)\ (12/25)^{1/3}$, так что отношение верхней границы к нижней приблизительно равно 11. Для W $(1, 1, C)$ это отношение порядка 30. Вне всякого сомнения, верхние границы не являются точными: они получены для достаточно примитивной оценки — ядерной с зависящим только от s, C и K параметром сглаживания h! Гораздо более эффективным представляется, конечно, выбор параметра h, зависящего от f (см. гл. 6, где рассматривается адаптивный выбор h).

Однако можно встать на пессимистическую позицию, выбирая оценку так, чтобы минимизировать максимальный риск для определенного класса, например, для W $(s - 1, 1, C)$. (Это имеет смысл, если требуется получить грубое начальное приближение или если никакие другие методы не приводят к успеху.) В таком случае следует выбирать h, как указано в теореме 11.

Кроме h требуется выбрать еще ядро K. Мы знаем, что в случаях $s = 1$ и $s = 2$ следует выбирать равнобедренную треугольную плотность K на $[-1, 1]$ и ядро Епанечникова. При $s \geqslant 3$ обязательно нужно, чтобы ядро K принимало отрицательные значения, и в связи с этим оценка f_n уже не может быть плотностью, несмотря на то что интеграл от нее по-прежнему равен единице. Впрочем, как показано в § 7.6, такие оценки можно легко нормализовать. Бартлетт (Bartlett (1963)) первым указал, что при выборе таких ядер, как в леммах 22 и 23, можно получить улучшение скоростей сходимости (см. также подробный анализ в § 7.6). Кроме того, он нашел оптимальный вид K при $s = 4$. Не представляет большого труда построить ядра K, удовлетворяющие условиям лемм 22 и 23 в общем случае. Например, рассмотрим

сначала некоторую исходную симметричную плотность K, сосредоточенную на $[-1, 1]$. При фиксированном s достаточно только найти действительные числа p_i (не обязательно положительные), такие, что функция

$$\sum_{i=1}^{N} p_i K_{1/i}$$

удовлетворяет требуемым условиям. (Кстати, выбор значений $1/i$ сделан по произволу, и их можно заменить на другие положительные числа из $(0, 1)$.) Если K, скажем, равномерная плотность на $[-1, 1]$, то получаются следующие условия:

$$\sum p_i = 1, \quad \sum p_i i^{-2} = 0, \quad \sum p_i i^{-4} = 0, \ldots, \quad \sum p_i i^{-(s-2)} = 0$$

при четных s, $s \geqslant 4$. Как правило, существует решение с $N = s/2$ компонентами смеси. См. также работу Bretagnolle, Huber (1979), где даны другие методы построения, основанные на полиномах Лежандра.

Из первого неравенства теоремы 11 видно, что верхняя граница для средней L_1-ошибки при фиксированной плотности f может быть гораздо меньше, чем минимаксная верхняя граница. В частности, величина $\int \sqrt{f}$ может быть значительно меньше единицы, несмотря на то что имеет место следующий факт.

Лемма 24. *Для любых f из класса $W(0, 1, C)$, $C \geqslant 4$, справедливы неравенства*

$$\frac{4}{3\sqrt{2}} \geqslant \int \sqrt{f} \geqslant \frac{4}{3C^{1/4}}.$$

Оба неравенства достигаются.

Доказательство. Верхняя граница достигается на равнобедренной треугольной плотности на $[0, 1]$, которая равна $4x$ на $[0, 1/2]$. (Эта плотность в определенном смысле является «самой гладкой» в классе $W(0, 1, C)$.) Нижняя граница достигается на равнобедренной треугольной плотности на $[1/2 - b, 1/2 + b]$, где $b = 1/\sqrt{C}$ (тангенс угла наклона сторон треугольника, конечно, равен C).

Таким образом, при больших C минимаксная верхняя граница теоремы 11 оказывается менее мощной — она не дает правильной информации о большинстве плотностей из данного класса Липшица.

Подход, принятый выше для классов $W(s, 1, C)$, нельзя использовать при получении равномерной верхней границы для

классов Бретаньоля — Юбер $F_{s,r}$, ибо для них значение h, правильным образом зависящее от r, оказывается не только функцией от s, r и n, но и от C_s и $\int \sqrt{f}$. Но так как величины C_s и $\int \sqrt{f}$ не известны, то их требуется оценить. Поэтому, строго говоря, следует рассматривать только ядерные оценки, у которых параметр h зависит от выборочных данных и асимптотически стремится к оптимальному h. Эта адаптивная стратегия использовалась Бретаньолем и Юбер (Bretagnolle, Huber (1979)) при поиске минимаксной верхней границы для $F_{s,r}$. Отметим также, что на классе $F_{s,r}$ величина $\int \sqrt{f}$ не ограничена равномерно снизу, а C_s не ограничена равномерно сверху. (Это следует из того, что для одной и той же плотности можно использовать различные масштабы!) Доказательство верхней границы для $F_{s,r}$ здесь не приводится, поскольку это довольно нудная процедура, состоящая в следующем. Сначала разобьем данные на несколько частей и воспользуемся одной из малых частей (объема $o(n)$) для оценивания h и большой частью (объема $\sim n$) для построения ядерной оценки с таким h. Затем получим верхнюю границу для $\sup_{f \in F_{s,r}} E\left(\int |f_n - f|\right)$, равную супремуму по $F_{s,r}$ математического ожидания правой части первого неравенства теоремы 11. После того убедимся, что эта граница не превосходит произведение величины $rn^{-s/(2s+1)}$ на постоянную, не зависящую от r и n.

Равномерное неравенство теоремы 11 полезно не только для получения минимаксных верхних границ, но и во многих других отношениях. Так, его можно применять при *любых* n, и потому оно представляет большую ценность для тех, кто работает с малой выборкой. Но более важно то, что с помощью этого неравенства можно находить скорость сходимости ядерной оценки со случайным параметром сглаживания h, не зависящим от $X_1, \ldots, X_n$. Типичен случай, когда величина h является функцией от $X_{n+1}, \ldots, X_{n+m}$. Имеет место

Теорема 12. *При условиях и обозначениях теоремы* 11 *для ядерной оценки* f_n *с параметром сглаживания* h, *не зависящим от* $X_1, \ldots, X_n$, *остается в силе граница теоремы* 11, *если только в правой ее части взять математическое ожидание по* h. *Пусть, в частности, существует последовательность положительных чисел* h_{n0}, *такая, что* $h_{n0} \to 0$, $nh_{n0} \to \infty$, *и*

$$E(h^s) \sim h_{n0}^s, \quad E(h^{-1/2}) \sim h_{n0}^{-1/2}, \quad E(\sqrt{h}) \to 0,$$

и пусть $\int |x| K^2 < \infty$. *Тогда*

$$E\left(\int |f_n - f|\right) \leqslant \left(h_{n0}^s C_s \int |L| + (nh_{n0})^{-1/2} \alpha \int \sqrt{f}\right)(1 + o(1)).$$

Если $C_s < \infty$, $\int \sqrt{f} < \infty$ *и*

$$h_{n0} = \left(\frac{\alpha \int \sqrt{f}}{2sC_s \int |L|} \right)^{2/(2s+1)} n^{-1/(2s+1)},$$

то

$$E\left(\int |f_n - f|\right) \leqslant$$
$$\leqslant \frac{2s+1}{2s} \left(2s \int |L| \alpha^{2s} n^{-s}\right)^{1/(2s+1)} D_s^*(f)\,(1 + o(1)).$$

В случае когда параметр сглаживания h зависит от данных, требуется более сильная теорема. Состоятельность оценок в этом случае изучается в гл. 6.

10. Плотности с неограниченным носителем

До сих пор мы откладывали рассмотрение задачи о поведении оценок для плотностей f с неограниченным носителем. Тому имеются две причины: меньшая важность этого случая (данные всегда можно монотонно отобразить на $[0, 1]$; плотности с неограниченным носителем реже встречаются на практике) и нежелание перегружать текст дополнительными обозначениями и понятиями, которые только отвлекали бы от основных идей.

В теоремах 2 и 5 получены нижние границы при любых f. Поэтому достаточно вывести лишь верхние границы для $E(J_n)$. Рассмотрим здесь только ядерную оценку, поскольку случай гистограммной оценки аналогичен. Если исходить из равномерной верхней границы теоремы 11, то можно сделать доказательства очень короткими. Однако следует отдавать себе отчет, что при этом мы несколько поступимся точностью по сравнению с границами теорем 1 и 6, полученными с помощью неравенств типа Берри — Эссеена для отклонений от нормального закона. Из леммы 23 и теоремы 11 получим следующий результат.

Лемма 25. *Пусть* f_n *— ядерная оценка на* R *с ядром* K *и функцией* L, *такими, как в леммах* 22 *и* 23, *и пусть* $s \geqslant 1$ *— целое число. Обозначим*

$$C_s = \sup_{a>0} \int |(f * \varphi_a)^{(s)}|,$$

где φ *— четная ограниченная плотность с компактным носителем, монотонная на* $[0, \infty)$ *и имеющая* s *абсолютно непрерывных производных. Тогда*

$$E\left(\int |f_n - f|\right) \leqslant h^s C_s \int |L| + (nh)^{-1/2} \int \sqrt{f * (K^2)_h}.$$

Равномерная граница леммы 25 указывает на важную роль вездесущего множителя $\int \sqrt{f * (K^2)_h}$, которому мы посвящаем отдельную лемму.

Лемма 26 (множитель $\int \sqrt{f * (K^2)_h}$. *Пусть K — произвольная измеримая функция на R, дополнительные условия на которую указываются в различных утверждениях данной леммы, и пусть f — плотность на R.*

A. *Существует плотность f на R, такая, что* $\int \sqrt{f} < \infty$, *но* $\int \sqrt{f * (K^2)_h} = \infty$ *при любых достаточно малых h и любых ядрах K, сосредоточенных на $[-1, 1]$, ограниченных по абсолютной величине некоторой постоянной и удовлетворяющих условию* $\int K = 1$.

B. $\int \sqrt{f * (K^2)_h} \geqslant \int \sqrt{f} \sqrt{\int K^2}$. *В частности,* $\int \sqrt{f * (K^2)_h} = \infty$ *при* $\int \sqrt{f} = \infty$.

C. $\int \sqrt{f * (K^2)_h} \to \int \sqrt{f} \sqrt{\int K^2}$ *при* $h \downarrow 0$, *если ядро K ограниченно и имеет компактный носитель, а функция f удовлетворяет следующим условиям:*

(i) *Существуют положительные числа t, T, такие, что*

$$\int\limits_{|x|>T} \sqrt{\sup_{|y-x|\leqslant t} f(y)}\, dx < \infty.$$

(ii) *Функция f непрерывна почти всюду.*

D. *Если функции f и K симметричны и унимодальны, то*

$$\int \sqrt{f * K^2} \leqslant 2 \int \sqrt{f} \sqrt{\int K^2} + \sqrt{2 \int |K|}.$$

Доказательство. A. Пусть f — индикаторная функция множества $\bigcup\limits_{i=1}^{\infty} [2^i, 2^i + 1/i\,(i+1)]$, $i \geqslant 1$. Ясно, что $\int \sqrt{f} = 1$. Кроме того, при $h \leqslant 1/2$ имеем

$$\int \sqrt{f * (K^2)_h} \geqslant \sum_{i=1}^{\infty} \int\limits_{2^i - h}^{2^i + (i\,(i+1))^{-1} + h} \sqrt{f * (K^2)_h} \geqslant$$

$$\geqslant \frac{\sum_{i=1}^{\infty} \int_{2^i-h}^{2^i+(i\,(i+1))^{-1}+h} f*(K^2)_h}{\sup_{-h\leqslant x-2^i\leqslant h+(i\,(i+1))^{-1}} \sqrt{f*(K^2)_h}} \geqslant$$

$$\geqslant \sum_{i=1}^{\infty} \frac{\sqrt{hi\,(i+1)}}{K^*} \cdot \frac{1}{i\,(i+1)} \int K^2 = \infty,$$

где K^* — равномерная граница для $|K|$.

B. По неравенству Йенсена $\sqrt{f*(K^2)_h} \geqslant \sqrt{f}*(K^2)_h \Big/ \sqrt{\int K^2}$.

C. Пусть t и T — числа из условия (i), и пусть K равно нулю вне $[-1, 1]$. Имеем

$$\sqrt{f*(K^2)_h\,(x)} \leqslant \sqrt{\sup_{|y-x|\leqslant h} f\,(y) \int K^2}.$$

Правая часть этого неравенства меньше, чем некоторая фиксированная интегрируемая функция, при $h \leqslant t$ (в силу (i)). По теореме 2.3 $f * (K^2)_h \to f \int K^2$ для почти всех x, так что теорема Лебега о мажорируемой сходимости дает

$$\int_{[-T,\,T]^c} \sqrt{f*(K^2)_h} \to \int_{[-T,\,T]^c} \sqrt{f}\,\sqrt{\int K^2}.$$

Кроме того, если положить $K^+ = K^2 \Big/ \int K^2$, то

$$\int_{-T}^{T} \left|\sqrt{f*(K^2)_h} - \sqrt{f}\,\sqrt{\int K^2}\right| \leqslant \sqrt{\int K^2} \int_{-T}^{T} |\sqrt{f*K_h^+} - \sqrt{f}\,| \leqslant$$

$$\leqslant \sqrt{\int K^2} \int_{-T}^{T} \sqrt{|f*K_h^+ - f\,|} \leqslant$$

$$\leqslant \sqrt{\int K^2}\,\sqrt{2T \int |f*K_h^+ - f\,|} = o\,(1),$$

где мы воспользовались теоремой 2.1.

D. Предположим, что $x \geqslant 0$. Тогда

$$\int f\,(y)\,K^2\,(x-y)\,dy \leqslant \int_{y<x/2} f\,(y)\,K^2\,(x/2)\,dy +$$

$$+ \int_{x/2\leqslant y} f\,(x/2)\,K^2\,(x-y)\,dy \leqslant$$

$$\leqslant K^2\,(x/2) + f\,(x/2) \int K^2.$$

Квадратный корень от правой части не превосходит величины $|K(x/2)| + \sqrt{f(x/2)}\sqrt{\int K^2}$. Интегрирование по x приводит к указанному в лемме неравенству.

Теорема 13. *Пусть $s \geqslant 1$ — целое число и K — симметричная ограниченная функция с компактным носителем, удовлетворяющая условиям леммы 22. Пусть L — функция из леммы 22 и C_s — постоянная из леммы 25. При плотностях f, удовлетворяющих условию* C *леммы 26, для ядерной оценки f_n выполнено неравенство*

$$E\left(\int |f_n - f|\right) \leqslant C_s \int |L|\, h^s + (nh)^{-1/2} \int \sqrt{f}\, \sqrt{\int K^2}\,(1 + o(1))$$

при $h \downarrow 0$. В частности, если $C_s < \infty$ и $\int \sqrt{f} < \infty$ (т. е. $D_s^*(f) < \infty$) и

$$h = \left(\frac{\sqrt{\int K^2} \int \sqrt{f}}{2sC_s \int |L|}\right)^{2/(2s+1)} n^{-1/(2s+1)},$$

то

$$E\left(\int |f_n - f|\right) \leqslant \frac{2s+1}{2s}\left(2s \int |L| \left(\sqrt{\int K^2}\right)^{2s} n^{-s}\right)^{1/(2s+1)} \times$$
$$\times D_s^*(f)\,(1 + o(1)).$$

Доказательство. Теорема 13 без труда следует из теоремы 11 и утверждения C леммы 26.

Таким образом, даже для плотностей f с неограниченным носителем величина $D_s^*(f)$, по-видимому, может рассматриваться как мера сложности. Наиболее важными снова являются случаи $s = 1$ и $s = 2$. В этих случаях ядро K, при котором граница минимальна, не зависит от f и снова представляет собой равнобедренную треугольную плотность при $s = 1$ и ядро Епанечникова при $s = 2$. При таком выборе ядер K верхние границы принимают вид

$$(3/2n)^{1/3} D_1^*(f) = (3/n)^{1/3} B_H^*(f) \quad (s = 1)$$

и

$$(225/512)^{1/5} n^{-2/5} D_2^*(f) = (225/256)^{1/5} n^{-2/5} B^*(f) \quad (s = 2).$$

Эти границы превышают соответствующие верхние границы теорем 10 и 1 только в знаках после запятой. Оптимальные значения h тоже очень мало отличаются от тех, которые были получены для плотностей с компактными носителями в теоремах 10 и 1.

Завершим этот параграф замечанием о том, что условие C леммы 26 выполнено для любых ограниченных унимодальных и непре-

рывных почти всюду плотностей f, таких, что $\int \sqrt{f} < \infty$. Действительно, если m — мода, то при $x > m$ имеем

$$\sup_{|y-x| \leqslant t} f(y) \leqslant f(\max(x-t,\ m)),$$

и квадратный корень от последнего выражения является интегрируемой функцией при любом t.

11. Несмещенность и достижимость скорости убывания ошибки $1/\sqrt{n}$

Скорость убывания средней ошибки ядерной оценки в L_1 равна $n^{-s/(2s+1)}$ при любых f из классов $F_{s,r}$ или $W(s-1, 1, C)$ (см. теорему 11) и при соответствующем выборе K и h. Если увеличивать s, то эти классы сужаются, а скорость приближается к $n^{-1/2}$. Вспомним еще, что наилучшая минимаксная скорость сходимости любой оценки плотности на таких простых однопараметрических классах, как $Q_1(g)$ (теорема 4.8) или $\Pi(g)$ (теорема 4.4), равна $1/\sqrt{n}$. В промежуточном случае имеется пробел, так как неизвестно, существует ли богатый класс плотностей $\mathcal{F}$, такой, что $\limsup_{n\to\infty} \sqrt{n} E\left(\int |f_n - f|\right) < \infty$ для некоторой оценки f_n при любых $f \in \mathcal{F}$. Под «богатым» подразумевается класс, который нельзя описать конечным числом параметров. Следует, однако, иметь в виду, что в силу нижних границ из гл. 4 класс $\mathcal{F}$ не может быть слишком большим.

В случае L_2 ответ на рассматриваемый вопрос — утвердительный. Именно Ибрагимов, Хасьминский (1982) получили следующий удивительный результат.

Теорема 14 (Ибрагимов, Хасьминский, 1982). *Пусть A_T — класс всех плотностей, характеристическая функция которых равна нулю вне $[-T, T]$. Тогда*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{f \in A_T} nE\left(\int (f_n - f)^2\right) = \frac{T}{\pi},$$

где через f_n обозначается произвольная оценка плотности.

Здесь надо сделать несколько замечаний. Во-первых, класс A_T состоит из исключительно гладких плотностей, ибо условия на хвосты характеристической функции соответствуют условиям гладкости плотности. Так как характеристическая функция φ абсолютно интегрируема, то можно выразить f через φ по формуле обратного преобразования

$$f(x) = (2\pi)^{-1} \int e^{-itx} \varphi(t)\, dt$$

(см. Lukacs (1970)). На самом деле можно вычислить и $f^{(s)}$, дифференцируя s раз под знаком данного интеграла. Используя это, получим

$$\int |f^{(s)}| \leqslant \frac{1}{\pi} \frac{T^{s+1}}{s+1}$$

при любых натуральных s.

К сожалению, не существует неравенств, непосредственно связывающих ошибки в L_1 и в L_2, и потому из теоремы 14 не следует, что

$$\limsup_{n\to\infty} \sqrt{n} \sup_{f \in A_T} E\left(\int |f_n - f|\right) < \infty$$

для некоторой оценки f_n. Имеется несколько косвенных неравенств, но они ничего не дают. Примеры таких неравенств приведены в теореме 8.3.

Тот факт, что скорость $n^{-1/2}$ достигается на A_T, не удивителен, поскольку этот класс является «почти» параметрическим. В самом деле, известно, что по теореме Найквиста [1]) функция f может быть полностью восстановлена (Feller (1971)) по своим значениям в точках iy, $i = 0, +1, -1, +2, -2, \ldots$, где y — достаточно малая положительная постоянная. Следовательно,можно рассматривать A_T как класс со счетным числом параметров.

Ибрагимов и Хасьминский, более того, показали, что предел в теореме 14 достигается на *оценке типа интеграла Фурье* (ОИФ), описанной Davis (1975, 1977) и Конаковым (1972). Достижимость скорости сходимости ошибки $1/n$ в L_2 для фиксированных плотностей $f \in A_T$ была также замечена Davis (1975, 1977), который опирался на результаты, полученные в L_2 Ватсоном и Лидбеттером (Watson, Leadbetter (1963)). ИОФ есть ядерная оценка с ядром

$$K(x) = \frac{\sin x}{\pi x}.$$

Заметим, что $\int K = 1$, но $\int |K| = \infty$. Кроме того, характеристическая функция ядра K равна $I_{[-1,\,1]}(t)$. Интеграл от оценки f_n с таким ядром в точности равен 1, но она не является абсолютно интегрируемой с вероятностью 1. Таким образом, нельзя «нормировать» f_n, полагая

$$f_n^*(x) = \frac{f_n(x)\, I_{[f_n(x) > 0]}}{\int\limits_{f_n > 0} f_n}.$$

поскольку значения f_n^* были бы равны нулю с вероятностью 1.

[1]) В отечественной литературе она называется теоремой Котельникова. — *Прим. перев.*

Если бы оценка f_n была абсолютно интегрируемой, то в результате такой нормировки получилась бы настоящая плотность f_n^* и, более того, имело бы место неравенство

$$\int |f_n^* - f| \leqslant \int |f_n - f|.$$

(Оно следует из теоремы 11.4 о неотрицательной проекции). Так как нам не известны какие-либо другие нормировки, при которых не увеличивается ошибка в L_1, то, придерживаясь общих принципов, установленных в начале этой книги, мы не склонны рекомендовать ОИФ в качестве оценки плотности. Дэвис, Ибрагимов и Хасьминский смогли получить в L_2 скорость сходимости $1/n$ по той причине, что у них при каждом x в оценивании участвовала вся выборка. Для ядерной оценки это означает, что h не стремится к 0. Но отсюда в свою очередь следует, что $\int |f - \hat{f} * * K_h| = 0$ при некотором положительном h, т. е. оценка является несмещенной! Если попытаться следовать этим рассуждениям, то ключом к решению будет существование функции K, такой, что $\int K = 1$, $\int |K| < \infty$ (чтобы ее можно было нормировать) и $\int |f - f * K_h| = 0$ при некотором $h > 0$ и любых $f \in A_T$. Такая функция K, в самом деле, существует, и, следовательно, можно надеяться, что на A_T скорость сходимости средней ошибки в L_1 получится равной $1/n$.

Рассмотрим сначала *плотность Валле-Пуссена*

$$K(x) = (2\pi)^{-1}\left(\frac{\sin(x/2)}{x/2}\right)^2,$$

характеристическая функция которой равна $(1 - |t|)_+$. Далее, при некотором постоянном $a > 0$ (которое будет выбрано ниже) определим функцию $g_a(x) = (a+1)K(x) - \frac{a^2}{a+1}K\left(\frac{ax}{a+1}\right)$. Ясно, что $\int g_a = 1$ и $\int |g_a| \leqslant 2a + 1$. Кроме того, характеристическая функция для g_a равна

$$\psi_a(t) = (a+1)(1 - |t|)_+ - a\left(1 - \left|\frac{a+1}{a}t\right|\right)_+,$$

т. е.

$$\psi_a(t) = \begin{cases} 1, & |t| \leqslant a/(1+a), \\ (a+1)(1-|t|), & a/(1+a) \leqslant |t| \leqslant 1, \\ 0, & |t| \geqslant 1. \end{cases}$$

Установим теперь несмещенность рассматриваемой ядерной оценки на A_T.

Теорема 15. *Пусть A_T — класс всех плотностей, характеристическая функция которых равна нулю вне $[-T, T]$. Ядерная*

оценка с ядром K и параметрам сглаживания h является несмещенной (т. е. $\int |f - f * K_h| = 0$) в каждом из следующих случаев:

1. $K(x) = \sin x/\pi x$, $h \leqslant 1/T$;
2. $K(x) = g_a(x)$ *при фиксированном* $a > 0$, $h \leqslant (a/(1 + a))\,(1/T)$.

Доказательство. Пусть φ — характеристическая функция для f. Известно, что характеристическая функция для $f * K_h$ равна $\varphi(t)\,\psi(th)$, где ψ — характеристическая функция для K (если K не является плотностью, то ψ определяется как $\int e^{itx} K(x)\,dx$). Для справедливости равенства $f = f * K_h$ при почти всех x достаточно, что $\varphi(t) = \varphi(t)\,\psi(th)$ при любых t. Так как $f \in A_T$, то нужно проверить лишь, что $\psi(th) = 1$ при любых $|t| \leqslant T$. Для первого ядра имеем $\psi(t) = I_{[-1,\,1]}(t)$, и достаточно только потребовать, чтобы было $h \leqslant 1/T$ (это рассуждение принадлежит Davis (1975, 1977)). Для второго ядра нужно потребовать, чтобы выполнялось неравенство $Th \leqslant a/(1 + a)$. Теорема 15 доказана.

Прежде чем приступить к исследованию свойств ядерной оценки с ядром g_a при $f \in A_T$, разберемся в том, чего не следует ожидать и чего нельзя сделать.

Теорема 16. *Пусть f_n — ядерная оценка с ядром K, таким, что $\int K = 1$, $\int |K| < \infty$ и $\int \sup_{|u| > |x|} |K(u)| < \infty$. Тогда*

1) *если $h \to 0$, то $\sqrt{n} E\left(\int |f_n - f|\right) \to \infty$ при любых f;*

2) *если $h \to 0$, $nh \to \infty$ и $\int \sqrt{f} = \infty$, то $\sqrt{nh} E\left(\int |f_n - f|\right) \to \infty$;*

3) *если характеристическая функция φ плотности f отлична от нуля всюду, кроме, быть может, множества лебеговой меры нуль, то $\sqrt{n} \inf_{h>0} E\left(\int |f_n - f|\right) \to \infty$;*

4) *если K является плотностью (но, возможно не имеет интегрируемой радиальной мажоранты), то*

$$\inf_f \liminf_{n\to\infty} n^{2/5} \inf_{h>0} E\left(\int |f_n - f|\right) > 0.$$

В теореме 16 утверждается, что для построения ядерной оценки, являющейся «нормируемой» и состоятельной при *любых* f, необходимо устремить h к 0, и если это сделать, то *скорость $1/\sqrt{n}$ не достигается ни при каких f*! По существу из утверждений 1 и 3 теоремы 16 следует, что нет смысла искать скорость сходимости ошибки $1/n$ вне класса A_T. Наконец, в утверждении 4, идентичном

теореме 2, дается еще более сильная нижняя граница при фиксированной плотности f для случая ядер с неотрицательными значениями. Доказательство теоремы 16 основано на следующих нижних границах.

Лемма 27. *Пусть $Z_1, \ldots, Z_n$ — независимые одинаково распределенные случайные величины с нулевым средним, такие, что $E(|Z_1|^p) < \infty$ при некотором $p \geqslant 2$. Тогда существуют положительные постоянные B_p и C_p, зависящие только от p, такие, что*

$$B_p E\left(\left(\sum_{i=1}^n Z_i^2\right)^{p/2}\right) \leqslant E\left(\left|\sum_{i=1}^n Z_i\right|^p\right) \leqslant C_p E\left(\left(\sum_{i=1}^n Z_i^2\right)^{p/2}\right). \quad (20)$$

Более того,

$$E\left(\sqrt{\sum_{i=1}^n Z_i^2}\right) \geqslant E\left(\left|\sum_{i=1}^n Z_i\right|\right) \geqslant \sqrt{n/8}\, E(|Z_1|).$$

Полагая $q = P(|Z_1| \geqslant u)$, $u > 0$, получим также

$$E\left(\left|\sum_{i=1}^n Z_i\right|\right) \geqslant u \frac{\sqrt{nq}}{4\sqrt{2}} \frac{nq}{8+nq} \geqslant \frac{u\sqrt{nq}}{20\sqrt{2}}$$

при $nq \geqslant 2$.

Замечание. Первое неравенство леммы 27 принадлежит Marcinkiewicz, Zygmund (1937) (см., например, работу Манставичуса (1982) и указанные в ней ссылки). Ограничимся доказательством других неравенств леммы 27.

Доказательство леммы 27. Мы будем неоднократно использовать неравенство $E(|U+V|) \geqslant E(|U+E(V|U)|)$, верное для произвольных случайных величин U, V. Сперва симметризуем задачу, используя тот факт, что

$$E(|U|) = E(|U|+|V|)/2 \geqslant E(|U-V|)/2$$

для независимых одинаково распределенных случайных величин U, V. Таким образом,

$$E\left(\left|\sum_{i=1}^n Z_i\right|\right) \geqslant E\left(\left|\sum_{i=1}^n (Z_i - Z_i')\right|\right)\Big/2,$$

где величины $Z_1', \ldots, Z_n', Z_1, \ldots, Z_n$ независимы и одинаково распределены. По неравенству Сарека (лемма 28) и неравенству Йенсена получим, что нижняя граница больше или равна

$$\frac{1}{2\sqrt{2}} E\left(\sqrt{\sum_{i=1}^n (Z_i - Z_i')^2}\right) \geqslant$$

$$\geqslant \sqrt{n/8}\, E(|Z_1 - Z_1'|) \geqslant \sqrt{n/8}\, E(|Z_1|).$$

Чтобы убедиться в этом, представим $Z_i - Z_i'$ в виде R_iB_i, где $R_i = |Z_i - Z_i'|$ и $B_i = \operatorname{sgn}(Z_i - Z_i')$ — не зависящие друг от друга величины, и возьмем условное математическое ожидание при фиксированных R_i.

Последнее неравенство леммы 27 получается следующим образом. Положим $v = E(R_1 \mid R_1 \geqslant u)$, заметим, что $v \geqslant u$, и введем $N = \sum_{i=1}^{n} I_{[R_i \geqslant u]}$. Имеем

$$E\left(\left|\sum_{i=1}^{n} R_iB_i\right|\right) = E\left(\left|\sum_{i=1}^{n} B_i\left(R_iI_{[R_i \geqslant u]} + R_iI_{[R_i < u]}\right)\right|\right) \geqslant$$

$$\geqslant E\left(\left|\sum_{i=1}^{n} B_i\left(vI_{[R_i \geqslant u]} + E(R_1 \mid R_1 < u)\, I_{[R_i < u]}\right)\right|\right) =$$

(это выражение получается, если взять условное математическое ожидание при фиксированных B_i, $I_{[R_i \geqslant u]}$ и применить условный вариант неравенства Йенсена)

$$= E\left(\left|\sum_{i=1}^{N} vB_i + \sum_{j=1}^{n-N} B_{j+N}E(R_1 \mid R_1 < u)\right|\right) \geqslant$$

$$\geqslant E\left(\left|\sum_{i=1}^{N} vB_i\right|\right) \geqslant \quad \text{(в силу независимости при фиксированном } N\text{)}$$

$$\geqslant uE\left(\left|\sum_{i=1}^{N} B_i\right|\right) \geqslant$$

$$\geqslant uE\left(\sqrt{N/2}\right) \qquad \text{(по неравенству Сарека).}$$

Положим $r = P(|Z_1 - Z_1'| \geqslant u)$ и заметим, что $r \geqslant P(|Z_1| \geqslant u)/2 = q/2$. Так как величина N имеет биномиальное распределение с параметрами (n, r), то, применяя при $nr \geqslant 1$ неравенство Чебышёва в форме Кантелли, получим

$$E\left(\left|\sum_{i=1}^{n} Z_i\right|\right) \geqslant uE\left(\sqrt{N/8}\right) \geqslant u\sqrt{nr}\, P(N \geqslant nr/2)/4 \geqslant$$

$$\geqslant u\frac{\sqrt{nr}}{4}\,\frac{(nr/2)^2}{(nr/2)^2 + nr(1-r)} \geqslant$$

$$\geqslant u\frac{\sqrt{nr}}{4}\,\frac{nr}{4+nr} \geqslant u\frac{\sqrt{nr}}{20} \geqslant u\frac{\sqrt{nq}}{20\sqrt{2}},$$

что и требовалось доказать.

Лемма 28 (неравенства для биномиального распределения; неравенство Хинчина). *Пусть $Y_1, \ldots, Y_n$ — независимые случайные величины, принимающие значения $+1$ и -1 с равными вероятностями, и $a_1, \ldots, a_n$ — действительные числа. Тогда существуют*

положительные постоянные B_p *и* C_p, *зависящие только от* $p > 0$, *такие, что*

$$B_p \sqrt{\sum_{i=1}^n a_i^2} \leqslant E^{1/p}\left(\left|\sum_{i=1}^n a_i Y_i\right|^p\right) \leqslant C_p \sqrt{\sum_{i=1}^n a_i^2}.$$

Оптимальными являются следующие значения постоянных:

$$B_p = 2^{1/2-1/p}, \quad 0 < p \leqslant p_0,$$
$$B_p = 2^{1/2}\left(\Gamma\left((p+1)/2\right)/\sqrt{\pi}\right)^{1/p}, \quad p_0 \leqslant p \leqslant 2,$$
$$B_p = 1, \quad p \geqslant 2,$$
$$C_p = 2^{1/2}\left(\Gamma\left((p+1)/2\right)/\sqrt{\pi}\right)^{1/p}, \quad 2 \leqslant p,$$
$$C_p = 1, \quad 0 < p \leqslant 2.$$

Здесь $p_0 = 1{,}84742\ldots$ — решение уравнения $\Gamma((p+1)/2) = \Gamma(3/2)$, принадлежащее интервалу $(1, 2)$.

Замечание. Оптимальная постоянная $B_1 = 1/\sqrt{2}$ получена в работе Szarek (1976) (мы называем соответствующее неравенство неравенством Сарека). Наилучшие значения C_p, $p \geqslant 3$, получены в работе Young (1976). Все случаи одновременно рассмотрены в работе Haagerup (1978). Кстати, заметим, что если величина Y имеет биномиальное распределение с параметрами $(n, 1/2)$, то $E(|Y - n/2|) \geqslant \sqrt{n/8}$.

Лемма 29. *Пусть* f — *плотность с характеристической функцией* φ, *и пусть* K — *борелевская функция, удовлетворяющая условиям* $\int K = 1$, $\int |K| < \infty$. *Предположим, что характеристическая функция для* K *равна* $\psi(t) = \int e^{itx} K(x)\, dx$, $t \in R$. *Тогда*

$$\int |f - f * K| \geqslant \sup_t |\varphi(t) - \varphi(t)\psi(t)|.$$

Если $h \to c$, *где* $c > 0$ — *некоторая постоянная, то* $\int |f * K_h - f * K_c| \to 0$.

Доказательство.

$$\sup_t |\varphi - \varphi\psi| = \sup_t \left|\int (f(x) - (f * K)(x))\, e^{itx}\, dx\right| \leqslant$$
$$\leqslant \int |f(x) - (f * K)(x)| \sup_t |e^{itx}|\, dx = \int |f - f * K|.$$

Для доказательства второго утверждения леммы 29 проверьте, что соотношение (2.4) из доказательства теоремы 2.4 остается в силе.

Доказательство теоремы 16. Утверждение 4 теоремы 16 содержится в теореме 2. Для доказательства остальных утверждений будут использоваться грубые неравенства

$$E\left(\int |f_n - f|\right) \geqslant \int |f - f * K_h|, \tag{21}$$

$$E\left(\int |f_n - f|\right) \geqslant E\left(\int |f_n - f * K_h|\right)/2 \tag{22}$$

(см., например, лемму 3.6). Предположим, что утверждение 1 доказано. Тогда утверждение 3 доказывается от противного. В самом деле, если существует подпоследовательность h, для которой величина $\sqrt{n}E\left(\int |f_n - f|\right)$ остается ограниченной, то существует постоянная c, такая, что $h \to c$ на этой подпоследовательности. Исходя из леммы 29, заключаем, что $\sup_t |\varphi(t) - \varphi(t)\psi(ct)| = 0$. Но отсюда следует, что $\psi(t) = 1$, т. е. мера, соответствующая ψ, является атомической с единичной массой в нуле. Мы пришли к противоречию.

Рассмотрим теперь одновременно утверждения 1 и 2. Применим лемму 27 при $Z_i = Z_i(x) = (K_h(x - X_i) - E(K_h(x - X_i)))/n$. Заметим, что $E(Z_i(x)) = 0$ и

$$E\left(\int |f_n - f * K_h|\right) = \int E\left(\left|\sum_{i=1}^{n} Z_i(x)\right|\right) dx \geqslant$$

$$\geqslant \int_B \frac{a}{nh}\sqrt{n}\,(20\sqrt{2})^{-1}\sqrt{P\left(|Z_1(x)| \geqslant \frac{a}{nh}\right)}\,dx,$$

где $a > 0$ — число, которое будет выбрано далее, и B — множество всех x, для которых вероятность $P(|Z_1(x)| \geqslant a/nh)$ больше или равна $2/n$. Здесь мы воспользовались леммой 27. Таким образом,

$$\sqrt{nh}\,E\left(\int |f_n - f|\right) \geqslant \frac{a}{40\sqrt{2h}} \int_B \sqrt{P\left(|Z_1(x)| \geqslant \frac{a}{nh}\right)}\,dx.$$

Пусть C — множество всех x, для которых $|K(x)| \geqslant a/2$, и пусть число $a > 0$ столь мало, что $\int_C dx = b > 0$ при некотором положительном b. Из теоремы 2.3 следует, что $E(|K_h(x - X_1)|) = f * |K|_h \to f\int |K|$ при почти всех x. При таких x и при h, достаточно малых для того, чтобы по крайней мере выполнялось неравенство $E(|K_h(x - X_1)|) \leqslant a/2h$, имеем

$$P(|Z_1(x)| \geqslant a/nh) \geqslant P(|K_h(x - X_1)| \geqslant$$

$$\geqslant a/h - E(|K_h(x - X_1)|)) \geqslant$$

$$\geqslant P(|K_h(x-X_1)| \geqslant a/2h) = P\left(\left|K\left(\frac{x-X_1}{h}\right)\right| \geqslant \frac{a}{2}\right) =$$
$$= P(X_1 \in x - hC) \sim$$
$$\sim f(x)\, h \int dx = \qquad \text{(при почти всех } x \text{ по теореме 2.3)}$$
$$= f(x)\, hb.$$

Таким образом, если $nh \to \infty$, то по лемме Фату

$$\liminf_{n\to\infty} \sqrt{nh}\, E\left(\int |f_n - f|\right) \geqslant$$
$$\geqslant \frac{a}{40\sqrt{2}} \int \liminf_{n\to\infty} I_B \sqrt{\frac{P(|Z_1(x)| > a/nh)}{h}}\, dx \geqslant$$
$$\geqslant \frac{a}{40\sqrt{2}} \int\limits_{f>0} b\sqrt{f(x)}\, dx.$$

Отсюда непосредственно вытекает утверждение 2 и частично утверждение 1.

Остается доказать лишь утверждение 1 в случае, когда $\liminf_{n\to\infty} nh < \infty$. Очевидно, можно считать, что $\limsup_{n\to\infty} nh \leqslant C_0 < \infty$, поскольку случай, когда $\limsup_{n\to\infty} nh = \infty$, можно затем рассмотреть тривиальным образом. В силу последнего неравенства леммы 27

$$\sqrt{n}\, E\left(\int |f_n - f|\right) \geqslant \frac{a}{4\sqrt{2}} \int \frac{1}{h}\sqrt{q}\, \frac{nq}{8+nq}\, dx,$$

где $q = P(|Z_1(x)| \geqslant a/nh) \geqslant f(x)\, hb\,(1 + o(1))$ при почти всех x. По лемме Фату получим, что нижний предел левой части этого неравенства равен ∞, если

$$\liminf_{n\to\infty} \frac{nh}{c+nh} \cdot \frac{1}{\sqrt{h}} = \infty$$

при любых постоянных $c > 0$. Но последнее следует из того, что $\liminf nh > 0$, $h \to 0$. Остается рассмотреть случай, когда $h = o(1/n)$. Утверждение 1 теоремы 16 и тем самым сама эта теорема будут доказаны, если показать, что из условия $h = o(1/n)$ следует

$$\liminf_{n\to\infty} E\left(\int |f_n - f|\right) \geqslant 1.$$

(Заметим, что это не вытекает из теоремы 3.1, поскольку здесь допускаются ядра, принимающие отрицательные значения.) Имеем

$$E\left(\int |f_n - f|\right) \geqslant E\left(\int (f - f_n)_+\right) =$$
$$= E\left(\int \left(f I_{[f_n \leqslant f]} - f_n I_{[f_n \leqslant f]}\right)\right) =$$
$$= \int f P(f_n \leqslant f) - \int E\left(f_n I_{[f_n \leqslant f]}\right).$$

В силу теоремы Лебега о мажорируемой сходимости для получения желаемого результата достаточно показать, что $f_n \to 0$ по вероятности при почти всех x (относительно плотности f). Для доказательства этого факта введем множество $C = \{x: | K(x) | \geqslant \varepsilon\}$, где $\varepsilon > 0$ — произвольное число, и заметим, что если множество $x + hC$ не содержит значений X_i, то $|f_n| \leqslant \varepsilon$. Но P (в $x + hC$ содержится хотя бы одно из X_i) $\leqslant n \int\limits_{x+hC} f \to nhf(x) \int\limits_C dy$ при почти всех x. (Это следует из теоремы 2.3, если функция I_C имеет интегрируемую радиальную мажоранту, и, конечно, является простым следствием того, что ядро K имеет интегрируемую радиальную мажоранту.) Далее, $\int\limits_C dx \leqslant \int | K |/\varepsilon < \infty$, откуда в силу произвольности ε и того, что $nh = o(1)$, следует желаемый результат.

Так как множитель $\int \sqrt{f}$ равен бесконечности для многих плотностей из A_T таких, как плотность Валле-Пуссена (в действительности, можно показать, что он бесконечен для любых плотностей с действительнозначными четными характеристическими функциями, выпуклыми на $[0, T]$ и равными нулю вне $[-T, T]$), то для того, чтобы можно было пользоваться равномерной границей леммы 25 и неравенством D леммы 26, необходимо в явном виде ввести требование конечности интеграла $\int \sqrt{f}$. По этой причине мы введем несколько более узкий класс плотностей $A_{T, s, C}$, где T, C — положительные постоянные и $s \geqslant 1$ — целое число.

$$A_{T,s,C} = \Big\{ f: f \text{ имеет характеристическую функцию } \varphi, \text{ такую, что (i) } \varphi = 0 \text{ вне } [-T, T] \text{ (т. е. } f \in A_T); \text{ (ii) функции } \varphi, \ldots, \varphi^{(s-1)} \text{ существуют и абсолютно непрерывны (их производные п. в. равны } \varphi^{(1)}, \ldots, \varphi^{(s)}); \quad \text{(iii)} \int | \varphi^{(s)} | \leqslant C \Big\}.$$

Это непустой класс, который, безусловно, нельзя описать конечным числом параметров. Он, например, содержит все плотности, характеристические функции которых представляют собой нормированные* (так, чтобы их значение в нуле равнялось 1) свертки достаточно большого числа функций вида $(1 - |t|)_+$ (эти свертки служат приближением для характеристической функции нормального распределения). Без учета масштабных множителей такие плотности имеют вид

$$C_p \left(\frac{\sin x}{x} \right)^{2p},$$

где p — целое число. Ясно, конечно, что при некоторых значениях T и C класс $A_{T, s, C}$ пуст. Предупредим со всей определен-

ностью, что нас интересует не сам по себе класс $A_{T,\,s,\,C}$, который, по-видимому, является довольно искусственным и нереалистичным, а только факт существования достаточно богатого класса плотностей $\mathcal{F}$, для которого можно построить такую оценку, что

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \sqrt{n} E\left(\int |f_n - f|\right) < \infty.$$

Теперь рассмотрим ядерную оценку с ядром $K(x) = g_a(x)$ и назовем ее *трапецеидальной ядерной оценкой* по причине того, что характеристическая функция для g_a имеет трапецеидальную форму. Пусть $a > 0$ — фиксированное число, и пусть $h > 0$ — фиксированное число, такое, что

$$h \leqslant (a/(1+a))\,(1/T). \tag{23}$$

Тогда по лемме 25 и теореме 15

$$E\left(\int |f_n - f|\right) = E\left(\int |f_n - f * K_h|\right) \leqslant$$
$$\leqslant (nh)^{-1/2} \int \sqrt{f * (K^2)_h}. \tag{24}$$

Теперь граница для последнего интеграла в (24) будет получена просто из следующей равномерной верхней границы для любых f, принадлежащих $A_{T,\,s,\,C}$:

Лемма 30 (неравенства, связывающие плотность f с ее характеристической функцией φ). *Для любой плотности f с характеристической функцией φ имеем*

$$\sup f \leqslant \frac{1}{2\pi} \int |\varphi|.$$

При любых $f \in A_{T,\,s,\,C}$ имеем

$$|x|^s f \leqslant \frac{1}{2\pi} \int |\varphi^{(s)}| \leqslant \frac{C}{2\pi}$$

и

$$f(x) \leqslant g(x) \triangleq \min\left(\frac{T}{\pi},\ \frac{C}{2\pi |x|^s}\right).$$

Доказательство. Первое неравенство следует из того, что

$$f(x) = (2\pi)^{-1} \int e^{-itx} \varphi(t)\, dt \leqslant (2\pi)^{-1} \int |\varphi|.$$

Заметим, что при $f \in A_T$ последняя величина не превосходит T/π. Далее, интегрируя по частям формулу обращения и учитывая абсолютную непрерывность функций $\varphi, \ldots, \varphi^{(s-1)}$, получим

$$f(x) = (2\pi)^{-1} \int (ix)^{-s} e^{-itx} \varphi^{(s)}(t)\, dt \leqslant$$
$$\leqslant (2\pi |x|^s)^{-1} \int |\varphi^{(s)}|.$$

Лемма 30 доказана.

Класс $A_{T,s,C}$ не замкнут относительно сдвигов, поскольку в противном случае не существовало бы интегрируемой равномерной границы для f. Легко, конечно, определить и класс, инвариантный относительно сдвигов, но мы не хотим загромождать этим изложение. Заметим, что величина $\int|\varphi|$ несет информацию о максимальном значении f, а величина $\int|\varphi^{(s)}|$ — о равномерной верхней границе для хвостов плотности f. Сформулируем теперь последний результат этого параграфа.

Теорема 17. *Пусть $s \geqslant 3$ — целое число и числа T и C достаточно велики для того, чтобы класс $A_{T,s,C}$ не был пуст. Тогда трапецеидальная ядерная оценка f_n, для которой параметр сглаживания h выбирается, исходя из* (23), *и ядро $K = g_1$, удовлетворяет неравенству*

$$\sup_{f \in A_{T,s,C}} \sqrt{n} E\left(\int|f_n - f|\right) \leqslant$$

$$\leqslant \left(\frac{16s}{s-2} \frac{T^{1/2-1/s}}{\pi^{3/2}\sqrt{2}} \left(\frac{C}{2}\right)^{1/s} + \frac{4}{\sqrt{\pi}}\right) \frac{1}{\sqrt{h}}.$$

Доказательство. Вначале убеждаемся в том, что

$$\int \min(\alpha, \beta/|x|^\gamma) = \frac{2\gamma}{\gamma-1} \alpha^{1-1/\gamma} \beta^{1/\gamma}$$

при положительных α, β, γ. Затем применяем в (24) следующие неравенства: $f \leqslant g$, где функция g определена в лемме 30, и $|K(x)| \leqslant 2\min(1/2\pi, (4/x^2)(1/\pi)) = \min(1/\pi, (4/\pi)x^{-2}) = K^*(x)$ (по определению функции K^*). Кроме того, $f * (K^2)_h \leqslant g * (K^{*2})_h$. Так как функции g и K^* симметричны и унимодальны, то можно применить утверждение D леммы 26. Результат теоремы получается, если заметить, что

$$\int|K^*| = 8/\pi, \quad \int K^{*2} = 16/3\pi^2,$$

$$\int\sqrt{g} = \frac{2s}{s-2} T^{1/2-1/s} \left(\frac{C}{2}\right)^{1/s} \pi^{-1/2}.$$

Чтобы сделать правую часть неравенства теоремы 17 как можно меньшей, нужно выбрать как можно большее h. Подставляя туда значение $h = 1/2T$, получим минимаксную верхнюю границу для $A_{T,s,C}$. При различных значениях параметра a трапецеидальной ядерной оценки получаются различные минимаксные границы. Формальная оптимизация этих границ по a здесь не проводится.

В этом параграфе остался нерешенным вопрос о том, может ли ядерная оценка с неслучайным h быть состоятельной для любых

плотностей, если h не стремится к 0 [1]). Так, не была построена оценка плотности, состоятельная при любых h и удовлетворяющая неравенству

$$\limsup_{n\to\infty} \sqrt{n}\, E\left(\int |f_n - f|\right) \leqslant c$$

при любых f из достаточно богатого семейства плотностей $\mathscr{F}$. Другой вопрос, оставшийся нерешенным, — существует ли оценка f_n, такая, что

$$\limsup_{n\to\infty} \sup_{f\in A_T} \sqrt{n} E\left(\int |f_n - f|\right) < \infty.$$

Ядро трапецеидальной ядерной оценки таково, что $\int x^i K = 0$ при любых $i > 0$. Занимательное свойство этой оценки состоит в том, что для нее в отличие от оценок Бартлета не нужно менять ядро K в зависимости от гладкости f. Параметр h, конечно, по-прежнему нужно настраивать, исходя из гладкости f, или, применяя альтернативный подход, оценивать h по данным одним из методов, описанных в гл. 6.

Наконец, следуя интересной L_2-теории, разработанной Watson, Leadbetter (1963) и переоткрытой Davis (1975, 1977), можно рассмотреть вопрос о том, какие скорости сходимости достигаются при различных условиях на хвост функции φ. При этом окажется, что существует континуальное число скоростей сходимости между $1/\sqrt{n}$ и $n^{-2/5}$ (таких, как $\log n/\sqrt{n}$ и т. д.), зависящих от того, как убывает хвост функции $|\varphi|$. Перечисление обычных условий на хвосты было бы равнозначно признанию того, что описываемые этими условиями классы важны. Но последнее неверно — они, так же как и класс $A_{T,s,c}$, не являются важными. Редки случаи, когда что-либо известно относительно гладкости f, и велика плата за выбор «неправильного» параметра h ядерной оценки! (Например, если ошибочно «сделать ставку» на то, что f принадлежит классу $A_{T,s,c}$, выбрать фиксированное значение h и воспользоваться трапецеидальной ядерной оценкой, то наказание оказывается суровым — f_n может вовсе не сходиться к f.)

ЛИТЕРАТУРА

Abou-Jaoude S.
(1977) La convergence L_1 et L_∞ de certains estimateurs d'une densité de probabilité, Thèse de Doctorat d'État, Université de Paris VI, Paris, France.

[1]) В случае, когда ядро оценки неотрицательно, ответ на этот вопрос следует из теоремы 3.1, где утверждается, что состоятельность в L_1 при $h \not\to 0$ невозможна. — *Прим. перев.*

Bartlett M. S.
(1963) Statistical estimation of density functions, Sankhya Series A, 25, pp. 245—254.
Beckenbach E. F. Bellman R.
(1965) Inequalities, Springer-Verlag, Berlin. (Имеется русский перевод: Беккенбах Э., Беллман Р. Неравенства. — М.: Мир, 1965.)
Bretagnolle J., Huber C.
(1979) Estimation de densités: risque minimax, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 47, pp. 119—137.
Carlson F.
(1934) Une inégalité, Arkiv foer Matematik, Astromi och Fysik, 25B, pp. 1—15.
Davis K. B.
(1975) Mean square error properties of density estimates, Annals of Statistics, 3, pp. 1025—1030.
(1977) Mean integrated square error properties of density estimates, Annals of Statistics, 5, pp. 530—535.
Deheuvels P.
(1977) Estimation non paramétrique de la densité par histogrammes généralisés, Revue de Statistique Appliquée, 25, pp. 5—42.
Deheuvels P., Hominal P.
(1980) Estimation automatique de la densité, Revue de Statistique Appliquée, 28, pp. 25—55.
Devroye L., Penrod C. S.
(1984) Distribution-free lower bounds in density estimation. Annals of Statistics, 12, pp. 1250—1262.
Feller W.
(1968) An Introduction to Probability Theory and Its Applications, Vol. 1, Wiley, New York. (Имеется русский перевод: Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1. — М.: Мир, 1984.)
(1971) An Introduction to Probability Theory and Its Applications, Vol. 2, Wiley, New York. (Имеется русский перевод: Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.)
Freedman D., Diaconis P.
(1981) On the histogram as a density estimator: L_2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 57, pp. 453—476.
Haagerup U.
(1978) Les meilleures constantes de l'inégalite de Khintchine, Comptes Rendus de l'Académie des Sciences de Paris A, 286, pp. 259—262.
Kulldorf G.
(1963, 1964) On the optimum spacing of sample quantiles from a normal distribution, Part 1, Skandinavisk Aktuarietidskrift, 46, pp. 143—161, 1963; Part 2, Skandinavisk Aktuarietidskrift, 47, pp. 71—87, 1964.
Lukacs E.
(1970) Characteristic Functions, Griffin, London. (Имеется русский перевод: Лукач Е. Характеристические функции. — М.: Наука, 1979.)
Marcinkiewicz J., Zygmund A.
(1937) Sur les fonctions indépendantes, Fundamentales de Mathematiques, 29, pp. 60—90.
Parzen E.
(1962) On estimation of a probability density function and the mode, Annals of Mathematical Statistics, 33, pp. 1065—1076.
Rosenblatt M.
(1956) Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, pp. 832—835.
(1971) Curve estimates, Annals of Mathematical Statistics, 42, pp. 1815—1842.
(1979) Global measures of deviation for kernel and nearest neighbor density estimates, in Smoothing Techniques for Curve Estimation, Th. Gasser and M. Rosen-

blatt (Eds.), Lecture Notes in Mathematics # 757, pp. 181—190, Springer-Verlag, Berlin.

Scott D. W.
(1979) On optimal data-based histograms, Biometrika, 66, pp. 605—610.

Szarek S. J.
(1976) On the best constants in the Khintchine inequality, Studia Mathematica, 63, pp. 197—208.

Tapia R. A., Thompson J. R.
(1978) Nonparametric Probability Density Estimation, The Johns Hopkins University Press, Baltimore.

Watson G. S., Leadbetter M. R.
(1963) On the estimation of the probability density, Annals of Mathematical Statistics, 34, pp. 480—491.

Wertz W.
(1972) Fehlerabschätzung für eine Klasse von nichtparametrischen Schätzfolgen, Metrika, 19, pp. 131—139.

Whittaker E. T., Watson G. N.
(1963) A Course of Modern Analysis, 4th ed., Cambridge University Press, Cambridge, U. K. (Имеется русский перевод: Уиттекер Э. Т., Ватсон Д. Н. Курс современного анализа (в 2-х томах). — М.: Физматгиз, 1962—1963.)

Young R. M. G.
(1976) On the best possible constants in the Khintchine inequality, Journal of the London Mathematical Society, 14, pp. 496—504.

Епанечников В. А.
(1969) Непараметрическая оценка многомерной плотности вероятности//Теория вероятн. и ее примен. — Т. 14. — 1. — С. 156—161.

Ибрагимов И. А., Хасьминский Р. З.
(1982) Об оценке плотности распределения, принадлежащей одному классу целых функций//Теория вероятн. и ее примен. — Т. 27. — 3. — С. 514—524.

Конаков В. Д.
(1972) Непараметрическая оценка плотности распределения вероятностей//Теория вероятн. и ее примен. — Т. 17. — 2. — С. 377—379.

Манставичус Э.
(1982) Неравенства для момента порядка p, $0 < p < 2$, суммы независимых случайных величин//Литовск. матем. сборник. — Т. 22. — 1. — С. 112—117.

Петров В. В.
(1972) Суммы независимых случайных величин. — М.: Наука.

Глава 6

Адаптивная ядерная оценка, ее сходимость в L_1 и поточечная сходимость

1. Основной результат

В этой главе изучается состоятельность адаптивной ядерной оценки

$$f_n(x) = (nh^d)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

где K — некоторая фиксированная плотность и $h = h(n, X_1, \ldots, X_n)$ — борелевская функция от n и от данных. В идеальном случае h не зависит от параметров, которые должен выбирать пользователь, хотя оценки с такой зависимостью (в том числе обычная ядерная оценка), строго говоря, тоже называются адаптивными ядерными оценками. Заметим, что зависимость h от x не допускается, поскольку при этом могут получаться оценки, не являющиеся плотностями на R^d.

Первый и основной результат этой главы аналогичен теореме 3.1 для обычной ядерной оценки. Он формулируется здесь без доказательства (доказательство см. в § 5). В § 4 приведено несколько примеров адаптивных ядерных оценок. В § 2 и 3 исследуется поточечная сходимость этих оценок.

Теорема 1. *Пусть f_n — адаптивная ядерная оценка, ядро K которой — произвольная плотность. Если $h + (nh^d)^{-1} \to 0$ вполне (почти наверное, по вероятности), то $\int |f_n - f| \to 0$ вполне (почти наверное, по вероятности) для любых плотностей f на R^d.*

2. Поточечная сходимость адаптивной ядерной оценки

Другой, менее мощный способ доказательства состоятельности в L_1 оценки плотности состоит в том, что устанавливается поточечная сходимость оценки при почти всех x и применяется теорема 2.7. Шеффе или ее обобщение, принадлежащее Глику (теорема 2.8). Имеется следующая теорема о поточечной сходимости адаптивной ядерной оценки при любых плотностях f на R^d.

Теорема 2. *Пусть K — интегрируемая по Риману плотность с компактным носителем, f_n — адаптивная ядерная оценка*

с параметром сглаживания h. Пусть f — произвольная фиксированная плотность на R^d. Тогда

A. *Если* $h + (nh^d)^{-1} \to 0$ *по вероятности, то* $f_n \to f$ *по вероятности при почти всех* x.

B. *Если* $h \to 0$ *и* $nh^d/(\log \log n) \to \infty$ *почти наверное, то* $f_n \to f$ *почти наверное при почти всех* x.

C. *Если* $h \to 0$ *и* $nh^d/(\log n) \to \infty$ *вполне, то* $f_n \to f$ *вполне при почти всех* x.

Доказательство теоремы 2 приведено в § 5. Из утверждений A и B непосредственно следует состоятельность в L_1 по вероятности и почти наверное, но они слабее, чем соответствующие утверждения теоремы 1. Указание «при почти всех x» относится ко всем точкам Лебега плотности f. Его нельзя опустить, поскольку плотность f определяется с точностью до множества лебеговой меры нуль. Теорема 2 по существу неулучшаема, ибо условия на h в утверждениях A, B и C необходимы для состоятельности обычной ядерной оценки (последнее было впервые установлено Дехёвельсом (Deheuvels (1974)) при различных условиях регулярности на K, f и h). Уместно привести результат Дехёвельса. В следующем параграфе сформулировано его обобщение, свободное от большей части условий регулярности. Доказательство приведено в § 5.

В теореме 2 допускается, что параметр h зависит от x, но в случае, когда такая зависимость есть, из этой теоремы нельзя вывести состоятельность оценки в L_1, пользуясь теоремами 2.7 и 2.8.

Относительно равномерной сходимости адаптивной ядерной оценки известно несколько результатов в духе теоремы 2, например Wagner (1975), Devroye, Wagner (1980), Deheuvels, Hominal (1980). В последней работе дан набросок доказательства следующего утверждения: если $h \to 0$ почти наверное, $nh^d/(\log n) \to \infty$ почти наверное, плотность f равномерно непрерывна и ядро K является интегрируемой по Риману плотностью, то $\sup_x |f_n(x) - f(x)| \to 0$ почти наверное.

3. Поточечная сходимость обычной ядерной оценки

Определение. Последовательность положительных чисел a_n называется *семимонотонной*, если существует постоянная $c > 0$, такая, что $a_{n+m} \geqslant ca_n$ при $m, n \geqslant 1$. Заметим, что отсюда следует, что либо $\liminf_{n\to\infty} a_n = \infty$, либо $\sup_n a_n < \infty$.

Теорема 3. *Пусть f_n — обычная ядерная оценка, у которой параметр сглаживания h зависит только от n, и K — ограниченная плотность с компактным носителем.*

1. Слабый вариант. *Следующие утверждения эквивалентны:*

A. $f_n \to f$ *по вероятности при почти всех* x *для некоторой плотности* f.

B. $f_n \to f$ *по вероятности при почти всех* x *для любых* f.

C. $h \to 0$ *и* $nh^d \to \infty$.

D. $\int |f_n - f| \to 0$ *по вероятности для некоторой плотности* f.

E. $\int |f_n - f| \to 0$ *вполне для любых* f.

2. Сильный вариант. *Пусть, кроме того, ядро* K *интегрируемо по Риману и последовательность* $nh^d/(\log\log n)$ *семимонотонна. Тогда следующие утверждения эквивалентны:*

A. $f_n \to f$ *почти наверное при почти всех* x *для некоторой плотности* f.

B. $f_n \to f$ *почти наверное при почти всех* x *для любых* f.

C. $h \to 0$ *и* $nh^d/(\log\log n) \to \infty$.

Интегрируемость по Риману ядра K *не требуется для импликации* $A \Rightarrow C$, *и условие семимонотонности не используется в доказательстве того, что* $C \Rightarrow B$.

3. Вариант со сходимостью вполне. *Пусть последовательность* $nh^d/(\log n)$ *семимонотонна. Тогда следующие утверждения эквивалентны:*

A. $f_n \to f$ *вполне при почти всех* x *для некоторой плотности* f.

B. $f_n \to f$ *вполне при почти всех* x *для любых* f.

C. $h \to 0$ *и* $nh^d/(\log n) \to \infty$.

Условие семимонотонности не требуется для импликации $C \Rightarrow B$.

Теорема 3 включена в эту книгу отчасти для того, чтобы показать, что обычная ядерная оценка может быть сильно состоятельной в L_1 (т. е. $\int |f_n - f| \to 0$ почти наверное) и в то же время не сходиться почти наверное при почти всех x. В частности, это имеет место для любых f, если nh^d равно $\log\log\log n$ или $\sqrt{\log\log n}$, или $c\log\log n$, где c — некоторая положительная постоянная.

4. Примеры адаптивных ядерных оценок

Бóльшая часть методов адаптивного ядерного оценивания относится к одному из следующих двух классов. Первый класс содержит все методы, связанные с оцениванием по выборке неизвестных параметров главного члена в асимптотическом разложении некоторого критерия качества. Ко второму классу относятся методы, в которых h получается путем *непосредственной* минимизации некоторого критерия (например, должным образом

видоизмененного критерия максимума правдоподобия). В последнем случае часто требуется большее количество вычислений, а также, как правило, усложняется теоретический анализ свойств адаптивной ядерной оценки.

Оценки, основанные на асимптотических разложениях. Первым и наиболее важным из методов, основанных на асимптотических разложениях, является параметрический метод, который был изложен в § 5.6 на примере ошибки в L_1. Параметрический метод дает превосходные результаты, если гипотеза о принадлежности плотности f к заданному параметрическому семейству плотностей верна. Например, если это параметрическое семейство — нормальное с неизвестным средним μ и дисперсией σ^2 и если $\hat{\sigma}^2$ — построенная по данным оценка для σ^2, то при параметрическом методе используется параметр сглаживания

$$h = \hat{\sigma}\left(\frac{15e\sqrt{2\pi}}{8n}\right)^{1/5} = \frac{1{,}6644\ldots\hat{\sigma}}{n^{1/5}}.$$

В силу теоремы 1 при любых f, таких, что $\hat{\sigma} \to c$ почти наверное, где c — некоторая постоянная (последнее выполнено почти для всех плотностей, если $\hat{\sigma}$ — разумная оценка, построенная по данным), имеем $\int |f_n - f| \to 0$ почти наверное. Это утверждение остается в силе, если величина $\hat{\sigma}$ отграничена почти наверное от нуля и от бесконечности (последнее выполнено при любых f в случае, когда $\hat{\sigma}$ — робастная оценка, основанная на квантилях; см. § 5.6).

В действительности можно утверждать большее. В качестве хорошего упражнения можно получить условия, при которых из сходимости $\hat{\sigma} \to c$ почти наверное (или слабой сходимости) следует, что $E\left(\int |f_n - f|\right) \sim E\left(\int |f_n^* - f|\right)$, где f_n^* — ядерная оценка, у которой в выражении для h_n вместо $\hat{\sigma}$ стоит c. Таким образом, имеется не только состоятельность, но и информация о скорости сходимости. Заметим, однако, что эта скорость оптимальна только в случае, когда f действительно принадлежит рассматриваемому параметрическому семейству; см. также § 7.

Параметрический метод впервые был глубоко исследован Дехёвельсом (Deheuvels (1977)) для плотностей на R^1 и критерия $E\left(\int (f_n - f)^2\right)$. Это исследование было основано на фундаментальном результате Розенблата о том, что если $K \in L_2$ — ограниченная симметричная плотность, $f \in L_2$ — ограниченная и дважды непрерывно-дифференцируемая плотность и $f'' \in L_2$, то обычная ядерная оценка удовлетворяет соотношению

$$E\left(\int (f_n - f)^2\right) \sim (nh)^{-1}\int K^2 + \frac{1}{4}h^4\left(\int x^2 K(x)\,dx\right)^2 \int (f'')^2 \qquad (2)$$

при $h \to 0$ и $nh \to \infty$ (Rosenblatt (1956, 1971)). Из (2) следует, что наилучшее значение h равно $h = \left[A/n \int (f'')^2\right]^{1/5}$, где коэффициент $A = \int K^2 \Big/ \left(\int x^2 K(x)\,dx\right)^2$ зависит только от K. Единственным неизвестным в этом выражении является интеграл $\int (f'')^2$, который равен $3/(8\sqrt{\pi}\sigma^5)$ для нормальной плотности. Дальнейшее обсуждение этого метода см. также в работе Deheuvels, Hominal (1980). Параметрический метод для гистограммной оценки разработан Scott (1979).

Другие авторы предложили оценивать неизвестный коэффициент $\int (f'')^2$ в (2) по данным, применяя непараметрические методы. При этом процедура становится двухшаговой — сначала оценивается $\int (f'')^2$, а затем — f, с использованием оценки интеграла $\int (f'')^2$ в формуле $h = \left[A/n \int (f'')^2\right]^{1/5}$ (Woodroofe (1970)), Надарая (1974), Scott др. (1977), Deheuvels, Hominal (1980), Scott, Factor (1981)). Эти результаты, по-видимому, непосредственно переносятся на случай, когда ошибка измеряется в L_1. Однако в обоих случаях мы снова сталкиваемся с выбором параметров непараметрических оценок (создавая, таким образом, новую задачу, эквивалентную по сложности той, которую мы пытаемся решить) и предполагаем, что существуют состоятельные оценки таких величин, как $\int (f'')^2$. С этой точки зрения параметрический метод является более робастным. Наконец, все указанные методы основаны на некоторых сильных предположениях относительно f, которые могут и не выполняться. Например, если ошибка измеряется в L_1, то требуется оценить величины $\int \sqrt{f}$ и $\int |f''|$ (или соответствующее обобщение последней). Первая величина бесконечна для плотности распределения Коши, а вторая — для равномерной плотности. Из рассмотрения сразу исключаются большие и важные классы плотностей, чего хотелось бы избежать.

Эвристические оценки. Среди оценок, не основанных на асимптотических разложениях, особый класс образуют оценки Wagner (1975) Silverman (1978) и др. В отличие от оценок, основанных на асимптотических разложениях, они не претендуют на достижение какого-либо теоретически предсказанного поведения и в отличие от оценок, рассматриваемых далее, — на оптимизацию какого-либо критерия.

Например, Wagner (1975) вводит расстояния $D_{n1}, \ldots, D_{nn}$ между точками $X_1, \ldots, X_n$ и их k-ми ближайшими соседями соответственно, где $k = [n^a]$, $0 < a < 1$, и предлагает несколько способов выбора h, таких, как

(i) h выбирается случайным образом из набора $D_{n1}, \ldots, D_{nn}$,

(ii) $h = \sum_{i=1}^{n} (D_{ni}/n)$,

(iii) $h = \max_i D_{ni}$,

(iv) $h = \min_i D_{ni}$.

Число возможных способов выбора h фактически неограниченно. Для случая (i) Вагнер показал, что $h \to 0$ почти наверное и $n^b h^{2d} \to \infty$ почти наверное при любых f, если $b < 1 - a$. Таким образом, для ядер, рассмотренных в теореме 2, при любых f имеем $f_n \to f$ почти наверное при почти всех x и $\int |f_n - f| \to 0$ почти наверное.

Оптимизация некоторого критерия. Если выбирать h так, чтобы максимизировать некоторый критерий, то можно надеяться получить оценку, заслуживающую доверия при любых (неизвестных) f даже на достаточно малых объемах выборки. Первый и наиболее важный критерий основан на принципе *максимума правдоподобия* (МП). Для ядерных оценок плотности он был впервые применен в работах Duin (1976) (она была сдана в журнал в 1973 г.) и Habbema и др. (1974). См. также недавний обзор Rudemo (1982). В этих работах предлагается выбирать h из условия максимума функции правдоподобия

$$L(h) = \prod_{i=1}^{n} f_{ni}(X_i), \tag{3}$$

где

$$f_{ni}(x) = \frac{1}{(n-1)h^d} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x - X_j}{h}\right), \tag{4}$$

и использовать это значение h в обычной ядерной оценке. Определяемая таким образом кросс-проверочная [1]) ядерная оценка плотности и аналогичная ей кросс-проверочная гистограммная оценка, по-видимому, в большинстве случаев (но не во всех) работают хорошо. Кросс-проверка [2]) нужна в первую очередь потому, что величина $\prod_{i=1}^{n} f_n(X_i)$ достигает максимума при $h = 0$.

[1]) От английского «cross-validated». — *Прим. перев.*

[2]) От английского «cross-validation». В отечественной литературе также используются термины «перекрестная проверка», «скользящий контроль». — *Прим. перев.*

Основная сложность состоит в доказательстве состоятельности кросс-проверочных оценок. Теорема 1 позволяет свести его к доказательству некоторых утверждений относительно параметра сглаживания h. Впервые доказательство состоятельности кросс-проверочной ядерной оценки на R^1 появилось в работе Chow и др. (1983). Так как основной результат этой работы важен для практики и в его доказательстве используется новая методика, то сформулируем здесь соответствующее его обобщение на R^d, свободное от ненужных условий на f. Доказательство будет дано в § 5.

Теорема 4. *Пусть значение h выбрано так, что $L(h) \geqslant a \sup_{h>0} L(h)$ при некотором $a \in (0, 1)$. Пусть f_n — кросс-проверочная ядерная оценка плотности и f — плотность с компактным носителем. Предположим, что K — интегрируемое по Риману ограниченное ядро с компактным носителем, такое, что $K \geqslant cI_{S_{0r}}$ при некоторых c, $r > 0$ (напомним, что S_{xr} — замкнутый шар радиуса r с центром в точке x). Тогда $h \to 0$ и $\liminf_{n\to\infty} nh^d/(\log n) > 0$ почти наверное.*

Таким образом (вследствие теоремы 2), $f_n \to f$ почти наверное при почти всех x и $\int |f_n - f| \to 0$ почти наверное.

Иногда для получения состоятельных оценок решения оптимизационных задач подвергают двустороннему усечению. Теорема 4 сильна тем, что в ней такое усечение не нужно. Единственным требованием к плотности f является компактность ее носителя. Вместе с тем в случае плотностей на R^1 можно рассмотреть следующую модификацию указанного выше метода:

Шаг 1. Применим к данным монотонное преобразование, переводящее R в $[-1, 1]$ (такое, как $T(x) = x/(1 + |x|)$), в результате чего получим значения $Y_1, \ldots, Y_n$.

Шаг 2. Построим кросс-проверочную ядерную оценку плотности на $[-1, 1]$. (Она *всегда* состоятельна в силу теоремы 4).

Шаг 3. Построим для f преобразованную кросс-проверочную ядерную оценку плотности.

Так как расстояние в L_1 между плотностями инвариантно относительно монотонных преобразований координатных осей, то ошибка этого метода в L_1 равна ошибке при оценивании плотности случайной величины Y_1 посредством кросс-проверочной оценки на шаге 2. Иными словами, рассмотренная здесь модифицированная оценка состоятельна для *любых* плотностей f.

Хотя теорема 4 является обнадеживающей, в ней ничего не говорится о скорости сходимости величины $E\left(\int |f_n - f|\right)$ к 0 — эта задача, по-видимому, ждет своего решения. Еще мы не объяс-

нили, откуда взялось жесткое условие на хвосты плотности f в формулировке теоремы 4. Пусть f — плотность на R^1, ядро K равно нулю вне $[-1, 1]$ и $X_{(1)} < \ldots < X_{(n)}$ — порядковые статистики для $X_1, \ldots, X_n$. Если $h < X_{(n)} - X_{(n-1)}$, то $f_{ni}(X_i) = 0$, где $X_i = X_{(n)}$, так что $L(h) = 0$. Следовательно, параметр h, выбранный методом кросс-проверки, удовлетворяет условию $h \geqslant X_{(n)} - X_{(n-1)}$ при любых n. Пусть теперь K — ядро, ограниченное величиной M. Для любого $\varepsilon > 0$ найдется $\delta > 0$, такое, что

$$\int\limits_{\delta f > 2M} f \geqslant \varepsilon > 0.$$

Далее,

$$\int |f_n - f| = 2 \int\limits_{f > f_n} f - f_n \geqslant 2 \int\limits_{f > 2f_n} (f/2) \geqslant \int\limits_{f > 2M/h} f$$

и последняя величина больше или равна ε при $h \geqslant \delta$. Здесь мы молчаливо воспользовались тем, что $f_n \leqslant M/h$. Таким образом,

$$P\left(\int |f_n - f| \geqslant \varepsilon\right) \geqslant P(h \geqslant \delta) \geqslant P(X_{(n)} - X_{(n-1)} \geqslant \delta). \quad (5)$$

Следовательно, если $X_{(n)} - X_{(n-1)} \not\to \infty$ по вероятности, то $\int |f_n - f|$ не может сходиться к 0 по вероятности. На это впервые указали Schuster, Gregory (1981). Примером плотности, для которой $X_{(n)} - X_{(n-1)} \not\to \infty$ по вероятности, является любая плотность, удовлетворяющая условиям $\lim\limits_{x \to \infty} f(x) \int\limits_x^\infty f(y)\, dy = 0$, $\int\limits_x^\infty f > 0$ при всех x (в частности, *любая* плотность, хвосты которой убывают с полиномиальной скоростью c/x^a, $a > 1$). Этим условиям удовлетворяют плотности распределений Коши, Парето, t-распределения Стьюдента и все устойчивые плотности, за исключением нормальной.

Шустер и Грегори (Schuster, Gregory (1981)) предложили метод устранения несостоятельности в случае распределений с длинными хвостами. По-видимому, пограничное положение между состоятельностью и несостоятельностью соответствует экспоненциальному распределению (для которого $h \not\to 0$ по вероятности, так что оценка не состоятельна). Для распределений с меньшими хвостами, по-видимому, имеет место состоятельность. В силу этого условие компактности носителя f из теоремы 4 является слишком жестким.

Экспериментальное обоснование метода кросс-проверки дается в работах Scott, Factor (1981), Rudemo (1982), Schuster, Gregory (1981). Chow и др. (1983) приводят результат типа теоремы 4 для кросс-проверочных гистограммных оценок. Geman (1981), Geman,

Hwang (1982) применяют принцип МП к ядерной оценке, у которой центры ядер расположены не в точках X_i, а в некоторых точках плана $x_1, \ldots, x_n$. Schuster, Gregory (1978) искусственно разбивают выборку на две части, определяют h из условия максимума функции

$$\prod_{i=1}^{n/2} f_{n/2}(X_i), \tag{6}$$

где $f_{n/2}(x) = (2/n) \sum_{j=n/2+1}^{n} h^{-d} K((x - X_j)/h)$, и используют это значение h в исходной ядерной оценке. При таком подходе требуется меньше вычислений, но результат, по-видимому, получается хуже, чем при кросс-проверке.

Наконец, Hall (1982a, b) показал, что если f — вогнутая функция на $[0, 1]$, то порядок величины параметра сглаживания h в методе кросс-проверки равен $n^{-1/3}$ (этот порядок, конечно, не всегда является оптимальным).

Кроме критерия максимума правдоподобия в методе кросс-проверки можно использовать и другие критерии. Так, можно стремиться выбирать h из условия минимума величины $\int (f_{nh} - f)^2$, где f_{nh} — ядерная оценка с параметром сглаживания h и ядром K. С этой целью Hall (1983a, b), Rudemo (1982) и Bowman (1982) предложили минимизировать критерий $\int f_{nh}^2 - 2M_{nh}$, где M_{nh} — выборочная кросс-проверочная оценка для $\int f_{nh} f$, например

$$M_{nh} = \frac{1}{n} \sum_{i=1}^{n} f_{nhi}(X_i),$$

где

$$f_{nhi}(x) = \frac{1}{n-1} \sum_{j \neq i} h^{-1} K\left(\frac{x - X_j}{h}\right), \quad i = 1, \ldots, n.$$

Stone (1984) заметил, что

$$M_{nh} = \frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j),$$

$$\int f_{nh}^2 = \frac{1}{n^2} \sum_{i,\, j} (K * K)_h(X_i - X_j).$$

Получаемое этим способом значение h^* является наилучшим в том смысле, что

$$\frac{\int (f_{nh^*} - f)^2}{\min\limits_h \int (f_{nh} - f)^2} \to 1$$

почти наверное при любых ограниченных плотностях f на R, если ядро K удовлетворяет условию Липшица (с любым положительным показателем степени), симметрично, имеет компактный носитель и $\int K^2 < 2K(0)$ (Stone (1984)). Ядру K не обязательно быть неотрицательным, но нужна нормировка $\int K = 1$. Подобный результат для гистограммной оценки см. в работе Stone (1983). Вопрос о состоятельности рассмотренной кросс-проверочной оценки при любых f остается открытым. Мы уже видели, что минимизация ошибки в L_2 может на деле привести к тому, что скорость сходимости в L_1 будет исключительно низкой. Не решена еще задача построения оценки h^*, такой, что

$$\frac{\int |f_{nh^*} - f|}{\inf_h \int |f_{nh} - f|} \to 1$$

почти наверное при любых f, или оценки, для которой вместо стремления к 1 в последнем соотношении имеет место неравенство «$\leqslant C + o(1)$».

5. Доказательства

Для доказательства теоремы 1 требуется несколько ключевых лемм.

Лемма 1. *Для любой плотности K на R^d имеем*

$$\lim_{h \to 1} \int |K_h - K| = 0,$$

где $K_h(x) = h^{-d} K(x/h)$.

Доказательство. Если функция K непрерывна, то утверждение леммы, очевидно, следует из сходимости $K_h \to K$ при любых x и теоремы 2.7 Шеффе.

Если K — произвольная плотность, то для любого $\varepsilon > 0$ можно найти непрерывную плотность $\tilde{K}$, такую, что $\int |K - \tilde{K}| < \varepsilon$. Но так как

$$\int |K_h - K| \leqslant \int |K_h - \tilde{K}_h| + \int |\tilde{K}_h - \tilde{K}| + \int |\tilde{K} - K|,$$

то

$$\limsup_{h \to 1} \int |K_h - K| \leqslant 2\varepsilon + \lim_{h \to 1} \int |\tilde{K}_h - \tilde{K}| = 2\varepsilon,$$

и лемма 1 доказана.

Для произвольной плотности K обозначим

$$\varphi(\delta) = \sup_{1-\delta<h<1+\delta} \int |K_h - K|.$$

При любых $\delta \in (0, 1)$ имеем $0 \leqslant \varphi(\delta) \leqslant 2$. Кроме того, $\lim\limits_{\delta \downarrow 0} \varphi(\delta) = 0$ в силу леммы 1. Далее зависимость f_n от h будет указываться явно — вместо f_n будет использоваться обозначение f_{nh}.

Лемма 2. *Пусть имеется последовательность отрезков* $H_n = [h_n', h_n'']$, *где* $h_n'' \to 0$ *и* $n(h_n')^d \to \infty$. *Тогда для каждого* $\varepsilon > 0$ *существуют положительные числа* n_0 *и* r, *такие, что*

$$P\left(\sup_{h \in H_n} \int |f_{nh} - f| > \varepsilon\right) \leqslant \exp(-rn\varepsilon^2)$$

при любых $n \geqslant n_0$.

Доказательство. В силу сделанных предположений

$$\lim_{n\to\infty} \frac{(h_n'')^d}{n(h_n')^d} = 0,$$

так что

$$1 \leqslant h_n''/h_n' = a_n n^{1/d},$$

где $\lim\limits_{n\to\infty} a_n = 0$. Пусть $\delta_n \geqslant 0$ — решение уравнения

$$(1+\delta_n)^n = a_n n^{1/d}.$$

Ясно, что $\delta_n \to 0$. Далее, положим

$$h_{ni} = (1+\delta_n)^i h_n', \quad i = 0, 1, 2, \ldots, n.$$

Здесь $h_{n0} = h_n'$ и $h_{nn} = h_n''$, так что

$$\sup_{h \in H_n} \int |f_{nh} - f| \leqslant \sup_{1 \leqslant i \leqslant n} \left(\int |f_{nh_{ni-1}} - f| + \right.$$

$$\left. + \sup_{h_{ni-1} \leqslant h \leqslant h_{ni}} \int |f_{nh} - f_{nh_{ni-1}}| \right). \qquad (7)$$

При каждом $u > h$ имеем

$$\int |f_{nh} - f_{nu}| \leqslant \frac{1}{n} \sum_{i=1}^{n} \int |K_h(x - X_i) - K_u(x - X_i)|\, dx =$$

$$= \int \left| \left(\frac{u}{h}\right)^d K\left(\frac{ux}{h}\right) - K(x) \right| dx \leqslant \varphi\left(\frac{u}{h} - 1\right).$$

Таким образом,

$$\sup_{h_{ni-1}\leqslant h\leqslant h_{ni}} \int |f_{nh} - f_{nh_{ni-1}}| \leqslant$$

$$\leqslant \sup_{h_{ni-1}\leqslant h\leqslant h_{ni}} \varphi\left(\frac{h}{h_{ni-1}} - 1\right) = \varphi\left(\frac{h_{ni}}{h_{ni-1}} - 1\right) = \varphi(\delta_n). \quad (8)$$

Для каждого $\varepsilon > 0$ существует число $n_1 > 0$, такое, что $\varphi(\delta_n) < \varepsilon$, $n \geqslant n_1$. Таким образом, ввиду (7) и (8)

$$P\left(\sup_{h\in H_n} \int |f_{nh} - f| > 2\varepsilon\right) \leqslant \sum_{i=1}^{n} P\left(\int |f_{nh_{ni-1}} - f| > \varepsilon\right). \quad (9)$$

Воспользуемся теперь замечанием 3.1. В обозначениях, принятых в замечании 3.1, получим, что каждое слагаемое в правой части (9) ограничено сверху величиной $\exp(-rn\varepsilon^2)$ при $n > n_0$, где r и n_0 — положительные постоянные, если

$$(c_0(\varepsilon)/n)^{1/d} \leqslant h_{ni} \leqslant h_0(\varepsilon)$$

при каждом $i \in \{1, \ldots, n\}$. Но последнее условие выполняется, если $h_n'' < h_0(\varepsilon)$ и $(c_0(\varepsilon)/n)^{1/d} < h_n'$. Следовательно, при всех достаточно больших n правая часть (9) ограничена сверху величиной $n\exp(-rn\varepsilon^2)$, что доказывает лемму 2.

Доказательство теоремы 1. Докажем сходимость вполне. Доказательство сильной и слабой состоятельности получается аналогичным образом при замене слова «вполне» на слова «почти наверное» или «по вероятности». Предположим, что $I_{[h+1/(nh^d)\geqslant\varepsilon]} \to 0$ вполне при каждом $\varepsilon > 0$. Это эквивалентно существованию последовательности $\varepsilon_n' \downarrow 0$, такой, что $I_{[h+1/(nh^d)\geqslant\varepsilon_n']} \to 0$ вполне. Положим теперь

$$\varepsilon_n = \max\left(\varepsilon_n', (1/n)^{1/(d+1)}\right)$$

и убедимся, что $\varepsilon_n \to 0$ и $I_{[h+1/(nh^d)\geqslant\varepsilon_n]} \to 0$ вполне. Далее, определим отрезок $H_n = [h_n', h_n'']$, где

$$h_n' = (n\varepsilon_n)^{-1/d}, \quad h_n'' = \varepsilon_n.$$

В силу определения ε_n имеем $h_n' \leqslant h_n''$. Более того, $h_n'' \to 0$ и $n(h_n')^d \to \infty$. Таким образом, учитывая, что

$$I_{[h+1/(nh^d)>\varepsilon_n]} = \left(I_{[h+1/(nh^d)>h_n'']} + I_{[h+1/(nh^d)>(n(h_n')^d)^{-1}]}\right)/2 \geqslant$$

$$\geqslant \left(I_{[h>h_n'']} + I_{[h\leqslant h_n']}\right)/2 = I_{[h\notin H_n]}/2,$$

убеждаемся, что $I_{[h \notin H_n]} \to 0$ вполне. Из этого замечания, леммы 2 и справедливого при любых $\varepsilon > 0$ неравенства

$$I_{[\int |f_n - f| > \varepsilon]} \leqslant I_{[h \notin H_n]} + I_{\left[\sup\limits_{h \in H_n} \int |f_{nh} - f| > \varepsilon\right]}$$

следует утверждение теоремы 1.

При доказательстве теоремы 1 по существу не потребовалось никаких новых технических приемов по сравнению с теоремой 3.1, что явно свидетельствует о силе и глубине последней. В доказательствах теорем 2 и 3 будет совсем немного новых элементов. До конца доказательства теоремы 3 предполагается, что K — плотность, ограниченная величиной K^*, и ее носитель содержится в замкнутом шаре S_{0c} радиуса c с центром в нуле.

Лемма 3 (сходимость смещения). *Пусть h_n'' — последовательность положительных чисел, стремящаяся к 0 при $n \to \infty$. Для любых плотностей f на R^d имеем*

$$\lim_{n \to \infty} \sup_{0 < h \leqslant h_n''} |f * K_h - f| = 0$$

при почти всех x.

Доказательство. Выражение, стоящее под знаком предела в утверждении леммы, ограничено сверху величиной

$$\sup_{0 \leqslant h \leqslant h_n''} \int\limits_{S_{0,ch}} |f(x-y) - f(x)| K_h(y)\,dy \leqslant$$

$$\leqslant K^* \sup_{0 < h \leqslant h_n''} \left(\int\limits_{S_{0,ch}} \frac{|f(x-y) - f(x)|\,dy}{\lambda(S_{0,ch})} \right) \lambda(S_{0c}),$$

где λ — мера Лебега. По теореме 2.2 Лебега о плотностях последняя величина стремится к 0 при почти всех x.

Лемма 4. *Для любой неотрицательной интегрируемой по Риману функции K, ограниченной величиной K^* на $[0, 1]^d$, и для любого $\varepsilon > 0$ существует целое N и неотрицательные числа $a_i \in [0, K^*]$, $1 \leqslant i \leqslant N^d$, такие, что функция*

$$K_1(x) = \sum_{i=1}^{N^d} a_i I_{A_i}(x), \quad x \in [0, 1]^d,$$

где A_i — прямоугольники, образованные произведениями интервалов вида $[(j-1)/N, j/N)$, $1 \leqslant j < N$, или $[(N-1)/N, 1]$, удовлетворяет условиям

(i) $|K_1(x) - K(x)| < \varepsilon$ *при любых* $x \notin A$,

(ii) $0 \leqslant K_1(x) \leqslant K^*$ *при любых* x,

где A — объединение нескольких множеств A_i, такое, что
(iii) $\lambda(A) < \varepsilon$.

Лемма 5 (основные неравенства для равномерного уклонения). *Пусть $\varepsilon > 0$ — произвольное число, x — точка Лебега для f (см. теорему 2.2) и h_n' и h_n'' — последовательности положительных чисел, такие, что $0 < h_n' \leqslant h_n'' \downarrow 0$. Пусть f_{nh} — оценка (1) с параметром сглаживания h. Тогда*

$$\sup_{h_n' \leqslant h \leqslant h_n''} P\left(|f_{nh}(x) - f * K_h(x)| \geqslant \varepsilon\right) \leqslant 2\exp\left(-bn\left(h_n'\right)^d\right),$$

где в качестве b можно выбрать величину $\varepsilon^2/(2K^*(f(x) + o(1) + \varepsilon))$.
Если ядро K интегрируемо по Риману, то также

$$P\left(\sup_{h_n' \leqslant h \leqslant h_n''} |f_{nh}(x) - f * K_h(x)| \geqslant \varepsilon\right) \leqslant \frac{a\exp\left(-bn\left(h_n'\right)^d\right)}{1 - \exp\left(-b'n\left(h_n'\right)^d\right)}$$

при некоторых положительных постоянных a, b, b', не зависящих от n.

Доказательство. В работе Bennett (1962) показано, что если Z_i — независимые одинаково распределенные случайные величины с нулевыми средними, такие, что $|Z_i| \leqslant t$, то при любых $\varepsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| > \varepsilon\right) \leqslant$$

$$\leqslant 2\exp\left(-\frac{n}{2t}\left(\left(1 + \frac{\sigma^2}{2t\varepsilon}\right)\log\left(1 + \frac{2t\varepsilon}{\sigma^2}\right) - 1\right)\right) \leqslant$$

$$\leqslant 2\exp\left(\frac{-n\varepsilon^2}{2(\sigma^2 + t\varepsilon)}\right),$$

где $\sigma^2 = E(Z_i^2)$. Последнее неравенство следует из того, что $\log(1 + u) \geqslant 2u/(2 + u)$ при любых $u > 0$.

Первое из доказываемых неравенств получается, если подставить в качестве Z_i случайную величину $h^{-d}(K((x - X_i)/h) - E(K((x - X_i)/h)))$, ограниченную по модулю значением $t = K^*/h^d$ и такую, что ее дисперсия σ^2 удовлетворяет неравенству $\sigma^2 \leqslant K^* f * K_h(x)/h^d = K^*(f(x) + o(1))/h^d$ равномерно на $[0, h_n'']$ (в силу леммы 3).

Для доказательства второго неравенства выберем положительное число δ (оно будет конкретизировано далее) и положим $h_{ni} = h_n'(1 + \delta)^i$, $i \geqslant 0$. Пусть номер i_0 таков, что $h_{ni_0 - 1} \leqslant h_n'' \leqslant h_{ni_0}$. Имеем

$$\sup_{h_n' \leqslant h \leqslant h_n''} |f_{nh}(x) - f * K_h(x)| \leqslant$$

$$\leqslant \sup_{0 < i \leqslant i_0} \left[\left|f_{nh_{n,\,i-1}}(x) - f * K_{h_{n,\,i-1}}(x)\right| +\right.$$

$$+ \sup_{h_{n,\,i-1} \leqslant h,\, h' \leqslant h_{ni}} |f * K_h(x) - f * K_{h'}(x)| +$$
$$+ \sup_{h_{n,\,i-1} \leqslant h,\, h' \leqslant h_{ni}} |f_{nh}(x) - f_{nh'}(x)| \Big] =$$
$$= \sup_{0 < i \leqslant i_0} [U_i + V_i + W_i].$$

В силу первого утверждения леммы 5

$$P(U_i \geqslant \varepsilon) \leqslant 2 \exp\left(\frac{-nh_{n,\,i-1}^d \varepsilon^2}{2K^*(f(x) + \varepsilon + o(1))}\right)$$

при $\varepsilon > 0$, где $o(1)$ не зависит от i, так как $0 \leqslant i \leqslant i_0$. По лемме 3

$$\sup_{0 < i \leqslant i_0} V_i \leqslant 2 \sup_{h \leqslant h_{ni_0}} |f * K_h(x) - f(x)| \to 0.$$

Фиксируем $\varepsilon > 0$ и найдем функцию K_1, числа N, a_1, ..., $a_N{}^d$ и множества A_i, указанные в лемме 4 (заменяя там $[0, 1]^d$ на $[-c, c]^d$). Пусть также через A обозначается то же множество, что в лемме 4. Обозначим μ и μ_n меру, индуцированную плотностью f, и эмпирическую меру для X_1, ..., X_n соответственно. Пусть также Δ — оператор симметрической разности множеств.

Не нарушая общности, предположим, что все множества A_i содержатся строго внутри одного ортанта, например внутри $[0, c]^d$. Приведем теперь несколько геометрических фактов. Пусть числа h, h' принадлежат отрезку $[h_{n,\,i-1}, h_{ni}]$ и A_j — фиксированное множество, например $A_j = [a_1, a_1'] \times \ldots \times [a_d, a_d']$. Тогда $(x + hA_j) \Delta (x + h'A_j) \subseteq (x + h_{ni}B_j)$, где B_j — множество фиксированной формы, размеры которого зависят только от A_j, d и δ. Кроме того, $\lambda(B_j) \leqslant 2c^d\delta$.

Для доказательства этого первого геометрического факта достаточно только показать, что $uA_j \Delta u'A_j \subseteq B_j$ при любых u, $u' \in [1/(1+\delta), 1]$. Положим

$$B_j = [a_1, a_1'] \times \cdots \times [a_d, a_d'] - \left[a_1, \frac{a_1'}{1+\delta}\right] \times \cdots \times \left[a_d, \frac{a_d'}{1+\delta}\right] +$$
$$+ \left[\frac{a_1}{1+\delta}, a_1'\right] \times \cdots \times \left[\frac{a_d}{1+\delta}, a_d'\right] - [a_1, a_1'] \times \cdots \times [a_d, a_d'],$$

где операторы «—» рассматриваются прежде, чем оператор объединения «+». Заметим также, что B_j содержится в $[-c, c]^d$. Кроме того,

$$\lambda(B_j) \leqslant \left(a_1'\left(1 - \frac{1}{1+\delta}\right) + a_1\left(1 - \frac{1}{1+\delta}\right)\right) a_2' a_3' \ldots a_d' \leqslant$$
$$\leqslant 2a_1' a_2' \ldots a_d'\delta \leqslant 2c^d\delta.$$

Второй геометрический факт состоит в следующем. Пусть A — множество из леммы 4, т. е. объединение M непересекающихся пря-

моугольников A_j, и пусть $B = \bigcup_{u \in [1/(1+\delta),\ 1]} uA$. Тогда $B \subseteq$ $\subseteq [-c, c]^d$ и в силу проведенной выше выкладки для одного прямоугольника $\lambda(B) \leqslant \lambda(A) + \sum_{j=1}^{M} \lambda(B_j) \leqslant \lambda(A) + 2Mc^d\delta$ [1]). Теперь мы в состоянии получить следующую основную верхнюю границу для W_i:

$$W_i \leqslant \sup_{h_{n,\,i-1} \leqslant h,\ h' \leqslant h_{ni}} \sum_{j=1}^{N^d} \int a_j \left| I_{x+hA_j}(y) - I_{x+h'A_j}(y) \right| \mu_n(dy)/h_{n,\,i-1}^d +$$

$$+ 2 \sup_{h_{n,\,i-1} \leqslant h \leqslant h_{ni}} \sum_{j=1}^{N^d} \int_{x+hA_j} \left| K\left(\frac{x-y}{h}\right) - K_1\left(\frac{x-y}{h}\right) \right| \frac{\mu_n(dy)}{h_{n,\,i-1}^d} \leqslant$$

$$\leqslant \sup_{h_{n,\,i-1} \leqslant h,\ h' \leqslant h_{ni}} \sum_{j=1}^{N^d} a_j \frac{\mu_n((x+hA_j)\,\Delta\,(x+h'A_j))}{h_{n,\,i-1}^d} +$$

$$+ \frac{2\varepsilon\mu_n(x + [-c, c]^d h_{ni})}{h_{n,\,i-1}^d} + \frac{2K^*\mu_n(x + h_{ni}B)}{h_{n,\,i-1}^d} \leqslant$$

$$\leqslant h_{n,\,i-1}^{-d} \left(\sum_{j=1}^{N^d} K^*\mu_n(x + h_{ni}B_j) + \right.$$

$$\left. + 2\varepsilon\mu_n(x + [-c, c]^d h_{ni}) + 2K^*\mu_n(x + h_{ni}B) \right) =$$

$$= \sum_{j=1}^{N^d} W_{ij} + W_i' + W_i''.$$

Для заданного $\eta > 0$ выберем ε, $\delta > 0$, такие, что математическое ожидание каждой из величин W_{ij} не превосходит $\eta/3N^d$ и математические ожидания W_i' и W_i'' не превосходят $\eta/3$. Это эквивалентно тому, что

$$f(x) + o(1))(1+\delta)^d K^*(2c^d\delta) < \eta/(3N^d),$$

$$(f(x) + o(1))(1+\delta)^d (2\varepsilon)(2c)^d < \eta/3,$$

$$(f(x) + c(1))(1+\delta)^d (2K^*)(\varepsilon + 2Mc^d\delta) < \eta/3.$$

Слагаемые $o(1)$ снова не зависят от i, так что все три неравенства выполняются при всех достаточно больших n равномерно по i. Отметим здесь одну небольшую техническую деталь: по-видимому,

---

[1]) Так как $uA = \bigcup_{j=1}^{M} uA_j \subseteq \bigcup_{j=1}^{M} (A_j \cup (uA_j\,\Delta A_j)) \subseteq A \cup \bigcup_{j=1}^{M} B_j$ при любом $u \in [1/(1+\delta), 1]$. — *Прим. перев.*

сначала нужно выбирать ε, считая, что $\delta \leqslant 2$. При этом фиксируются значения N и M, так что на следующем шаге можно выбрать δ.

При каждом i легко получим

$$P(W_i > 2\eta) \leqslant \sum_{j=1}^{N^d} P\left(W_{ij} - E(W_{ij}) > \frac{\eta}{3N^d}\right) +$$

$$+ P\left(W_i' - E(W_i') > \frac{\eta}{3}\right) + P\left(W_i'' - E(W_i') > \frac{\eta}{3}\right). \qquad (10)$$

Как показано выше, все математические ожидания здесь меньше, чем $\eta/3$, равномерно по i и j при любых $n \geqslant n_0$. Кроме того, каждую из величин W_{ij}, W_i' и W_i'' можно записать в виде $(1/n)\sum_{m=1}^{n} Y_m$, где Y_m — независимые ограниченные неотрицательные случайные величины, не превосходящие по абсолютной величине r/h_{ni}^d, где

$$r = \max(2K^*, 2\varepsilon)(1+\delta)^d.$$

Таким образом, применяя еще раз неравенство Беннета (Bennett (1962)), убеждаемся, что каждая из вероятностей в правой части (10) не превосходит

$$2\exp\left(-\frac{n(\eta/3N^d)^2}{2((\eta/3)(r/h_{ni}^d) + (r/h_{ni}^d)(\eta/3N^d)}\right) \triangleq 2\exp(-bnh_{ni}^d).$$

Объединяя все полученные выше границы, мы видим, что при любых n, бо́льших некоторого n_1,

$$P\left(\sup_{h_n' \leqslant h \leqslant h_n''} |f_{nh}(x) - f * K_h(x)| > 4\eta\right) \leqslant$$

$$\leqslant \sum_{i=1}^{i_0} 2\exp(-sn(h_n')^d(1+\delta)^{d(i-1)}) +$$

$$+ (N^d+2)\,2\exp(-bn(h_n')^d(1+\delta)^{di}), \qquad (11)$$

где $s = \eta^2/4K^*(f(x)+\eta)$. Правая часть (11) в свою очередь ограничена сверху (хотя и очень грубо) величиной

$$\sum_{i=0}^{\infty} b'\exp(-b''n(h_n')^d(1+\delta)^i) \leqslant$$

$$\leqslant \sum_{i=0}^{\infty} b'\exp(-b''n(h_n')^d(1+\delta i)) = \frac{b'\exp(-b''n(h_n')^d)}{1-\exp(-b''\delta n(h_n')^d)},$$

где b', b'' — некоторые положительные постоянные. Лемма 5 доказана.

Лемма 6 (неравенство на хвосты биномиального распределения). *Пусть случайная величина Z имеет биномиальное распределение с параметрами (n, p), где $p = p(n) \in (0, 1)$ зависит от n так, что $p + np^2 = o(1)$, но $\lim\limits_{n\to\infty} np = \infty$. Тогда при $\delta > 0$ имеем*

$$P(Z - np \geqslant \delta np) \geqslant \frac{1 + o(1)}{(2\pi(1+\delta)^3 np)^{1/2}} \exp(-npH(\delta)),$$

где $0 < H(\delta) = (1+\delta)\log(1+\delta) - \delta \to 0$ при $\delta \downarrow 0$.

Доказательство. Пусть $k = \lceil np(1+\delta) \rceil$. Тогда

$$P(Z - np \geqslant \delta np) \geqslant \binom{n}{k} p^k (1-p)^{n-k} \geqslant$$

$$\geqslant \frac{(n-k+1)^k}{k!} p^k (1-p)^n e^{pk} \geqslant$$

$$\geqslant \frac{(np)^k}{k!} (1-p)^n e^{pk} (1 - k^2/n).$$

Так как $k^2 = o(n)$, $pk = o(1)$ и $k! \sim (k/e)^k \sqrt{2\pi k}$, то нижняя граница равна

$$(1+o(1)) \left(\frac{npe}{k}\right)^k \frac{e^{-np}}{\sqrt{2\pi k}} = (1+o(1)) \frac{e^{k-np}(1+\delta)^{-k}}{\sqrt{2\pi k}} \geqslant$$

$$\geqslant (1+o(1)) \frac{e^{\delta np - k\log(1+\delta)}}{\sqrt{2\pi k}} \geqslant$$

$$\geqslant (1+o(1)) \frac{e^{-npH(\delta)}}{(1+\delta)\sqrt{2\pi k}},$$

откуда следует искомое неравенство.

Лемма 7 (экспоненциальные нижние границы для больших уклонений). *Пусть f — произвольная плотность на R^d и x — точка Лебега плотности f, такая, что $f(x) > 0$. Пусть $\varepsilon > 0$ — некоторая постоянная и $h = h_n$ — последовательность положительных чисел, удовлетворяющая условиям $h + nh^{2d} = o(1)$, $\lim\limits_{n\to\infty} nh^d = \infty$. Пусть, далее, $H(\cdot)$ — функция, определенная в лемме 6, и $\delta = 2\varepsilon/f(x)$.*

Тогда для ядерной оценки (1) выполнено неравенство

$$P(f_n(x) - f * K_h(x) \geqslant \varepsilon) \geqslant \frac{1+o(1)}{(2\pi nh^d f(x)(2c)^d (1+\delta)^3)^{1/2}} \times$$

$$\times \exp(-nh^d H(\delta)(f(x) + o(1))(2c)^d).$$

Доказательство. Обозначим через Y случайный вектор, являющийся сужением X на $x + [-c, c]^d h$. Положим

$$g_n(x) = \frac{1}{n} \sum_{i=1}^{n} h^{-d} K\left(\frac{x - Y_i}{h}\right),$$

где случайные векторы $Y_1, Y_2, \ldots$ независимы и имеют то же распределение, что Y. Ясно, что $f_n(x)$ совпадает по распределению с $(N/n)\, g_N(x)$, где случайная величина N не зависит от векторов Y_i и распределена как количество X_i, попавших в $x + [-c, c]^d h$. Кроме того, $E(f_n(x)) = pE(g_n(x))$, где $p = P(X_1 \in A = x + [-c, c]^d h) = (2ch)^d (f(x) + o(1))$. При всех достаточно больших n имеет место следующее неравенство:

$$P(f_n(x) \geqslant E(f_n(x)) + \varepsilon) \geqslant P(N \geqslant np(1+\delta)) \times \inf_{k \geqslant np(1+\delta)} P\left(g_k(x) \geqslant E(g_k(x)) - \frac{\varepsilon}{2p(1+\delta)}\right). \quad (12)$$

В самом деле, на достаточно богатом вероятностном пространстве можно считать, что $f_n(x)$ равно $(N/n)\, g_N(x)$, а $Y_1, \ldots, Y_N$ — часть множества $X_1, \ldots, X_n$, попавшая в A. Если $N \geqslant np(1+\delta)$ и $g_N(x) \geqslant E(g_N(x)) - \varepsilon/2p(1+\delta)$, то

$$f_n(x) = \frac{N}{n} g_N(x) \geqslant \frac{np(1+\delta)}{n}\left(E(g_N(x)) - \frac{\varepsilon}{2p(1+\delta)}\right) =$$

$$= p(1+\delta)\left(\frac{E(f_n(x))}{p} - \frac{\varepsilon}{2p(1+\delta)}\right) =$$

$$= E(f_n(x)) + \delta E(f_n(x)) - \frac{\varepsilon}{2} \geqslant E(f_n(x)) + \varepsilon$$

при достаточно больших n. Отсюда следует (12). Используя неравенство Чебышёва и то, что $\operatorname{Var}(g_k(x)) \leqslant K^*(f(x) + o(1))/kh^d p$, получим, что правая часть (12) больше или равна

$$P(N - np \geqslant \delta np) \inf_{k \geqslant np(1+\delta)} \left(1 - \left(\frac{2p(1+\delta)}{\varepsilon}\right)^2 \operatorname{Var}(g_k(x))\right) \geqslant$$

$$\geqslant P(N - np \geqslant \delta np)\left(1 - \frac{K^*(f(x) + o(1))(2p)^2(1+\delta)^2}{np(1+\delta)\varepsilon^2 h^d p}\right) =$$

$$= P(N - np \geqslant \delta np)(1 - o(1)). \quad (13)$$

К последнему выражению можно применить лемму 6, поскольку случайная величина N имеет биномиальное распределение с параметрами (n, p), где $p + np^2 = o(1)$ и $\lim_{n \to \infty} np = \infty$. Лемма 7 доказана.

В дальнейшем очень важную роль играет величина

$$D_n(x) = \sup_{H_n} |f_{nh}(x) - f(x)|,$$

где верхняя грань берется по всем h из отрезка $H_n = [h_n', h_n'']$ и величины $0 < h_n' \leqslant h_n'' < \infty$ зависят только от n.

Лемма 8. *Пусть K — ограниченная интегрируемая по Риману плотность с компактным носителем и $h_n'' = o(1)$. Тогда для каждой плотности f на R^1 выполнены следующие утверждения.*

A. *Если $n(h_n')^d \to \infty$, то $D_n(x) \to 0$ по вероятности при почти всех x.*

B. *Если h_n' — правильно меняющаяся последовательность с коэффициентом $r \leqslant 0$ (т. е. $h'_{[tn]}/h_n' \to t^r$ при любых $t > 0$) и $n(h_n')^d/\log\log n \to \infty$, то $D_n(x) \to 0$ почти наверное при почти всех x.*

C. *Если $n(h_n')^d/\log n \to \infty$, то $D_n(x) \to 0$ вполне при почти всех x.*

Доказательство. Утверждения A и C следуют непосредственно из лемм 3 и 5 и тривиального неравенства

$$D_n(x) \leqslant \sup_{H_n} |f_{nh}(x) - f * K_h(x)| + \sup_{H_n} |f * K_h(x) - f(x)|.$$

Докажем утверждение B. Фиксируем малое $\delta > 0$ и введем последовательность $n_i = [(1+\delta)^i]$, $i = 0, 1, 2, \ldots$. Положим

$$E_i = \sup_{n_i \leqslant n < n_{i+1}} \sup_{h \in H_i^*} |f_{nh}(x) - f * K_h(x)|, \tag{14}$$

где

$$H_i^* = \left[\inf_{n_i \leqslant n < n_{i+1}} h_n', \sup_{n_i \leqslant n < n_{i+1}} h_n''\right] = [h_i^*, h_i^{**}].$$

В силу леммы 3 ясно, что

$$\sup_{n_i \leqslant n < n_{i+1}} D_n(x) \leqslant E_i + o(1), \quad i \to \infty$$

в любых точках Лебега x. Таким образом, в силу леммы Бореля — Кантелли для доказательства сходимости $D_n(x) \to 0$ почти наверное при почти всех x достаточно показать, что

$$\sum_{i=0}^{\infty} P(E_i > \varepsilon) < \infty$$

для любых точек Лебега, любых $\varepsilon > 0$ и некоторого $\delta(\varepsilon) > 0$. При фиксированном h и любых $n_i \leqslant n < n_{i+1}$ легко получить границы

$$|f_{nh} - f * K_h| \leqslant |f_{nh} - f_{n_i h}| + |f_{n_i h} - f * K_h| \leqslant$$

$$\leqslant \left(\frac{1}{n_i} - \frac{1}{n_{i+1}}\right) \sum_{j=1}^{n_i} K_h(x - X_j) + \frac{1}{n_i} \sum_{j=n_i+1}^{n_{i+1}} K_h(x - X_j) +$$

$$+ |f_{n_i h} - f * K_h| \leqslant$$

$$\leqslant (\delta + o(1))\,(f_{n_i h} + \bar{f}_{n_{i+1}-n_i h}) + |f_{n_i h} - f * K_h| \leqslant$$
$$\leqslant (1 + \delta + o(1))\,|f_{n_i h} - f * K_h| +$$
$$+ (\delta + o(1))\,|\bar{f}_{n_{i+1}-n_i,\ h} - f * K_h| +$$
$$+ (\delta + o(1))\,2f * K_h. \tag{15}$$

Здесь f_{nh} — оценка, не зависящая от f_{nh}, но распределенная так же, как f_{nh}. Ясно, что E_i не превосходит верхней грани по $h \in H_i^*$ правой части (15). Так как $h_i^{**} \to 0$ при $i \to \infty$, то последнее слагаемое полученной таким образом верхней границы для E_i равно $2\delta f(x) + o(1)$ в силу леммы 3. Теперь фиксируем $\varepsilon > 0$, выберем δ достаточно малым, чтобы выполнялись неравенства $\delta \leqslant 1/2$, $2\delta f(x) < \varepsilon/4$, и i достаточно большим, чтобы все члены $o(1)$ в (15) не превосходили $1/2$ и член $o(1)$ в слагаемом $2\delta f(x) + o(1)$ не превосходил $\varepsilon/12$ (при этом данное слагаемое не превосходит $\varepsilon/3$). При таких больших i имеем

$$E_i \leqslant 2 \sup_{h \in H_i^*} |f_{n_i,\ h} - f * K_h| + \sup_{h \in H_i^*} |\bar{f}_{n_{i+1}-n_i,\ h} - f * K_h| + \frac{\varepsilon}{3}. \tag{16}$$

В силу леммы 5 существуют положительные постоянные a, a', a'', b, b', b'', такие, что вероятности того, что первое и второе слагаемые в правой части (16) больше, чем $\varepsilon/3$, не превосходят

$$\frac{a \exp(-a' n_i (h_i^*)^d)}{1 - \exp(-a'' n_i (h_i^*)^d)} \tag{17}$$

и

$$\frac{b \exp(-b'(n_{i+1} - n_i)(h_i^*)^d)}{1 - \exp(-b''(n_{i+1} - n_i)(h_i^*)^d)}$$

соответственно. Постоянные не зависят от i.

Для любого $M > 0$ найдется достаточно большое i, такое, что $n_j (h_j^*)^d \geqslant M \log\log n_j \geqslant M \log(j \log(1 + \delta))$ при $j \geqslant i$. Границы (17) при $j \geqslant i$ меньше, чем величины

$$\frac{a + o(1)}{(j \log(1 + \delta))^{Ma'}}, \quad \frac{b + o(1)}{(j \log(1 + \delta))^{Mb'\delta}} \tag{18}$$

соответственно. Но оба выражения в (18) суммируемы по j при $Ma' > 1$ и $Mb'\delta > 1$. Следовательно, можно найти такое $\delta(\varepsilon) > 0$, что

$$\sum_{i=0}^{\infty} P(E_i > \varepsilon) < \infty$$

для любых $\varepsilon > 0$ и любых точек Лебега плотности f. Лемма 8 доказана.

Доказательство теоремы 2 основано на неравенстве

$$|f_n(x) - f(x)| \leqslant \sup_{H_n} |f_{nh}(x) - f(x)| + \infty \cdot I_{[h_n \notin H_n]}, \qquad (19)$$

где I — индикаторная функция событий и $\infty \cdot 0$ равно 0. Из утверждений A и B о поточечной сходимости следуют утверждения об интегральной сходимости, если заметить, что f_n является плотностью на R^d при каждом n и что можно применить слабое и сильное обобщения теоремы Шеффе (теорема 2.8).

Доказательства утверждений о поточечной сходимости получаются, если построить соответствующие последовательности $H_n = [h_n', h_n'')$. При этом используются возрастающие последовательности целых чисел n_k' и n_k'' соответственно. Во всех случаях (A, B и C) считается, что $n_1' = n_1'' = 1$. Кроме того, $h_n'' = 1/k$ на $[n_k'', n_{k+1}'')$, так что $h_n'' \to 0$ при $k \to \infty$. Наконец, на $[n_1', n_2')$ и $[n_1'', n_2'')$ величины h_n' и h_n'' определяются произвольным образом.

Утверждение A. Пусть

$$n_k'' = \inf\left(n : n > n_{k-1}'', \sup_{m \geqslant n} P\left(h_m \geqslant \frac{1}{k}\right) \leqslant \frac{1}{k}\right), \quad k \geqslant 2,$$

$$n_k' = \inf\left(n : n > n_{k-1}', \sup_{m \geqslant n} P\left(mh_m^d \leqslant k\right) \leqslant \frac{1}{k}\right), \quad k \geqslant 2,$$

$$h_n' = (k/n)^{1/d} \text{ на } [n_k', n_{k+1}'), \quad k \geqslant 2.$$

Ясно, что $n(h_n')^d \to \infty$. Кроме того, на $[n_k'', n_{k+1}'')$ имеем $P(h_n \geqslant h_n'') = P(h_n \geqslant 1/k) \leqslant 1/k \to 0$ при $k \to \infty$. Точно так же на $[n_k', n_{k+1}')$ имеем $P\left(nh_n^d \leqslant n(h_n')^d\right) = P\left(nh_n^d \leqslant k\right) \leqslant 1/k$ и последняя величина стремится к 0 при $k \to \infty$. Применяя теперь (19) и лемму 8, получим утверждение A.

Утверждение C. Пусть

$$n_k'' = \inf\left(n : n > n_{k-1}'', \sum_{m \geqslant n} m^k P\left(h_m \geqslant \frac{1}{k}\right) \leqslant 2^{-k}\right), \quad k \geqslant 2,$$

$$n_k' = \inf\left(n : n > n_{k-1}', \sum_{m \geqslant n} m^k P\left(\frac{mh_m^d}{\log m} \leqslant k\right) \leqslant 2^{-k}\right), \quad k \geqslant 2,$$

$$h_n' = \left(k \frac{\log n}{n}\right)^{1/d} \text{ на } [n_k', n_{k+1}'), \quad k \geqslant 2.$$

Ясно, что $n\,(h_n')^d/\log n \to \infty$. Кроме того,

$$\sum_{n=1}^{\infty} P\,(h_n \geqslant h_n'') \leqslant n_2'' + \sum_{k\geqslant 2} \sum_{n=n_k''}^{n_{k+1}''-1} n^k P\left(h_n \geqslant \frac{1}{k}\right) \leqslant$$

$$\leqslant n_2'' + \sum_{k\geqslant 2} 2^{-k} < \infty.$$

Из таких же в точности рассуждений следует, что

$$\sum_{n=1}^{\infty} P\,(h_n \leqslant h_n') \leqslant n_2' + \sum_{k\geqslant 2} 2^{-k} < \infty.$$

Таким образом, $\sum P\,(h_n \notin H_n) < \infty$, и, следовательно, с учетом леммы 8 получим, что правая часть (19) стремится к 0 вполне.

Утверждение В. Пусть

$$n_k'' = \inf\left(n : n > n_{k-1}'',\ P\left(\bigcup_{m\geqslant n}\left[\frac{m h_m^d}{\log\log m} \leqslant k\right]\right) \leqslant 2^{-k}\right), \quad k \geqslant 2,$$

$$n_k' = \inf\left(n : n > n_{k-1}',\ P\left(\bigcup_{m\geqslant n}\left[h_m \geqslant \frac{1}{k}\right]\right) \leqslant 2^{-k}\right), \quad k \geqslant 2,$$

$$h_n' = (k(\log\log n)/n)^{1/d} \text{ на } [n_k', n_{k+1}'), \quad k \geqslant 2.$$

Можно проверить, что $n\,(h_n')^d/\log\log n \to \infty$ и что $h_n \geqslant h_n''$ почти наверное конечное число раз, поскольку на $[n_k'', n_{k+1}'')$

$$P\left(\bigcup_{m\geqslant n}[h_m \geqslant h_m'']\right) \leqslant \sum_{j=k}^{\infty} P\left(\bigcup_{m=n_j''}^{n_{j+1}''}[h_m \geqslant 1/j]\right) \leqslant \sum_{j=k}^{\infty} 2^{-j} = 2^{-k+1} \to 0$$

при $k \to \infty$.

Аналогичным образом можно проверить, что $h_n \leqslant h_n'$ почти наверное конечное число раз. Утверждение В будет доказано, если найдется *правильно меняющаяся* последовательность положительных чисел $h_n^* \leqslant h_n'$, такая, что $n\,(h_n^*)^d/\log\log n \to \infty$. Тогда доказательство завершается применением леммы 8 и (19). Последовательность $\varphi\,(n) = n\,(h_n')^d/\log\log n$ по построению монотонно не убывает и стремится к ∞. Определим функцию $\varphi\,(t)$ на действительной оси посредством линейной интерполяции значений $\varphi\,(n)$. Попытаемся найти такую функцию $\psi\,(t)$, что $0 \leqslant \psi \leqslant \varphi$, $\psi\,(t) \uparrow \infty$ при $t \uparrow \infty$ и $t\psi'\,(t)/\psi\,(t) \to 0$ при $t \to \infty$. Эта функция ψ является медленно меняющейся (Seneta (1976) с. 10 по русскому изданию). Затем положим $h_n^* = (\psi\,(n)\,(\log\log n)/n)^{1/d}$ и заметим, что эта последовательность удовлетворяет всем указанным выше требованиям.

В качестве ψ выберем непрерывную кусочно-линейную функцию с узлами в точках $t_1 < t_2 < \dots$, где $t_k \to \infty$. Положим $t_1 = 1$ и $\psi(t) = \varphi(t)$ на $[0, 1]$. При заданных t_k и $\psi(t_k)$ определим t_{k+1} и $\psi(t_{k+1})$ следующим образом:

$$\psi(t_{k+1}) = \min\left(\varphi(t_k),\ \psi(t_k)\left(1 + \frac{1}{2\log k}\right)\right),$$

$$t_{k+1} = \inf\{t : t \geqslant t_k + 1,\ t/t_k \geqslant \psi(t_{k+1})/\psi(t_k),$$

$$t - t_k \geqslant \frac{(\psi(t_{k+1}) - \psi(t_k))\, t \log k}{\psi(t_{k+1})}\Big).$$

Заметим, что $t_k \geqslant k \to \infty$ при $k \to \infty$, функция $\psi(t)/t$ монотонно не возрастает и что

$$\psi'(t) \leqslant \frac{\psi(t_{k+1})}{t_{k+1}\log k} \leqslant \frac{\psi(t)}{t \log k}$$

на $[t_k, t_{k+1})$. Существование t_{k+1} следует из того, что всегда найдется $t \geqslant t_k + 1$, такое, что

$$t \geqslant \frac{t_k}{1 - (\log k)(1 - \psi(t_k)/\psi(t_{k+1}))}.$$

В самом деле, знаменатель последнего выражения всегда больше или равен $1/2$ (иными словами, данное условие всегда выполнено при $t \geqslant 2t_k$). Наконец, $0 \leqslant \psi \leqslant \varphi$ и $\psi(t) \uparrow \infty$, поскольку $\varphi(t) \to \infty$ и

$$\prod_{k=2}^{\infty}\left(1 + \frac{1}{2\log k}\right) = \infty.$$

Доказательство теоремы 3. Для доказательства утверждения 1 не требуется ничего нового. Эквивалентность C, D и E установлена в теореме 3.1. Очевидно, $C \Rightarrow B \Rightarrow A$ в силу лемм 3 и 5. Наконец, $A \Rightarrow D$ в силу обобщения теоремы Шеффе, принадлежащего Глику (теорема 2.8).

Утверждение 3 частично доказано в леммах 3 и 5 (имеются в виду включения $C \Rightarrow B \Rightarrow A$). Для доказательства того, что $A \Rightarrow C$, заметим, что необходимость условия $h_n = o(1)$ следует из утверждения 1 и необходимость условия $nh^d/\log n \to \infty$ следует из леммы 7. В самом деле, так как можно ограничиться точками Лебега функции f, в которых $f(x) > 0$, то, в силу леммы 7, из условий $h_n = o(1)$, $nh_n^d \to \infty$ (оба этих условия вытекают из утверждения 1 данной теоремы) и

$$\sum_{n=1}^{\infty} P(|f_n(x) - E(f_n(x))| > \varepsilon) < \infty,$$

при любых $\varepsilon > 0$ и почти всех x следует

$$\sum_{n=1}^{\infty} \min\left(1,\ (nh_n^d)^{-1/2} \exp\left(-anh_n^d\right)\right) < \infty \qquad (20)$$

при любых $a > 0$. Если величина $nh_n^d/\log n$ не превосходит M, то сумма в (20) больше или равна

$$\sum_{n > e^{1/M}}^{\infty} (M \log n)^{-1/2} n^{-aM}.$$

Последняя сумма бесконечна при $a \leqslant 1/M$. Если же последовательность $nh_n^d/\log n$ не ограничена, то в силу ее семимонотонности имеем $\lim\limits_{n\to\infty} (nh_n^d/\log n) = \infty$. Следовательно, A $\Rightarrow$ C.

Единственной нетривиальной частью теоремы является утверждение 2. Очевидно, B $\Rightarrow$ A. Кроме того, если ядро K интегрируемо по Риману, то C $\Rightarrow$ B в силу теоремы 2. Покажем теперь, что для доказательства импликации A $\Rightarrow$ C достаточно применить лемму 7. Фиксируем постоянную $a > 0$ и введем последовательность $n_i = [\exp(ai \log i)]$, $i \geqslant 1$. Заметим, что $(n_{i+1} - n_i)/n_i \sim (ei)^a$. Пусть уже доказано, что если $nh_n^d/\log\log n \leqslant M < \infty$, $h_n \to 0$, $nh_n^d \to \infty$ и x — точка Лебега функции f, такая, что $f(x) > 0$, то

$$P\left(|f_{n_i}(x) - E\left(f_{n_i}(x)\right)| > \varepsilon \text{ бесконечно часто}\right) = 1 \qquad (21)$$

при достаточно малых ε. Тогда в силу семимонотонности последовательности $nh_n^d/\log\log n$ имеем $\lim\limits_{n\to\infty} (nh_n^d/\log\log n) = \infty$, так как в противном случае получается противоречие. Необходимость условий $h_n = o(1)$ и $\lim\limits_{n\to\infty} nh_n^d = \infty$ следует из утверждения 1 данной теоремы. Таким образом, остается доказать (21) при сформулированных выше условиях. Имеем

$$\left[|f_{n_i}(x) - E\left(f_{n_i}(x)\right)| > \varepsilon \text{ б. ч.}\right] \supseteq$$

$$\supseteq \left[|\bar{f}_i(x) - E\left(\bar{f}_i(x)\right)| > 2\varepsilon \text{ б. ч.}\right] \cap$$

$$\cap \left[\frac{n_i}{n_{i+1}} |f_i^*(x) - E\left(f_i^*(x)\right)| > \varepsilon \text{ конечное число раз}\right], \qquad (22)$$

где

$$\bar{f}_i(x) = (n_{i+1} - n_i)^{-1} \sum_{j=n_i+1}^{n_{i+1}} \frac{K\left((x - X_j)/h_{n_{i+1}}\right)}{h_{n_{i+1}}^d},$$

$$f_i^*(x) = n_i^{-1} \sum_{j=1}^{n_i} \frac{K\left((x - X_j)/h_{n_{i+1}}\right)}{h_{n_{i+1}}^d}.$$

Включение (22) следует из неравенства

$$|\hat{f}_{n_{i+1}}(x) - E(\hat{f}_{n_{i+1}}(x))| \geqslant \frac{n_{i+1}-n_i}{n_{i+1}} |\hat{f}_i(x) - E(\hat{f}_i(x))| - \frac{n_i}{n_{i+1}} |\hat{f}_i^*(x) - E(\hat{f}_i^*(x))|.$$

По лемме 5

$$P\left(\frac{n_i}{n_{i+1}} |\hat{f}_i^*(x) - E(\hat{f}_i^*(x))| > \varepsilon\right) \leqslant 2\exp\left(-n_i h_{n_{i+1}}^d \frac{\varepsilon^2 (n_{i+1}/n_i)^2}{2K^*(f(x)+\varepsilon+o(1))}\right) = 2\exp\left(-n_{i+1} h_{n_{i+1}}^d \frac{(\varepsilon^2+o(1))(ei)^a}{2K^*(f(x)+\varepsilon+o(1))}\right)$$

Так как $nh_n^d \to \infty$, то эта последовательность суммируема по i при любых a, $\varepsilon > 0$ и по лемме Бореля — Кантелли последнее событие в (22) имеет вероятность 1. В силу независимости своих элементарных составляющих среднее событие в (22) происходит с вероятностью 1 тогда и только тогда, когда

$$\sum_{i=1}^{\infty} P(|\hat{f}_i(x) - E(\hat{f}_i(x))| > 2\varepsilon) = \infty. \tag{23}$$

Нижняя граница для i-й вероятности в (23) дается леммой 7 при замене там n и h на $n_{i+1} - n_i$ и $h_{n_{i+1}}$ соответственно. В силу сделанных предположений имеем $h_{n_{i+1}} = o(1)$, $(n_{i+1} - n_i) h_{n_{i+1}}^{2d} = o(1)$ и $(n_{i+1} - n_i) h_{n_{i+1}}^d \to \infty$, так что действительно можно воспользоваться леммой 7. Нижняя граница для i-го слагаемого при достаточно больших i равна

$$c_1 \left(n_{i+1} h_{n_{i+1}}^d\right)^{-1/2} \exp\left(-c_2 n_{i+1} h_{n_{i+1}}^d\right), \tag{24}$$

где при любых $\varepsilon > 0$ постоянные c_1, c_2 положительны и таковы, что $\liminf\limits_{\varepsilon \downarrow 0} c_1 > 0$, $\liminf\limits_{\varepsilon \downarrow 0} c_2 = 0$. Ясно, что величина (24) больше или равна

$$c_1 (M \log\log n_{i+1})^{-1/2} \exp(-c_2 M \log\log n_{i+1}) \sim c_1 (M \log i)^{-1/2} (ai \log i)^{-c_2 M},$$

так что при $c_2 < 1/M$ (т. е. при достаточно малых ε) хвосты суммы бесконечны. Тем самым доказаны соотношения (23), (21) и теорема 3.

Доказательство теоремы 4. Пусть T — компактный носитель f, M — граница для значений K, $K = 0$ вне $S_{0,r}$ и $K \geqslant cI_{S_{0,r}}$. Пусть A_n — множество всех h_n, для которых $L(h_n) > a \sup\limits_{h>0} L(h)$.

Имеются некоторые сложности с измеримым выбором h_n из A_n. Их можно обойти несколькими способами. Здесь предполагается, что процесс выбора построен так, что h_n — случайная величина. Через F обозначается функция распределения, соответствующая плотности f, через F_n — эмпирическая функция распределения выборки $X_1, \ldots, X_n$.

Доказательство теоремы 4 основано на ряде важных лемм. Они выделены для удобства чтения.

Лемма 9 (неравенства больших уклонений для распределения Пуассона). *Если случайная величина X имеет распределение Пуассона с параметром λ, то*

$$P(|X-\lambda| \geqslant \lambda\varepsilon) \leqslant 2\exp(-\lambda\varepsilon^2/2(1+\varepsilon))$$

при любых $\varepsilon > 0$.

Доказательство. См. соотношение (3.3) и его доказательство.

Лемма 10. *Пусть $\{g_\theta\}$ — семейство функций из R^d в R^1, зависящих от параметра θ, и пусть f — плотность с компактным носителем. Тогда*

$$\sup_\theta \left| \int g_\theta \, dF_n - \int g_\theta \, dF \right| \to 0$$

почти наверное при следующих условиях:

(i) $\sup\limits_\theta \sup\limits_x |g_\theta(x)| < \infty$ (*семейство $\{g_\theta\}$ равномерно ограничено*),

(ii) *семейство $\{g_\theta\}$ равностепенно непрерывно.*

Доказательство. Пусть $\varepsilon > 0$ — произвольное число. Разобьем T на прямоугольники R_i, $1 \leqslant i \leqslant N$, такие, что $\sup\limits_i \sup\limits_{x,\, y \in R_i} \|x - y\| \leqslant \delta$, где $\delta > 0$ выбрано столь малым, что $|g_\theta(x) - g_\theta(y)| < \varepsilon$ при любых θ, $\|x-y\| < \delta$. Пусть $x_i \in R_i$, $1 \leqslant i \leqslant N$, — произвольные точки из прямоугольников. Тогда

$$\sup_\theta \left| \int g_\theta \, dF_n - \int g_\theta \, dF \right| \leqslant$$

$$\leqslant \sum_{i=1}^N \sup_\theta \left| \int\limits_{R_i} g_\theta \, dF_n - \int\limits_{R_i} g_\theta \, dF \right| \leqslant$$

$$\leqslant \sum_{i=1}^N \sup_\theta \left(\int\limits_{R_i} |g_\theta(x) - g_\theta(x_i)| (dF_n + dF) + g_\theta(x_i) \left| \int\limits_{R_i} dF_n - \int\limits_{R_i} dF \right| \right) \leqslant$$

$$\leqslant 2\varepsilon + N \sup_\theta \sup_x |g_\theta(x)| \sup_R \left| \int\limits_R dF_n - \int\limits_R dF \right|,$$

где R — некоторый прямоугольник. Но в силу d-мерного варианта теоремы Гливенко — Кантелли (см., например, Kiefer, Wolfowitz (1958) или Kiefer (1961)) имеем

$$\sup_R \left| \int_R dF_n - \int_R dF \right| \leqslant 2^d \sup_x |F_n(x) - F(x)| \to 0$$

почти наверное. Лемма 10 доказана.

Лемма 11. *При любых постоянных* $0 < c_1 \leqslant c_2 < \infty$ *имеем*

$$\inf_{c_1 \leqslant h \leqslant c_2} \inf_{x \in T} f * K_h(x) > 0$$

и

$$\sup_{c_1 \leqslant h \leqslant c_2} \sup_{x \in T} f * K_h(x) \leqslant M/c_1^d < \infty.$$

Доказательство. Второе утверждение леммы 11 очевидно. Для доказательства первого утверждения заметим, что

$$f * K_h(x) \geqslant c \int_{S_{x, rh}} \frac{f}{\lambda(S_{x, rh})} \geqslant c \int_{S_{x, rc_1}} \frac{f}{(rc_2)^d \lambda(S_{0,1})},$$

где λ — мера Лебега. Так как T — компакт, то семейство $\{S_{x, rc_1/2},\ x \in T\}$ содержит конечное множество сфер с центрами в $x_1, \ldots, x_N$, являющееся покрытием для T. Следовательно,

$$\inf_{x \in T} f * K_h(x) \geqslant c_3 \inf_{1 \leqslant i \leqslant N} \int_{S_{x_i, rc_1/2}} f > 0,$$

где c_3 — положительная постоянная. Так как эта нижняя граница не зависит от h, то лемма 11 доказана.

Лемма 12. *Для любых плотностей* f *и* K, *удовлетворяющих условиям теоремы* 4, *имеем*

$$\int \frac{f}{f * K_h} \leqslant M_1 + M_2 h^d,$$

где M_1, M_2 — *постоянные, зависящие только от* f, K *и* d.

Доказательство. Прежде всего заметим, что $f * K_h \geqslant h^{-d} c \int_{S_{x, hr}} f$. Всегда можно найти покрытие T множествами вида $S_{x, hr/2}$, содержащее самое большее $M_3 = M_4 + M_5/h^d$ множеств, где M_i — постоянные, зависящие только от f и K. Это следует из того, что

наименьший замкнутый куб, содержащий T, имеет такое покрытие. Пусть $x_1, x_2, \ldots, x_{M_3}$ — центры M_3 множеств покрытия. Ясно, что при любых $x \in S_{x_i,\, hr/2}$ имеем $\int\limits_{S_{x,hr}} f \geqslant \int\limits_{S_{x_i,\, hr/2}} f$. Таким образом,

$$\int\limits_T \frac{f}{f * K_h} \leqslant \sum_{i=1}^{M_3} \int\limits_{S_{x_i,\, hr/2}} \frac{h^d f(x)}{c \int\limits_{S_{x,\, hr}} f}\, dx \leqslant \sum_{i=1}^{M_3} \frac{h^d}{c} = \frac{M_3 h^d}{c},$$

что и требовалось доказать.

Лемма 13 (свойства нормы $L \log L$). *Пусть f и K — плотности, удовлетворяющие условиям теоремы 4. Тогда справедливы следующие утверждения*:

(i) $\int f \log (f * K_h) < \int f \log f$ *при любых* $h > 0$,

(ii) $\lim\limits_{h \downarrow 0} \int f \log (f * K_h) = \int f \log f$,

(iii) *Функция* $\int f \log (f * K_h)$ *непрерывна по* h *на* $(0, \infty)$.

Доказательство. Заметим сначала, что $\int f \log (f * K_h) < \infty$ при любых $h > 0$, но возможно, что $\int f \log f = \infty$. По неравенству Йенсена

$$\int\limits_T f \log \left(\frac{f * K_h}{f} \right) \leqslant \log \left(\int\limits_T f \cdot \frac{f * K_h}{f} \right) = \log \int\limits_T f * K_h < 0$$

при любых $h > 0$.

Доказательство утверждения (ii) состоит из двух частей. Обозначим через $\log_+$ и $\log_-$ положительную и отрицательную части функции $\log$ соответственно. Для первой части в силу леммы Фату и теоремы 2.3 имеем

$$\liminf_{h \downarrow 0} \int f \log_+ (f * K_h) \geqslant \int f \liminf_{h \downarrow 0} \log_+ (f * K_h) = \int f \log_+ f.$$

Далее нам потребуется следующий факт из анализа: для любого действительного числа $u \in (0, 1)$ имеем

$$\left| \log u + \sum_{j=0}^{J} \frac{(1-u)^j}{j} \right| = \sum_{j=J+1}^{\infty} \frac{(1-u)^j}{j} \leqslant \frac{1}{J+1} \cdot \frac{1}{u}.$$

Из этой оценки хвоста ряда и леммы 12 следует, что для любых целых J

$$\left| \int f \log_-(f * K_h) + \int f \sum_{j=0}^{J} \frac{(1 - f * K_h)_+^j}{j} \right| \leqslant$$

$$\leqslant \int \frac{f}{f * K_h} \cdot \frac{1}{J+1} \leqslant \frac{M_1 + M_2 h^d}{J+1}$$

и

$$\left| \int f \log_- f + \int f \sum_{j=0}^{J} \frac{(1-f)_+^j}{j} \right| \leqslant \frac{\lambda(T)}{J+1},$$

где λ — мера Лебега. С другой стороны, для любых целых j имеем $\int f (1 - f * K_h)_+^j \to \int f (1-f)_+^j$ (по теореме Лебега о мажорируемой сходимости и теореме 2.3). Объединяя все эти утверждения, получим, что $\int f \log_-(f * K_h) \to \int f \log_- f$. Но так как $\log = \log_+ + \log_-$, то $\liminf_{h \downarrow 0} \int f \log (f * K_h) \geqslant \int f \log f$, что вместе с утверждением (i) данной теоремы влечет за собой (ii).

Для доказательства утверждения (iii) рассмотрим произвольные h, $h' > 0$ и запишем сначала следующее неравенство:

$$\left| \int f \log (f * K_h) - \int f \log (f * K_{h'}) \right| \leqslant \int f \left| \log \left(\frac{f * K_h}{f * K_{h'}} \right) \right|. \qquad (25)$$

Если $h' > 0$ фиксировано и $h \in [h'/2, 2h']$, то в силу леммы 11 величина $\sup_{x \in T} | \log (f * K_h / f * K_{h'}) |$ ограничена равномерно по h. Обозначим верхнюю границу для этой величины через c_0. Тогда, в силу того что $| \log u - \log v | \leqslant | u - v | / \max (u, v)$ при $u, v > 0$, подынтегральное выражение в (25) не превосходит

$$c_1 f | f * K_h - f * K_{h'} |. \qquad (26)$$

Здесь c_1 — верхняя граница для $\sup_{h'/2 \leqslant h \leqslant 2h'} \sup_{x \in T} 1/(f * K_h)$. Интеграл от функции (26) не превосходит

$$c_1 c_2 \int_{f > c_3} f + c_1 c_3 \int | f * K_h - f * K_{h'} | \qquad (27)$$

при любых $c_3 > 0$, где $c_2 = \sup_{h'/2 \leqslant h \leqslant 2h'} \sup_{x \in T} f * K_h$ (это выражение конечно также в силу леммы 11). Рассуждая так же, как при выводе (2.4) в теореме 2.4, получим, что последнее слагаемое в (27) есть $o(1)$ при $h \to h'$. Первое слагаемое в (27) можно сделать сколь угодно малым за счет выбора c_3. Утверждение (iii) леммы 13 доказано.

Лемма 14. *Пусть C принадлежит отрезку $[c_1, c_2] \subseteq (0, \infty)$ и выполнены условия теоремы 4. Тогда*

$$\sup_{h \in C} \left| \frac{1}{n} \log L(h) - \int f \log (f * K_h) \right| \to 0$$

почти наверное.

Доказательство. Покажем сначала, что

$$\sup_{h \in C} \left| \frac{1}{n} \log L(h) - \frac{1}{n} \sum_{i=1}^{n} \log (f * K_h(X_i)) \right| \to 0$$

почти наверное. Действительно, заметим, что

$$\sup_{h \in C} \sup_i |f_{ni}(X_i) - f_n(X_i)| \leqslant \sup_{h \in C} \sup_{x \in T} \sup_i |f_{ni}(x) - f_n(x)| \leqslant$$

$$\leqslant n^{-1} \sup_{h \in C} \sup_{x, i} f_{ni}(x) + \sup_{h \in C} M (nh^d)^{-1} = o(1)$$

и что

$$\sup_{h \in C} \sup_x |f_n(x) - f * K_h(x)| \to 0$$

почти наверное в случае, когда ядро K ограниченно, имеет компактный носитель и непрерывно п. в. (см., например, Devroye, Wagner (1980) или Bertrand-Retali (1978)). Используя равномерные верхнюю и нижнюю границы для $f * K_h$ из леммы 11, заключаем, что

$$\sup_{h \in C} \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{f_{ni}(X_i)}{f * K_h(X_i)} \right) \to 0$$

почти наверное. Лемма 14 будет доказана, если мы покажем, что

$$\sup_{h \in C} \left| \frac{1}{n} \sum_{i=1}^{n} \log (f * K_h(X_i)) - \int f \log (f * K_h) \right| \to 0$$

почти наверное. Положим $g_h = \log (f * K_h)$. Требуется показать, что $\sup_{h \in C} \left| \int g_h dF_n - \int g_h dF \right| \to 0$ почти наверное. Для этого проверим условия леммы 10. Прежде всего семейство $\{g_h, h \in C\}$ равномерно ограничено ввиду леммы 11. Таким образом, достаточно проверить только равностепенную непрерывность. Применяя снова лемму 11, видим, что достаточно проверить равностепенную непрерывность функций $f * K_h$. Пусть $x, y \in T$. Тогда

$$\sup_{h \in C} |f * K_h(x) - f * K_h(y)| \leqslant \sup_{h \in C} \int K_h(z) |f(x - z) - f(y - z)| \, dz \leqslant$$

$$\leqslant (M / c_1^d) \int |f(x - z) - f(y - z)| \, dz. \quad (28)$$

Но последнее выражение равно $o(1)$ при $y \to x$, что можно показать, аппроксимируя f в L_1 некоторой равномерно непрерывной функцией f^* с компактным носителем. Так как интеграл в (28), таким образом, непрерывен по y при каждом фиксированном x и так как x, y принадлежат компакту T, то этот интеграл равномерно непрерывен по x и y. Лемма 14 доказана.

Лемма 15 (первое утверждение теоремы 4). *Если выполнены условия теоремы 4, то* $h_n \to 0$ *почти наверное.*

Доказательство. Множество A_n почти наверное не пусто, так как $L(h) = 0$ при всех достаточно малых h (это следует из компактности носителя K) и $L(h) > 0$ при всех достаточно больших h (это следует из неравенства $K \geqslant cI_{S_{0,r}}$).

Для доказательства леммы достаточно установить, что для любого $\varepsilon > 0$ существует $\delta \in (0, \varepsilon)$, такое, что

$$\liminf_{n\to\infty} (aL(\delta) - \sup_{h\geqslant\varepsilon} L(h)) > 0 \tag{29}$$

почти наверное. (Этого достаточно, так как тогда если $\gamma \in [\varepsilon, \infty)$, то $L(\gamma) \leqslant \sup_{h\geqslant\varepsilon} L(h) < aL(\delta) \leqslant a \sup_{h>0} L(h)$ почти наверное при всех достаточно больших n, и, следовательно, $\gamma \notin A_n$). Но (29) имеет место, если

$$\limsup_{n\to\infty} \sup_{h\geqslant\varepsilon} \frac{1}{n} \log L(h) <$$

$$< \liminf_{n\to\infty} \frac{1}{n} \log(aL(\delta)) = \liminf_{n\to\infty} \frac{1}{n} \log L(\delta). \tag{30}$$

В силу леммы 14 правая часть (30) почти наверное равна $\int f \log(f * K_\delta)$. Если M_1 — большое положительное число, то также

$$\limsup_{n\to\infty} \sup_{h\geqslant\varepsilon} \frac{1}{n} \log L(h) \leqslant$$

$$\leqslant \max\left(\limsup_{n\to\infty} \sup_{\varepsilon\leqslant h\leqslant M_1} \frac{1}{n} \log L(h),\ \limsup_{n\to\infty} \sup_{h>M_1} \log\left(\frac{M}{h^d}\right)\right) \leqslant$$

$$\leqslant \max\left(\sup_{\varepsilon\leqslant h\leqslant M_1} \int f \log(f * K_h),\ \log(M/M_1^d)\right) <$$

почти наверное (в силу леммы 14)

$$< \int f \log(f * K_\delta)$$

при некотором $0 < \delta < \varepsilon$. Последнее неравенство следует из леммы 13 и того, что M_1 можно выбрать сколь угодно большим. Лемма 15 доказана.

Лемма 16 (второе утверждение теоремы 4). *Если выполнены условия теоремы 4, то* $\liminf_{n\to\infty} (nh_n^d/\log n) > 0$ *почти наверное.*

Доказательство. При любом $a > 0$ имеем

$$P\left(\sup_{h\in A_n} h < a\right) \leqslant P\left(\max_i \min_{j\neq i} \|X_i - X_j\| < \tilde{r}a\right). \qquad (31)$$

Это следует из того, что при $h < (1/\tilde{r}) \max_i \min_{j\neq i} \|X_i - X_j\|$ имеем $L(h) = 0$, так что, конечно, $h \notin A_n$ (в самом деле, $L(h) > 0$ при всех достаточно больших h). Неравенство (31) является отправной точкой доказательства. Чтобы не накладывать на f слишком много ограничений, потребуются довольно тонкие рассуждения. Предположим, что носитель T плотности f содержится в замкнутом квадрате Q, который, не нарушая общности, можно считать равным $[0, 1]^d$. Воспользуемся неравенством (31) при $a = a_n = ((\varepsilon \log n)/n)^{1/d}$, где $\varepsilon > 0$ — некоторое малое число, которое будет выбрано в дальнейшем. Разобьем каждую сторону Q на $[1/(\tilde{r}a)]$ интервалов равной длины (таким образом, длина каждого интервала больше или равна $\tilde{r}a$, и пропорциональна $\tilde{r}a$ при $n \to \infty$). Соответствующая этому разбиению сетка квадратов состоит из m_n ячеек B_i, причем, снова не нарушая общности, предполагается, что m_n кратно 3^d. Каждая ячейка имеет d координат, а каждая координата представляет собой целое число, лежащее между 1 и $[1/(\tilde{r}a)]$ и равное рангу интервала, который является соответствующей проекцией ячейки. Пусть C_i' — ячейки, все координаты которых имеют вид $2 + 3j$, $j = 0, 1, 2, \ldots$, и пусть C_i — совокупность ячеек, имеющих хотя бы одну общую вершину с C_i'. Таким образом, C_i является сверхъячейкой, состоящей из 3^d первоначальных ячеек, и существует ровно $m_n/3^d$ таких ячеек C_i. Предполагается, что первые $m_n/3^d$ целых чисел в линейной нумерации ячеек являются индексами ячеек типа C_i'. Пусть $p_i = \int_{C_i} f$, $p_i' = \int_{C_i'} f$.

Проведем теперь пуассонизацию объема выборки. Этот прием — исключительно дело вкуса. Так, доказательство в R^1 можно легко провести без пуассонизации, поскольку хорошо изучены свойства спейсингов на действительной оси. Пусть N_0, N_1 — независимые пуассоновские случайные величины с параметрами $n - b_n$ и $2b_n$ соответственно, где $b_n = \sqrt{M^* n \log n}$ и M^* — большое число, которое будет выбрано в дальнейшем. Случайная величина $N = N_0 + N_1$ имеет, таким образом, распределение Пуассона с параметром $n + b_n$. При заданном N извлечем выборку из совокуп-

ности с плотностью f, состоящую из независимых случайных векторов $X_1, \ldots, X_N$. Далее, правая часть (31) равна

$$P\left(\bigcap_{i=1}^{n}[S_{X_i,\bar{r}a} \text{ содержит хотя бы одно } X_j,\ j\neq i,\ j\leqslant n]\right)\leqslant$$

$$\leqslant P\left(\bigcap_{i=1}^{m_n/3^d}[C_i'=\varnothing]\cup\right.$$

$$\left.\cup[C_i'\neq\varnothing,\ C_i \text{ содержит хотя бы два } X_j,\ j\leqslant n]\right)\leqslant$$

$$\leqslant P(N<n)+P(N_0>n)+$$

$$+P\left(\begin{array}{l}\bigcap\limits_i[C_i' \text{ не содержит } X_j,\ j\leqslant N_0]\cup\\ \cup[C_i' \text{ содержит хотя бы одно } X_j\\ \text{при } j\leqslant N;\ C_i \text{ содержит хотя бы}\\ \text{две точки с индексами } j\leqslant N];\\ N_0\leqslant n,\ N\geqslant n\end{array}\right). \qquad (32)$$

В силу леммы 9 первые два слагаемых в правой части (32) не превосходят

$$2\exp\left(-\frac{b_n^2}{2(n+2b_n)}\right)+2\exp\left(-\frac{b_n^2}{2n}\right)=4n^{-M^*/(2+o(1))},$$

и эта последовательность суммируема по n при $M^*>2$. Покажем, что при фиксированном M^* последнее слагаемое в (32) ограничено сверху функцией, суммируемой по n. Отсюда в силу (31) и леммы Бореля — Кантелли будет следовать, что $\sup\limits_{h\in A_n} h<((\varepsilon\log n)/n)^{1/d}$ конечное число раз почти наверное, что и требуется доказать.

Исключим в последнем слагаемом в (32) ограничения $N_0\leqslant n$, $N\geqslant n$, тем самым увеличив его. В результате получим вероятность пересечения независимых событий:

$$\prod_{i=1}^{m_n/3^d}\left(e^{-(n-b_n)p_i'}+\left(1-e^{-(n+b_n)p_i'}-\right.\right.$$

$$\left.\left.-(n+b_n)p_i'e^{-(n+b_n)p_i'-(n+b_n)(p_i-p_i')}\right)\right)\leqslant$$

$$\leqslant\exp\left(\sum_{i=1}^{m_n/3^d}\left(e^{-(n-b_n)p_i'}-e^{-(n+b_n)p_i'}\right)-\right.$$

$$\left.-\sum_{i=1}^{m_n/3^d}(n+b_n)p_i'e^{-(n+b_n)p_i}\right), \qquad (33)$$

где мы воспользовались неравенством $1 + u \leqslant e^u$, верным при любых u. При $0 \leqslant u \leqslant v$ имеем $e^{-u} - e^{-v} \leqslant e^{-u}(v - u) \leqslant v - u$. Следовательно, первая сумма под экспонентой в (33) не превосходит

$$2b_n \sum_{i=1}^{m_n/3^d} p_i' \leqslant 2b_n.$$

Введем теперь следующие гистограммные аппроксимации плотности:

$$g_n(x) = \frac{p_i}{(3\tilde{r}a)^d}, \quad g_n'(x) = \frac{p_i'}{(\tilde{r}a)^d}, \quad x \in C_i.$$

Для последней суммы под экспонентой в (33) имеем

$$\sum_{i=1}^{m_n/3^d} (n + b_n) p_i' e^{-(n+b_n) p_i} \sim$$

$$\sim 3^{-d} \sum_{i=1}^{m_n/3^d} \int_{C_i} (n + b_n) g_n'(x) e^{-(n+b_n)(3\tilde{r}a)^d g_n(x)} dx \sim$$

$$\sim 3^{-d} \int_Q n g_n'(x) e^{-(n+b_n)(3\tilde{r}a)^d g_n(x)} dx. \tag{34}$$

Но в силу леммы Фату и теоремы 2.2

$$\liminf_{n\to\infty} \frac{(34)}{2b_n} \geqslant \frac{1}{2} 3^{-d} \int_Q f(x) \liminf_{n\to\infty} \sqrt{\frac{n}{M^* \log n}} \times$$

$$\times e^{-(n+b_n)(3\tilde{r}a)^d g_n(x)} dx.$$

Здесь нижний предел бесконечен на множестве ненулевой f-меры:

$$\liminf_{n\to\infty} \sqrt{\frac{n}{\log n}} \exp\left(-(n + b_n)(3\tilde{r}a)^d g_n(x)\right) = \infty. \tag{35}$$

В самом деле, по теореме 2.2 $g_n \to f$ при почти всех x. Кроме того, $(n + b_n)(3\tilde{r}a)^d \sim (3\tilde{r})^d \varepsilon \log n$, так что выражение под знаком нижнего предела в (35) равно

$$\sqrt{\frac{n}{\log n}} \exp\left(-(3\tilde{r})^d f(x)(\varepsilon + o(1)) \log n\right)$$

при почти всех x. Эта величина стремится к ∞ на множестве положительной f-меры, если ε достаточно мало. Таким образом, при всех достаточно больших n правая часть (33) не превосходит значений последовательности $\exp\left(-\sqrt{n \log n}\right)$, суммируемой по n. Лемма 16 доказана.

6. Инвариантное оценивание плотности

Проблема инвариантности оценок плотности относительно определенных преобразований впервые была затронута Wertz (1974a, 1976). В этом параграфе будет показано, что использование адаптивных оценок плотности в общем случае мотивируется соображениями инвариантности.

Пусть φ — монотонно возрастающая функция, взаимно однозначно отображающая R на R, и пусть φ и обратная к ней функция абсолютно непрерывны на конечных интервалах. Говорят, что оценка плотности f_n на R является φ-*инвариантной*, если

$$f_n(\varphi(x), \varphi(x_1), \ldots, \varphi(x_n)) = (\varphi^{-1}(x))' f_n(x, x_1, \ldots, x_n)$$

при любых n и любых $x, x_1, \ldots, x_n \in R^{n+1}$. Это определение можно перефразировать следующим образом. Предположим, что построена оценка плотности $f_n(x, X_1, \ldots, X_n)$. Тогда оценку плотности величины $Y = \psi(X_1)$ можно получить двумя способами:

(i) Обычным преобразованием плотностей; при этом получим

$$\varphi'(y) f_n(\varphi(y), \varphi(y_1), \ldots, \varphi(y_n)), \quad y, y_1, \ldots, y_n \in R^{n+1},$$

где $\varphi = \psi^{-1}$.

(ii) Построением новой оценки по данным $\psi(X_1), \ldots, \psi(X_n)$,

$$f_n(\psi(x), \psi(x_1), \ldots, \psi(x_n)) = f_n(y, y_1, \ldots, y_n),$$
$$y, y_1, \ldots, y_n \in R^{n+1}.$$

По существу φ-инвариантность означает, что эти оценки совпадают друг с другом. Иначе говоря, φ-инвариантность означает, что

$$\int_B f_n(y, y_1, \ldots, y_n)\, dy = \int_{\varphi[B]} f_n(y, \varphi(y_1), \ldots, \varphi(y_n))\, dy$$

для любых борелевских множеств B, где $\varphi[B]$ — множество значений $\varphi(y)$, $y \in B$.

В частности, представляет интерес *трансляционная инвариантность* ($\varphi(x) = x + a$, $a \in R$) и *масштабная инвариантность* ($\varphi(x) = bx$, $b > 0$). Трансляционная инвариантность почти всегда рассматривается как само собой разумеющееся, и оценки, не обладающие этим свойством, кажутся несколько странными. Трансляционная инвариантность по существу эквивалентна требованию, что f_n является функцией только от $x - X_1, \ldots, x - X_n$. Очевидно, что все ядерные оценки с фиксированным h трансляционно инвариантны. Это утверждение остается в силе, если допустить, что h зависит от всех попарных разностей $X_i - X_j = (x - X_j) - (x - X_i)$. К сожалению, гистограммную оценку можно сделать трансляционно инвариантной, только если не-

сколько изменить ее определение. *Оценки типа дельта-функции Дирака* (Walter, Blum (1979)) имеют вид

$$f_n(x) = \sum_{i=1}^{n} w_{ni} K_{ni}(x, X_i),$$

где w_{ni} — веса и K_{ni} — заданные действительнозначные функции. Оценки такого вида трансляционно инвариантны, только если $K_{ni}(x, y) = K_{ni}^*(x - y)$ для некоторых функций K_{ni}^*. Поэтому только в такой форме эти оценки рассматриваются в данной книге. В гл. 12 будет показано, что проекционные оценки на действительной оси не являются трансляционно инвариантными. Заметим также, что без трансляционной инвариантности трудно добиться выполнения условия $\int f_n = 1$.

Для масштабно инвариантных оценок плотности имеем

$$bf_n(bx, bX_1, \ldots, bX_n) = f_n(x, X_1, \ldots, X_n)$$

при любых $b > 0$. К сожалению, обычная ядерная оценка с фиксированным h не является масштабно инвариантной при любых ядрах K, поскольку условие

$$\frac{1}{b}\left(\frac{1}{h} K\left(\frac{x-y}{h}\right)\right) = \frac{1}{h} K\left(\frac{bx-by}{h}\right), \quad \forall x, y \in R, b, h > 0,$$

выполнено, только если $K(x) = c/x$ или $K(x) = c/|x|$ при некотором $c \in R$. Ядерные оценки с такими ядрами обладают очень плохими свойствами: например, при $K(x) = c/|x|$ имеем $\int |f_n| = \infty$ при любых n и $E(f_n(x)) = \infty$ при почти всех x, для которых $f(x) > 0$. При обычных условиях вида $\int |K| < \infty$, $\int K = 1$ на ядро K ядерные оценки могут быть масштабно инвариантными, только если h зависит от данных. В частности, масштабная инвариантность имеет место, если h зависит от данных таким образом, что

$$h(bX_1, \ldots, bX_n) = bh(X_1, \ldots, X_n).$$

Это один из основных аргументов в пользу изучения адаптивных ядерных оценок. Примерами функций h указанного вида являются

$$c\left(\sum_{i,j} |X_i - X_j|^p\right)^{1/p}, \quad p > 0,$$

$$c(X_{(b_n)} - X_{(a_n)}),$$

где $1 \leqslant a_n \leqslant b_n \leqslant n$ и $X_{(i)}$ есть i-я порядковая статистика выборки $X_1, \ldots, X_n$. В обоих случаях c — соответствующим образом выбранная функция только от n.

Инвариантность по отношению к другим преобразованиям требуется редко. Так, существуют приложения, где нужны оценки плотности в линейном и логарифмическим масштабах. Было бы неприятно обнаружить, что обычное логарифмическое преобразование оценки с линейным масштабом не согласуется с оценкой, построенной по логарифмически преобразованным данным. Если разница есть, то какую оценку следует выбирать? Другой вопрос, который здесь не был затронут, касается существования разумных (состоятельных и т. д.) φ-инвариантных оценок при заданном преобразовании φ.

7. Скорость сходимости адаптивных ядерных оценок

В этом параграфе приводится общая теорема, которая позволяет делать выводы о скорости сходимости адаптивной ядерной оценки при условии, что имеется некоторая информация о поведении h, в частности, о близости h к некоторой детерминированной последовательности a_n. Подчеркнем, что новое условие, накладываемое здесь на ядро, вводится только для удобства выкладок.

Теорема 5. *Предположим, что f_n — адаптивная ядерная оценка с параметром сглаживания h и ядром K, где $\int K = 1$, и*

$$\int |K_u - K| \leqslant C(1 - u^{-d}), \quad u \geqslant 1,$$

с некоторой постоянной C. Если f_{na} — обычная ядерная оценка с параметром сглаживания a и тем же самым ядром K, то

$$\left| \int |f_n - f| - \int |f_{na} - f| \right| \leqslant C\left(1 - \min^d\left(\frac{a}{h}, \frac{h}{a}\right)\right).$$

Доказательство. Не нарушая общности, предположим, что $a \leqslant h$. Тогда

$$\int |f_n - f| - \int |f_{na} - f| \Big| \leqslant \int |f_n - f_{na}| \leqslant$$

$$\leqslant \frac{1}{n} \sum_{i=1}^{n} \int \left| h^{-d} K\left(\frac{x - X_i}{h}\right) - a^{-d} K\left(\frac{x - X_i}{a}\right) \right| dx =$$

$$= \int \left| \left(\frac{a}{h}\right)^d K\left(\frac{ax}{h}\right) - K(x) \right| dx \leqslant C\left(1 - \left(\frac{a}{h}\right)^d\right).$$

Замечание. Условие, накладываемое на K, выполнено для большинства плотностей K, в частности для всех плотностей K на R^d, которые не возрастают вдоль лучей, т. е. таких, что $K(ux) \leqslant K(x)$ при любых $x \in R^d$, $u \geqslant 1$. В этом случае C можно

положить равным двум, что явствует из следующей простой выкладки:

$$\int |K_u - K| \leqslant \int \left| u^{-d} K\left(\frac{x}{u}\right) - u^{-d} K(x) \right| dx + \\ + \int |u^{-d} K(x) - K(x)|\, dx = u^{-d} \int \left(K\left(\frac{x}{u}\right) - \right. \\ \left. - K(x) \right) dx + (1 - u^{-d}) = 2(1 - u^{-d}), \quad u \geqslant 1.$$

Неравенство теоремы 5 получено без каких-либо предположений о параметре сглаживания, как то: сходимость к нулю и т. д. Оно имеет много применений, лишь небольшая часть которых будет перечислена ниже в леммах.

Лемма 17. *Для адаптивной ядерной оценки* f_n *с ядром* K, *удовлетворяющим условиям теоремы* 5, *имеем*

$$E\left(\int |f_n - f|\right) \leqslant E\left(\int |f_{na_n} - f|\right) + C\varepsilon_n + \\ + 2P\left(1 - \min^d\left(\frac{a_n}{h}, \frac{h}{a_n}\right) \geqslant \varepsilon_n\right),$$

где a_n, ε_n — *последовательности положительных чисел и* f_{na_n} — *обычная ядерная оценка с параметром сглаживания* a_n.

Эта лемма исключительно важна. Например, рассмотрим параметры сглаживания вида

$$h = a_n \frac{2}{n} \sum_{i=1}^{n/2} |X_{2i-1} - X_{2i}|.$$

Здесь a_n — произвольная последовательность положительных чисел. Можно было бы рассмотреть и двойную сумму по всем $|X_i - X_j|$, но это было бы нелегко в численном отношении и, конечно, запутало бы изложение. Справедлива

Лемма 18. *Если ядро адаптивной ядерной оценки неотрицательно, равно нулю вне* $[-1, 1]$ *и удовлетворяет условиям теоремы* 5, *параметр сглаживания* h *выбирается, как указано в предыдущем абзаце, при некоторой произвольной последовательности положительных чисел* a_n, *и если* $E(|X_1|^p) < \infty$ *при некотором* $p > 4$, *то*

$$E\left(\int |f_n - f|\right) \sim E\left(\int |f_{na_n E(|X_1 - X_2|)} - f|\right)$$

в обозначениях леммы 17.

Заметим, что условия на f исчерпываются указанным моментным ограничением.

Доказательство. Применим непосредственно лемму 17. В силу теоремы 5.2 мы получим требуемый результат, если укажем последовательность ε_n, такую, что $\varepsilon_n = o(n^{-2/5})$ и $P(|h/(a_n E(|X_1 - X_2|)) - 1| > \varepsilon_n) = o(n^{-2/5})$. По неравенству Чебышева последняя вероятность не превосходит

$$E\left(\varepsilon_n^{-p}\left|\frac{2}{n}\sum_{i=1}^{n/2}(Y_i - E(Y_i))\right|^p\right) \leqslant$$

$$\leqslant C_p \varepsilon_n^{-p}\left(\frac{2}{n}\right)^p E\left(\left(\sum_{i=1}^{n/2}(Y_i - E(Y_i))^2\right)^{p/2}\right)$$

(это имеет место при $p \geqslant 2$ и некотором $C_p > 0$ в силу неравенства Марцинкевича и Зигмунда, см. лемму 5.27)

$$\leqslant C_p \varepsilon_n^{-p}\left(\frac{2}{n}\right)^p \left(\frac{n}{2}\right)^{p/2-1}\sum_{i=1}^{n/2} E(|Y_i - E(Y_i)|^p)$$

$$\leqslant C_p \varepsilon_n^{-p}\left(\frac{2}{n}\right)^{p/2} 2^{p-1} E(|Y_1|^p) \leqslant A\varepsilon_n^{-p} n^{-p/2},$$

где $Y_i = |X_{2i-1} - X_{2i}|$ и A — некоторая конечная постоянная. Если выбрать $\varepsilon_n \sim n^{-p/2(p+1)}$, то $\varepsilon_n = o(n^{-2/5})$ при $p > 4$ и последовательность $\varepsilon_n^{-p} n^{-p/2}$ убывает так же, как ε_n, с точностью до постоянного коэффициента пропорциональности. Лемма 18 доказана.

Замечание. Из теорем 5, 3.1 и рассуждений, сходных с теми, которые приводят к лемме 17, явствует, что если интеграл от ядра равен 1, ядро абсолютно интегрируемо и удовлетворяет условиям теоремы 5, и если существует последовательность a_n, такая, что $a_n \to 0$, $na_n^d \to \infty$ и $h/a_n \to 1$ по вероятности (почти наверное), то $\int |f_n - f| \to 0$ по вероятности (почти наверное). Теорема 6.1 является более общей, поскольку в ней не накладывается никаких дополнительных условий на K и не предполагается существования центрирующей последовательности a_n (тем самым допускается бóльшая изменчивость случайной величины h).

ЛИТЕРАТУРА

Bennett G.
(1962) Probability inequalities for the sum of independent random variables, Journal of the American Statistical Association, 57, pp. 33—45.

Bertrand-Retali M.
(1978) Convergence uniforme d'un estimateur de la densité par la méthode du noyau, Revue Roumaine de Mathematiques Pures et Appliquées, 23, pp. 361—385.

Bowman A. W.
(1982) A comparative study of some kernel-based nonparametric density estimators, Manchester-Sheffield School of Probability and Statistics, Research Report No. 84/AWB/1.
Bretagnolle J., Huber C.
(1979) Estimation des densités: risque minimax, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 47, pp. 119—137.
Chow Y. S., Geman S., Wu L. D.
(1983) Consistent cross-validated density estimation, Annals of Statistics, 11, pp. 25—38.
Deheuvels P.
(1974) Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, Comptes Rendus Académie des Sciences de Paris Série A, 178, pp. 1217—1220.
(1977) Estimation non paramétrique de la densité par histogrammes généralisés, Revue de Statistique Appliquée, 25, pp. 5—42.
Deheuvels P., Hominal P.
(1980) Estimation automatique de la densité, Revue de Statistique Appliquée, 28, pp. 25—55.
Devroye L., Wagner T. J.
(1980) The strong uniform consistency of kernel density estimates, in Multivariate Analysis V, P. R. Krishnaiah (Ed.), North-Holland, New York, pp. 59—77.
Duin R. P. W.
(1976) On the choice of smoothing parameters for Parzen estimators of probability density functions, IEEE Transactions on Computers, C—25, pp. 1175—1179.
Geman S.
(1981) Sieves for nonparametric estimation of densities and regressions, Reports in Pattern Analysis No. 99, Division of Applied Mathematics, Brown University, Providence, Rhode Island.
Geman S., Hwang C.-R.
(1982) Nonparametric maximum likelihood estimation by the method of sieves, Annals of Statistics, 10, pp. 401—414.
Habbema J. D. F., Hermans J., Vandenbroek K.
(1974) A stepwise disriminant analysis program using density estimation in COMPSTAT 1974, G. Bruckmann (Ed.), Physica Verlag, Wien, pp. 101—110.
Hall P.
(1982a) Cross-validation in density estimation, Biometrika, 69, pp. 383—390.
(1982b) Limit theorems for stochastic measures of the accuracy of nonparametric density estimators, Stochastic Processes and Applications, 13, pp. 11—25.
(1983a) Large-sample optimality of least squares cross-validation in density estimation, Annals of Statistics, 11, pp. 1156—1174.
(1983b) Asymptotic theory of minimum integrated square error for multivariate density estimation, Proceedings of the Sixth International Symposium on Multivariate Analysis, Pittsburgh.
Kiefer J.
(1961) On large deviations of the empiric d. f. of vector chance variables and a law of the iterated logarithm, Pacific Journal of Mathematics, 11, pp. 649—660.
Kiefer J., Wolfowitz J.
(1958) On the deviations of the empiric distribution function of vector chance variables, Transactions of the American Mathematical Society, 87, pp. 173—186.
Rosenblatt M.
(1956) Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, pp. 832—837.
(1971) Curve estimates, Annals of Mathematical Statistics, 42, pp. 1815—1842.
Rudemo M.
(1982) Empirical choice of histogram and kernel density estimators, Scandinavian Journal of Statistics, 9, pp. 65—78.

Schuster E. F., Gregory G. G.
(1978) Choosing the shape factor (s) when estimating a density, Technical Report, Department of Mathematics, University of Texas, El Paso, Texas.
(1981) On the nonconsistency of maximum likelihood nonparametric density estimators, in Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface, W. F. Eddy (Ed), Springer-Verlag, New York, pp. 295—298.
Scott D. W.
(1979) Optimal data-based histograms. Biometrika, 66, pp. 605—610.
Scott D. W., Tapia R. A., Thompson J. R.
(1977) Kernel density estimation revisited, Journal of Nonlinear Analysis, Theory, Methods and Applications, 1, pp. 339—372.
Scott D. W., Factor L. E.
(1981) Monte-Carlo study of three data-based nonparametric probability density estimators, Journal of the American Statistical Association, 76, pp. 9—15.
Seneta E.
(1976) Regularly Varying Functions, Lecture Notes in Mathematics, No. 508, Springer-Verlag, Heidelberg. (Имеется русский перевод: Сенета Е. Правильно меняющиеся функции. — М.: Наука, 1985.)
Silverman B. W.
(1978) Choosing the window width when estimating a density, Biometrika, 65, pp. 1—11.
Stone C. J.
(1983) An asymptotically efficient histogram selection rule, Proceedings of the Neyman — Kiefer meeting..
(1984) An asymptotically optimal window selection rule for kernel density estimates, Annals of Statistics, 12, pp. 1285—1297.
Tapia R. A., Thompson J. R.
(1978) Nonparametric Probability Density Estimation, The John Hopkins University Press, Baltimore, Maryland.
Wagner T. J.
(1975) Nonparametric estimates of probability densities, IEEE Transactions on Information Theory, IT—21, pp. 438—440.
Walter G., Blum J.
(1979) Probability density estimation using delta-sequences, Annals of Statistics, 7, pp. 328—340.
Wertz W.
(1974a) Invariante und Optimale Dichteschätzungen, Mathematica Balkanica, 4, pp. 707—722.
(1974b) On the existence of density estimators, Studia Scientiarium Mathematicarum Hungarica, 9, pp. 45—50.
(1976) Invariant density estimation, Monatshefte für Mathematik, 81, pp. 315—324.
Woodroofe M.
(1970) On choosing a delta-sequence, Annals of Mathematical Statistics, 41, pp. 1665—1671.
Надарая Э. А.
(1974) Об интегральной среднеквадратичной ошибке некоторых непараметрических оценок плотности вероятностей//Теория вероятн. и ее примен. — Т. 19. — 1. — С. 131—140.

Глава 7

Оценки, близкие к ядерной и гистограммной оценкам

1. Введение

Все оценки плотности, рассматриваемые в этой главе, являются плотностями на R^d. Они в основном представляют собой обобщения ядерной и гистограммной оценок, обладающие некоторыми дополнительными свойствами:

(i) лучшее поведение при малых выборках;
(ii) робастность;
(iii) простое рекуррентное задание;
(iv) локально настраиваемое сглаживание.

Вообще говоря, легко убедиться в состоятельности этих оценок, хотя результаты об эквивалентности в полном объеме теоремы 3.1 в настоящее время не известны. В гл. 3—5 была создана прочная база для сравнения ядерных и гистограммных оценок. Гораздо меньше известно о скоростях сходимости для обобщений этих оценок. В отличие от состоятельности исследование скорости сходимости, как правило, является нестандартной задачей.

Рассматриваемые оценки будут довольно произвольным образом разбиты на следующие группы:

1. Ядерные оценки с переменным параметром сглаживания.
2. Рекуррентные ядерные оценки.
3. Оценки максимума правдоподобия.
4. Переменные гистограммные оценки.
5. Ядерные оценки с пониженным смещением.
6. Оценка Гренандера для монотонных плотностей.

2. Ядерные оценки с переменным параметром сглаживания

В 1977 г. (Breiman, Meisel, Purcell (1977)) была предложена следующая оценка:

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} H_{ni}^{-d} K\left(\frac{x - X_i}{H_{ni}}\right), \tag{1}$$

где K — заданное ядро и H_{ni} — расстояние от X_i до k_n-го ближайшего соседа точки X_i среди $X_1, \ldots, X_n$ [1]). Оценка (1) является плотностью как функция от x и, по-видимому, обладает локально настраиваемым параметром сглаживания, поскольку, грубо говоря, значение H_{ni} велико там, где плотность f мала, и наоборот. Конечно, нужно еще выбирать k_n, так что эту оценку едва ли можно назвать адаптивной.

Оценка (1), вероятно, произошла от *оценки ближайших соседей*, которую можно задать с помощью (1), заменяя там H_{ni} на расстояние $H_n(x)$ от x до k_n-го ближайшего соседа точки x среди $X_1, \ldots, X_n$; см., например, Moore, Yackel (1977), Mack, Rosenblatt (1979) и Loftsgaarden, Quesenberry (1965) (последние авторы впервые дали определение оценки ближайших соседей при частном выборе ядра $K(x) = I_{S_{01}}(x)/\lambda(S_{01})$). Однако если вместо H_{ni} использовать $H_n(x)$, то оценка не является плотностью: действительно, $\int f_n = \infty$ при любых n.

Идея локальной настройки параметра сглаживания заслуживает дальнейшего изучения. Анализ, приведенный в гл. 5, показывает, что параметр сглаживания должен быть большим, если величины $1/f$ и $|f''|$ малы. В областях высокой кривизны f (т. е. больших значений $|f''|$) или малых значений f нужно уменьшать параметр сглаживания. Так как параметр сглаживания H_{ni} не зависит от кривизны, он по сути не является асимптотически оптимальным. В самом деле, при фиксированной кривизне параметр сглаживания должен возрастать с ростом значений f, а не убывать, как это имеет место в методах ближайших соседей.

Тем не менее экспериментальные результаты для оценок (1), приведенные в работах Breiman и др. (1977), Habbema и др. (1978) и Raatgever и др. (1978) по меньшей мере можно считать обнадеживающими. Однако большинство исследователей избегало подробного изучения свойств оценки (1), представляющей собой сумму зависимых случайных величин. Деврой (Devroye (1985)) показал, что оценка (1) сходится в среднем в L_1 к f при любых f, если K не возрастает вдоль лучей, $k_n = o(n)$ и $\lim\limits_{n\to\infty} k_n = \infty$. Абрамсон (Abramson (1982)) предложил еще один метод выбора величин H_{ni}, зависящих от x. Однако его метод нетрудно видоизменить так, чтобы зависимости от x не было. При этом потребуется теорема состоятельности в духе теоремы 6.2 (которую можно рассматривать как частный случай, когда $H_{n1} = \ldots = H_{nn} = H_n$, H_n — случайная величина), где условия будут налагаться, скажем, на квантили последовательности H_{ni}.

---

[1]) k_n-м ближайшим соседом точки x среди $X_1, \ldots, X_n$ называется точка X_j, такая, что $\|x - X_j\|$ есть k_n-й элемент вариационного ряда для $\|x - X_1\|, \ldots, \|x - X_n\|$. — *Прим. перев.*

Дальнейшего исследования заслуживает еще одна задача — выбор H_{ni}. В методах ближайших соседей, таких, как в работе Breiman и др. (1977), параметр сглаживания возрастает с ростом f. При этом нет непосредственной зависимости от кривизны. Но, как было показано в гл. 5 (хотя и для нелокального случая), параметр сглаживания в идеале должен зависеть от f и f''.

3. Рекуррентные ядерные оценки

При рекуррентном определении оценок появляется два преимущества: данные не нужно запоминать и оценки легко перестраивать при получении новых данных. Последнее относится к случаю, когда представляют интерес только значения f в ряде фиксированных точек x. Едва ли следует ожидать, что простые рекуррентные аналоги ядерной оценки Парзена—Розенблата ведут себя так же хорошо, как исходная нерекуррентная оценка.

Чаще всего рассматривается оценка, определяемая соотношением

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} h_i^{-d} K\left(\frac{x - X_i}{h_i}\right). \tag{2}$$

Эта оценка введена в работе Wolverton, Wagner (1969a, b) и Yamato (1971) (теоретический анализ ее см. в работах Davies (1973), Deheuvels (1973a, b, 1974), Carroll (1976), Ahmad, Lin (1976), Devroye (1979), Wegman, Davies (1979), Györfi (1981)). Здесь h_n — последовательность положительных параметров сглаживания.

Deheuvels (1973a, b, 1974) предложил также другие обобщения, в частности

$$f_n(x) = \frac{\sum_{i=1}^{n} K\left((x - X_i)/h_i\right)}{\sum_{i=1}^{n} h_i^d}, \tag{3}$$

$$f_n(x) = \frac{\sum_{i=1}^{n} a_i K\left((x - X_i)/h_i\right)}{\sum_{i=1}^{n} a_i h_i^d}, \tag{4}$$

где $a_i = g(h_i)$ для некоторой функции g с положительными значениями. Один из его результатов состоит в том, что асимптотическая дисперсия оценки (3) меньше, чем для оценок (2) или (4) (при $a_i \neq 1$).

В работе Deheuvels (1979) найдены асимптотически оптимальные значения h_i и a_i при критерии интегральной среднеквадра-

тичной ошибки и показано, что если K — симметричная плотность на R^1, то оценка (2) оптимальна. Исходя из этого, ниже будут подробно рассмотрены оценки (2) и (3). Хотелось бы также привлечь внимание к общим классам оценок, введенным Banon (1976) и Rejtö, Révész (1973).

Если $a_i = 1/\sqrt{h_i^d}$, то оценка (4) близка к оценке из Wegman, Davies (1979). Наконец, хотелось бы отметить работы Isogai (1978, 1980, 1982), который рассмотрел оценки, определяемые рекуррентным соотношением

$$f_{n+1}(x) = f_n(x) + a_{n+1}\left(h_{n+1}^{-d} K\left(\frac{x - X_{n+1}}{h_{n+1}}\right) - f_n(x)\right), \qquad (5)$$

где $a_1 = 1$, $0 < a_n \leqslant 1$, $a_n \to 0$ и $\sum_{n=1}^{\infty} a_n = \infty$.

Он привел достаточные условия для различных типов состоятельности. Заметим, что при выборе $a_n = 1/n$ соотношение (5) определяет оценку (2).

Теорема 1. *Пусть K — ограниченная плотность с интегрируемой радиальной мажорантой (см. теорему 2.3), и пусть $\{h_n\}$ — последовательность положительных чисел. Пусть f_n — оценка (3). Тогда следующие утверждения эквивалентны:*

A. $f_n \to f$ *почти наверное при почти всех x для любых f;*

B. $f_n \to f$ *по вероятности при почти всех x для некоторой плотности f;*

C. $\sum_{n=1}^{\infty} h_n^d = \infty$ и $\lim_{n\to\infty} \sum_{i=1}^{n} h_i^d I_{[h_i > \varepsilon]} \Big/ \sum_{i=1}^{n} h_i^d = 0$ *при любых $\varepsilon > 0$;*

D. $\int |f_n - f| \to 0$ *почти наверное для любых f;*

E. $\int |f_n - f| \to 0$ *по вероятности для некоторой плотности f.*

Приведем полное доказательство теоремы 1. Доказательство взято по частям из работ Deheuvels (1973a, b, 1974) и Devroye (1979). Заметим, что утверждается эквивалентность слабой и сильной поточечной сходимости, которая для обычной ядерной оценки не имеет места (см. теорему 6.3).

Заметим также, что утверждение C следует из условий $h_n = o(1)$ и $\sum_{n=1}^{\infty} h_n^d = \infty$, но из него не следует, что $h_n = o(1)$. Доказательство теоремы 1 основано на нескольких важных леммах.

Лемма 1. *Для любой случайной величины X, имеющей абсолютные моменты $\mu_r = E(|X|^r)$, выполнено неравенство*

$$\mu_1 \geqslant \sqrt{\mu_2^3/\mu_4}.$$

Доказательство. При фиксированном $a > 0$ минимум функции $1/x + ax^2$ на $(0, \infty)$ достигается при $x^3 = 1/2a$. Таким образом,

$$\frac{x + ax^4}{x^2} \geqslant (2a)^{1/3} + a\,(1/2a)^{2/3} = (3/2)\,(2a)^{1/3}.$$

Отсюда после замены x на $|X|$ и взятия математических ожиданий получим

$$\mu_1 \geqslant (3/2)\,(2a)^{1/3}\,\mu_2 - a\mu_4.$$

Эта нижняя граница, рассматриваемая как функция от a, максимальна при $a = (\mu_2/\mu_4)^{3/2}/2$. Подстановка этого значения a в границу дает

$$\mu_1 \geqslant \mu_2^{3/2}/\mu_4^{1/2}.$$

Лемма 2. *Если K — ограниченная плотность с интегрируемой радиальной мажорантой и $h \to 0$, то*

$$f * K_h^p \to f \int K^p$$

при почти всех x и любых $p > 0$.

Доказательство. Эта лемма следует из теоремы 2.3.

Доказательство теоремы 1. В силу теоремы 2.8 имеем $A \Rightarrow B \Rightarrow E$ и $A \Rightarrow D \Rightarrow E$. Покажем сначала, что $C \Rightarrow A$, а затем, что $E \Rightarrow C$.

Положим

$$f_n^*(x) = \frac{\sum_{i=1}^n K\left((x - X_i)/h_i\right) I_{[h_i \leqslant \varepsilon]}}{\sum_{i=1}^n h_i^d}$$

при некотором $\varepsilon > 0$. Тогда в силу C

$$|f_n - f_n^*| \leqslant M \sum_{i=1}^n I_{[h_i > \varepsilon]} \Big| \sum_{i=1}^n h_i^d = o(1),$$

где M — верхняя граница для K. Кроме того,

$$|f - f_n^*| \leqslant \left| \sum_{i=1}^n V_i \right| \Big| \sum_{i=1}^n h_i^d \Big| + \\ + \frac{\sum_{i=1}^n h_i^d \left| f * K_{h_i} - f \right| I_{[h_i \leqslant \varepsilon]} + \sum_{i=1}^n h_i^d I_{[h_i > \varepsilon]} f}{\sum_{i=1}^n h_i^d},$$

где

$$V_i = h_i^d \left(h_i^{-d} K\left((x - X_i)/h_i\right) - f * K_{h_i}(x)\right) I_{[h_i \leqslant \varepsilon]}, \quad 1 \leqslant i \leqslant n,$$

— независимые случайные величины. Последнее слагаемое в правой части этого неравенства состоит из двух частей, одна из которых может быть сделана малой за счет выбора малого ε (ввиду леммы 2), а вторая равна o (1) при любых $\varepsilon > 0$ (в силу C) при почти всех x. Таким образом, утверждение A будет доказано, если показать, что $\sum_{i=1}^{n} V_i \Big/ \sum_{i=1}^{n} h_i^d \to 0$ почти наверное при почти всех x. Последнее имеет место, если

$$\sum_{n=1}^{\infty} h_n^d = \infty, \quad \sum_{n=1}^{\infty} \frac{E(V_n^2)}{\left(\sum_{i=1}^{n} h_i^d\right)^2} < \infty \tag{6}$$

(см., например, Loève (1963, с. 267 по русскому изданию). Но заметим, что можно выбрать ε так, что

$$E(V_n^2) \leqslant h_n^{2d} I_{[h_n \leqslant \varepsilon]} h_n^{-d} f * K_{h_n}^2 \leqslant h_n^d I_{[h_n \leqslant \varepsilon]} \left(f \int K^2 + 1\right)$$

при почти всех x по лемме 2. Следовательно, надо только проверить, что

$$\sum_{n=1}^{\infty} \frac{h_n^d}{\left(\sum_{i=1}^{n} h_i^d\right)^2} < \infty.$$

Но это неравенство имеет место, поскольку

$$\sum_{n=1}^{\infty} \frac{h_n^d}{\left(\sum_{i=1}^{n} h_i^d\right)^2} \leqslant h_1^{-d} + \sum_{n=2}^{\infty} \frac{h_n^d}{\left(\sum_{i=1}^{n} h_i^d \sum_{i=1}^{n-1} h_i^d\right)} =$$

$$= h_1^{-d} + \sum_{n=2}^{\infty} \left(\left(\sum_{i=1}^{n-1} h_i^d\right)^{-1} - \left(\sum_{i=1}^{n} h_i^d\right)^{-1}\right) = 2h_1^{-d} < \infty.$$

Тем самым доказана импликация C $\Rightarrow$ A.

Покажем теперь, что E $\Rightarrow \sum_{n=1}^{\infty} h_n^d = \infty$. Будем рассуждать от противного. Предположим, что эта сумма конечна и равна s. Тогда $h_n \to 0$ и из леммы 2 следует, что $|E(f_n) - f| \to 0$ при почти всех x. Далее, в силу леммы Фату и утверждения E имеем $0 = \lim\limits_{n\to\infty} \inf \int E(|f_n - f|) \geqslant \int \lim\limits_{n\to\infty} \inf E(|f_n - f|)$ и, следовательно,

$$\lim_{n\to\infty} \inf E(|f_n - E(f_n)|) = 0 \tag{7}$$

при почти всех x. Ввиду леммы 1 противоречие с (7) будет получено, если показать, что

$$\limsup_{n\to\infty} E(|f_n - E(f_n)|^4) < \infty \tag{8}$$

при почти всех x и

$$\liminf_{n\to\infty} E(|f_n - E(f_n)|^2) > 0 \tag{9}$$

при почти всех x. Положим $Y_i = K((x - X_i)/h_i)$, $Z_i = Y_i - E(Y_i)$. В силу леммы 2 $E(Y_i^r) \sim h_i^d f(x) \int K^r$ при любых $r > 0$ и почти всех x. Отсюда заключаем, что $E(Z_i^2) \sim E(Y_i^2) \sim h_i^d \times f(x) \int K^2$ при почти всех x и $E(Z_i^4) \sim E(Y_i^4) \sim h_i^d f(x) \int K^4$ при почти всех x. Чтобы проверить (8), заметим, что при почти всех x при $n \to \infty$

$$E(|f_n - E(f_n)|^4)\left(\sum_{i=1}^n h_i^d\right)^4 \sim s^4 E\left(\left(\sum_{i=1}^n Z_i\right)^4\right) \leqslant$$

$$\leqslant s^4 E\left(\sum_{i=1}^n Z_i^4 + 6\sum_{i=1}^n \sum_{j=1}^n Z_i^2 Z_j^2\right) \leqslant cs^8,$$

где c — некоторая постоянная, зависящая от x. Здесь мы воспользовались тем, что величины $E(Z_i^2)/h_i^d$ и $E(Z_i^4)/h_i^d$ ограничены равномерно по i при почти всех x.

Кроме того, $E(Z_i^2) \geqslant \left(f(x) h_i^d \int K^2\right)/2$ для любых $i \geqslant N(x)$ при почти всех x. Таким образом,

$$E(|f_n - E(f_n)|^2) \geqslant s^{-2} \sum_{i=N(x)}^n E(Z_i^2) \geqslant$$

$$\geqslant \frac{f(x)\int K^2}{2s^2} \sum_{i=N(x)}^n h_i^d \to \frac{f(x) K^2}{2s^2} \sum_{i=N(x)}^\infty h_i^d > 0$$

при почти всех x. Отсюда следует (9). Таким образом, должно выполняться первое условие из С.

Для доказательства второго условия из С воспользуемся характеристическими функциями. Пусть φ и ψ — характеристические функции для f и K. Так как $E\left(\int |f_n - f|\right) \to 0$ и $E\left(\int |f_n - f|\right) \geqslant \int |E(f_n) - f|$, то характеристическая функ-

ция $E(f_n)$ стремится к характеристической функции f. Характеристическая функция $E(f_n)$ равна

$$\varphi_n(t) = \frac{\sum_{i=1}^{n} h_i^d \varphi(t)\, \psi(h_i t)}{\sum_{i=1}^{n} h_i^d}, \quad t \in R^d.$$

Из E следует, что $\sum_{n=1}^{\infty} h_n^d = \infty$.

Предположим, что существуют положительные числа a и b, такие, что $\sum_{i=1}^{n} h_i^d I_{[h_i > a]} \Big/ \sum_{i=1}^{n} h_i^d \geqslant b$ для n, принадлежащих некоторой подпоследовательности натурального ряда. Так как

$$\varphi_n(t) - \varphi(t) = \varphi(t) \frac{\sum_{i=1}^{n} h_i^d (\psi(h_i t) - 1)}{\sum_{i=1}^{n} h_i^d} \to 0,$$

то

$$\sum_{i=1}^{n} h_i^d (\psi(h_i t) - 1) \Big/ \sum_{i=1}^{n} h_i^d \to 0$$

при достаточно малых t. Но действительная часть последнего выражения не превосходит

$$\sup_{h > a} (\mathrm{Re}(\psi(ht) - 1)) \frac{\sum_{i=1}^{n} h_i^d I_{[h_i > a]}}{\sum_{i=1}^{n} h_i^d} \leqslant -b \sup_{h \geqslant a} (\mathrm{Re}(\psi(ht) - 1))$$

для n из рассматриваемой подпоследовательности. Таким образом, $\sup_{h \geqslant a} \mathrm{Re}(\psi(ht)) = 1$ при всех достаточно малых t. В силу непрерывности ψ отсюда следует, что $\mathrm{Re}(\psi(ht)) = 1$ при некотором $h \geqslant a$, $t \neq 0$. Но последнее невозможно, так как ψ — характеристическая функция случайной величины, имеющей плотность. Таким образом, получено противоречие и, следовательно, E $\Rightarrow$ C.

Оценка (2) обладает другими сильными поточечными свойствами и сильными свойствами в L_1. По-видимому, все типы сходимости этой оценки в L_1 эквивалентны, но здесь такое утверждение не доказывается. Дается лишь кратная формулировка теоремы о достаточных условиях сходимости. Необходимость этих условий

доказана в Deheuvels (1974) при различных условиях регулярности на f, K и h_n. Во всей общности необходимость можно установить, используя методы теорем 3.1 и 7.1.

Теорема 2. *Пусть f_n — оценка (2), для которой K — ограниченная плотность с интегрируемой радиальной мажорантой. Условия*

$$\lim_{n\to\infty} h_n = 0, \quad \lim_{n\to\infty} nh_n^d = \infty \tag{10}$$

достаточны для сходимости по вероятности к 0 последовательности $|f_n - f|$ при почти всех x. Если, кроме того,

$$\lim_{n\to\infty} \frac{nh_n^d}{\log\log n} = \infty, \tag{11}$$

то $|f_n - f| \to 0$ почти наверное при почти всех x. Наконец, если, кроме того,

$$\lim_{n\to\infty} \frac{nh_n^d}{\log n} = \infty,$$

то $|f_n - f| \to 0$ вполне при почти всех x.

Доказательство. Введем случайные величины $Y_n = h_n^{-d} \times K((x - X_n)/h_n)$. В силу леммы 2 $E(Y_n) = f * K_{h_n} \to f$ при почти всех x, если $h_n \to 0$. Далее, $E(f_n) = n^{-1} \sum_{i=1}^{n} E(Y_i) \to f$ при почти всех x по лемме Теплица (Hall, Heyde (1980, с. 31)). Пусть $m_n = \inf_{i\leqslant n} h_i^d$ и M — верхняя грань $K(x)$ по всем x. Пусть $c = \sup_i E(Y_i)$ (эта величина зависит от x). Заметим, что $E(Y_i^2) \leqslant cM/h_i^d$ и $|Y_i - E(Y_i)| \leqslant M/h_i^d$. По неравенству Беннета, которое использовалось также в доказательстве леммы 6.5, при произвольном $\varepsilon > 0$ и любых x, для которых $c < \infty$, имеем

$$P(|f_n(x) - E(f_n(x))| \geqslant \varepsilon) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(Y_i - E(Y_i))\right| \geqslant \varepsilon\right) \leqslant$$
$$\leqslant 2\exp\left(-\frac{n\varepsilon^2}{2(cM/m_n + \varepsilon M/m_n)}\right).$$

Последнее выражение стремится к 0 при любых $\varepsilon > 0$, так как из (10) следует, что $nm_n \to \infty$ (см. лемму 3, которая приведена сразу после данного доказательства). Таким образом, сходимость по вероятности имеет место при почти всех x, и первое утверждение теоремы доказано.

Для доказательства сильной сходимости применим вариант усиленного закона больших чисел (Loève (1963, с. 267 по

русскому изданию)): если $|Y_n| \leqslant an$ при любых n и некотором $a < \infty$ (это условие в рассматриваемом случае выполнено, поскольку h_n — положительные числа и $nh_n^d/\log\log n \to \infty$), то $n^{-1}\sum_{i=1}^{n}(Y_i - E(Y_i)) \to 0$ почти наверное тогда и только тогда, когда

$$\sum_{k=0}^{\infty} P\left(\left|2^{-k}\sum_{i=2^k+1}^{2^{k+1}}(Y_i - E(Y_i))\right| \geqslant \varepsilon\right) < \infty$$

при любых $\varepsilon > 0$ (это утверждение называется также критерием сходимости Прохорова (1949)). Применяя снова неравенство Беннета, заключаем, что $f_n - E(f_n) \to 0$ почти наверное при почти всех x, если

$$\sum_{k=0}^{\infty} 2\exp\left(-\frac{2^k\varepsilon^2}{2\left(cM/m_{2^k} + \varepsilon M/m_{2^k}\right)}\right) < \infty.$$

Это неравенство выполнено, если $2^k m_{2^k}/\log k \to \infty$ при $k \to \infty$. Последнее в свою очередь следует из условия $nm_n/\log\log n \to \infty$ и, следовательно, из условия $nh_n^d/\log\log n \to \infty$ (см. лемму 3). Теорема 2 доказана.

Лемма 3. *Если* a_n, $b_n \geqslant 0$, $a_n \uparrow \infty$, *то* $a_n/b_n \to \infty$ *тогда и только тогда, когда* $a_n/\sup_{i\leqslant n} b_i \to \infty$.

Доказательство. Заметим, что

$$\frac{a_n}{b_n} \geqslant \frac{a_n}{\sup\limits_{i\leqslant n} b_i} \geqslant \min\left(\inf_{i>N}\frac{a_i}{b_i},\ \frac{a_n}{\sup\limits_{i\leqslant N} b_i}\right)$$

Чтобы получить утверждение леммы 3, нужно сначала выбрать достаточно большое N, а затем устремить n к бесконечности.

4. Оценки максимума правдоподобия

Согласно классическому принципу оценивания по максимуму правдоподобия, в рассматриваемой задаче оценкой максимума правдоподобия f_n является плотность g, максимизирующая функционал

$$\prod_{i=1}^{n} g(X_i). \tag{12}$$

Максимум этого функционала не достигается, если нет ограничений на класс допустимых g. Грубо говоря, g может стремиться к дискретному распределению с атомами в точках X_i. Имеется несколько способов борьбы с этим. Например, способ, предложенный Гренандером (Grenander (1981)), состоит в том, чтобы выбирать, g из подходящего множества плотностей C_n, которое может медленно расширяться с ростом n. Последовательность множеств C_n называется *решетом*, а соответствующий метод оценивания — *методом решета*. Ниже приводятся некоторые примеры решет. В последнем примере проводится связь метода решета с другими известными методами оценивания плотности по максимуму правдоподобия.

Пример 1 (гистограммная оценка). Если

$$C_n = \{g\colon\ g \text{ постоянна на } [(j-1)h_n,\ jh_n),\ j \text{ — целое}\},$$

где $h_n \to 0$ при $n \to \infty$, то максимум функционала достигается на гистограммной оценке с фиксированным шагом

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} I_{[X_i \in [(j-1)h_n,\ jh_n)]}, \quad x \in [(j-1)h_n,\ jh_n)$$

(см. Tapia, Thompson (1978, разд. 3.2)).

Пример 2 (сверточное решето). В работе Geman, Hwang (1982) предложено решето

$$C_n = \{g\colon\ g = K_{h_n} * \nu \text{ для какой-либо вероятной меры } \nu\},$$

где $h_n \to 0$ при $n \to \infty$ и K — плотность нормального распределения. Последовательность C_n называется *сверточным решетом*. Для выбора плотности нормального распределения нет никаких особых причин, кроме, быть может, удобства в теоретических рассмотрениях. Так, максимум функционала (12) на нормальном сверточном решете достигается для оценки вида

$$f_n(x) = \sum_{i=1}^{n} p_i \frac{1}{h_n} K\left(\frac{x - y_i}{h_n}\right),$$

где $(p_1, \ldots, p_n)$ — некоторый вероятностный вектор и $(y_1, \ldots, y_n)$ — некоторые действительные числа, заключенные строго внутри интервала $(\min X_i, \max X_i)$ (этот факт был доказан Гиманом и Маклюром; их доказательство можно найти в работе Geman (1981)). Заметим, что, хотя ядерная оценка Парзена—Розенблата принадлежит решету, она не является оптимальным решением.

Вычисление оптимальных значений $y_1, \ldots, y_n, p_1, \ldots, p_n$ является сложной задачей. Если $p_i = 1/n$ при всех i, то решение получить несколько проще, чем в общем случае. Однако следует иметь в виду, что вычислительные трудности неизбежны при на-

хождении любых оценок, основанных на максимуме правдоподобия. Так как решения получаются в неявном виде, то анализ их свойств также затруднителен. Результаты, касающиеся скорости сходимости, отсутствуют, но в литературе можно найти некоторые результаты о состоятельности оценок. Например, в случае нормального сверточного решета

$$C_n = \left\{ g\colon\ g(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - y_i}{h_n}\right) \text{ при некоторых } y_1, \ldots, y_n \in R^1 \right\}$$

в работе Geman (1981) показано, что

$$\sup_{\substack{\text{все} \\ \text{решения задачи} \\ \text{на максимум (12)} \\ \text{на } C_n}} \int |f_n - f| \to 0 \tag{13}$$

почти наверное при следующих условиях: $h_n \to 0$, $n^a h_n \to \infty$ при некотором $0 < a < 1/4$, f имеет компактный носитель и

$$\int f \log f < \infty. \tag{14}$$

Одно из этих условий означает, что множеству C_n не разрешается слишком быстро расширяться. Условие (14) на пики плотности f возникает естественным образом, поскольку задача максимизации функционала (12) эквивалентна максимизации функционала $\sum_{i=1}^{n} \log g(X_i)$, среднее значение которого равно $n \int f \log g$. Но, ввиду того что

$$\int f \log g \leqslant \int f \log f, \quad \forall f,\ g,$$

(это неравенство следует из неравенства Йенсена), можно надеяться что максимизация (12) дает функцию g, близкую к f, по крайней мере в случае, если применим какой-либо закон больших чисел, т. е. если имеет место (14). Этот метод удобен тем, что между его погрешностью и L_1 -ошибкой имеется некоторая слабая односторонняя связь, а именно:

$$\sqrt{2\left(\int f \log f - \int f \log g\right)} \geqslant \int |f - g|$$

(см. теорему 8.2).

Предположим теперь, что последовательность h_n также является параметром сверточного решета, подлежащим оптимизации. Тогда максимум функционала (12) достигается, если $h_n = 0$ и $y_i = X_i$ при любых i. Таким образом, если мы хотим максимизировать и по h_n, то нужны другие приемы. Например, в работе

Geman, Hwang (1982) предложено ограничивать сложность смеси, вводя решето

$$C_n = \left\{g:\ g(x) = \frac{1}{k_n}\sum_{i=1}^{k_n}\frac{1}{h}K\left(\frac{x-y_i}{h}\right) \text{ для каких-либо } h>0,\ y_1,\ldots,y_{k_n}\in R^1\right\}.$$

Для этого решета утверждение (13) остается в силе, если плотность f ограниченна и имеет компактный носитель и если $k_n \to \infty$ и $k_n/n^a \to 0$ при $n \to \infty$ для некоторого $a < 1/5$ (Geman, Hwang (1982)). К сожалению, остается еще задача выбора k_n.

Пример 3 (метод максимума правдоподобия со штрафами). Пусть $\mathcal{G}$ — подходящий класс плотностей и C_n определяется соотношением

$$C_n = \{g:\ g \in \mathcal{G},\ \Psi(g) \leqslant M\},$$

где M — постоянная, возможно зависящая от n, и $\Psi(g)$ — функция штрафа за колебательное поведение. Метод с таким решетом был предложен, но не исследован в работе Geman, Hwang (1982). Соответствующий ему метод множителей Лагранжа состоит в отыскании функции g из $\mathcal{G}$, максимизирующей функционал

$$\sum_{i=1}^{n}\log g(X_i) - \lambda_n\Psi(g), \tag{15}$$

где $\lambda_n \geqslant 0$ — множитель Лагранжа. Этот множитель играет роль параметра сглаживания. В качестве Ψ и $\mathcal{G}$ предлагаются различные функции и классы, например

(i) $\Psi(g) = \int (g')^2/g$, $\mathcal{G} = \{g:\ (\sqrt{g})' \in L_2\}$ (Good, Gaskins (1971); см. также Tapia, Thompson (1978, с. 108—109));

(ii) $\Psi(g) = a\int (g')^2 + b\int (g'')^2$, $\mathcal{G} = \{g: (\sqrt{g})' \in L_2,\ (\sqrt{g})'' \in L_2\}$,

$a \geqslant 0$, $b > 0$ (Good, Gaskins (1971)).

Снова основное место здесь занимает проблема вычислений: как найти максимумы по этим классам? Определенное решение этой проблемы, основанное на квантовании выборочного пространства, дается в работе Scott и др. (1980). Хотя оценка из этой работы состоятельна для больших классов плотностей f, к сожалению, экспериментально найденная для нее асимптотическая скорость сходимости в L_2 несколько хуже, чем для обычной ядерной оценки в ряде простых задач.

Для состоятельности необходимо, чтобы $\lambda_n \to 0$ с ростом n. Должны быть некоторые условия на скорость сходимости λ_n, так как при $\lambda_n = 0$ получается вырожденное решение. Различные условия на Ψ и f, обеспечивающие состоятельность, приведены в работах de Montricher (1980), Klonias (1982) и Silverman (1982).

5. Переменная гистограммная оценка

Гистограммная оценка, изученная в гл. 3 и 5, не обладает локальной чувствительностью — размер ячеек, по которым она строится, не может зависеть от x. Эта оценка есть не что иное, как частный случай *переменных гистограммных оценок*, определяемых следующим образом.

(i) Зададим счетное (возможно, конечное), разбиение P_{1n}, P_{2n}, ... пространства R^d. Это разбиение может зависеть от данных $X_1, \ldots, X_n$.

(ii) Оценим плотность f на P_{in} постоянной c_{in} таким образом, чтобы полученная оценка тоже была плотностью, т. е. чтобы было $c_{in} \geqslant 0$ при любых i и $\sum_i c_{in}\lambda(P_{in}) = 1$ (λ — мера Лебега). Обычно (хотя и не обязательно) $c_{in} = N_{in}/n\lambda(P_{in})$, где N_{in} — число точек выборки, попавших в P_{in}.

Таким образом, ядерная оценка с равномерным на $[-1, 1]$ ядром K классифицируется как переменная гистограммная оценка Но более существенно, что класс переменных гистограммных оценок достаточно широк, чтобы для некоторых f на нем достигалась асимптотическая скорость сходимости в L_1 порядка $n^{-2/5}$. Исходная гистограммная оценка с фиксированной решеткой в принципе не может обладать скоростью, лучшей, чем $n^{-1/3}$.

Наиболее популярным типом переменной гистограммной оценки, по-видимому, является оценка, основанная на статистически эквивалентных блоках [1]). Так, в R^1 рассматриваются порядковые статистики $X_{(1)}, \ldots, X_{(n)}$, соответствующие данным, и производится разбиение пространства на интервалы $P_{1n} = [X_{(1)}, X_{(k)}]$, $P_{2n} = (X_{(k)}, X_{(2k)}]$, ..., каждый из которых содержит k или менее точек выборки. В качестве оценки для $f(x)$ на множестве $x \in P_{in}$ принимается $k/n\lambda(P_{in})$. Оценка определена при любых x из $[X_{(1)}, X_{(n)}]$.

Эта оценка была в неявной форме предложена в работе Anderson (1965), формально определена и исследована в Van Ryzin (1970, 1973). Ее обобщения, представляющие собой гладкие сплайны (возможно, не являющиеся плотностями), можно найти в ра-

[1]) Используется также термин «полиграмма», см. Тарасенко (1976*). — *Прим. ред.*

ботах Wahba (1971, 1975, 1976). Ее состоятельность в L_1 была впервые доказана при очень общих условиях в Abou-Jaoude (1976), а скорость ее сходимости в L_1 была исследована в работе Hanna, Abou-Jaoude (1981).

Теорема 3 (Abou-Jaoude (1976)). *Для гистограммной оценки, основанной на порядковых статистиках и имеющей по k точек выборки в каждом интервале, следующие утверждения эквивалентны:*

A. $\int |f_n - f| \to 0$ *по вероятности для любых интегрируемых по Риману плотностей* f;

B. $\int |f_n - f| \to 0$ *вполне для любых интегрируемых по Риману плотностей* f;

C. $\lim_{n\to\infty} k = \infty$ *и* $\lim_{n\to\infty} (k/n) = 0$.

Существует много возможных обобщений рассматриваемой оценки на R^d. В одном из них (Gessaman (1970)), первая координатная ось разбивается приблизительно на $(n/k)^{1/d}$ интервалов, таких, что соответствующие им цилиндрические множества содержат примерно одинаковое число точек выборки (такое разбиение называется однородным). Каждое из полученных цилиндрических множеств снова подвергается однородному разбиению на $(n/k)^{1/d}$ частей, но теперь уже вдоль второй координатной оси. После d однородных разбиений каждая из n/k полученных «ячеек» содержит примерно k точек выборки. В работе Gessaman (1970) приводятся некоторые результаты о поточечной состоятельности и вполне справедливо обращается внимание на то, что оценку f_n можно быстро вычислить после того, как разбиение уже построено.

Другой способ — двигаясь по кругу, перебирать одну за другой координатные оси, разбивая их всякий раз так, чтобы оставшиеся данные делились ровно пополам, и прекращая этот процесс, когда в каждой из ячеек останется приблизительно по k точек выборки. С вычислительной точки зрения этот метод обладает тем преимуществом, что его можно реализовать в виде сбалансированного двоичного дерева, число уровней которого приблизительно равно $\log_2 (n/k)$.

6. Ядерные оценки с пониженным смещением

В некоторых случаях модификации обычных оценок плотности обладают лучшими скоростями сходимости ошибки $E(J_n)$, чем полученные в гл. 5 скорости сходимости для ядерной и гистограммной оценок. Улучшение, как правило, происходит из-за понижения смещения $\int |E(f_n) - f|$, и оно возможно только

для очень гладких f. В этом параграфе будут указаны некоторые общие методы понижения смещения. Эти методы будут продемонстрированы на примере ядерной оценки.

В дальнейшем предполагается, что f_n — оценка плотности f (f_n сама является плотностью по x) и g_n — другая функция на R^d. *Аддитивная модификация* f_n определяется следующим образом:

$$f_n^* = \frac{(f_n + g_n)_+}{\int (f_n + g_n)_+}; \quad \int g_n = 0, \quad \int |g_n| < \infty,$$

при любых n. Нормировка необходима здесь для того, чтобы функция f_n^* была плотностью. *Мультипликативная модификация* f_n определяется соотношением

$$f_n^* = \frac{f_n g_n}{\int f_n g_n}; \quad g_n \geqslant 0,$$

при любых n. Если речь идет о сходимости в L_1, то нормировками можно пренебречь в силу следующей леммы.

Лемма 4. *Для любых плотностей f, любых оценок плотности f_n на R^d и любой аддитивной модификации выполнено неравенство*

$$\int |f_n^* - f| \leqslant \int |(f_n + g_n) - f|.$$

Точно так же для любой мультипликативной модификации выполнено неравенство

$$\int |f_n^* - f| \leqslant \int |f_n g_n - f| + \left| \int f_n g_n - 1 \right|.$$

Доказательство. По поводу аддитивной модификации сошлемся на теорему 11.4 о неотрицательной проекции. В случае мультипликативной модификации рассуждаем следующим образом. Если $\int f_n g_n \leqslant 1$, то $f_n^* \geqslant f_n g_n$, так что

$$\int |f_n^* - f| = 2 \int (f - f_n^*)_+ \leqslant 2 \int (f - f_n g_n)_+ =$$

$$= \int |f - f_n g_n| + \left(1 - \int f_n g_n\right).$$

Если $\int f_n g_n \geqslant 1$, то $f_n^* \leqslant f_n g_n$, и поэтому

$$\int |f_n^* - f| = 2 \int (f_n^* - f)_+ \leqslant 2 \int (f_n g_n - f)_+ =$$

$$= \int |f_n g_n - f| + \left(\int f_n g_n - 1\right).$$

Основным примером оценки с пониженным смещением является оценка Бартлета (Bartlett (1963)) на R,

$$f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

где K — борелевская функция со следующими свойствами:

(i) K — симметричная ограниченная функция с компактным носителем;

(ii) $\int K = 1$;

(iii) $\int x^{2i} K = 0, \quad i = 1, 2, \ldots, s-1$;

(iv) $\int |x|^{2s} |K| < \infty$; (16)

где $s \geqslant 1$ — фиксированное целое число. При $s > 1$ ядро K, а потому и оценка f_n могут принимать отрицательные значения. Легко проверить, что оценка f_n представима в виде суммы обычной ядерной оценки с ядром $K_+ / \int K_+$ и некоторой функции g_n, интеграл от которой равен нулю. Следовательно, плотность $f_n^* = (f_n)_+ / \int (f_n)_+$ в действительности является аддитивной модификацией обычной ядерной оценки. Чтобы избежать неоднозначности, будем называть f_n *оценкой Бартлета*, а f_n^* — *нормированной оценкой Бартлета*; см. § 5.9, где эта оценка рассмотрена с точки зрения, отличной от принятой здесь.

Теорема 4. *Пусть* $h \to 0$ *и* $nh \to \infty$. *Тогда при любом целом* $s \geqslant 1$ *оценка Бартлета экспоненциально сходится, т. е. для любого* $\varepsilon > 0$ *существуют положительные числа* r *и* n_0, *такие, что*

$$P\left(\int |f_n - f| > \varepsilon\right) \leqslant e^{-rn}, \quad n \geqslant n_0.$$

Это свойство выполняется при любых f, *но* r *можно выбрать не зависящим от* f. *Теорема* 4 *остается в силе для нормированной оценки Бартлета.*

Доказательство. Теорема 4 следует непосредственно из теоремы 3.1, в которой на K накладывались лишь следующие два условия: $\int K = 1$, $\int |K| < \infty$. В случае нормированной оценки Бартлета воспользуемся леммой 4.

Теорема 5. *Пусть* $s \geqslant 1$ — *фиксированное целое число и* $h \to 0$, $nh \to \infty$. *Пусть* f — *плотность с компактным носителем, имеющая* $2s - 1$ *абсолютно непрерывных производных и непрерывную*

производную $f^{(2s)}$. *В обозначениях гл. 5 для оценки Бартлета выполнено соотношение*

$$E(J_n) \sim \int \frac{\alpha \sqrt{f}}{\sqrt{nh}} \psi\left(\frac{h^{2s}}{(2s)!} \beta_{2s} |f^{(2s)}|\right),$$

причем для нее и для нормированной оценки Бартлета —

$$E(J_n) \lesssim \frac{\alpha \int \sqrt{f}}{\sqrt{nh}} \sqrt{\frac{2}{\pi}} + \frac{h^{2s}}{(2s)!} \beta_{2s} \int |f^{2s}|.$$

Здесь $\alpha = \sqrt{\int K^2}$ *и* $\beta_{2s} = \int |x|^{2s} K \neq 0$. *Кроме того,*

$$\limsup_{n\to\infty} \inf_h n^{2s/(4s+1)} E(J_n) \leqslant C_{2s} A_{2s}(K) D_{2s}(f),$$

где

$$C_{2s} = \frac{1+4s}{(2s)!} \left(\frac{2}{\pi}\right)^{2s/(4s+1)} \left(\frac{(2s)!}{4s}\right)^{4s/(4s+1)},$$

$$A_{2s}(K) = (\alpha^{4s} |\beta_{2s}|)^{1/(4s+1)} = \left(\left(\int K^2\right)^{2s} \left|\int x^{2s} K\right|\right)^{1/(4s+1)}$$

и

$$D_{2s}(f) = \left(\left(\int \sqrt{f}\right)^{4s} \int |f^{(2s)}|\right)^{1/(4s+1)}.$$

Данная верхняя граница не нарушается, если выбрать

$$h = \left(\sqrt{\frac{2}{\pi}} \frac{\alpha (2s-1)!}{2|\beta_{2s}|} \frac{\int \sqrt{f}}{\int |f^{(2s)}|}\right)^{2/(4s+1)} n^{-1/(4s+1)}.$$

Если $\beta_{2s} = 0$ *и* $D_{2s}(f) < \infty$, *то*

$$\limsup_{n\to\infty} \inf_h n^{2s/(4s+1)} E(J_n) = 0.$$

Если $\beta_{2s} = 0$ *и* f — *произвольная плотность с компактным носителем, такая, что* $B^*(f) < \infty$ (*в обозначениях теоремы* 5.1), *то*

$$\limsup_{n\to\infty} \inf_h n^{2/5} E(J_n) = 0.$$

Величина $D_{2s}(f)$ из верхней границы теоремы 5 присутствует также в минимаксных нижних границах теорем 4.2 и 4.3. Более существенно, что скорость сходимости для отдельных f с конечными значениями $D_{2s}(f)$ может быть гораздо лучше, чем скорость $n^{-2s/(4s+1)}$, даваемая минимаксной нижней границей, если ядро K просто удовлетворяет условиям (16) *и* $\beta_{2s} = \int x^{2s} K = 0$. Это очевидное противоречие объясняется тем, что улучшение не равномерно на классе всех f компактным носителем, таких, что

$D_{2s}(f) \leqslant r$, где r — фиксированная постоянная. В действительности на этом классе достигаются все медленные скорости сходимости последовательности $n^{2s/(4s+1)}E(J_n)$ к нулю. Однако в последнем утверждении теоремы 5 нет указания на то, как выбирать h. Из доказательства видно, что лучше выбирать $h = Mn^{-1/(4s+1)}$, чем $h = M^*n^{-1/(4s+1)}$, если $M > M^*$, но, кроме этого, практически ничего нельзя сказать без дополнительных предположений относительно f. Такая же проблема возникает для K, и потому, видимо, имеет смысл выбирать h и K так, чтобы минимизировать минимаксную границу (такую, как в теореме 5.12). Например, в теореме 5 можно было бы получить грубую, но удобную для анализа верхнюю границу, заменяя β_{2s} (в определении h и $A_{2s}(K)$) на $\int x^{2s}|K|$, и выбрать K из условия минимума $A_{2s}(K)$. Отметим здесь, что ядра

$$K(x) = (9/8)(1 - 5x^2/3), \quad |x| \leqslant 1 \text{ (Bartlett (1963))},$$
$$K(x) = (3/2)(1 - x^2/3)e^{-x^2/2} \text{ (Rosenblatt (1971))}$$

удовлетворяют пп. (ii), (iii) условия (16) и таковы, что $\int x^{2s}K = 0$ при $s = 1$. Легко построить и другие такие ядра (Deheuvels (1977)).

Выбор ядра определяется предполагаемой степенью гладкости f. Однако могут учитываться и другие соображения. Например, в некоторых приложениях требуется одновременно оценивать плотность f и одну или несколько ее производных. Если производные плотности f оцениваются соответствующими производными оценки f_n, то очевидно, что ядро K должно быть гладким. Об этом см. работы Müller (1984) и Gasser и др. (1983). Заметим, наконец, что если $h \sim n^{-1/(4s+1)}$ и при этом f не обладает требуемой в теореме 5 гладкостью, то на самом деле ухудшение асимптотических свойств оценки может быть существенным.

Доказательство теоремы 5. Существенное отличие от теоремы 5.1 заключается в выражении для смещения. Используя разложение f в ряд Тейлора в окрестности точки x,

$$f(y) = \sum_{i=0}^{2s}(y-x)^i \frac{f^{(i)}(x)}{i!} + \frac{(y-x)^{2s}}{(2s)!}(f^{(i)}(\xi) - f^{(i)}(x)) \quad y \geqslant \xi \geqslant x,$$

и симметричное разложение при $y < x$, получим

$$\left| \int \frac{1}{h} K\left(\frac{x-y}{h}\right)(f(y) - f(x))\,dy - \int \frac{h^{2s}}{(2s)!}\beta_{2s}f^{(2s)}(x) \right| \leqslant$$

$$\leqslant \int \frac{1}{h} K\left(\frac{x-y}{h}\right) \frac{|y-x|^{2s}}{(2s)!} |f^{2s}(\xi) - f^{(2s)}(x)|\,dy,$$

где ξ зависит от x, y и f. Иными словами, в силу условия (16) все члены разложения в ряд Тейлора, кроме $2s$-го, исчезают после свертки с K_h. Правая часть последнего неравенства равна $o(h^{2s})$ равномерно по x из некоторого большого интервала T (в силу равномерной непрерывности $f^{(2s)}$ и компактности носителя K) и равна нулю вне T. Отсюда следует, что вместо леммы 5.11 имеем утверждение

$$\int \left| |B_n| - \frac{h^{2s}}{(2s)!} |\beta_{2s} f^{(2s)}| \right| = o(h^{2s}),$$

где, так же как в гл. 5, $B_n = E(f_n) - f$.

Очевидно, что лемма 5.10 остается в силе, так что можно дословно повторить доказательство теоремы 5.1, заменяя там z на $h^{2s} |\beta_{2s}| \int |f^{(2s)}|/(2s)!$. Тем самым доказано первое утверждение теоремы 5.

Заметим, что асимптотическая верхняя граница имеет вид $uh^{-1/2} + vh^{2s}$, где u, v — некоторые положительные числа, не зависящие от h. Формальной минимизацией по h получим, что минимальное значение этой границы равно

$$(u^{4s}v)^{1/(4s+1)} (4s+1)/(4s)^{4s/(4s+1)}.$$

Это значение достигается при $h = (u/4sv)^{2/(4s+1)}$. Заменяя u на $\alpha \int \sqrt{f}\sqrt{2/\pi n}$ и v на $|\beta_{2s}| \int |f^{(2s)}|/(2s)!$, получим указанную в теореме верхнюю границу.

В случае $\beta_{2s} = 0$, $D_{2s}(f) < \infty$ положим $h = Mn^{-1/(4s+1)}$, где M — сколь угодно большое число, и заметим, что $\int |B_n| = o(n^{-2s/(4s+1)})$. Кроме того, в обозначениях леммы 5.10 имеем

$$\frac{\int \sigma_n}{\sqrt{\int K^2}} \leqslant \frac{\int \sqrt{f} + o(1)}{\sqrt{nh}} \leqslant \frac{\int \sqrt{f} + o(1)}{\sqrt{M}} n^{-2s/(4s+1)}.$$

Теорема доказана.

В качестве второго примера, частично перекрывающегося с предыдущим, укажем испытанный *метод складного ножа*, применяемый в статистике для понижения смещения оценок (Quenouille (1956); см. также Schucany и др. (1971)). Для оценивания плотности этот метод был впервые разработан и продемонстрирован в действии в работах Sommers (1972) и Schucany, Sommers (1977).

Пусть $f_{n1}, \ldots, f_{nM}$ суть M ядерных оценок плотности f, основанных на одной и той же выборке $X_1, \ldots, X_n$, но, возможно, имеющих различные ядра $K_1, \ldots, K_M$ и различные параметры сглаживания $h_1, \ldots, h_M$. Предположим, что $h_i = ha_i$, где a_i —

некоторые постоянные, а h зависит только от n. Рассмотрим линейную комбинацию

$$f_n = \frac{\sum_{i=1}^{M} b_i f_{ni}}{\sum_{i=1}^{M} b_i},$$

где $b_1, \ldots, b_M$ — постоянные, сумма которых не равна 0. Ясно, что если интегралы от всех ядер равны единице, то интеграл от f_n также равен единице. Таким образом, в нормированном виде эта оценка снова представляет собой аддитивную модификацию обычной ядерной оценки. Нетрудно проверить, что f_n совпадает с оценкой Бартлета с ядром

$$K(x) = \frac{\sum_{i=1}^{M} b_i \left((1/a_i) K_i (x/a_i)\right)}{\sum_{i=1}^{M} b_i}$$

при надлежащем выборе $b_1, \ldots, b_M$, и потому к f_n применимы теоремы 4 и 5. Пусть теперь все K_i удовлетворяют только п. (i) условия Бартлета (16) (и не удовлетворяют пп. (ii), (iii), (iv)), все K_i являются плотностями и f удовлетворяет условиям теоремы 5. Тогда для поточечного смещения имеем следующее разложение в ряд Тейлора:

$$E(f_{ni}) - f = \sum_{j=1}^{s} \frac{h^{2j}}{(2j)!} f^{(2j)} \int x^{2j} K_i + o(h^{2s}).$$

Первые $s-1$ членов в смещении оценки f_n равны нулю, если

$$\sum_{i=1}^{M} b_i a_i^{2j} \int x^{2j} K_i = 0, \quad j = 1, 2, \ldots, s-1.$$

Эта система уравнений имеет много степеней свободы. Например, можно положить все K_i равными K и a_i — равными i. Тогда рассматриваемая система уравнений принимает вид

$$\sum_{i=1}^{M} b_i i^{2j} = 0, \quad j = 1, 2, \ldots, s-1.$$

Данная система имеет ненулевые решения относительно b_i при $M \geqslant s$. Так, при $s = M = 2$ решением является $b_1 = 1$, $b_2 = -1/4$ и ему соответствует *оценка складного ножа*

$$f_n = (4f_{n1} - f_{n2})/3 = f_{n1} + (f_{n1} - f_{n2})/3.$$

Последняя запись наводит на мысль о том, что $(f_{n1} - f_{n2})/3$ является смешанным поправочным членом для обычной ядерной оценки f_{n1}.

В качестве далеко не единственного примера мультипликативной модификации обычной ядерной оценки рассмотрим *оценку Террела — Скотта* (Terrell, Scott (1980)). Пусть есть две ядерные оценки f_{n1}, f_{n2} с одним и тем же ограниченным симметричным ядром K, имеющим компактный носитель, и параметрами сглаживания h и $2h$ соответственно. Обе оценки по-прежнему основаны на одной и той же выборке $X_1, \ldots, X_n$. Образуем оценку

$$f_{n1}\left(\frac{f_{n1}}{f_{n2}}\right)^{1/3}.$$

Если ядро K выбрано унимодальным, то поправочный множитель этой оценки всегда корректно определен и заключен между 0 и $2^{1/3}$. Доказательство состоятельности данной оценки стоит определенного труда, поскольку требуется установить сходимость

$$\int \left| f_{n1}\left(\frac{f_{n1}}{f_{n2}}\right)^{1/3} - f \right| \to 0.$$

Но левая часть здесь не превосходит

$$2^{1/3}\int |f_{n1} - f| + \int f \left| \left(\frac{f_{n1}}{f_{n2}}\right)^{1/3} - 1 \right|.$$

Таким образом, в силу теорем 3.1, 6.3 и теоремы Лебега о мажорируемой сходимости получается следующий результат.

Теорема 6. *Пусть K — симметричная ограниченная унимодальная плотность на R с компактным носителем, и пусть $h \to 0$, $nh \to \infty$. Тогда для любых плотностей f*

$$\int |f_n^* - f| \to 0$$

по вероятности при $n \to \infty$, где f_n^* — нормированная оценка Террела—Скотта. То же самое верно при любых поправочных множителях g_n [1], равномерно ограниченных снизу нулем и сверху некоторой положительной постоянной и сходящихся поточечно почти всюду к 1.

Существует вариант теоремы 6 со сходимостью почти наверное, где требуется только добавить условие $nh/\log\log n \to \infty$. В работе Terrell, Scott (1980) показано, что если плотность f — достаточно гладкая, то смещение равно $O(h^4)$ в каждой точке x, а дисперсия равна $O((nh)^{-1})$ в каждой точке x. По-видимому, оценка Террела и Скотта асимптотически ведет себя как оценка Бартлета при $s =$

[1]) То есть теорема 6 верна для оценок вида $f_n^* = f_{n1}g_n \Big/ \int f_{n1}g_n$, где g_n удовлетворяет перечислеиным далее условиям. — *Прим. перев.*

$= 2$, но строго это не доказано. Основное преимущество мультипликативных модификаций перед аддитивными заключается в их хорошем поведении на хвостах — аддитивные модификации ядерной оценки часто спадают резко к нулю из-за нормировки. Нормировка мультипликативной модификации ядерной оценки является менее жесткой.

7. Оценка Гренандера для монотонных плотностей

В этом параграфе будут рассматриваться только монотонные плотности на $[0, \infty)$ (класс всех таких плотностей обозначается через M) и монотонные плотности на $[0, 1]$, такие, что $f(0) \leqslant B$ (множество таких плотностей обозначается через M_B). Выше было показано, что минимаксная нижняя граница для M больше или равна 1/8 (теорема 4.1), но минимаксная нижняя граница для M_2 равна $(1/16 + o(1))\,(4/n)^{1/3}$ (теорема 4.9). Люсьен Бирже доказал, что при $B \geqslant 1{,}3$, $B/n \leqslant 0{,}026$ существует минимаксная нижняя граница для M_B, равная 0,198 $(\log(B+1)/n)^{1/3}$ (это неопубликованный результат, и потому он не включен в данную книгу). Учитывая перечисленные результаты, можно сравнивать различные оценки. Конечно, нет смысла рассматривать оценки, состоятельные при любых f, поскольку представляют интерес только плотности f из классов M и M_B. Одной из оценок, предназначенных для этих классов, является *оценка максимума правдоподобия Гренандера* (Grenander (1956)). Она представляет собой плотность из M, для которой максимально произведение

$$\prod_{i=1}^{n} f_n(X_i).$$

Оптимизационная задача отыскания такой плотности имеет замечательно простое решение: если $F_n(x) = (1/n)\sum_{i=1}^{n} I_{[X_i \leqslant x]}$ — эмпирическая функция распределения, то $f_n = G_n'$, где G_n — наименьшая вогнутая мажоранта для F_n (т. е. G_n — кусочно линейная функция, которая получается, если взять огромную эластичную ленту, обвязать ею первый квадрант и отпустить, зафиксировав ее часть, лежащую вдоль оси x, — в результате лента окажется натянутой на график F_n).

Глубокий анализ поточечных свойств этой оценки проведен в работе Prakasa-Rao (1969). Однако основной интерес для нас представляет поразительно красивый результат из работы Groeneboom (1983):

$$n^{1/6}\left(n^{1/3}\int |f_n - f| - C(f)\right) \xrightarrow{\mathscr{L}} N(0, \sigma^2),$$

где символом $\xrightarrow{\mathscr{L}}$ обозначается сходимость по распределению, $N(0, \sigma^2)$ — нормально распределенная случайная величина с параметрами $(0, \sigma^2)$, $\sigma > 0$ — постоянная, не зависящая от f, и

$$C(f) = c_0 \int_0^1 (|f'|f/2)^{1/3},$$

где c_0 — еще одна универсальная постоянная (из экспериментальных результатов, сообщенных Грёнебомом, следует, что она приблизительно равна 0,82). Результат Грёнебома справедлив для любых строго убывающих функций f на $[0, 1]$, имеющих непрерывную и ограниченную вторую производную и таких, что $f' < 0$ на $(0, 1)$. Для таких f имеем

$$n^{1/3} E\left(\int |f_n - f|\right) \to c_0 \int (|f'|f/2)^{1/3}.$$

Поведение оценки Гренандера еще лучше в случае, когда f имеет участки постоянства. Для равномерной плотности на $[0, 1]$ достигается скорость сходимости $n^{-1/2}$.

Сравним этот результат, полученный для фиксированных f, с результатами о ядерных оценках. Для плотностей f, удовлетворяющих условиям теоремы Грёнебома,

$$\int (|f'|f/2)^{1/3} \leqslant \left(\left(\int \sqrt{f}\right)^2 \left(\int |f'|\right)/2\right)^{1/3} = B_H(f).$$

Мы знаем, что для ядерной оценки с параметром сглаживания h, выбранным согласно теореме 5.10, выполнено неравенство $n^{1/3} E\left(\int |f_n - f|\right) \leqslant (1{,}24 \ldots + o(1))\, B_H(f)$. Это, как видим, немного хуже, чем для оценки Гренандера. С другой стороны, следующая теорема показывает, что соответствующим образом подправленная в окрестности нуля ядерная оценка имеет скорость сходимости $n^{-1/3}$ при любых f из M_B.

Теорема 7. *Пусть f — произвольная плотность из M_B и параметр сглаживания h выбран следующим образом:*

$$h = \left(\frac{6}{\pi n B^2}\right)^{1/3}.$$

Пусть K — равнобедренная треугольная плотность на $[-1, 1]$. Пусть, далее, g_n — ядерная оценка, построенная по $Y_1, \ldots, Y_n$, где величина Y_i равна X_i со случайным знаком (определяемым бросанием монеты), и оценка f_n задана соотношением

$$f_n(x) = g_n(x) + g_n(-x), \quad x > 0.$$

Тогда

$$\limsup_{n\to\infty} n^{1/3} E\left(\int |f_n - f|\right) \leqslant (6/\pi)^{1/3} B^{1/3}.$$

Доказательство. Имеем

$$\int |f_n - f| \leqslant \int_0^\infty |g_n(x) - f(x)/2|\,dx + \int_0^\infty |g_n(-x) - f(x)/2|\,dx \leqslant$$

$$\leqslant \int_{-\infty}^{\infty} |g_n(x) - f(|x|)/2|\,dx.$$

Теперь применим теорему 5.10 непосредственно к величине последней L_1-ошибки. Эта довольно искусственная симметризация проведена для того, чтобы обеспечить равномерную ограниченность величины $B_H^*(f)$ на рассматриваемом классе плотностей. Заметим, что если выбрать в качестве φ из определения $B_H^*(f)$ симметричную и унимодальную функцию, то функция $f * \varphi_a$ совсем не обязательно будет унимодальной. Вместе с тем если $g(x) = f(|x|)/2$, то функция $g * \varphi_a$ действительно унимодальна (Feller (1971)), так что $\int |(g * \varphi_a)'| = 2g * \varphi_a(0) = B$ при любых a. Ввиду того что $\int \sqrt{g} = \sqrt{2} \int \sqrt{f}$, имеем $B_H^*(g) = \left(\left(\int \sqrt{f}\right)^2 B\right)^{1/3} \leqslant B^{1/3}$. Оставшаяся часть доказательства следует непосредственно из теоремы 5.10.

Для ядерной оценки ситуация даже более благоприятна, так как если плотность f удовлетворяет условиям теоремы Грёнебома, то можно выбрать h и K таким образом, что величина $n^{2/5} E\left(\int |f_n - f|\right)$ стремится к постоянной. Чтобы получить этот результат, нужно обязательно воспользоваться приемом симметризации из теоремы 7 (иначе результат невозможен, так как имеется разрыв в нуле). Иными словами, для отдельных плотностей подходящим образом модифицированная ядерная оценка может быть гораздо лучше оценки Гренандера.

Также имеет место следующая минимаксная верхняя граница, которая достигается на модифицированной ядерной оценке.

Теорема 8. *Для оценки f_n из теоремы 7*

$$\limsup_{n\to\infty} n^{1/3} \sup_{f \in M_B} E\left(\int |f_n - f|\right) \leqslant (6/\pi)^{1/3} B^{1/3}$$

при любых $B \geqslant 1$.

Доказательство. Достаточно убедиться в равномерности члена $o((nh)^{-1/2})$ из теоремы 5.10 по классу симметричных унимодальных плотностей на $[-1, 1]$, ограниченных величиной $B/2$. Остальные рассуждения такие же, как в доказательстве теоремы 7.

К сожалению, эта граница не самая лучшая, хотя для M_2 отношение минимаксной верхней границы к нижней всего лишь около 16. Бирже (Birgé (1985)) показал, что минимаксная верхняя граница для M_B меньше, чем $1{,}89\,(\log(B+1)/n)^{1/3}$ при $B \geqslant 1{,}3$, $B/n \leqslant 0{,}026$. Очень правдоподобно, что эта улучшенная минимаксная верхняя граница достигается на оценке Гренандера. Это предположение основано на следующем факте.

Лемма 5. *Пусть f — абсолютно непрерывная плотность из класса M_B, имеющая почти всюду производную f'. Тогда*

$$\int (|f'|/f)^{1/3} \leqslant 1 + (\log B)^{1/3}.$$

Доказательство. Пусть точка $u \in [0, 1]$ такова, что $f(x) \geqslant 1$, $x < u$ и $f(x) \leqslant 1$, $x > u$. Тогда по неравенству Йенсена

$$\int (|f'|/f)^{1/3} \leqslant \int_0^u (|f'|/f)^{1/3} + \int_u^1 |f'|^{1/3} \leqslant \int_0^u f\,(|f'|/f^2)^{1/3} + f(u)^{1/3} \leqslant$$

$$\leqslant \left(\int_0^u f\,(|f'|/f^2)\right)^{1/3} + 1 = \left(-\int_0^u d(\log f)\right)^{1/3} + 1 = (\log B)^{1/3} + 1.$$

В работе Birgé (1985) построена гистограммная оценка с неодинаковыми длинами интервалов, для которой равномерная верхняя граница на M_B не превосходит $1{,}89\,(\log(B+1)/n)^{1/3} + 0{,}20\,(\log(B+1)/n)^{2/3}$.

В заключение приведем несколько ссылок относительно других оценок, предназначенных для класса всех унимодальных плотностей. Здесь укажем, например, работы Robertson (1967) и Wegman (1969, 1970, 1975).

ЛИТЕРАТУРА

Abou-Jaoude S.
(1976) Sur la convergence L_1 et L_∞ de l'estimateur de la partition aléatoire pour une densité, Annales de l'Institut Henri Poincaré, B 12, pp. 299—317.
Abramson I. S.
(1982) On bandwidth variation in kernel estimates — a square root law, Annals of Statistics, 10, pp. 1217—1223.
Ahmad I. A., Lin P.
(1976) Nonparametric sequential estimation of a multiple regression function, Bulletin of Mathematical Statistics, 17, pp. 63—75.
Anderson T. W.
(1965) Some nonparametric multivariate procedures based on statistically equivalent blocks, in Multivariate Analysis I, P. R. Krishnaiah (Ed).
Banon G.
(1976) Sur un estimateur non paramétrique de la densité de probabilité, Revue de Statistique Appliquée, 24, pp. 61—73.

Bartlett M. S.
(1963) Statistical estimation of density functions, Sankhya, Series A, 25, pp. 245—254.

Birgé L.
(1985a) Estimating a density under order restrictions. Technical Report, UER de Sciences Economiques, Université Paris X, Nanterre.
(1985b) On the risk of histograms for estimating decreasing densities. Technical Report, UER de Sciences Economiques, Université Paris X, Nanterre.

Breiman L., Meisel W., Purcell E.
(1977) Variable kernel estimates of multivariate densities, Technometrics, 19, pp. 135—144.

Carrol R. J.
(1976) On sequential density estimation, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 36, 136—151.

Davies H. I.
(1973) Strong consistency of a sequential estimator of a probability density function, Bulletin of Mathematical Statistics, 15, pp. 49—53.

Deheuvels P.
(1973a) Sur une famille d'estimateurs de la densité d'une variable aléatoire, Comptes Rendus de l'Académie des Sciences de Paris, 276, pp. 1013—1015.
(1973b) Sur l'estimation séquentielle de la densité, Comptes Rendus de l'Academie des Sciences de Paris, 276, pp. 1119—1121.
(1974) Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, Comptes Rendus de l'Académie des Sciences de Paris, 278, pp. 1217—1220.
(1977) Estimation nonparamétrique de la densité par histogrammes généralisés, Revue de Statistique Appliquée, 25, pp. 5—42.
(1979) Estimation séquentielle de la densité, Contrib. en Prob. y Est. Mat. Ens. de la Mat. y Análisis, pp. 156—169, University of Granada.

De Montricher G. M.
(1980) On the consistency of maximum penalized likelihood density estimation, Technical Report, Department of Mathematics, Rice University, Houston, Texas.

De Montricher G. M., Tapia R. A., Thompson J. R.
(1975) Nonparametric maximum likelihood estimation of probability densities by penalty function methods, Annals of Statistics, 3, pp. 1329—1348.

Devroyé L.
(1979) On the pointwise and the integral convergence of recursive kernel estimates of probability densities, Utilitas Mathematica, 15, pp. 113—128.
(1985) A note on the L_1 consistency of variable kernel estimates, Annals of Statistics, 13, pp. 1041—1049.

Feller W.
(1971) An Intoduction To Probability Theory and Its Applications, Wiley, New York. (Имеется русский перевод: Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. — М.: Мир, 1984.)

Gasser T., Müller H.-G., Mammitzsch V.
(1983) Kernels for nonparametric curve estimation, Sonderforschungsbereich 123, Stochastische Mathematische Modelle, Preprint 210, Universität Heidelberg.

Geman S.
(1981) Sieves for nonparametric estimation of densities and regressions, Reports in Pattern Analysis No. 99, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

Geman S., Hwang C.-R.
(1982) Nonparametric maximum likelihood estimation by the method of sieves, Annals of Statistics, 10, pp. 401—414.

Gessaman M. P.
(1970) A consistent nonparametric multivariate density estimator based on statistically equivalent blocks, Annals of Mathematical Statistics, 41, pp. 1344—1346.

Good I. J., Gaskins R. A.
(1971) Nonparametric roughness penalties for probability densities, Biometrika, 58, pp. 255—277.
Grenander U.
(1956) On the theory of mortality measurement. Part II, Skandinavisk Aktuarietidskrift, 39, pp. 125—153.
(1981) Abstract Inference, Wiley, New York.
Groeneboom P.
(1983) Estimating a monotone density, Proceedings of the Neyman-Kiefer Conference.
Györfi L.
(1981) Strong consistent density estimate from ergodic sample, Journal of Multivariate Analysis, 11, pp. 81—84.
Habbema J. D. F., Hermans J., Remme J.
(1978) Variable kernel density estimation in discriminant analysis, in COMPSTAT 78, L. C. A. Corsten and J. Hermans (Eds.), Physica Verlag, Wien.
Hall P., Heyde C. C.
(1980) Martingale Limit Theory and Its Application, Academic Press, New York.
Hanna B., Abou-Jaoude S.
(1981) Sur la vitesse de convergence de l'estimateur de la partition aléatoire d'une densité de probabilité, Publications de l'Institut de Statistique des Universités de Paris, 26, pp. 51—67.
Isogai E.
(1978) On strong consistency of a sequential estimator of probability density, Science Reports of Niigata Univeisity, A 15, pp. 25—33.
(1980) Strong consistency and optimality of a sequential density estimator, Bulletin of Mathematical Statistics, 19, pp. 55—69.
(1982) Strong uniform consistency of recursive kernel density estimators, Science Reports of Niigata University, A 18, pp. 15—27.
Klonias V. K.
(1982) Consistency of two nonparametric maximum penalized likelihood estimators of the probability density function, Annals of Statistics, 10, pp. 811—824.
Loève M.
(1963) Probability Theory, Van Nostrand, Princeton, New Jersey. (Имеется русский перевод: Лоэв М. Теория вероятностей. — М.: ИЛ, 1962.)
Loftsgaarden D. O., Quesenberry C. P.
(1965) A nonparametric estimate of a multivariate probability density function, Annals of Mathematical Statistics, 28, pp. 1049—1051.
Mack Y. P., Rosenblatt M.
(1979) Multivariate k nearest-neighbor density estimates, Journal of Multivariate Analysis, 9, pp. 1—15.
Moore D. S., Yackel J. W.
(1977) Consistency properties of nearest-neighbor density estimates, Annals of Statistics, 5, pp. 143—154.
Müller H.-G.
(1984) Smooth optimum kernel estimators of densities, regression curves and modes, Annals of Statistics, 12, pp. 766—774.
Prakasa-Rao B. L. S.
(1969) Estimation of a unimodal density, Sankhya, Series A, 31, pp. 23—36.
Quenouille M.
(1956) Notes on bias in estimation, Biometrika, 43, pp. 353—360.
Raatgever J. W., Duin R. P. W.
(1978) On the variable kernel model for multivariate nonparametric density estimation, in COMPSTAT 78, L. C. A. Corsten and J. Hermans (Eds.), Physica Verlag, Wien.

Rejtö L., Révész P.
(1973) Density estimation and pattern classification, Problems of Control and Information Theory, 2, pp. 67—80.
Robertson T.
(1967) On estimating a density which is measurable with respect to a σ-lattice, Annals of Mathematical Statistics, 38, pp. 482—493.
Rosenblatt M.
(1971) Curve estimates, .Annals of Mathematical Statistics, 42, pp. 1815—1842.
Schucany W. R., Gray H. L., Owen D. B.
(1971) On bias reduction in estimation, Journal of the American Statistical Association, 66, pp. 524—533.
Schucany W. R; Sommers J. P.
(1977) Improvement of kernel type density estimators, Journal of the American Statistical Association, 72, pp. 420—423.
Scott D. W., Tapia R. A., Thompson J. R.
(1980) Nonparametric probability density estimation by discrete maximum penalized likelihood criteria, Annals of Statistics, 8, pp. 820—832.
Silverman B. W.
(1982) On the estimation of a probability density function by the maximum penalized likelihood method, Annals of Statistics, 10, pp. 795—810.
Sommers J. P.
(1972) Improved density estimation, Technical Report 114, Department of Statistics, Southern Methodist University, Dallas, Texas.
Tapia R. A., Thompson J. R.
(1978) Nonparametric Probability Density Estimation, The Johns Hopkins University Press, Baltimore.
Terrel G. R., Scott D. W.
(1980) On improving convergence rates for nonnegative kernel density estimators, Annals of Statistics, 8, pp. 1160—1163.
Van Ryzin J.
(1970) On a histogram method of density estimation, Technical Report 226, Statistics Department, University of Wisconsin, Madison, Wisconsin.
(1973) A histogram method of density estimation, Communications in Statistics, 2, pp. 493—506.
Wahba G.
(1971) A polynomial algorithm for density estimation, Annals of Mathematical Statistics, 42, pp. 1870—1886.
(1975) Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation, Annals of Statistics, 3, pp. 15—29.
(1976) Histosplines with knots which are order statistics, Journal of the Royal Statistical Society, B 38, pp. 140—151.
Wegman E. J.
(1969) A note on estimating a unimodal density, Annals of Mathematical Statistics, 40, pp. 1661—1667.
(1970a) Maximum likelihood estimation of a unimodal density function, Annals of Mathematical Statistics, 41, pp. 457—471.
(1970b) Maximum likelihood estimation of a unimodal density, II, Annals of Mathematical Statistics, 41, pp. 2160—2174.
(1975) Maximum likelihood estimation of a probability density function, Sankhya, Series A, 37, pp. 211—224.
Wegman E. J., Davies H. I.
(1979) Remarks on some recursive estimators of a probability density, Annals of Statistics, 7, pp. 316—327.
Wolverton C. T., Wagner T. J.
(1969a) Asymptotically optimal discriminant functions for pattern classification, IEEE Transactions on Information Theory, IT-15, pp. 258—265.

(1969b) Recursive estimates of probability densities, IEEE Transactions on Systems, Science and Cybernetics, 5, p. 307.

Yamato H.

(1971) Sequential estimation of a continuous probability density function and the mode, Bulletin of Mathematical Statistics, 14, pp. 1—12.

Прохоров Ю. В.

(1949) Об усиленном законе больших чисел//ДАН СССР. — Т. 69. — 5. С. 607—610.

КОММЕНТАРИЙ ПЕРЕВОДЧИКА

Среднеквадратичная ошибка оценки Волвертона—Вагнера (2) исследовалась Кошкиным, Тарасенко (1976*), Булдаковым, Кошкиным (1977*). Оценка плотности, основанная на статистической эквивалентности выборочных блоков (полиграмма), введена Tarasenko (1968*). Асимптотическое распределение полиграммы изучено в книге Тарасенко (1976*).

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Булдаков В. М., Кошкин Г. М.

(1977*) О рекуррентных оценках плотности вероятности и линии регрессии// Проблемы передачи информации. — Т. 13. — 1. — С. 58—66.

Кошкин Г. М., Тарасенко Ф. П.

(1976*) Рекуррентное оценивание плотности вероятности и линии регрессии по зависимой выборке//Матем. статист. и ее приложения. Вып. 4. — Томск: Изд-во Томского гос. ун-та. — С. 135—136.

Тарасенко Ф. П.

(1976*) Непараметрическая статистика. — Томск: Изд-во Томского гос. ун-та.

Tarasenko F. P.

(1968*) On the evaluation of an unknown probability density function, the direct estimation of the entropy from independent observations of a continuous random variable and the distribution-free test of goodness-of-fit//Proc. IEEE. — V. 56. — 11. — P. 2052—2053.

Глава 8

Моделирование, неравенства и генерирование случайных величин

1. Выбор критерия

Рассмотрим ситуацию, когда требуются случайные величины с функцией распределения F на R^d, но вместо них используются случайные величины с функцией распределения G. Причины такой замены бывают как экономическими (случайные величины, распределенные согласно G, можно получить за меньшее время или с меньшей затратой памяти ЭВМ), так и практическими (в рассматриваемой прикладной задаче достаточно иметь лишь хорошее приближение для F). Иногда функция F неизвестна, и ее нужно оценить по данным. Во многих же случаях просто не хочется тратить время на написание сложной программы генерирования случайных величин с функцией распределения F. Какова бы ни была причина замены, необходимо хорошо разобраться в том, к каким последствиям такая замена приводит. Что нужно использовать в качестве меры согласия приближения при моделировании?

Один из классических критериев,

$$\Delta_1 = \sup_x |F(x) - G(x)|,$$

обладает тем недостатком, что он нечувствителен к локальным расхождениям между распределениями. Например, если вся масса распределения F сосредоточена равномерно на отрезках $[0, 1]$, $[2, 3]$, ..., $[2n-2, 2n-1]$, а вся масса G — равномерно на отрезках $[1, 2]$, $[3, 4]$, ..., $[2n-1, 2n]$, то $\Delta_1 = 1/n$. При больших n значение этого критерия достаточно мало, но ясно, что при моделировании никто не захочет заменять F на G.

Предположим, что $d = 1$. Если функции F и G непрерывны и U — случайная величина, равномерно распределенная на $[0, 1]$, то $F^{-1}(U)$ и $G^{-1}(U)$ — случайные величины с функциями распределения F и G соответственно. Этот факт, как известно, лежит в основе *метода инверсии* при генерировании случайных величин, приводящего к критерию

$$\Delta_2 = \sup_{0<u<1} |F^{-1}(u) - G^{-1}(u)|.$$

К сожалению, критерий Δ_2, так же как Δ_1, не является локально чувствительным и, кроме того, придает слишком большое значе-

ние хвостам распределений. Например, если носитель F бесконечен, а носитель G — компакт, то $\Delta_2 = \infty$.

Рассмотрим теперь *критерий полной вариации*

$$J = \frac{1}{2}\int |f - g| = \sup_B \left| \int_B f - \int_B g \right|,$$

где f и g — плотности, соответствующие F и G. Как показано в гл. 1, величина J дает верхнюю границу для абсолютной величины ошибки, возникающей при замене любой вероятности вида $\int_A f$ на ее приближение $\int_A g$. Если случайные величины нужны для оценки функционала $\int h dF$ (где $h \geqslant 0$) методом Монте-Карло, то

$$\left| \int h\, dF - \int h\, dG \right| =$$

$$= \left| \int_0^\infty \int_{h(x) > t} dF(x)\, dt - \int_0^\infty \int_{h(x) > t} dG(x)\, dt \right| \leqslant$$

$$\leqslant \int_0^\infty \left| \int_{h(x) > t} dF(x) - \int_{h(x) > t} dG(x) \right| dt \leqslant J \sup_x h(x).$$

Следовательно, при ограниченных h имеется четкая верхняя граница ошибки в случае, когда интеграл $\int h dG$ можно вычислить точно. Часто величина J определяется без большого труда, но иногда ее вычисление оказывается очень сложным. В § 2 приводятся некоторые неравенства, которые могут быть полезны при получении верхних границ для J.

2. Неравенства

В этом параграфе приводятся неравенства, устанавливающие связь величины $\int |f - g|$ с другими мерами расстояния между f и g. Одни из этих неравенств использовались при доказательстве сходимости оценок в предыдущих главах; другие помогают пользователям узнать нечто о свойствах оценок при мерах расстояния, отличных от L_1.

Начнем с неравенств, которые находят применение при генерировании случайных величин.

Теорема 1.

$$\int |f - g| \leqslant 2 \min(K_r, K_c),$$

где

$$K_r = \sup\,(1 - g/f), \quad K_c = \sup\,(1 - f/g).$$

Доказательство. По теореме 1.1

$$\int |f - g| = 2 \int\limits_{f \geqslant g} (f - g) = 2 \int\limits_{f \geqslant g} f\,(1 - g/f) \leqslant 2K_r,$$

что в силу соображений симметрии влечет утверждение теоремы 1.

Постоянные K_r и K_c применяются при генерировании случайных величин методами удаления и композиции. Например, в алгоритме удаления величина K_r используется следующим образом:

Шаг 1. Генерировать случайную величину X с плотностью g и не зависящую от нее случайную величину U, равномерно распределенную на $[0, 1]$.

Шаг 2. Если $Ug\,(X) \leqslant (1 - K_r) f\,(X)$, то вывести X; в противном случае возвратиться к шагу 1.

Метод композиции работает следующим образом:

Шаг 1. Генерировать случайную величину U, равномерно распределенную на $[0, 1]$.

Шаг 2. Если $U \leqslant 1 - K_c$, то вывести случайную величину X с плотностью g; в противном случае вывести случайную величину X с плотностью $(f - (1 - K_c)g)/K_c$.

Случайные величины, генерируемые обоими методами, имеют плотность f, если арифметические операции выполняются с бесконечной точностью. В обоих алгоритмах предполагается, что плотность g близка к f (т. е. значения K_c или K_r малы) и что можно легко генерировать случайные величины с плотностью g.

В гл. 4 нам потребовалось теоретико-информационное неравенство между L_1-ошибкой и величиной $\int f \log (f/g)$.

Теорема 2.

$$\int f \log\left(\frac{f}{g}\right) \geqslant \frac{1}{2}\left(\int |f - g|\right)^2,$$

$$\frac{1}{2}\int |f - g| \leqslant \sqrt{1 - \exp\left(-\int f \log\left(\frac{f}{g}\right)\right)},$$

$$\int \min(f, g) \geqslant \frac{1}{2}\exp\left(-\int f \log\left(\frac{f}{g}\right)\right).$$

Доказательство. Пусть $A = \{f \geqslant g\}$, $B = \{f < g\}$, $h = gI_A \Big/ \int\limits_A g$. Тогда по неравенству Йенсена

$$\int\limits_A f \log\left(\frac{f}{g}\right) = \int h \frac{f}{g} \log\left(\frac{f}{g}\right) \int\limits_A g \geqslant$$

$$\geqslant \int h \frac{f}{g} \log\left(\int h \frac{f}{g}\right) \int\limits_A g = \int\limits_A f \log\left(\int\limits_A f \Big/ \int\limits_A g\right).$$

Положим $p = \int\limits_A f$, $q = \int\limits_A g$. В силу симметрии имеем

$$\int f \log (f/g) \geqslant p \log\left(\frac{p}{q}\right) + (1-p) \log\left(\frac{1-p}{1-q}\right) = H(p, q).$$

Не нарушая общности, предположим, что $p = q + r$ при некотором $r > 0$. Запишем $H(p, q)$ в виде $H(q, r) = (q + r) \log(1 + r/q) + (1 - q - r) \log(1 - r/(1-q))$ и заметим, что $H'(q, r) = \log(1 + r/q) - \log(1 - r/(1-q))$ и $H''(q, r) = 1/p\,(1-p) \geqslant 4$, где производные берутся по r. Таким образом, используя разложение в ряд Тейлора с остаточным членом, получим

$$H(p, q) \geqslant 4\,(r^2/2) = 2r^2 = 2\left(\int\limits_{f>g} (f-g)\right)^2 = \frac{1}{2}\left(\int |f-g|\right)^2.$$

Первое неравенство доказано.

Для доказательства второго неравенства применим еще раз неравенство Йенсена:

$$-\int f \log\left(\frac{f}{g}\right) = \int f \left(\log\left(\min\left(\frac{g}{f}, 1\right)\right) + \log\left(\max\left(\frac{g}{f}, 1\right)\right)\right) \leqslant$$

$$\leqslant \log\left(\int \min(f, g)\right) + \log\left(\int \max(f, g)\right).$$

Следовательно,

$$\exp\left(-\int f \log\left(\frac{f}{g}\right)\right) \leqslant \int \min(f, g) \int \max(f, g) =$$

$$= \left(1 - \frac{1}{2}\int |f-g|\right)\left(1 + \frac{1}{2}\int |f-g|\right) = 1 - \left(\frac{1}{2}\int |f-g|\right)^2.$$

Третье неравенство тривиальным образом следует из предыдущего рассуждения.

Первое неравенство теоремы 2 доказано в работах Kullback (1967), Csiszar (1967) и Kemperman (1969). Другие неравенства и их доказательства см. в Bretagnolle, Huber (1979). Существует другая мера расстояния, равная $\int f^2/g - 1 = \int (f^2 - g^2)/g$, которая тесно связана с величиной $\int f \log(f/g)$ в том смысле, что

обе они конечны или бесконечны одновременно. В теореме 3 указывается связь этой новой меры расстояния с введенными ранее.

Теорема 3.

$$\int \frac{f^2}{g} - 1 \geqslant \log\left(\int \frac{f^2}{g}\right) \geqslant \int f \log\left(\frac{f}{g}\right),$$

$$\int \frac{f^2}{g} - 1 \geqslant \left(\int |f - g|\right)^2.$$

Кроме того, для произвольной функции g, не обязательно плотности, имеем

$$\int \sqrt{f}\,|f - g| \leqslant \sqrt{\int (f - g)^2}.$$

Доказательство. Левое неравенство в первой цепочке получается, если заметить, что $\log u \leqslant u - 1$ при $u > 0$. Правое неравенство следует непосредственно из неравенства Йенсена. Для получения второго утверждения теоремы применим неравенство Гёльдера. Пусть числа a, $q > 1$ таковы, что $1/p + 1/q = 1$. Тогда

$$\int |f - g| = \int \frac{|f - g|}{f^{1/p}} f^{1/p} \leqslant \left(\int \frac{|f - g|^q}{f^{q/p}}\right)^{1/q} \left(\int f\right)^{1/p}.$$

Полагая $p = q = 2$, получим искомое неравенство.

Последнее утверждение следует непосредственно из неравенства Коши—Шварца.

Расстояние Хеллингера $H_p = \left(\int |f^{1/p} - g^{1/p}|^p\right)^{1/p}$, $p \geqslant 1$, так же как и L_1-ошибка H_1, обладает рядом хороших свойств, а именно оно всегда конечно и инвариантно относительно строго монотонных преобразований. Питмен (Pitman (1979)) использовал расстояние H_2 при исследовании оценок максимума правдоподобия. К сожалению, линейной связи между H_2 и $H_1 = \int |f - g|$ нет, т. е. не существует универсальной постоянной a, такой, что $H_2 \sim aH_1$ при $H_1 \to 0$. В самом деле, справедливы следующие неравенства:

Теорема 4.

$$H_2^2 \leqslant H_1 \leqslant H_2\sqrt{4 - H_2^2} \leqslant 2H_2.$$

Кроме того, для любых f существуют последовательности плотностей f_n и g_n, такие, что

$$H_1(f, f_n) \sim 2H_2^2(f, f_n) \to 0,$$

$$H_1(f, g_n) \sim 2H_2(f, g_n) \to 0.$$

Доказательство. Имеем

$$H_1 = \int |f-g| = \int |\sqrt{f}-\sqrt{g}|\,(\sqrt{f}+\sqrt{g}) \geqslant$$

$$\geqslant \int |\sqrt{f}-\sqrt{g}|^2 = H_2^2,$$

$$H_1^2 \leqslant \int (\sqrt{f}-\sqrt{g})^2 \int (\sqrt{f}+\sqrt{g})^2 =$$

$$= H_2^2 \left(2 + 2\int \sqrt{fg}\right) = H_2^2\,(4 - H_2^2),$$

где мы воспользовались неравенством Коши—Шварца.

Последовательность g_n строится так, чтобы она сильно перекрывалась с f. Пусть m — медиана для f. Положим $g_n = (1 + p_n) f$ на $(-\infty, m]$ и $g_n = (1 - p_n) f$ на (m, ∞), где $p_n \downarrow 0$ — некоторая последовательность. Ясно, что при каждом n функция g_n является плотностью. Кроме того, $H_1 = p_n$ и $H_2^2 = 2 - 2\int \sqrt{fg_n} = 2 - \sqrt{1-p_n} - \sqrt{1+p_n} \sim p_n^2/4$. Здесь использовано то, что $\sqrt{1-x} = 1 - x/2 - x^2/8 + O(x^3)$ при $x \downarrow 0$ и $\sqrt{1+x} = 1 + x/2 - x^2/8 + O(x^3)$ при $x \downarrow 0$.

Равенство $H_1 = H_2^2$ достигается для неперекрывающихся f и g, т. е. в случае, когда $\int \sqrt{fg} = 0$. Поэтому последовательность f_n строится частично на основе последовательности плотностей, не перекрывающихся с f. Пусть m_n есть $(1/n)$-квантиль для f. Определим f_n такими соотношениями: $f_n = \sqrt{n} f$ на $(-\infty, m_n]$, $f_n = (1 - 1/\sqrt{n})\, f/(1 - 1/n)$ на (m_n, ∞). Снова f_n является плотностью при каждом n. Легко проверить, что $H_1 = 2(\sqrt{n}-1)/n \sim 2/\sqrt{n}$. Кроме того, $H_2^2 = 2 - 2\int \sqrt{f f_n} = 2 - 2/n^{3/4} - 2\sqrt{(1-1/n)(1-1/\sqrt{n})} \sim 2 - 2n^{-3/4} - 2(1 - 1/2\sqrt{n}) \sim 1/\sqrt{n}$.

В теореме 5 формулируется *неравенство Ле Кама* (Le Cam (1973)), которое существенно использовалось при доказательстве леммы Ассуа (теорема 4.5).

Теорема 5.

$$\int \min(f, g) \geqslant \frac{1}{2}\left(\int \sqrt{fg}\right)^2$$

для любых плотностей f и g на R^d.

Доказательство. По неравенству Коши—Шварца

$$\left(\int_{f<g} \sqrt{fg}\right)^2 = \left(\int_{f<g} f\sqrt{\frac{g}{f}}\right)^2 \leqslant \int_{f<g} f \frac{g}{f} \int_{f<g} f \leqslant \int_{f<g} f.$$

В силу симметрии

$$\left(\int \sqrt{fg}\right)^2 \leqslant 2\int\limits_{f<g} f + 2\int\limits_{g\leqslant f} g = 2\int \min(f, g).$$

Рассмотрим, наконец, L_∞-норму ess sup $|f-g|$, где существенная верхняя грань берется относительно меры Лебега. Ясно, что интеграл $\int |f-g|$ может быть мал, а при этом величина sup $|f-g|$ велика и, возможно, бесконечна. Обратно, из малости значений L_∞-нормы не следует малость расстояния в L_1, если нет никаких предположений о хвостах плотностей f или g. В качестве примера того, какие предположения на хвосты следует накладывать, приведем несколько неравенств (Serfling (1979)).

Теорема 6. *Пусть f и g — плотности на R^d и r — положительная постоянная. Положим $v_d = \pi^{d/2}\Gamma(d/2+1)$ и*

$$A = \begin{cases} \sup\limits_t t^r \int\limits_{\|x\|>t} f \left(\leqslant \int \|x\|^r f\right), & 0 < r < \infty, \\ \inf\left\{t: \int\limits_{\|x\|\leqslant t} f = 1\right\}, & r = \infty. \end{cases}$$

Тогда

$$\int |f-g| \leqslant 4A^{d/(r+d)} (v_d \operatorname{ess\,sup} |f-g|)^{r/(r+d)}, \quad 0 < r < \infty,$$

и

$$\int |f-g| \leqslant 2A^d v_d \operatorname{ess\,sup} |f-g|, \quad r = \infty.$$

В определении величины A можно заменить f на g.

Доказательство.

$$\frac{1}{2}\int |f-g| = \int\limits_{f>g} (f-g) = \int\limits_{\substack{\|x\|\leqslant t \\ f>g}} (f-g) + \int\limits_{\substack{\|x\|>t \\ f>g}} (f-g) \leqslant$$

$$\leqslant v_d t^d \operatorname{ess\,sup} |f-g| + At^{-r}$$

при любых $t > 0$. Слагаемые в правой части неравенства равны между собой при $t^{r+d} = A/(v_d \operatorname{ess\,sup} |f-g|)$. Подставляя значение t, удовлетворяющее этому условию, получим первое неравенство теоремы. Второе неравенство очевидно.

3. Обобщение выборки для генерирования случайных величин

Пусть задана выборка $X_1, \ldots, X_n$, состоящая из независимых случайных векторов со значениями в R^d, которые имеют одну и ту же неизвестную плотность f, и требуется генерировать (т. е.

получить с помощью вычислительной машины) новую, не зависящую от исходной выборку независимых случайных векторов $Y_1, \ldots, Y_m$ с той же плотностью f. Очевидно, что поставленная таким образом задача не имеет решения. Некоторые из препятствий на пути к решению можно преодолеть, если ввести следующие удобные и достаточно реалистичные предположения:

(i) вычислительная машина может хранить действительные числа — в противном случае понятие «плотность» потеряло бы смысл;

(ii) имеется источник, способный генерировать последовательность $U_1, U_2, \ldots$ независимых случайных величин, равномерно распределенных на $[0, 1]$.

Так как плотность f неизвестна, то она должна быть явно или неявно) оценена по выборке $X_1, \ldots, X_n$. В самом общем случае нас будут интересовать процедуры следующего вида:

Шаг 1. Построить оценку $f_n(x) = f_n(x; X_1, \ldots, X_n)$ плотности $f(x)$.

Шаг 2. При $i = 1, \ldots, m$ выполнять следующее:

генерировать новые равномерно распределенные на $[0, 1]$ случайные величины U_i;

вычислять Y_i, используя f_n и U_i.

Ясно, что имеется зависимость между двумя такими выборками. Кроме того, оценка f_n совпадает с f только в случае невероятного везения. В данном параграфе обсуждается, до какой степени можно ограничить проявление этих нежелательных эффектов.

Особый интерес здесь представляют вопросы независимости выборок, состоятельности, неразличимости выборок, согласования моментов, построения генераторов для f_n.

3.1. Независимость выборок

С зависимостью между $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ почти ничего нельзя поделать — остается только надеяться, что происходит приближение к асимптотической независимости при достаточно больших n. Еще отметим, что в некоторых приложениях независимости выборок вовсе не требуется.

Так как $Y_1, \ldots, Y_m$ условно независимы при фиксированных $X_1, \ldots, X_n$, то достаточно учитывать только зависимость между $Y = Y_1$ и $X_1, \ldots, X_n$. Мерой этой зависимости является

$$D_n = \sup_{A, B} |P(Y \in A, X \in B) - P(Y \in A) P(X \in B)|,$$

где верхняя грань берется по всем борелевским множествам A из R^d и всем борелевским множествам B из R^{nd} и X — краткое

обозначение для $(X_1, \ldots, X_n)$. Будем говорить, что выборки *асимптотически независимы*, если

$$\lim_{n\to\infty} D_n = 0.$$

В случаях когда векторы $X_1, \ldots, X_n$ используются для проектирования или построения системы, а $Y_1, \ldots, Y_m$ — для ее тестирования, зависимость выборок часто приводит к оптимистичным оценкам. Без асимптотической независимости нет надежды на то, что с ростом n эта тенденция к излишнему оптимизму будет уменьшаться.

Приводимое ниже неравенство теоремы 7 показывает, что условие $\lim\limits_{n\to\infty} E(J_n) = 0$ достаточно для асимптотической независимости.

Теорема 7.

$$D_n \leqslant E(J_n) = E\left(\int |f_n - f|\right).$$

Доказательство. Имеем

$$\begin{aligned} D_n \leqslant & \sup_{A,B} |P(Y \in A,\ X \in B) - P(X_{n+1} \in A,\ X \in B)| + \\ & + \sup_{A,B} |P(X_{n+1} \in A,\ X \in B) - P(X_{n+1} \in A)\,P(X \in B)| + \\ & + \sup_{A,B} |P(X_{n+1} \in A)\,P(X \in B) - P(Y \in A)\,P(X \in B)|. \end{aligned} \tag{1}$$

Последнее слагаемое в (1) равно

$$\sup_A |P(X_{n+1} \in A) - P(Y \in A)| = \sup_A \left|\int_A E(f_n) - \int_A f\right| = \tfrac{1}{2}\int |E(f_n) - f|$$

в силу теоремы 1.1 Шеффе. Второе слагаемое в (1), очевидно, равно нулю, а первое слагаемое не превосходит

$$\sup_{A,B} E\left(I_{[X\in B]}\left|\int_A f_n - \int_A f\right|\right) \leqslant$$

$$\leqslant \sup_A E\left(\left|\int_A f_n - \int_A f\right|\right) \leqslant E\left(\sup_A \left|\int_A f_n - \int_A f\right|\right) = E\left(\tfrac{1}{2}\int |f_n - f|\right)$$

Теорема 7 доказана.

3.2. Состоятельность

Теорема 7 показывает, что асимптотическая независимость выборок следует из состоятельности оценки, т. е из условия $\lim\limits_{n\to\infty} E(J_n) = 0$. Но более существенно, что состоятельность

нужна для получения хороших приближений всевозможных вероятностей, ибо

$$\sup_A \left| \int_A f_n - \int_A f \right| = \frac{1}{2} \int |f_n - f| = J_n/2 \qquad (2)$$

(см. теорему 1.1, обсуждение в гл. 1 и § 1 данной главы).

3.3. Неразличимость выборок

Одной из важных мер качества метода моделирования является неразличимость $X_1, \ldots, X_m$ и $Y_1, \ldots, Y_m$ *при заданном объеме выборки* m. Если обозначить через Card (A) и Card* (A) количества элементов выборок $X_1, \ldots, X_m$ и $Y_1, \ldots, Y_m$ соответственно, попавших в A, где A — произвольное подмножество R^d, то в качестве меры неразличимости можно предложить

$$S_n = \sup_A | E(\text{Card}(A)) - E(\text{Card}^*(A) | X_1, \ldots, X_n) |.$$

Если случайная величина S_n меньше, чем 1, то *любые* множества A в среднем содержат от $u - 1/2$ до $u + 1/2$ точек каждой выборки, где u — некоторое число. Такой строгий критерий необходим, например, если важно избавиться от экстремальных множеств.

Нетрудно видеть, что

$$S_n = m \sup_A \left| \int_A f_n - \int_A f \right| = \frac{m}{2} \int |f_n - f| = \frac{m}{2} J_n.$$

Таким образом, мы снова приходим к L_1-критерию J_n.

Назовем оценку f_n *k-превосходной для выборок объема* m, если

$$E(S_n) = \frac{m}{2} E(J_n) \leqslant k. \qquad (3)$$

Понятие 1-превосходной оценки является очень сильным. Чтобы продемонстрировать это, достаточно будет показать, сколь плохо ведут себя произвольные, параметрические или непараметрические, оценки плотности. В табл. 1 для определенных комбинаций оценок плотности f_n и плотностей f вычислены пороговые значения n, ниже которых оценки не могут быть 1-превосходными при заданных m. Рассматриваются следующие комбинации:

А. *Любые* оценки f_n и некоторая плотность f вида $p\tilde{f} + (1 - p) g$, где $\tilde{f}$ и g — известные плотности с непересекающимися носителями и p — единственный неизвестный параметр — число между 0 и 1.

B. *Любые* оценки f_n и некоторая плотность $f \in F_{2,r}$, где r — произвольное число, большее или равное $2r^*$ (см. гл. 4). По существу это класс всех плотностей с конечными значениями $B^*(f)$.

C. *Любые* ядерные оценки (допустимы любые K и h) и *любые* плотности f.

D. *Любые* гистограммные оценки и любые плотности $f \in \mathscr{F}$, т. е. любые абсолютно непрерывные f, имеющие ограниченную и п. в. непрерывную производную f'.

Числа из табл. 1 получены по результатам гл. 4 и 5, где отброшены члены $o(1)$ асимптотических разложений и произведено округление постоянных до первых двух значащих десятичных цифр. Таким образом, эти числа являются лишь приближенными. В столбцах A и B использовались соответственно приближение

$$E(J_n) \geqslant \frac{0{,}0849856\ldots}{\sqrt{n}} \quad (\text{или } n \geqslant (0{,}0424928\ldots m)^2) \qquad (4)$$

Таблица 1

m	A	B	C	D
10	1	1	40	85
100	18	1	13000	85000
1000	1800	1	4000000	85000000
10000	180000	77	1300000000	85000000000
100000	18000000	24000	400000000000	85000000000000

для неравенства из теоремы 4.4 и приближение

$$E(J_n) \geqslant (2e)^{-4}\, n^{-2/5} \quad (\text{или } n \geqslant ((2e)^{-4}\, m/2)^{5/2}) \qquad (5)$$

для неравенства из теоремы 4.3.

Числа в столбце B меньше чисел в A, что объясняется грубостью рассуждений, проведенных в теореме 4.4. Если заменить $2r^*$ на $5r^*$, то пороговое значение n увеличивается в $4^{5/2} = 32$ раза. Отсюда следует, что при $m = 10\,000$ для любой сколь угодно хорошей оценки плотности найдется плотность из класса $F_{2,r}$, $r \geqslant 5r^*$, такая, что данная оценка не является 1-превосходной при $n < 32 \times 77 = 2\,464$. Даже в случае простого класса A, где можно использовать очень простые параметрические оценки, для существования 1-превосходных (равномерно по f из этого класса) оценок нужно, чтобы исходная выборка содержала по меньшей мере $n = 180\,000$ точек при $m = 10\,000$. Столбцы A и B таблицы свидетельствуют о том, что существуют плотности из определенных классов, такие, что для них не может быть 1-превосходных оценок при слишком малых n.

Какие именно эти плотности, неизвестно — они зависят от используемой оценки f_n и от n. Здесь как раз проявляется слабость нижних границ из гл. 4. Действительно, данный результат все еще не является удовлетворительным, поскольку вполне может оказаться, что, несмотря на таблицу 1, для оцениваемой плотности f существуют 1-превосходные оценки. В этой связи более мощными являются границы для фиксированных плотностей, полученные в гл. 5.

Так, ни при каких f никакая обычная ядерная оценка не может быть 1-превосходной, если $m = 1000$ и $n < 4\,000\,000$. Это мощное утверждение следует из неравенства

$$E(J_n) \geqslant C_3 n^{-2/5} \quad (\text{или } n \geqslant (0{,}43933402\ldots m)^{5/2}) \qquad (6)$$

теоремы 5.2. Исключений здесь нет — неравенство применимо к любым плотностям. Таким образом, четырехмиллионная нижняя граница объема выборки является абсолютной нижней границей. Данные для гистограммной оценки (столбец D) еще более разочаровывают, чем данные для ядерной оценки (столбец C), поскольку нижняя граница при $m = 1000$ становится уже равной 85 миллионам и числа в столбце D возрастают со скоростью порядка m^3. Эти числа получаются из неравенства

$$E(J_n) \geqslant 0{,}880261\ldots n^{-1/3} \quad (\text{или } n \geqslant (0{,}4401305\ldots m)^3) \qquad (7)$$

теоремы 5.5.

Возможно, пользователю захочется предъявить более слабое требование — чтобы оценка была k-превосходной при $k > 1$. В самом деле, высокое качество оценки, подразумеваемое условием $E(S_n) \leqslant 1$, часто не требуется. В этом случае значения m из табл. 1 следует умножить на k.

Тем не менее табл. 1 очень ясно показывает, что при больших значениях m не следует использовать гистограммную оценку, за исключением случаев, когда имеются другие, более важные, чем $E(J_n)$, показатели, такие, как монотонность генератора случайных чисел (см. п. 3.5 ниже).

Границы (4)—(7) представляют собой отрицательные результаты. Позитивный же и обнадеживающий результат состоит в том, что можно построить 1-превосходную при любых m ядерную оценку, если определенной ценой заплатить за это. В теореме 5.1 было показано, что для любых f с компактным носителем в первом приближении верно неравенство

$$E(J_n) \leqslant 1{,}3768102\ldots (9/125)^{1/5} B^*(f)\, n^{-2/5},$$

если h и K выбраны оптимальным образом.

Еще одна граница,

$$E(J_n) \leqslant 1{,}240701\ldots B_H^*(f)\, n^{-1/3},$$

дается в теореме 5.10. Используя эти два приближенных неравенства, можно убедиться в том, что если f — равнобедренная треугольная плотность, то оценка является 1-превосходной при

$$n \geqslant \left(\frac{1{,}17624440\ldots}{2}\, m\right)^{5/2},$$

и если f — равномерная плотность на $[0, 1]$, то оценка является таковой при

$$n \geqslant \left(\frac{1{,}240701\ldots}{2}\, m\right)^{3}.$$

Это утверждение верно, во всяком случае если значение h выбрано оптимальным образом и K — ядро Епанечникова или равнобедренное треугольное ядро соответственно. Таблица 2 показывает, как работают данные неравенства. В случае k-превосходных оценок значения m нужно умножить на k.

3.4. Согласование моментов

Некоторые статистики и инженеры придают большое значение моментам плотностей f_n и f. При $d = 1$ *невязкой i-го момента* называется случайная величина

$$M_{ni} = \int x^i f_n - \int x^i f, \quad i = 1, 2, 3, \ldots, \tag{8}$$

определенная в случае, когда функция $|x|^i$ интегрируема относительно плотностей f_n и f.

В теореме 8 даются невязки M_{n1} и M_{n2} для ядерной оценки.

Теорема 8. *Для ядерной оценки на R^1, ядро которой удовлетворяет условиям $\int xK = 0$, $\int x^2K = \sigma^2$, имеем*

$$M_{n1} = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)),$$

$$M_{n2} = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - E(X_i^2)) + h^2\sigma^2.$$

Кроме того,

$$E(M_{n1}) = 0, \quad \mathrm{Var}(M_{n1}) = \mathrm{Var}(X_1)/n, \quad E(M_{n2}) = h^2\sigma^2$$

и

$$\mathrm{Var}(M_{n2}) = \mathrm{Var}(X_1^2)/n.$$

Таблица 2

m	Равномерная на [0, 1] плотность f	Равнобедренная треугольная f
10	239	83,9
100	239000	26500
1000	239000000	8390000
10000	239000000000	2650000000
100000	239000000000000	839000000000

Доказательство. Воспользуемся тем, что f_n — плотность случайной величины $Y = X_Z + hW$, где Z, W — независимые между собой и с X_i случайные величины, Z распределена равномерно на $\{1, \ldots, n\}$ и W имеет плотность K. Утверждение теоремы 8 получается, если заметить, что

$$E(Y \mid X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i + hE(W) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и

$$E(Y^2 \mid X_1, \ldots, X_n) = E(X_Z^2 + 2hWX_Z + h^2W^2 \mid X_1, \ldots, X_n) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 + h^2\sigma^2.$$

Заметим прежде всего, что на распределение величины M_{n1} не влияет выбор h и K. По слабому закону больших чисел M_{n1} стремится к 0 по вероятности, если $E(|X_1|) < \infty$, вне зависимости от того, как h меняется с ростом n. Однако невязка второго момента складывается из случайной величины, не зависящей от h и K, и постоянной составляющей $h^2\sigma^2$. Только из-за этой составляющей возникает положительное смещение $E(M_{n2}) = h^2\sigma^2$. Так как мы не можем управлять величиной $\mathrm{Var}(M_{n2})$, то самое лучшее — максимально уменьшить смещение. Но это привело бы к выбору столь малых h, что значение $E(J_n)$ стало бы возрастать. При оптимальном в смысле критерия $E(J_n)$ выборе параметра h для многих гладких распределений он зависит от n как $n^{-1/5}$. В этом случае величина $E(M_{n2})$ является хорошей мерой невязки второго момента, ибо $\sqrt{\mathrm{Var}(M_{n2})} = O(n^{-1/2}) = o(E(M_{n2}))$.

Если, например, K — ядро Епанечникова и параметр h выбран оптимальным образом (см. (5.14)), то нормированная невязка второго момента равна

$$\frac{E(M_{n2})}{\mathrm{Var}(X_1)} = \frac{1}{5}\left(\frac{15}{2\pi}\right)^{2/5} \frac{n^{-2/5}\left(\int \sqrt{f} \Big/ \int |f''|\right)^{4/5}}{\mathrm{Var}(X_1)}. \qquad (9)$$

Если использовать параметр h, оптимальный при заданном ядре K, то получится выражение, достигающее минимума на ядре Епанечникова. Поэтому выбор ядра Епанечникова вполне обоснован. Выражение (9) трансляционно и масштабно инвариантно. В нем существенна только форма плотности f. Чтобы грубо прикинуть, каково численное значение (9), будем полагать, что f — плотность нормального распределения. Тогда

$$\frac{E(M_{n2})}{\mathrm{Var}(X_1)} = \left(\frac{225\pi e^2}{32}\right)^{1/5} \frac{1}{5} n^{-2/5} = 0{,}5540591\ldots n^{-2/5}. \qquad (10)$$

В табл. 3 приведены полученные из (10) значения процентных уровней относительной невязки второго момента в случае плотности нормального распределения в зависимости от n. Заметим, что при обычных значениях n эти уровни заключены между 1 % и 10 %. Следовательно, сглаживание, необходимое для обеспечения состоятельности и малых значений $E(J_n)$, оказывает нежелательное побочное воздействие на невязку второго момента, и это воздействие сильнее всего проявляется при малых n. Заметим также, что, согласно табл. 3, почти безнадежно требовать, чтобы относительная ошибка была порядка 0,1 % или меньше. При $d > 1$ ситуация, разумеется, еще сложнее (см., например, Shanmugam (1977), где обсуждается похожая проблема).

Таблица 3

n	Нормированная невязка второго момента (f — плотность нормального распределения, h — оптимальный параметр сглаживания)
10	0,2205...
100	0,08781...
1000	0,03496...
10000	0,01391...
100000	0,005540...
1000000	0,002205...

Таблица 4

n	Абсолютная верхняя граница нормированной невязки второго момента
10	3,652...
100	1,453...
1000	0,5788...
10000	0,2304...
100000	0,09173...
1000000	0,03652...

Заметим еще, что в силу (5.18) нормированная невязка второго момента не превосходит величины

$$(1/5) \times (6{,}7726100\ldots)^2\, n^{-2/5} = 9{,}1736492\ldots n^{-2/5},$$

если K — ядро Епанечникова и h выбирается согласно (5.14). Значения данной универсальной границы при различных n приведены в табл. 4.

3.5. Генераторы случайных величин с плотностью f_n

Для генерирования случайных величин с плотностью, представляющей собой ядерную оценку

$$f_n(x) = (nh^d)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

можно применить следующую процедуру:

Шаг 1. Генерировать случайную величину Z, распределенную равномерно на $\{1, \ldots, n\}$, и генерировать не зависящий от нее случайный вектор W с плотностью K.

Шаг 2. Вывести величину $Y \leftarrow X_Z + hW$.

Единственное осложнение может возникать из-за ядра K. Как было показано, при $d = 1$ имеется ряд аргументов в пользу выбора ядра Епанечникова $K(x) = (3/4)(1 - x^2)$, $|x| \leqslant 1$. Существуют два очень быстрых алгоритма генерирования случайной величины W с такой плотностью распределения — *метод удаления* и *метод порядковых статистик*.

Метод удаления с прямоугольной доминирующей плотностью.

Шаг 1. Многократно генерировать случайную величину W, равномерно распределенную на $[-1, 1]$, и не зависящую от нее случайную величину U, равномерно распределенную на $[0, 1]$, до тех пор, пока не окажется выполненным неравенство $U \leqslant 1 - W^2$.

Шаг 2. Вывести W.

Метод порядковых статистик.

Шаг 1. Генерировать три независимые равномерно распределенные на $[-1, 1]$ случайные величины V_1, V_2 и V_3. Положить $W \leftarrow V_3$.

Шаг 2. Если $|V_3| > |V_1|$ и $|V_3| > |V_2|$, то положить $W \leftarrow V_2$. Вывести W.

В методе удаления величина W принимается на шаге 2 с вероятностью 2/3, так что на каждое получен-

ное на выходе значение W в среднем используются по три равномерно распределенных на $[-1, 1]$ случайные величины. При этом, однако, требуется производить умножения. В методе порядковых статистик также требуются три равномерно распределенные случайные величины, но умножение заменяется несколькими операциями взятия абсолютной величины.

Дехёвельс (Deheuvels (1977b)) нашел оптимальное в смысле критерия L_2 ядро K в R^d. Оно имеет вид

$$K(x) = C_d(d + 4 - \|x\|_2^2), \quad \|x\|_2^2 \leqslant d + 4,$$

где C_d — нормировочная постоянная, зависящая только от d. Случайные векторы с такой плотностью распределения (это многомерная плотность Пирсона II-го типа) представимы в виде

$$\sqrt{d+4}\,\sqrt{\text{Beta}\,(d/2,\ 2)}\,T_d,$$

где Beta $(d/2,\ 2)$ — случайная величина, имеющая бета-распределение и T_d — не зависящий от нее случайный вектор, равномерно распределенный на единичной сфере в R^d. Генерированию случайных величин с бета-распределением посвящены работы (Schmeiser, Shalaby (1980), Schmeiser, Babu (1980)). В этих работах, а также в Schmeiser (1980) можно найти дополнительные ссылки. Вектор T_d легко генерировать различными методами, например методом спейсингов (Sibuya (1962), Tashiro (1977)), полярным методом или специальными методами для малых значений d (см. обзорные статьи Deak (1979) и Rubinstein (1980)).

В полярном методе, например, используется тот факт, что вектор T_d распределен так же, как $(N_1/N, \ldots, N_d/N)$, где $N_1, \ldots, N_d$ — независимые нормальные случайные величины с параметрами (0, 1) и $N = \sqrt{N_1^2 + \ldots + N_d^2}$.

По поводу выбора параметра h, зависящего от данных, сошлемся на § 5.6 и § 6.2. Иногда требуются и другие модификации ядерной оценки, при которых необходимо модифицировать и генератор случайных чисел. В гл. 9 вводится преобразованная ядерная оценка, обладающая, как правило, пониженной величиной ошибки $E(J_n)$ при $d = 1$. Преобразованная ядерная оценка не имеет аналога при $d > 1$. Качество ядерной оценки можно повысить также за счет использования информации о направлениях разброса данных. Так, в работе Shanmugam (1977) (см. также Deheuvels (1977b)) обсуждаются последствия замены $Y = X_Z + hW$ на $Y = X_Z + hAW$, где A — матрица размера $d \times d$, такая, что обратная к выборочной ковариационной матрице равна A^tA. Для оценки из работы Breiman, Meisel, Purcell (1977) нужно изменить исходный алгоритм, полагая $Y = X_Z + h_Z W$, где h_i — расстояние от точки X_i до ее k-го ближайшего соседа среди $X_1, \ldots, X_n$, причем целое число k выбирается заранее (вместе с $X_1, \ldots, X_n$ нужно хранить и значения $h_1, \ldots, h_n$).

Если известно, что вся масса плотности f сосредоточена на $[0, \infty)$ или на $[0, 1]$, требуется еще одна модификация (см. гл. 9 ниже или работу Hominal, Deheuvels (1979)). В этом случае f_n можно заменить функцией $\bar{f}_n = f_n \Big/ \int_0^\infty f_n$ на $[0, \infty)$ и $\bar{f}_n = 0$ на $(-\infty, 0)$. В теореме 11.3 будет показано, что такую замену можно делать для любой оценки. Алгоритм генерирования случайных чисел также легко видоизменить:

Шаг 1. Многократно генерировать случайные величины X с плотностью f_n до тех пор, пока не окажется выполненным неравенство $X \geqslant 0$.

Шаг 2. Вывести X.

Среднее число повторений шага 1 в данном алгоритме удаления равно $1 \Big/ \int_0^\infty f_n$, что, как правило, близко к 1.

Приемы уменьшения смещения из § 7.6 усложняют дело лишь незначительно. Рассмотрим сначала нормированную оценку Бартлета $g_n = (f_n)_+ \Big/ \left(\int f_n\right)_+$, где f_n — ядерная оценка с ядром K, принимающим как положительные, так и отрицательные значения. Заметим лишь, что $(f_n)_+ \leqslant f_n^*$, где

$$f_n^*(x) = (nh)^{-1} \sum_{i=1}^{n} K_+\left(\frac{x - X_i}{h}\right).$$

Применим теперь метод удаления, принадлежащий фон Нейману (von Neumann (1951); см. также Rubinstein (1981)):

Шаг 1. Многократно выполнять следующее:

Генерировать три независимые случайные величины I, W и U, где I — случайное целое между 1 и n, W — имеет плотность $K_+ \Big/ \int K_+$ и величина U распределена равномерно на $[0, 1]$.

Положить $X \leftarrow X_I + hW$. Теперь X имеет плотность $f_n^* \Big/ \int f_n^*$.

Указанные действия выполнять до тех пор, пока не окажется, что $U \sum_{i=1}^{n} K_+((X - X_i)/h) \leqslant \left(\sum_{i=1}^{n} K((X - X_i)/h)\right)_+$.

Шаг 2. Вывести X.

Среднее число итераций цикла на первом шаге равно $\int f_n^* \Big/ \int (f_n)_+ = \int K_+ \Big/ \int (f_n)_+ \leqslant \int K_+$. Данная верхняя граница

точна, поскольку $\int (f_n)_+ \to 1$ при $h \to 0$, $nh \to \infty$ (это является следствием теоремы 7.4). Например, для ядра Бартлета

$$K(x) = (9/8)(1 - 5x^2/3), \quad |x| \leqslant 1,$$

получим

$$\int K_+ = 2 \int_0^{\sqrt{3/5}} \frac{9}{8}\left(1 - \frac{5x^2}{3}\right) dx = \sqrt{27/20} = 1{,}161895\ldots.$$

Вместе с тем вычисление сумм на первом шаге требует бóльших затрат машинного времени. Очевидно, что их не следует вычислять непосредственно, так как при этом длительность вычисления будет линейно возрастать с ростом n. Если ядро K сосредоточено на $[-1, 1]$, как в рассмотренном примере, то можно, в частности, хранить данные в виде массива, упорядоченного по возрастанию их значений. Методом бинарного поиска (см. Knuth (1975)) можно определить, в какой из интервалов (X_i, X_{i+1}) попадает X. Двигаясь по массиву вверх и вниз от данного интервала, можно найти все X_i, расстояние которых от X не превосходит h. Другие X_i не влияют на рассматриваемые суммы. При некоторых условиях на f можно показать, что среднее время работы такого алгоритма равно $O(nh) + O(\log n)$.

Если учитывать и вид ядра K, то можно еще больше сократить время вычислений. Например, если K — квадратичная функция на $[-1, 1]$ (как ядра Епанечникова и Бартлета), то известно, что ядерная оценка f_n является кусочно-квадратичным сплайном с изломами в точках $X_i - h$, $X_i + h$, $i = 1, \ldots, n$. Таким образом, достаточно запомнить в упорядоченном виде эти точки, а также по три коэффициента квадратичного полинома для каждого интервала. После того как найден интервал, которому принадлежит X (бинарным поискам это делается за время $O(\log n)$), суммы на первом шаге вычисляются мгновенно (за время $O(1)$). Конечно, при этом методе значительно возрастает время предварительной обработки, но и его можно ограничить разумными рамками, если сначала упорядочивать данные, а затем, просматривая их еще раз слева направо, вычислять все коэффициенты. В случае равномерного ядра K оценка f_n кусочно-постоянна, что позволяет еще больше упростить указанную процедуру. В дальнейшем есть все основания считать, что ядерную оценку можно вычислить за время порядка $O(\log n)$, не учитывая время предварительной обработки.

Оценка Террела — Скотта, также рассмотренная в § 7.6, при отсутствии нормировки имеет вид $f_n = f_{n1}(f_{n1}/f_{n2})^{1/3}$, где поправочный коэффициент $(f_{n1}/f_{n2})^{1/3}$ всегда заключен между 0 и $2^{1/3}$. Таким образом, $f_n \leqslant 2^{1/3} f_{n1}$. По сравнению с алгоритмом для

оценки Бартлета здесь нужно внести совсем небольшие изменения — на шаге 1 полагать, что X имеет плотность f_{n1}, а W — плотность K, и проводить итерации до тех пор, пока не окажется выполненным неравенство $2^{1/3}U \leqslant (f_{n1}/f_{n2})^{1/3}$. Среднее число итераций равно $2^{1/3} + o(1)$ (это следует из условия $\int f_n \to 1$, см. теорему 7.6), и вычисление величины $(f_{n1}(X)/f_{n2}(X))^{1/3}$ проводится за время $O(\log n)$ равномерно для любых X.

Рассмотрим теперь гистограммную оценку, определяемую разбиениями $\mathscr{P}_n = \{A_{nj},\ j$ — целое$\}$. Легко найти алгоритм генерирования случайных величин с плотностью f_n. Приведем пример такого алгоритма.

Шаг 0. Предварительная обработка. Вычислить и запомнить вероятности $p_i = \mu_n(A_{ni})$, такие, что $p_i \neq 0$ (μ_n — обычная эмпирическая мера для $X_1, \ldots, X_n$). Заметим, что требуется запомнить не более n вероятностей p_i.

Шаг 1. Генерировать целочисленную случайную величину I, такую, что $P(I = i) = p_i$.

Шаг 2. Генерировать и вывести случайную величину X, равномерно распределенную на A_{nI}.

В отличие от приведенного выше алгоритма для ядерной оценки здесь требуется предварительная обработка. Если использовать подходящие программы, то ее можно выполнить за время $O(n \log n)$, поскольку для каждой точки X_i надо проверить, содержит ли множество A_{nj}, которому принадлежит X_i, другую точку X_m. Аналогичный алгоритм без предварительной обработки имеет такой вид:

Шаг 1. Генерировать случайную величину Z, равномерно распределенную на множестве $\{1, \ldots, n\}$, и найти множество A_{nj}, которому принадлежит X_Z.

Шаг 2. Вывести случайную величину Y, равномерно распределенную на A_{nj}.

В большинстве случаев этот алгоритм предпочтительнее алгоритма с предварительной обработкой. Однако в особых случаях это не так, например если требуется получить случайные величины с максимальными отрицательными взаимными корреляциями для уменьшения дисперсии при моделировании методом Монте-Карло. Известно, что если F — непрерывная функция распределения, то две случайные величины с этой функцией распределения максимально отрицательно коррелированы, если они равны $F^{-1}(U)$ и $F^{-1}(1 - U)$, где U — равномерно распределенная на $[0, 1]$ случайная величина. В связи с этим Фокс (Fox (1980)) предлагает всегда пытаться применять метод инверсии (т. е. генерировать случайную величину с функцией распре-

деления F в виде $F^{-1}(U)$). К сожалению, ни один из приведенных выше алгоритмов для ядерной и гистограммной оценок не основан на инверсии.

Закончим этот параграф обсуждением различных алгоритмов инверсии для оценок плотности. Большинство интересных оценок, таких, как гистограммная и обычная ядерная с равномерным, треугольным или квадратичным ядром, сосредоточенным на $[-1, 1]$, являются кусочно-полиномиальными. Предположим, что действительная ось разбита на части точками излома $a_1 < a_2 < \cdots < a_n$ и оценка равна нулю вне $[a_1, a_n]$. На $[a_i, a_{i+1})$ оценка имеет вид

$$f_n(x) = b_{i0} + b_{i1}x + b_{i2}x^2 + \ldots + b_{ip}x^p.$$

Предположим также, что значения соответствующей функции распределения F_n в точках излома известны и равны $F_n(a_i) = c_i$. Нетрудно убедиться, что указанные коэффициенты, точки излома и значения в точках излома для рассматриваемых оценок вычисляются довольно просто.

Алгоритм инверсии работает следующим образом:

Шаг 1. Генерировать равномерно распределенную на $[0, 1]$ случайную величину U. Найти целое I, такое, что $c_I \leqslant U < c_{I+1}$. (Так как $c_1 = 0$ и $c_n = 1$, то значение I заключено между 1 и $n - 1$.)

Шаг 2. Вывести решение X уравнения

$$U - c_I = b_{I0}(X - a_I) + \frac{1}{2} b_{I1}(X^2 - a_I^2) + \\ + \ldots + \frac{1}{p+1} b_{Ip}(X^{p+1} - a_I^{p+1}).$$

Для кусочно постоянной оценки это уравнение решается очень просто. Для кусочно-квадратичных оценок решение предполагает нахождение корней полинома третьей степени. Во всех случаях время, необходимое для выполнения шага 2, не зависит от n и равно $O(1)$. Если на шаге 1 используется последовательный поиск интервала, то время счета растет линейно с ростом n. Не рекомендуется и бинарный поиск интервала, потому что в худшем случае он занимает время порядка $\log n$. Для генерирования целочисленной случайной величины I, распределенной так же, как I на шаге 1, можно использовать метод эталона (Walker (1977), Kronmal, Peterson (1979)), который тратит время $O(1)$ в худшем случае. К сожалению, этот метод неприемлем, поскольку при инверсии U не получается целое число. По-видимому, наиболее подходящим для шага 1 является метод «путеводителей» (Chen, Asau, (1974); см. также исчерпывающий обзор Ahrens, Kohrt (1981)). Этот метод в среднем занимает время $O(1)$,

но в худшем случае может работать гораздо дольше. Он основан на простом принципе. При предварительной обработке строится путеводитель, т. е. набор g_i, $1 \leqslant i \leqslant n$, где

$$g_i = \max \{j: c_j < i/n\}.$$

В частности, ясно, что $g_n = n - 1$. Поиск числа I, такого, что $c_I \leqslant U < c_{I+1}$, осуществляется последовательно, начиная с места, указанного в путеводителе. Таким образом, путеводитель *почти* является таблицей инверсии.

Шаг 1. $I \leftarrow [nU + 1]$. (Теперь число I таково, что $(I - 1)/n \leqslant U < I/n$.)

Шаг. 2 $I \leftarrow g_{I-1}$. (Заглянуть в путеводитель.)

Шаг 3. Последовательно присваивать I назначение $I + 1$ до тех пор, пока не окажется, что $c_{I+1} < U$.

Шаг 4. Вывести I.

Самое интересное в этом алгоритме то, что цикл на шаге 3 в среднем выполняется почти мгновенно. Это вызвано тем, что n значений c_i распределены по n интервалам и что величина U распределена равномерно на [0, 1]. Сам путеводитель можно построить за время порядка n следующим образом:

Шаг 1. При $i = 1, \ldots, n$ положить $g_i \leftarrow 0$.

Шаг 2. При $j = 1, \ldots, n$ положить $i \leftarrow [nc_j + 1]$, $g_i \leftarrow j$. (Заметим, что $(i - 1)/n \leqslant c_j < i/n$.)

Шаг 3. При $i = 2, \ldots, n$ последовательно присваивать g_i значение $\max(g_{i-1}, g_i)$. (Доопределение на пустых интервалах).

Заметим, что для гистограммных оценок с зависящими от данных точками разрыва (точнее точками разрыва, совпадающими с k-й, $2k$-й и т. д. порядковыми статистиками выборки) метод инверсии исключительно прост в реализации (Fox (1980)). В работе Archer (1980) также рассматриваются простые генераторы для f_n и предлагается кусочно постоянная оценка плотности, точки разрыва и величины скачков которой определяются из условия совпадения моментов оценки f_n с выборочными моментами. К сожалению, в общем случае такой подход не приводит к состоятельным оценкам. Метод генерирования случайных величин в R^d без построения f_n в явном виде указан в Thompson, Taylor (1982).

ЛИТЕРАТУРА

Ahrens J. H., Kohrt K. D.
(1981) Computer methods for efficient sampling from largely arbitrary statistical distributions, Computing, 26, pp. 19—31.

Archer N. P.
(1980) The generation of piecewise linear approximations of probability distribution functions, Journal of Statistical Computation and Simulation, 11, pp. 21—40.

Breiman L., Meisel W., Purcell E.
(1977) Variable kernel estimates of multivariate densities, Technometrics, 19, pp. 135—144.

Bretagnolle J., Huber C.
(1979) Estimation des densités: risque minimax, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 47, pp. 119—137.

Chen H. C., Asau Y.
(1974) On generating random variates from an empirical distribution, AIIE Transactions, 6, pp. 163—166.

Csiszar I.
(1967) Information-type measures of difference of probability distributions and indirect observations, Studia Scientiarium Mathematicarum Hungarica, 2, pp. 299—318.

Deak I.
(1979) Comparison of methods for generating uniformly distributed random points in and on a hypersphere, Problems of Control and Information Theory, 8, pp. 105—113.

Deheuvels P.
(1977a) Estimation non paramétrique de la densité par histogrammes généralisés, Revue de Statistique Appliquée, 25, pp. 5—42.
(1977b) Estimation non paramétrique de la densité par histogrammes généralisés, Publications de l'ISUP, 22, pp. 1—23.

Devroye L.
(1982) A note on approximations in random variate generation, Journal of Statistical Computation and Simulation, 14, pp. 149—158.

Fox B. L.
(1980) Monotonicity, extremal correlations, and synchronization: implications for nonuniform random numbers, Technical Report, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Canada.

Hominal P., Deheuvels P.
(1979) Estimation non paramétrique de la densité compte-tenu d'informations sur le support, Revue de Statistique Appliquée, 27, pp. 47—68.

Kemperman J. H. B.
(1969) On the optimum rate of transmiting information, Probability and Information Theory, Springer Lecture Notes in Mathematics, 89, Springer-Verlag, Berlin, pp. 126—169.

Knuth D. E.
(1975) The Art of Computer Programming, Vol. 3; Sorting and Searching, Addison-Wesley, Reading, Massachusetts. (Имеется русский перевод: Кнут Д. Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск. — М.: Мир, 1978.)

Kronmal R. A., Peterson A. V.
(1979) On the alias method for generating random variables from a discrete distribution, The American Statistician, 33, pp. 214—218.

Kullback S.
(1967) A lower bound for discrimination information in terms of variation, IEEE Transactions on Information Theory, 13, pp. 126—127.

LeCam L.
(1973) Convergence of estimates under dimensionality restrictions, Annals of Statistics, 1, pp. 38—53.

Pitman E. J. G.
(1979) Some Basic Theory for Statistical Inference, Chapman and Hall, London. (Имеется русский перевод: Питмен Э. Основы теории статистических выводов. — М.: Мир, 1986.)

Rubinstein R.
(1980) Generating random vectors uniformly distributed inside and on the surface of different regions, IBM Thomas J. Watson Research Center, Technical Report RC 8409.
(1981) Simulation and the Monte Carlo Method, Wiley, New York.
Schmeiser B. W.
(1980) Random variate generation: a survey, Proceedings of the 1980 Winter Simulation Conference, Orlando, Florida.
Schmeiser B. W., Shalaby M. A.
(1980) Acceptance/rejection methods for beta variate generation, Journal of the American Statistical Association, 75, pp. 673—678.
Schmeiser B. W., Babu A. J. G.
(1980) Beta variate generation via exponential majorizing functions, Operations Research, 28, pp. 917—926.
Serfling R. J.
(1979) A variation on Scheffé's theorem, with application to nonparametric density estimation, Report M502, Department of Statistics, Florida State University.
Shanmugam K. S.
(1977) On a modified form of Parzen estimator for nonparametric pattern recognition, Pattern Recognition, 9, pp. 167—170.
Sibuya M.
(1962) A method for generating uniformly distributed points on n-dimensional spheres, Annals of the Institute of Statistical Mathematics, 44, pp. 81—85.
Tashiro Y.
(1977) On methods for generating uniform random points on the surface of a sphere, Annals of the Institute of Statistical Mathematics, 29, pp. 295—300.
Thompson J. R., Taylor M. S.
(1982) A data-based random number generator for a multivariate distribution, Proceedings of the NASA Workshop on Density Estimation and Function Smoothing, held at Texas A & M University, College Station, Texas, pp. 214—225.
Von Neumann J.
(1951) Various techniques in connection with random digits, National Bureau of Standards AMS, 12, pp. 36—38.
Walker A. J.
(1977) An efficient method for generating discrete random variables with general distributions, ACM Transactions on Mathematical Software, 3, pp. 253—256.

Глава 9

Преобразованная ядерная оценка

1. Введение

Ядерная оценка

$$f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

обладает тем недостатком, что параметр h в ней не настраивается локально. Это находит отражение в результатах гл. 5, где показано, что поведение ядерной оценки ухудшается для менее гладких плотностей f и для плотностей с тяжелыми хвостами. Можно несколько смягчить возникающие здесь проблемы, если строить оценку плотности преобразованной случайной величины, а затем брать обратное преобразование.

В *преобразованной ядерной оценке* (Devroye и др. (1983)) используется строго монотонно возрастающее взаимно однозначное непрерывно дифференцируемое преобразование T, отображающее R^1 на $[0, 1]$ и имеющее непрерывно дифференцируемое обратное преобразование. Преобразованная последовательность данных имеет вид $Y_1, \ldots, Y_n$, где $Y_i = T(X_i)$. Заметим, что плотностью случайной величины Y_1 является

$$g(x) = f(T^{-1}(x))\,(T^{-1}(x))'.$$

Если g_n — оценка плотности g по данным $Y_1, \ldots, Y_n$, то оценка для f имеет вид

$$f_n(x) = g_n(T(x))\,T'(x). \tag{2}$$

Самое главное, что если g_n — плотность на $[0, 1]$, то f_n — плотность на R^1 и, более того,

$$\int |f_n - f| = \int |g_n - g|.$$

Иными словами, L_1-ошибка инвариантна относительно монотонных преобразований. Далее, если g_n — ядерная оценка, то ее L_1-ошибка пропорциональна величине $B^*(g)$, введенной в гл. 5. Следовательно, надо выбирать T так, чтобы минимизировать $B^*(g)$. В связи с этим возникает естественный вопрос: какая плотность g является наилучшей? Ответ на этот вопрос, данный в теореме 5.3, гласит: минимальное значение $B^*(g)$ равно $(2^9/3^4)^{1/5}$ и оно достигается для равнобедренной треугольной плот-

ности на [0, 1]. Поэтому надо выбирать T так, чтобы плотность g была равнобедренной треугольной. Если бы функции распределения F, соответствующая плотности f, была известна, то такое оптимальное преобразование имело бы вид

$$T(x) = \begin{cases} \sqrt{F(x)/2}, & F(x) \leqslant 1/2, \\ 1 - \sqrt{(1-F(x))/2}, & F(x) > 1/2. \end{cases} \tag{3}$$

Вспомним, что если g — треугольная плотность и K — ядро Епанечникова, то оптимальное значение h из табл. 5.1 равно $(5/192\pi n)^{1/5}$. Таким образом, оценки g_n и f_n полностью определены при известной F. К сожалению, функция F не известна и должна быть заменена некоторой оценкой. Кроме того, функция g_n обычно не является плотностью на [0, 1], поскольку часть массы g_n находится за пределами отрезка [0, 1]. Учитывая это, будем пользоваться оценкой

$$\tilde{g}_n(x) = \frac{g_n(x)}{\int\limits_0^1 g_n(y)\,dy} \tag{4}$$

вместо g_n. Вычисление интеграла в (4) не составляет труда, ибо

$$\int\limits_0^1 g_n(x) = \frac{1}{n} \sum_{i=1}^{n} \int\limits_{(Y_i-1)/h}^{Y_i/h} K(x)\,dx. \tag{5}$$

Если положить $f_n(x) = \tilde{g}_n(T(x))\,T'(x)$, то, очевидно, $\int |f_n - f| = \int |\tilde{g}_n - g|$. И хотя теоретические результаты, исходя из которых выбиралась плотность g, верны для g_n, их можно использовать и для $\tilde{g}_n$, так как

$$\int |\tilde{g}_n - g| \leqslant \int |g_n - g| \tag{6}$$

при любых g и g_n (см. теорему 11.3).

Единственным неизвестным элементом конструкции осталось преобразование T. Заметим, что в случае преобразованной гистограммной оценки оптимальное преобразование T должно приводить к равномерной плотности на [0, 1], так что оно имеет вид $T(x) = F(x)$ при любых x. Значение h, которое следует использовать в гистограммной оценке, равно $(2\pi n)^{-1/3}$ (табл. 5.1).

Так как T-гладкая функция, то плотность g наследует гладкость плотности f. Поэтому преобразованные оценки едва ли могут сильно помочь в борьбе с разрывами. Преобразования используются в основном, чтобы улучшить поведение исходных оценок на хвостах. Оптимальное обратное преобразование растягивает

ядра на хвостах. Визуальный эффект при этом такой же, как в методе с переменным h, зависящим от x: «видно», что параметр h, как правило, больше на хвостах и меньше вблизи моды.

Еще один важный момент состоит в том, что оценку $\tilde{g}_n$ легко строить графически, поскольку она имеет компактный носитель. Из аналогичных соображений была введена и квантильная функция Парзена для плотностей (Parzen (1979)).

Имеется довольно мало работ, посвященных выбору h. Вместе с тем мы знаем, что средняя L_1-ошибка не может убывать быстрее, чем произведение $n^{-2/5}\ B^*(f)$ на некоторую постоянную. В данной главе предлагается не обращать внимания на h, а попытаться расширить возможности ядерной оценки, работая с множителем $B^*(f)$.

2. Выбор преобразования

Выбор преобразования — дело не простое. Во многих приложениях предполагают, что плотность f принадлежит некоторому семейству плотностей (как правило, параметрическому) или по крайней мере близка к определенному элементу такого семейства. Если θ — параметр семейства и F_θ — соответствующая функция распределения, то естественно найти некоторую *робастную* оценку $\hat{\theta}$ параметра θ и подставить $F_{\hat{\theta}}$ в выражение для оптимального преобразования T. Здесь и везде далее используется одно и то же значение h, оптимальное для равнобедренной треугольной плотности на $[0, 1]$.

Особенно привлекательны так называемые «простые и грубые» [1]) робастные оценки, идея которых принадлежит Гаствирту (Gastwirth (1966)). Например, если $X_{(1)} < \ldots < X_{(n)}$ — порядковые статистики выборки $X_1, \ldots, X_n$, то оценка Гаствирта для среднего нормального семейства равна

$$\hat{\mu} = 0{,}3X_{(n/3)} + 0{,}4X_{(n/2)} + 0{,}3X_{(2n/3)}.$$

В работах Huber (1972) и Andrews и др. (1972) приводятся дальнейшие примеры таких простых робастных оценок сдвига. Для робастного оценивания параметра масштаба можно использовать двухквантильный метод из § 6 гл. 5. Этот метод дает, например, следующую оценку параметра масштаба семейства Коши:

$$\hat{\sigma} = (X_{(3n/4)} - X_{(n/4)})/2.$$

За исключением тривиальных случаев, нормальное семейство и семейство Коши слишком бедны для описания статистических моделей. Обзор семейств с большим числом параметров дается в работе Schmeiser (1977). Мы не рекомендуем непараметрические

[1]) В оригинале quick and dirty. — *Прим. перев.*

семейства, поскольку для них трудно (а быть может, и невозможно) показать, что получаемые оценки состоятельны и обладают некоторой скоростью сходимости. Если параметров мало, то функция T — всегда примерно одна и та же, и потому нет риска потерять состоятельность.

3. Оценивание плотностей с тяжелыми хвостами

Эффективность ядерной оценки определяется двумя факторами: во-первых, наличием разрывов или резких осцилляций у оцениваемой плотности и, во-вторых, ее тяжелыми хвостами. Первый фактор для гладких плотностей количественно выражается величиной $\int |f''|$, а для плотностей с простыми разрывами, таких, как равномерная плотность на $[0, 1]$, равен бесконечности. Второй фактор, измеряемый величиной $\int \sqrt{f}$, равен бесконечности для плотностей с тяжелыми хвостами, таких, как плотность распределения Коши. Выше было показано, что если один или оба этих фактора бесконечны, то для обычной ядерной оценки имеем $n^{2/5}E(J_n) \to \infty$ независимо от выбора h как функции от n.

Преобразование к треугольной плотности устраняет разрывы и хвосты, и потому должно улучшать качество оценки в смысле критерия L_1. При обратном преобразовании разрывы восстанавливаются и создается впечатление, что преобразованная оценка работает так же, как ядерная оценка с локально настраиваемым параметром сглаживания h. Так, значения h, как правило, кажутся больше на хвостах. Для описания этого явления воспользуемся понятием изолированного холма [1]). С помощью данного понятия будет отчасти объяснено получаемое уменьшение L_1-ошибки.

Изолированный холм *произвольной* оценки плотности ассоциируется с одной из точек выборки $X_1, \ldots, X_n$, а именно считается, что изолированный холм расположен в точке X_i, если существует отрезок $[a, b]$, такой, что $X_i \in [a, b]$, ни одна из других точек X_j не принадлежит $[a, b]$, $\int_a^b f_n > 0$ и $f_n = 0$ на $[a - \varepsilon, a) \cup (b, b + \varepsilon]$ при некотором $\varepsilon > 0$. Предположим, например, что используется ядерная оценка с ядром Епанечникова. Тогда изолированный холм расположен в точке X_i в том и только том случае, когда отрезок $[X_i - 2h, X_i + 2h]$ не содержит никаких точек выборки, кроме X_i. Таким образом, часть графика функции f_n, относящаяся к отрезку $[X_i - h, X_i + h]$, имеет вид изолированного холма, и создается впечатление, что X_i — как бы бес-

[1]) В оригинале isolated bump. — *Прим. перев.*

полезная точка выборки. Заметим также, что число изолированных холмов инвариантно относительно строго монотонных преобразований, рассматриваемых в данной главе.

Общее число изолированных холмов B_n дает границу снизу для числа локальных максимумов графика оценки. Например, при оценивании унимодальной плотности желательно, чтобы был лишь один изолированный максимум и чтобы выполнялось условие $B_n = 0$. В этом параграфе мы покажем, что такого обычно не бывает. Так, если для нормальной плотности рассматривается оценка с оптимальным h, то $E(B_n)$ возрастает по крайней мере как $n^{1/5}/\sqrt{\log n}$, а для плотностей с тяжелыми хвостами ситуация еще хуже. Мы покажем также, что для треугольной плотности $E(B_n) = o(1)$.

Основная исходная формула имеет вид [1])

$$E(B_n) = nP(\text{отрезок } [X_1 - 2h,\ X_1 + 2h] \text{ не содержит точек } X_2, \ldots, X_n) = n \int f(x) \left(1 - \int_{x-2h}^{x+2h} f(y)\,dy\right)^{n-1} dx.$$

Теорема 1 (общий результат). *При любых f имеем $E(B_n) = o(n)$, если $h \to 0$, $nh \to \infty$.*

Доказательство. Заметим, что $E(B_n)/n = \int f(x)\, r_n(x)\, dx$, где функция $r_n(x)$ со значениями в $[0, 1]$ такова, что

$$r_n(x) \leqslant \exp\left(-(n-1)\int_{x-2h}^{x+2h} f(y)\,dy\right) \to 0 \text{ при почти всех } x$$

(это следует из теоремы Лебега о плотностях (теорема 2.2), в силу которой выражение под экспонентой асимптотически равно $4nhf(x)$ при почти всех x). Теперь для доказательства теоремы 1 достаточно применить теорему Лебега о мажорируемой сходимости.

Теорема 2 (случай плотностей с правильно меняющимися хвостами). *Пусть плотность f строго монотонно убывает на $[0, \infty)$ и имеет там однозначно определенную обратную функцию; положим для удобства $f = 0$ на $(-\infty, 0)$. Предположим также, что f - правильно меняющаяся функция порядка $r < -1$ на ∞, т. е.*

$$\lim_{x \to \infty} f(tx)/f(x) = t^r$$

---

[1]) Эта формула и последующие результаты данного параграфа получены в предположении, что f_n — ядерная оценка, носитель ядра которой содержится в $[-1, 1]$. — *Прим. перев.*

при любых $t > 0$. *Если* $h \to 0$, $nh \to \infty$, *то*

$$E(B_n) \geqslant \frac{L(n)}{(nh)^{1/r} h},$$

где L — *некоторая медленно меняющаяся функция* (*т. е. правильно меняющаяся функция порядка* 0).

Доказательство. Воспользуемся следующими фактами:

(i) $\lim\limits_{x \to \infty} xf(x) \Big/ \int\limits_x^\infty f(y)\, dy = -r - 1$ (Dehaan (1975), теорема 1.2.1);

(ii) функция $f^{-1}(1/x)$ — правильно меняющаяся порядка $-1/r$ на ∞ (Dehaan (1975), с. 22);

(iii) $f(x) = x^r L(x)$, где L — некоторая медленно меняющаяся функция (Seneta (1976), лемма 2.1).

Символ L будем использовать для обозначения любых медленно меняющихся функций. Обозначим A_n множество всех x, для которых $n \int\limits_{x-2h}^{x+2h} f(y)\, dy \leqslant 1/2$, и A_n^* — множество всех $x > 2h$ для которых $4nhf(x - 2h) < 1/2$. Теорема 2 следует из соотношений

$$E(B_n) \geqslant n \int\limits_0^\infty f(x) \left(1 - n \int\limits_{x-2h}^{x+2h} f(y)\, dy\right)_+ dx \geqslant \frac{n}{2} \int\limits_{A_n} f \geqslant \frac{n}{2} \int\limits_{A_n^*} f \sim$$

$$\sim \frac{-n}{2(r+1)} \left(2h + f^{-1}\left(\frac{1}{8nh}\right)\right) f\left(2h + f^{-1}\left(\frac{1}{8nh}\right)\right) \sim$$

$$\sim \frac{-n}{2(r+1)} f^{-1}\left(\frac{1}{8nh}\right) f\left(f^{-1}\left(\frac{1}{8nh}\right)\right) =$$

$$= \frac{-n}{2(r+1)} f^{-1}\left(\frac{1}{8nh}\right) \frac{1}{8nh} = \frac{L(8nh)\,(8nh)^{-1/r}}{-16(r+1)h}.$$

Пример. Для плотности $c/(1 + x^2)^{(a+1)/2}$, $x \geqslant 0$, $a > 0$, положительного t-распределения Стьюдента имеем $r = -(a + 1)$, так что

$$E(B_n) \geqslant L(n)\, h^{-1} (nh)^{1/(a+1)} = L(n)\, n^{1/(a+1)}\, h^{-a/(a+1)}.$$

Для фиксированных $\varepsilon > 0$, $\beta \in (0, 1)$ найдется a, такое, что $(1 + a\beta)/(1 + a) > 1 - \varepsilon/2$. Следовательно, для любого $\varepsilon > 0$ и любой последовательности $h \sim c/n^\beta$, $\beta \in (0, 1)$, существует a, такое, что $E(B_n) \geqslant L(n)\, n^{1-\varepsilon/2} > n^{1-\varepsilon}$ при достаточно больших n. Последнее неравенство следует из свойства медленно меняющихся функций (Seneta (1976) с. 38 по русскому изданию). Таким об-

разом, для любой полиномиально убывающей и удовлетворяющей условиям $h \to 0$, $nh \to \infty$ последовательности h можно выбрать соответствующую плотность f из семейства t-распределений Стьюдента так, чтобы величина $E(B_n)$ возрастала с заданной полиномиальной скоростью $n^{1-\varepsilon}$. В частности, если h убывает как $n^{-1/5}$, то имеем $E(B_n) \geqslant L(n)\, n^{(1+a/5)/(1+a)}$. Здесь показатель степени изменяется в пределах от 1 ($a \downarrow 0$) до 1/5 ($a \to \infty$).

Теорема 3 (случай равнобедренной треугольной плотности). *Если f — равнобедренная треугольная плотность на* [0, 1], *то*

$$E(B_n) \leqslant (1 + o(1))\left((32nh^2)^{-1} + 32nh^2 e^{-8nh^2}\right)$$

при $h \to 0$, $nh \to \infty$. В частности, если $nh^2 \to \infty$, то $E(B_n) \to 0$.

Доказательство. Прежде всего имеем

$$E(B_n) \leqslant 8n \int_0^{1/2} x \exp\left(-(n-1)\int_{x-2h}^{x+2h} f(y)\,dy\right) dx.$$

Разобьем этот интеграл на три части, соответствующие отрезкам [0, $2h$], [$2h$, 1/2 — $2h$] и [1/2 — $2h$, 1/2]. Величина первой части не превосходит $8n \cdot 2h \exp(-2(n-1)(2h)^2) \cdot 2h = (32 + o(1))\, nh^2 \exp(-8\, nh^2)$. Третья часть не превосходит $2n \cdot 2 \times 2h \exp(-(n-1) \cdot 4h \cdot 4\ (1/2 - 2h)) = (8 + o(1))\, nh \exp(-8\, nh)$. Вторая часть не превосходит

$$2\int_{2h}^{\infty} 4nxe^{-16xh\,(n-1)}\,dx =$$

$$= (32h^2(n-1))^{-1}\,\frac{n}{n-1}\int_{32\,(n-1)\,h^2}^{\infty} ye^{-y}\,dy \leqslant \frac{1+o(1)}{32nh^2}.$$

Теорема 3 доказана

В силу теоремы 3 ядерная оценка с большой вероятностью не имеет изолированных холмов, если f — треугольная плотность (в самом деле, $E(B_n) = O(n^{-3/5})$ при $h \sim n^{-1/5}$). То же самое верно для преобразованной ядерной оценки, если преобразование «совершенное». При этом не только сокращается число изолированных холмов, но и уменьшаются осцилляции оценки.

Наконец, заметим без доказательства, что в случае нормальной плотности величина $E(B_n)$ при $h \to 0$, $nh \to \infty$ больше или равна отношению некоторой постоянной к $h\sqrt{\log(nh)}$ и, значит, при оптимальном выборе h эта величина возрастает не медленнее, чем $n^{1/5}/\sqrt{\log n}$. Таким образом, преобразование к нормальной плотности не дает выигрыша по сравнению с преобразованием к треугольной плотности.

4. Состоятельность

При фиксированном преобразовании T: $R^1 \to R^1$, удовлетворяющем условиям § 1, имеем $J_n = \int |g_n - g|$, так что, несомненно, можно воспользоваться экспоненциальной границей для J_n из теоремы 3.1. Эта граница имеет вид $P(J_n \geqslant \varepsilon) \leqslant \leqslant \exp(-cn)$ при любых $n > n_0$, где $c > 0$ — некоторая функция от ε и n_0 — число, зависящее от g и ε. Более того, остается в силе нижняя граница теоремы 5.2, тогда как верхняя граница $C^*A(K)B^*(f)$ для $E(J_n)\, n^{2/5}$ (теорема 5.1) уже неприменима, поскольку значение h может быть неоптимальным для g. Все же, выбирая $h = (5/192\pi n)^{1/5}$, можно гарантировать, что $E(J_n)$ убывает как $n^{-2/5}$ при $B^*(g) < \infty$. Напомним также, что величина $E(J_n)$ относительно нечувствительна к малым отклонениям от оптимального выбора h (см. § 5.6).

При переменном преобразовании T следует позаботиться, чтобы получаемая оценка была состоятельной.

Преобразование $Y_i = T(X_i)$ обычно имеет вид

$$Y_i = T_n(X_i;\ X_1, \ldots, X_n),$$

где T_n — борелевская функция, такая, что по первому аргументу она строго монотонно возрастает, непрерывно дифференцируема и отображает R^1 на $[0, 1]$. Обратная к ней функция тоже непрерывно дифференцируема.

Рассмотрим преобразованную ядерную оценку с ядром K Епанечникова и параметром сглаживания $h = (5/6\pi n)^{1/5}/2$, оптимальным для треугольной плотности на $[0,1]$. Не будем пока интересоваться выбором преобразований T_n: $R^1 \to [0, 1]$ и соответствующими нормировками, поскольку, как показано выше, это не сказывается на асимптотике. Имеем следующие плотности:

f — плотность распределения данных $X_1, \ldots, X_n$.

g — плотность распределения случайной величины $Y = T_n(X)$ при фиксированных $X_1, \ldots, X_n$, где случайная величина X не зависит от $X_1, \ldots, X_n$ и распределена так же, как X_1

g^* — плотность распределения случайной величины $T(X_1)$, где T — некоторое заданное преобразование.

g_n — преобразованная ядерная оценка, основанная на $Y_1, \ldots, Y_n$.

g_n^* — преобразованная ядерная оценка, основанная на $Z_i = T(X_i)$, $1 \leqslant i \leqslant n$.

При доказательстве сходимости $\int |g_n - g|$ к 0 оказывается полезным следующее неравенство:

$$\int |g_n - g| \leqslant \int |g^* - g| + \int |g_n^* - g^*| + 6 \sup_x |T_n(x) - T(x)|.$$

Чтобы убедиться в справедливости этого неравенства, заметим, что K — ядро Епанечникова, и потому

$$\int |g_n - g_n^*| \leqslant (nh)^{-1} \int \sum_{i=1}^{n} \left| K\left(\frac{y - Y_i}{h}\right) - K\left(\frac{y - Z_i}{h}\right) \right| dy \leqslant$$

$$\leqslant \frac{6}{n} \sum_{i=1}^{n} |Y_i - Z_i| \leqslant 6 \sup_x |T_n(x) - T(x)|$$

В качестве преобразования T рассматривается определенный предел последовательности T_n. В частности, если T_n получается путем оценивания параметров, то истинная форма преобразования T известна. Хочется надеяться (и эта надежда обычно оправдана), что основной вклад в оценку величины $\int |g_n - g|$ дает среднее из трех слагаемых в правой части неравенства и что потому при исследовании скоростей сходимости можно по существу заменить T_n фиксированным преобразованием T.

Некоторые сложности возникают со слагаемыми $\int |g^* - g|$. Известно, что оно стремится к 0 по вероятности, если подынтегральное выражение стремится к 0 по вероятности при почти всех x (теорема 2.8). Но последнее имеет место, если плотность f п. в. непрерывна, $T_n^{-1} \to T^{-1}$ п. в. и $(T_n^{-1})' \to (T^-)'$ п. в. по вероятности.

ЛИТЕРАТУРА

Andrews D. F., Bickel P. J., Hampel F. R., Huber P. J., Rogers W. H., Tukey J. W.
(1972) Robust Estimates of Location: Survey and Advances, Princeton University Press, Princeton.

Dehaan L.
(1975) On Regular Variation and Its Applications to the Weak Convergence of Sample Extremes, Mathematisch Centrum Tracts 32, Mathematisch Centrum Amsterdam.

Devroye L., Machell F., Penrod C. S.
(1983) The transformed kernel estimate, Technical Report, Applied Research Laboratories, University of Texas, Austin, Texas.

Gastwirth J. L.
(1966) On robust procedures, Journal of the American Statistical Association, 61, pp. 929—948.
Huber P. J.
(1972) Robust statistics: a review, Annals of Mathematical Statistics, 43, pp. 1041—1067.
Parzen E.
(1979) Nonparametric statistical data modeling, Journal of the American Statistical Association, 74, pp. 105—131.
Schmeiser B. W.
(1977) Methods for modelling and generating probabilistic components in digital computer simulation when the standard distributions are not adequate: a survey, Proceedings of the Winter Simulation Conference, pp. 51—55.
Seneta E.
(1976) Regularly Varying Functions, Lecture Notes in Mathematics, 508, Springer-Verlag, Berlin. (Имеется русский перевод: Сенета Е. Правильно меняющиеся функции. — М.: Наука, 1985).

Глава 10

Применения к задаче классификации

1. Задача классификации

Задача классификации (дискриминантного анализа, статистического распознавания образов) обычно формулируется следующим образом. Пусть наблюдение X — случайная величина со значениями в R^d и метка Y — случайная величина со значениями в $\{1, \ldots, M\}$. Требуется по заданному X угадать значение Y, т. е. принять решение. *Решением* [1]) называется измеримая функция g: $R^d \to \{1, \ldots, M\}$, а вероятностью ошибки — величина $P(g(X) \neq Y)$. Распределение пары (X, Y) определяется вероятностной мерой μ, с которой распределен вектор X, и функциями регрессии

$$p_i(x) = P(Y = i \mid X = x), \quad x \in R^d, \quad 1 \leqslant i \leqslant M,$$

которые также называют *апостериорными вероятностями*. Решение g^* называется *байесовским*, если

$$p_{g^*(x)}(x) = \max_i p_i(x) \tag{1}$$

при почти всех x по мере μ. Если случайный вектор X имеет плотность распределения f и условные плотности f_i при фиксированном $Y = i$, $1 \leqslant i \leqslant M$, то

$$p_i(x) = \frac{p_i f_i(x)}{f(x)}$$

при почти всех x по мере μ, где $p_i = P(Y = i)$. Таким образом, для байесовского решения имеем

$$p_{g^*(x)} f_{g^*(x)}(x) = \max_i p_i f_i(x) \tag{2}$$

при почти всех x по мере, соответствующей плотности f.

В задаче классификации требуется минимизировать вероятность ошибки в случае, когда p_i и f_i не известны и имеется выборка $D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, состоящая из n независимых случайных векторов, распределенных так же, как (X, Y). Предполагается, что D_n и (X, Y) независимы. Оценкой для Y

---

[1]) Наряду с термином *решение* в литературе используются термины *решающая функция* и *решающее правило*. — *Прим. перев.*

теперь является $g_n(X)$ — измеримая функция от X и D_n (зависимость от D_n в обозначении не указывается), и представляет интерес величина условной вероятности ошибки

$$L_n = P(g_n(X) \neq Y \mid D_n). \tag{3}$$

В частности, хотелось бы найти последовательности функций g_n, для которых

$$L_n \to L^* = \min_g P(g(X) \neq Y) \tag{4}$$

почти наверное. Здесь L^* обычно называется *байесовской вероятностью ошибки*. Данная глава посвящена именно этому, т. е. различным способом выбора последовательностей g_n, для которых выполнено (4). Иные свойства последовательностей g_n подробно рассматриваться не будут. Наша цель, скорее, состоит в том, чтобы показать, что (4) следует из полученных в гл. 1—9 результатов об оценках плотности. При этом мы будем считать, что случайный вектор X имеет плотность распределения f. Вместе с тем подчеркнем, что большинство результатов, сформулированных в данной главе, остается в силе для случая, когда вектор X распределен с произвольной вероятностной мерой μ на борелевских множествах в R^d.

Байесовское решение аппроксимируется функцией g_n, такой, что

$$\tilde{p}_{g_n(x)}(x) = \max_i \tilde{p}_i(x), \tag{5}$$

где $\tilde{p}_i$ — некоторые предварительные оценки для p_i, т. е. функции от D_n со значениями в $[0, 1]$.

Если случайный вектор X имеет плотность распределения f и $\tilde{p}_i \tilde{f}_i(x)$ — оценка для $p_i f_i(x)$ по выборке D_n, то $g_n(x)$ можно определить соотношением

$$\tilde{p}_{g_n(x)}(x)\, \tilde{f}_{g_n(x)}(x) = \max_i \tilde{p}_i \tilde{f}_i(x). \tag{6}$$

Теорема 1. (i) *Если* g^* *— байесовское решение, то* $L^* = P(g^*(X) \neq Y)$.

(ii) *Если* g_n *удовлетворяет условию* (5), *то*

$$0 \leqslant L_n - L^* \leqslant \sum_{i=1}^{M} \int |p_i(x) - \tilde{p}_i(x)|\, \mu(dx).$$

(iii) *Если случайный вектор* X *имеет плотность распределения и* g_n *удовлетворяет условию* (6), *то*

$$0 \leqslant L_n - L^* \leqslant \sum_{i=1}^{M} \int |p_i f_i(x) - \tilde{p}_i \tilde{f}_i(x)|\, dx.$$

Замечание 1. Различные варианты соотношений (ii) и (iii) были доказаны в работах Van Ryzin (1966), Wolverton, Wagner (1969), Csibi (1975), Györfi (1974, 1978), Devroye, Wagner (1976) и Devroye (1982b).

Доказательство. В силу (1)

$$P(g^*(X) \neq Y) = 1 - \sum_{i=1}^{M} P(Y = i,\ g^*(X) = i) =$$

$$= 1 - \sum_{i=1}^{M} \int_{[g^*(x)=i]} p_i(x)\,\mu(dx) =$$

$$= 1 - \int \max_i p_i(x)\,\mu(dx) \leqslant 1 - \int p_{g(x)}(x)\,\mu(dx) \qquad (7)$$

для любой функции g: $R^d \to \{1, \ldots, M\}$. Взяв нижнюю грань по всем g, получим (i). Кроме того

$$L_n = 1 - \sum_{i=1}^{M} \int_{[g_n(x)=i]} p_i(x)\,\mu(dx) = 1 - \int p_{g_n(x)}(x)\,\mu(dx), \qquad (8)$$

так что, объединяя (7), (i) и (8), будем иметь

$$L_n - L^* = \int \left(\max_i p_i(x) - p_{g_n(x)}(x)\right)\mu(dx) =$$

$$= \int \left(\max_i p_i(x) - \max_i \tilde{p}_i(x)\right)\mu(dx) +$$

$$+ \int \left(\tilde{p}_{g_n(x)}(x) - p_{g_n(x)}(x)\right)\mu(dx) \leqslant$$

$$\leqslant \sum_{i=1}^{M} \int |p_i(x) - \tilde{p}_i(x)|\,\mu(dx), \qquad (9)$$

откуда следует (ii). Утверждение (iii) легко следует из (ii), ибо $\tilde{p}_i(x) = \tilde{p}_i \tilde{f}_i(x)/f(x)$.

2. Медленные скорости сходимости

То, что вероятность ошибки ограничена сверху L_1-ошибками оценок плотности $\tilde{f}_i$, довольно интересно. Но существуют и разного рода обратные утверждения, например результат о том, что вероятность ошибки L_n может стремиться к L^* со сколь угодно медленной скоростью (Devroye (1982a)).

Теорема 2. *Пусть a_n — последовательность положительных чисел, стремящаяся к 0, $M = 2$ и $c \in [0, 1/2)$. Пусть g_n — произвольное решение. Тогда существует распределение пары*

(X, Y), *где* X — *равномерно распределенная на* $[0, 1]$ *случайная величина, такое, что*

$$\limsup_{n\to\infty} \frac{E(L_n) - c}{a_n} = \infty$$

при $L^* = c$.

Доказательство следует общему плану доказательства утверждения (ii) теоремы 4.1. Опишем лишь в основных чертах конструкцию рандомизированного семейства для частного случая $c = 0$. Пусть случайная величина X имеет равномерную плотность распределения на $[0, 1]$, и пусть $b = 0, b_1b_2b_3 \ldots \in [0, 1]$ (b_i — коэффициенты в двоичном разложении для b). Положим

$$p_2(x) = f_b(x) = \sum_{i=1}^{\infty} b_i I_{[x_i, x_{i+1})}(x), \quad p_1(x) = 1 - p_2(x),$$

$$Y = 1 + f_b(X), \quad Y_n = 1 + f_b(X_n),$$

где $0 = x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \uparrow 1$. Положим $q_i = x_{i+1} - x_i$, $i = 1, 2, \ldots$. Предположим, что случайная величина $B = 0, B_1B_1 \ldots$ равномерно распределена на $[0, 1]$ и не зависит от $X, X_1, X_2, \ldots, X_n$, и обозначим $R_n(b) = E(L_n)$, $b \in [0, 1]$. Имеем

$$\sup_{b\in[0,1]} R_n(b) \geqslant E(R_n(b)) = P(g_n(X, D_n) \neq Y)$$

$$= E(P(g_n(X, D_n) \neq Y \mid X, X_1, \ldots, X_n)) \geqslant$$

$$\geqslant E(I_A P(g_n(X, D_n) \neq Y \mid X, X_1, \ldots, X_n)),$$

где через A обозначено событие $\bigcap_{i=1}^{n} [f_B(X) f_B(X_i) = 0]$. На A случайные величины Y и $g_n(X, X_1, Y_1, \ldots, X_n, Y_n)$ независимы при фиксированных $X, X_1, X_2, \ldots, X_n$. Следовательно, на A имеем $P(g_n(X, D_n) \neq Y \mid X, X_1, \ldots, X_n) = 1/2$, и потому

$$\sup_{b\in[0,1]} R_n(b) \geqslant \frac{1}{2} P\left(\bigcap_{i=1}^{n} [f_B(X) f_B(X_i) = 0]\right) =$$

$$= \frac{1}{2} \sum_{i=1}^{\infty} q_i (1 - q_i)^n.$$

Остальные элементы доказательства такие же, как для утверждения (ii) теоремы 4.1.

Из теоремы 2 следует, что результаты о скорости сходимости величины $E(L_n)$ можно получить лишь в случае, когда предполагается определенная гладкость функции регрессии. Мы не будем исследовать вопрос об оптимальных скоростях сходимости

для различных семейств распределений пары (X, Y), а сосредоточим внимание на результатах о состоятельности наиболее широко используемых методов непараметрической классификации.

3. Ядерный метод классификации

Пусть K — функция из $L_1(R^d)$, такая, что $\int K = 1$. Рассмотрим модифицированную ядерную оценку плотности

$$\tilde{p}_i \tilde{f}_i(x) = (nh_n^d)^{-1} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right) I_{[Y_j = i]}, \quad 1 \leqslant i \leqslant M. \tag{10}$$

Соответствующее решение определяется соотношением (6). При некоторых дополнительных условиях на K в работах Devroye, Wagner (1980) и Spiegelman, Sacks (1980) показано, что $L_n \to L^*$ по вероятности при $n \to \infty$, если $h_n \to 0$, $nh_n^d \to \infty$. В работе Devroye (1981) получена сильная сходимость при дополнительном условии $nh_n^d / \log n \to \infty$. Во всех этих работах нет никаких условий на распределение (X, Y). Теорема 3, которая приводится ниже, верна, если случайный вектор X имеет плотность распределения; утверждается, что L_n сходится к L^* с экспоненциальной скоростью. Условия сходимости в ней неулучшаемы.

Теорема 3. *Если случайный вектор X имеет плотность распределения, $h_n \to 0$ и $nh_n^d \to \infty$, то для ядерного метода, определяемого соотношениями* (6) *и* (10), *верно следующее*

Для любого $\varepsilon \in (0, 1)$ существует $n_0 > 0$, такое, что

$$P(L_n - L^* > \varepsilon) \leqslant \exp(-c_1 n\varepsilon^2), \quad n \geqslant n_0.$$

Здесь $c_1 > 0$ — постоянная, зависящая только от K.

Доказательство. В силу теоремы 1 достаточно лишь показать, что

$$P\left(\sum_{i=1}^{M} \int |\tilde{p}_i \tilde{f}_i(x) - p_i f_i(x)|\, dx > \varepsilon\right) \leqslant \exp(-c_1 n\varepsilon^2), \quad n \geqslant n_0.$$

Так как $E(\tilde{p}_i \tilde{f}_i(x)) = p_i K_{h_n} * f_i(x)$, по теореме 2.1 имеем $\int |p_i f_i - E(\tilde{p}_i \tilde{f}_i)| \to 0$ при любых i. Таким образом, достаточно установить экспоненциальное неравенство для вероятности

$$P\left(\sum_{i=1}^{M} \int |\tilde{p}_i \tilde{f}_i(x) - E(\tilde{p}_i \tilde{f}_i(x))|\, dx > \varepsilon\right).$$

Как и в доказательстве леммы 3.2, достаточно рассматривать лишь ядра K, являющиеся индикаторными функциями прямоугольников A. Определим меры μ_i и μ_{ni} на борелевских множествах B следующим образом:

$$\mu_i(B) = P(Y_1 = i,\ X_1 \in B),$$

$$\mu_{ni}(B) = \frac{1}{n}\sum_{j=1}^{n} I_{[Y_j=i,\ X_j\in B]}, \quad 1 \leqslant i \leqslant M.$$

Тогда

$$\sum_{i=1}^{M} \int |\tilde{p}_i\tilde{f}_i(x) - E(\tilde{p}_i\tilde{f}_i(x))|\,dx =$$

$$= \sum_{i=1}^{M} |\mu_{ni}(x + h_nA) - \mu_i(x + h_nA)|\,dx.$$

Дальнейшее доказательство очевидно, поскольку можно применить тот же метод, что в лемме 3.2.

Предположительно теорема 3 остается в силе при *любых* распределениях пары (X, Y).

4. Гистограммный метод классификации

Рассмотрим теперь гистограммные решения, в которых пространство R^d разбивается на множества $A_{n1}, A_{n2}, \ldots$ и оценки в (6) имеют вид

$$\tilde{p}_i\tilde{f}_i(x) = \frac{1}{n}\sum_{m=1}^{n} \frac{I_{[X_m\in A_{nj},\ Y_m=i]}}{\lambda(A_{nj})}, \quad x \in A_{nj}. \tag{11}$$

Теорема 4. *Если случайный вектор X имеет плотность распределения, последовательность разбиений удовлетворяет условиям* (3.13)—(3.15) *и используется решение* (6) *с гистограммной оценкой* (11), *то верно следующее:*

Для любого $\varepsilon \in (0, 1)$ *существует* $n_0 > 0$, *такое, что*

$$P(L_n - L^* > \varepsilon) \leqslant \exp(-c_2 n\varepsilon^2), \quad n \geqslant n_0.$$

Здесь $c_2 > 0$ — *универсальная постоянная.*

Доказательство этой теоремы повторяет доказательство леммы 3.4 и потому здесь не приводится. Сходимость L_n к L^* вполне для любых распределений пары (X, Y) показана в работе Devroye, Györfi (1983). Условия сходимости по существу те же, что для слабой состоятельности, чем опять демонстрируется эквива-

лентность всех типов сходимости. По поводу гистограммных решений, в которых разбиение пространства R^d зависит от данных, см. работу Gordon, Olshen (1978) и ссылки в ней. В этой работе получены результаты о слабой состоятельности.

5. Метод ближайших соседей

Еще одно распространенное непараметрическое решение основано на понятии k ближайших соседей (Cover, Hart (1967)). При заданном X выборка D_n упорядочивается по возрастанию величин расстояний $\| X_i - X \|$. В результате получается вектор рангов $(R_1(X), \ldots, R_n(X))$, где $X_{R_i(X)}$ есть i-й ближайший сосед точки X. В случае совпадения расстояний выбор ближайшего соседа производится исходя из сравнения индексов. Заметим, что если случайный вектор X имеет некоторую плотность f, то вероятность совпадения равна нулю. Решение принимается большинством голосов среди $Y_{R_j(X)}$, $1 \leqslant j \leqslant k_n$, где k_n — некоторая последовательность целых чисел. Стоун (Stone (1977)) также рассматривает случай взвешенного голосования: i-й ближайший сосед голосует с весом v_{ni}, для каждого класса подсчитывается общая сумма голосов и решение принимается в пользу класса, для которого такая сумма максимальна. В случае когда решение принимается при одинаковых весах среди k_n ближайших соседей, заметим, что данная процедура эквивалентна использованию в (6) оценки плотности методом ближайших соседей (Fix, Hodges (1951, 1952), Loftsgaarden, Quesenberry (1965)), имеющей вид

$$\tilde{p}_i \tilde{f}_i(x) = \frac{1}{n} \sum_{j=1}^{k_n} \frac{I_{[Y_{R_j(x)} = i]}}{\lambda\left(S_x, \left\| x - X_{R_{k_n}(x)} \right\|\right)}. \tag{12}$$

Эквивалентность имеет место в силу того, что знаменатель в (12) — один и тот же при любых i. Докажем следующую теорему.

Теорема 5. *Пусть случайный вектор X имеет плотность распределения f, $k_n \to \infty$ и $k_n/n \to 0$. Тогда для решения* (6), *основанного на оценке ближайших соседей* (12), *верно следующее: Для любого $\varepsilon \in (0, 1)$ существует $n_0 > 0$, такое, что*

$$P(L_n - L^* > \varepsilon) \leqslant \exp(-c_3 n \varepsilon^2), \quad n \geqslant n_0,$$

где $c_3 > 0$ зависит только от размерности.

Прежде чем привести доказательство, дадим краткую историческую справку. Стоун (Stone (1977)) показал, что $L_n \to L^*$ по вероятности при любом распределении пары (X, Y), если $k_n \to \infty$ и $k_n/n \to 0$. Заметим здесь, что эти условия на k_n являются необ-

ходимыми. Исходя из поточечной сходимости, Деврой (Devroye (1982b)) показал, что $L_n \to L^*$ почти наверное при любом распределении пары (X, Y), если $k_n/n \to 0$ и $k_n/\log\log n \to \infty$. Бек (Beck (1979)) доказал теорему 5 при некоторых дополнительных условиях гладкости на f и на функции f_i. Приводимое здесь доказательство более краткое и общее. Кроме того, в нем не накладывается никаких условий на f. Итак, следующие условия эквивалентны:

(i) $\lim_{n\to\infty} k_n = \infty$, $\lim_{n\to\infty} k_n/n = 0$;

(ii) $L_n \to L^*$ по вероятности, если случайный вектор X имеет плотность распределения;

(iii) $L_n \to L^*$ с экспоненциальной скоростью (т. е. для любого $\varepsilon > 0$ существует $c > 0$, такое, что $P(L_n - L^* > \varepsilon) \leqslant e^{-cn}$ при любых n), если случайный вектор X имеет плотность распределения.

Заметим также, что экспоненциальное неравенство теоремы 5 не обобщается на случай произвольных распределений пары (X, Y) (при тех же условиях на k_n). Пусть, например, $X = 0$ с вероятностью 1, а Y принимает значения 1 и 2 с вероятностями 1/3 и 2/3 соответственно. Так как $(Y_{R_1(x)}, \ldots, Y_{R_{k_n}(x)}) = (Y_1, \ldots, Y_{k_n})$, то

$$L_n = 1/3 + (1/3)\, I_A, \quad L^* = 1/3,$$

где событие $A = \left[(1/k_n) \sum_{i=1}^{k_n} I_{[Y_i=1]} > 1/2\right]$. Следовательно, в силу экспоненциальной нижней границы Колмогорова (Stout (1974, с. 262) или лемма 6.6) имеем $P(L_n - L^* \geqslant \varepsilon) \geqslant \exp(-ck_n)$ при любых n и некотором $c > 0$.

Для доказательства теоремы 5 требуется одно геометрическое свойство, которое также применялось в работах Fritz (1975) и Stone (1977). Определим *конус* с углом θ и с центром в x как множество всех точек $y \in R^d$, таких, что угол $(y - x, z - x) < \theta$, где $z \in R^d$ — некоторая заданная точка. Это множество обозначается $\mathrm{Cone}(x, z, \theta)$. Выберем теперь столь малое θ, что при каждом $v \in \mathrm{Cone}(x, z, \theta)$ выполнено

$$\mathrm{Cone}(x, z, \theta) \cap S_{x, \|x-v\|} \subseteq S_{v, \|x-v\|}. \tag{13}$$

Зафиксировав θ, определим целое M_d — минимальное число конусов вида $\mathrm{Cone}(x, z_i, \theta)$, $1 \leqslant i \leqslant M_d$, необходимое для того, чтобы покрыть R^d.

Лемма 1. *Пусть* μ — *некоторая вероятностная мера на* R^d *и*

$$B_a(x) = \{z\colon \mu(S_{z, \|x-z\|}) \leqslant a\}, \quad x \in R^d.$$

Тогда

$$\mu(B_a(x)) \leqslant M_d a.$$

Доказательство. Пусть C_i, $1 \leqslant i \leqslant M_d$ — множество конусов вида Cone (x, z_i, θ), покрывающее R^d, где каждый конус C_i обладает свойством (13). Тогда

$$\mu(B_a(x)) \leqslant \sum_{i=1}^{M_d} \mu(C_i \cap B_a(x)). \tag{14}$$

При фиксированном i выберем произвольную точку $y \in C_i \cap \cap B_a(x)$. В силу (13)

$$\mu(S_{x, \|x-y\|} \cap C_i \cap B_a(x)) \leqslant \mu(S_{y, \|y-x\|}) \leqslant a, \tag{15}$$

где мы воспользовались тем, что $y \in B_a(x)$. Так как точка y произвольна, то

$$\mu(C_i \cap B_a(x)) \leqslant a. \tag{16}$$

Утверждение леммы следует из (14) и (16).

Доказательство теоремы 5. Обозначим

$$\tilde{p}_i(x) = \frac{1}{k} \sum_{j=1}^{k} I_{[Y_{R_j(x)} = i]}$$

и

$$p_i^*(x) = \frac{1}{k} \sum_{j=1}^{n} I_{[Y_j = i]} I_{[\|X_j - x\| \leqslant r_n(x)]},$$

где $k = k_n$ и $r_n(x)$ — корень уравнения

$$k/n = \mu(S_{x, r_n(x)}). \tag{17}$$

Заметим, что решение (5), где $\tilde{p}_i(x)$ имеет указанный выше вид, эквивалентно решению, определяемому соотношениями (6) и (12). Заметим еще, что корень $r_n(x)$ положителен, поскольку мера μ имеет плотность. Кроме того, из условия $k/n \to 0$ следует, что $r_n(x) \to 0$ при почти всех x по мере μ. Если C_d — мера Лебега единичного шара в R^d, то (17) эквивалентно соотношению

$$k = n r_n^d(x) C_d \frac{\mu(S_{x, r_n(x)})}{\lambda(S_{x, r_n(x)})}. \tag{18}$$

Таким образом, учитывая теорему 2.2 и то, что $k \to \infty$, получим: $n r_n^d(x) \to \infty$ при почти всех x по мере μ.

Очевидно,

$$|p_i(x) - \tilde{p}_i(x)| \leqslant |p_i(x) - E(p_i^*(x))| + \\ + |E(p_i^*(x)) - p_i^*(x)| + |p_i^*(x) - \tilde{p}_i(x)|. \tag{19}$$

Слегка обобщая теорему 2.2, получим

$$E(p_i^*(x)) = \frac{P(\|X_1 - x\| \leqslant r_n(x),\ Y_1 = i)}{k/n} =$$

$$= \int\limits_{S_{x,\,r_n(x)}} \frac{p_i(z)\,\mu(dz)}{\mu\left(S_{x,\,r_n(x)}\right)} \to p_i(x) \qquad (20)$$

при почти всех x по мере μ. Таким образом, теорема Лебега о мажорируемой сходимости дает: $\int |E(p_i^*(x)) - p_i(x)|\,\mu(dx) \to 0$ при любых i.

Пусть, далее, μ_n — эмпирическая мера для $X_1, \ldots, X_n$ и i — целое число из множества $\{1, \ldots, M\}$. Тогда

$$\int |p_i^*(x) - E(p_i^*(x))|\,\mu(dx) = \int \frac{\left|\nu_n\left(S_{x,\,r_n(x)}\right) - \nu\left(S_{x,\,r_n(x)}\right)\right|}{\mu\left(S_{x,\,r_n(x)}\right)}\,\mu(dx),$$

где

$$\nu_n(A) = \frac{1}{n}\sum_{j=1}^{n} I_{[X_j \in A,\ Y_j = i]},$$

$$\nu(A) = E(\nu_n(A)),$$

A — борелевское подмножество R^d. Пусть $h = (k/n)^{1/d}$, и пусть $\mathscr{P}$ — разбиение пространства R^d на кубы со сторонами h/N, где N — некоторое большое целое число, которое будет указано далее. Элементы $\mathscr{P}$ будут обозначаться через B, а центры этих множеств B — через b. Пусть T — фиксированный шар вида S_{0r}, r — некоторое большое число. Имеем

$$\int |p_i^*(x) - E(p_i^*(x))|\,\mu(dx) \leqslant$$

$$\leqslant \int \sum_{\substack{B:\ B \cap T \neq \varnothing,\\ B \subseteq S_{x,\,r_n(x)}}} \left(\frac{|\nu_n(B) - \nu(B)|}{\mu\left(S_{x,\,r_n(x)}\right)}\right)\mu(dx) +$$

$$+ \int \sum_{\substack{B:\ B \cap T = \varnothing,\\ B \subseteq S_{x,\,r_n(x)}}} \left(\frac{\nu_n(B) + \nu(B)}{\mu\left(S_{x,\,r_n(x)}\right)}\right)\mu(dx) +$$

$$+ \int \sum_{\substack{B:\ B \cap S_{x,\,r_n(x)} \neq \varnothing,\\ B \cap S^c_{x,\,r_n(x)} \neq \varnothing}} \frac{\left(\nu_n\left(B \cap S_{x,\,r_n(x)}\right) + \nu\left(B \cap S_{x,\,r_n(x)}\right)\right)}{\mu\left(S_{x,\,r_n(x)}\right)}\,\mu(dx). \qquad (21)$$

Применяя лемму 1 при $a = k/n$, получим, что первое слагаемое в правой части (21) не превосходит

$$\sum_{B:\ B \cap T \neq \varnothing} |\nu_n(B) - \nu(B)| \int \left(\frac{I_{S_{x,\, r_n(x)}}(b)}{\mu\left(S_{x,\, r_n(x)}\right)} \right) \mu(dx) \leqslant$$

$$\leqslant M_d \sum_{B:\ B \cap T \neq \varnothing} |\nu_n(B) - \nu(B)|. \quad (22)$$

Последняя величина стремится к 0 с экспоненциальной скоростью, но показатель экспоненты зависит от M_d (см. лемму 3.1). Второе слагаемое в правой части (21) не больше, чем

$$M_d(\mu_n(T^c) + \mu(T^c)). \quad (23)$$

При заданном $\varepsilon > 0$ выберем сначала r столь большим, что $\mu(T^c) < \varepsilon/M_d$. Тогда в силу неравенства Хёфдинга (Hoeffding (1963)) имеем $P((23) > 3\varepsilon) \leqslant P(|\mu_n(T^c)| > \varepsilon/M_d) \leqslant 2 \cdot \times \exp(-2n(\varepsilon/M_d)^2)$. Таким образом, остается рассмотреть лишь последнее слагаемое в (21).

При оценке сверху последнего слагаемого в (21) введем обозначения $A = S_{x,\, r_n(x)}$, $A^* = S_{x,\, (r_n(x) - h/N)_+}$ и $A_N = \{x: (C_d f(x))^{1/d} > \sqrt{N}\}$, где символом $(\ )_+$ обозначается положительная часть функции. Последнее слагаемое в (21) не превосходит

$$\int \left(\frac{\mu_n(A) - \mu_n(A^*)}{\mu(A)} \right) \mu(dx) + \int \left(\frac{\mu(A) - \mu(A^*)}{\mu(A)} \right) \mu(dx)$$

$$= \int \left(\int \left(\frac{I_A(z) - I_{A^*}(z)}{\mu(A)} \right) \mu(dx) \right) \mu_n(dz) + \int \left(\frac{\mu(A) - \mu(A^*)}{\mu(A)} \right) \mu(dx). \quad (24)$$

Первое слагаемое в правой части (42) можно записать в виде $\int \zeta(z)\, \mu_n(dz)$, где $|\zeta(z)| \leqslant M_d$. Последнее слагаемое тогда равно $\int \zeta(z)\, \mu(dz)$. Из неравенства Хёфдинга следует сходимость $\left| \int \zeta(z)\, \mu_n(dz) - \int \zeta(z)\, \mu(dz) \right| \to 0$ с экспоненциальной скоростью, где показатель экспоненты зависит от M_d. Таким образом, остается лишь показать, что величина $\int \zeta(z)\, \mu(dz)$ (последнее слагаемое в (24)) может быть сделана сколь угодно малой.

Выберем N так, что $\mu(A_N) < \varepsilon$, $d/\sqrt{N} < \varepsilon$, где $\varepsilon > 0$ — некоторое произвольное число. Ввиду (18) и определения h имеем

$$(h/r_n(x))^d \to C_d f(x) \quad (25)$$

при почти всех x. Следовательно, при почти всех $x \notin A_N$ имеем $(1 - h/Nr_n(x))_+ \to (1 - (C_d f(x))^{1/d}/N)_+ \geqslant (1 - 1/\sqrt{N})_+ > 0$ и $r_n(x) > h/N$, где n достаточно велико. При таких x

$$\frac{\mu(A) - \mu(A^*)}{\mu(A)} \to 1 - \left(1 - \frac{(C_d f(x))^{1/d}}{N}\right)^d \leqslant \frac{d}{\sqrt{N}} < \varepsilon. \qquad (26)$$

Отсюда и из теоремы Лебега о мажорируемой сходимости следует, что

$$\limsup_{n\to\infty} \int \left(\frac{\mu(A) - \mu(A^*)}{\mu(A)}\right) \mu(dx) \leqslant$$

$$\leqslant \mu(A_N) + \limsup_{n\to\infty} \int\limits_{A_N^c} \left(\frac{\mu(A) - \mu(A^*)}{\mu(A)}\right) \mu(dx) < 2\varepsilon. \qquad (27)$$

Таким образом, величина $\int |p_i^*(x) - E(p_i^*(x))| \mu(dx)$ стремится к 0 с экспоненциальной скоростью.

Рассмотрим, наконец, последнее слагаемое в (19). Вновь, фиксируя $i \in \{1, \ldots, M\}$, получим, что почти наверное

$$|p_i^*(x) - \tilde{p}_i(x)| =$$

$$= \frac{1}{k}\left|\sum_{j=1}^n I_{[Y_j = i,\ \|X_j - x\| \leqslant r_n(x)]} - \sum_{j=1}^n I_{[Y_j = i,\ \|X_j - x\| \leqslant \|X_{R_k(x)} - x\|]}\right| \leqslant$$

$$\leqslant \frac{1}{k}\sum_{j=1}^n \left| I_{[\|X_j - x\| \leqslant r_n(x)]} - I_{[\|X_j - x\| \leqslant \|X_{R_k(x)} - x\|]}\right| =$$

$$= \left|\frac{1}{k}\sum_{j=1}^n I_{[\|X_j - x\| \leqslant r_n(x)]} - 1\right|. \qquad (28)$$

Рассмотрим новую задачу классификации, в которой данные имеют вид (X_j, W_j), где $W_j = i$ при любых j, и обозначим соответствующие этой задаче функции $p_i^*(x)$ через $q_i^*(x)$. Тогда из (28) следует, что

$$\int |p_i^*(x) - \tilde{p}_i(x)| \mu(dx) \leqslant \int |q_i^*(x) - E(q_i^*(x))| \mu(dx), \qquad (29)$$

а последняя величина, как нам известно, стремится к 0 с экспоненциальной скоростью.

Чтобы убедиться, что показатель экспоненты скорости сходимости не зависит от M, поступим следующим образом. Просуммируем по всем i левую часть (21). Тогда сперва получим сумму по всем i выражений типа (22), к которой можно применить оценку

леммы 3.1. Второе и третье слагаемые в правой части (21) рассматриваются совершенно так же, как выше. Что касается оценки последнего слагаемого в правой части (19), то знак суммирования вносится в первые две строки в (28). Неравенство в (28) остается в силе без изменений. Наконец, остается просуммировать по всем i неравенство (29). Теорема доказана.

ЛИТЕРАТУРА

Beck J.
(1979) The exponential rate of convergence of error for k_n—NN nonparametric regression and decision, Problems of Control and Information Theory, 8, pp. 303—312.

Cover T. M., Hart P. E.
(1967) Nearest neighbor pattern classification, IEEE Transactions on Information Theory, IT-13, pp. 21—27.

Csibi S.
(1975) Stochastic Processes With Learning Properties, Springer-Verlag, Berlin.

Devroye L.
(1981) On the almost everywhere convergence of nonparametric regression function estimates, Annals of Statistics, 9, pp. 1310—1319.
(1982a) Any discrimination rule can have an arbitrarily bad probability of error for finite sample size, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-4, pp. 154—157.
(1982b) Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 61, pp. 467—481.

Devroye L., Györfi L.
(1983) Distribution-free exponential bound on the L_1 error of partitioning estimates of a regression function, in Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics, G. Pflug, W. Grossmann, and W. Wertz (Eds.), D. Reidel, Hingham, MA.

Devroye L., Wagner T. J.
(1976) Nonparametric discrimination and density estimation, Technical Report 183, Electronics Research Centre, University of Texas, Austin, Texas.
(1980) Distribution-free consistency results in nonparametric discrimination and regression function estimation, Annals of Statistics, 8, pp. 231—239.

Fix E., Hodges J. L.
(1951) Discriminatory analysis, nonparametric discrimination, consistency properties, Report No. 4, Project 21-49-004, School of Aviation Medicine, Randolph Field, Texas.
(1952) Nonparametric discrimination: small sample performance, Report No. 11, Project 21-49-004, School of Aviation Medicine, Randolph Field, Texas.

Fritz J.
(1975) Distribution-free exponential error bound for nearest neighbor pattern classification, IEEE Transactions on Information Theory, IT-21, pp. 552—557.

Gordon L., Olshen R. A.
(1978) Asymptotically efficient solutions to the classification problem, Annals of Statistics, 6, pp. 515—533.

Györfi L.
(1974) Estimation of probability density and optimal decision function in RKHS, in Progress in Statistics, J. Gani, K. Sarkadi, and I. Vincze (Eds.), North-Holland, Amsterdam, pp. 281—301.
(1978) On the rate of convergence of nearest neighbor rules, IEEE Transactions on Information Theory, IT-24, pp. 509—512.

Hoeffding W.
(1963) Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association, 58, pp. 13—30.
Loftsgaarden D. O., Quesenberry C. P.
(1965) A nonparametric estimate of a multivariate density function, Annals of Mathematical Statistics, 36, pp. 1049—1051.
Spiegelman C., Sacks J.
(1980) Consistent window estimation in nonparametric regression, Annals of Statistics, 8, pp. 240—246.
Stone C. J.
(1977) Consistent nonparametric regression, Annals of Statistics, 5, pp. 595—645.
Stout W. F.
(1974) Almost Sure Convergence, Academic Press, New York.
Van Ryzin J.
(1966) Bayes risk consistency of classification procedures using density estimation, Sankhya, 28, pp. 261—270.
Wolverton C. T., Wagner T. J.
(1969) Asymptotically optimal discriminant functions for pattern classifications, IEEE Transactions on Information Theory, IT-15, pp. 258—265.

Глава 11

Операции над оценками плотности

Основной темой данной главы является связь между L_1-ошибкой оценки плотности и L_1-ошибкой той же оценки после применения к ней таких операций, как взятие маргинальной плотности, формирование произведения плотностей, свертка плотностей, усечение плотностей и формирование неотрицательной проекции плотности. Все эти операции применяются довольно часто. Мы получим полезные неравенства во всех случаях, когда это возможно.

1. Маргинальные плотности

Пусть f^* и g^* — плотности на R^d (вообще говоря, они мыслятся здесь как неизвестная плотность f^* и некоторая построенная по выборке оценка g^*, но подразумеваемая случайность функции g^* несущественна). Пусть f и g — соответствующие маргинальные плотности на подпространстве R^s из R^d. Тогда справедлива

Теорема 1.

$$\int\limits_{R^s} |f - g| \leqslant \int\limits_{R^d} |f^* - g^*|.$$

Доказательство. Несколько пренебрегая точностью обозначений, имеем

$$\int\limits_{R^s} |f - g| = \int\limits_{R^s} \left| \int\limits_{R^{d-s}} f^* - \int\limits_{R^{d-s}} g^* \right| \leqslant \int\limits_{R^d} |f^* - g^*|.$$

Теорема 1 часто почти ничего не дает, ибо существуют примеры, когда при $d = 2, s = 1$ имеем $f = g$, но $\int |f^* - g^*| = 2$. Один из таких примеров прост: f^* — равномерная плотность на $[0, 1]^2 \cup [1, 2]^2$ и g^* — равномерная плотность на $[0, 1] \times [1, 2] \cup [1, 2] \times [0, 1]$.

Неравенство теоремы 1, казалось бы, свидетельствует о том, что если сначала оценивать f^* посредством g^*, а затем использовать маргинальную плотность g для оценивания f, то потери точности не происходит. Однако не надо забывать, что значение

$\int |f^* - g^*|$ обычно слишком велико, ибо если g^* — непараметрическая оценка для f^*, то скорость сходимости величины $\int |f^* - g^*|$ к 0, как правило, зависит от d.

2. Композиция (смесь) плотностей

Если две плотности f и g на R^d можно записать в виде конечных смесей $\sum p_i f_i$ и $\sum p_i g_i$, где f_i и g_i — плотности на R^d и $(p_1, \ldots, p_i, \ldots)$ — некоторый вероятностный вектор, то имеем следующий результат.

Теорема 2.

$$\int \left| \sum p_i f_i - \sum p_i g_i \right| \leqslant \sum p_i \int |f_i - g_i|.$$

Неравенство теоремы 2 тривиально, но оно имеет интересные следствия. Пусть, например, известно, что $f = pf_1 + (1 - p) f_2$, где f_1 — известная плотность, принадлежащая некоторому небольшому параметрическому семейству (такому, как семейство нормальных плотностей), f_2 — неизвестная плотность и параметр p для простоты считается известным. Если предъявляются выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$, извлеченные из f_1 и f_2 соответственно, то можно строить оценку для f в виде подходящей комбинации параметрической оценки плотности f_1 с непараметрической оценкой плотности f_2. Так как ошибка последней оценки обычно доминирует, то произведение этой ошибки для объема выборки m на $(1 - p)$ служит приближенной верхней границей общей ошибки. С другой стороны, если бы мы пренебрегли информацией относительно f_1 и смешали бы выборки, то могли бы построить непараметрическую оценку для f по выборке объема $n + m$. Если m/n — величина порядка $(1 - p)/p$, то ошибка последней оценки, как правило, больше, чем для рассмотренной сначала.

Если дана выборка из f и плотность f_1 частично или полностью известна, то из такой дополнительной информации, по-видимому, можно извлечь пользу, хотя сразу и не ясно, как это сделать.

Наконец, заметим, что неравенство теоремы 2 может быть в большой мере неадекватным. Если, например, f_1 и g_2 — равномерные плотности на $[0, 1]$, f_2 и g_1 — равномерные плотности на $[1, 2]$ и $p_1 = p_2 = 1/2$, то $\int |f - g| = 0$ и $\int |f_i - g_i| = 2$ при любых i.

3. Сужения плотностей

Рассмотрим некоторую плотность f и некую ее оценку g, которая также является плотностью. Иногда известно, что носитель f — множество A. Чтобы исключить бессмысленные случаи

(например, такие, когда часть массы оценки плотности лежит на отрицательной полуоси, а оценивается плотность положительной случайной величины), можно заменить плотность g ее *сужением* на A, т. е. функцией

$$g^*(x) = \frac{g(x)\, I_A}{\int\limits_A g}.$$

Такое сужение встречалось уже при рассмотрении преобразованной ядерной оценки (тл. 9). Сужение всегда лучше, чем исходная оценка g, о чем свидетельствует

Теорема 3.

$$\int |f - g^*| \leqslant \int |f - g|.$$

Доказательство.

$$\int |f - g^*| = 2 \int\limits_A \left(f - g \Big/ \int\limits_A g\right)_+ \leqslant$$

$$\leqslant 2 \int\limits_A (f - g)_+ \leqslant 2 \int (f - g)_+ = \int |f - g|.$$

4. Неотрицательные проекции

Предположим снова, что f и g — плотность и ее оценка на R^d. Хотя при этом $\int g = 1$, но g может принимать отрицательные значения, как, например, в случае оценки Бартлета (§ 7.5). Функция

$$g^* = \frac{g I_A}{\int\limits_A g}, \quad A = \{x:\ g(x) > 0\}$$

является уже настоящей плотностью. Она называется *неотрицательной проекцией* g. Она снова оказывается лучше, чем исходная оценка g.

Теорема 4.

$$\int |f - g^*| \leqslant \int |f - g|.$$

Доказательство. Так как $\int g = 1$, то $\int\limits_A g \geqslant 1$. Следовательно, $g \geqslant g^*$ на A. Таким образом,

$$\int |f - g^*| = 2 \int (g^* - f)_+ = 2 \int\limits_A (g^* - f)_+ + 2 \int\limits_{A^c} (g^* - f)_+ =$$

$$= 2 \int\limits_A (g^* - f)_+ \leqslant 2 \int\limits_A (g - f)_+ = 2 \int\limits_{g>f} (g - f)_+ = \int |g - f|.$$

5. Произведения плотностей

Здесь рассматривается вопрос о том, насколько хороша оценка $\prod_{i=1}^{d} g_i(x_i)$ для $\prod_{i=1}^{n} f_i(x_i)$, $x_i \in R$, если f_i, $1 \leqslant i \leqslant d$, — одномерные плотности и g_i — одномерные плотности, являющиеся оценками для f_i. Предлагаются следующие неравенства.

Теорема 5. *Пусть H_{pi} — расстояние Хеллингера между f_i и g_i, т. е.*

$$H_{pi} = \left(\int |f_i^{1/p} - g_i^{1/p}|^p\right)^{1/p}, \quad 1 \leqslant i \leqslant d,$$

и пусть

$$L_i = \int f_i \log\left(\frac{f_i}{g_i}\right), \quad 1 \leqslant i \leqslant d.$$

Тогда, вспоминая, что

$$\int \min\left(\prod f_i, \prod g_i\right) = 1 - \frac{1}{2}\int \left|\prod f_i - \prod g_i\right|,$$

будем иметь

$$\int \min\left(\prod f_i, \prod g_i\right) \leqslant \exp\left(-\frac{1}{2}\sum H_{2i}^2\right) \leqslant \exp\left(-\frac{1}{8}\sum H_{1i}^2\right),$$

$$\int \min\left(\prod f_i, \prod g_i\right) \geqslant \exp\left(-\sum \frac{H_{1i}}{2 - H_{1i}}\right),$$

$$\int \min\left(\prod f_i, \prod g_i\right) \geqslant \frac{1}{2}\exp\left(-\sum L_i\right).$$

Доказательство. Первое неравенство следует из того, что

$$\int \min\left(\prod f_i, \prod g_i\right) \leqslant \prod_i \int \sqrt{f_i g_i} = \prod_i \left(1 - \frac{1}{2}H_{2i}^2\right) \leqslant$$

$$\leqslant \exp\left(-\frac{1}{2}\sum H_{2i}^2\right),$$

и неравенства $H_{2i} \geqslant H_{1i}/2$ (теорема 8.4). Второе неравенство следует из того, что

$$\int \min\left(\prod f_i, \prod g_i\right) \geqslant \int \prod \min(f_i, g_i) =$$

$$= \prod \int \min(f_i, g_i) = \prod \left(1 - \frac{1}{2}H_{1i}\right) \geqslant \exp\left(-\sum \frac{H_{1i}}{2 - H_{1i}}\right).$$

Наконец, последнее неравенство следует непосредственно из теоремы 8.2.

Заметим, что если $f_i = f$, $g_i = g$ и $\int |f - g| \neq 0$, то $\int \left| \prod f_i - \prod g_i \right| \to 2$ при $d \to \infty$. По этой причине неравенства теоремы 5 сформулированы для величины $\int \min\left(\prod f_i, \prod g_i\right)$.

6. Радиально-симметричные плотности

Назовем f^* *радиально-симметричной плотностью* на R^d, если это плотность распределения случайного вектора YZ, где Y — случайная величина с некоторой плотностью f на $[0, \infty)$ и Z — случайный вектор, не зависящий от Y и равномерно распределенный на поверхности единичной сферы в R^d (так что $\| Z \| = 1$ с вероятностью 1). Если g^* — другая радиально-симметричная плотность, которой отвечает плотность g на $[0, \infty)$, то имеет место

Теорема 6.

$$\int |f^* - g^*| = \int |f - g|.$$

Доказательство. Пусть B — борелевское подмножество R^d. Тогда, обозначая через μ равномерную меру на единичной сфере в R^d, имеем $P(YZ \in B) = \int f(x)\, \mu(B/x)\, dx$. Таким образом,

$$\int |f^* - g^*| = 2 \sup_B \left| \int_B f^* - \int_B g^* \right| =$$

$$= 2 \sup_B \left| \int f(x) \mu(B/x) dx - \int g(x)\, \mu(B/x) dx \right| =$$

$$= 2 \sup_B \left| \int\limits_{f > g} (f - g) \mu(B/x) dx - \int\limits_{f < g} (g - f) \mu(B/x) dx \right| \leqslant$$

$$\leqslant 2 \max \left(\int\limits_{f > g} (f - g), \int\limits_{f < g} (g - f) \right) = \int |f - g|.$$

Кроме того, полагая $B = \{x\colon x \in R^d, f(\| x \|) \geqslant g(\| x \|)\}$, замечаем, что верно также обратное неравенство.

Теорема 6 важна, поскольку она показывает, что все наши одномерные результаты по оцениванию плотности переносятся на задачу оценивания радиально-симметричных плотностей на R^d с известным центром симметрии. Так, если задано условие радиальной симметрии, то всегда надо пытаться оценивать не f^*, а плотность f случайной величины Y. Оценка для f^* всегда восстанавливается по одномерной оценке g плотности f на основе

радиальной симметрии. Не надо объяснять, что это всего лишь один из многих случаев, когда априорную информацию можно использовать для понижения размерности (а значит, и сложности) задачи.

7. Свертки

Рассмотрим теперь задачу оценки плотности распределения величины $Y_1 + \ldots + Y_d$, где Y_i — независимые случайные величины с неизвестной общей плотностью f, по данной последовательности $X_1, \ldots, X_n$ независимых случайных величин с плотностью f. В большинстве интересных случаев значение d ограниченно либо по крайней мере очень мало по сравнению с n — если это не так, то лучше воспользоваться локальной центральной предельной теоремой. Возможно, существуют методы решения задачи, более хорошие, чем оценивание плотности f^{*d} (d-кратной свертки f) посредством g^{*d}, где g — обычная оценка плотности f. Тем не менее имеет место

Теорема 7.

$$\int |f^{*d} - g^{*d}| \leqslant d \int |f - g|.$$

Доказательство. Данное неравенство следует из того, что

$$\int |f * \tilde{f} - g * \tilde{g}| \leqslant \int |f * (\tilde{f} - \tilde{g})| + \int |(f - g) * \tilde{g}| \leqslant$$

$$\leqslant \int |f - g| + \int |\tilde{f} - \tilde{g}|$$

для любых четырех плотностей f, g, $\tilde{f}$ и $\tilde{g}$.

Таким образом, не гарантируется, что L_1-ошибка сохраняет одно и то же значение. С другой стороны, неравенство, использованное в доказательстве теоремы 7, очень грубое. Чтобы понять это, достаточно рассмотреть плотности четырех гамма-распределений с различными параметрами a, b, c, d, удовлетворяющими условию $a + c = b + d$.

Если d велико по сравнению с n, то возрастает роль локальных предельных теорем. Пусть, например, f — плотность со средним μ, дисперсией σ^2 и третьим центральным моментом α, а g — плотность нормального распределения с параметрами $(d\mu, d\sigma^2)$. Тогда

$$\int |f^{*d} - g| = \frac{|\alpha|}{3\sqrt{2\pi d}}(1 + 4e^{-3/2}) + o\left(\frac{1}{\sqrt{d}}\right)$$

(см., например, Петров (1972, с. 263) или Сираждинов и Маматов (1962)).

8. Унимодальные плотности

Рассмотрим унимодальную плотность f на R с модой в нуле. *Теорема Хинчина* (Feller (1971, с. 187 по русскому изданию)) утверждает, что существует функция распределения F, такая, что

$$f(x) = \int_x^\infty \frac{1}{y}\, dF(y), \quad f(-x) = -\int_{-\infty}^{-x} \frac{1}{y}\, dF(y), \quad x > 0.$$

Таким образом, если f и g — унимодальные плотности на $[0, \infty)$ с модой в нуле и с соответствующими функциями распределения Хинчина F и G, то

$$\int |f - g| = \int \left| \int_x^\infty \frac{dF(y) - dG(y)}{y} \right| dx \leqslant$$

$$\leqslant \int \int_x^\infty \frac{|dF(y) - dG(y)|}{y}\, dx \leqslant \int |dF(y) - dG(y)|.$$

Если F и G имеют плотности f^* и g^*, то заключаем следующее:

Теорема 8.

$$\int |f - g| \leqslant \int |f^* - g^*|.$$

Плотности f и f^* связаны соотношением $f(x) = \int_x^\infty (f^*(y)/y)\, dy$, и потому ясно, что плотность g ближе к f, чем g^* к f^*, какова бы ни была пара унимодальных плотностей f, g. К сожалению, обычно мы не имеем в своем распоряжении выборки из f^*.

9. Применения в задаче обнаружения

Рассмотрим одну из простейших постановок задачи обнаружения. Пусть f и g — известные плотности на R^d и задана выборка $X_1, \ldots, X_n$, состоящая из независимых одинаково распределенных случайных векторов, плотность которых равна либо f, либо g. Требуется принять решение в пользу одной из этих двух альтернатив. Подходя более формально, введем число Z, принимающее значения 1 или 2 в зависимости от того, какова плотность наблюдений: f или g. Тогда решающее правило Y — борелевская функция от $X_1, \ldots, X_n$, принимающая значения 1 или 2. Представляет интерес индикаторная функция ошибки

$$L_n = I_{[Y \neq Z]}.$$

В классической теории обнаружения задача фактически асимметрична: там интересуются значениями $I_{[Y \neq 1]}$ и $I_{[Y \neq 2]}$ при $Z = 2$ $Z = 1$ соответственно, но ошибка одного рода хуже, чем другого.

Здесь такой случай не рассматривается. Не рассматриваются также задачи различения f и «не f». Эти задачи лучше всего изложены в книгах по критериям согласия. Дополнительные сведения о задаче обнаружения можно почерпнуть в книге Rao (1973, с. 399 по русскому изданию). Заметим, что теорема 9 ниже тесно связана с результатами разд. 7а, 3 книги Rao (1973). В дальнейшем считается, что Z — случайная величина, для которой $P(Z = 1) = p \in [0, 1]$. Все границы и утверждения теорем 9—11 справедливы при *любых* p, в том числе при $p = 0$, $p = 1$. Заметим, что $E(L_n) = pP_1(Y \neq 1) + (1-p)P_2(Y \neq 2)$, где вероятности P_1, P_2 вычисляются по условным распределениям $X_1, \ldots, X_n$ при фиксированном $Z = 1$ и $Z = 2$ соответственно. Если $p = 1/2$, то величина $E(L_n)$ минимальна при

$$Y = \begin{cases} 1, & \text{если } \prod_{i=1}^{n} f(X_i)/g(X_i) > 1, \\ 2 & \text{в противном случае.} \end{cases}$$

Это правило называется *оптимальным детектором* или *детектором максимума правдоподобия*. Для удобства перепишем его в виде

$$Y = \begin{cases} 1, & \text{если } (1/n)\sum_{i=1}^{n} \log(f(X_i)/g(X_i)) > c, \\ 2 & \text{в противном случае,} \end{cases}$$

считая пока, что c — произвольное действительное число. Отметим, что оптимальный детектор зависит только от отношения f/g и, следовательно, обладает желательным для любых детекторов свойством инвариантности к монотонным преобразованиям координатных осей. Не исключено, что сумма в определении оптимального детектора принимает значения $+\infty$ или $-\infty$, но она всегда имеет смысл, поскольку $P(f(X_i) = g(X_i)$ при некотором $i) = 0$.

Теорема 9. *Если* $c \in \left(-\int g \log(g/f), \int f \log(f/g)\right)$, *то* $L_n \to 0$ *почти наверное при любых* $f \neq g$ (*т. е. при* f, g, *таких, что* $\int |f - g| > 0$). *Указанный интервал для* c *может быть бесконечным в одну или в обе стороны. Он всегда содержит значение* 0. *При* $c = 0$ *имеем*

$$P(Y \neq Z) = E(L_n) \begin{cases} \geqslant \frac{q}{2} \exp\left(-n \min\left(\int g \log\left(\frac{g}{f}\right), \int f \log\left(\frac{f}{g}\right)\right)\right) \\ \geqslant q \exp\left(-n \int |f-g| \Big/ \left(2 - \int |f-g|\right)\right) \\ \leqslant \exp\left(-(n/8)\left(\int |f-g|\right)^2\right), \end{cases}$$

где $q = \min(p, 1-p)$.

Доказательство. Эти три неравенства получаются, если заметить, что

$$E(L_n) = p \int\limits_{\pi g_i < \pi f_i} \pi g_i + (1-p) \int\limits_{\pi f_i \leqslant \pi g_i} \pi f_i,$$

где $f_i = f(x_i)$, $g_i = g(x_i)$ и интеграл берется по $dx_1\, dx_2 \ldots dx_n$. Таким образом, $E(L_n) \Big/ \int \min(\pi f_i, \pi g_i) \in (q, 1]$. Затем применяется теорема 5.

Далее, вспомним, что при $f \neq g$ значения $\int f \log(f/g)$ принадлежат интервалу $(0, \infty]$ (теорема 8.2). Разобьем этот интеграл на положительную и отрицательную части, представляя функцию log в виде $\log_+ + \log_-$, и заметим, что отрицательная часть ограниченна, ибо

$$0 \geqslant \int f \log_-(f/g) \geqslant -1/e.$$

Чтобы убедиться в этом, воспользуемся неравенством $\log_+(g/f) \leqslant g/ef$, из которого следует, что

$$0 \leqslant \int f \log_+(g/f) \leqslant \int g/e = 1/e.$$

Нам также понадобится усиленный закон больших чисел в следующей форме: если $Z_1, \ldots, Z_n, \ldots,$ — независимые одинаково распределенные случайные величины, такие, что $E(Z_{1-}) > -\infty$, то $(1/n) \sum_{i=1}^{n} Z_i \to E(Z_1)$ почти наверное при $n \to \infty$, даже если $E(Z_1) = \infty$.

Отсюда следует, что при $Z = 1$ выполнено соотношение

$$\frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{f(X_i)}{g(X_i)}\right) \to \int f \log\left(\frac{f}{g}\right)$$

почти наверное при $n \to \infty$. Таким образом, если $c < \int f \log(f/g)$, то $P(Z = 1, Y = 2 \mid X_1, \ldots, X_n) \to 0$ почти наверное при $n \to \infty$. Оставшаяся часть доказательства следует из соображений симметрии.

Известно, что $\log P(Y \neq Z) = n(\log H + o(1))$, где $H = \inf\limits_{0<a<1} \int f^a g^{1-a}$ — энтропия Реньи. Таким образом, из теоремы 9 следуют неравенства

$$\frac{1}{8}\left(\int |f-g|\right)^2 \lesssim -\log H \lesssim \frac{\int |f-g|}{2 - \int |f-g|}.$$

Можно применять и другие детекторы. Так как они не оптимальны, необходимо указать случаи, когда это следует делать. Например, если хотя бы одна из плотностей f и g точно не известна и вместо неизвестных плотностей используются лишь хорошие выборочные оценки, то оптимальный детектор может привести к катастрофическим результатам из-за своей чувствительности к событиям $f(X_i) = 0$ и $g(X_i) = 0$. Другими словами, в таких случаях желательно иметь более робастные детекторы. Здесь мы введем и обсудим два типа таких детекторов — ДРО (детектор, основанный на критериях распознавания образов) и ДL1 (детектор, основанный на критерии L_1).

ДРО определяется следующим образом:

$$Y = \begin{cases} 1, & \text{если } (1/n)\sum_{i=1}^{n} I_{[f(X_i)/g(X_i) > 1]} - I_{[f(X_i)/g(X_i) < 1]} > c, \\ 2 & \text{в противном случае.} \end{cases}$$

По существу в ДРО производится суммирование решающих правил, оптимальных для отдельных наблюдений. Заметим, что существуют случаи, когда при выборе $c = 0$ получается детектор, вероятность ошибки которого $L_n \to 1$ почти наверное для $Z = 1$. Например, предположим, что g — равномерная плотность на $[0, 1]$, $f = 2$ на $[0, a]$ и $f = (1 - 2a)/(1 - a)$ на $(a, 1]$, где $0 < a < 1/4$. Тогда $g(X_i) > f(X_i)$ с вероятностью $1 - 2a > 1/2$.

Теорема 10. *Если* $c \in \left(\int_{f>g} g - \int_{f<g} g, \int_{f>g} f - \int_{f<g} f \right)$ *(этот интервал не содержит значение* 0*), то* $L_n \to 0$ *почти наверное при* $n \to \infty$ *для любых* $f \neq g$. *Если выбрать*

$$c = \frac{1}{2}\left(\int_{f>g} (f+g) - \int_{f<g} (f+g) \right),$$

то

$$E(L_n) \leqslant \exp\left(-\frac{n}{8} \left(\int |f - g| \right)^2 \right).$$

Доказательство. Не нарушая общности, положим $p = 1$. Слагаемые в сумме из определения ДРО независимы, одинаково распределены, принимают значения из $[-1, 1]$ и имеют средние $\int_{f>g} f - \int_{f<g} f$. Первое утверждение теоремы следует из усиленного закона больших чисел и симметрии задачи.

Далее, если выбрать c, как указано в теореме, то

$$E(L_n) = P\left(\frac{1}{n}\sum_{i=1}^{n} (W_i - E(W_i)) < c - E(W_1) \right),$$

где $W_i = I_{[f(X_i)/g(X_i) > 1]} - I_{[f(X_i)/g(X_i) < 1]}$ и $E(W_1) = \int\limits_{f>g} f - \int\limits_{f<g} f$.

Имеем

$$c - E(W_1) = \frac{1}{2}\left(\int\limits_{f>g}(g-f) - \int\limits_{f<g}(g-f)\right) = -\frac{1}{2}\int|f-g|.$$

Так как случайные величины $\{W_i - E(W_i)\}$ независимы и одинаково распределены с нулевыми средними и принимают значения из $[-1 - E(W_1),\ 1 - E(W_1)]$, то в силу неравенства Хёфдинга (Hoeffding (1963)) получим

$$E(L_n) \leqslant \exp\left(-2n\left(\int|f-g|^2/2\right)\Big/4\right) = \exp\left(-\frac{n}{8}\left(\int|f-g|\right)^2\right).$$

ДЛ1 определяется соотношением

$$Y = \begin{cases} 1, & \text{если } \frac{1}{n}\sum\limits_{i=1}^{n}\left(\left(1 - \frac{g(X_i)}{f(X_i)}\right)_+ - \left(1 - \frac{f(X_i)}{g(X_i)}\right)_+\right) > c, \\ 2 & \text{в противном случае,} \end{cases}$$

где $c \in R$ — некоторая постоянная. По своим свойствам этот детектор очень похож на ДРО, что видно из следующей теоремы.

Теорема 11. *Если* $c \in \left(-\int f(1-g/f)_+^2,\ \int g(1-f/g)_+^2\right)$, *то* $L_n \to 0$ *почти наверное при* $n \to \infty$ *для любых* $f \neq g$. *Если выбрать*

$$c = \frac{1}{2}\left(\int g\left(1 - \frac{f}{g}\right)_+^2 - \int f\left(1 - \frac{g}{f}\right)_+^2\right),$$

то

$$E(L_n) \leqslant \exp\left(-\frac{n}{24}\left(\int|f-g|\right)^2\right).$$

Доказательство. Слагаемые в определении Y являются случайными величинами со средними

$$\int f\left(1 - \frac{g}{f}\right)_+ - \int f\left(1 - \frac{f}{g}\right)_+ = \int g\left(1 - \frac{f}{g}\right)_+^2$$

при $Z = 1$ и

$$\int g\left(1 - \frac{g}{f}\right)_+ - \int g\left(1 - \frac{f}{g}\right)_+ = -\int f\left(1 - \frac{g}{f}\right)_+^2$$

при $Z = 2$, так что первое утверждение теоремы легко следует из усиленного закона больших чисел для независимых ограниченных случайных величин.

Для доказательства неравенства предположим, не нарушая общности, что $Z = 1$. Заметим, что

$$E(L_n) = P\left(\frac{1}{n}\sum_{i=1}^{n}(W_i - E(W_i)) < -\varepsilon/2\right),$$

где W_i — i-е слагаемое в определении Y и

$$\varepsilon = \int g\left(1 - \frac{f}{g}\right)_+^2 + \int f\left(1 - \frac{g}{f}\right)_+^2.$$

Далее, в силу одной из разновидностей неравенства Беннета (Bennett (1962)) получим

$$E(L_n) \leqslant \exp\left(-\frac{n\varepsilon^2}{8(\sigma^2 - \varepsilon/2)}\right),$$

где $\sigma^2 = \mathrm{Var}(W_1)$. Это неравенство верно только для независимых слагаемых W_i при $|W_i| \leqslant 1$ (в данном случае они таковы). Но

$$\sigma^2 \leqslant E(W_1^2) = \int f\left(1 - \frac{g}{f}\right)_+^2 + \int f\left(1 - \frac{f}{g}\right)_+^2 \leqslant \varepsilon,$$

так что $E(L_n) \leqslant \exp(-n\varepsilon/12)$. По неравенству Коши

$$\varepsilon \geqslant \left(\int g\left(1 - \frac{f}{g}\right)_+\right)^2 + \left(\int f\left(1 - \frac{g}{f}\right)_+\right)^2 = \frac{1}{2}\left(\int |f - g|\right)^2,$$

и мы получаем требуемый результат.

Существует бесчисленное множество других примеров детекторов. Так, $L_n \to 0$ почти наверное для детектора

$$Y = \begin{cases} 1, & \text{если } \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(X_i)}{g(X_i)} - \frac{g(X_i)}{f(X_i)}\right) > c, \\ 2 & \text{в противном случае} \end{cases}$$

при c, принадлежащих интервалу $\left(-\int g^2/f + 1,\ \int f^2/g - 1\right)$, который при $f \neq g$ содержит интервал $\left(-\left(\int |f - g|\right)^2, \left(\int |f - g|\right)^2\right)$ (теорема 8.3). Неограниченные слагаемые очень чувствительны к различиям между f и g, но эта чувствительность несколько усложняет ситуацию в случае, когда хотя бы одна из функций f и g точно не известна.

На практике f и g известны редко. Обычно имеются случайные векторы $Y_1, \ldots, Y_k$ и $Z_1, \ldots, Z_k$ с плотностями f и g соответственно. По этим данным сначала строятся оценки f_k и g_k для f и g, представляющие собой плотности на R^d. Затем предлагается принять решение о том, какую общую плотность — f или g — имеют

случайные векторы $X_1, \ldots, X_n$. Индикаторная функция ошибки здесь равна

$$L_n = I_{[Y \neq Z]}.$$

В большинстве прикладных задач теории связи и теории информации можно считать, что число k фиксировано, а n, возможно, и нет, что соответствует случаю, когда $X_1, \ldots, X_n$ — отсчеты переменного во времени сигнала. Таким образом, величина

$$L_k^* = \limsup_{n \to \infty} (L_n) \quad \text{(п. н.)}$$

является хорошим показателем качества оценок f_k и g_k. Очевидно, что L_k^*, так же как L_n, может принимать лишь значения 0 и 1. Ожидать, что $P(L_k^* = 1) = 0$, нереалистично. Вместе с тем назовем детектор *состоятельным*, если

$$\lim_{k \to \infty} P(L_k^* = 1) = 0.$$

Попытаемся теперь доказать состоятельность для некоторых широких классов детекторов. Все приводимые ниже рассуждения легко распространить на случай сильной состоятельности, т. е. сходимости $L_k^* \to 0$ почти наверное при $k \to \infty$. Рассматриваемые далее детекторы основаны на существовании функции H: $[0, \infty] \to [-1, 1]$, такой, что

$$\int f H\left(\frac{f}{g}\right) > \int g H\left(\frac{f}{g}\right)$$

при любых $f \neq g$ (т. е. при $\int |f - g| > 0$). Примерами таких функций H являются $H(u) = I_{[u>1]} - I_{[u<1]}$ (случай ДРО) и $H(u) = (1 - 1/u)_+ - (1 - u)_+$ (случай $\Delta L1$).

Детектор строится следующим образом:

1. Определяются оценки плотности f_k и g_k по выборкам $Y_1, \ldots, Y_k$ и $Z_1, \ldots, Z_k$ соответственно. Вычисляется порог детектора $c_k = \int (f_k + g_k) H(f_k/g_k)/2$.

2. Решающее правило Y определяется соотношением

$$Y = \begin{cases} 1, & \text{если } \frac{1}{n} \sum_{i=1}^{n} H(f_k(X_i)/g_k(X_i)) > c_k, \\ 2 & \text{в противном случае.} \end{cases}$$

Теорема 12. *Основанный на выборочных данных детектор, который определен выше, состоятелен, если*

(i) $f_k \to f$ *и* $g_k \to g$ *по вероятности при* $k \to \infty$ *для почти всех* x,

(ii) *функция* H *непрерывна* (*это условие выполнено для* $\Delta L1$).

Доказательство. Не нарушая общности, предположим, что $Z = 1$. Так как H — ограниченная функция, то сразу ясно, что

$$L_k^* = I_{[\int fH(f_k/g_k) \leqslant c_k]}$$

почти наверное. Из определения H следует, что

$$\int fH\left(\frac{f}{g}\right) > c = \int \frac{f+g}{2} H\left(\frac{f}{g}\right).$$

Таким образом, остается доказать, что $c_k \to c$ по вероятности и $\int fH(f_k/g_k) \to \int fH(f/g)$ по вероятности при $k \to \infty$. Последнее следует непосредственно из теоремы Лебега о мажорируемой сходимости и условий (i) и (ii). Кроме того,

$$|c_k - c| \leqslant \left(\int |f_k - f| + \int |g_k - g|\right)\Big/2 + \\ + \int \frac{f+g}{2}\left|H\left(\frac{f_k}{g_k}\right) - H\left(\frac{f}{g}\right)\right|,$$

где правая часть стремится к 0 по вероятности в силу (i), теоремы 2.8 Глика и (ii).

Детектор максимума правдоподобия не имеет непосредственного обобщения на случай, когда f и g неизвестны. Одно из препятствий здесь состоит в том, что при некоторых i величина $\log(f_k(X_i)/g_k(X_i))$ равна $-\infty$, а при некоторых других i она равна $+\infty$. Эту неустойчивость, конечно, можно устранить, усекая логарифмическую функцию сверху и снизу, т. е. применяя процедуру так называемой *винзоризации* (см. обсуждение винзоризованных детекторов максимума правдоподобия в книге Huber (1981) [1]). Ясно, что функция H для винзоризованного детектора максимума правдоподобия удовлетворяет условиям теоремы 12, так что состоятельность этого детектора получается без труда.

10. Симметризация и перестановочная инвариантность

В этом параграфе f — произвольная плотность на R^d и, если не оговорено противное, f_n — произвольная оценка плотности, основанная на выборке объема n из совокупности с плотностью f. Хорошо известно, что оценки можно улучшить, увеличивая число наблюдений, и что оценки, не являющиеся симметричными функциями от данных, можно улучшить за счет симметризации. Однако

[1]) О робастных детекторах см. обзор Kassam S. A., Poor H. V. Robust techniques for signal processing: a survey. Proc. IEEE, 1985, v. 73, No. 3, p. 433—481. — *Прим. перев.*

нужно еще и количественно указать, каковы эти улучшения при фиксированном n. Этому посвящен данный параграф. Одновременно мы воспользуемся случаем, чтобы продемонстрировать красоту и изящество теории выпуклости по Шуру. С этой целью построим доказательства всех результатов на основе неравенства Маршалла — Прошана (Marshall, Proshan (1965)).

Лемма 1 (Marshall, Proschan (1965)). *Пусть* φ *— выпуклая функция* n *переменных, симметричная по своим аргументам. Пусть весовые векторы* $a = (a_1, \ldots, a_n)$ *и* $b = (b_1, \ldots, b_n)$ *таковы, что* a *мажорирует* b $(a \succ b)$, *т. е.*

$$\sum_{i=1}^{k} a_{[i]} \geqslant \sum_{i=1}^{k} b_{[i]}, \quad k = 1, \ldots, n,$$

с равенством при $k = n$, *где* $a_{[1]} \geqslant \ldots \geqslant a_{[n]}$ *и* $b_{[1]} \geqslant \ldots \geqslant b_{[n]}$ *— упорядоченные значения компонент векторов* a *и* b. *Если* $X_1, \ldots, X_n$ *— случайные векторы с перестановочно инвариантными распределениями, то*

$$E(\varphi(a_1 X_1, \ldots, a_n X_n)) \geqslant E(\varphi(b_1 X_1, \ldots, b_n X_n)).$$

Теорема 13. *Пусть* f_n *— оценка вида*

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_n(x, X_i),$$

где K_n *— произвольные измеримые функции. Если* $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ *— независимые одинаково распределенные случайные величины, то*

$$E\left(\int |g_n - f|\right) \leqslant E\left(\int |f_n - f|\right),$$

где

$$g_n(x) = \frac{1}{n+m} \sum_{i=1}^{n+m} K_n(x, X_i).$$

Замечание. Для ядерной оценки (Parzen (1962), Rosenblatt (1956)) с фиксированными ядром K и параметром сглаживания h L_1-ошибка является невозрастающей функцией от n.

Отложим доказательство теоремы 13 до того момента, когда будут сформулированы все результаты.

Теорема 14. *Пусть* f_n *— оценка вида*

$$f_n(x) = \sum_{i=1}^{n} w_{ni} K_n(x, X_i),$$

где w_{ni} *— веса, сумма которых равна 1, и функции* K_n *такие же, как в теореме* 13. *Тогда*

$$E\left(\int |g_n - f|\right) \leqslant E\left(\int |f_n - f|\right),$$

где

$$g_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_n(x, X_i).$$

Теорема 15. *Пусть* f_n *— оценка вида*

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_i(x - X_i),$$

где K_i *— измеримые функции. Тогда*

$$E\left(\int |g_n - f|\right) \leqslant E\left(\int |f_n - f|\right),$$

где

$$g_n(x) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_j(x - X_i).$$

Замечание. Оценка g_n из теоремы 15 имеет вид $\frac{1}{n}\sum_{i=1}^{n} L_n(x - X_i)$, так что теорема 15 указывает на улучшение, которое можно получить при замене рекуррентных оценок типа дельта-функции Дирака (см., например, Walter, Blum (1979)) их нерекуррентными аналогами.

Все сформулированные выше теоремы подчеркивают тот факт, что наилучшие оценки всегда являются перестановочно инвариантными функциями от данных (см., например, Wertz (1976)). Имеет место

Теорема 16. *Если* f_n *— произвольная оценка плотности, то оценка*

$$g_n(x, X_1, \ldots, X_n) = \frac{1}{n!}\sum_{\substack{\text{все перестановки}\\ \sigma(1), \ldots, \sigma(n)\\ \text{набора } 1, \ldots, n}} f_n(x, X_{\sigma(1)}, \ldots, X_{\sigma(n)})$$

не хуже, чем f_n, *в том смысле, что*

$$E\left(\int |g_n - f|\right) \leqslant E\left(\int |f_n - f|\right).$$

Доказательство теорем 13—16. Из доказательства будет ясно, что улучшение, получаемое за счет симметризации, не только глобальное. В действительности оно имеет место при *каждом* x.

Пусть $Y_\sigma = f_n(x, X_{\sigma(1)}, \ldots, X_{\sigma(n)}) - f(x)$ и S — множество всех $n!$ перестановок σ набора $1, \ldots, n$. Ясно, что $\{Y_\sigma, \sigma \in S\}$ — множество случайных величин с перестановочно инвариантными распределениями. Положим

$$\varphi(u_1, \ldots, u_{n!}) = \left| \sum_{i=1}^{n!} u_i \right|$$

и заметим, что эта функция выпукла и симметрична. Следовательно, в силу леммы 1 и того, что $a = (1, 0, 0, \ldots, 0) \succ b = (1/n!, \ldots, 1/n!)$, имеем

$$E\left(\left|\sum_{\sigma=1}^{n!} a_\sigma Y_\sigma\right|\right) = E(|Y_1|) = E(|f_n(x, X_1, \ldots, X_n) - f(x)|) \geqslant$$

$$\geqslant E\left(\left|\sum_{\sigma=1}^{n!} \frac{1}{n!} Y_\sigma\right|\right) = E(|g_n(x, X_1, \ldots, X_n) - f(x)|).$$

Интегрируя это неравенство по dx, получим результат теоремы 16. (Заметим попутно, что $E(|g_n - f|^p) \leqslant E(|f_n - f|^p)$ при любых x и любых $p \geqslant 1$.)

Теорема 15 легко следует из теоремы 16. Для доказательства теоремы 13 построим оценку g_n так же, как в теореме 16, и заметим, что

$$g_n(x, X_1, \ldots, X_{n+m}) = \frac{1}{(n+m)!} \sum_\sigma \frac{1}{n} \sum_{i=1}^{n} K_n(x, X_{\sigma(i)}) =$$

$$= \frac{1}{(n+m)!} \frac{1}{n} \sum_{i=1}^{n+m} K_n(x, X_i) \frac{(n+m)!}{n+m} n = \frac{1}{n+m} \sum_{i=1}^{n+m} K_n(x, X_i).$$

Для доказательства теоремы 14 можно непосредственно применить лемму 1, полагая там $a = (w_{n1}, \ldots, w_{nn})$, $b = (1/n, \ldots, 1/n)$, $\varphi(u_1, \ldots, u_n) = \left|\sum_{i=1}^{n} u_i\right|$ и формально подставляя $K_n(x, X_i) - f(x)$ вместо X_i.

ЛИТЕРАТУРА

Bennett G.
(1962) Probability inequalities for the sum of independent random variables, Journal of the American Statistical Association, 57, pp. 33—45.

Feller W.
(1971) An Introduction to Probability Theory and Its Applications, Vol. 2, Wiley, New York. (Имеется русский перевод: Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.)

Hoeffding W.
(1963) Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association, 58, pp. 13—30.
Huber P. J.
(1981) Robust Statistics, Wiley, New York. (Имеется русский перевод: Хьюбер П. Дж. Робастность в статистике. — М.: Мир, 1984.)
Marshall A. W., Proshan F.
(1965) An inequality for convex functions involving majorization, Journal of Mathematical Analysis and Applications, 12, pp. 87—90.
Parzen E.
(1962) On estimation of a probability density function and mode, Annals of Mathematical Statistics, 33, pp. 1065—1076.
Rao C. R.
(1973) Linear Statistical Inference and Its Applications, Wiley, New York. (Имеется русский перевод: Рао С. Р. Линейные статистические методы и их применения. — М.: Наука, 1968.)
Rosenblatt M.
(1956) Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, pp. 832—837.
Walter G., Blum J.
(1979) Probability density estimation using delta sequences, Annals of Statistics, 7, pp. 328—340.
Wertz W.
(1976) Invariant density estimation, Monatshefte für Mathematik, 81, pp. 315—324.
Петров В. В.
(1972) Суммы независимых случайных величин. — М.: Наука.
Сираждинов С. Х., Маматов М.
(1962) О сходимости в среднем для плотностей//Теория вероятн. и ее примен. — Т. 7. — 4. — С. 433—437.

Глава 12

Проекционные оценки

1. Определения

Для построения оценок плотности можно использовать богатую теорию ортогональных функций (Sansone (1977), Szegö (1975)). При этом, конечно, возникает ряд проблем, поскольку в исходной математической постановке не учитывается, что оцениваемая функция является плотностью и приближения функций частичными суммами оргогональных разложений по большей части не являются плотностями — они либо не принадлежат L_1, либо не удовлетворяют условию положительности.

Введем сначала *ортонормальную систему* на множестве B, которое обычно считается равным R или $[-\pi, \pi]$. Функции $p_0, p_1, \ldots$, составляющие ортонормальную систему, по определению удовлетворяют условию

$$\int_B p_i p_j = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

Коэффициенты Фурье a_i функции f на B определяются соотношением

$$a_i = \int_B f p_i.$$

Функция f на B допускает или не допускает разложение по системе p_i в зависимости от того, сходится ли ряд $\sum_{i=0}^{\infty} a_i p_i(x)$ и равна ли его сумма $f(x)$ или нет. Если этот ряд существует, то он называется *рядом Фурье* для f. Если ряд Фурье не существует, то функцию f нельзя восстановить по разложению в ортогональный ряд. Таким образом, важно охарактеризовать случаи, когда f допускает разложение по системе функций p_i.

Ортонормальная система называется *полной* в $L_p(B)$, если для любой функции $f \in L_p(B)$ из условия $\int f p_i = 0$, $\forall i$, следует, что $f = 0$ почти всюду. Система называется *базисом* в $L_p(B)$, если для любой функции $f \in L_p(B)$ существует единственное

разложение в сходящийся ряд вида $\sum a_i p_i$. Известно, что если B — компакт, то система полна в $L_2(B)$ тогда и только тогда, когда

$$\int_B f^2 = \sum_{i=0}^{\infty} a_i^2, \quad \forall f \in L_2(B).$$

Это соотношение называется *равенством Бесселя* [1]). Если оно выполнено, то имеет место сходимость частичных сумм в $L_2(B)$:

$$\int_B \left(f - \sum_{i=0}^{m} a_i p_i\right)^2 \to 0, \quad m \to \infty.$$

(См., например, Sansone (1977, с. 23)). В L_1 нет аналога этого свойства. При изучении сходимости в L_1 *частичных сумм*

$$S_m(f) = \sum_{i=0}^{m} a_i p_i$$

(иногда будет использоваться обозначение $S_m(f, x)$, чтобы подчеркнуть зависимость от x) нельзя воспользоваться даже неравенством Коши—Шварца

$$\int_B |S_m(f) - f| \leqslant \sqrt{\lambda(B) \int_B (S_m(f) - f)^2},$$

чтобы перекинуть мост к L_2-теории. Действительно, при этом необходимо было бы ввести условие $f \in L_2(B)$. Условие $\lambda(B) < \infty$ менее ограничительно, поскольку всегда можно монотонным преобразованием отобразить данные в некоторый компактный интервал (см., например, гл. 9).

Одно из замечательных свойств ортогональных разложений состоит в том, что если функция допускает разложение с конечным числом членов, то ее очень легко приблизить и, по-видимому, оценить. Если задана выборка $X_1, \ldots, X_n$ из f, то несмещенной оценкой коэффициента $a_i = \int f p_i$ является

$$a_{ni} = \frac{1}{n} \sum_{j=1}^{n} p_i(X_j),$$

а оценкой для $f(x)$ — функция

$$f_n(x) = \sum_{i=0}^{m} a_{ni} p_i(x).$$

[1]) Чаще используется термин *равенство Парсеваля. — Прим. перев.*

Если $f(x) = \sum_{i=0}^{m} a_i p_i(x)$, где m — конечное число, то f_n — несмещенная оценка для f. В общем случае функция f разлагается в ряд с бесконечным числом членов, так что необходимо, чтобы параметр m стремился к бесконечности с ростом n. Таким образом, *проекционную оценку* f_n можно рассматривать как непосредственное обобщение параметрических оценок.

Можно записать f_n иначе:

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} K_m(x, X_j),$$

где функция

$$K_m(x, y) = \sum_{i=0}^{m} p_i(x) p_i(y)$$

называется *ядром*. Эта запись называется записью в *правильном виде*, поскольку она напоминает определение ядерной оценки (больше о связи с ядерной оценкой см. в § 8 ниже).

Полезным результатом из фурье-анализа является *формула суммирования Кристоффеля — Дарбу* для ортогональных полиномов (см., например, Szegö (1975) с. 42—43 по русскому изданию):

$$K_m(x, y) = b_m \frac{p_{m+1}(x) p_m(y) - p_m(x) p_{m+1}(y)}{x - y},$$

где $b_m = k_m/k_{m+1}$ и k_m — коэффициент при старшей степени x в полиноме p_m.

Целое число m можно рассматривать как параметр сглаживания. В этой главе в основном будет исследоваться состоятельность и скорость сходимости проекционных оценок в случае, когда $m = m_n$ — некоторая последовательность положительных чисел. В частности, будет показано, что некоторые системы, такие, как тригонометрическая или системы Эрмита, Лагерра и Лежандра, недостаточно богаты для оценивания любых плотностей на B. Этот недостаток, конечно, уравновешивается рядом больших достоинств проекционных оценок, например их превосходным поведением в случае, когда разложение функции f в ряд содержит конечное число членов или является бесконечным с быстро убывающими коэффициентами.

При оценивании плотности важно, чтобы оценки удовлетворяли условию $\int_B f_n = 1$. Для проекционной оценки на компактном множестве B это условие выполняется, если первая функция p_0 выбрана в виде

$$p_0 = I_B/\sqrt{\lambda(B)},$$

где I — индикаторная функция. Ясно, что

$$\int_B f_n = \sum_{i=0}^{m} a_{ni} \int_B p_i = \sum_{i=0}^{m} a_{ni} \sqrt{\lambda(B)} \int_B p_i p_0 = a_{n0} \sqrt{\lambda(B)} =$$

$$= a_0 \sqrt{\lambda(B)} = \int_B f = 1.$$

Если постоянная функция включается в некоторую ортонормальную систему на R, то в общем случае получаются оценки, не принадлежащие L_1.

Исследование проекционных оценок проведено в ряде работ: Ченцов (1962), Van Ryzin (1966), Schwartz (1967), Kronmal, Tarter (1968), Bosq (1969), Watson (1969), Földes, Révész (1974). Многомерные проекционные оценки в данной главе не рассматриваются. По поводу многомерных оценок с тригонометрическим рядом читатель может обратиться к работам Kronmal, Tarter (1968), Schüler (1976), Sterbuchner (1980), Stégbuchner (1980), Greblicki, Pawlak (1981), Krzyzak, Pawlak (1982).

2. Примеры ортонормальных систем

Тригонометрическую систему на $B = [-\pi, \pi]$ образуют функции

$$p_0 = \frac{1}{\sqrt{2\pi}}, \quad p_{2i-1}(x) = \frac{\cos(ix)}{\sqrt{\pi}}, \quad p_{2i}(x) = \frac{\sin(ix)}{\sqrt{\pi}}, \quad i \geqslant 1.$$

Соответствующая проекционная оценка называется *оценкой с тригонометрическим рядом* или *оценкой с рядом Фурье* (см., например, Kronmal, Tarter (1968)). Тригонометрическая система полна в $L_1[-\pi, \pi]$, но не является базисом в $L_1[-\pi, \pi]$. Иногда удобно записывать оценку с тригонометрическим рядом в виде

$$f_n(x) = \frac{1}{2\pi} + \sum_{i=1}^{m} \left(a_{n,2i-1} \frac{\cos(ix)}{\sqrt{\pi}} + a_{n,2i} \frac{\sin(ix)}{\sqrt{\pi}} \right),$$

а иногда — в виде

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} D_m(x, X_j),$$

где D_m — *ядро Дирихле*,

$$D_m(x, y) = \frac{\sin\left(\frac{2m+1}{2}(x-y)\right)}{2\pi \sin\left(\frac{x-y}{2}\right)}.$$

Заметим, что определение числа m здесь несколько отличается от принятого в § 1, поскольку на самом деле рассматривается разложение в ряд с $2m+1$ членами.

Полиномы Лежандра образуют ортонормальную систему на $[-1, 1]$. Эти полиномы можно определять многими способами. Например, можно определить их *формулой Родрига*

$$p_i(x) = \sqrt{\frac{2i+1}{2}} \frac{1}{2^i i!} \frac{d^i}{dx^i}((x^2-1)^i), \quad i \geqslant 0.$$

Эта система полна в $L_1[-1, 1]$ (Sansone (1977, с. 191)). Соответствующее ядро $K_m(x, y)$ равно

$$K_m(x, y) = \sum_{i=0}^{m} p_i(x) p_i(y) =$$

$$= \frac{m+1}{\sqrt{2m+1}\sqrt{2m+3}} \frac{p_m(x) p_{m+1}(y) - p_{m+1}(x) p_m(y)}{y-x}.$$

Оценки с рядом Лежандра обсуждались в работах Crain (1974), Viollaz (1980) и Hall (1982). Различные явные выражения для p_i и вывод формулы для K_m имеются, например, в книгах Sansone (1977) или в Szegö (1975). Существует ряд ортонормальных систем, служащих обобщением системы Лежандра, например функции Феррера (Sansone (1977, с. 246—253)) и полиномы Якоби (Szegö (1975, гл. 4)).

Разложение в ряд Эрмита производится с помощью функций

$$p_i(x) = \frac{e^{x^2/2}}{\sqrt{2^i i!}\sqrt{\pi}} \frac{d^i}{dx^i}(e^{-x^2}), \quad i \geqslant 0.$$

Эти функции образуют ортонормальную систему, полную в $L_2(R)$. *Оценка с рядом Эрмита* изучалась применительно к задаче оценивания плотности в работах Schwartz (1967), Walter (1977), Bleuez, Bosq (1979) и Greblicki (1981). Используя формулу (5.5.9) из Szegö (1975), получим ядро

$$K_m(x, y) = \sqrt{\frac{m+1}{2}} \frac{p_{m+1}(x) p_m(y) - p_m(x) p_{m+1}(y)}{x-y}.$$

Оценка с рядом Лагерра на $B = [0, \infty)$ основана на ортонормальной и полной в $L_2[0, \infty)$ системе функций

$$p_i(x) = \left(\frac{\Gamma(i+1)}{\Gamma(i+\alpha+1)} x^{-\alpha} e^x\right)^{1/2} \cdot \frac{1}{i!} \frac{d^i}{dx^i}(x^{i+\alpha} e^{-x}), \quad i \geqslant 0.$$

Здесь $\alpha > -1$ — параметр системы. Например, при $\alpha = 0$ получим

$$p_i(x) = e^{-x/2} \sum_{j=0}^{i} (-1)^j \frac{1}{j!} \binom{i}{j} x^j.$$

Соответствующее ядро равно

$$K_m(x, y) = \frac{\Gamma(m+2)}{\Gamma(m+\alpha+1)} \frac{p_{m+1}(x)\, p_m(y) - p_m(x)\, p_{m+1}(y)}{y-x}.$$

Ортонормальная система Хаара отличается от всех предыдущих систем тем, что она является базисом в любых L_p [0, 1]. При заданном целом m функции этой системы определяются следующим образом. Пусть целые $k \geqslant 0$ и j, $1 \leqslant j \leqslant 2^k$ таковы, что $m = 2^k + j$. Тогда

$$p_m(x) = \begin{cases} 2^{k/2}, & x \in \left(\frac{j-1}{2^k}, \frac{j-1/2}{2^k}\right), \\ -2^{k/2}, & x \in \left(\frac{j-1/2}{2^k}, \frac{j}{2^k}\right), \\ 0 & \text{в противном случае.} \end{cases}$$

Эта система обладает тем нужным свойством, что $S_m(f) \to f$ почти всюду и $\int |S_m(f) - f| \to 0$ при любых $f \in L_1$ [0, 1]. Оценка по этой системе, записанная в *правильном виде*, почти совпадает с гистограммной оценкой (см., например, Bleuez, Bosq (1979)). В самом деле, ее ядро принимает только неотрицательные значения, так что легко видеть, что сама оценка f_n является плотностью. Единственное отличие от гистограммной оценки с одинаковыми интервалами состоит в том, что интервалы при различных значениях m соответствующим образом вложены (благодаря двоичной конструкции функций p_m). Оценка с рядом Хаара наследует все свойства гистограммной оценки, включая неотъемлемое ограничение на скорость сходимости средней L_1-ошибки, которая больше или равна величине, пропорциональной $n^{-1/3}$, и включая состоятельность при любых плотностях f на [0, 1]. В данной главе эта оценка больше рассматриваться не будет.

3. Общие свойства

В следующей лемме даются полезные, хотя и грубые верхняя и нижняя границы для средней L_1-ошибки (см. также лемму 3.6).

Лемма 1. *Пусть f_n — проекционная оценка для f с $m = m_n$ членами разложения, и пусть f имеет формальное разложение в, возможно, не сходящийся ряд*

$$f(x) \sim \sum_{i=0}^{\infty} a_i p_i(x),$$

где функции p_i образуют ортонормальную систему на некотором множестве B из R и $a_i = \int\limits_B f p_i$. Предположим также, что все p_i абсолютно интегрируемы на B. Тогда

$$E\left(\int |f_n - f|\right) \leqslant \int |S_m(f) - f| + E\left(\int |f_n - E(f_n)|\right) \leqslant$$

$$\leqslant \int |S_m(f) - f| + \int \sqrt{E((f_n - E(f_n))^2)} \leqslant$$

$$\leqslant \int |S_m(f) - f| + \frac{1}{\sqrt{n}} \int \sqrt{E\left(K_m^2(x, X_1)\right)}\, dx;$$

$$E\left(\int |f_n - f|\right) \geqslant \max\left(\int |S_m(f) - f|,\ \frac{1}{2}\int E(|f_n - E(f_n)|)\right).$$

Лемма 2. *Пусть $f \in L_2(B)$, и пусть функции p_i, $i \geqslant 0$, образуют ортонормальную систему на B. Тогда проекционная оценка f_n с m членами разложения имеет следующую среднюю L_2-ошибку:*

$$E\left(\int (f_n - f)^2\right) = \frac{1}{n}\sum_{i=0}^{m}\left(\int f p_i^2 - a_i^2\right) + \sum_{i=m+1}^{\infty} a_i^2 \leqslant$$

$$\leqslant \frac{1}{n}\sum_{i=0}^{m}\int f p_i^2 + \sum_{i=m+1}^{\infty} a_i^2 = \frac{1}{n}\int f(x) K_m(x, x)\, dx + \sum_{i=m+1}^{\infty} a_i^2,$$

где все интегралы берутся по множеству B.

Доказательство. Заметим сначала, что

$$\int (f_n - f)^2 = \int (f_n - E(f_n))^2 + \int (S_m(f) - f)^2.$$

В силу равенства Бесселя последнее слагаемое равно

$$\int\left(\sum_{i=m+1}^{\infty} a_i p_i\right)^2 = \sum_{i=m+1}^{\infty} a_i^2.$$

Здесь мы воспользовались свойством ортонормальности. Кроме того,

$$\int (f_n - E(f_n))^2 = \int\left(\sum_{i=0}^{m}(a_i - a_{ni})\, p_i\right)^2 = \sum_{i=0}^{m}(a_i - a_{ni})^2,$$

откуда без труда получается нужный результат.

Так как мы требуем интегрируемости оценок, разумно потребовать, чтобы все функции p_i ортонормальной системы были абсолютно интегрируемы. Однако это ограничение приводит

к некоторым плохим побочным эффектам в случае, когда рассматривается проекционная оценка на неограниченном множестве. Эти побочные эффекты описаны в следующей лемме.

Лемма 3. *Пусть* p_i, $i \geqslant 0$, — *ортонормальная система на* R (*или на* $[0, \infty)$) *и все функции* p_i *абсолютно интегрируемы. Тогда*

(i) *если оценка* f_n *трансляционно инвариантна* (см. § 6.6), *то* $\int f_n = 0$;

(ii) *невозможно, чтобы равенство* $\int f_n = 1$ *выполнялось почти наверное при любых* f.

Доказательство. Если оценка f_n трансляционно инвариантна, то функция $K_m(x, y)$ должна иметь вид $K_m^*(x - y)$, где K_m^* — некоторая функция, и интеграл $\int K_m(x, y)\, dy$ должен не зависеть от x. Следовательно, величина

$$p_i(x) \sum_{i=0}^{m} p_i(y)\, dy$$

должна не зависеть от x. Так как все функции p_i невырожденны $\left(\int p_i^2 = 1\right)$ и ни одна из них не равна постоянной почти всюду (в силу того что $\int p_i^2 = 1$), то ясно, что $\sum_{i=0}^{m} \int p_i(y)\, dy = 0$, и потому $\int K_m(x, y)\, dx = 0$ при любых y. Таким образом, $\int f_n = 0$.

Докажем утверждение (ii) от противного. Если $\int f_n = 1$ почти наверное при любых f, то $\sum_{i=0}^{m} \int p_i(x)\, p_i(y)\, dx = 1$ при почти всех y. Возводя это равенство в квадрат и интегрируя по dy, получим

$$\infty = \int \left(\sum_{i=0}^{m} p_i(y) \int p_i \right)^2 dy = \sum_{i=0}^{m} \int p_i^2 \left(\int p_i \right)^2 = \sum_{i=0}^{m} \left(\int p_i \right)^2,$$

что, очевидно, невозможно в силу сделанных предположений.

Лемма 3 не оставляет сомнений в том, что возможности проекционных оценок на R или на $[0, \infty)$ ограниченны. По этой причине мы будем в основном рассматривать свойства проекционных оценок на компактных множествах.

4. Состоятельность оценки с тригонометрическим рядом

В этом параграфе преследуются три цели. Строятся плотности на $[-\pi, \pi]$ со сходящимися рядами Фурье, такие, что для них непригодна оценка с тригонометрическим рядом при сколь угодно больших n. Чтобы уравновесить этот результат, приводятся слабые достаточные условия состоятельности. Наконец, очень кратко рассматривается вопрос о необходимости этих условий.

Теорема 1 (несостоятельность оценки с тригонометрическим рядом). *Пусть* $a_m \downarrow 0$ — *последовательность чисел, выпуклая как функция от* m *и такая, что* $a_0 = 1/\sqrt{\pi}$. *Тогда ряд*

$$\frac{1}{2\pi} + \sum_{i=1}^{\infty} a_i \frac{\cos(ix)}{\sqrt{\pi}}$$

сходится всюду, кроме, быть может, одной точки, и является рядом Фурье некоторой плотности f *на* $[-\pi, \pi]$.

Пусть f_n — *оценка с тригонометрическим рядом, построенная по выборке объема* n *и имеющая параметр* m. *Если*

$$\liminf_{m\to\infty} a_m \log m > \pi^{3/2}/2,$$

то

$$\inf_{n,\, m} E\left(\int |f_n - f|\right) > 0.$$

Если $\lim\limits_{n\to\infty} m = \infty$ *и* $\lim\limits_{m\to\infty} a_m \log m = \infty$, *то*

$$\liminf_{n\to\infty} \frac{E\left(\int |f_n - f|\right)}{a_m \log m} \geqslant \frac{4}{\pi^{3/2}}.$$

В теореме 1 утверждается, что многие плотности на $[-\pi, \pi]$, даже такие, ряд Фурье которых сходится (всюду, кроме одной точки), невозможно оценить ни при каком выборе m и n. Причина тому — уход смещения $\left(\int |S_m(f) - f| \to \infty\right.$ при $n \to \infty$, если $m \to \infty$, $\left.a_m \log m \to \infty\right)$. Таким образом, величина $E\left(\int |f_n - f|\right)$ может возрастать с любой заданной скоростью, равной $o(\log m)$ при $m \to \infty$.

Для доказательства теоремы 1 требуются некоторые вспомогательные результаты, в частности ряд свойств *ядра Дирихле*

$D_m(x, y)$. Так как оно является функцией только от разности $x - y$, то в дальнейшем мы позволим себе писать

$$D_m(u) = \frac{\sin((m + 1/2)u)}{2\pi \sin(u/2)}.$$

Другая важная функция — *ядро Фейера*

$$F_m(u) = \frac{1}{m+1}\sum_{i=0}^{m} D_i(u) = \frac{1}{2\pi(m+1)}\left(\frac{\sin((m+1)u/2)}{\sin(u/2)}\right)^2.$$

Лемма 4 (свойства ядер Дирихле и Фейера).

A. $\int_{-\pi}^{\pi} D_m(u)\,du = \int_{-\pi}^{\pi} F_m(u)\,du = 1.$

B. $|D_m(u)| \leqslant 1/2\,|u|,\ |u| \leqslant \pi;$
$|D_m(u)| \leqslant (m+1)/2,\ |u| \leqslant \pi.$

C. $\int |D_m| \sim (4/\pi^2)\log m$ при $m \to \infty$. $\left(\int |D_m|\right.$ называется постоянной Лебега$\left.\right)$.

D. $\int |D_m| \leqslant 2 + (4/\pi^2)\log m,\ m \geqslant 1.$

E. $F_m(u) \leqslant \pi/2\,(m+1)u^2,\ |u| \leqslant \pi,\ F_m(u) \leqslant (m+1)/4.$

Доказательство. Свойство A хорошо известно. Оно следует непосредственно из определения D_m и F_m и условий ортонормальности. Свойство B следует из неравенств

$$|D_m(u)| \leqslant (2\pi|\sin(u/2)|)^{-1} \leqslant \sup_v \left|\frac{v}{\sin v}\right| \frac{1}{\pi|u|} \leqslant \frac{1}{2|u|},$$

$$|D_m(u)| \leqslant \sup_v \left|\frac{\sin v}{v}\right| \left|\frac{(m+1/2)}{\pi}\right| \frac{|u/2|}{\sin(|u/2|)} \leqslant \frac{1}{2}\left(m + \frac{1}{2}\right),$$

где u, v — любые числа, принадлежащие $[-\pi, \pi]$.

Свойство C читатель может найти у Бари (1961, с. 115). Оно наполовину следует, конечно, из свойства D, полное доказательство которого будет сейчас приведено:

$$\int |D_m| = \frac{1}{\pi}\int_0^{\pi} \left|\frac{\sin((m+1/2)u)}{\sin(u/2)}\right| du = \frac{2}{\pi}\int_0^{\pi/2} \left|\frac{\sin((2m+1)y)}{\sin y}\right| dy \leqslant$$

$$\leqslant \frac{2}{\pi}\int_0^{\pi/2} \left|\frac{\sin((2m+1)y)}{y}\right| dy + \frac{2}{\pi}\int_0^{\pi/2} \left|\frac{1}{\sin y} - \frac{1}{y}\right| dy.$$

Обозначим последние два слагаемых I_1 и I_2. Легко проверить, что $I_2 \leqslant \pi^2/48$. Это получается интегрированием неравенства

$$\left|\frac{y-\sin y}{y\sin y}\right| \leqslant \left|\frac{y^3/6}{(2/\pi)\,y^2}\right| \leqslant \frac{\pi y}{12}, \quad 0 \leqslant y \leqslant \frac{\pi}{2}.$$

Рассматривая слагаемое I_1, поступим так же, как Бари (1961, 114) или Edwards (1979, с. 80—81). Полагая $(2m+1)\,y = t$, получим

$$I_1 = \frac{2}{\pi}\int_0^{(2m+1)\,\pi/2}\left|\frac{\sin t}{t}\right| dt = \frac{2}{\pi}\sum_{k=0}^{2m}\int_{k\pi/2}^{(k+1)\,\pi/2}\frac{|\sin t|}{t}\,dt \leqslant$$

$$\leqslant \frac{2}{\pi}\sum_{k=1}^{2m}\int_{k\pi/2}^{(k+1)\,\pi/2}\left|\frac{\sin t}{k/2}\right| dt + 1 =$$

$$= \frac{4}{\pi^2}\sum_{k=1}^{2m}\frac{1}{k} + 1 \leqslant \frac{4}{\pi^2}\,(1+\log(2m)) + 1.$$

Замечая теперь, что $\pi^2/48 + 1 + 4/\pi^2 + 4\log 2/\pi^2 \leqslant 2$, получим свойство D.

Наконец, свойство E следует из неравенств

$$F_m(u) \leqslant (2\pi\,(m+1)\sin^2(u/2))^{-1} \leqslant (2\,(m+1)\,u^2)^{-1}\,\pi$$

и

$$F_m(u) \leqslant \frac{1}{m+1}\sum_{i=0}^{m}\frac{1}{2}\left(i+\frac{1}{2}\right) =$$

$$= \frac{1}{2\,(m+1)}\left(\frac{m+1}{2} + \frac{m\,(m+1)}{2}\right) = \frac{m+1}{4}.$$

Лемма 5 (теорема Фейера—Лебега, см., например, Бари (1961, с. 143)). *Если положить*

$$\sigma_m(f) = \int_{-\pi}^{\pi} f(t)\,F_m(t-x)\,dt,$$

то для любых плотностей f на $[-\pi, \pi]$ имеем $\sigma_m(f) \to f$ при $m \to \infty$ для почти всех x и $\int|\sigma_m(f) - f| \to 0$ при $m \to \infty$.

Замечание. Удобно доопределить f вне $[-\pi, \pi]$ по периодичности. Это упростит обозначения. Заметим также, что D_m и F_m — периодические функции с периодом 2π.

Доказательство леммы 5.

$$|\sigma_m(f) - f| \leqslant \int |f(u+x) - f(x)|\, F_m(u)\, du \leqslant$$

$$\leqslant \delta^{-1} \int\limits_{|u|\leqslant\delta} |f(u+x) - f(x)|\, du \left(\frac{m+1}{4}\right)\delta + \int\limits_{|u|\geqslant\delta} \frac{|f(u+x) - f(x)|\,\pi}{2(m+1)u^2}\, du.$$

Здесь $\delta > 0$ — произвольное число. Выберем $\delta = 1/(m+1)$ и заметим, что первое слагаемое в правой части тогда равно $o(1)$ при почти всех x по теореме Лебега о плотностях. При анализе второго слагаемого введем обозначения $g(u) = |f(u+x) - f(x)|$ и $G(u) = \int\limits_0^u g(v)\, dv$. Имеем $G(u) = o(u)$ при $u \downarrow 0$ для почти всех x. Интегрируя по частям, заметим, что

$$\int\limits_{\pi\geqslant u\geqslant\delta} g(u) \frac{1}{(m+1)u^2}\, du \leqslant \frac{1}{m}\left(\frac{G(\pi)}{\pi^2} - \frac{G(\delta)}{\delta^2} + \int\limits_\delta^\pi \frac{2G(u)}{u^3}\, du\right) =$$

$$= o(1) + \frac{2}{m} \int\limits_{1/(1+m)}^{\pi} \frac{G(u)}{u^3}\, du = o(1) + \frac{2}{m}\, o\left(\int\limits_{1/m}^{\infty} u^{-2}\, du\right) = o(1)$$

при почти всех x. Второе утверждение леммы следует из первого, если заметить, что σ_m является плотностью при любых m.

Доказательство теоремы 1. Приведем сначала некоторые хорошо известные свойства рядов Фурье. Рассмотрим ряд

$$\frac{a_0}{\sqrt{2\pi}} + \sum_{i=1}^{\infty} a_i \frac{\cos(ix)}{\sqrt{\pi}},$$

где $a_0\sqrt{2}$, a_1, a_2, a_3, ... — невозрастающая последовательность, стремящаяся к 0. Этот ряд сходится всюду на $[-\pi, \pi]$, за исключением, быть может, точки 0, и сходимость его равномерна на $[\varepsilon, \pi]$ при любых $\varepsilon > 0$ (простое доказательство этого см. Бари (1961, с. 95)). Если, кроме того, эта последовательность выпукла, то ряд сходится к некоторой неотрицательной интегрируемой

функции f на $[-n, n]$ (всюду, кроме, быть может, точки $x = 0$) и является ее рядом Фурье, т. е.

$$a_i = \int_{-\pi}^{\pi} f(x) \frac{\cos(ix)}{\sqrt{\pi}}\, dx, \quad i \geqslant 1;$$

$$a_0 = \int_{-\pi}^{\pi} f(x) \frac{1}{\sqrt{2\pi}}\, dx$$

(см. Бари (1961, с. 100)). Так как в нашем случае a_0 здесь равно $1/\sqrt{2n}$, то f действительно является плотностью на $[-\pi, \pi]$.

Для удобства обозначений положим $b_0 = a_0\sqrt{2/\pi}$, $b_i = a_i/\sqrt{n}$, $i \geqslant 1$, так что

$$S_m(f) = b_0/2 + \sum_{i=1}^{m} b_i \cos(ix)$$

и последовательность b_i выпукла и монотонно стремится к 0 при $i \to \infty$. Таким образом, $\Delta b_i = b_i - b_{i+1} \geqslant 0$ при любых i и $\Delta^2 b_i = \Delta b_i - \Delta b_{i+1} \geqslant 0$ при любых i.

Заметим, что $D_m(x) = 1/2 + \sum_{i=1}^{m} \cos(ix)$ и $\sum_{i=0}^{m} D_i(x) = (m+1) \times$ $\times F_m(x)$. Используя эти тождества и преобразование Абеля, получим

$$S_m(f)/\pi = \sum_{i=0}^{m-1} \Delta b_i D_i(x) + b_m D_m(x) =$$

$$= \Delta b_{m-1} \sum_{i=0}^{m-1} D_i(x) + \sum_{i=0}^{m-2} \Delta^2 b_i \sum_{k=0}^{i} D_k(x) + b_m D_m(x) =$$

$$= \Delta b_{m+1} m F_{m-1}(x) + \sum_{i=0}^{m-2} \Delta^2 b_i (i+1) F_i(x) + b_m D_m(x).$$

Первые два слагаемых в последнем выражении неотрицательны. (Кстати, тем самым показано, что частичные суммы сходятся к неотрицательной функции, поскольку последнее слагаемое в последнем выражении равно $o(1)$ при $x \neq 0$.) Сумма интегралов по $[-\pi, \pi]$ от первых двух слагаемых равна

$$m\Delta b_{m-1} + \sum_{i=0}^{m-2} (i+1)\Delta^2 b_i = \sum_{i=0}^{m-1} \Delta b_i = b_0 - b_m.$$

Следовательно,

$$\int |S_m(f)| \geqslant \left(b_m \int |D_m| + b_m - b_0\right)\pi =$$

$$= \sqrt{\pi}\left(a_m \int |D_m| + a_m - a_0\sqrt{2}\right) = a_m\sqrt{\pi}\left(1 + \int |D_m|\right) - 1,$$

и, значит, в силу леммы 1

$$E\left(\int |f_n - f|\right) \geqslant \int |S_m(f) - f| \geqslant a_m \sqrt{\pi}\left(1 + \int |D_m|\right) - 2.$$

По лемме 4 эта нижняя граница равна $(4/\pi^{3/2} + o(1))\, a_m \log m - 2$. Отсюда непосредственно следует последнее утверждение теоремы 1. Докажем второе утверждение теоремы 1. Заметим, что существуют положительные постоянные c, M, такие, что

$$\inf_{m \geqslant M} \int |S_m(f) - f| > c.$$

Но так как, очевидно, $\inf_{m \leqslant M} \int S_m(f) - f| > 0$ при любых конечных M, то второе утверждение доказано.

Примерами к теореме 1 являются хорошие плотности, потому что их ряды Фурье сходятся при всех $x \neq 0$. Следует отметить, что существуют плотности $f \in L_1[-\pi, \pi]$, для которых суммы $S_m(f)$ не сходятся ни в одной точке. Такие f в некотором смысле хуже, чем плотности из теоремы 1. Так как сама по себе поточечная сходимость нас не интересует, то мы приведем лишь формулировки некоторых известных результатов о поточечной сходимости частичных сумм интегрируемых функций f.

Лемма 6 (поточечная сходимость рядов Фурье).

A. *При любых* $f \in L_1[-\pi, \pi]$ *имеет место сходимость* $S_{m_k}(f) \to f$ *почти всюду, где* m_k *— некоторая подпоследовательность. Более того,* $S_m(f) = o(\log m)$ *при почти всех* x *и* $\int |S_m(f)| = o(\log m)$.

B. *Если* $f \in L_p[-\pi, \pi]$ *при некотором* $p > 1$, *то* $S_m(f) \to f$ *при почти всех* x.

C. *Существует функция* f *из* $L_1[-\pi, \pi]$, *такая, что* $\limsup_{m\to\infty} S_m(f) = \infty$ *при любых* x. *В действительности,* $\limsup_{m\to\infty} |S_m \times (f)|/\log\log m = \infty$ *при любых* x *для некоторой функции* f *из* $L_1[-\pi, \pi]$.

D. *Для любой последовательности* $m_k \uparrow \infty$ *существует функция* f *из* $L_1[-\pi, \pi]$, *такая, что* $\limsup_{k\to\infty} S_{m_k}(f) = \infty$ *при почти всех* x.

E. *Для любой последовательности* $c_m \downarrow 0$ *существует функция* f *из* $L_1[-\pi, \pi]$, *такая, что* $\limsup_{m\to\infty} \int |S_m(f)|/(c_m \log m) = \infty$.

Замечание. Утверждение A можно найти в Zygmund (1959, разд. 7.3) и Edwards (1979, с. 167, 180). При $p = 2$ свойство B известно под названием *теоремы Карлесона* (Carleson (1966)).

Общее утверждение при $p > 1$ было доказано Hunt (1968) (см. также Mozzochi (1971) и книгу Jorsboe, Mejlbro (1982), где приведены другие доказательства этой глубокой *теоремы Карлесона—Ханта*). Первая часть утверждения C известна как *контрпример Колмогорова* (Kolmogorov (1926); см. Zygmund(1959, разд. 8.4)). Вторая часть утверждения C связана с уточнением, принадлежащим Körner (1981), который также доказал D, опираясь на идеи Кахана и Стейна. Свойство E можно найти у Edwards (1979, с. 180). Заметим также, что теорема Карлесона—Ханта была обобщена на d-мерный случай в работах Fefferman (1971) и Sjölin (1971).

Установив, что оценка с тригонометрическим рядом не является универсально состоятельной, получим все же некоторые позитивные результаты о ее состоятельности для определенных классов плотностей. Сильную сходимость L_1-ошибки оценки изучать не будем. Следующая лемма поможет нам при анализе смещения $\int |S_m(f) - f|$.

Лемма 7. *Пусть f — плотность на $[-\pi, \pi]$. Тогда*

(i) $\lim_{m\to\infty} \int |S_m(f) - f|^p = 0$ *при любых* $p \in (0, 1)$.

(ii) $\lim_{m\to\infty} \int |S_m(f) - f| = 0$, *если* $f \in L_p$ *при некотором* $p > 1$.

(iii) $\lim_{m\to\infty} \int |S_m(f) - f| = 0$, *если* $\int f \log_+ f < \infty$,

и $\int |S_m(f)| \leqslant A \int f \log_+ f + B$, *где* A, B — *некоторые универсальные постоянные.*

Замечание. За доказательством леммы 7 мы отсылаем читателя к разд. 7.3 книги Zygmund (1959).

Лемма 8. *Для любых плотностей f на $[-\pi, \pi]$ и оценки с тригонометрическим рядом, имеющей параметр m, выполнено неравенство*

$$E\left(\int |f_n - f|\right) \leqslant \sqrt{\frac{2m+1}{n}} + \int |S_m(f) - f|.$$

Если $\lim_{n\to\infty} m = \infty$, *то*

$$E\left(\int |f_n - f|\right) \leqslant \sqrt{\frac{m}{\pi n}}\left(\int \sqrt{f} + o(1)\right) + \int |S_m(f) - f|.$$

Теорема 2 (состоятельность оценки с тригонометрическим рядом). *Пусть f — плотность на $[-\pi, \pi]$, удовлетворяющая усло-*

вию на пики $\int f \log_+ f < \infty$ *(это условие выполнено, если* $f \in L_p[-\pi, \pi]$ *при некотором* $p > 1$*). Пусть* f_n *— оценка с тригонометрическим рядом для* f*, имеющая параметр* m*, и пусть*

$$\lim_{n\to\infty} m = \infty, \quad \lim_{n\to\infty} (m/n) = 0.$$

Тогда $E\left(\int |f_n - f|\right) \to 0$ *при* $n \to \infty$*. Если, кроме того,* $f \in L_p[-\pi, \pi]$ *при некотором* $p > 1$*, то* $E((f_n - f)^2) \to 0$ *при* $n \to \infty$ *для почти всех* x*.*

Доказательство леммы 8 *и теоремы* 2. Первое неравенство леммы 8 следует из леммы 1. Член

$$\frac{1}{\sqrt{n}} \int \sqrt{E\left(D_m^2(x - X_1)\right)}\, dx$$

равен

$$\sqrt{\frac{2m+1}{2\pi n}} \int \sqrt{E(F_{2m}(x - X_1))}\, dx \leqslant$$

$$\leqslant \sqrt{\frac{2m+1}{2\pi n}} \sqrt{2\pi} \sqrt{\int E(F_{2m}(x - X_1))\, dx} = \sqrt{\frac{2m+1}{n}},$$

где мы воспользовались неравенством Коши — Шварца и леммой 4. Используя обозначение $\sigma_{2m}(f) = E(F_{2m}(x - X_1))$ из леммы 5, вспоминая, что $\sigma_{2m}(f)$ — плотность по x при каждом m и что $\int |\sigma_{2m}(f) - f| \to 0$, перепишем этот же член в виде

$$\sqrt{\frac{2m+1}{2\pi n}} \int \sqrt{\sigma_{2m}(f)}$$

и получим

$$\int \sqrt{\sigma_{2m}(f)} = \int \sqrt{(\sigma_{2m}(f) - f) + f} \leqslant \int \sqrt{|\sigma_{2m}(f) - f|} + \int \sqrt{f} \leqslant$$

$$\leqslant \sqrt{2\pi \int |\sigma_{2m}(f) - f|} + \int \sqrt{f} = o(1) + \int \sqrt{f}.$$

Лемма 8 доказана.

Первое утверждение теоремы 2 следует из лемм 7 и 8. Остается доказать лишь утверждение о сходимости почти всюду. В силу теоремы Карлесона — Ханта (лемма 6, утверждение B) достаточно показать, что $E((f_n - E(f_n))^2) \to 0$ при почти всех x. Но $f_n - E(f_n) = (1/n) \sum_{j=1}^{n} Y_j$, где $Y_j = D_m(x - X_j) - E(D_m(x - X_j))$. Следовательно, достаточно, чтобы было $E(Y_1^2)/n \to 0$ почти всюду. Это в свою очередь следует из того, что $E(D_m^2(x - X_1))/n \to 0$ почти всюду. Но $D_m^2(u) = ((2m + 1)/2) F_{2m}(u)$ при любых u. Таким образом, достаточно условия $m/n \to 0$ и того, чтобы вели-

чина $E(F_{2m}(x - X_1)) = \sigma_{2m}(f)$ была ограничена при почти всех x. Но последнее является следствием теоремы Фейера — Лебега (лемма 5).

Наконец, кратко рассмотрим вопрос о необходимости условий на m. Ясно, что если f не имеет конечного разложения в ряд Фурье, то условие $m \to \infty$ необходимо для того, чтобы было $\int |S_m(f) - f| \to 0$, и, следовательно, необходимо для сходимости $E\left(\int |f_n - f|\right) \to 0$. Следующее простое построение показывает, что в действительности существует много плотностей с конечным разложением в ряд Фурье: используя многократно формулу $\cos^2 x = (1 + \cos 2x)/2$ и формулу бинома, убеждаемся, что функции вида $(\cos x)^{2^r}$, где r — натуральное число, имеют разложение со старшим ненулевым коэффициентом Фурье a_{2^r+1}. Таким образом, если нормировать эти функции, превращая их в плотности, и взять фиксированное m, большее или равное 2^r, то получится

$$E\left(\int |f_n - f|\right) \leqslant \sqrt{\frac{2m+1}{n}}.$$

Правая часть этого неравенства стремится к 0 со скоростью $1/\sqrt{n}$. Заметим, что ядерная оценка не может достичь такой скорости сходимости для таких плотностей. Конечно, это связано с тем, что оценка с тригонометрическим рядом как бы специально предназначена для плотностей такого вида.

Лемма 9 (достижимость ошибки порядка $O(1/\sqrt{n})$). *Пусть f — плотность на $[-\pi, \pi]$, и пусть f_n — оценка с тригонометрическим рядом, имеющая параметр m. Тогда из неравенства* $\limsup_{n\to\infty} E\left(\int |f_n - f|\right)\sqrt{n} < \infty$ *следует, что* $\limsup_{n\to\infty} m < \infty$.

Доказательство. Достаточно показать, что для любых плотностей f на $[-\pi, \pi]$, таких, что $\int |S_m(f) - f| \to 0$ при $m \to \infty$, из условия $\lim_{n\to\infty} m = \infty$ следует, что

$$E\left(\int |f_n - f|\right) \geqslant \frac{\log m}{\sqrt{n}}(A + o(1)),$$

где $A > 0$ — некоторая универсальная постоянная.

При $m \to \infty$ для любых f имеем

$$E\left(\int |f_n - f|\right) \geqslant E\left(\int |f_n - E(f_n)|\right)/2 \text{ (по лемме 1)}$$

$$\geqslant (32n)^{-1/2} \int E\left(|D_m(x - X_1) - E(D_m(x - X_1))|\right) dx \text{ (по лемме 5.27)}$$

$$\geqslant (32n)^{-1/2} \int (E(|D_m(x - X_1)|) - |S_m(f)|)\,dx$$

$$\geqslant (32n)^{-1/2} \left(\int |D_m| - \int |S_m(f) - f| - 1 \right).$$

Если $\int |S_m(f) - f| \to 0$ при $m \to \infty$, то в силу леммы 4 полученная нижняя граница асимптотически ведет себя как $(4/\pi^2) \log m/(\sqrt{32n})$. Лемма 9 доказана.

Из леммы 5.27 с помощью некоторых рассуждений можно также получить, что условие $m/n = o(1)$ необходимо для состоятельности оценок. Здесь эти рассуждения приводиться не будут. Отметим, что Bosq, Bleuez (1978) и Bleuez, Bosq (1979) показали, что для плотностей $f \in L_2[-\pi, \pi]$, ряд Фурье которых сходится всюду, и в предположении, что $\lim_{n\to\infty} m = \infty$, следующие условия эквивалентны:

(i) $f_n \to f$ по вероятности при любых x и любых заданных f;
(ii) $E(|f_n - f|) \to 0$ при любых x и любых заданных f;
(iii) $E((f_n - f)^2) \to 0$ при любых x и любых заданных f;
(iv) $E\left(\int (f_n - f)^2\right) \to 0$ при любых заданных f;
(v) $\lim_{n\to\infty} m/n = 0$.

Используя имеющуюся информацию (см., например, лемму 2), читатель может легко доказать этот результат. В цитированных работах Bleuez и Bosq результаты такого типа получены для многих проекционных оценок как следствие из очень общей теоремы.

5. Скорость сходимости оценки с тригонометрическим рядом

Из предыдущего параграфа непосредственно следует, что скорости сходимости оценки с тригонометрическим рядом и ядерной оценки несравнимы. В случае равномерной плотности на $[-\pi, \pi]$ для оценки с тригонометрическим рядом имеем $f_n = f$ при $m = 0$, тогда как ядерная оценка не может сходиться к этой плотности быстрее, чем со скоростью $n^{-1/3}$. Это же ограничение снизу на скорость сходимости имеет место для плотностей, представляющих собой смеси равномерной плотности с плотностями, пропорциональными $(\cos x)^{2r}$, где r — натуральное число, $|x| \leqslant \pi$. Для таких смесей оценка с тригонометрическим рядом достигает скорости сходимости $1/\sqrt{n}$, если последовательность m ограниченна и $m \geqslant 2r$. С другой стороны, оценка с тригонометрическим рядом часто не является даже состоятельной (теорема 1).

В этом параграфе нашей первой целью является доказательство того, что при подходящем выборе m оценка с тригонометрическим рядом имеет равномерно ограниченную на классах Липшица $W(s, \alpha, C)$ (см. § 4.2) L_1-ошибку и что граница для ошибки совпадает с точностью до константы с минимаксной нижней границей теоремы 4.6. Этим свойством обладает также и ядерная оценка. В конце этого параграфа сделаем некоторые замечания о поведении оценки с тригонометрическим рядом на классах Бретаньоля — Юбер и классах Соболева. В определении класса $W(s, \alpha, C)$ заменим интервал $[0, 1]$ на $[-\pi, \pi]$ и будем предполагать, что условия Липшица и гладкости на f имеют место на всей действительной оси (так же, как в гл. 4). Это предположение очень важно, поскольку оно влечет за собой, что $f = 0$ в точках $-\pi$ и $+\pi$ и что функция f является достаточно гладкой вблизи этих точек. Таким образом, нас не будет интересовать эффект Гиббса (см., например, Hall (1981)).

Исследование скоростей сходимости для фиксированной плотности f оказывается гораздо более трудным, чем в случае ядерной оценки, поскольку оно зависит от того, с какой скоростью стремится к 0 величина $\int |S_m(f) - f|$ (см. лемму 1), а эта скорость не связана, во всяком случае явно, с такими обычными количественными характеристиками, как $\int |f^{(r)}|$. Например, для плотностей с конечным разложением в ряд Фурье трудности начинаются при получении нижних границ для средней L_1-ошибки.

Лемма 10. *Если f — плотность из $L_2[-\pi, \pi]$ с коэффициентами Фурье a_i, то*

$$\int |S_m(f) - f| \leqslant \sqrt{2\pi} \sqrt{\sum_{i=2m+1}^{\infty} a_i^2}.$$

Доказательство. В силу неравенства Коши — Шварца и условия ортонормальности

$$\int |S_m(f) - f| \leqslant \sqrt{2\pi} \sqrt{\int (S_m(f) - f)^2} = \sqrt{2\pi} \sqrt{\sum_{i=2m+1}^{\infty} a_i^2}.$$

Лемма 11 (неравенство Лоренца; см. также Бари (1961, с. 208)). *Пусть $f \in W(0, \alpha, C)$ при некотором $\alpha \in (0, 1]$, и пусть коэффициенты Фурье функции f равны a_i, $i \geqslant 0$. Тогда*

$$\sum_{i=2m-1}^{\infty} a_i^2 \leqslant \gamma C^2 / m^{2\alpha},$$

где $\gamma = \pi^{2\alpha+1}/(4^\alpha - 1)$, $m \geqslant 1$.

Доказательство. Для удобства обозначений предположим, что f — периодическая функция с периодом 2π (доопределим ее вне $[-\pi, \pi]$ по периодичности). Если разложение функции f в ряд Фурье имеет вид

$$\frac{a_0}{\sqrt{2\pi}} + \sum_{i=1}^{\infty} \left(a_{2i-1} \frac{\cos ix}{\sqrt{\pi}} + a_{2i} \frac{\sin ix}{\sqrt{\pi}} \right),$$

то разложением в ряд Фурье для $f(x+h) - f(x-h)$ как функции от x (h — постоянная) будет

$$2 \sum_{i=1}^{\infty} \left(a_{2i} \frac{\sin ix}{\sqrt{\pi}} - a_{2i-1} \frac{\cos ix}{\sqrt{\pi}} \right) \sin ih.$$

В силу равенства Бесселя и условия Липшица

$$4 \sum_{i=1}^{\infty} (a_{2i-1}^2 + a_{2i}^2) \sin^2 ih = \int_{-\pi}^{\pi} (f(x+h) - f(x-h))^2 dx \leqslant$$

$$\leqslant \int_{-\pi}^{\pi} (C(2h)^{\alpha})^2 dx = 2\pi C^2 (2h)^{2\alpha}.$$

Следовательно, при каждом m

$$\sum_{i=m}^{2m-1} (a_{2i-1}^2 + a_{2i}^2) \sin^2 ih \leqslant (\pi/2)\, 4^{\alpha} C^{\alpha} h^{2\alpha}.$$

Полагая $h = \pi/4m$ и замечая, что $\sin ih \geqslant \sin(\pi/4) = 1/\sqrt{2}$ при $m \leqslant i \leqslant 2m-1$, получим

$$\sum_{i=m}^{2m-1} (a_{2i-1}^2 + a_{2i}^2) \leqslant \frac{\pi^{2\alpha+1} C^2}{4^{\alpha} m^{2\alpha}}.$$

Таким образом,

$$\sum_{i=2m-1}^{\infty} a_i^2 = \sum_{i=m}^{\infty} (a_{2i-1}^2 + a_{2i}^2) = \sum_{j=0}^{\infty} \sum_{i=m2^j}^{m2^{j+1}-1} (a_{2i-1}^2 + a_{2i}^2) \leqslant$$

$$\leqslant \sum_{j=0}^{\infty} \frac{\pi^{2\alpha+1} C^2}{4^{\alpha} (m2^j)^{2\alpha}} = \frac{\pi^{2\alpha+1} C^2}{4^{\alpha} m^{2\alpha}} \cdot \frac{1}{1 - 4^{-\alpha}},$$

что и требовалось доказать.

Лемма 12. *Пусть* $f \in W(s, \alpha, C)$, *где* $s \geqslant 0$ — *некоторое целое число и* $\alpha \in (0, 1]$. *В обозначениях леммы* 11

$$\sum_{i=2m-1}^{\infty} a_i^2 \leqslant \gamma C^2 / m^{2s+2\alpha},$$

где

$$\gamma = \frac{\pi^{2\alpha+1}}{4^{\alpha} - 4^{-s}}, \quad m \geqslant 1.$$

Доказательство. Заметим сначала, что если f имеет $s-1$ абсолютно непрерывных производных и функция $f^{(s)}$ удовлетворяет условию Липшица, то допустимо формальное дифференцирование ряда Фурье и $f^{(s)}$ раскладывается в ряд Фурье

$$\sum_{i=1}^{\infty} i^s \left((-1)^{s/2} a_{2i-1} \frac{\cos ix}{\sqrt{\pi}} + (-1)^{s/2} a_{2i} \frac{\sin ix}{\sqrt{\pi}} \right)$$

при четном s и в ряд Фурье

$$\sum_{i=1}^{\infty} i^s \left((-1)^{(s+1)/2} a_{2i-1} \frac{\sin ix}{\sqrt{\pi}} + (-1)^{(s-1)/2} a_{2i} \frac{\cos ix}{\sqrt{\pi}} \right)$$

при нечетном s. Из равенства Бесселя следует

$$\sum_{i=1}^{\infty} (a_{2i-1}^2 + a_{2i}^2) i^{2s} = \int (f^{(s)})^2.$$

Это равенство полезно для дальнейшего. Нетрудно проверить, что функция $f^{(s)}(x+h) - f^{(s)}(x-h)$ при четном s и фиксированном h раскладывается в ряд Фурье

$$2 \sum_{i=1}^{\infty} i^s \left((-1)^{s/2} a_{2i-1} \left(\frac{-\sin ix}{\sqrt{\pi}} \right) + (-1)^{s/2} a_{2i} \frac{\cos ix}{\sqrt{\pi}} \right) \sin ih.$$

При нечетном s имеет место аналогичное разложение. Следовательно, рассуждая так же, как в доказательстве леммы 11, получим

$$4 \sum_{i=1}^{\infty} (a_{2i-1}^2 + a_{2i}^2) i^{2s} \sin^2 ih = \int_{-\pi}^{\pi} (f^{(s)}(x+h) - f^{(s)}(x-h))^2 dx \leqslant$$

$$\leqslant \int_{-\pi}^{\pi} (C(2h)^{\alpha})^2 dx = 2\pi C^2 (2h)^{2\alpha}.$$

Так же как в лемме 11, получаем

$$\sum_{i=m}^{2m-1}(a_{2i-1}^2+a_{2i}^2)\,i^{2s}\leqslant\frac{\pi^{2\alpha+1}C^2}{4^\alpha m^{2\alpha}}.$$

Отсюда

$$\sum_{i=2m-1}^{\infty}a_i^2=\sum_{i=m}^{\infty}(a_{2i-1}^2+a_{2i}^2)\leqslant$$

$$\leqslant\sum_{j=0}^{\infty}\sum_{i=m2^j}^{m2^{j+1}-1}(a_{2i-1}^2+a_{2i}^2)\,i^{2s}(m2^j)^{-2s}\leqslant$$

$$\leqslant\sum_{j=0}^{\infty}\frac{\pi^{2\alpha+1}C^2}{4^\alpha}(m2^j)^{-2\alpha-2s}=\frac{\pi^{2\alpha+1}C^2}{4^\alpha(1-4^{-\alpha-s})\,m^{2\alpha+2s}},$$

что и требовалось доказать.

Теорема 3. *Пусть фиксированы* $\alpha\in(0,1]$, $C>0$ *и целое* $s\geqslant 0$. *Тогда для оценки с тригонометрическим рядом* f_n, *имеющей параметр* m, *выполнено неравенство*

$$\sup_{f\in W(s,\alpha,C)}E\left(\int|f_n-f|\right)\leqslant\sqrt{\frac{2m+1}{n}}+C\sqrt{2\pi\gamma}\left(\frac{1}{m+1}\right)^{s+\alpha},$$

где

$$\gamma=\frac{\pi^{2\alpha+1}}{4^\alpha-4^{-s}}$$

— *постоянная из леммы* 12. *В частности, если* $m\sim(C\sqrt{\pi\gamma}2(\alpha+s))^{2/(1+2(\alpha+s))}n^{1/(1+2(\alpha+s))}$, *то*

$$\limsup_{n\to\infty}\sup_{f\in W(s,\alpha,C)}E\left(\int|f_n-f|\right)n^{(\alpha+s)/(1+2(\alpha+s))}\leqslant$$

$$\leqslant(C\sqrt{\pi\gamma}2(\alpha+s))^{1/(1+2)(\alpha+s))}\left(\frac{1+2(\alpha+s)}{2(\alpha+s)}\right)\sqrt{2}.$$

Доказательство. Первое неравенство следует из лемм 8, 10 и 12. Его правая часть представима в виде суммы $u\sqrt{m}+vm^{-(\alpha+s)}$ и членов более высокого порядка малости, если $\lim\limits_{n\to\infty}m=\infty$. Здесь $u=\sqrt{2/n}$, $v=C\sqrt{2\pi\gamma}$. Минимум правой части по m достигается, если m — решение уравнения

$$\frac{u}{2\sqrt{m}}-\frac{v(a+s)}{m^{\alpha+s+1}}=0.$$

Отсюда следует, что желательно выбирать значения $m \sim (2v(\alpha + s)/u)^{2/(1+2(\alpha+s))}$. Подстановка таких m в исходное неравенство дает требуемый результат.

Замечание. Для важного класса $W(0, 1, C)$ получаем

$$\limsup_{n\to\infty} \sup_{f \in W(0,1,C)} n^{1/3} E\left(\int |f_n - f|\right) \leqslant \left(2C\frac{\pi^2}{\sqrt{3}}\right)^{1/3}\sqrt{2}$$

при $m \sim (4C^2\pi^4/3n)^{1/3}$. Правую часть этого неравенства следует сравнить с нижней границей теоремы 4.7, заменяя там C на $2\pi C$ (так как теорема 3 верна для плотностей на $[-\pi, \pi]$, а не на $[0, 1]$). Заметим, что верхняя граница теоремы 3 в общем случае зависит от C и n так же, как минимаксная нижняя граница теоремы 4.6. Чтобы получить аналогичный результат для ядерной оценки, мы должны были менять ядро в зависимости от s и α. Здесь, напротив, достаточно лишь настраивать параметр сглаживания m.

Из теоремы 3 следует, что оценка с тригонометрическим рядом в принципе может иметь любую скорость сходимости вплоть до $1/\sqrt{n}$ — эта скорость зависит от гладкости f. Классы Бретаньоля — Юбер определяются величинами $\int |f^{(s)}|$ и $\int \sqrt{f}$. Из леммы 8 видно, что для получения на этих классах скоростей сходимости, сравнимых в асимптотике со скоростями убывания минимаксных нижних границ, достаточно при некоторых условиях гладкости на f оценить сверху величину $\int |S_m(f) - f$ выражением, пропорциональным $\int |f^{(s)}|/m^s$. К сожалению, в полной мере достичь этого не удается из-за наличия дополнительного множителя $\log m$. Такого рода верхнюю границу можно быстро получить следующим образом. Пусть AC_s — класс всех плотностей на $[-\pi, \pi]$, имеющих $s-1$ абсолютно непрерывных производных (на действительной оси) и таких, что $\int |f^{(s)}| < \infty$. Пусть T_m — пространство всех тригонометрических полиномов степени m, т. е. линейных функций от $\cos ix$ и $\sin ix$, $0 \leqslant i \leqslant m$. Тогда по *второй теореме Джексона* (см., например, Butzer, Nessel (1971, с. 97—99))

$$\inf_{t_m \in T_m} \int |f - t_m| \leqslant \left(\frac{36}{m}\right)^s \int |f^{(s)}|, \quad f \in AC_s.$$

Отсюда сразу следует верхняя граница для смещения

$$\int |S_m(f) - f| \leqslant \int |S_m(f) - t_m| + \int |t_m - f| \leqslant \qquad (t_m \in T_m)$$

$$\leqslant \int |D_m| \int |f - t_m| + \int |f - t_m| \leqslant \quad \text{(неравенство Юнга)}$$

$$\leqslant \left(3 + \tfrac{4}{\pi^2} \log m\right) \int |f - t_m|, \quad \text{(лемма 4)}$$

и, выбирая наилучшее t_m из класса T_m, приходим к следующему результату.

Теорема 4. *Пусть f — плотность на $[-\pi, \pi]$ из класса AC_s, где $s \geqslant 0$ — фиксированное целое число. Тогда для оценки с тригонометрическим рядом, имеющей параметр m, справедливо неравенство*

$$E\left(\int |f_n - f|\right) \leqslant \sqrt{\frac{2m+1}{n}} + \left(3 + \tfrac{4}{\pi^2} \log m\right) \left(\frac{36}{m}\right)^s \int |f^{(s)}|.$$

Первое слагаемое в правой части этого неравенства можно заменить на $(\sqrt{m/\pi n})\left(\int \sqrt{f} + o(1)\right)$.

Таким образом, возникает вопрос о том, действительно ли существен дополнительный множитель $\log m$. Мы знаем, что при $s = 0$ это так (см., например, теорему 1). По-видимому, и при $s > 0$ этот множитель нельзя уменьшить (см., например, Butzer, Nessel (1971, с. 108) или Quade (1937)). Именно это послужило для нескольких исследователей препятствием к изучению поведения оценки с тригонометрическим рядом в терминах характеристики $\int |f^{(s)}|$. Так, в Wahba (1975) рассматриваются *классы Соболева*, т. е. пространства плотностей f, имеющих $s - 1$ абсолютно непрерывных производных и таких, что $\int |f^{(s)}|^p \leqslant M < \infty$, где $p > 1$ — еще один параметр, определяющий класс Соболева. В своем знаменитом исследовании она сравнивает поведение нескольких оценок плотности на этих классах. Поведение оценки с тригонометрическим рядом при любых $p > 1$ легко проанализировать с помощью неравенства Хаусдорфа — Юнга (см. Бари (1961, с. 211)), связывающего q-ю норму коэффициентов Фурье с p-й нормой функции при $1/p + 1/q = 1$. При $p = 2$ оно совпадает с равенством Бесселя, и на этом частном случае мы покажем, как оно применяется.

Лемма 13. *Пусть f — абсолютно непрерывная плотность с носителем, содержащимся в $[0, 1]$, и пусть $\int (f')^2 < \infty$. Тогда для оценки с тригонометрическим рядом, имеющей параметр m, справедливо неравенство*

$$E\left(\int |f_n - f|\right) \leqslant \sqrt{\frac{2m+1}{n}} + \gamma \sqrt{\int (f')^2}\, \frac{1}{m+1},$$

где $\gamma = \pi^{3/2}/\sqrt{3}$.

Замечание. Верхняя граница леммы 13 убывает со скоростью $O(n^{-1/3})$, если m возрастает как $n^{1/3}$. Обобщая приведенные ниже рассуждения по типу обобщения леммы 11 в лемме 12, можно рассмотреть все классы Соболева при $p = 2$ и получить границы в терминах величин $\int (f^{(s)})^2$. Сравнение этого результата с теоремой 4 сразу показывает, что множитель $\log m$ действительно устранен за счет введения дополнительного условия $f^{(s)} \in L_2[-\pi, \pi]$.

Доказательство леммы 13. Воспользуемся леммами 8 и 10. Кроме того, применим рассуждения из леммы 11 с некоторыми изменениями. Иначе, чем в лемме 11, оценим величину

$$\int_{-\pi}^{\pi} (f(x+h) - f(x-h))^2 \, dx.$$

Продолжая f снова до периодической функции, мы видим, что эта величина равна

$$\int_{-\pi}^{\pi} \left(\int_{x-h}^{x+h} f' \right)^2 dx \leqslant \int_{-\pi}^{\pi} 2h \left(\int_{x-h}^{x+h} (f')^2 \right) dx = (2h)^2 \int (f')^2.$$

Таким образом, можно повторить остаток доказательства леммы 11, заменяя там формально α на 1 и $2\pi C^2$ на $\int (f')^2$. В частности,

$$\sum_{i=2m-1}^{\infty} a_i^2 \leqslant \frac{\pi^2}{3} \int (f')^2 \frac{1}{2\pi m^2} = \frac{\pi^2}{6m^2} \int (f')^2.$$

Лемма 13 доказана.

6. Оценка с рядом Эрмита

Оценка с рядом Эрмита, несомненно, является наиболее популярной проекционной оценкой плотности на действительной оси (Schwartz (1967), Bosq, Bleuez (1978), Bleuez, Bosq (1979), Walter (1977), Greblicki (1981)). В этом параграфе мы вкратце исследуем основные свойства этой оценки. Почти все из приводимых ниже результатов остаются в силе для оценки с рядом Лагерра на $[0, \infty)$.

Теорема 5 (несостоятельность оценки с рядом Эрмита). *Ортонормальная система Эрмита не является базисом в L_p при $p \in [1, 4/3]$ и $p \in [4, \infty)$. Если f_n — оценка с рядом Эрмита, имеющая параметр m_n и частичные суммы $S_{m_n}(f)$, и если*

$$\lim_{n \to \infty} m_n = \infty, \quad |m_n - m_{n-1}| \leqslant 1, \quad \forall n,$$

то найдется плотность f, такая, что

$$\limsup_{n\to\infty} \int |S_{m_n}(f) - f| = \infty$$

и, следовательно,

$$\limsup_{n\to\infty} E\left(\int |f_n - f|\right) = \infty.$$

От условия на L_p здесь нельзя легко избавиться, что явствует из следующего классического результата анализа.

Лемма 14 (Askey, Wainger (1965); см. также Muckenhoupt (1970)). *При любых $f \in L_p$, $p \in (4/3, 4)$ имеем*

$$\lim_{m\to\infty} \int |S_m(f) - f|^p = 0.$$

При любых $p \notin (4/3, 4)$ существует функция $f \in L_p$, такая, что

$$\limsup_{m\to\infty} \int |S_m(f) - f|^p > 0.$$

Лемма 15 (границы Сковгарда; см. Askey, Wainger (1965, с. 700)). *Пусть p_i есть i-я функция ортонормальной системы Эрмита. Тогда существуют положительные постоянные C_1, C_2, C_3, C_4, не зависящие от i и x, такие, что*

$$|p_i(x)| \leqslant \begin{cases} C_1 i^{-1/12} & \text{при любых } x, i, \\ C_2 i^{-1/4} & \text{при любых } |x| \leqslant \sqrt{4i},\ \left||x| - \sqrt{2i}\right| \geqslant (2i)^{-1/6}, \\ C_2 \exp(-C_4 x^2) & \text{при любых } |x| \geqslant \sqrt{4i}. \end{cases}$$

Доказательство теоремы 5. Так как S_m — линейный оператор, то для любых функций f_1, f_2 имеем

$$\int |S_m(f_1 + f_2) - (f_1 + f_2)| \leqslant \int |S_m(f_1) - f_1| + \int |S_m(f_2) - f_2|.$$

Полагая $f_1 = (f)_+$, $f_2 = (f)_-$, где $f \in L_1$ — некоторая функция, легко видеть, что достаточно доказать соотношение $\limsup_{n\to\infty} \int |S_{m_n}(f) - f| = \infty$, где $f \in L_1$ — некоторая функция. В силу принципа равномерной ограниченности [1]) (см., например,

---

[1]) Теорема Банаха—Штейнгауза (см., например, Люстерник Л. А., Соболев В. И. Краткий курс функционального анализа. — М.: Высшая школа, 1982, с. 116). — *Прим. перев.*

Butzer, Nessel (1971, с. 18—19)) доказательство будет завершено, если показать, что

$$\sup_m \sup_{f \in L_1} \frac{\int |S_m(f)|}{\int |f|} = \infty.$$

Будем теперь рассуждать от противного. Если предположить, что существует конечное M, такое, что $\int |S_m(f)| \leqslant M \int |f|$ при любых m, $f \in L_1$, то

$$\int |S_m(f) - S_{m-1}(f)| = \int |a_m p_m| \leqslant 2M \int |f|.$$

Но $\int |a_m p_m| = \left| \int f p_m \right| \int |p_m|$. Мы знаем, что для некоторой функции $f \in L_1$ выполнено

$$\left| \int f p_m \right| \geqslant \tfrac{1}{2} \int |f| \operatorname{ess\,sup} |p_m|,$$

так что

$$\operatorname{ess\,sup} |p_m| \int |p_m| \leqslant 4M.$$

Ясно, что последнее невозможно, поскольку $\operatorname{ess\,sup} |p_m| \geqslant cm^{-1/12}$ при достаточно больших m и некотором $c > 0$ (точная верхняя грань достигается приблизительно при $x = \sqrt{2m}$) и $\int |p_m| \geqslant cm^{1/4}$, $m \geqslant 1$, где $c > 0$ — некоторая другая постоянная. (Эти соотношения следуют не из границ Сковгарда, поскольку последние неточны, а из точных границ (Askey, Wainger (1965)).) В этом заключается требуемое противоречие. Заменим теперь m на m_n и потребуем, чтобы последовательность m_n стремилась к бесконечности и модуль разности $|m_n - m_{n-1}|$ не превосходил 1 при каждом n. Тогда $\limsup_{n \to \infty} \int |S_{m_n}(f) - f| = \infty$ для некоторой функции $f \in L_1$.

Первое утверждение теоремы 5 здесь не доказывается.

Лемма 16. *Пусть f_n — оценка с рядом Эрмита, построенная по независимой выборке $X_1, \ldots, X_n$ из совокупности с плотностью f. Тогда*

$$\lim_{a \to \infty} \int |f_n(x + a, X_1 + a, \ldots, X_n + a)| = 0$$

почти наверное при любых $n \geqslant 1$ и любых параметрах $m \geqslant 1$ и

$$\lim_{a \to \infty} E\left(\int |f_n(x + a, X_1 + a, \ldots, X_n + a)| \right) = 0.$$

Доказательство.

$$\int |f_n(x+a,\ X_1+a,\ \ldots,\ X_n+a)| \leqslant$$

$$\leqslant \sum_{i=0}^{m} \left| \frac{1}{n} \sum_{j=1}^{n} p_i(X_j+a) \right| \int |p_i(x+a)| \leqslant$$

$$\leqslant \frac{1}{n} \sum_{j=1}^{n} \sum_{i=0}^{m} |p_i(X_j+a)| \int |p_i| \leqslant (m+1) \sup_i \int |p_i| \sup_{i,\ j} |p_i(X_j+a)|.$$

Но величина $\sup_i \int |p_i|$ конечна в силу границ Сковгарда (лемма 15) и $\sup_{i,\ j} |p_i(X_j+a)| \to 0$ при $a \to \infty$ тоже в силу леммы 15. Лемма 16 доказана.

То, что оценка с рядом Эрмита, возможно, не является трансляционно-инвариантной, следует из леммы 3 о проекционных оценках на действительной оси. Свойство, указанное в лемме 16, довольно неожиданно, поскольку, казалось бы, нельзя ничего гарантировать относительно величины $\int |f_n|$, не говоря уже о $\int f_n$. Учитывая еще и теорему 5, заключаем, что оценка с рядом Эрмита, по-видимому, плохо приспособлена для оценивания плотностей общего вида на действительной оси.

Оценка с рядом Эрмита состоятельна в L_p при любых $p \in (4/3, 4)$, $f \in L_p$. В лемме 17 это будет показано при $p = 2$. Доказать сходимость в L_1 затруднительно по той причине, что, так как интегрирование ведется по R, нельзя оценить сверху $\int |S_m(f) - f|$ с помощью неравенства Гёльдера через некоторую L_p-норму функции $S_m(f) - f$. Можно близко подойти к сходимости в L_1, используя другие приемы. Muckenhoupt (1970) показал, что при фиксированных $b > 0$, $B \geqslant b + 1/3$ выполнено соотношение

$$\int \frac{|S_m(f) - f|}{|1 + |x||^b} \to 0, \quad m \to \infty,$$

где предполагается, что $\int f \log_+ f < \infty$, $\int f |x|^B \log_+ f < \infty$, $\int f |x|^{B+2} < \infty$. К сожалению, здесь нельзя положить $b = 0$, $B = 1/3$.

Лемма 17 (сходимость в L_2 оценки с рядом Эрмита). *Пусть* f_n *— оценка с рядом Эрмита, имеющая параметр* m, *и пусть*

$$\lim_{n\to\infty} m = \infty, \quad \lim_{n\to\infty} \frac{m}{n^2} = 0.$$

Тогда $E\left(\int (f_n - f)^2\right) \to 0$ *при* $n \to \infty$ *для любых плотностей* f *из* L_2.

Доказательство. В силу леммы 2 и равенства Бесселя (см. также лемму 14) видно, что достаточно доказать сходимость к 0 дисперсионного члена $E\left(\int (f_n - E(f_n))^2\right)$. Она следует из того, что

$$\int f \sum_{i=0}^{m} p_i^2 = o(n).$$

При $|x| \geqslant \sqrt{4m}$ в силу границ Сковгарда имеем

$$\sum_{i=0}^{m} p_i^2 \leqslant (m+1) C_3^2 \exp(-2C_4 x^2) \leqslant (m+1) C_3^2 \exp(-8C_4 m),$$

так что

$$\int\limits_{|x| \geqslant \sqrt{4m}} f \sum_{i=0}^{m} p_i^2 = O\left(e^{-C_4 m}\right) = o(n).$$

Далее, введем

$$q_m(x) = \sum_{i=1}^{m} \left(C_1 i^{-1/12}\right)^2 I_{[|x| \leqslant \sqrt{4m},\ |\,|x| - \sqrt{2i}\,| \leqslant (2i)^{-1/6}]}.$$

Тогда в силу леммы 15

$$\int\limits_{|x| \leqslant \sqrt{4m}} f \sum_{i=1}^{m} p_i^2 \leqslant \int f q_m + \sum_{i=1}^{m} \left(C_2 i^{-1/4}\right)^2 + \sum_{i=1}^{m} C_3^2 \exp(-8C_4 i) = I_1 + I_2 + I_3.$$

Ясно, что $I_3 = O(1)$. Кроме того, $I_2 = O\left(\sqrt{m}\right) = o(n)$. Наконец, нетрудно показать, что индикаторная функция в определении $q_m(x)$ отлична от нуля для почти $C_5 m^{1/3}$ индексов i равномерно по всем $|x| \leqslant \sqrt{4m}$, где $C_5 > 0$ — заданная постоянная. Таким образом, $q_m(x) = O\left(m^{5/18}\right) = o(n)$ равномерно по таким x, и лемма доказана.

Замечание. Лемма 17 была получена в работе Schwartz (1967) при более сильном условии $m = o(n)$ и в работе Greblicki (1981) при условии $m = o(n^{6/5})$. Необходимость условий на m, указанных в лемме 17, для сходимости при любых $f \in L_2$ получена Bleuez, Bosq (1979).

Замечание. До сих пор мы не рассматривали вопрос о поточечной сходимости оценки с рядом Эрмита. На основании теоремы Карлесона — Ханта в Muckenhoupt (1970) доказано, что $S_m(f) \to f$ при почти всех x, если $\int f (\log_+ f)^2 < \infty$. Этот результат вместе с границами Сковгарда можно использовать для доказательства поточечной сходимости данной оценки.

Замечание. Оценка с рядом Лагерра очень похожа по своему поведению на оценку с рядом Эрмита. Важные сведения о смещении оценки с рядом Лагерра имеются в работах Askey, Wainger (1965) и Muckenhoupt (1970a, *b*, c).

7. Оценка с рядом Лежандра

Предложение использовать при оценивании плотности оценку с рядом Лежандра принадлежит Crain (1974) и Hall (1982) (см. также Viollaz (1980)). Эта оценка сходится и расходится при условиях, во многих отношениях похожих на соответствующие условия для оценки с рядом Эрмита. Последнее становится очевидным после сравнения леммы 14 со следующим результатом.

Лемма 18. *Полиномы Лежандра образуют базис в* L_p $[-1, 1]$ *тогда и только тогда, когда* $3/4 < p < 4$. *Для* $f \in L_p$ $[-1, 1]$, $3/4 < p < 4$, *имеем*

$$\int |S_m(f) - f|^p \to 0$$

при $m \to \infty$. *Для каждого* $p \notin (3/4, 4)$ *существует функция* f *из* L_p $[-1, 1]$, *такая, что*

$$\limsup_{m \to \infty} \int |S_m(f) - f|^p > 0.$$

Лемма 18 принадлежит Pollard (1947) и Newman, Rudin (1952). Она была распространена на полиномы Якоби в работах Pollard (1948, 1949); см. также Muckenhoupt (1969). Так как мы в основном интересуемся метрикой L_1, то, по-видимому, полезно будет привести пример плотности, которую нельзя оценить посредством оценки с рядом Лежандра.

Теорема 6 (несостоятельность оценки с рядом Лежандра). *Рассмотрим плотность*

$$f(x) = \frac{7}{2^{15/4}}(1-x)^{-3/4}, \quad |x| < 1.$$

Тогда плотность f принадлежит L_p $[-1, 1]$ при любых $p \in [1, 4/3)$,

$$\liminf_{m\to\infty}\left(\int |S_m(f) - f| + \int |S_{m+1}(f) - f|\right) > 0,$$

$$\limsup_{m\to\infty} \int |S_m(f) - f| > 0,$$

$$\inf_{n,\,m}\left(E\left(\int |f_{n,m} - f|\right) + E\left(\int |f_{n,m+1} - f|\right)\right) \geqslant 0,$$

где $f_{n,m}$ — оценка с рядом Лежандра, имеющая параметр m и объем выборки n.

Доказательство. Легко проверить, что $f \in L_p$, $1 \leqslant p < 4/3$. Остальные утверждения теоремы следуют из леммы 1 и соотношения

$$\int |S_m(f) - f| + \int |S_{m-1}(f) - f| \geqslant \int |S_m(f) - S_{m-1}(f)| = |a_m| \int |p_m|,$$

где $a_m \geqslant A_1 + o(1)$, $\int |p_m| \geqslant A_2 + o(1)$ при $m \to \infty$ и A_1, A_2 — положительные постоянные (см. Szegö (1975, с. 256 и с. 173 соответственно)).

Замечание. В теореме 6 мы подошли близко к доказательству того, что

$$\inf_{n,\,m} E\left(\int |f_{n,m} - f|\right) > 0$$

для указанной плотности f. В этом доказательстве нужны гораздо более сложные рассуждения.

Теорема 7 (состоятельность оценки с рядом Лежандра). *Если f — плотность на $[-1, 1]$, $f \in L_p$ $[-1, 1]$ при некотором $p > 4/3$,*

$$\int_{-1}^{1} (1-x^2)^{-1/2} f(x)\,dx < \infty,$$

$$\lim_{n\to\infty} m = \infty, \quad \lim_{n\to\infty} \frac{m}{n} = 0,$$

то

$$\lim_{n\to\infty} E\left(\int |f_n - f|\right) = 0,$$

где f_n — *оценка с рядом Лежандра, имеющая параметр* m.

Доказательство. Если $p \geqslant 4$, то положим $p^* = 2$, в противном случае $p^* = p$. Определим q равенством $1/p^* + 1/q = 1$. В силу леммы 18 и неравенства Гёльдера

$$\int |S_m(f) - f| \leqslant 2^{1/q}\left(\int |S_m(f) - f|^{p^*}\right)^{1/p^*} \to 0$$

при $m \to \infty$. Учитывая лемму 1, неравенство Коши — Шварца и лемму 2, получим

$$E\left(\int |f_n - E(f_n)|\right) \leqslant \int \sqrt{E((f_n - E(f_n))^2)} \leqslant$$

$$\leqslant \sqrt{2\int E((f_n - E(f_n))^2)} \leqslant \sqrt{2}\sqrt{\frac{1}{n}\int f \sum_{i=0}^{m} p_i^2}.$$

По первой теореме Стилтьеса (Sansone (1977, с. 199))

$$|p_i| \leqslant \sqrt{\frac{2i+1}{2}}\, 4\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{i}}(1-x^2)^{-1/4} \leqslant$$

$$\leqslant \sqrt{1+\frac{1}{i}}\, 4\sqrt{\frac{2}{\pi}}(1-x^2)^{-1/4} \leqslant \frac{8}{\sqrt{\pi}}(1-x^2)^{-1/4}, \quad i \geqslant 1.$$

Так как $p_0 = 1/\sqrt{2}$, эта граница верна и при $i = 0$. Следовательно,

$$\sum_{i=0}^{m} p_i^2 \leqslant (m+1)\frac{64}{\pi}(1-x^2)^{-1/2},$$

так что $E\left(\int |f_n - E(f_n)|\right) \to 0$.

Ни одно из условий сходимости в теореме 7 нельзя полностью исключить. Холл (Hall (1982)) приводит убедительные аргументы в пользу оценки с рядом Лежандра, в частности то, что для нее достаточно небольшого числа членов разложения, чтобы обеспечить хорошую скорость сходимости в L_2 $[-1, 1]$ для определенных классов плотностей. Вопрос о скорости сходимости здесь не рассматривается. Отметим еще, что оценка с рядом Лежандра не является трансляционно-инвариантной, но тем не менее $\int f_n = 1$ при любых n, m.

8. Сингулярно-интегральные оценки

Сингулярно-интегральная оценка плотности f с *ядром* K_m определяется соотношением

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} K_m(x - X_j).$$

Частными случаями оценок такого вида являются ядерная оценка и оценка с тригонометрическим рядом (для которой K_m — ядро Дирихле). Для большей общности можно было бы писать $K_m(x, X_j)$ (при некоторых условиях на K_m такие оценки называются оценками типа дельта-функции Дирака, см. Walter, Blum (1979)), но требование трансляционной инвариантности заставляет рассматривать только случай, когда слагаемые имеют вид $K_m(x - X_j)$. Мы будем заниматься оцениванием только плотностей на $[-\pi, \pi]$, накладывая на K_m следующие ограничения:

$$K_m(x) = K_m(-x); \quad \int_{-\pi}^{\pi} K_m = 1; \quad K_m \text{ — периодическая функция с периодом } 2\pi; \quad \int_{-\pi}^{\pi} |K_m| < \infty. \quad (1)$$

Ввиду периодичности K_m интеграл от K_m по действительной оси равен ∞, так что ядерная оценка уже не является частным случаем сингулярно-интегральной оценки с ядром, удовлетворяющим условиям (1).

Исследование сингулярно-интегральных оценок проведем обычным образом. Сначала покажем, что можно выбрать такие последовательности ядер, что сингулярно-интегральная оценка состоятельна при любых $f \in L_1[-\pi, \pi]$. Можно даже выбрать все ядра K_m неотрицательными, так что f_n будет плотностью на $[-\pi, \pi]$ при любых n, m. Затем исследуем скорость сходимости оценок и заметим, в частности, что если $K_m \geqslant 0$, то так же, как в случае ядерной оценки, средняя L_1-ошибка не может стремиться к 0 быстрее, чем $n^{-2/5}$.

Прежде всего дадим несколько определений. Определим *сингулярный интеграл* $S_m(f)$ (или $S_m(f, x)$) соотношением

$$S_m(f) = \int_{-\pi}^{\pi} f(x - u) K_m(u)\, du,$$

где функция f продолжена на R по периодичности. Остается в силе неравенство Юнга

$$\int |S_m(f)| \leqslant \int |f| \int |K_m|.$$

Будем говорить, что K_m — *дельтаобразная последовательность* [1]), если выполнено условие (1) и

$$\sup_m \int |K_m| \leqslant C < \infty, \tag{2}$$

$$\lim_{m\to\infty} \int\limits_{\delta \leqslant |u| \leqslant \pi} |K_m(u)|\, du = 0 \tag{3}$$

при любых $\delta > 0$. Последовательность K_m называется *сильно дельтаобразной последовательностью* [2]), если вместо (3) выполнено условие

$$\lim_{m\to\infty} \sup_{\delta \leqslant |u| \leqslant \pi} |K_m(u)| = 0 \tag{4}$$

при любых $\delta > 0$.

Лемма 19. *Для любых дельтаобразных последовательностей и любых* $f \in L_1\,[-\pi, \pi]$ *имеем*

$$\int |S_m(f) - f| \to 0$$

при $m \to \infty$.

Доказательство.

$$S_m(f) - f = \int\limits_{-\pi}^{\pi} (f(x-u) - f(x))\, K_m(u)\, du.$$

Пусть $g(u) = \int\limits_{-\pi}^{\pi} |f(x-u) - f(x)|\, dx$. По неравенству Юнга

$$\int |S_m(f) - f| \leqslant \int\limits_{-\pi}^{\pi} g(u)\, |K_m(u)|\, du \leqslant$$

$$\leqslant \int\limits_{|u| \leqslant \delta} \sup_{v \leqslant \delta} g(v)\, |K_m(u)|\, du +$$

$$+ \int\limits_{\delta \leqslant |u| \leqslant \pi} |K_m(u)|\, du \cdot 2 \int |f| \leqslant C \sup_{v \leqslant \delta} g(v) + o(1).$$

---

[1]) В оригинале approximate identity (этот термин дается со ссылкой на Butzer, Nessel (1971, с. 31)). — *Прим. перев.*

[2]) В оригинале strong approximate identity. — *Прим. перев.*

Но мы знаем, что $\lim\limits_{v \downarrow 0} g(v) = 0$ при любых $f \in L_1[-\pi, \pi]$. Лемма 19 доказана.

Теорема 8 (состоятельность сингулярно-интегральных оценок). *Пусть f_n — сингулярно-интегральная оценка с параметром m, таким, что $\lim\limits_{n\to\infty} m = \infty$. Предположим, что последовательность ядер K_m — дельтаобразная и что $\int K_m^2 = o(n)$ при $n \to \infty$. Тогда*

$$E\left(\int |f_n - f|\right) \to 0$$

при $n \to \infty$ для любых плотностей f на $[-\pi, \pi]$.

Доказательство. В силу леммы 19 достаточно доказать, что $E\left(\int |f_n - E(f_n)|\right)$ при $n \to \infty$. (Заметим, что $E(f_n) = S_m(f)$.) Применяя дважды неравенство Коши — Шварца, получим

$$E\left(\int |f_n - E(f_n)|\right) =$$

$$= \int E\left(\left|\frac{1}{n}\sum_{j=1}^{n}(K_m(x - X_j) - E(K_m(x - X_j)))\right|\right) \leqslant$$

$$\leqslant \int \sqrt{\frac{1}{n}\operatorname{Var}(K_m(x - X_1))} \leqslant \sqrt{2\pi}\sqrt{\frac{1}{n}\int \operatorname{Var}(K_m(x - X_1))} \leqslant$$

$$\leqslant \sqrt{\frac{2\pi}{n}}\sqrt{\int E(K_m^2(x - X_1))} = \sqrt{\frac{2\pi}{n}}\sqrt{\int K_m^2},$$

где все интегралы берутся по $[-\pi, \pi]$. Теорема 8 доказана.

Ранее мы уже встречались с дельтаобразными последовательностями. Например, *ядра Фейера*

$$K_m(u) = F_m(u) = \frac{1}{2\pi(m+1)}\left(\frac{\sin((m+1)u/2)}{\sin(u/2)}\right)^2$$

неотрицательны и образуют дельтаобразную последовательность. Чтобы убедиться в этом, вспомним, что в силу леммы 4 $\int F_m = 1$, $\int |F_m| = 1$ и $\int\limits_{\delta \leqslant |u| \leqslant \pi} F_m(u)\,du \leqslant \int\limits_{\delta \leqslant |u| \leqslant \pi} \pi(2mu^2)^{-1}\,du \leqslant \pi \times$ $\times (2m\delta)^{-1} \to 0$ при $m \to \infty$ для любых $\delta > 0$.

Последовательность D_m, напротив, не является дельтаобразной ввиду теоремы 1 или леммы 4. Приведем краткий перечень

дельтаобразных последовательностей. Все они, за исключением двух, тоже неотрицательны.

(i) *Ядро Рогозинского* (Butzer, Nessel (1971, с. 56))

$$\frac{1}{2}\left(D_m\left(x+\frac{\pi}{2m+1}\right)+D_m\left(x-\frac{\pi}{2m+1}\right)\right).$$

(ii) *Ядро Джексона* (Butzer, Nessel (1971, с. 60—61))

$$\frac{3}{2\pi m\,(2m^2+1)}\left(\frac{\sin\,(mx/2)}{\sin\,(x/2)}\right)^4.$$

(iii) *Ядро Фейера — Коровкина* (Butzer, Nessel (1971, с. 79—80))

$$\frac{\sin^2\,(\pi/(m+2))}{\pi\,(m+2)}\left(\frac{\cos\,((m+2)\,x/2)}{\cos\,(\pi/(m+2))-\cos x}\right)^2.$$

(iv) *Ядро Валле-Пуссена* (Butzer, Nessel (1971, с. 112))

$$\frac{m!^2}{2\pi\,(2m)!}\left(2\cos\left(\frac{x}{2}\right)\right)^{2m}.$$

(v) *Ядро Джексона — Валле-Пуссена* (Butzer, Nessel (1971, с. 131)

$$\frac{2+\cos x}{4\pi m^3}\left(\frac{\sin\,(mx/2)}{\sin\,(x/2)}\right)^4.$$

(vi) *Второе ядро Валле-Пуссена* (Butzer, Nessel (1971, с. 108))

$$(1+2\cos mu)\,F_{m-1}\,(u).$$

Сингулярно-интегральные оценки Фейера, Рогозинского и Фейера — Коровкина исследованы и сравнены друг с другом в L_2 в работе Hall (1983). Некоторые свойства сингулярно-интегральной оценки Фейера получены в работе Krzyzak, Pawlak (1982). При определенных условиях непрерывности на f и при оптимальном выборе m скорости сходимости средней L_2-ошибки ядерной оценки и сингулярно-интегральной оценки Рогозинского совпадают, но асимптотическая константа для последней меньше (Hall (1983)). Это неожиданное замечание вдохновляет нас на исследование сингулярно-интегральных оценок и побуждает провести анализ их скоростей сходимости в L_1. Прежде чем приступить к этому, мы хотели бы указать на тесную связь между сингулярно-интегральными оценками и оценкой с тригонометрическим рядом. Сингулярно-интегральные оценки часто можно записать в виде

$$f_n(x)=\frac{1}{2\pi}+\sum_{i=1}^{\infty}\left(a_{n,\,2i-1}\,\frac{\cos ix}{\sqrt{\pi}}+a_{n,\,2i}\,\frac{\sin ix}{\sqrt{\pi}}\right)\lambda_{mi}\sqrt{\pi},$$

где a_{ni} — те же самые, что и для оценки с тригонометрическим рядом, т. е. это обычные оценки коэффициентов Фурье a_i функции f по тригонометрической системе. Такое представление сингулярно-интегральных оценок следует из разложения

$$K_m(u) = \frac{1}{2\pi} + \sum_{i=1}^{\infty} \lambda_{mi} \frac{\cos iu}{\sqrt{\pi}},$$

справедливого для любых четных ядер, ряд Фурье которых сходится, и формулы $\cos(i(x-y)) = \cos ix \cos iy + \sin ix \sin iy$. Таким образом, величины λ_{mi} в определении f_n являются коэффициентами Фурье ядра K_m.

Данное представление для f_n особо интересно с практической точки зрения в случае, когда коэффициенты λ_{mi} равны нулю при всех достаточно больших i. При этом удобно вычислять и запоминать значения a_{ni} вместо X_j.

В работах Watson (1969) и Rosenblatt (1971) предложены сглаженные проекционные оценки вида $\sum_{i=0}^{\infty} \lambda_{mi} a_{ni} p_i$, где веса λ_{mi} играют роль сглаживателей. Ясно, что при $\lambda_{mi} = 1$, $i \leqslant m$, $\lambda_{mi} = 0$, $i \geqslant m$, никакого сглаживания нет, и мы получаем снова проекционные оценки. Сглаживание полезно во многих отношениях: грубо говоря, при правильном сглаживании получаются оценки, состоятельные в L_1 для любых плотностей. Примером являются сингулярно-интегральные оценки с дельтаобразными последовательностями ядер. Однако мы теряем в точной настройке, т. е. для определенных классов плотностей скорость сходимости сглаженной оценки ниже, чем скорость сходимости исходной проекционной оценки. Это тоже будет показано в данном параграфе. Интересно, что Ватсон (Watson (1969)) нашел наилучшие коэффициенты λ_{mi} при фиксированном n в смысле минимума критерия $E\left(\int (f_n - f)^2\right)$. Они имеют вид

$$\lambda_{mi} = \frac{a_i^2}{a_i^2 + (1/n)\left(\int f p_i^2 - a_i^2\right)}$$

(что проверяется в две строчки). К сожалению, от этого мало пользы, поскольку величины a_i не известны и речь идет об L_2-ошибке. В литературе имеются различные предложения относительно выбора коэффициентов λ_{mi} в общем случае и для конкретных ортонормальных систем; см., например, Whittle (1958), Fellner (1974), Brunk (1977, 1978), Kronmal, Tarter (1968) и Wahba (1978). Зачастую дается различное обоснование этих предложений. Так, в работах Brunk (1977, 1978) применяется

байесовский подход, использующий априорную информацию, что λ_{mi} имеют вид $c_i/(c_i + 1/n)$, $1 \leqslant i \leqslant n$. В Wahba (1978) рассматриваются оценки, для которых $\lambda_{mi} = (1 + ci^p)^{-1}$, $1 \leqslant i \leqslant n$, где c, $p > 0$ — постоянные. Формула Ватсона сразу наводит на мысль об адаптивных методах выбора коэффициентов λ_{mi}. Адаптивные варианты проекционных оценок и их сглаженных версий рассматриваются, например, в работах Kronmal, Tarter (1968), Tarter, Kronmal (1976), Crain (1973), Asselin de Beauville (1978) и Wahba (1977, 1978).

Вернемся теперь к сингулярно-интегральным оценкам, т. е. сглаженным оценкам с тригонометрическим рядом. Рассматривая ядро Дирихле

$$D_m(x) = \frac{1}{2\pi} + \sum_{i=1}^{m} \frac{1}{\pi} \cos ix,$$

заметим сразу, что

$$\lambda_{mi} = \begin{cases} 1/\sqrt{\pi}, & 1 \leqslant i \leqslant m, \\ 0, & i > m. \end{cases}$$

Приведем еще несколько примеров.

(i) *Ядро Фейера*

$$\lambda_{mi}\sqrt{\pi} = \begin{cases} 1 - i/(m+1), & 1 \leqslant i \leqslant m, \\ 0, & i > m. \end{cases}$$

(ii) *Ядро Рогозинского*

$$\lambda_{mi}\sqrt{\pi} = \begin{cases} \cos\left(\frac{i\pi}{2m+1}\right), & 1 \leqslant i \leqslant m, \\ 0, & i > m \end{cases}$$

(iii) *Ядро Фейера — Коровкина*

$$\lambda_{mi}\sqrt{\pi} = \begin{cases} \dfrac{(m-i+3)\sin\left(\pi\frac{i+1}{m+2}\right) - (m-i+1)\sin\left(\pi\frac{i-1}{m+2}\right)}{2(m+2)\sin\left(\frac{\pi}{m+2}\right)}, & 1 \leqslant i \leqslant m, \\ 0, & i > m. \end{cases}$$

(iv) *Ядро Валле-Пуссена*

$$\lambda_{mi}\sqrt{\pi} = \begin{cases} \dfrac{m!^2}{(m-i)!\,(m+i)!}, & 1 \leqslant i \leqslant m, \\ 0, & i > m. \end{cases}$$

(v) *Второе ядро Валле-Пуссена*

$$\lambda_{mi}\sqrt{\pi} = \begin{cases} 1, & 1 \leqslant i \leqslant m, \\ 2 - i/(2m+1), & m < i \leqslant 2m-1, \\ 0, & i > 2m. \end{cases}$$

Веса λ_{mi} играют важную роль в исследовании скоростей сходимости сингулярно-интегральных оценок. В следующей лемме приводится основное неравенство, из которого мы будем исходить.

Лемма 20. *Пусть* $S_m(f)$, $S_m^*(f)$ — *сингулярные интегралы для* f *с ядрами* K_m *и* $K_m^2 \Big/ \int K_m^2$ *соответственно, и пусть* f_n — *сингулярно-интегральная оценка с ядром* K_m. *Тогда*

$$\int |S_m(f) - f| \leqslant E\left(\int |f_n - f|\right) \leqslant$$

$$\leqslant \int |S_m(f) - f| + \frac{1}{\sqrt{n}} \int \sqrt{f}\, \sqrt{\int K_m^2} + \left(\frac{2\pi}{n} \int K_m^2 \int |S_m^*(f) - f|\right)^{1/2} \tag{5}$$

для любых плотностей f *на* $[-\pi, \pi]$. *Если* $K_m^2 \Big/ \int K_m^2$ — *дельтаобразная последовательность, то граница сверху равна* $\int |S_m(f) - f| + \left(\int \sqrt{f} + o(1)\right)\left(\int K_m^2/n\right)^{1/2}$. *Все интегралы берутся по* $[-\pi, \pi]$.

Доказательство. Вернемся к доказательству теоремы 8 и заметим, что справедливы неравенства

$$E\left(\int |f_n - E(f_n)|\right) \leqslant \int \sqrt{E\left(\frac{1}{n} K_m^2(x - X_1)\right)}\, dx =$$

$$= n^{-1/2} \int \sqrt{\int K_m^2(x-y) f(y)\, dy}\, dx \leqslant$$

$$\leqslant n^{-1/2} \int \sqrt{f}\, \sqrt{\int K_m^2} + n^{-1/2} \int \sqrt{\left|\int K_m^2(x-y)(f(y) - f(x))\, dy\right|}\, dx$$

Последнее слагаемое здесь не превосходит величины

$$n^{-1/2} \sqrt{\int K_m^2} \int \sqrt{|S_m^*(f) - f|} \leqslant n^{-1/2} \sqrt{\int K_m^2}\, \sqrt{2\pi}\, \sqrt{\int |S_m^*(f) - f|},$$

которая равна o малому от первого слагаемого, если $K_m^2 \Big/ \int K_m^2$ — дельтаобразная последовательность (лемма 19).

Замечание. Если K_m — дельтаобразная последовательность и

$$\sup_m \sup_u K_m(u) \Big/ \int K_m^2 < \infty, \tag{6}$$

то $K_m^2 \Big/ \int K_m^2$ — дельтаобразная последовательность.

Начиная с этого места, мы будем заниматься только оценкой величины $\int | S_m(f) - f |$, а стандартные рассуждения о выборе m из условия минимума верхней границы леммы 20 предоставим читателю. Из леммы 20 можно, разумеется, получить нижние границы как для фиксированной плотности f, так и равномерные. Например, если $\int | S_m(f) - f | = O(m^{-\alpha})$ и $\sup_u | K_m(u)| = O(m^{\beta})$ при некоторых $\alpha, \beta \geqslant 0$, то при $m \sim n^{1/(\beta+2\alpha)}$ получим

$$E\left(\int |f_n - f|\right) = O(n^{-\alpha/(\beta+2\alpha)}).$$

Для всех ядер (за исключением ядра Валле-Пуссена, для которого $\beta = 1/2$) имеем $\beta = 1$. Напомним, что важными значениями α являются $\alpha = 1$ и $\alpha = 2$. При этих значениях α и ядрах, для которых $\beta = 1$, достигаются скорости сходимости $n^{-1/3}$ и $n^{-2/5}$ соответственно. Оставшаяся часть этого параграфа в большой степени посвящена вычислению α для широких классов плотностей и ядер.

Напомним, что AC_s — класс всех функций $f \in L_1[-\pi, \pi]$, имеющих $s-1$ абсолютно непрерывных производных и таких, что $\int | f^{(s)} | < \infty$. Хотя все последующие утверждения верны, если эти условия выполнены для периодического продолжения f, мы будем считать, что они выполнены для исходной плотности f как функции на действительной оси. Это позволит затем провести содержательное сравнение сингулярно-интегральных оценок с другими.

Следуя Butzer, Nessel (1971), определим *модуль непрерывности в* L_1 функции $f \in L_1[-\pi, \pi]$ соотношением

$$\omega(f, \delta) = \sup_{|h| \leqslant \delta} \int_{-\pi}^{\pi} |f(x+h) - f(x)|\, dx, \quad \delta > 0,$$

и *модуль непрерывности в* L_1 *второго порядка* соотношением

$$\omega^*(f, \delta) = \sup_{|h| \leqslant \delta} \int_{-\pi}^{\pi} |f(x+h) + f(x-h) - 2f(x)|\, dx, \quad \delta > 0.$$

Напомним также определение классов Липшица $W(s, \alpha, C)$ при $\alpha \in (0, 1]$, $C > 0$ и целых $s \geqslant 0$: $W(s, \alpha, C)$ — класс всех плотностей из AC_s, для которых

$$|f^{(s)}(x) - f^{(s)}(y)| \leqslant C|y - x|^{\alpha}$$

при любых $x, y \in R$.

Лемма 21 (границы для модулей непрерывности в L_1).

A. $\omega^*(f, \delta) \leqslant 2\omega(f, \delta)$ *при любых* $\delta > 0$ *и* $\omega(f, \delta) \to 0$ *при* $\delta \downarrow 0$ *для любых* $f \in L_1[-\pi, \pi]$.

B. *При* $f \in W(0, \alpha, C)$ *имеем*

$$\omega(f, \delta) \leqslant 2\pi C\delta^{\alpha}, \quad \omega^*(f, \delta) \leqslant 4\pi C\delta^{\alpha}, \quad \delta > 0.$$

C. *При* $f \in W(1, \alpha, C)$ *имеем*

$$\omega^*(f, \delta) \leqslant 2\pi C\delta^{\alpha+1}, \quad \delta > 0.$$

D. *При* $f \in AC_1$ *имеем*

$$\omega(f, \delta) \leqslant \delta \int |f'|, \quad \omega^*(f, \delta) \leqslant \delta\omega(f', \delta),$$

$$\omega^*(f, \delta) \leqslant 2\delta \int |f'|, \quad \delta > 0.$$

E. *При* $f \in AC_2$ *имеем*

$$\omega^*(f, \delta) \leqslant \delta^2 \int |f''|, \quad \delta > 0.$$

Доказательство этой простой леммы опускается. В лемме 22 будут получены верхние границы для $\int |S_m(f) - f|$ в терминах модуля непрерывности в L_1 второго порядка. Нам понадобятся величины

$$r_{mi} = 1 - \sqrt{\pi}\lambda_{mi}, \quad i \geqslant 1.$$

Лемма 22. *Если* K_m *— четное ядро, то*

$$\int |S_m(f) - f| \leqslant \int_0^{\pi} \omega^*(f, u) |K_m(u)| du$$

при любых $f \in L_1[-\pi, \pi]$. *Если* K_m *— четное неотрицательное ядро и* $f \in L_1[-\pi, \pi]$, *то*

$$\int |S_m(f) - f| \leqslant A\omega^*(f, \sqrt{r_{m1}}),$$

где $A = (1 + \pi/\sqrt{2})^2/2$ *— универсальная постоянная.*

Доказательство. Заметим, что

$$S_m(f) - f = \int_0^{\pi} (f(x+u) + f(x-u) - 2f(x)) K_m(u)\, du,$$

так что

$$\int |S_m(f) - f| \leqslant \int_0^{\pi} \omega^*(f,\ u) |K_m(u)|\, du.$$

Для доказательства второго утверждения леммы 22 потребуются некоторые дополнительные рассуждения. Имеем

$$\int_{-\pi}^{\pi} u^2 K_m(u)\, du \leqslant \int_{-\pi}^{\pi} \pi^2 \sin^2(u/2) K_m(u)\, du =$$

$$= \int_{-\pi}^{\pi} \frac{\pi^2}{2} (1 - \cos u) K_m(u)\, du = \frac{\pi^2}{2} (1 - \lambda_{m1} \sqrt{\pi}) = \pi^2 r_{m1}/2$$

и

$$\int_{\pi-}^{\pi} |u| K_m(u)\, du \leqslant \left(\int_{-\pi}^{\pi} u^2 K_m(u)\, du \int_{-\pi}^{\pi} |K_m| \right)^{1/2} \leqslant \pi \sqrt{r_{m1}/2}.$$

Используя тот факт, что $\omega(f, t\delta) \leqslant (1+t)\,\omega(f, \delta)$, $\omega^*(f, t\delta) \leqslant (1+t)^2\,\omega^*(f, \delta)$, $t, \delta > 0$, получим

$$\int |S_m(f) - f| \leqslant \omega^*\left(f, \frac{1}{t}\right) \int_0^{\pi} (1+tu)^2 K_m(u)\, du \leqslant$$

$$\leqslant \omega^*\left(f, \frac{1}{t}\right) \left(\frac{1}{2} + 2t \int_0^{\pi} u K_m(u)\, du + t^2 \int_0^{\pi} u^2 K_m(u)\, du \right) \leqslant$$

$$\leqslant \omega^*\left(f, \frac{1}{t}\right) \left(\frac{1}{2} + \pi t \sqrt{\frac{r_{m1}}{2}} + \frac{\pi^2}{4} t^2 r_{m1} \right).$$

Подставим сюда значение $1/t = \sqrt{r_{m1}}$.

Величина r_{m1} характеризует точность оценивания с неотрицательным ядром: меньшим значениям r_{m1} соответствует меньшее смещение при одних и тех же условиях непрерывности на f. Приведем значения r_{m1} для некоторых из указанных выше ядер.

(i) Ядро Фейера $r_{m1} = 1/(m+1)$.

(ii) Ядро Джексона $r_{m1} = 3/(2m^2+1)$.

(iii) Ядро Фейера — Коровкина $r_{m1} = 1 - \cos(\pi/(m+2)) \leqslant \pi^2/2m^2$ (и $r_{m1} \sim \pi^2/2m^2$).

(iv) Ядро Валле-Пуссена $r_{m1} = 1/(m+1)$.
(v) Ядро Рогозинского $r_{m1} \sim \pi^2/8m^2$.
(vi) Второе ядро Валле-Пуссена $r_{m1} = 0$.

Ядра (v) и (vi) принимают отрицательные значения, и потому их нельзя сравнивать на основе второго утверждения леммы 22. Исходя из леммы 22 и перечисленных значений r_{m1}, заключаем, что ядра Джексона и Фейера — Коровкина — более мощные, чем ядра Фейера и Валле-Пуссена. Объединяя леммы 21 и 22, получим следующие явные границы:

Теорема 9 (границы для смещения). *Пусть* $K_m \geqslant 0$ — *четное ядро и* $r_{m1} \leqslant C_1/m^p$ *при некоторых* C_1, $p > 0$. *Тогда смещение* $\int |S_m(f) - f|$ *не больше, чем*

$$4\pi ACC_1^{\alpha/2} m^{-p\alpha/2}, \quad f \in W(0, \alpha, C),$$

$$2\pi ACC_1^{(\alpha+1)/2} m^{-p(\alpha+1)/2}, \quad f \in W(1, \alpha, C),$$

$$2A\sqrt{C_1} \int |f'| m^{-p/2}, \quad f \in AC_1,$$

$$AC_1 \int |f''| m^{-p}, \quad f \in AC_2,$$

где A — *постоянная из леммы* 21.

Объединяя теорему 9 с границами теоремы 8 или леммы 20, убеждаемся, что при некотором выборе m сингулярно-интегральные оценки Джексона и Фейера — Коровкина удовлетворяют соотношению

$$E\left(\int |f_n - f|\right) = O\left(n^{-(s+\alpha)/(1+2(s+\alpha))}\right),$$

где $\alpha \in (0, 1]$, $s = 0$ или $s = 1$. Итак, на классах Липшица с параметрами $s = 0$ и $s = 1$ величина $E\left(\int |f_n - f|\right)$ равномерно ограничена сверху произведением минимаксной нижней границы для этих классов на постоянную, не зависящую от C и n. Таким образом, эти оценки ведут себя так же, как ядерная оценка с неотрицательным ядром. При фиксированной плотности f асимптотическое поведение границ из леммы 20 и теоремы 9 тоже сходно с поведением аналогичных границ для неотрицательных ядерных оценок, например они одинаковым образом зависят от $\int \sqrt{f}$, $\int |f'|$ ($f \in AC_1$) и от $\int |f''|$ ($f \in AC_2$). К сожалению, при $K_m \geqslant 0$ мы сталкиваемся с теми же ограничениями, что и в случае ядерной оценки. Это вытекает из следующей леммы, взятой из Butzer, Nessel (1971).

Лемма 23 (ограниченные возможности сингулярно-интегральных оценок). *Пусть ядро K_m удовлетворяет следующим условиям:*

(i) $\lambda_{mi} = 0$ *при* $i > m$ (*или при* $i > cm$ *с некоторым* $c > 0$),

(ii) $\liminf\limits_{n\to\infty} m^p r_{mk} > 0$ *при некотором* $p > 0$ *и всех* $k \geqslant 1$.

Тогда если

$$\liminf_{m\to\infty} m^p \int |S_m(f) - f| = 0$$

для некоторой плотности f *на* $[-\pi, \pi]$, *то* $f(x) = 1/2\pi$ *почти всюду,* $|x| \leqslant \pi$. *Если ядро* K_m *удовлетворяет условию* (i) *и неотрицательно, то для любых функций* $f \in L_1[-\pi, \pi]$, *не равных почти всюду постоянной,*

$$\liminf_{m\to\infty} m^2 \int |S_m(f) - f| > 0.$$

Доказательство. Заметим, что $S_m(f)$ — тригонометрический полином степени, не большей m (т. е. линейная комбинация функций $\sin kx$, $\cos kx$, $k \leqslant m$). Кроме того,

$$\int_{-\pi}^{\pi} S_m(f, x) \frac{\cos kx}{\sqrt{\pi}}\, dx = \begin{cases} \sqrt{\pi}\lambda_{mk} a_{2k-1}, & 1 \leqslant k \leqslant m, \\ 0, & k > m, \end{cases}$$

$$\int_{-\pi}^{\pi} S_m(f, x) \frac{\sin kx}{\sqrt{\pi}}\, dx = \begin{cases} \sqrt{\pi}\lambda_{mk} a_{2k}, & 1 \leqslant k \leqslant m, \\ 0, & k > m. \end{cases}$$

Таким образом,

$$\int (S_m(f, x) - f(x)) \left(\frac{\cos kx - i \sin kx}{\sqrt{\pi}}\right) dx = (\sqrt{\pi}\lambda_{mk} - 1)(a_{2k-1} - ia_{2k}),$$

где i — мнимая единица. Следовательно,

$$\frac{1}{\sqrt{\pi}} \int |S_m(f) - f| \geqslant |\sqrt{\pi}\lambda_{mk} - 1| \sqrt{a_{2k-1}^2 + a_{2k}^2} = |r_{mk}| \sqrt{a_{2k-1}^2 + a_{2k}^2}.$$

Пусть сначала $a_{2k-1}^2 + a_{2k}^2 = 0$ при любых $k \neq 0$. Тогда ясно, что $S_m(f, x) = 1/2\pi$ при любых m и $f(x) = 1/2\pi$ при почти всех x, поскольку функция f однозначно определяется своими коэффициентами Фурье. С другой стороны, если $a_{2k-1}^2 + a_{2k}^2 > 0$ при данном $k \neq 0$, то

$$\liminf_{m\to\infty} \frac{\int |S_m(f) - f|}{|r_{mk}|} \geqslant \sqrt{\pi} \sqrt{a_{2k-1}^2 + a_{2k}^2} > 0$$

при этом k, откуда следует первое утверждение леммы. Для доказательства второго утверждения воспользуемся неравенством из Boas, Kac (1945), согласно которому

$$\lambda_{mk}\sqrt{\pi} \leqslant \cos\left(\frac{\pi}{[m/k]+2}\right), \quad 1 \leqslant k \leqslant m,$$

если ядро K_m удовлетворяет условиям леммы 23. Таким образом,

$$\liminf_{m\to\infty} m^2 |r_{mk}| \geqslant \liminf_{m\to\infty} \frac{m^2\pi^2}{2([m/k]+2)^2} = \frac{\pi^2 k^2}{2} > 0.$$

Теперь применим первое утверждение леммы при $p = 2$.

Таким образом, в случае неотрицательных K_m величина $\int |S_m(f) - f|$ не может быть меньше, чем $O(m^{-2})$, ни при каких условиях гладкости на f. Это значит, что при использовании таких ядер можно выиграть по сравнению с ядрами Джексона или Фейера — Коровкина разве лишь в постоянном множителе. Остается надеяться только на то, что смещение уменьшается при ядрах K_m с отрицательными значениями. Это будет показано далее.

В лемме 23 содержится много информации о наилучших скоростях сходимости. В частности, для ядер $K_m \geqslant 0$, удовлетворяющих условию (i) леммы 23, имеем

$$\liminf_{m\to\infty} m^2 \int |S_m(f) - f| \geqslant \sup_{k>1} \sqrt{\pi}\,\frac{\pi^2 k^2}{2}\sqrt{a_{2k-1}^2 + a_{2k}^2},$$

где правая часть равна бесконечности, если

$$\limsup_{k\to\infty} k^2 |a_k| = \infty.$$

В случае ядер Фейера и Валле-Пуссена и не постоянной f смещение ограничено снизу некоторой постоянной, деленной на m. В теореме 9 такие же границы получены только для классов $W(1, 1, C)$ и AC_2. На самом же деле сингулярно-интегральная оценка Фейера имеет смещение $O(1/m)$ при $f \in AC_2$ и при $f \in W(1, \alpha, C)$ для любых $\alpha > 0$. Таким образом, применение теоремы 9 иногда приводит к неоптимальным границам. Для получения верхних границ правильного порядка можно использовать первое неравенство леммы 22. Это будет показано ниже на примере сингулярного интеграла Фейера.

Лемма 24. *Пусть* $S_m(f)$ — *сингулярный интеграл Фейера для плотности* f *на* $[-\pi, \pi]$. *Тогда*

$$\int |S_m(f) - f| =$$

$$= \begin{cases} O(1/m), & f \in W(1, \alpha, C) \text{ при любых } \alpha \in (0, 1] \text{ или } f \in AC_2, \\ O(\log m/m), & f \in W(0, 1, C) \text{ или } f \in AC_1, \\ O(m^{-\alpha}), & f \in W(0, \alpha, C),\ \alpha \in (0, 1). \end{cases}$$

Доказательство. Воспользуемся такими же оценками, как в Butzer, Nessel (1971, с. 81). В силу леммы 4

$$2\pi u^{\alpha} F_m(u) \leqslant \begin{cases} 2^{\alpha-1}\pi (m+1)^{1-\alpha}, & 0 < u \leqslant 1/m,\ 0 < \alpha \leqslant 1, \\ \pi^2 u^{\alpha-2} (m+1)^{-1}, & 0 < u \leqslant \pi,\ 0 < \alpha \leqslant 2. \end{cases}$$

Отсюда $\int_0^{1/m} u^{\alpha} F_m(u)\, du \leqslant m^{-\alpha}$, $0 < \alpha \leqslant 1$, и

$$\int_{1/m}^{\pi} u^{\alpha} F_m(u)\, du \leqslant \begin{cases} \frac{\pi}{2}(1-\alpha)^{-1} m^{-\alpha}, & 0 < \alpha \leqslant 1, \\ \frac{\pi}{2} \frac{\log(\pi m)}{m+1}, & \alpha = 1. \end{cases}$$

Отсюда следуют границы

$$\int_0^{\pi} u^{\alpha} F_m(u)\, du \leqslant \begin{cases} m^{-\alpha}\left(1 + \frac{\pi}{2(1-\alpha)}\right), & \alpha \in (0, 1), \\ \frac{1}{m} + \frac{\pi}{2} \frac{\log(\pi m)}{m+1}, & \alpha = 1, \\ \frac{\pi^{\alpha}}{2(\alpha-1)} \frac{1}{m}, & \alpha \in (1, 2]. \end{cases}$$

Из этих границ, леммы 21 и первого неравенства леммы 22 следует лемма 24. (Для всех интересующих нас классов плотностей выполнено неравенство $\int |S_m(f) - f| \leqslant C \int_0^{\pi} u^{\alpha} F_m(u)\, du$, где C и α — некоторые постоянные.)

Для сингулярно-интегральной оценки Фейера соотношение $E\left(\int |f_n - f|\right) = o(n^{-1/3})$ по существу можно получить разве лишь в случае, когда f — равномерная плотность на $[-\pi, \pi]$. В этом смысле сингулярно-интегральные оценки Фейера и Валле-Пуссена ведут себя так же, как гистограммная оценка. На другом конце шкалы точности находятся сингулярно-интегральные оценки неограниченной мощности, т. е. такие оценки, что

$$\int |S_m(f) - f| = O(m^{-p})$$

для каждой степени $p \geqslant 1/2$ при «достаточно гладких» f. Из леммы 23 видно, что в этом случае ядро K_m обязательно должно принимать отрицательные значения и при любых $i \neq 0$, $p > 0$ должно выполняться условие

$$\liminf_{m\to\infty} m^p r_{mi} = 0.$$

Последнее условие выполнено для второго ядра Валле-Пуссена (поскольку $r_{mi} = 0$, $1 \leqslant i \leqslant m$) и для ядра Дирихле. По существу нужно, чтобы значения λ_{mi} были постоянны вблизи нуля ($\sqrt{\pi}\lambda_{mi} = 1$ при любых i, меньших, чем некоторое число, стремящееся к бесконечности при $m \to \infty$). Есть надежда, что, используя такие ядра, не меняющиеся в зависимости от гладкости плотности, можно получить любую скорость сходимости. Аналогичным образом ведет себя трапецеидальная ядерная оценка.

Заметим, что принимающее отрицательные значения ядро Рогозинского удовлетворяет условиям (i) и (ii) леммы 23 со степенью $p = 2$. Таким образом, оно обладает такими же ограниченными возможностями, как ядро Джексона и другие ядра с положительными значениями. Это приносит определенное разочарование, поскольку, жертвуя положительностью оценки плотности f_n, можно было бы выбирать и ядра неограниченной мощности, указанные выше. Холл (Hall (1983)) сравнил сингулярно-интегральную оценку Рогозинского с другими оценками, основанными на положительных ядрах, и нашел, что она обладает такой же скоростью сходимости в L_2, но с меньшей константой. В некотором смысле это «надувательство». На самом деле, как мы увидим далее, при использовании ядер неограниченной мощности достигается лучшая скорость сходимости при *тех же самых* условиях гладкости на f.

Завершим этот параграф описанием свойств одного ядра неограниченной мощности. Для получения уточненных скоростей сходимости воспользуемся первой и второй теоремами Джексона. Это еще один технический прием, которым может вооружиться читатель (напомним, что в § 5 при анализе поведения оценки с тригонометрическим рядом применялось неравенство Лоренца).

Лемма 25 (теоремы Джексона). *Пусть T_m — класс всех тригонометрических полиномов степени, не превосходящей m, и пусть $f \in L_1[-\pi, \pi]$.*

Первая теорема Джексона.

$$\inf_{t_m \in T_m} \int |t_m - f| \leqslant 2A^2\omega^*(f, 1/m),$$

где A — постоянная из леммы 22.

Вторая теорема Джексона.

$$\inf_{t_m \in T_m} \int |t_m - f| \leqslant \begin{cases} \left(\frac{36}{m}\right)^s \int |f^{(s)}|, \\ \frac{36^s (36^2+1)}{m^s} \omega^* \left(f^{(s)}, \frac{1}{m}\right) \end{cases}$$

при любых $f \in AC_s$, $s > 0$.

Доказательство. Не будем стремиться получить наилучшие константы (см., например, Butzer, Nessel (1971), где имеются ссылки на работы, посвященные наилучшим константам, и полное доказательство второй теоремы Джексона). Докажем только первую теорему Джексона.

Заметим, что в случае ядра Фейера—Коровкина $S_m(f) \in T_m$, так что

$$\inf_{t_m \in T_m} \int |t_m - f| \leqslant \int |S_m(f) - f|.$$

Кроме того, в силу леммы 22 получим

$$\int |S_m(f) - f| \leqslant A\omega^*\left(f, \left(1 - \cos\left(\frac{\pi}{m+2}\right)\right)^{1/2}\right) \leqslant A\omega^*\left(f, \frac{\pi}{(m+2)\sqrt{2}}\right) \leqslant$$

$$\leqslant A\left(1 + \pi/\sqrt{2}\right)^2 \omega^*(f, 1/m) = (1/2)\left(1 + \pi/\sqrt{2}\right)^4 \omega^*(f, 1/m).$$

Первая теорема Джексона доказана.

Теорема 10 (сингулярно-интегральная оценка со вторым ядром Валле-Пуссена). *Пусть* f_n *— сингулярно-интегральная оценка со вторым ядром Валле-Пуссена (обозначаемым здесь через* K_m*), и пусть* $S_m(f)$ *— соответствующий сингулярный интеграл для плотности* f *на* $[-\pi, \pi]$.

A. K_m *— дельтаобразная последовательность, и потому*

$$\int |S_m(f) - f| = o(1), \quad \forall f.$$

B. $\int |S_m(f)| \leqslant 3 \int |g|$ *при любых* $g \in L_1[-\pi, \pi]$.

C. $\int |S_m(f) - f| \leqslant 4 \inf_{t_m \in T_m} \int |t_m - f|$.

D. $\int K_m^2/n \leqslant 9m/4n$.

E. $K_m^2 \Big/ \int K_m^2$ *— дельтаобразная последовательность.*

F. *Если* $\lim_{n\to\infty} m = \infty$, $\lim_{n\to\infty} (m/n) = 0$, *то*

$$E\left(\int |f_n - f|\right) = o(1), \quad \forall f.$$

G. $E\left(\int |f_n - f|\right) \leqslant 4 \inf_{t_m \in T_m} \int |t_m - f| + \left\{ \begin{array}{l} \left(\int \sqrt{f} + o(1)\right) \sqrt{9m/4n} \\ \sqrt{9\pi m/2n}. \end{array} \right.$

H. *Если* M *— целочисленная постоянная,* $f \in T_M$ *и* m *— постоянная,* $m \geqslant M$, *то*

$$\int |S_m(f) - f| = 0, \quad E\left(\int |f_n - f|\right) = O(n^{-1/2}).$$

I. *Если* $f \in AC_s$, *где* $s > 0$ *— целое, то*

$$\inf_m E\left(\int |f_n - f|\right) = o(n^{-s/(2s+1)}).$$

J. *Если* $f \in AC_s$, *где* $s > 0$ *— целое, то*

$$E\left(\int |f_n - f|\right) \leqslant 4\left(\frac{36}{m}\right)^s \int |f^{(s)}| + \left\{ \begin{array}{l} \left(\int \sqrt{f} + o(1)\right) \sqrt{9m/4n}, \\ \sqrt{9\pi m/2n} \end{array} \right.$$

K. *Если* $f \in W(s, \alpha, C)$ *при некотором целом* $s \geqslant 0$, $\alpha \in (0, 1]$, $C > 0$, *то*

$$E\left(\int |f_n - f|\right) \leqslant 4(36)^s (36^2 + 1) \frac{4\pi C}{m^{s+\alpha}} + \sqrt{\frac{9\pi m}{2n}},$$

$$\inf_m E\left(\int |f_n - f|\right) = O(n^{-(s+\alpha)/(2(s+\alpha)+1)}).$$

Доказательство. Утверждение A следует из леммы 19. Для доказательства B заметим, что

$$\int |K_m| \leqslant \int |(1 + 2\cos(mu))| F_{m-1}(u)\, du \leqslant 3 \int F_{m-1}(u)\, du = 3,$$

и потому

$$\int |S_m(g)| \leqslant \int |g| \int |K_m| \leqslant 3 \int |g|.$$

При $t_m \in T_m$ имеем $S_m(t_m) = t_m$. Таким образом,

$$\int |S_m(f) - f| \leqslant \inf_{t_m \in T_m} \left(\int |S_m(f) - S_m(t_m)| + \int |t_m - f|\right) \leqslant$$

$$\leqslant 4 \inf_{t_m \in T_m} \int |t_m - f|,$$

откуда следует C. Утверждение D получается, если применить грубые границы из леммы 4:

$$\int K_m^2 \leqslant 9 \int F_{m-1}^2 \leqslant 9 \sup F_{m-1} \leqslant 9m/4.$$

Утверждение E тривиально. Состоятельность (утверждение F) следует из A, D и теоремы 8. Утверждения C и D вместе с неравенствами из теоремы 8 и леммы 20 дают утверждение G. Из него непосредственно вытекает утверждение H. Учитывая, что $\omega^*(f^{(s)}, 1/m) = o(1)$ при $f \in AC_s$, получим, что I следует из G и последнего неравенства во второй теореме Джексона. Используя первое неравенство во второй теореме Джексона, приходим к утверждению J. Наконец, утверждение K вытекает из G, второй теоремы Джексона и утверждения B леммы 21.

Сингулярно-интегральная оценка из теоремы 10 — лишь незначительно видоизмененная оценка с тригонометрическим рядом, но она надежнее последней, ибо обладает свойством универсальной состоятельности. Кроме того, граница смещения (утверждение C) лучше, чем соответствующая граница для оценки с тригонометрическим рядом (теорема 4), на множитель порядка $\log m$. Средняя L_1-ошибка рассматриваемой здесь оценки, так же как и оценки с тригонометрическим рядом или трапецеидальной ядерной оценки, достигает минимаксной нижней границы для классов $W(s, \alpha, C)$ *при любых* s с точностью до константы (см. утверждение K). Поэтому мы называем ее оценкой неограниченной мощности. Она достигает даже скорости сходимости $O(n^{-1/2})$ и в этом случае является несмещенной при любых n (утверждение H). Наконец, утверждение I показывает, что даже на классе AC_2 данная оценка лучше любых сингулярно-интегральных оценок с ядрами $K_m \geqslant 0$ и оценки Рогозинского — для последних ошибка в лучшем случае равна $O(n^{-2/5})$, так как их смещение не меньше, чем величина порядка m^{-2} (лемма 23).

ЛИТЕРАТУРА

Anderson G. L., de Figueiredo R. J. P.
(1980) An adaptive orthogonal-series estimator for probability density functions, Annals of Statistics, 8, pp. 347—376.

Askey R., Wainger S.
(1965) Mean convergence of expansions in Laguerre and Hermite series, American Journal of Mathematics, 87, pp. 695—708.

Asselin de Beauville J. P.
(1978) Estimation de la densité de probabilité par une série de polynômes d'Hermite. Détermination du nombre optimal de termes de la série, Comptes Rendus de l'Académie des Sciences de Paris, 286, pp. 309—311.

Bleuez J., Bosq D.
(1976) Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité, Comptes Rendus de l'Académie des Sciences de Paris, 282, pp. 63—66.

(1979) Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales, Revue Roumaine de Mathématiques Pûres et Appliquées, 24, pp. 869—886.

Boas R. P., Kac M.

(1945) Inequalities for Fourier transforms of positive functions, Duke Mathematical Journal, 12, pp. 189—206.

Bosq D.

(1969) Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales, Comptes Rendus de l'Academie des Sciences de Paris, 268, pp. 555—557.

Bosq D., Bleuez J.

(1978) Etude d'une classe d'estimateurs non-paramétriques de la densité, Annales de l'Institut Henri Poincaré, 14, pp. 479—498.

Brunk H. D.

(1977) Univariate density estimation by orthogonal series, with application to estimation of wildlife populations by line transect surveys, unpublished manuscript.

(1978) Univariate density estimation by orthogonal series, Biometrika, 65, pp. 521—528.

Butzer P. L., Nessel R. J.

(1971) Fourier Analysis and Approximation, Vol. 1, Birkhäuser Verlag, Basel and Stuttgart.

Carleson L.

(1966) On convergence and growth of partial sums of Fourier series, Acta Mathematica, 116, pp. 135—157.

Crain B. R.

(1973) A note on density estimation using orthogonal expansions, Journal of the American Statistical Association, 68, pp. 964—965.

(1974) Estimation of distributions using orthogonal expansions, Annals of Statistics, 2, pp. 454—463.

Edwards R. E.

(1979) Fourier Series. A Modern Introduction, Vol. 1, Springer-Verlag, Berlin.

Fellner W. H.

(1974) Heuristic estimation of probability densities, Biometrika, 61, pp. 485—492.

Fefferman C.

(1971) On the convergence of multiple Fourier series, Bulletin of the American Mathematical Society, 77, pp. 744—755.

Földes A., Révész P.

(1974) A general method for density estimation, Studia Scientiarium Mathematicarum Hungarica, 9, pp. 81—92.

Greblicki W.

(1981) Asymptotical efficiency of classifying procedures using the Hermite series estimate of multivariate probability densities, IEEE Transactions on Information Theory, IT-27, pp. 364—366.

Greblicki W., Pawlak M.

(1981) Classification using the Fourier series estimate of multivariate density functions, IEEE Transactions on Systems, Man and Cybernetics, SMC-11, pp. 726—730.

Hall P.

(1981) On trigonometric series estimates of densities, Annals of Statistics, 9, pp. 683—685.

(1982) Comparison of two orthogonal series methods of estimating a density and its derivatives on an interval, Journal of Multivariate Analysis, 12, pp. 432—449.

(1983) Measuring the efficiency of trigonometric series estimates of a density, Journal of Multivariate Analysis, 13, pp. 234—256.

Hunt R. A.

(1968) On the convergence of Fourier series. Orthogonal expansions and their continuous analogues, Proceedings of a Conference held at Edwardsville, Illinois, 1967. Southern Illinois University Press, Carbondale, pp. 235—255.

Jorsboe O. G., Mejlbro L.
(1982) The Carleson—Hunt theorem of Fourier Series, Springer-Verlag, Berlin.
Kolmogorov A. N.
(1926) Une série de Fourier—Lebesgue divergente partout, Comptes Rendus de l'Académie des Sciences de Paris, 183, pp. 1327—1328. (Имеется русский перевод: Колмогоров А. Н. Избранные труды. Математика и механика. — М.: Наука, 1985, с. 8—12.)
Körner T. W.
(1981) Everywhere divergent Fourier series, Colloquium in Mathematics, 45, pp. 103—118.
Kronmal R., Tarter M.
(1968) The estimation of probability densities and cumulatives by Fourier series methods, Journal of the American Statistical Association, 63, pp. 925—952.
Krzyzak A., Pawlak M.
(1982) Estimation of a multivariate density by orthogonal series, in Probability and Statistical Inference, W. Grossman et al. (Eds.), Reidel, Hingham, MA, pp. 211—221.
Lorentz G. G.
(1948) Fourier-Koeffizienten und Funktionenklassen, Mathematische Zeitschrift, 51, pp. 135—149.
Mozzochi C. J.
(1971) On the Pointwise Convergence of Fourier Series, Springer-Verlag, Berlin.
Muckenhoupt B.
(1969) Mean convergence of Jacobi series, Proceedings of the American Mathematical Society, 23, pp. 306—310.
(1970a) Equiconvergence and almost everywhere convergence of Hermite and Laguerre series, SIAM Journal of Mathematical Analysis, 1, pp. 295—321.
(1970b) Mean convergence of Hermite and Laguerre series. I, Transactions of the American Mathematical Society, 147, pp. 419—431.
(1970c) Mean convergence of Hermite and Laguerre series. II, Transactions of the American Mathematical Society, 147, pp. 433—460.
Newman J., Rudin W.
(1952) Mean convergence of orthogonal series, Proceedings of the American Mathematical Society, 3, pp. 219—222.
Olevskii A. M.
(1975) Fourier Series with Respect to General Orthogonal Systems, Springer-Verlag, Berlin.
Pollard H.
(1947) The mean convergence of orthogonal series. I, Transactions of the American Mathematical Society, 62, pp. 387—403.
(1948) The mean convergence of orthogonal series. II, Transactions of the American Mathematical Society, 63, pp. 355—367.
(1949) The mean convergence of orthogonal series. III, Duke Mathematical Journal, 16, pp. 189—191.
Quade E. S.
(1937) Trigonometric approximation in the mean, Duke Mathematical Journal, 3, pp. 529—543.
Rosenblatt M.
(1971) Curve estimates, Annals of Mathematical Statistics, 42, pp. 1815—1841.
Sansone G.
(1977) Orthogonal Functions, Krieger, Huntington, NY.
Schüler L.
(1976) Uber die Konsistenz einer Schätzung mehrdimensionaler Dichten auf der Basis trigonometrischer Reihen, Metrika, 23, pp. 77—82.
Schwartz S. C.
(1967) Estimation of a probability density by an orthogonal series, Annals of Mathematical Statistics, 38, pp. 1262—1265.

Sjölin P.
(1971) Convergence almost everywhere of certain singular integrals and multiple Fourier series, Arkiv für Mathematik, 9, pp. 65—90.

Stegbuchner H.
(1980) Dichteschätzungen mit Gleichverteilungsmethoden, Periodica Mathematica Hungarica, 11, pp. 161—175.

Stein E. M.
(1981) On limits of sequences of operators, Annals of Mathematics, 74, pp. 140—170.

Sterbuchner H.
(1980) On nonparametric multivariate density estimation, Revue Roumaine de Mathématiques Pûres et Appliquées, 25, pp. 111—118.

Szegö G.
(1975) Orthogonal Polynomials, Vol. 23, 4th Ed., American Mathematical Society Colloquia Publications, Providence, RI. (Имеется русский перевод: Сегё Г. Ортогональные многочлены. — М.: Физматгиз, 1962.)

Tarter M. E., Kronmal R. A.
(1976) An introduction to the implementation and theory of nonparametric density estimation, The American Statistician, 30, pp. 105—112.

Van Ryzin J.
(1966) Bayes risk consistency of classification procedures using density estimation, Sankhya, Series A, 28, pp. 261—270.

Viollaz A. J.
(1980) Asymptotic distribution of L_2 norms of the deviations of density function estimates, Annals of Statistics, 8, pp. 322—346.

Wahba G.
(1975) Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, Annals of Statistics, 3, pp. 15—29.
(1977) Optimal smoothing of density estimates, in Classification and Clustering, J. Van Ryzin (Ed.) Academic Press, New York, pp. 423—458.
(1978) Data-based optimal smoothing of orthogonal series density estimates, Department of Statistics, University of Wisconsin, Technical Report 509.

Walter G. G.
(1977) Properties of Hermite series estimation of probability density, Annals of Statistics, 5, pp. 1258—1264.

Walter G. G., Blum J. R.
(1979) Probability density estimation using delta sequences, Annals of Statistics, 7, pp. 328—340.

Watson G. S.
(1969) Density estimation by orthogonal series, Annals of Mathematical Statistics, 40, pp. 1496—1498.

Whittle P.
(1958) On the smoothing of probability density functions, Journal of the Royal Statistical Society, B 20, pp. 334—343.

Zygmund A.
(1959) Trigonometric Series, Vols. 1, 2, Cambridge University Press, Cambridge, U. K. (Имеется русский перевод: Зигмунд А. Тригонометрические ряды (в 2-х томах). — М.: Мир, 1965.)

Бари Н. К.
(1961) Тригонометрические ряды. — М.: Физматгиз.

Ченцов Н. Н.
(1962) Оценка неизвестной плотности распределения по наблюдениям//Докл. АН СССР. — Т. 147. — 1. — С. 45—48.

КОММЕНТАРИЙ ПЕРЕВОДЧИКА

Сингулярно-интегральные оценки плотности введены в работе Watson, Leadbetter (1964*). Необходимые и достаточные условия сходимости сингулярно-интегральных оценок в L_1 получены Мнацакановым, Хмаладзе (1981*). Оценка с рядом Лагерра исследована Хашимовым (1973*). Равномерная сходимость проекционных оценок плотности доказана в работе Mirzahmedov, Hašimov (1972*).

Ефроймович, Пинскер (1982*) нашли сглаженные проекционные оценки плотности, асимптотически эффективные в точном минимаксном смысле в метрике L_2. Приведем формулировку их результата, используя обозначения, принятые в данной главе. Пусть $f \in L_2[0,1]$, $\{p_j\}$ — ортонормальная тригонометрическая система на $[0,1]$. Обозначим $\mathcal{F}(\{b_j\}, Q)$ класс плотностей $f \in L_2[0,1]$, для которых

$$\sum_{j=0}^{\infty} a_j^2 b_j \leqslant Q,$$

где $\{b_j\}$ — заданная положительная числовая последовательность, a_j — коэффициенты Фурье для f и $Q < \infty$. Рассмотрим сглаженную проекционную оценку

$$f_n(x) = \sum_{i=0}^{m_n} \lambda_{ni} a_{ni} p_i(x),$$

где

$$a_{ni} = \frac{1}{n} \sum_{j=1}^{n} p_i(X_j),$$

$$m_n = \max\left\{k: \sum_{j=0}^{k} [(b_k b_j)^{1/2} - b_j] \leqslant nQ\right\},$$

$$\lambda_{ni} = 1 - (\mu_n b_i)^{1/2}, \quad i = 1, \ldots, m_n,$$

$$\mu_n = b_{m_n}^{-1} \left[\sum_{j=0}^{m_n} (b_{m_n} b_j)^{1/2} \Big/ \left(nQ + \sum_{j=0}^{m_n} b_j\right)\right]^2.$$

Теорема (Ефроймович, Пинскер (1982*)). *Пусть последовательность* $\{b_j\}$ *монотонно не убывает и* $\lim\limits_{j\to\infty} b_j = \infty$. *Тогда*

$$\sup_{f \in \mathcal{F}(\{b_j\}, Q)} E\left(\int (f_n - f)^2\right) \leqslant \nu_n^2 (1 + o(1)), \quad n \to \infty,$$

где $\nu_n^2 = n^{-1} \sum\limits_{i=1}^{m_n} \lambda_{ni}$. *Если, кроме того,* $\lim\limits_{j\to\infty} b_j \ln^{-1} j = \infty$, *то*

$$\inf_{T_n} \sup_{f \in \mathcal{F}(\{b_j\}, Q)} E\left(\int (T_n - f)^2\right) = \nu_n^2 (1 + o(1)), \quad n \to \infty,$$

где $\inf\limits_{T_n}$ — *нижняя грань по всевозможным оценкам* $T_n = T_n(x, X_1, \ldots, X_n)$.

Пример. Пусть $b_0 = 0$, $b_{2j} = b_{2j-1} = (2\pi j)^{2\alpha}$, $j = 1, 2, \ldots$, α — натуральное число. Тогда $\mathcal{F}(\{b_j\}, Q)$ — множество плотностей, имеющих квадратично-

интегрируемые периодические производные до порядка $\alpha - 1$ включительно и α-ю производную $f^{(\alpha)}$, такую, что

$$\int_0^1 (f^{(\alpha)}(x))^2 dx \leqslant Q.$$

В этом случае

$$v_n^2 = \left[Q(2\alpha+1) \left(\frac{2\alpha}{2\pi(\alpha+1)} \right)^{2\alpha} \right]^{1/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} (1 + o(1)), \quad n \to \infty.$$

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Ефроймович С. Ю., Пинскер М. С.
(1982*) Оценивание квадратично-интегрируемой плотности вероятности случайной величины//Проблемы передачи информации. — Т. 18. — 3. — С. 19—38.

Мнацаканов Р. М., Хмаладзе Э. В.
(1981*) Об L_1-сходимости статистических ядерных оценок плотности распределений//Докл. АН СССР. — Т. 258. — 5. С. 1052—1055.

Хашимов Ш. А.
(1973*) Оценка плотности вероятности полиномами Лагерра//Случайные процессы и статист. выводы. Вып. 3. — Ташкент: Фан, с. 186—191.

Mirzahmedov M. A., Hašimov S. A.
(1972*) On some properties of density estimation//Colloq. Math. Soc. J. Bolyai. — V. 9. Budapest. — P. 535—545.

Watson G. S., Leadbetter M. R.
(1964*) Hazard analysis II//Sankhya, ser. A. — V. 26. — 1. — P. 101—116.

Дополнение 1

Почему L_1-подход и что за горизонтом?

Н. Н. Ченцов

Монография Л. Девроя и Л. Дьёрфи посвящена задаче статистической оценки плотности распределения по наблюдениям. В книге сделан обзор современного состояния исследований в этой области, доведенный в ряде мест до 1983 г., подробно рассмотрены наиболее интересные и важные результаты, а также вопросы, остающиеся нерешенными. Название книги точно отражает ее содержание, так что неискушенному читателю может показаться, что книга двух сравнительно молодых специалистов посвящена узкому, даже экзотическому специальному вопросу. Между тем L_1-подход является наиболее простым и естественным в рассматриваемой проблеме, и можно дать тому строгое математическое доказательство. Для этого надо лишь обратиться к общей теории статистических выводов, заложенной А. Вальдом [1]. В предлагаемом ниже кратком очерке некоторых основных понятий математической статистики выясняется место рассмотренной в книге проблематики и тем самым дополняются и уточняются суждения авторов на этот счет (в частности, по поводу L_1-подхода и его сравнения с L_2-подходом).

В теории вероятностей случайное явление задается своим вероятностным пространством $(\Omega, \mathscr{A}, P)$. В этой идеализированной схеме Ω — пространство всех мыслимых исходов $\omega \in \Omega$ эксперимента, $\mathscr{A}$ есть σ-алгебра подмножеств $A \subseteq \Omega$, называемых событиями, P — вероятностная мера на $\mathscr{A}$, называемая распределением вероятностей подходов. Измеримое пространство $(\Omega, \mathscr{A})$ задает качественное описание случайного явления, а мера P — количественное. Пусть $\omega^1, \ldots, \omega^N$ — последовательность независимых наблюдений данного явления, т. е. $P^N\{d\omega\} = P\{d\omega^1\} \ldots P\{d\omega^N\}$. Тогда по закону больших чисел для любого события $A \in \mathscr{A}$ частота $\frac{\nu(A)}{N} \approx P\{A\}$, где $\nu_N(A) = \mathrm{Card}\{\omega^k : \omega^k \in A,\ k = 1, \ldots, N\}$, причем точность и надежность (по P^N-вероятности) этого предсказания растут с увеличением N. Задача предсказания частот является одной из основных в теории вероятностей и во многом определяет ее практическую ценность. В математической статистике мы сталкиваемся с обратной задачей: нам дана последовательность наблюдений $\omega^1, \ldots,$

ω^N некоторого случайного явления с известным качественным описанием $(\Omega, \mathscr{A})$ и требуется оценить «наблюдаемое» распределение P или какую-то его характеристику либо выяснить наличие у этого неизвестного наблюдателю закона P тех или иных свойств. В первом случае говорят о *задаче статистической точечной оценки* (с. т. о.), и именно одна из ее постановок и составляет предмет исследования Девроя и Дьёрфи, а также нашего рассказа. Здесь только следует добавить, что при решении иных статистических задач процедура точечного оценивания закона P используется как удобный промежуточный этап, когда о свойствах P судят по построенной его оценке P^*.

Обратные задачи математической физики, как правило, некорректны. Не составляет в этом отношении исключения и обратная задача теории вероятностей. Для ее корректной сильной разрешимости нужна дополнительная априорная информация. Например, может быть заранее известно, что наблюдаемое P принадлежит семейству $\mathscr{P} = \{P_\theta,\ \theta \in \Theta\}$. Когда «координатизирующее» множество Θ конечно, задачу статистической точечной оценки называют *задачей проверки* (нескольких простых) *гипотез*. Когда законы P_θ гладко зависят от конечномерного векторного параметра θ, говорят о *параметрической задаче оценивания*. Основное внимание в этой книге уделяется случаю, когда такое априорное семейство $\mathscr{P}$ законов может быть гладко запараметризовано только счетным числом вещественных координат. Так будет, например, если известно лишь, что наблюдаемая случайная величина ограниченна, а ее распределение имеет плотность $p(x)$, принадлежащую функциональному пространству $C^{(2)}$. Наконец, семейство $\mathscr{P}$ может быть столь обширно, что оно не допускает даже счетномерной гладкой параметризации. Так будет, например, когда известно лишь, что координаты наблюдаемой двумерной случайной величины независимы (друг от друга), и ничего более. По традиции обе последние постановки объединяют под названием *задачи непараметрического оценивания*, хотя они резко отличаются и по подходам, и по методам решения. В частности, последняя задача корректна лишь в слабом смысле. Приведенный перечень не дает, разумеется, исчерпывающей классификации всех возможных постановок задачи с. т. о. Мы указали лишь наиболее важные из них, расположив их в порядке убывания априорной информации и вытекающего отсюда убывания точности решения. Так, при проверке простых гипотез вероятность ошибки убывает экспоненциально, в конечно-параметрической задаче с. т. о. погрешность в значении параметра и закона имеет порядок $N^{-1/2}$, а в счетно-параметрической порядок убывания не достигает и этой величины. Как мы уже отмечали, именно последняя задача и связанные с нею аспекты конечно-параметрической задачи являются предметом нашего рассмотрения.

Согласно догме Вальда [1], в каждой статистической задаче кроме вводных данных должно быть указано измеримое пространство $(\Delta, \mathscr{B})$ выводов $\delta \in \Delta$, к одному из которых мы должны прийти в результате статистического анализа данных эксперимента. Тогда каждое детермированное решающее правило (т. е. избранная статистиком процедура обработки наблюдений) задается своей функцией $\delta = f(\omega^1, \ldots, \omega^N)$, а стохастическое (или рандомизированное) — функцией $\delta = f(\omega_1, \ldots, \omega^N; \eta)$, где η — случайный параметр, который приходится дополнительно разыгрывать. Теоретически рандомизированные (и даже детерминированные) решающие правила удобнее описывать неявно, переходным распределением вероятностей $Ш(\omega^1, \ldots, \omega^N; d\delta)$, не фиксируя способа конструирования дополнительного случайного эксперимента, в котором будет «разыгран» вывод.

Точность выбранного решения δ Вальд предлагает определять функцией потерь $\mathscr{L}(\theta, \delta)$ — убытком, который терпит статистик, принявший решение δ, когда на самом деле он наблюдает явление P_θ. Тогда качество решающего правила $Ш$ можно охарактеризовать математическим ожиданием потерь — риском

$$\mathscr{R}_{Ш\,(N)} = E\mathscr{L}(\theta, \delta) = \int_\Delta (P_\theta^N Ш)\{d\delta\}\,\mathscr{L}(\theta, \delta), \qquad (1)$$

где вероятностная мера $P_{\theta, Ш}^{(N)} = P_\theta^N Ш$ задается, как

$$P_{\theta, Ш}^{(N)}\{\cdot\} = \int_{\Omega^N} Ш(\omega^1, \ldots, \omega^N; \cdot)\, P_\theta\{d\omega^1\} \ldots P_\theta\{d\omega^N\}. \qquad (2)$$

Заметим в скобках, что вместо риска можно рассматривать какой-либо квантиль случайных потерь. Так или иначе, зная семейство $\mathscr{P} = \{P_\theta\}$, мы можем заранее вычислять распределения $P_{\theta, Ш}^{(N)}$ и подбирать по ним наиболее «выгодное» решающее правило $Ш$.

В обратной задаче теории вероятностей пространством Δ всевозможных ответов P^* является в принципе вся полная совокупность Сар $(\Omega, \mathscr{A})$ всех распределений вероятностей P на $(\Omega, \mathscr{A})$. Для лебеговых измеримых пространств $(\Omega, \mathscr{A})$, преодолев, следуя Ю. В. Прохорову, некоторые технические трудности, см. [2, § 29], можно снабдить эту совокупность σ-алгеброй событий и построить измеримое пространство случайных вероятностных мер Сар (Сар $(\Omega, \mathscr{A},)$ $\mathscr{K}(\mathscr{A})$). При этом параметрами закона P служат значения $P\{A_i\}$ на некоторой системе образующих A_i алгебры $\mathscr{A}$ и зависимость P от параметров $\mathscr{K}(\mathscr{A})$ измерима, но не гладка.

В качестве величины погрешности оценки P^* искомой вероятностной меры P естественно в первую очередь взять вариацию их разности

$$|P^* - P| = \sup_{A \in \mathcal{A}} [P^*\{A\} - P\{A\}] - \inf_{A \in \mathcal{A}} [P^*\{A\} - P\{A\}], \quad (3)$$

т. е. задать ее через сильную норму в линейном пространстве $L_1(\Omega, \mathcal{A})$ всех (знакопеременных, вообще говоря) счетно-аддитивных мер на $(\Omega, \mathcal{A})$; $L_1(\Omega, \mathcal{A}) = \operatorname{Lin} \operatorname{Cap}(\Omega, \mathcal{A})$.

Следующий точный результат принадлежит автору этих строк, см. [3, теорема 4]:

Теорема 1. *Когда погрешность оценки с. т. о. измеряется сильной нормой, задача с. т. о. для совокупности* $\operatorname{Cap}(E, \mathcal{A}^*)$, *где E — единичный отрезок, а $\mathcal{A}^*$ — алгебра всех его абсолютно измеримых лебеговых подмножеств, без дополнительной априорной информации некорректна.*

Оказывается, что для любой последовательности решающих правил

$$Ш(N): \operatorname{Cap}(E^N, \mathcal{A}^{*N}) \to \operatorname{Cap}(\operatorname{Cap}(E, \mathcal{A}^*), \mathcal{K}(\mathcal{A}^*))$$

существует такой вероятностный закон $P \in \operatorname{Cap}(E, \mathcal{A}^*)$, что

$$\lim_{N \to \infty} \mathcal{R}_{Ш(N)}(P) \geqslant 1. \quad (4)$$

Наряду с сильной на совокупности $\operatorname{Cap}(E, \mathcal{A}^*)$ и ей подобных существуют и слабые метрики, выражающиеся через функции распределения. Простейшая из них задается С-расстоянием

$$\rho(P, Q) = \sup_x |F(x) - G(x)|,$$

где $F(x) = P\{[0, x)\}$, $G(x) = Q\{[0, x)\}$. По известной теореме Гливенко [4] при $N \to \infty$

$$\sup_x |F_N^*(x) - F(x)| \to 0$$

почти наверно, так что эмпирическая функция распределения F_N^* является состоятельной оценкой для теоретической F. Таким образом, некорректность обратной задачи теории вероятностей не слишком сильна — примерно такая же, как и у операции численного дифференцирования в вещественном анализе.

Мы не будем здесь выяснять, насколько приближения в слабой метрике менее информативны приближений в сильной метрике, отсылая за этим читателя к [5]. Ограничимся примером. Как известно, всякая мера на прямой может быть разложена в линейную комбинацию мер трех типов — дискретной, непрерывной и сингулярной. Мера первого типа сосредоточена на не более чем счетном множестве точек, имеющих каждая положитель-

ную меру, мера второго типа имеет плотность (по лебеговой длине), а мера третьего типа, не имея точек положительной меры, сосредоточена на каком-то подмножестве лебеговой длины нуль. Эти три типа мер резко различаются по виду функции распределения. Разделены они между собой и в сильной метрике — вариация разности двух мер разных типов равна 2. А в слабых метриках непрерывный и сингулярный типы не разделены. И несколько препарируя доказательство теоремы 1, можно показать, что не существует состоятельных решающих правил, позволяющих по возрастающей выборке определить, непрерывен или сингулярен «наблюдаемый» закон P. Любопытно отметить, что отличить дискретный тип от остальных или выделить дискретную компоненту можно, только обладая способностью устанавливать абсолютно точное совпадение двух результатов эксперимента, и невозможно в противном случае.

Из приведенных соображений нетрудно вывести, что максимальным семейством, для которого задача с. т. о. может иметь смысл в сильной метрике, является подсовокупность всех доминированных мер Capd $(\Omega, \mathscr{A}, Z)$ на каком-либо измеримом пространстве $(\Omega, \mathscr{A})$ с фиксированным идеалом Z нуль-множеств, в частности подсовокупность всех распределений вероятностей на вещественной прямой или единичном отрезке, имеющих плотность (по лебеговой длине). Для таких подсовокупностей расстояние по вариации превращается в обычное L_1-расстояние между плотностями распределений:

$$|Q - P| = \int_{\Omega} |q(\omega) - p(\omega)| \, \mu\{d\omega\}, \tag{5}$$

где μ — доминирующая мера, например $\mu\{dx\} = dx$ для единичного отрезка, а оценка вероятностного закона сводится к оценке его плотности. Корректность такой постановки задачи с. т. о. была установлена в 1976 г. независимо Абу-Жаудом [6] (его результаты приведены Девроем и Дьёрфи) и Э. А. Надарая [7]. Первый в качестве оценки плотности использовал гистограммы, а второй — оценки типа ядра. Из инвариантности подсовокупности всех распределений вероятностей непрерывного типа относительно взаимно измеримых сохраняющих длину отображений отрезка на себя легко вытекает, см. [3], что универсальное состоятельное решающее правило не может быть на этой подсовокупности равномерно состоятельным. Однако для более узких априорных семейств $\mathscr{P}$ законов те же решающие процедуры могут приводить к равномерно на всем семействе $\mathscr{P}$ убывающему риску.

Дело теоретика — искать в первую очередь наиболее общие закономерности, наиболее общие подходы. Попытаемся сейчас выяснить, чем же L_2-подход в этом отношении хуже L_1-подхода.

Каждое статистическое решающее правило $Ш$ задает аффинное отображение совокупности Сар $(\Omega^N, \mathscr{A}^N)$ в Сар $(\Delta, \mathscr{B})$. Это связано с тем и только тем, что правило $Ш$ задается переходным распределением вероятностей. Рассмотрим все возможные объекты Сар $(\Omega, \mathscr{A})$ и все возможные переходные распределения вероятностей $Ш$ из какого-либо одного измеримого пространства $(\Omega, \mathscr{A})$ в какое-то другое $(\Omega', \mathscr{A}')$. Они образуют алгебраическую категорию САР, объектами которой служат совокупности Сар $(\Omega, \mathscr{A})$, а морфизмы (здесь это марковские отображения) задаются переходными распределениями вероятностей по правилу

$$Q\{\cdot\} = (PШ)\{\cdot\} = \int_{\Omega} P\{d\omega\}\, Ш(\omega; \cdot).$$

При этом выполнены все аксиомы категории:

1°. Тождественное отображение $И$ каждого объекта на себя принадлежит категории, $И(\omega; B) = \chi_B(\omega)$, где $\chi_B(\cdot)$ — индикатор множества B.

2°. Композиция $ШЖ$ двух марковских отображений снова есть марковское отображение П:

$$П(\omega; \cdot) = \int_{\Omega'} Ш(\omega;\ d\omega')\, Ж(\omega'; \cdot).$$

3°. Композиция является ассоциативной операцией, $Ш(ЖЦ) = (ШЖ)Ц$, как композиция отображений.

Этот факт был впервые отмечен Н. Н. Ченцовым [8] и независимо Н. Морсом и Р. Сакстедером [9]. Кроме того, в этой категории САР имеется дополнительная операция умножения объектов и связанные с ней операции умножения и усреднения морфизмов, [10]. Отсюда нетрудно установить, см. [10], [11], что два семейства распределений вероятностей $\{P_\theta,\ \theta \in \Theta\}$ и $\{Q_\theta,\ \theta \in \Theta\}$, координатизированные одним и тем же множеством Θ, обладают одинаковыми статистическими свойствами тогда и только тогда, когда существуют такие два марковских морфизма $Ш$ и $Ж$, что

$$P_\theta Ш = Q_\theta, \quad Q_\theta Ж = P_\theta, \quad \forall \theta \in \Theta.$$

Во всякой теории общий закон должен допускать эквивариантную формулировку. Другими словами, его утверждение не должно изменяться при переходе от одной ситуации к другой, эквивалентной первой (эквивалентной в рамках этой теории), иначе он не будет общим законом. В классической геометрии такие переходы образуют группу, в теории статистического вывода — категорию, порождающую своеобразную геометрию. Семейства распределений играют в ней роль «фигур», а марковские морфизмы — «движений». И многие основные понятия математи-

ческой статистики получают истолкования как инварианты, ковариантны или более сложные эквивариантны этой геометрии. Так как марковские морфизмы, вообще говоря, необратимы, кроме обычного для групповой геометрии понятия инварианта — величины, принимающей одинаковые значения для конгруэнтных фигур (т. е. статистически эквивалентных семейств), возникает понятие монотонного инварианта. Нам оно понадобится только для случая пары вероятностных законов, определение для случая более обширных семейств совершенно аналогично.

Определение. Функцию $f(P, Q)$ от двух аргументов, определенную на квадратах всех объектов Cap $(\Omega, \mathcal{A})$, назовем *монотонным инвариантом*, если

$$f(PШ, QШ) \leqslant f(P, Q) \tag{6}$$

при всех допустимых наборах P, Q, $Ш$ аргументов.

Примерами монотонных инвариантов от пары распределений могут служить вариация разности $|P - Q|$, см. (3), относительная энтропия

$$\mathcal{Э}(P, Q) = \int_{\Omega} \left[\frac{dQ_s}{dP}(\omega) \ln \frac{dQ}{dP}(\omega)\right] P\{d\omega\} = \int_{\Omega} \left[\ln \frac{dQ}{dP}(\omega)\right] Q\{d\omega\},$$

расстояние Бхаттачария

$$s(P, Q) = 2 \arccos \int_{\Omega} \sqrt{P\{d\omega\} Q\{d\omega\}}, \tag{7}$$

порождаемое квадратичной формой Фишера (фишеровской информацией). Кстати, это расстояние — единственная с точностью до постоянного множителя инвариантная риманова метрика на объектах категории CAP. И говоря о гладких конечномерных семействах $\{P_\theta, \theta \in \Theta\}$, мы почти всегда молчаливо предполагаем, что функция P_θ аргумента θ в области $\Theta \subseteq R^n$ дифференцируема в смысле этой метрики. К классу монотонных инвариантов относятся, как нетрудно установить, и φ-дивергенции Чисара [12, 13]

$$I_\varphi(P, Q) = \int \varphi\left(\frac{dP}{dQ}(\omega)\right) Q\{d\omega\},$$

где φ — выпуклая функция на R^+. Для взаимно абсолютно непрерывных мер определение энтропии упрощается до

$$\mathcal{Э}(P, Q) = \int_{\Omega} \left[\ln \frac{q(\omega)}{p(\omega)}\right] q(\omega)\, \mu\{d\omega\}, \tag{8}$$

а форма Фишера на семействе $\{P_\theta\}$ записывается как

$$ds^2 = \sum_{\alpha, \beta} d\theta_\alpha d\theta_\beta \int_{\Omega} \frac{\partial \ln p(\omega; \theta)}{\partial \theta_\alpha} \cdot \frac{\partial \ln p(\omega; \theta)}{\partial \theta_\beta} P_\theta\{d\omega\}. \tag{9}$$

Исключительная роль L_1-нормы в классе инвариантных метрик устанавливается следующей теоремой, см. [14].

Теорема 2. *Если метрика* ρ *на объектах категории* CAP *монотонна в категории, то*

$$\rho(P, Q) \geqslant \frac{1}{8}\rho(R_{1/2}, R_{1/4}) \cdot |P - Q|, \tag{10}$$

где R_θ *— распределения вероятностей на* $\Omega_2 = \{\omega_1, \omega_2\}$,

$$R_\theta\{\omega_1\} = \theta, \quad R_\theta\{\omega_2\} = 1 - \theta, \quad 0 \leqslant \theta \leqslant 1.$$

Таким образом, если статистическая задача некорректна в расстоянии по вариации, то она остается некорректной и при измерении величины погрешности любой другой инвариантной метрикой (обратное неверно). В то же время, см. [2], расстояние по вариации является единственным (с точностью до множителя) инвариантным расстоянием, определяемым нормой разности. Точнее если метрика ρ на объектах категории CAP инвариантна относительно категории и однородна, т. е.

$$P - Q = \lambda(P' - Q') \Rightarrow \rho(P, Q) = \lambda \cdot \rho(P', Q'),$$

то

$$\rho(P, Q) = \frac{1}{2}\rho(R_{1/2}, R_{1/4}) \cdot |P - Q|.$$

Сходная ситуация имеет место и для функций потерь самого общего вида (для φ-дивергенций это было доказано Чисаром), лишь только бы они являлись монотонными инвариантами, см. [14]:

Теорема 3. *Для монотонного инварианта* $\mathscr{L}(P, Q)$ *со свойством*

$$Q \neq P \Rightarrow \mathscr{L}(P, Q) \neq \mathscr{L}(P, P)$$

существуют константа $c = \mathscr{L}(P, P)$, $\forall P$, *и монотонная числовая функция*

$$g(z) = \inf_{2 \,|\, x-y \,| \,> z} \rho(R_x, R_y), \tag{11}$$

где $0 \leqslant z \leqslant 2$, $g(z) > 0$ *при* $z > 0$, $g(0) = 0$, *такие, что*

$$\mathscr{L}(P, Q) \geqslant c + g(|P - Q|).$$

Отсюда вытекает, что L_1-подход, излагаемый авторами этой книги, является наиболее общим инвариантным подходом в статистической задаче оценки плотности, что мы и обещали строго доказать.

Что же касается L_2-подхода, который также рассматривают Деврой и Дьёрфи, то он неинвариантен, поскольку квадрат нормы

$$\int_{\Omega} [p(\omega) - q(\omega)]^2 \mu \{d\omega\} \qquad (12)$$

уже зависит от выбора доминирующей меры μ. И вполне может случиться, что при замене меры конечная норма разности превратится в бесконечную (в то же время характеристики (5) и (8) близости двух вероятностных законов инвариантны относительно этого выбора). Тем не менее для некоторых классов априорных семейств $\mathscr{P}$ удается развить квазиинвариантный подход, использующий целый «букет» эквивалентных L_2 (P)-норм. Этот подход будет указан несколько далее.

Несмотря на неинвариантность, L_2-подход удобен тем, что позволяет строить простые алгоритмы восстановления неизвестной плотности по независимым наблюдениям и легко характеризовать их точность в выбранной L_2-норме. С ним, например, естественно связаны так называемые *проекционные оценки плотности*, см. гл. 12, предложенные впервые нами в 1958 г. и исследованные затем в [15], [16]. Как показали эксперименты В. В. Статулявичюса [17], при их использовании обработка данных может идти на порядок быстрее, чем при использовании ядерных оценок Розенблата—Парзена с фиксированной формой ядра, а результаты обработки требуют на порядок меньшего объема памяти для хранения. Но они, как и всякие ядерные методы со знакопеременными ядрами, обладают существенным недостатком. Отрезок ряда Фурье с подставленными оценками коэффициентов может не оказаться неотрицательной функцией, т. е. построенная проекционная оценка не будет плотностью вероятностной меры. Разумеется, можно заменить ее на участках отрицательности нулем и отнормировать к единичному интегралу. Как показывают Деврой и Дьёрфи, см. с. 277, L_1-погрешность при этом только уменьшится. Но при такой дополнительной процедуре теряются и простота конструкции и простота хранения. Возникает естественное желание построить метод, лишенный указанного недостатка, достаточно точный и к тому же эквивариантный. Заметим, что требование точности является существенным. Как известно, полигон частот, построенный по гистограмме сходится по вероятности к искомой гладкой плотности класса $C^{(2)}$ со скоростью $N^{-2/5}$. Таков же порядок погрешности ядерных оценок с неотрицательными ядрами ввиду их смещенности.

Отмеченных выше недостатков лишен экспонентный метод оценивания плотности, предложенный Р. Л. Стратоновичем [18] и независимо автором этих строк [2]. Идея состоит в том, чтобы приближать линейным агрегатом не саму плотность, а ее лога-

рифм. Поскольку интеграл от функции $p_0(\omega)\exp[s^j q_j(\omega)]$, где по совпадающим верхнему и нижнему индексам предполагается суммирование, не обязан быть тождественно равен единице, эту функцию надо отнормировать, что дает формулы

$$p(\omega;\ \mathbf{s}) = p_0(\omega)\exp[s^j q_j(\omega) - \Psi(\mathbf{s})], \tag{13}$$

$$\Psi(\mathbf{s}) = \ln \int_{\Omega} \exp[s^j q_j(\omega)]\, p_0(\omega)\, \mu\{d\omega\}, \tag{14}$$

где $\psi(\mathbf{s})$ — логарифм нормирующего делителя. Формула (13) имеет смысл, когда интеграл в (14) конечен. Нетрудно показать, что функция $\psi(\mathbf{s})$ выпуклая, так что область $\{\mathbf{s} : \psi(\mathbf{s}) < +\infty\}$ выпукла (хотя и может быть пустой).

Таким образом, мы построили экспонентное семейство γ распределений вероятностей с каноническим параметром $\mathbf{s}$ и областью задания $\mathrm{Dom}\,\gamma = \{\mathbf{s} : \psi(\mathbf{s}) < +\infty\}$. Естественно, что каноническая параметризация определена с точностью до аффинного преобразования параметров и направляющих статистик $q_j(\omega)$.

Экспонентные семейства возникают в очень многих вопросах математической статистики (и статистической физики). За их последовательной теорией мы отсылаем к [2, 19, 20]. Введем вектор-функцию

$$\mathbf{t} = \mathbf{T}(\mathbf{s}) = \int_{\Omega} \mathbf{q}(\omega)\, p(\omega;\ \mathbf{s})\, \mu\{d\omega\}. \tag{15}$$

Хорошо известно, что $\mathbf{T}(\mathbf{s}) = \mathrm{grad}\,\psi(\mathbf{s})$ — взаимно однозначная и аналитическая вектор-функция внутри области $\mathrm{Dom}\,\gamma$, являющаяся преобразованием Лежандра, так что

$$\mathbf{s} = \mathrm{grad}\,[s^j(\mathbf{t})\, t_j - \Psi(\mathbf{s}(\mathbf{t}))].$$

Несколько более сложны связи $\mathbf{s}$ и $\mathbf{t}$ на границе области $\mathrm{Dom}\,\gamma$, см. [2].

Параметр $\mathbf{t}$ мы будем называть *натуральным параметром экспонентного семейства*. Для него существует простая несмещенная статистическая оценка

$$\mathbf{t}^* = N^{-1}[\mathbf{q}(\omega^1) + \cdots + \mathbf{q}(\omega^N)]. \tag{16}$$

Эта оценка эффективна, т. е. неравенство информации для нее обращается в равенство. Отметим, что справедливо и обратное: если параметр семейства допускает эффективную оценку, то это семейство является экспонентным, а параметр — его натуральным параметром. Самое сложное в процедуре оценивания — интерпретация $\mathbf{t}^*$, т. е. вычисление значения $\mathbf{s}^*$ канонического параметра, отвечающего $\mathbf{t}^*$, поскольку плотность вероятностного закона задана нам формулой именно через $\mathbf{s}$. Но эту трудность нельзя считать принципиальной, поскольку решать систему урав-

нений $\operatorname{grad} \Psi(s) = t^*$ надо лишь один раз в конце обработки. В наши дни для этой цели можно эффективно использовать персональный компьютер с универсальной стандартной программой решения таких систем.

Плотность $p_0(\omega)$ определена с точностью до постоянного множителя, а направляющие статистики $q_j(\omega)$ — с точностью до постоянного слагаемого. Если же принять $p_0(\omega) = p(\omega;\ 0)$, $\mathbf{q}(\omega) = \mathbf{q}(\omega;\ 0)$, где $\int \mathbf{q}(\omega;\ 0)\, p(\omega;\ 0)\, d\mu = 0$, то $\Psi(s) = \mathcal{I}(P_s,\ P_0)$, где относительная энтропия определена формулой (8), и

$$p(\omega;\ s) = p(\omega;\ 0) \exp\left[s^i q_j(\omega;\ 0) - \mathcal{I}(P_s,\ P_0)\right]. \qquad (17)$$

Сравним это представление (17) и представление плотности семейства гауссовских законов с единичной матрицей ковариаций и переменным средним s:

$$p(x,\ s) = p(x,\ 0) \exp\left[s^i x_j - (s,\ s)/2\right]. \qquad (18)$$

Из этого сравнения видно, во-первых, [2], что гауссовские семейства являются единственными, в которых канонические параметризации совпадают с натуральными (для примера, в статистической физике натуральный параметр — температура $t = s^{-1}$). И во-вторых, что относительная энтропия является обобщением половины квадрата евклидова расстояния. Аналогия заходит столь далеко, что для относительной энтропии справедлива, см. [21], несимметричная теорема Пифагора.

Теорема 4. *Если* $P_\sigma = \arg\min_s \mathcal{I}(R,\ P_s)$, *где* $\gamma = \{P_s\}$ — *экспонентное семейство*, $\sigma \in \operatorname{Int} \operatorname{Dom} \gamma$, *то*

$$\mathcal{I}(R,\ P_s) = \mathcal{I}(R,\ P_\sigma) + \mathcal{I}(P_\sigma,\ P_s) \quad \forall s \in \operatorname{Dom} \gamma. \qquad (19)$$

Когда направляющие статистики $q_j(\omega;\ 0)$ *в* (17) *ограниченны, дополнительное условие на* σ *излишне.*

В свое время Лаплас предложил измерять потери статистика расстоянием от сделанной им оценки до истинного значения аргумента. Но, как заметил Гаусс, теория сильно упрощается, если считать потери $\mathscr{L}$ квадратичной функцией ошибки оценивания, см. [22]. Для негауссовских вероятностных законов, как мы видим, естественно брать $\mathscr{L} = 2\mathcal{I}(P^*,\ P_\theta)$, а не какие-либо $L_2(\mu)$-нормы, зависящие от выбора μ. Дальнейшая цель теоретика — указывать в задаче с. т. о. разумные априорные семейства с гарантируемой скоростью убывания характеристики погрешности (максимума риска, байесовского риска и т. п.). Отметим, что в отличие от L_1-подхода задача с. т. о. с энтропийной

функцией потерь для всей доминированной подсовокупности Capd $(E, \mathscr{A}^*, Z)$ уже некорректна, см. [3, теорема 6].

Выше мы уже упоминали, что задача статистического оценивания плотности по наблюдениям в самой широкой постановке не является равномерно состоятельной [3]. Построить равномерно сходящиеся алгоритмы ее восстановления удается только для более узких семейств $\mathscr{P}$ априори возможных законов. Проще всего предположить, что

$$C^{-1} \leqslant \frac{dQ}{dP}(\omega) \leqslant C: \quad \forall \omega \in \Omega, \quad \forall P, Q \in \mathscr{P}: \tag{20}$$

при некоторой положительной константе c. Заметим, что это условие является весьма сильным. В частности, все нормы $L_p(R)$ $1 \leqslant p < \infty$, $R \in \mathscr{P}$, задают на $\mathscr{P}$ одну и ту же топологию, все нормы $L_2(R)$ эквивалентны при $R \in \mathscr{P}$. В таких условиях квазиинвариантности мер удается указывать нижние границы для точности восстановления плотности и строить алгоритмы, оптимальные по порядку точности [2, 23]. Однако класс семейств, удовлетворяющих (20), слишком узок. К нему не принадлежит, например, семейство (18) гауссовских законов.

Если же вместо $L_2(R)$-нормы разности в качестве функции потерь принять энтропию, для которой, как известно, всегда

$$\mathscr{I}(P, Q) \leqslant \|P - Q\|_P^2, \tag{21}$$

то разумным оказывается следующее достаточное условие квазиоднородности семейства $\mathscr{P}$, предложенное нами еще в [2]:

$$C^{-1} \leqslant D_{R'}\left[\ln \frac{dP}{dQ}(\omega)\right] \Big/ D_{R''}\left[\ln \frac{dP}{dQ}(\omega)\right] \leqslant C \tag{22}$$

при некоторой фиксированной константе C и любых P, Q, R', $R'' \in \mathscr{P}$. Этому условию локально удовлетворяют, например, семейства, зависящие от векторного параметра, регулярно гладкие в смысле Крамера—Рао. Наметим для них сейчас, следуя [2], основные результаты теории оптимального оценивания.

Теорема 5. *Пусть* $\{P_\theta, \ \theta \in \Theta\}$ — *компактное гладкое семейство вероятностных мер,* Θ — *замкнутая область в* R^n. *Тогда при* $\mathscr{L}(\theta, P^*) = 2\mathscr{I}(P^*, P)$

$$\lim_{N \to \infty} N \cdot \inf_{Ш(N)} \sup_{\theta \in \Theta} \mathscr{R}_{Ш(N)}(\theta) = \dim \Theta, \tag{23}$$

где не обязательно ограничиваться оценками $P^* \in \{P_\theta\}$, *а асимптотически оптимальной является оценка максимума правдоподобия* П:

$$\sup_{\theta \in \Theta} \left| N \cdot \mathscr{R}_{П(N)}(\theta) - \dim \Theta \right| \underset{N \to \infty}{\to} 0. \tag{24}$$

Впрочем, и саму оценку максимума правдоподобия можно истолковать как оценку минимума относительной энтропии, см. [24, 25], также [2].

Комбинируя теорему 5 с теоремой 4, мы развили в [2] для квазиоднородных семейств с правильно убывающими информационными поперечниками теорию почти оптимального «непараметрического» (точнее счетно-параметрического) оценивания плотности. Как мы видели, этот наш подход близок к L_2-подходу, даже совпадает с ним на этапе построения средних значений от базисных (направляющих) статистик $q_j(\omega)$, но отличается в «интерпретации» построенных средних (и только при построении гистограммы оба подхода дают совпадающие результаты). Заинтересовавшийся читатель может ознакомиться с деталями по монографии [2]. Нам кажется, что благодаря своей эквивариантности экспонентный метод оценивания плотности не менее перспективен, чем ставшие традиционными, см, например, монографию [26], метод сглаживающего ядра или метод ортогонального разложения, рассмотренные в книге Девроя и Дьёрфи.

ЛИТЕРАТУРА

1. Wald A. Contributions to the theory of statistical estimation and testing hypotheses//Ann. Math. Statist. — 1939. — V. 10. — 4. — P. 299—326.
2. Ченцов Н. Н. Статистические решающие правила и оптимальные выводы. — М.: Наука, 1972.
3. Ченцов Н. Н. О корректности задачи статистического точечного оценивания//Теория вероятн. примен. — 1981. — Т. 26. — 1. С. 15—31. Препр. Ин-та прикл. матем. им. М. В. Келдыша АН СССР. — 1979. — 193.
4. Гливенко В. И. Sulla determinatione empirica di una legge di distribytione// Giorn. dell'Istituto Italiano degli Attuari. — 1933. — V. 4. — 1. — P. 83—91.
5. Соболь И. М. Многомерные квадратурные формулы и функции Хаара. — М.: Наука, 1969.
6. Abou-Jaoude S. Conditions nécessaires et suffisantes de convergence L_1 en probabilité de l'histogramme pour une densité//Ann. de l'Inst. H. Poincaré. — 1976. — Ser. B. — V. 12. — P. 213—231.
7. Надарая Э. А. О непараметрической оценке байесовского риска в задаче классификации//Сообщ. АН Груз. ССР. — 1976. — Т. 82. — 2. — С. 277—280.
8. Ченцов Н. Н. Категории математической статистики//Докл. АН СССР. — 1965. — Т. 164. — 3, 511—514.
9. Morse N., Sacksteder R. Statistical isomorphism//Ann. Math. Statist. — 1966. — V. 37. — 1. — P. 203—214. (Имеется русский перевод: Морзе Н., Сакстедер Р. Статистический изоморфизм//Сб. «Математика». — 1968. — Т. 12. — 6. — С. 147—160.)
10. Čencov N. N. Algebraic foundations of mathematical statistics//Math. Operationsforsch. Statist., ser. Statistics. — 1978. — 9. — 2. — P. 267—276.
11. Čencov N. N. On basic concepts of mathematical statistics//Banach Center Publ. — 1980. — 6. — P. 85—94.
12. Csiszar I. Information-type indices of the divergence of distributions. I, II// Magyar Tud. Akad. Mat. — 1967. — V. 17. — 2. — P. 123—149. — 3. — P. 267—291.

13. Csiszar I. On topological properties of f-divergence//Studia Sci. Math. Hungar. — 1967. — V. 2. — 3—4. — P. 329—339.
14. Morozova E. A., Čencov N. N. Markov maps in noncommutative probability theory and mathematical statistics//Proc. 4th Intern. Vilnius Conf. Probability Theory Mathem. Statistics (1985). — V. 2. — VNU Science Press.— 1987.— P. 287 — 310.
15. Ченцов Н. Н. Оценка неизвестной плотности распределения по наблюдениям//Докл. АН СССР. — 1962. — Т. 147. — 1. — С. 45—48.
16. Фролов А. С., Ченцов Н. Н. Использование зависимых испытаний в методе Монте Карло для получения гладких кривых//Тр. 6-го Всесоюзн. совещ. по теории вероятн. матем. статистике (Вильнюс, 1960). — Вильнюс, 1962. — С. 425—437.
17. Статулявичюс В. В. Некоторые свойства асимптотически минимаксных статистических оценок плотности распределения//Дисс. к. ф.—м. н., Вильнюсский гос. университет им. В. Капсукаса. — 1986. — 109 с.
18. Стратонович Р. Л. Быстрота сходимости алгоритмов оценки плотности распределения вероятностей//Изв. АН СССР, сер. техн. киберн. — 1969. — 6. — 1. — С. 3—15.
19. Ченцов Н. Н. К систематической теории экспоненциальных семейств распределений вероятностей//Теор. вероятн. примен. — 1966. — 11. — 3. — С. 483—494.
20. Barndorff-Nielsen O. Information and Exponential Families in Statistical theory. — Wiley, 1978.
21. Ченцов Н. Н. Несимметричное расстояние между распределениями вероятностей, энтропия и теорема Пифагора//Матем. заметки. — 1968. — 4. — 3. — С. 323—332.
22. LeCam L. On some asymptotic properties of maximum likelihood estimates and related Bayes estimates//Univ. Calif. Publ. Statistics. — 1953. — 1. — P. 267—329. (Имеется русский перевод: ЛеКам Л. О некоторых асимптотических свойствах оценок максимального правдоподобия и относительных байесовских оценках//Сб. «Математика». — 1960. — Т. 4. — 2. — С. 69—119.)
23. Ибрагимов И. А., Хасьминский Р. З. Об оценке плотности распределения// Сб. Исследования по математической статистике, IV. — Записки научн. семинаров ЛОМИ. — 1980. — Т. 98. — С. 61—86.
24. Hartigan J. A. The likelihood and invariance principles//Ann. Math. Statist. — 1967. — Т. 38. — P. 533—539.
25. Kriz T. A., Talacko J. V. Equivalence of the maximum likelihood estimator to a minimum entropy estimator//Trab. Estadist. Invest. Oper. — 1968. — Т. 19. — 1—2. — P. 55—65.
26. Надарая Э. А. Непараметрическое оценивание плотности вероятностей и кривой регрессии. — Тбилиси: ТГУ, 1983.

Дополнение 2

Принципы оценивания плотности распределения вероятностей

В. Н. Вапник

Говорят, что всякая проблема включает в себя три элемента: постановку проблемы, ее разрешение и анализ качества разрешения проблемы. В задаче оценивания плотности вероятностей эта структура имеет особенности, которые могут вызвать неудовлетворенность исследователя: слишком много существует способов разрешения проблемы оценивания плотности (метод ядерных функций, проекционные методы, метод гистограмм, метод максимума правдоподобия со штрафными функциями, метод решета и др.) и нет оснований думать, что все возможные методы оценивания исчерпаны.

В таких случаях уместен вопрос: «А не существует ли общего принципа построения оценок, следуя которому можно получить известные методы оценивания и, быть может, конструировать новые?.»

Попытке ответить на этот вопрос и посвящено настоящее дополнение. В нем мы рассмотрим две постановки задачи оценивания плотности вероятностей: косвенную и прямую, и покажем, что в рамках каждой из них на основе одной и той же системы идей могут быть получены различные методы оценивания плотности.

1. Косвенная постановка задачи оценивания плотности вероятностей

Косвенная постановка задачи оценивания плотности вероятностей восходит к схеме, предложенной Р. Фишером [1] и впервые в достаточно общем случае исследованной А. Вальдом [2].

Пусть задано множество плотностей [1]) $p(x, \alpha)$, $\alpha \in \Lambda$, содержащее искомую плотность $p(x, \alpha_0)$. Рассмотрим функционал

$$H_{\alpha_0}(\alpha) = -\int \ln p(x, \alpha)\, p(x, \alpha_0)\, dx, \qquad (1)$$

определенный на функциях этого множества.

---

[1]) Для простоты мы ограничимся случаем $x \in X \subset R^1$. Однако все приведенные здесь результаты переносятся и на случай $X \subset R^n$.

Известно, что минимум функционала (1) (пусть минимум существует) достигается на функциях $p(x, \alpha^{\wedge})$, которые могут отличаться от $p(x, \alpha_0)$ разве лишь на множестве меры нуль.

Более того, известно (см. теорему 8.2 книги), что

$$\int |p(x, \alpha^*) - p(x, \alpha_0)|\, dx \leqslant \sqrt{2(H_{\alpha_0}(\alpha^*) - H_{\alpha_0}(\alpha_0))}. \quad (2)$$

Поэтому функции $p(x, \alpha^*)$, доставляющие функционалу (1) значения, ε-близкие к минимальному, будут $\sqrt{2\varepsilon}$-близки в метрике L_1 к оцениваемой плотности.

Таким образом, задача оценивания плотности $p(x, \alpha_0)$ (в случае когда для нее существует энтропия Шеннона $H_{\alpha_0}(\alpha_0) = -\int \ln p(x, \alpha_0)\, p(x, \alpha_0)\, dx < \infty$) эквивалентна задаче минимизации функционала (1). Однако в нашей постановке функционал (1) не определен явно: плотность $p(x, \alpha_0)$ неизвестна, но задана случайная независимая выборка

$$x_1, \ldots, x_l, \quad (3)$$

распределенная согласно $p(x, \alpha_0)$.

Проблема как раз и заключается в том, чтобы в этих условиях минимизировать функционал (1), т. е. предложить метод построения последовательности функций $p(x, \alpha_l)$, $\alpha_l = \alpha(x_1, \ldots, x_l)$, обеспечивающий сходимость

$$H_{\alpha_0}(\alpha_l) - H_{\alpha_0}(\alpha_0) = -\int \ln \frac{p(x, \alpha_l)}{p(x, \alpha_0)}\, p(x, \alpha_0)\, dx \xrightarrow[l\to 0]{p} 0. \quad (4)$$

Функционал $-\int \ln \frac{p(x, \alpha_l)}{p(x, \alpha_0)}\, p(x, \alpha_0)\, dx$ получил название *метрики Кульбака* (согласно (2), сходимость в метрике Кульбака обеспечивает сходимость в L_1).

Р. Фишер указал на следующий путь минимизации функционала (1) по эмпирическим данным (3). Он предложил в качестве искомой оценки выбирать плотности, минимизирующие на заданном множестве $p(x, \alpha)$, $\alpha \in \Lambda$ эмпирический функционал

$$H_э(\alpha) = -\frac{1}{l} \sum_{i=1}^{l} \ln p(x_i, \alpha), \quad (5)$$

построенный по выборке (3).

Как мы увидим ниже, такой путь разрешения проблемы минимизации функционала (1) (он получил название *метода максимума правдоподобия*) является не самым удачным: он гарантирует состоятельность оценки лишь для достаточно узких классов функций. Ниже мы остановимся на условиях состоятельности метода максимума правдоподобия и рассмотрим его обобщение. Сейчас же для нас важно, что задача оценивания плотности была сведена

к другой задаче — минимизации функционала среднего риска (1) по эмпирическим данным (3), и в этом смысле постановка Фишера—Вальда является косвенной в отличие от возможной прямой. Однако, прежде чем перейти к прямой постановке задачи оценивания плотности вероятностей, заметим, что, используя параметрическую форму записи множества $p(x, \alpha)$, $\alpha \in \Lambda$, мы никаких ограничений на множество параметров Λ не налагаем: оно может быть как множеством конечномерных векторов, так и множеством абстрактных элементов.

2. Прямая постановка задачи оценивания плотности вероятностей

Прямая постановка задачи оценивания плотности вероятностей сформулирована сравнительно недавно [3].

Согласно определению, плотность вероятностей есть решение интегралъного уравнения

$$\int_a^b \theta(x - x')\, p(x', \alpha)\, dx' = P_{\alpha_0}(x), \tag{6}$$

$$\theta(z) = \begin{cases} 0, & \text{если } z < 0, \\ 1, & \text{если } z \geqslant 0, \end{cases}$$

определенного на некотором множестве $\mathcal{P}$ функций $p(x, \alpha)$, $\alpha \in \Lambda$, где $P_{\alpha_0}(x)$ — функция распределения вероятностей.

Оценить плотность $p(x, \alpha_0)$ на основе выборки (3) — значит найти приближения $p(x, \alpha_l)$ к решению $p(x, \alpha_0)$ интегрального уравнения (6) в условиях, когда функция распределения вероятностей $P_{\alpha_0}(x)$ неизвестна, но задана выборка, полученная согласно $P_{\alpha_0}(x)$.

Построим на основе выборки (3) эмпирическую функцию распределения

$$F_l(x) = \frac{1}{l} \sum_{i=1}^{l} \theta(x - x_i). \tag{7}$$

Согласно теореме Гливенко—Кантелли с ростом объема выборки l эмпирическая функция распределения сходится п. н. к истинной в метрике C:

$$\sup_x |F_l(x) - P_{\alpha_0}(x)| \xrightarrow[l \to \infty]{\text{п. н.}} 0.$$

Таким образом, проблема состоит в том, чтобы, используя последовательность функций $F_l(x)$, $l = 1, 2, \ldots$, сходящуюся к $P_{\alpha_0}(x)$, найти последовательность функций $p(x, \alpha_l)$, сходящуюся в метрике множества E_1 к решению $p(x, \alpha_0)$ уравнения (6).

Принципиальная трудность решения уравнения (6) на основе последовательности эмпирических функций $F_l(x)$, $l = 1, 2, \ldots$, состоит в том, что для сколько-нибудь широкого класса функций $p(x, \alpha)$, $\alpha \in \Lambda$, отыскание решения уравнения (6) является некорректно поставленной задачей.

В 60-х годах были найдены различные методы решения некорректно поставленных задач: метод квазирешений В. К. Иванова [4], метод регуляризации А. Н. Тихонова [5], метод «невязки» Д. Филлипса [6]. Все они, как потом выяснилось, идейно близки друг другу. Однако еще раньше многие оценки плотности, являющиеся решением прямой задачи, были предложены разными авторами. В разд. 12 мы укажем на связь хорошо известных методов оценивания плотности с решением интегрального уравнения (6) на основе последовательности функций $F_l(x)$, $l = 1, 2, \ldots$.

3. Состоятельность метода минимизации эмпирического риска

Рассмотрим задачу минимизации среднего риска по эмпирическим данным, охватывающую задачу оценивания плотности вероятностей в косвенной постановке [7]. Пусть на множестве измеримых функций $Q(x, \alpha)$, $\alpha \in \Lambda$, задан функционал среднего риска

$$I(\alpha) = \int Q(x, \alpha)\, dP(x). \tag{8}$$

Требуется минимизировать (8), если вероятностная мера $P(x)$ неизвестна, но задана случайная независимая выборка $x_1, \ldots, x_l$, полученная согласно $P(x)$.

Введем эмпирический функционал (функционал эмпирического риска)

$$I_{\text{э}}(\alpha) = \frac{1}{l} \sum_{i=1}^{l} Q(x_i, \alpha), \tag{9}$$

минимум которого $Q(x, \alpha_l)$ примем за оценку функции $Q(x, \alpha_0)$, минимизирующей средний риск (8). Такой путь минимизации среднего риска называется *методом минимизации эмпирического риска*.

Говорят, что решение задачи минимизации среднего риска на множестве $Q(x, \alpha,)$, $\alpha \in \Lambda$ методом минимизации эмпирического риска является *состоятельным*, если

$$\int Q(x, \alpha_l)\, dP(x) \xrightarrow[l \to \infty]{P} \int Q(x, \alpha_0)\, dP(x)$$

и

$$\frac{1}{l}\sum_{i=1}^{l} Q(x_i, \alpha_l) \xrightarrow[l\to\infty]{P} \int Q(x, \alpha_0)\, dP(x).$$

Однако приведенное определение состоятельности допускает следующую возможность. Пусть на множестве функций $Q(x, \alpha)$, $\alpha \in \Lambda$, метод минимизации эмпирического риска не является состоятельным. Рассмотрим новое множество $Q(x, \alpha)$, $\alpha \in \Lambda^*$, состоящее из множества $Q(x, \alpha)$, $\alpha \in \Lambda$, и еще одной функции $\varphi(x)$, такой, что $\varphi(x) < Q(x, \alpha)$, $\alpha \in \Lambda$, $\forall x \in X$. На этом множестве минимум функционала (9) будет всегда достигаться на функции $\varphi(x)$ и, следовательно, метод минимизации эмпирического риска будет состоятельным. Из приведенного примера следует, что необходимые условия состоятельности должны содержать сравнение функций множества. Для того чтобы построить общую теорию, не содержащую сравнения функций, введем определение строгой состоятельности [8].

Определение. Решение задачи минимизации среднего риска на множестве $Q(x, \alpha)$, $\alpha \in \Lambda$, методом минимизации эмпирического риска является *строго состоятельным*, если на любом подмножестве $Q(x, \alpha)$, $\alpha \in \Lambda(c)$, таком, что

$$\Lambda(c) = \left\{\alpha: \int Q(x, \alpha)\, dP(x) > c \text{ и } \alpha \in \Lambda\right\}, \quad c \in (-\infty, \infty),$$

имеет место сходимость по вероятности

$$\inf_{\alpha\in\Lambda(c)} \frac{1}{l}\sum_{i=1}^{l} Q(x_i, \alpha) \xrightarrow[l\to\infty]{P} \inf_{\alpha\in\Lambda(c)} \int Q(x, \alpha)\, dP(x).$$

Замечание. Из строгой состоятельности метода минимизации эмпирического риска на $Q(x, \alpha)$, $\alpha \in \Lambda$, следует состоятельность метода

$$\int Q(x, \alpha_l)\, dP(x) \xrightarrow[l\to\infty]{P} \int Q(x, \alpha_0)\, dP(x).$$

Для того чтобы сформулировать условия, обеспечивающие строгую состоятельность метода минимизации эмпирического риска, нам понадобятся два понятия.

Определения. На множестве $Q(x, \alpha)$, $\alpha \in \Lambda$, имеет место *равномерная сходимость средних к их математическим ожиданиям*, если

$$\sup_{\alpha\in\Lambda}\left|\int Q(x, \alpha)\, dP(x) - \frac{1}{l}\sum_{i=1}^{l} Q(x_i, \alpha)\right| \xrightarrow[l\to\infty]{P} 0.$$

На множестве функций $Q(x, \alpha)$, $\alpha \in \Lambda$, имеет место *равномерная односторонняя сходимость средних к их математическим ожиданиям*, если

$$\sup_{\alpha \in \Lambda} \left(\int Q(x, \alpha)\, dP(x) - \frac{1}{l} \sum_{i=1}^{l} Q(x_i, \alpha) \right)_+ \xrightarrow[l \to \infty]{P} 0,$$

где

$$(Z)_+ = \begin{cases} z, & \text{если } z \geqslant 0, \\ 0, & \text{если } z < 1. \end{cases}$$

Пусть теперь $Q(x, \alpha)$, $\alpha \in \Lambda$, — множество функций, удовлетворяющих для некоторых $-\infty < a < A < +\infty$ условию

$$a \leqslant \int Q(x, \alpha)\, dP(x) \leqslant A. \tag{10}$$

Теорема 1 [8]. *Для функций, удовлетворяющих* (10), *следующие два утверждения эквивалентны;*

1) *на множестве функций* $Q(x, \alpha)$, $\alpha \in \Lambda$ *метод минимизации эмпирического риска является строго состоятельным;*

2) *на множестве функций* $Q(x, \alpha)$, $\alpha \in \Lambda$, *имеет место равномерная односторонняя сходимость.*

Таким образом, условия строгой состоятельности метода минимизации эмпирического риска полностью определяются условиями равномерной односторонней сходимости.

4. Необходимые и достаточные условия равномерной односторонней сходимости для множества равномерно ограниченных функций

Введем дальнейшие ограничения на множества функций $Q(x, \alpha)$, $\alpha \in \Lambda$: пусть $Q(x, \alpha)$, $\alpha \in \Lambda$, — множество равномерно ограниченных по модулю функций

$$|Q(x, \alpha)| \leqslant B < +\infty,$$

удовлетворяющих подходящим условиям измеримости. Пусть $x_1, \ldots, x_l$ — случайная независимая выборка объема l, полученная согласно неизвестной мере $P(x)$. Поставим в соответствие каждому набору $x_1, \ldots, x_l$ множество l-мерных векторов

$$q(\alpha) = (Q(x_1, \alpha), \ldots, Q(x_l, \alpha)), \quad \alpha \in \Lambda,$$

принадлежащих l-мерному кубу с ребром $2B$.

Определим число элементов минимальной ε-сети множества $q(\alpha)$, $\alpha \in \Lambda$, в метрике C:

$$\rho_C(q_1, q_2) = \sup_{1 \leqslant i \leqslant l} |Q(x_i, \alpha_1) - Q(x_i, \alpha_2)|,$$

которое будем обозначать $N^{\Lambda}(\varepsilon; x_1, \ldots, x_l)$.

Величину $H^{\Lambda}(\varepsilon; x_1, \ldots, x_l) = \ln N^{\Lambda}(\varepsilon; x_1, \ldots, x_l)$ называют *энтропией множества* $Q(x, \alpha)$, $\alpha \in \Lambda$, на выборке $x_1, \ldots, x_l$, а величину

$$H^{\Lambda}(\varepsilon, l) = E \ln N^{\Lambda}(\varepsilon; x_1, \ldots, x_l) \tag{11}$$

средней энтропией множества функций $Q(x, \alpha)$, $\alpha \in \Lambda$, на выборках объема l. В определении (11) математическое ожидание берется по мере $P^l(x)$.

В следующих теоремах устанавливается, что условия равномерной сходимости средних к их математическим ожиданиям на $Q(x, \alpha)$, $x \in \Lambda$, определяются особенностями асимптотического (по l) поведения функции $H^{\Lambda}(\varepsilon, l)$.

Теорема 2 [9]. *Для того чтобы на множестве равномерно ограниченных по модулю функций $Q(x, \alpha)$, $\alpha \in \Lambda$, имела место равномерная сходимость средних к их математическим ожиданиям, необходимо и достаточно, чтобы для любого $\varepsilon > 0$ выполнялось равенство*

$$\lim_{l \to \infty} \frac{H^{\Lambda}(\varepsilon, l)}{l} = 0$$

(т. е. средняя энтропия на элемент выборки стремится к нулю с ростом l).

Теорема 3 [8]. *Для того чтобы на множестве равномерно ограниченных по модулю функций $Q(x, \alpha)$, $\alpha \in \Lambda$, имела место равномерная односторонняя сходимость, необходимо и достаточно, чтобы для любых положительных ε, δ и η нашлось множество функций $R(x, \beta)$, $\beta \in \mathscr{B}$, удовлетворяющее двум условиям:*

1) для всякой функции $Q(x, \alpha^*)$ существует функция $R(x, \beta^*)$, такая, что

$$Q(x, \alpha^*) \geqslant R(x, \beta^*), \quad \int (Q(x, \alpha^*) - R(x, \beta^*))\, dP(x) < \delta;$$

2) для любого $\varepsilon > 0$ энтропия множества $R(x, \beta)$, $\beta \in \mathscr{B}$ на выборках объема l удовлетворяет условию

$$\lim_{l \to \infty} \frac{H^{\mathscr{B}}(\varepsilon, l)}{l} \leqslant \eta.$$

Замечание [10]. Если имеет место равномерная сходимость, то она имеет место п. н.

Итак, согласно теореме 1, условия теоремы 3 являются необходимыми и достаточными условиями строгой состоятельности метода минимизации эмпирического риска на множестве ограниченных функций.

Эти условия для заданной вероятностной меры $P(x)$ описывают степень разнообразия множества функций и никак не связаны с особенностями функций, образующих множество. Грубо говоря, теоремы утверждают, что если множество функций «не очень разнообразно», то метод минимизации эмпирического риска является строго состоятельным, в противном случае не является.

5. Достаточные условия равномерной сходимости

Приведенные в предыдущем разделе необходимые и достаточные условия равномерной сходимости средних к их математическим ожиданиям опирались на свойства энтропии, которая конструировалась с помощью неизвестной вероятностной меры $P(x)$. Однако часто необходимо иметь более грубую характеристику разнообразия множества $Q(x, \alpha)$, $\alpha \in \Lambda$, в терминах которой можно было бы установить достаточные условия равномерной сходимости для неизвестной вероятностной меры. Такой характеристикой мог бы служить факт существования конечной ε-сети множества $Q(x, \alpha)$, $\alpha \in \Lambda$, в метрике C. Пусть $N^{\Lambda}(\varepsilon)$ — минимальное число элементов ε-сети множества $Q(x, \alpha)$, $\alpha \in \Lambda$, в метрике C. Тогда для любого l справедливо неравенство

$$N^{\Lambda}(\varepsilon; x_1, \ldots, x_l) \leqslant N^{\Lambda}(\varepsilon)$$

и, следовательно,

$$\frac{H^{\Lambda}(\varepsilon, l)}{l} \leqslant \frac{\ln N^{\Lambda}(\varepsilon)}{l} \xrightarrow[l \to \infty]{} 0,$$

что, согласно теореме 2, влечет за собой существование равномерной сходимости. Более того, для этого случая справедлива

Теорема 4 [7]. *Пусть множество ограниченных по модулю функций* $|Q(x, \alpha)| \leqslant B$, $\alpha \in \Lambda$, *имеет* ε*-сеть в метрике* C, *состоящую из* $N^{\Lambda}(\varepsilon)$ *элементов. Тогда*

$$P\left\{\sup_{\alpha \in \Lambda}\left|\int Q(x, \alpha)\, dP(x) - \frac{1}{l}\sum_{i=1}^{l} Q(x_i, \alpha)\right| > \varepsilon\right\} <$$

$$< 2N^{\Lambda}\left(\frac{\varepsilon}{2}\right) e^{-\frac{\varepsilon^2 l}{4(2B)^2}}. \qquad (12)$$

Из неравенства (12) следует, что

$$P\left\{\left|\int Q(x, \alpha_l)\, dP(x) - \int Q(x, \alpha_0)\, dP(x)\right| > 2\varepsilon\right\} < 2N^{\Lambda}\left(\frac{\varepsilon}{2}\right) e^{-\frac{\varepsilon^2 l}{4(2B)^2}}.$$

Приведенная характеристика позволяет установить существование равномерной сходимости лишь для множеств непрерывных функций. Поэтому рассмотрим другую, весьма общую характеристику разнообразия множества $Q(x, \alpha)$, $\alpha \in \Lambda$ — *емкость множества*. Определим ее сначала для множества $\omega(x, \alpha)$, $\alpha \in \Lambda$, индикаторных функций ($\omega = 0, 1$).

Как и при построении энтропии, поставим в соответствие выборке $x_1, \ldots, x_l$ множество (бинарных) векторов

$$q_\omega(\alpha) = (\omega(x_1, \alpha), \ldots, \omega(x_l, \alpha)), \quad \alpha \in \Lambda.$$

Для множества $q_\omega(\alpha)$, $\alpha \in \Lambda$, количество элементов минимальной ε-сети не зависит от ε и равно числу $\Delta^{\Lambda}(x_1, \ldots, x_l)$ различных элементов множества $q_\omega(\alpha)$, $\alpha \in \Lambda$.

Введем функцию

$$m^{\Lambda}(l) = \max_{x_1, \ldots, x_l} \Delta^{\Lambda}(x_1, \ldots, x_l), \tag{13}$$

которая обладает следующим замечательным свойством.

Теорема 5 [10]. *Функция $m^{\Lambda}(l)$ либо определяется равенством*

$$m^{\Lambda}(l) = 2^l, \tag{14а}$$

либо, если найдется такое $l = h + 1$, что $m^{\Lambda}(h+1) \neq 2^{h+1}$, оценивается неравенством

$$m^{\Lambda}(l) < l^h. \tag{14б}$$

Определения. *Емкость множества индикаторных функций* $\omega(x, \alpha)$, $\alpha \in \Lambda$, равна h, если имеет место оценка (14б), и равна бесконечности, если для всех l имеет место равенство (14а).

Пусть теперь $Q(x, \alpha)$, $\alpha \in \Lambda$, — множество произвольных функций. По всякому $c \in (-\infty, \infty)$ построим множество индикаторных функций

$$\omega_c(x, \alpha) = \theta(Q(x, \alpha) + c), \quad \alpha \in \Lambda.$$

Емкость множества функций $Q(x, \alpha)$, $\alpha \in \Lambda$, равна h, если

$$h = \sup_c h_c,$$

где h_c есть емкость множества индикаторных функций $\omega_c(x, \alpha)$, $\alpha \in \Lambda$.

Теорема 5 дает возможность оценить емкость любого множества. В частности, емкость множества функций, заданных в виде

$$Q(x, \alpha) = \sum_{i=0}^{n-1} \alpha_i \varphi_i(x),$$

где $\varphi_0(x), \ldots, \varphi_{n-1}(x)$ — линейно независимые функции, равна n [10].

Для множества равномерно ограниченных по модулю функций $|Q(x, \alpha)| \leqslant B$, $\alpha \in \Lambda$, имеющего емкость h, справедлива оценка

$$N^{\Lambda}(\varepsilon; x_1, \ldots, x_l) < (l^h)^{\frac{2B}{\varepsilon}}$$

и, следовательно,

$$\frac{H^{\Lambda}(\varepsilon, l)}{l} < \frac{2Bh}{\varepsilon} \frac{\ln l}{l} \underset{l \to \infty}{\longrightarrow} 0.$$

Согласно же теореме 2, условие $\frac{H^{\Lambda}(\varepsilon, l)}{l} \underset{l \to \infty}{\longrightarrow} 0$ влечет за собой равномерную сходимость. Более того, справедлива

Теорема 6 [7]. *Для множества равномерно ограниченных по модулю функций* $|Q(x, \alpha)| \leqslant B$, $\alpha \in \Lambda$, *имеющего емкость* h, *справедливо неравенство*

$$P\left\{\sup_{\alpha \in \Lambda} \left| \int Q(x, \alpha)\, dP(x) - \frac{1}{l} \sum_{i=1}^{l} Q(x_i, \alpha) \right| > \varepsilon \right\} < 6(2l)^h e^{-\frac{\varepsilon^2 l}{4(2B)^2}}. \tag{15}$$

Из неравенства (15) следует, что

$$P\left\{\left| \int Q(x, \alpha_0)\, dP(x) - \int Q(x, \alpha_l)\, dP(x) \right| > 2\varepsilon \right\} < 6(2l)^h e^{-\frac{\varepsilon^2 l}{4(2B)^2}}. \tag{16}$$

6. Принцип структурной минимизации риска

Приведенные в теореме 3 необходимые и достаточные условия строгой состоятельности устанавливают, в частности, что метод минимизации эмпирического риска применим лишь в ограниченных случаях. Поэтому рассмотрим следующее его обобщение. Пусть по-прежнему стоит задача о минимизации функционала среднего риска (8) по эмпирическим данным $x_1, \ldots, x_l$, и пусть наряду с $Q(x, \alpha)$, $\alpha \in \Lambda$, на X задана структура

$$S_1 \subset S_2 \subset \ldots \subset S_n \subset \ldots \tag{17}$$

функций $S_n = \{R(x, \beta) : \beta \in \mathcal{B}_n\}$, такая, что

$$\inf_{\beta \in \mathcal{B}_n} \int R(x, \beta)\, dP(x) \xrightarrow[n\to\infty]{} \int Q(x, \alpha_0)\, dP(x).$$

Зададим закон $n = n(l)$, определяющий номер n элемента S_n структуры (17) в зависимости от объема l выборки $x_1, \ldots, x_l$.

В качестве приближения к искомой функции $Q(x, \alpha_0)$, минимизирующей на $Q(x, \alpha)$, $\alpha \in \Lambda$, средний риск, будем выбирать такую функцию $R(x, \beta_l)$, которая на элементе $S_{n(l)}$ структуры (17) минимизирует эмпирический риск

$$I_{\text{э}}(\beta) = \frac{1}{l} \sum_{i=1}^{l} R(x_i, \beta)$$

(т. е. $R(x, \beta_l) = \arg\min_{\beta \in \mathcal{B}_n} \sum_{i=1}^{l} R(x_i, \beta)$). Такой метод минимизации среднего риска назовем *методом структурной минимизации риска* [1] [8,10].

Будем требовать, чтобы каждое из множеств $R(x, \beta)$, $\beta \in \mathcal{B}_n$, удовлетворяло достаточным условиям равномерной сходимости, приведенным в предыдущем разделе. В этом случае можно указать такой закон $n = n(l)$, для которого метод структурной минимизации риска будет состоятельным.

В частности, если потребовать, чтобы каждый элемент S_n структуры (17) удовлетворял условиям теоремы 4:

1) функции из $R(x, \beta)$, $\beta \in \mathcal{B}_n$, ограничены по модулю величиной B_n $(B_{n-1} \leqslant B_n)$;

2) число элементов минимальной ε-сети в метрике C множества $R(x, \beta)$, $\beta \in \mathcal{B}_n$, не превосходит $N_n^{\Lambda}(\varepsilon)$ $\left(N_{n-1}^{\Lambda}(\varepsilon) \leqslant N_n^{\Lambda}(\varepsilon)\right)$, то состоятельность (фактически сходимость п. н.) метода структурной минимизации риска обеспечит такой закон $n = n(l)$, для которого при любом $\varepsilon > 0$ будет выполнено равенство

$$\lim_{l\to\infty} \frac{B_{n(l)}^2 \left(\ln N_{n(l)}^{\Lambda}(\varepsilon) + \ln l\right)}{l} = 0. \qquad (18)$$

Используя теорему 4, можно оценить и асимптотическую скорость сходимости $\int R(x, \beta_l)\, dP(x) \xrightarrow[l\to\infty]{\text{п. н.}} \int Q(x, \alpha_0)\, dP(x)$.

Пусть теперь каждый элемент S_n структуры (17) удовлетворяет условиям теоремы 6:

1) функции $R(x, \beta)$, $\beta \in \mathcal{B}_n$, ограничены по модулю величиной B_n $(B_{n-1} \leqslant B_n)$;

---

[1] В этом названии подчеркивается приоритет определения подходящего элемента структуры перед определением функции, минимизирующей эмпирический риск. В книге аналогичный метод называется *методом решета*.

2) емкость множества функций $R(x, \beta)$, $\beta \in \mathscr{B}_n$, равна h_n ($h_{n-1} \leqslant h_n$).

Для такой структуры справедлива

Теорема 7 [7]. Метод структурной минимизации риска обеспечивает сходимость п. н. с асимптотической скоростью, имеющей порядок

$$V(l) = \left(\inf_{\beta \in \mathscr{B}_{n(l)}} \int R(x, \beta)\, dP(x) - \int Q(x, \alpha_0)\, dP(x)\right) + \\ + \sqrt{\frac{B^2_{n(l)} h_{n(l)} \ln l}{l}} \tag{19}$$

$\left(\text{т. е. } P\left\{\lim_{l\to\infty} V^{-1}(l)\left(\int R(x, \beta_l)\, dP(x) - \int Q(x, \alpha_0)\, dP(x)\right) = \right.\right.$

$\left.\left. = \text{const}\right\} = 1\right)$, если закон $n = n(l)$ таков, что

$$\frac{B^2_{n(l)} h_{n(l)} \ln l}{l} \xrightarrow[l\to\infty]{} 0.$$

Заметим, что скорость сходимости (19) представлена суммой двух слагаемых. Первое слагаемое

$$r_{n(l)} = \inf_{\beta \in \mathscr{B}_{n(l)}} \int R(x, \beta)\, dP(x) - \int Q(x, \alpha_0)\, dP(x)$$

определяет величину уклонения среднего риска, доставляемого искомым решением, от среднего риска, доставляемого наилучшим в S_n приближением (чем больше $n = n(l)$, тем меньше уклонение). Второе слагаемое

$$\Delta_{n(l)} = \sqrt{\frac{B^2_{n(l)} h_{(n)\,l} \ln l}{l}}$$

оценивает величину уклонения среднего риска наилучшего в S_n приближения от среднего риска, доставляемого выбранной функцией (чем меньше $n = n(l)$, тем меньше уклонение). Таким образом, скорость сходимости определяется двумя противоречащими друг другу требованиями к закону $n = n(l)$.

7. Условия состоятельности метода максимума правдоподобия

Используем приведенные в разд. 5 и 6 теоремы для выяснения условий состоятельности метода максимума правдоподобия.

Определения. Метод максимума правдоподобия является *состоятельным методом оценивания* плотности из $p(x, \alpha)$, $\alpha \in \Lambda$,

в метрике $\rho(p_1, p_2)$, если для любого $p(x, \alpha_0)$, $\alpha_0 \in \Lambda$, справедливо

$$\rho(p(x, \alpha_l), p(x, \alpha_0)) \xrightarrow[l\to\infty]{p_0} 0.$$

Метод максимума правдоподобия является *строго состоятельным методом оценивания* плотностей множества $p(x, \alpha)$, $\alpha \in \Lambda$, если для любой плотности $p(x, \alpha_0)$, $\alpha_0 \in \Lambda$, имеет место сходимость по вероятности

$$\inf_{\alpha\in\Lambda}\left(-\frac{1}{l}\sum_{i=1}^{l}\ln p(x_i, \alpha)\right) \xrightarrow[l\to\infty]{p_0} -\int \ln p(x, \alpha_0)\, p(x, \alpha_0)\, dx.$$

Замечание. Если метод максимума правдоподобия является строго состоятельным методом оценивания плотностей из $p(x, \alpha)$, $\alpha \in \Lambda$, то он является состоятельным методом оценивания в метрике Кульбака

$$-\int \ln \frac{p(x, \alpha_l)}{p(x, \alpha_0)}\, p(x, \alpha_0)\, dx \xrightarrow[l\to\infty]{p} 0.$$

Изучение условий строгой состоятельности метода максимума правдоподобия мы начнем с частного случая: класс плотностей $p(x, \alpha)$, $\alpha \in \Lambda$, равномерно ограничен и равномерно отделим от нуля:

$$0 < a \leqslant p(x, \alpha) \leqslant A < \infty \tag{20}$$

(a и A не зависят от x и α). С помощью теоремы 3 для заданного класса плотностей (20) устанавливается

Теорема 8 [8]. *Для того, чтобы на множестве равномерно ограниченных и равномерно отделимых от нуля плотностей метод максимума правдоподобия был строго состоятельным, необходимо и достаточно, чтобы на множестве (равномерно ограниченных по модулю) функций* — $\ln p(x, \alpha)$, $\alpha \in \Lambda$, *имела место равномерная односторонняя сходимость относительно меры, заданной некоторой (любой) плотностью* $p(x, \alpha_0)$, $\alpha_0 \in \Lambda$.

Иначе говоря, чтобы на множестве функций $Q(x, \alpha) = -\ln p(x, \alpha)$, $\alpha \in \Lambda$ выполнялись условия теоремы 3.

Как следует из замечаний, приведенных в разд. 6, достаточным условием строгой состоятельности метода максимума правдоподобия является существование на множестве $\ln p(x, \alpha)$, $\alpha \in \Lambda$, конечной ε-сети в метрике C. Можно показать, что необходимым условием строгой состоятельности является факт существования конечной ε-сети в метрике L_1.

8. Обобщение на множество неограниченных функций

Перейдем теперь к общему случаю: $p(x, \alpha)$, $\alpha \in \Lambda$, — произвольное множество плотностей. Для того чтобы сформулировать условия состоятельности метода максимума правдоподобия в метрике L_1, нам понадобится новое понятие.

Определение. Для плотности $p(x, \alpha_0)$, $\alpha_0 \in \Lambda$, существует *регулярная оценка метода максимума правдоподобия*, если найдется такое число $l_0 = l(\alpha_0)$, для которого оценка величины максимального правдоподобия

$$\sup_{\alpha \in \Lambda} \sum_{i=1}^{l_0} \ln p(x_i, \alpha) \leqslant L(x_1, \ldots, x_{l_0})$$

удовлетворяет условию

$$EL(x_1, \ldots, x_{l_0}) < \infty,$$

где математическое ожидание берется по мере $P_{\alpha_0}^{l_0}(x)$.

Замечание. Если для плотности $p(x, \alpha_0)$ регулярная оценка максимума правдоподобия существует для $l = l_0$, то она существует и для любого $l > l_0$.

Легко убедиться, что

1) для любой ограниченной сверху плотности регулярная оценка существует, начиная с $l = 1$;

2) для плотности нормального закона

$$p(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad a \in (-\infty, \infty), \sigma \in (0, \sigma),$$

(дисперсия неотделима от нуля) регулярная оценка максимума правдоподобия существует, начиная с $l = 2$, так как

$$\sup_{a, \sigma}\left(-\frac{(x_1-a)^2}{2\sigma^2} - \frac{(x_2-a)^2}{2\sigma^2} - \frac{1}{2}\ln 2\pi\sigma^2\right) < \ln\frac{1}{|x_2 - x_1|}$$

и

$$\frac{1}{2\pi\sigma_0^2}\int_{-\infty}^{\infty} \ln\frac{1}{|x_2 - x_1|}\, e^{-\frac{(x_1-a_0)^2 + (x_2-a_0)^2}{2\sigma_0^2}}\, dx_1\, dx_2 < \infty.$$

Введем функцию

$$[Q(x, \alpha)]_B = \begin{cases} Q(x, \alpha), & \text{если } |Q(x, \alpha)| \leqslant B, \\ B \operatorname{sgn} Q(x, \alpha), & \text{если } |Q(x, \alpha)| > B, \end{cases}$$

с помощью которой сформулируем теорему.

Теорема 9 [8]. *Для того чтобы метод максимума правдоподобия был состоятелен в метрике L_1, достаточно, чтобы для любого $\alpha_0 \in \Lambda$*

1) *существовала регулярная оценка метода максимума правдоподобия (начиная с* $l_0 = l(\alpha_0)$);

2) *для любого* B *на множестве функций* $\left[-\sum_{i=1}^{l_0} \ln p(x^i, \alpha)\right]_B$, $\alpha \in \Lambda$ ($l_0 = l(\alpha_0)$ *переменных* $x^1, \ldots, x^{l_0}$) *существовала равномерная односторонняя сходимость относительно меры* $P_{\alpha_0}^{l_0}(x)$.

Теорема 9 устанавливает, по-видимому, достаточно общие условия состоятельности метода максимума правдоподобия. Ее формулировка содержит два требования: существование регулярной оценки и емкостную ограниченность множества, в котором ведется оценивание. (Для равномерно ограниченных и отделимых от нуля плотностей эти два условия совпадают с условиями теоремы 8, т. е. являются необходимыми и достаточными условиями строгой состоятельности.) Здесь уместно еще раз отметить, что параметр α, задающий элемент множества $p(x, \alpha)$, $\alpha \in \Lambda$, — не обязательно конечномерный вектор, а потому метод максимума правдоподобия есть инструмент не только параметрических методов оценивания (т. е. методов оценивания в классе плотностей $p(x, \alpha)$, $\alpha \in \Lambda$, заданных с точностью до конечного числа параметров).

Приведем пример множества гладких плотностей, для которого метод максимума правдоподобия является состоятельным.

Пример 1. Рассмотрим множество $\mathscr{P}_{k,\Lambda}$ плотностей $p(x, \alpha)$, $\alpha \in \Lambda$, заданных на отрезке $[0, 1]$, таких, что

$$|\ln^{(k)} p(x, \alpha)| \leqslant A, \quad k \geqslant 1.$$

Для этого множества существует регулярная оценка метода максимума правдоподобия, начиная с $l = 1$, а множество функций $\ln p(x, \alpha)$, $\alpha \in \Lambda$, образует компакт в метрике C.

Следовательно, метод максимума правдоподобия на множестве $\mathscr{P}_{k,\Lambda}$ состоятелен в метрике Кульбака. Свойства оценки метода максимума правдоподобия на $\mathscr{P}_{k,\Lambda}$ устанавливает

Теорема 10 [11]. *Оценка максимума правдоподобия на* $\mathscr{P}_{k,\Lambda}$
1) *единственна*;

2) *представима в виде* $p(x, \alpha_l) = \exp u(x)$, *где* $u(x)$ — *сплайн порядка* k, *такой, что* $|u^{(k)}(x)| = A$, *причем* $u^{(k)}(x)$ *меняет знак между двумя элементами выборки не более* k *раз*;

3) *удовлетворяет равенству*

$$\int x^m p(x, \alpha_l)\, dx = \frac{1}{l} \sum_{i=1}^{l} x_i^m, \quad m = 0, 1, 2, \ldots, k-1,$$

где x_i — *элементы выборки.*

Для $k = 1$ оценка определяется решением уравнения

$$u'(x) = -A \operatorname{sgn}\left(\int_0^x e^{u(x')}\,dx' - F_l(x)\right),$$

$$\int_0^1 e^{u(k)}\,dx = 1.$$

Можно найти порядок скорости сходимости (в метрике Кульбака) оценки $p(x, \alpha_l)$ к искомой плотности $p(x, \alpha_0)$ [11]. Оценка сходится почти наверное со скоростью $\sqrt{\frac{\ln \ln l}{l}}$, т. е.

$$P\left\{\overline{\lim_{l\to\infty}} \sqrt{\frac{l}{\ln \ln l}}\left(-\int \ln \frac{p(x, \alpha_l)}{p(x, \alpha_0)}\, p(x, \alpha_0)\,dx = \text{const}\right\} = 1.\right.$$

9. Метод максимума структурного правдоподобия

К сожалению, метод максимума правдоподобия оказывается состоятельным лишь для узких классов плотностей. Так, он не состоятелен уже при оценивании смеси

$$p(x; a, \sigma) = \frac{1}{2\sqrt{2\pi}}\left(e^{-\frac{x^2}{2}} + \frac{1}{\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}\right), \quad a \in (-\infty, \infty),$$

$$\sigma \in (0, \sigma),$$

(не существует оценки) или при оценивании плотностей, принадлежащих множеству индикаторных функций (для этих плотностей существует регулярная оценка, но не выполнены емкостные условия теоремы 9).

Поэтому модифицируем метод максимума правдоподобия, используя принцип структурной минимизации риска. Пусть для плотностей $p(x, \alpha)$, $\alpha \in \Lambda$, существует энтропия Шеннона. Тогда если принять $Q(x, \alpha) = -\ln p(x, \alpha)$, $\alpha \in \Lambda$, $R(x, \beta) = -\ln q(x, \beta)$, $\beta \in \mathscr{B}$ ($q(x, \beta)$, $\beta \in \mathscr{B}$, — множество плотностей, всюду плотное в $p(x, \alpha)$, $\alpha \in A$, в метрике Кульбака), построить на множестве $R(x, \beta)$, $\beta \in \mathscr{B}$, структуру, удовлетворяющую требованиям разд. 7, то принцип структурной минимизации риска (который для задачи оценивания плотности назовем *принципом максимума структурного правдоподобия*) в соответствии с теоремой 7 (или условием (18)) обеспечит сходимость п. н. оценки к искомой плотности.

Пример 2. Для множества плотностей, заданных на $[-\pi, \pi]$ и имеющих энтропию, рассмотрим структуру

$$S_1 \subset S_2 \subset \ldots \subset S_n \subset \ldots,$$

где элемент S_n содержит функции $q(x, \beta)$, $\beta \in \mathcal{B}_n$, такие, что

$$R(x, \beta) = \ln q(x, \beta) = a_0 + \sum_{r=1}^{n} (a_r \sin rx + b_r \cos rx),$$

$$\sup_x |R(x, \beta)| < 1 + \sqrt{\ln n}. \tag{21}$$

Построенная структура удовлетворяет условиям теоремы 7 с параметрами

$$h_n = 2n, \quad B_n = 1 + \sqrt{\ln n},$$

и, следовательно, метод максимума структурного правдоподобия будет сходиться к искомой плотности п. н., если закон $n = n(l)$ таков, что

$$\frac{n(l) \ln n(l) \ln l}{l} \xrightarrow[l \to \infty]{} 0.$$

При этом будет достигнута асимптотическая скорость сходимости в метрике Кульбака, имеющая порядок

$$V(l) = r_{n(l)} + \sqrt{\frac{n(l) \ln n(l) \ln l}{l}}, \tag{22}$$

где $r_n = \inf_{\beta \in \mathcal{B}_n} \left(- \int \ln \frac{q(x, \beta)}{p(x, \alpha_0)} p(x, \alpha_0) dx \right)$ — скорость аппроксимации искомой плотности на элементах структуры.

Для того чтобы оценить порядок скорости сходимости (22), надо уметь оценить скорость аппроксимации r_n. Это можно сделать, лишь имея априорную информацию об оцениваемой плотности.

Так, если плотность $p(x, \alpha_0)$ такова, что функция $\ln p(x, \alpha_0)$ разложима в конечный ряд Фурье, то, начиная с некоторого $n = n_0$, будет выполнено равенство $r_n = 0$, и, следовательно, выбрав $n = n(l)$ достаточно медленно растущей функцией, можно приблизиться к асимптотической скорости $\sqrt{\frac{\ln l}{l}}$.

Если искомая плотность — гладкая функция, например

$$|\ln^{(k)} p(x, \alpha_0)| \leqslant M, \quad k \geqslant 1,$$

то, согласно теореме об аппроксимации рядами Фурье гладких функций [12], существует такая константа A, что

$$r_n \leqslant AM \frac{1}{n^k}.$$

Выбрав $n = l^{\frac{1}{2k+1}}$, получим

$$V(l) = (\ln l) l^{-\frac{k}{2k+1}}.$$

Замечание. Для того чтобы построить структуру, обеспечивающую состоятельность оценки на бесконечном интервале, достаточно в (21) вместо тригонометрических полиномов использовать ортонормальные полиномы Эрмита.

Пример 3. Пусть $p(x, \alpha)$, $\alpha \in \Lambda$ — множество плотностей, заданных на (a, b), $-\infty < a < b < \infty$. Рассмотрим структуру, элемент которой S_n состоит из плотностей, удовлетворяющих ограничению

$$|\ln^{(k)} q(x, \beta)| \leqslant B_n; \quad \beta \in \mathcal{B}_n. \tag{23}$$

Как показано в теореме 10, функция $u(x) = \ln q(x, \beta_l)$, максимизирующая на множестве (23) правдоподобие, единственна и является сплайном порядка k.

Для того чтобы определить закон $n = n(l)$, обеспечивающий состоятельность оценки плотности $q(x, \beta_l) = e^{u(x)}$, найдем число элементов минимальной ε-сети множества (23).

Известно [13], что число ячеек минимальной ε-сети в метрике C множества функций $\ln q(x, \beta)$, $\beta \in \mathcal{B}_n \left(\int q(x, \beta)\, dx = 1\right)$, удовлетворяющих (23), ограничено величиной

$$N^{\mathcal{B}_n}(\varepsilon) \leqslant \exp\left\{A \left(\frac{B_n}{\varepsilon}\right)^{\frac{1}{k}} (b-a)\right\}, \tag{24}$$

где A — абсолютная константа.

Согласно (18), метод максимума структурного правдоподобия обеспечит сходимость оценки к искомой плотности п. н., если будет выбран такой закон $n = n(l)$, что для любого ε выполнится равенство

$$\lim_{l\to\infty} \frac{B_{n(l)}^2 \left[\left(\frac{B_{n(l)}}{\varepsilon}\right)^{\frac{1}{k}} + \ln l\right]}{l} = 0.$$

При этом будет достигнута асимптотическая скорость сходимости, равная по порядку величины

$$V(l) = r_{n(l)} + \varepsilon_l^*,$$

где ε_l^* — корень уравнения

$$B_{n(l)}^2 \left[\left(\frac{B_{n(l)}}{\varepsilon}\right)^{\frac{1}{k}} + \ln l\right] = \varepsilon^2 l.$$

Пусть теперь плотность $p(x, \alpha_0)$ — гладкая функция, такая, что $|\ln^{(k)} p(x, \alpha_0)| < \infty$, $k \geqslant 1$. Тогда, начиная с некоторого $n = n_0$, скорость аппроксимации будет равна нулю. Положив $B_n = n$ и $n = \chi(l)$, где $\chi(l$ — сколь угодно медленно стремящаяся к беско-

нечности функция, получим порядок асимптотической скорости сходимости

$$V(l) = \chi(l)\, l^{-\frac{k}{2k+1}}.$$

Замечание. Для плотностей, заданных на интервале $(-\infty, \infty)$ и имеющих абсолютный момент порядка $\tau > 0 \left(\int\limits_{-\infty}^{\infty} |x|^{\tau} p(x, \alpha)\, dx < \right.$ $\left. < \infty \right)$, состоятельный метод максимума структурного правдоподобия может быть получен с помощью структуры с элементами $S_n = \bigcup\limits_{m=1}^{n} s_m$, где s_m — множество плотностей $q(x, \beta)$, $\beta \in \mathcal{B}_m$, удовлетворяющих на интервале $(-m, m)$ условию $\sup\limits_x |\ln^{(k)} q(x, \beta)| \leqslant B_m$ и равных $|x|^{-(1+\tau)}$ вне этого интервала.

Если $\sup\limits_x |\ln^{(k)} p(x, \alpha_0)| < \infty$, то для $B_m = \ln m$, $m = l^{\frac{k}{k+\tau(2k+1)}}$ порядок асимптотической скорости сходимости составит $V(l) = l^{-\frac{k}{2k+1+\frac{k}{\tau}}} \ln l$.

10. Методы решения некорректно поставленных задач

Перейдем теперь к изложению прямых методов оценивания плотности, т. е. методов решения интегрального уравнения (6) на основе эмпирических данных (7). Как уже говорилось в разд. 3, трудности здесь связаны с тем, что задача решения интегрального уравнения (6) относится к так называемым некорректно поставленным задачам математической физики.

Рассмотрим операторное уравнение

$$Af = \mathcal{F}, \tag{25}$$

взаимно однозначно отображающее элементы $f \in \mathcal{M}$ метрического пространства E_1 в элементы $\mathcal{F} \in \mathcal{N}$ метрического пространства E_2. Пусть A — непрерывный оператор, т. е. близкие (в метрике E_1) элементы $f \in \mathcal{M}$ он переводит в близкие (в метрике E_2) элементы $\mathcal{F} \in \mathcal{N}$.

Говорят, что решение операторного уравнения (25) *устойчиво*, если обратный оператор A^{-1} является непрерывным.

Говорят также, что задача решения операторного уравнения (25) поставлена *корректно по Адамару*, если решение существует, единственно, устойчиво.

К сожалению, большое количество задач (и к ним относится задача оценивания плотности (6)) оказываются некорректно поставленными из-за отсутствия устойчивости решения.

Идея устранения неустойчивости решения связана с сужением класса возможных решений до множества $\mathcal{M}^* \subset \mathcal{M}$ (множество корректности). Задачу решения операторного уравнения (25) называют *корректной по Тихонову* на множестве $\mathcal{M}^*$, если решение уравнения существует (и принадлежит $\mathcal{M}^*$), единственно и устойчиво относительно $\mathcal{N}^* = A\mathcal{M}^*$.

При $\mathcal{M}^* = \mathcal{M}$, $\mathcal{N}^* = \mathcal{N}$ понятие корректности по Тихонову совпадает с понятием корректности по Адамару. Смысл определения корректности по Тихонову заключается в том, что корректность может быть достигнута за счет сужения рассматриваемого множества решений $\mathcal{M}$ до множества корректности $\mathcal{M}^*$. Основная идея в решении некорректно поставленных задач состоит в том, что если сузить множество возможных решений $\mathcal{M}$ до компакта $\mathcal{M}^*$, то оно образует множество корректности.

Лемма (об обратном операторе). *Если на компакте $\mathcal{M}^* \subset \mathcal{M}$ задан непрерывный взаимно однозначный оператор A, то обратный оператор A^{-1} непрерывен на множестве $\mathcal{N}^* = A\mathcal{M}^*$.*

Этот факт лежит в основе различных методов решения некорректно поставленных задач. Рассмотрим один из них — метод регуляризации [5].

Пусть $\Omega(f)$ — полунепрерывный снизу функционал (стабилизатор), обладающий следующими свойствами:

1) решение операторного уравнения (25) принадлежит области определения функционала $\Omega(f)$;

2) функционал $\Omega(f)$ принимает неотрицательные значения;

3) все множества $\mathcal{M}_c = \{f: \Omega(f) \leqslant c\}$ являются компактами.

Метод регуляризации состоит в том, чтобы, используя последовательность функций f_γ, минимизирующих функционалы

$$R(f, \mathcal{F}) = \rho^2_{E_2}(Af, \mathcal{F}) + \gamma\Omega(f),$$

где $\gamma > 0$, $\gamma \to 0$, получить решение

$$f = \lim_{\gamma \to 0} f_\gamma.$$

Приближения f_γ называются *регуляризованными решениями уравнения* (25).

Пусть теперь вместо правой части $\mathcal{F}$ уравнения (25) дана последовательность функций $\mathcal{F}_\delta$, такая, что $\rho_{E_2}(\mathcal{F}, \mathcal{F}_\delta) \xrightarrow[\delta \to 0]{} 0$. В этом случае последовательностью регуляризованных решений f_γ будет последовательность функций $f_{\gamma(\delta)}$, минимизирующих функционалы

$$R_\delta(f, \mathcal{F}_\delta) = \rho^2_{E_2}(Af, \mathcal{F}_\delta) + \gamma(\delta)\Omega(f). \qquad (26)$$

Справедливы следующие две теоремы.

Теорема 11 [5]. *Пусть E_1 и E_2 — метрические пространства, и пусть для $\mathcal{F} \in \mathcal{N}$ существует решение уравнения* (25). *Тогда если вместо правой части $\mathcal{F}$ уравнения* (25) *известны приближения $\mathcal{F}_\delta$, такие, что $\rho_{E_2}(\mathcal{F}, \mathcal{F}_\delta) < \delta$, а значения параметра $\gamma = \gamma(\delta)$ таковы, что*

$$\gamma(\delta) \xrightarrow[\delta \to 0]{} 0, \quad \lim_{\delta \to 0} \frac{\delta^2}{\gamma(\delta)} = a < \infty,$$

то последовательность регуляризованных решений $f_{\gamma(\delta)}$ сходится в метрике E_1 к искомому решению уравнения при $\gamma(\delta) \xrightarrow[\delta \to 0]{} 0$.

В гильбертовом пространстве для линейного оператора A можно выбрать функционал $\Omega(f) = \|f\|^2$. И хотя множества $\mathcal{M}_c = \{f: \Omega(f) \leqslant c\}$ оказываются только слабо компактными, сходимость регуляризованного решения к искомому имеет место.

Теорема 12 [5]. *Пусть E_1 — гильбертово пространство, $\Omega(f) = \|f\|^2$ и для $\mathcal{F} \in \mathcal{N}$ существует решение уравнения* (25). *Тогда если вместо правой части $\mathcal{F}$ уравнения* (25) *известны приближения $\mathcal{F}_\delta$, такие, что $\rho_{E_2}(\mathcal{F}, \mathcal{F}_\delta) \xrightarrow[\delta \to 0]{} 0$, а значения параметра $\gamma = \gamma(\delta)$ таковы, что*

$$\gamma(\delta) \xrightarrow[\delta \to 0]{} 0, \quad \frac{\delta^2}{\gamma(\delta)} \xrightarrow[\delta \to 0]{} 0,$$

то регуляризованные решения $f_{\gamma(\delta)}$ сходятся к искомому f в метрике L_2.

11. Теоремы о стохастической регуляризации и задача оценивания плотности вероятностей

Метод регуляризации может быть распространен на решение некорректно поставленных стохастических задач [3].

Пусть стоит задача о решении операторного уравнения (25) в ситуации, когда вместо правой части $\mathcal{F}$ уравнения задана последовательность случайных функций $\mathcal{F}_l$, $l = 1, 2, \ldots$, сходящаяся к $\mathcal{F}$ в метрике E_2 по вероятности:

$$P\{\rho_{E_2}(\mathcal{F}, \mathcal{F}_l) > \varepsilon\} \xrightarrow[l \to \infty]{p} 0. \tag{27}$$

Пусть γ_l — последовательность положительных величин, стремящихся к нулю. Определим последовательность функций f_l, минимизирующих функционал

$$R(f, \mathcal{F}_l) = \rho^2_{E_2}(Af, \mathcal{F}_l) + \gamma_l \Omega(f), \tag{28}$$

где $\Omega(f)$ — стабилизирующий функционал, удовлетворяющий соответствующим условиям разд. 10.

Справедливы следующие две теоремы.

Теорема 13 [3]. *Для любых* $\varepsilon > 0$ *и* $\delta > 0$ *найдется такое* $l(\varepsilon, \delta)$, *что для всех* $l > l(\varepsilon, \delta)$ *выполнятся неравенства*

$$P\{\rho_{E_1}(f, f_l) > \varepsilon\} < P\{\rho^2_{E_2}(\mathscr{F}, \mathscr{F}_l) > \gamma\delta\}.$$

Теорема 14 [3]. *Пусть* E_1 — *гильбертово пространство* $L_2(a, b)$, A — *линейный оператор*, $\Omega(f) = \int_a^b f^2(x)\,dx$. *Тогда для всякого* $\varepsilon > 0$ *найдется такое число* $l(\varepsilon)$, *что для всех* $l > l(\varepsilon)$ *будут выполнены неравенства*

$$P\left\{\int_a^b (f(x) - f_l(x))^2\,dx > \varepsilon\right\} < 2P\left\{\rho^2_{E_2}(\mathscr{F}, \mathscr{F}_l) > \frac{\varepsilon}{2}\gamma\right\}.$$

Воспользуемся этими теоремами для построения методов оценивания плотности. Для того чтобы получить не одно, а множество приближений к искомой плотности, рассмотрим наряду с уравнением

$$\int_a^b \theta(x - x')\,p(x', \alpha)\,dx' = F(x) \tag{29}$$

уравнение

$$B\int_a^b \theta(x - x')\,p(x', \alpha)\,dx' = BF(x), \tag{29a}$$

где B — линейный невырожденный оператор. Хотя точные решения уравнений (29) и (29а) совпадают, приближенные решения этих уравнений, найденные на основе одних и тех же эмпирических функций распределения $F_l(x)$, будут различными в зависимости от оператора B.

Будем искать решение уравнения (29а) методом регуляризации в условиях, когда расстояние $\rho_{E_2}(F, F_l)$ от оценок $F_l(x)$ до функции $F(x)$ подчинено следующему достаточно слабому ограничению:

$$\rho_{E_2}(BF, BF_l) \leqslant \Phi_B(F) \sup_x |F(x) - F_l(x)|,$$

где $\Phi_B(F)$ — функционал, зависящий от B. Пусть $p(x, \alpha_l)$ — последовательность регуляризованных решений (29а), найденных на основе эмпирических функций распределения $F_l(x)$, т. е. последовательность минимумов функционалов

$$R(p, F_l) = \rho^2_{E_2}\left(B\int_0^b \theta(x - x')\,p(x', \alpha)\,dx',\ BF_l(x)\right) + \gamma\Omega(p(x, \alpha)). \tag{30}$$

Тогда из оценки Колмогорова

$$P\left\{\sup_x |F(x) - F_l(x)| > \varepsilon\right\} < 2e^{-2\varepsilon^2 l} \tag{31}$$

и теоремы 13 следует, что для всяких $\varepsilon > 0$, $\delta > 0$ найдется такое $l(\varepsilon, \delta)$, что при $l > l(\varepsilon, \delta)$ выполнятся неравенства

$$P\{\rho_{E_1}(p(x, d_l), p(x, \alpha_0)) > \varepsilon\} < 2e^{-\frac{2l\gamma_l\delta}{\Phi_B^2(F)}}. \tag{32а}$$

Если же $p(x, \alpha_0) \in L_2(a, b)$, то из теоремы 14 и неравенства (31) следует, что для всех $l > l(\varepsilon)$ выполнятся неравенства

$$P\left\{\int_a^b (p(x, \alpha_l) - p(x, \alpha_0))^2 dx > \varepsilon\right\} < 4e^{-\frac{\varepsilon l\gamma_l}{\Phi_B^2(F)}}. \tag{32б}$$

Из (32а), (32б) вытекает, что условия

$$\gamma_l \xrightarrow[l\to\infty]{} 0, \quad l\gamma_l \xrightarrow[l\to\infty]{} \infty$$

обеспечивают сходимость по вероятности регуляризованных решений к искомым, а условие

$$\gamma_l \xrightarrow[l\to\infty]{} 0, \quad \frac{l}{\ln l}\gamma_l \xrightarrow[l\to\infty]{} \infty$$

— сходимость п. н.

12. Методы оценивания плотности вероятностей

Приведем различные оценки плотности вероятностей, полученные на основе метода стохастической регуляризации.

Пример 4 (тригонометрические оценки проекционного типа). Пусть искомая плотность принадлежит $L_2(-\pi, \pi)$. В соотношении (30) определим B — единичный оператор, $\rho_{E_2}^2(F, F_l) = \int_{-\pi}^{\pi} (F(x) - F_l(x))^2 dx$ и $\Omega(p) = \int_{-\pi}^{\pi} p^2(x, \alpha) dx$. В результате получим функционал

$$R(p, F_l) = \int_{-\pi}^{\pi} \left(\int_{-\pi}^{x} p(x', \alpha) dx' - F_l(x)\right)^2 dx + \gamma_l \int_{-\pi}^{\pi} p^2(x, \alpha) dx. \tag{33}$$

Будем искать функцию, минимизирующую (33), в виде разложения в ряд Фурье. Получим оценку

$$p(x) = p_1(x) + p_2(x), \tag{34}$$

где

$$p_1(x) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} \frac{\left(\frac{1}{l}\sum_{i=1}^{l} \sin nx_i\right)\sin nx + \left(\frac{1}{l}\sum_{i=1}^{l} \cos nx_i\right)\cos nx}{\pi(1+\gamma_l n^2)},$$

$$p_2(x) = \frac{2\left(\frac{1}{l}\sum_{i=1}^{l} x_i - \int_{-\pi}^{\pi} x p_1(x)\,dx\right)}{\pi(1+2\Gamma_0(\gamma_l))} \sum_{n=1}^{\infty} \frac{(-1)^n n \sin nx}{1+\gamma_l n^2},$$

$$\Gamma_0(\gamma_l) = \sum_{n=1}^{\infty} \frac{1}{1+\gamma_l n^2}.$$

Найденная оценка сходится к искомой плотности в метрике $L_2(-\pi, \pi)$.

Если же k-я производная искомой плотности ($k \geqslant 1$) принадлежит $L_2(-\pi, \pi)$, а разложение плотности в ряд Фурье допускает k-кратное почленное дифференцирование, то используем функционал [1])

$$R(p, F_l) = \int_{-\pi}^{\pi} \left(\int_{-\pi}^{x} p(x, \alpha)\,dx - F(x)\right)^2 dx + \gamma \int_{-\pi}^{\pi} (p^{(k)}(x, \alpha))^2\,dx, \tag{33a}$$

минимум которого достигается на оценке

$$p(x) = p_1(x) + p_2(x), \tag{34a}$$

где

$$p_1(x) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} \frac{\left(\frac{1}{l}\sum_{i=1}^{l} \sin nx_i\right)\sin nx + \left(\frac{1}{l}\sum_{i=1}^{l} \cos nx_i\right)\cos nx}{\pi(1+\gamma_l n^{2(k+1)})},$$

[1]) Для задачи оценивания плотности можно использовать стабилизатор $\Omega(p) = \int_a^b (p^{(k)}(x, \alpha))^2\,dx$ с $k > 0$.

$$p_2(x) = \frac{2\left(\frac{1}{l}\sum_{i=1}^{l} x_i - \int_{-\pi}^{\pi} x p_1(x)\,dx\right)}{\pi(1+2\Gamma_k(\gamma_l))} \sum_{n=1}^{\infty} \frac{(-1)^n n \sin nx}{1+\gamma_l n^{2(k+1)}},$$

$$\Gamma_k(\gamma_l) = \sum_{n=1}^{\infty} \frac{1}{1+\gamma_l n^{2(k+1)}}.$$

Оценка сходится к искомой плотности в метрике $C^{k-1}(-\pi, \pi)$.

Пример 5 (проекционные оценки). Пусть $p(x, \alpha) \in L_2(a, b)$. Найдем оценку плотности в виде разложения в ряд по различным базисным функциям. Пусть B — линейный невырожденный оператор, действующий из L_2. Тогда

$$BT(x) = \int_a^b K(z, x)\, T(x)\,dx.$$

Обозначим

$$Gp(x', \alpha) = \int_a^b K(z, x) \int_a^b \theta(x - x')\, p(x', \alpha)\,dx',$$

$$\Phi_l(z) = BF_l(x) = \frac{1}{l}\sum_{i=1}^{l} \int_{x_i}^{b} K(z, x)\,dx.$$

Пусть

$$\varphi_1(x), \ \ldots, \ \varphi_n(x), \ \ldots$$

— собственные функции самосопряженного оператора GG^*;

$$\psi_1(z), \ \ldots, \ \psi_n(z), \ \ldots$$

— собственные функции оператора G^*G;

$$\lambda_1^2, \ \ldots, \ \lambda_n^2, \ \ldots$$

— соответствующие им собственные числа. Функции $\varphi_i(x)$, $\psi_i(z)$ и числа λ_i связаны соотношением

$$G\varphi_i(x) = \lambda_i \psi_i(z), \quad G^*\psi_i(z) = \lambda_i \varphi_i(x).$$

Будем искать минимум функционала

$$R(p, F_l) = \int_a^b (Gp(x, \alpha) - \Phi_l(z))^2\,dz + \gamma_l \int_a^b p^2(x, \alpha)\,dx \quad (35)$$

в виде разложения в ряд по системе $\varphi_i(x)$, $i = 1, 2, \ldots$,

$$p(x, \alpha_l) = \sum_{n=1}^{\infty} \alpha_n \varphi_n(x).$$

Функционал (35) достигает минимума, когда

$$\alpha_n = \frac{\lambda_n b_n}{\lambda_n^2 + \gamma_l},$$

где $b_n = \int_a^b \Phi_l(z)\,\psi_n(z)\,dz$.

Задавая различные операторы B (ядра $K(z, x)$), можно получать разные системы функций $\varphi_i(x)$ и $\psi_i(z)$. В частности, для единичного оператора B и интервала $(0, \pi)$ получим

$$p(x, \alpha_l) = p_1(x) + p_2(x),$$

где

$$p_1(x) = 2\sum_{n=1}^{\infty} \frac{\frac{1}{l}\sum_{i=1}^{l} \sin nx_i}{\pi(1+\gamma_l n^2)} \sin nx,$$

$$p_2(x) = \frac{1}{\sqrt{\gamma_l}}\left(\frac{1}{l}\sum_{i=1}^{l} \frac{\operatorname{sh}\frac{\pi - x}{\sqrt{\gamma_l}}}{\operatorname{sh}\frac{\pi}{\sqrt{\gamma_l}}}\right) \frac{\operatorname{sh}\frac{\pi - x}{\sqrt{\gamma_l}}}{\operatorname{ch}\frac{\pi}{\sqrt{\gamma_l}}}.$$

Пример 6 (сплайн-оценки). Пусть k-я производная плотности интегрируема с квадратом на (a, b) $(k \geqslant 0)$. Выберем в качестве B единичный оператор, стабилизатор $\Omega(p) = \int_a^b (p^{(k)}(x, \alpha))^2\,dx$ и метрику

$$\rho_{E_2}(F, F_l) = \int_a^b |F(x) - F_l(x)|\,dx.$$

Будем искать оценку плотности $p(x, \alpha_l)$, минимизируя функционал

$$R(p, F_l) = \left(\int_a^b |F(x) - F_l(x)|\,dx\right)^2 + \gamma_l \int_a^b (p^{(k)}(x, \alpha))^2\,dx,$$

или, что то же самое, функционал

$$R^*(p, F_l) = \int_a^b |F(x) - F_l(x)|\,dx + \mu_l \int_a^b (p^{(k)}(x, \alpha))^2\,dx.$$

Оценка плотности $p(x, \alpha_l)$, найденная методами оптимального управления, будет сплайном порядка $2k+1$, являющимся решением системы дифференциальных уравнений

$$\begin{cases} p^{(2k+1)}(x, \alpha_l) = \dfrac{(-1)^k}{\mu_l} \operatorname{sgn}(F(x) - F_l(x)), \\ F^{(1)}(x) = p(x, \alpha_l), \end{cases}$$

удовлетворяющим условиям

$$p^{(r)}(a, \alpha_l) = 0, \quad p^{(r)}(b, \alpha_l) = 0, \quad r = k+1, \ldots, 2k+1.$$

При $k \geqslant 1$ и $-\infty < a < b < \infty$ оценки сходятся к искомой плотности в метрике $C^{k-1}(a, b)$, при $k = 0$ — в метрике $L_2(a, b)$.

Пример 7 (оценки Парзена—Розенблатта). Пусть $p(x, \alpha) \in L_2(-\infty, \infty)$ и B — линейный оператор, действующий из L_2 с ядром $K(z-x)$

$$BT(x) = \int_{-\infty}^{\infty} K(z-x)\, T(x)\, dx, \quad \rho^2_{E_2}(F, F_l) = \int_{-\infty}^{\infty} (F(x) - F_l(x))^2 dx,$$

$$\Omega(p) = \int_{-\infty}^{\infty} p^2(x, \alpha)\, dx.$$

В этих условиях функционал (30) примет вид

$$R(p, F_l) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} K(z-x) \int_{-\infty}^{\infty} \theta(x-x')\, p(x', \alpha)\, dx'\, dx - \right.$$
$$\left. - \int_{-\infty}^{\infty} K(z-x)\, F_l(x)\, dx \right)^2 dz + \gamma_l \int_{-\infty}^{\infty} p^2(x, \alpha)\, dx. \quad (36)$$

Приравнивая нулю вариацию (по $p(x, \alpha)$) этого функционала и решая полученное уравнение методом преобразования Фурье, получим

$$p(x, \alpha_l) = \frac{1}{l\sqrt{\gamma_l}} \sum_{i=1}^{l} g\left(\frac{x - x_i}{\sqrt{\gamma_l}}\right)$$

— ядерную оценку Парзена — Розенблатта, где ядро $g(\cdot)$ связано с ядром $K(\cdot)$ оператора B соотношением

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(\omega)\, e^{i\omega u}\, d\omega, \quad \tilde{g}(\omega) = \frac{\tilde{K}(\omega)\, \tilde{K}(-\omega)}{\gamma_l \omega^2 + \tilde{K}(\omega)\, \tilde{K}(-\omega)},$$

$$\tilde{K}(\omega) = \int_{-\infty}^{\infty} K(u)\, e^{-iu\omega}\, du.$$

(Используя в (36) вместо $\int\limits_{-\infty}^{\infty} p^2(x, \alpha)dx$ стабилизатор вида $\int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{\infty} K_1(x-x')\,p(x',\alpha)\,dx'\right)^2 dx$, получим

$$\tilde{g}(\omega) = \frac{\tilde{K}(\omega)\,\tilde{K}(-\omega)}{\gamma_l\omega^2\tilde{K}_1(\omega)\,\tilde{K}_1(-\omega) + \tilde{K}(\omega)\,\tilde{K}(-\omega)},$$

где $\tilde{K}_1(\omega) = \int\limits_{-\infty}^{\infty} K_1(u)\,e^{-iu\omega}du$.) В частности, если B — единичный оператор, то минимум (36) определяет оценку

$$p(x, \alpha_l) = \frac{1}{l}\sum_{i=1}^{l}\frac{1}{2\sqrt{\gamma_l}}e^{-\frac{|x-x_i|}{\sqrt{\gamma_l}}}.$$

Для плотностей, у которых существует интегрируемая с квадратом производная

$$\int\limits_{-\infty}^{\infty}(p^{(1)}(x, \alpha))^2\,dx < \infty$$

функция, минимизирующая функционал

$$R(p, F_l) = \int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{\infty}\theta(x-x')\,p(x', \alpha)\,dx' - F_l(x)\right)^2 dx + \\ + \gamma_l\int\limits_{-\infty}^{\infty}(p^{(1)}(x, \alpha))^2\,dx, \qquad (37)$$

($\tilde{K}_1(\omega) = -i\omega$) определяет оценку

$$p(x, \alpha_l) = \frac{1}{2\sqrt[4]{4\gamma_l}}\sum_{i=1}^{l}e^{-\frac{|x-x_i|}{\sqrt[4]{4\gamma_l}}}\cos\frac{(x-x_i)}{\sqrt[4]{4\gamma_l}}.$$

Если плотность задана на ограниченном носителе, эта оценка сходится в метрике C.

Пример 8. Пусть $p(x, \alpha_0) \in L_2(a, b)$. Будем искать оценку $p(x, \alpha_l)$, минимизируя функционал

$$R(p, F_l) = \int\limits_a^b (F(x) - F_l(x))^2\,(d + p(x, \alpha))\,dx + \gamma_l\int\limits_a^b p^2(x, \alpha)\,dx,$$

где $d > 0$. Минимизация этого функционала сводится к задаче оптимального управления: найти управление $u(x)$, доставляющее минимум функционалу

$$R(u, F_l) = \int_a^b (F(x) - F_l(x))^2 (d + u(x))\, dx + \gamma_l \int_a^b u^2(x)\, dx$$

при ограничении $\dot{F}(x) = u(x)$. Оптимальная траектория здесь определяется решением системы дифференциальных уравнений

$$\begin{cases} \dot{F}(\alpha) = \frac{1}{2\gamma_l} (\psi(x) - (F(x) - F_l(x))^2), \\ \dot{\psi}(x) = 2(F(x) - F_l(x)) \left(d + \frac{1}{2\gamma_l} (\psi(x) - (F(x) - F_l(x))^2\right), \end{cases}$$

удовлетворяющим условиям

$$\psi(a) = 0, \quad \psi(b) = 0.$$

Оценка $p(x, \alpha_l) = \dot{F}(x)$ сходится к искомой плотности в метрике $L_2(a, b)$.

Пример 9 (гистограммные оценки). Пусть плотность $p(x, \alpha_0) \in L_2(a, b)$ имеет абсолютно интегрируемую производную. Будем искать оценку, минимизируя функционал

$$R(p, F_l) = \int_a^b (F(x) - F_l(x))^2 (d + |p^{(1)}(x, \alpha)|)\, dx + \gamma_l \int_a^b p^2(x, \alpha)\, dx, \tag{38}$$

где $d > 0$. Для этого будем решать задачу оптимального управления: найти управление $u(x)$, доставляющее минимум функционалу

$$R(u, F_l) = \int_a^b (z_2(x) - F_l(x))^2 (d + |u(x)|)\, dx + \gamma_l \int_a^b z_1^2(x)\, dx$$

при ограничениях

$$\begin{cases} \dot{z}_2(x) = z_1(x), \\ \dot{z}_1(x) = u(x). \end{cases}$$

Функция $p(x, \alpha_l) = z_1(x)$, полученная на основе неособого решения этой задачи, будет кусочно-постоянной. Она сходится к искомой плотности в метрике $L_2(a, b)$.

13. Замечания о прямых методах оценивания плотности вероятностей

1°. Итак, задавая разные способы определения расстояния от функций $F_l(x)$ до функции $F(x)$ и разные функционалы $\Omega(p)$, можно конструировать различные оценки плотности. Приведенные в предыдущем разделе примеры были подобраны так, чтобы получить конструктивные идеи основных методов оценивания, изложенных в этой книге: метода ядерных оценок (пример 7), проекционных методов (примеры 4, 5), метода гистограмм (пример 9). Однако лишь в одном случае оценка, полученная методом стохастической регуляризации, совпала с оценкой, используемой для разрешения проблемы оценивания плотности вероятностей. Это ядерная оценка Парзена — Розенблатта. В остальных случаях используются «упрощенные» аналоги оценок стохастической регуляризации.

Так, отбросив второе слагаемое в оценках плотностей (34), (34а) (что мало повлияет на результат при оценивании гладких плотностей), получим оценки с проекционными окнами [14]. Для достаточно гладких плотностей они близки к классическому варианту проекционных оценок [15].

Гистограммная оценка, определенная в примере 9, задает кусочно-постоянную функцию так, что по выборке определяются интервалы постоянных значений функции (в классических гистограммных оценках интервалы постоянства обычно фиксируются заранее).

2°. Несмотря на универсальный характер оценок Парзена — Розенблатта, их применение на практике при восстановлении плотности на заданном ограниченном носителе (особенно в многомерном пространстве) приводит к искажениям в точках, близких к граничным [16]. В этих случаях, видимо, преимущество имеют оценки проекционного типа (примеры 4,6), полученные путем минимизации того же самого функционала, что и ядерные оценки Парзена — Розенблатта, но определенного на заданном (а не бесконечном) интервале.

3°. Как следует из общей схемы, все приведенные в предыдущем разделе оценки являются состоятельными, коль скоро выполнятся условия

$$\gamma_l \xrightarrow[l\to\infty]{} 0, \quad l\gamma_l \xrightarrow[l\to\infty]{} \infty.$$

Однако при работе с выборками ограниченного объема возникает необходимость определения подходящего параметра регуляризации γ_l (параметра сглаживания). Для определения этого параметра в оценке Парзена — Розенблатта используются различные варианты *метода скользящего контроля* (cross-validation) [7, 17, 18].

В некоторых (но не всех) ситуациях они обеспечивают состоятельность оценивания [18, 17].

При поиске приближенного решения некорректно поставленных задач оптимальную константу регуляризации γ_l определяют на основе принципа невязки [19], суть которого состоит в том, чтобы выбрать константу γ_l, обеспечивающую равенство

$$\rho_{E_2}(F_\gamma, F_l) = \rho_{E_2}(F, F_l) = \delta, \tag{39}$$

где $\rho_{E_2}(F, F_l)$ — близость искаженной правой части $F_l(x)$ к истинной $F(x)$, $F_\gamma(x) = Af_\gamma$, f_γ — решение, полученное методом регуляризации с константой γ.

Основная трудность в реализации метода невязки состоит в том, чтобы оценить величину δ. При решении задачи восстановления плотности вероятностей этой трудности часто удается избежать. Так, известно распределение ω^2-статистики Смирнова

$$\omega^2 = l \int (F_l(x) - F(x))^2 \, dF(x)$$

или распределение D_l-статистики Колмогорова

$$\sqrt{l} D_l = \sqrt{l} \sup_x |F(x) - F_l(x)|.$$

Для приближенной [1]) реализации метода невязки (где в (39) метрика $\rho_{E_2}(.)$ заменяется метрикой $L_2(F)$ или C) следует задать нужную квантиль θ соответствующего распределения и определить такое γ, чтобы получить равенство

$$l \int (F_\gamma(x) - F_l(x))^2 \, dF_\gamma(x) = \theta, \tag{40}$$

или, если используется статистика Колмогорова, равенство

$$\sqrt{l} \sup_x |F_\gamma(x) - F_l(x)| = \theta_1.$$

Левая часть равенства (40) представима в виде, удобном для вычисления

$$l \int (F_\gamma(x) - F_l(x))^2 \, dF_\gamma(x) = \sum_{i=1}^{l} \left(F_\gamma(\hat{x}_i) - \frac{i - 1/2}{l}\right)^2 + \frac{1}{12l},$$

где $\hat{x}_i$ есть i-й элемент вариационного ряда выборки $x_1, \ldots, x_l$.

Равенство

$$\rho_{E_2}(F_\gamma, F_l) = \sqrt{\frac{\ln \ln l}{l}} \chi(l),$$

---

[1]) Для оценки, полученной в примере 8 при малых d, статистика ω^2 обеспечит почти точную реализацию метода невязки.

где $\chi(l)$ — произвольная (как угодно медленно) стремящаяся к бесконечности функция, обеспечит состоятельность любой оценке плотности вероятностей, полученной методом стохастической регуляризации.

14. О принципах индуктивного вывода

В этом дополнении мы рассмотрели две постановки задачи оценивания плотности вероятностей: косвенную и прямую. В каждой из них можно получать различные состоятельные методы оценивания.

В обоих случаях успех определила мысль о том, что, вообще говоря, оценку плотности надо искать не на заданном множестве функций, а на более узком множестве, зависящем от объема выборки. В разных постановках эта идея реализовалась по-разному: в косвенной — с помощью метода структурной минимизации риска, в прямой — с помощью метода стохастической регуляризации. Однако в обеих реализациях неизменным оставалось то, что на заданной структуре выбиралось подмножество с подходящей емкостной характеристикой, определяемой особенностями компакта (или емкости множества) $R(x, \beta)$, $\beta \in \mathscr{B}_n$, в методе структурной минимизации риска и особенностями компакта $\Omega(f) \leqslant c$ в методе стохастической регуляризации.

В такой ситуации, вероятно, оправданна мысль о том, что эта ключевая идея отражает общие принципы индуктивного вывода.

Принято считать, что существуют два способа рассуждений, дедуктивный, осуществляющий движение от общего к частному: и индуктивный — от частного к общему.

Идеалом дедуктивного пути является схема, в которой задается система аксиом и правил вывода, с помощью которых получают различные следствия. Дедуктивный путь должен гарантировать, что из истинных посылок будут получены истинные следствия.

Индуктивный путь рассуждений состоит в том, что из отдельных частных высказываний создаются общие суждения. При этом полученные на основе истинных частных высказываний общие суждения не всегда оказываются истинными. Тем не менее считается, что существуют такие методы индуктивного вывода, что обобщающие суждения являются оправданными.

Формулировка этих методов вывода традиционно была предметом обсуждения в философии, где в конце концов было принято, что индуктивные методы должны быть условно-кумулятивными, т. е. такими, что содержание выбранного обобщающего суждения принадлежит как можно большему числу частных высказываний при условии, что оно удовлетворяет некоторому специальному

свойству. Вопрос об этом специальном свойстве и явился предметом дискуссий. Отказ от учета особенностей обобщающего суждения определили кумулятивный принцип индукции (наивный эмпиризм), провозглашенный Ф. Бэконом: содержание обобщающего суждения должно принадлежать всем высказанным частным суждениям. Напротив, если предъявлять к особенностям обобщающего суждения слишком сильные требования, то будет отдан приоритет априорному знанию перед эмпирическим.

Обычно эти споры носили весьма общий характер. Однако в последнее время ситуация изменилась. В XVIII в. появилась теория вероятностей, которая в наше время с помощью аксиоматики А. Н. Колмогорова обрела классическую дедуктивную форму. Основную задачу теории вероятностей можно сформулировать так: известно пространство исходов эксперимента с заданной на нем вероятностной мерой, требуется определить распределение на исходах эксперимента. В начале нашего века возник новый раздел математики — математическая статистика, основная задача которой формулируется как задача, обратная теории вероятностей: по результатам эксперимента оценить вероятностную меру. (Для случая, когда вероятностная мера имеет плотность, задача оценивания плотности вероятностей по наблюдениям является, таким образом, основной задачей статистики.)

С появлением математической статистики сложилась новая ситуация в изучении процессов индукции: возник раздел математики, в котором строится и изучается простейшая модель индукции. При этом исследованию подлежат задачи, обратные задачам теории вероятностей — раздела математики, построенного по классической дедуктивной схеме. Однако этому обстоятельству, по-видимому, долгое время не уделялось достаточного внимания, и математическая статистика развивалась вне связи с анализом методов индукции [1]).

Между тем в первой половине нашего века интерес к вопросам логики науки, в том числе и процессам индукции, оказался исключительно большим благодаря работам Венской школы философов. Одним из наиболее значительных достижений здесь явилась концепция К. Поппера [21].

Основная проблема, которую пытался разрешить К. Поппер, была проблема демаркации И. Канта: чем отличаются методы с оправданным индуктивным шагом от методов, где индуктивный шаг не оправдан? К. Поппер считал, что необходимым условием оправданности индуктивного шага является принципиальная возможность фальсификации метода, т. е. существование такого набора высказываний, для которого с помощью данного метода

[1]) Хотя во многих книгах по математической статистике указывается на такую связь. См., например, [20].

нельзя отыскать общее суждение с содержанием, принадлежащим всем высказываниям набора.

Применительно к задаче оценивания функциональной зависимости по эмпирическим данным в классе $\omega(x, \alpha)$, $\alpha \in \Lambda$ индикаторных функций фальсифицируемость метода оценивания означает, что существует такая n-ка пар $x_1^*, \omega_1^*, \ldots, x_n^*, \omega_n^*$, что ни при какой выборке $x_1, \omega_1, \ldots, x_l, \omega_l$ с помощью данного метода не может быть получена оценка $\omega(x, \alpha_l)$, удовлетворяющая равенствам

$$\omega(x_i^*, \alpha_l) = \omega_i^*, \quad i = 1, 2, \ldots, n.$$

В этом случае n-ка $x_1^*, \omega_1^*, \ldots, x_n^*, \omega_n^*$ фальсифицирует метод оценивания в классе функций $\omega(x, \alpha)$, $\alpha \in \Lambda$.

При доказательстве необходимых условий равномерной односторонней сходимости в классе равномерно ограниченных функций $Q(x, \alpha)$, $\alpha \in \Lambda$ (теорема 3) центральным оказалось утверждение о том, что отсутствие (двусторонней) равномерной сходимости влечет за собой нефальсифицируемость метода минимизации эмпирического риска на некотором множестве пар x, y. А именно: если условия теоремы 2 не будут выполнены, то найдутся две функции $\psi_1(x) \geqslant \psi_0(x)$ и число $r > 0$, удовлетворяющие неравенству

$$\int (\psi_1(x) - \psi_0(x))\, dP(x) \geqslant r,$$

такие, что для всякой двоичной последовательности $\omega_1, \ldots, \omega_l$ всякого $\delta > 0$ и почти всякой выборки $x_1, \ldots, x_l$ в классе $Q(x, \alpha)$, $\alpha \in \Lambda$, отыщется функция $Q(x, \alpha^*)$, для которой выполнятся неравенства

$$|\psi_{\omega_i}(x_i) - Q(x_i, \alpha^*)| < \delta, \quad i = 1, 2, \ldots, l.$$

Согласно результатам разд. 5, условия фальсифицируемости метода минимизации эмпирического риска определяются ограниченностью емкости множества функций, в котором ведется оценивание, причем в классе функций ограниченной емкости метод минимизации эмпирического риска (отражающий кумулятивный принцип индукции в статистике) является состоятельным.

При оценивании плотности более глубоким, чем кумулятивный, оказался условно-кумулятивный принцип индукции, основанный на «правильном» соотнесении емкостной характеристики элемента структуры с количеством имеющихся эмпирических данных.

Является ли используемая здесь мера «правильного» соотнесения выражением общего условия оправданности индуктивного шага? Если принять эту мысль, то, учитывая, что понятие «емкость множества» определено для произвольных множеств функций, быть может, следует придать принципу структурной минимиза-

ции более общий смысл с тем, чтобы использовать его для создания различных индуктивных систем, в том числе и индуктивных логик.

ЛИТЕРАТУРА

1. Fisher R. A. Theory of statistical estimation//Proc. Cambridge Phil. Soc. — 1925. — P. 700—725.
2. Wald A. Note of consistency of M. L. estimate//Ann. Math. Stat. — 1945. — 20. — P. 595—601.
3. Вапник В. Н., Стефанюк А. Р. Непараметрические методы восстановления плотности вероятностей//АиТ. — 1978. — 8. — С. 38—52.
4. Иванов В. К., Васин В. В., Танана В. П. Теория линейных некорректных задач и ее приложения. — М.: Наука, 1978.
5. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. — М.: Наука, 1974.
6. Phillips D. L. A technique for numerical solution of certain integral equations of the first kind//J. Assoc. Comput. Math. — 1962. — V. 9. — N 1. — P. 84—87.
7. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.
8. Vapnik V. N., Chervonenkis A. J. Estimation of expected risk from empirical data//Тезисы 1-го Всемирного конгресса общества математической статистики и теории вероятностей им. Бернулли. — 1986. — Т. 2. — С. 580—583.
9. Вапник В. Н., Червоненкис А. Я. Необходимые и достаточные условия равномерной сходимости средних к их математическим ожиданиям.//Теория вероятн. и ее примен. — 1981. — Т. XXVI. — 3. — С. 543—563.
10. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. — М.: Наука, 1974.
11. Айду Ф. А. Сплайны в задаче непараметрической оценки плотности.//АиТ. — 1987.
12. Тиман А. Ф. Теория приближения функций. — М.: ГИФМЛ, 1960.
13. Витушкин А. Г. Оценка сложности задачи табулирования. — М.: ГИФМЛ, 1959.
14. Бенткус Р., Казбарас А. Оптимальные статистические оценки плотности распределения в присутствии априорной информации//Литовский математический сборник. — 1982. — Т. XXII. — 3. — С. 29—40.
15. Ченцов Н. Н. Статистические решающие правила и оптимальные выводы. — М.: Наука, 1972.
16. Shuster E. F. Incorporating Support constrains into nonparametric estimators of densities//Communication in statistics. — Theory Math. — 1985. — 14. — 5. — P. 1123—1136.
17. Shuster E. F., Gregory C. On the nonconsistency of maximum likelihood nonparametric density estimators//Computer science and statistics: Proceding of the 13 Symposium on the interface. — Springer-Verlag 1981.
18. Hall P. Large sample optimality of least squares cross—validation in density estimation//Ann. Stat. — 1985. N 11. — P. 1156—1174.
19. Морозов В. А. О регуляризации некорректно поставленных задач и выборе параметра регуляризации//ЖВМ и МФ. — 1966. — Т. 6. — 1. — С. 170—175.
20. Боровков А. А. Математическая статистика. — М.: Наука, 1984.
21. Поппер К. Логика и рост научного знания. — М.: Прогресс, 1983.

Именной указатель [1)]

Abou-Jaoude S. 19, 20, 21, 23, 31, 32, 35, 36, 38, 39, 42, 152, 211, 222, 224
Abramson I. S. 198, 222
Ahmad I. A. 199, 222
Ahrens J. H. 247, 248
Anderson G. L. 342
Anderson T. W. 210, 222
Andrews D. F. 253, 259
Archer N. P. 248, 249
Asau Y. 247, 249
Askey R. 318, 319, 322, 342
Asselin de Beauville J. P. 330, 342
Assouad P. 50, 51, 82, 84

Babu A. J. G. 250
Banon G. 200, 222
Bartlett M. S. 89, 121, 133, 153, 213, 215, 223
Bean S. J. 14
Beck J. 268, 273
Beckenbach E. F. 90, 153
Bellman R. 90, 153
Bennett G. 168, 194, 286, 291
Bertrand-Retali M. 185, 194
Bickel P. J. 43, 44, 259
Birgé L. 15, 47, 50, 51, 55, 82, 222, 223
Bleuez J. 297, 298, 310, 317, 322, 342, 343
Blum J. R. 191, 196, 290, 292, 345
Boas R. P. 337, 343
Bosq D. 296, 297, 298, 310, 317, 322, 342, 343
Bowman A. W. 163, 195
Boyd D. W. 59, 83
Breiman L. 197—199, 223, 243, 249
Bretagnolle J. 25, 42, 48, 58, 59, 60, 75, 83, 84, 129, 134, 135, 153, 195, 230, 249
Brunk H. D. 329, 343
Butzer P. L. 315, 316, 319, 326, 328, 332, 335, 338, 340, 343

Cacoullos T. 23, 42
Carleson L. 306, 343
Carlson F. 90, 153
Carroll R. J. 199, 223
Chen H. C. 247, 249
Chow Y. S. 161, 162, 195
Cover T. M. 14, 267, 273
Crain B. R. 297, 322, 330, 343
Csibi S. 263, 273
Csiszar I. 230, 249

Davies H. I. 199, 200, 223, 225
Davis K. B. 141, 143, 152, 153
Deak I. 243, 249
De Figueriedo R. J. P. 342
De Guzman 16, 22
Dehaan L. 256, 259
Deheuvels P. 60, 83, 89, 115, 153, 156, 158, 195, 199, 200, 205, 215, 223, 243, 244, 249
De Montricher M. 210, 223
Devroye L. 15, 22, 23, 42, 43, 44, 57, 83, 153, 156, 185, 195, 198—200, 223, 249, 251, 259, 263, 265, 266, 268, 273
Diaconis P. 105, 115, 153
Duin R. P. W. 160, 195, 224

Edwards R. E. 303, 306, 307, 343
Efron B. 43, 44

Factor L. E. 159, 162, 196
Farrell R. H. 59, 83, 84
Fefferman C. 307, 343
Feller W. 109, 141, 153, 221, 223, 281, 291
Fellner W. H. 329, 343
Fix E. 267, 273
Földes A. 14, 296, 343
Fox B. L. 239, 248, 249
Freedman D. 105, 115, 153
Fritz J. 268, 273
Fryer M. J. 14

Gaskins R. A. 209, 224
Gasser T. 215, 223
Gastwirth J. L. 253, 260
Geman S. 162, 195, 207—209, 223
Gessaman M. P. 211, 223
Glick N. 21, 22
Good I. J. 209, 224
Gordon L. 267, 273
Gray H. L. 225
Greblicki W. 296, 297, 317, 322, 343
Gregory G. G. 162, 163, 196
Grenander U. 207, 219, 224
Groeneboom P. 219, 224
Györfi L. 199, 224, 263, 266, 273

Haagerup U. 146, 153

[1)] К с. 1—347.

Habbema J. D. F. 160, 195, 198, 224
Hall P. 40, 43, 163, 195, 205, 224, 297, 311, 322, 324, 328, 339, 343
Hampel F. 259
Hanna B. 211, 224
Hart P. E. 267, 273
Hašimov S. A. 346, 347
Hayes C. A. 16, 22
Hermans J. 195, 224
Heyde C. C. 205, 224
Hodges J. L. 267, 273
Hoeffding W. 25, 43, 271, 274, 285, 292
Hominal P. 115, 153, 156, 159, 195, 244, 249
Huder C. 25, 42, 48, 58, 59, 60, 75, 83, 84, 129, 134, 135, 153, 195, 230, 249
Huber P. J. 253, 259, 260, 288, 292
Hunt R. A. 307, 343
Hwang C.-R. 162, 195, 207, 209, 223

Isogai E. 200, 224

Jorsboe O. G. 307, 343

Kac M. 337, 343
Kassam S. A. 288
Kemperman J. H. B. 230, 249
Kiefer J. 59, 83, 182, 195
Klonias V. K. 210, 224
Knuth D. E. 245, 249
Kohrt K. D. 247, 248
Kolmogorov A. N. 307, 343
Korner T. W. 307, 344
Kronmal R. A. 14, 247, 249, 296, 329, 330, 344, 345
Krzyzak A. 296, 328, 344
Kullback S. 230, 249
Kulldorf G. 118, 153

Leadbetter M. R. 141, 152, 154, 346, 347
Le Cam L. 232, 249
Leonard T. 14
Lin P. 199, 222
Loève M. 202, 205, 224
Loftsgaarden D. O. 198, 224, 267, 274
Lorentz G. G. 344
Lukacs E. 141, 153

Machell F. 259
Mack Y. P. 198, 224
Mammitzch V. 223
Marshall A. W. 289, 292
Marzinkiewicz J. 144, 153
Meilbro L. 307, 343
Meisel W. 197, 223, 243, 249
Mirzahmedov M. A. 346, 347
Moore D. S. 198, 224
Mozzochi C. J. 307, 344
Muckenhoupt B. 318, 320, 322, 344
Müller H.-G. 215, 223, 224

Nessel R. J. 315, 316, 319, 326, 328, 332, 335, 338, 340, 343
Neveu J. 22
Newman J. 322, 344

Olevskii A. M. 344
Olshen R. A. 267, 273
Owen D. B. 225

Parzen E. 23, 43, 85, 153, 253, 260, 289, 292
Pauc C. Y. 16, 22
Pawlak M. 296, 328, 343, 344
Penrod C. S. 153, 259
Peterson A. V. 247, 249
Pitman E. J. 231, 249
Pollard H. 322, 344
Poor H. V. 288
Prakasa Rao B. L. S. 14, 219, 224
Proshan F. 289, 292
Purcell E. 197, 223, 243, 249

Quade E. S. 316, 344
Quenouille M. 216, 224
Quesenberry C. P. 198, 224, 267, 274

Raatgever J. W. 198, 224
Rao C. R. 282, 292
Rejtö L. 200, 225
Remme J. 224
Révész P. 14, 200, 225, 296, 343
Robertson T. 222, 225
Rogers W. H. 259
Rosenblatt M. 12, 23, 43, 44, 56, 57, 60, 83, 85, 87, 89, 153, 159, 195, 198, 215, 224, 225, 289, 292, 329, 344
Rubinstein R. 243, 244, 250
Rudemo M. 160, 162, 163, 195
Rudin W. 322, 344

Sacks J. 265, 274
Sansone G. 293, 294, 297, 324, 344
Scheffé H. 11, 22
Schmeiser B. W. 243, 250, 253, 260
Schneider B. 14, 15
Schucany W. R. 216, 225
Schüler L. 296, 344
Schuster E. F. 162, 163, 196
Schwartz S. C. 297, 317, 322, 344
Scott D. W. 97, 115, 154, 159, 162, 196, 209, 218, 225

Seneta E. 177, 196, 256, 260
Serfling R. J. 233, 250
Shalaby M. A. 243, 250
Shanmugam K. S. 241, 243, 250
Shapiro H. S. 16, 22
Sibuya M. 243, 250
Silverman B. W. 15, 159, 196, 210, 225
Sjölin P. 307, 345
Slud E. V. 73, 83
Sommers J. P. 216, 225
Spiegelman C. 265, 274
Steele J. M. 59, 83
Stegbuchner H. 296, 345
Stein E. M. 16, 18, 22, 345
Stein C. 43, 44
Sterbuchner H. 296, 345
Stone C. J. 59, 83, 84, 163, 164, 196, 267, 268
Stout W. F. 268, 274
Szarek S. J. 146, 154
Szegö G. 293, 295, 297, 323, 345

Tapia R. A. 14, 89, 154, 196, 207, 209, 223, 225
Tarasenko F. P. 226
Tarter M. E. 14, 296, 329, 330, 344, 345
Tashiro Y. 243, 250
Taylor M. S. 248, 250
Terrell G. R. 218, 225
Thompson J. R. 14, 89, 154, 196, 207, 209, 223, 225, 248, 250
Tsokos C. P. 14
Tukey J. W. 259

Vandenbroek K. 195
Van Ryzin J. 210, 225, 263, 274, 296, 345
Viollaz A. J. 297, 322, 345
Von Neumann J. 244, 250

Wagner T. J. 156, 159, 185, 195, 196, 199, 225, 263, 265, 273, 274
Wahba G. 59, 60, 83, 211, 225, 316, 329, 330, 345
Wainger S. 318, 319, 322, 342
Walker A. J. 247, 250
Walter G. G. 191, 196, 290, 292, 297, 317, 345
Watson G. N. 109, 154
Watson G. S411, 152, 154, 296, 329, 345—347
Wegman E. J. 14, 199, 200, 222, 225
Wertz W. 14, 15, 125, 154, 190, 196, 290, 292
Wheeden R. L. 16, 17, 18, 22
Wittaker E. T. 109, 154
Whittle P. 329, 345
Wolfowitz J. 182, 195
Wolverton C. T. 199, 225, 263, 274
Woodroofe M. 159, 196
Wu L. D. 195

Yackel J. W. 198, 224
Yamato H. 199, 226
Young R. M. G. 146, 154

Zygmund A. 16—18, 22, 144, 153, 306, 345

Бари Н. К. 302—304, 311, 316, 345
Булдаков В. М. 226

Епанечников В. А. 89, 121, 154
Ефроймович С. Ю. 346, 347

Ибрагимов И. А. 59, 60, 83, 84, 140, 154

Колмогоров А. Н. 51, 83, 344
Конаков В. Д. 141, 154
Кошкин Г. М. 226

Маматов М. 280, 292
Мания Г. М. 59, 83
Манставичус Э. 144, 154
Мнацаканов Р. М. 346, 347

Надарая Э. А. 14, 15, 43, 44, 56, 60, 83, 159, 196

Петров В. В. 99, 154, 280, 292
Прохоров Ю. В. 206, 226

Самаров А. М. 59, 83
Сираждинов С. Х. 280, 292

Тарасенко Ф. П. 15, 210, 226

Фано Р. 84

Хасьминский Р. З. 59, 60, 83, 84, 140, 154
Хашимов Ш. А. 346, 347
Хмаладзе Э. В. 346, 347

Ченцов Н. Н. 84, 296, 345

Предметный указатель

Адаптивная ядерная оценка 155—196
— — — асимптотически оптимальная в L_2 163
— — — масштабная инвариантность 191
— — — основанная на асимптотических разложениях 158
— — — — — оптимизации критерия 160
— — — скорость сходимости 192—194
— — — состоятельность 155, 156
— — — эвристическая 159
Адаптивное оценивание 88, 135
Аддитивная модификация оценки плотности 212
Анализ чувствительности 119
Апостериорная вероятность 261
Ассуа лемма 50, 84

Бесселя равенство *см. Парсеваля* равенство
Биномиальная случайная величина, неравенство для абсолютного уклонения 35
— — — — — максимальной вероятности 109
— — — — — p-го момента 145, 146
— — — — на хвосты 172
Бхаттачария расстояние 354

Вариация оценки плотности в метрике L_1 34
— — — — точке 100
— — — равномерные границы 131
Вероятность ошибки байесовская 262
— — условная 262
Верхняя граница минимаксная 45, 132
— — — для классов аналитических плотностей 140, 151
— — — — — — — — *Бретаньоля — Юбер* 135, 136
— — — — — — — — *Липшица* 132, 314, 340, 341
— — — — — — — — монотонных плотностей 221
— — — — — — — — *Соболева* 316
Винзоризация 288
Выпуклость по *Шуру* 289

Генерирования случайных величин метод инверсии 227, 247
— — — — композиции 229
— — — — порядковых статистик 242
— — — — путеводителей 247, 248
— — — — удаления 229, 242, 244
— — — — эталона 247
Гиббса эффект 311
Гистограммная оценка 13, 19, 30—34, 85, 89, 97, 128, 159, 207, 210, 237, 298, 338, 352, 356, 390, 391
— — выбор параметра сглаживания 114—120
— — генерирование случайных величин 246—248
— — кросс-проверочная 160, 162
— — кубическая 31, 41
— — нижняя граница для L_1-ошибки 106
— — относительная устойчивость 38, 41, 43
— — переменная 210—211
— — преобразованная 252
— — скорость сходимости 105—107
— — смещение 111
— — состоятельность 31, 32
— — трансляционная инвариантность 190

Данные 10
Дельтаобразная последовательность 326
Детектор 282—288
— максимума правдоподобия 282, 284
— — — винзоризованный 288
— оптимальный 282, 284
— основанный на выборочных данных 287

— — — критерии L_1 (ДL1) 284—287
— — — критериях распознавания образов (ДРО) 284, 285, 287
— робастный 284, 288
— состоятельный 287
Дискриминантный анализ *см.* Классификация
Достаточная статистика 72

Емкость множества функций 370

Инвариантность оценок плотности масштабная 121, 190, 191
— — — к монотонным преобразованиям 10, 11, 251, 252, 254, 282, 354, 355
— — — перестановочная 288—291
— — — трансляционная 190, 191, 300, 320, 324, 325
Изолированный холм оценки плотности 254

Классификация 261—273
— ближайших соседей метод 267
— гистограммный метод 266
— ядерный метод 265
Класс плотностей аналитических 84, 140, 149
— — *Бретаньоля — Юбер* 48, 56, 128—136, 311, 315
— — *Гельдера* 84
— — *Липшица* 52, 128, 129, 134, 311, 333, 335
— — монотонно невозрастающих 46, 55, 219—222, 255
— — *Соболева* 311, 316, 317
Колмогорова контрпример 307
— нижняя граница 268
Корректность по *Адамару* 380
— — *Тихонову* 381
Кристоффеля — Дарбу формула суммирования 295
Кросс-проверка 160—162
Кульбака метрика 363

Лебега постоянная 302
— точка 17, 156, 168, 172

Максимума правдоподобия метод 363, 373—377
— — — со штрафами 209
— — оценка 206, 359
— — — *Гренандера* 219
— — принцип 160
— структурного правдоподобия метод 377—380
Минимаксный риск 45, 84
Минимизации эмпирического риска метод 365
Модуль непрерывности в L_1 332
— — верхние границы 333
— — второго порядка 332
Монотонный инвариант 354
Монте-Карло метод
— — оценка функционала 228
— — уменьшение дисперсии 246
Мультипликативная модификация оценки плотности 212

Невязка i-го момента 239
Независимость выборок 234
Некорректные задачи оценивания 351, 352, 365, 380—384
Неотрицательная проекция плотности 277
Неравенство
— *Беннета* 168, 206, 286
— *Берри — Эссеена* 98, 99, 136
— для равномерного уклонения 168
— *Карлсона* 90
— *Крамера — Рао* 50
— *Ле Кама* 232
— *Лоренца* 311
— *Марцинкевича — Зигмунда* 144
— *Маршалла — Прошана* 289
— *Сарека* 144—146
— *Фано* 84
— *Хаусдорфа — Юнга* 316
— *Хефдинга* 28, 271, 285
— *Хинчина* 145
— *Юнга* 16
Неразличимость выборок 236
Несмещенность 140—152, 342
Нижняя граница
— — для фиксированной плотности 45—47, 57, 59, 238
— — минимаксная 45—55, 57—60, 84, 128
— — равномерная 45
— — экспоненциальная для больших уклонений 172

Обнаружения задача 281—288
Обобщение выборки 233
Ортонормальная система 293
— — базис 293
— — *Лагерра* 295

— — *Лежандра* 295
— — полная 293
— — тригонометрическая 296
— — *Хаара* 298
— — *Эрмита* 295, 297
Относительная устойчивость 34, 43
Оценка плотности 10
— *Бартлета* 49, 152, 213, 244, 277
— — нормированная 213
— ближайших соседей 198, 267
— *Волвертона — Вагнера* 199, 226
— гистограммная *см.* Гистограммная оценка
— *Гренандера* 219
— *k*-превосходная для выборки объема *m* 236
— максимума правдоподобия 206, 359, 363, 373—377
— *Парзена — Розенблата см.* Ядерная оценка
— проекционная *см.* Проекционная оценка
— сингулярно-интегральная *см.* Сингулярно-интегральная оценка
— складного ножа 217
— сплайн 387, 388
— *Террела — Скотта* 218, 245
— типа дельта-функции *Дирака* 191, 290, 325
— — интеграла *Фурье* 141, 142
— ядерная *см.* Ядерная оценка
— — адаптивная *см.* Адаптивная ядерная оценка
— — преобразованная 251—259
— — рекуррентная 199—206
— — с переменным параметром сглаживания 197—199

Параметр масштаба 116, 118
— сглаживания
— — адаптивный выбор 155—196
— — локальная настройка 198
— — минимаксная стратегия выбора 125—128
— — оптимальный выбор для гистограммной оценки 114—120
— — — — — — ядерной оценки 114—120
— — параметрический метод выбора 115, 158—159
— сдвига 116
Парсеваля равенство 294
Плотность аналитическая 84
— бета-распределения 121, 125
— *Валле-Пуссена* 142, 149
— *Коши* 91, 116—118, 159, 162
— *Лапласа* 118, 177
— маргинальная 275
— монотонная 46, 55, 219—222, 255
— мультимодальная 119
— нормальная 59, 109, 115—118, 159, 162, 207, 241, 257, 280
— *Парето* 162
— равнобедренная треугольная 52, 88, 95, 97, 117, 119, 127, 133, 134, 139, 239, 251, 252, 257
— равномерная 106, 107, 109, 117—119, 121, 123—126, 128, 159, 239, 252, 310, 338
— радиально-симметричная 279
— с неограниченным носителем 136—140
— — полиномиальными хвостами 162
— — правильно меняющимися хвостами 91, 255—257
— *Стьюдента* 116, 117, 162, 256
— с тяжелыми хвостами 254—257
— унимодальная 46, 55, 66, 67, 109, 116, 119, 137, 139, 221, 222, 255, 281
— устойчивая 152
— экспоненциальная 117, 118
Полиграмма 210, 226
Полиномы *Лежандра* 297
— *Якоби* 297, 322
Полная вариация 11, 228, 351
Поперечник класса плотностей 84
Порядковые статистики 118, 162, 191, 210, 242, 248, 253
Правильно меняющаяся функция 91, 255—256
— — последовательность 174, 177
Проекционная оценка 295, 293—342, 356, 384—387, 391
— — адаптивная 330
— — многомерная 296
— — сглаженная 329, 346
— — с рядом *Лагерра* 297, 317, 322, 346
— — — — — *Лежандра* 297, 322—324
— — — — — — несостоятельность 323
— — — — — — состоятельность 323
— — — — — — трансляционная инвариантность 324
— — — — — *Фурье* 296, 301—317, 327, 342, 384—386
— — — — — — достижимость ошибки $O(1/\sqrt{n})$ 309
— — — — — — несостоятельность 301
— — — — — — связь с сингулярно-интегральными оценками 328, 329
— — — — — — сглаженная 330
— — — — — — скорость сходимости 310—317

— — — — — — смещение 307
— — — — — — состоятельность 307
— — — — — *Хаара* 298
— — — — — *Эрмита* 297, 317, 322
— — — — — — несостоятельность 317
— — — — — — состоятельность 321, 322
— — — — — — трансляционная инвариантность 320
— — — тригонометрическим рядом *см.*
Проекционная оценка с рядом *Фурье*
Произведение плотностей 278
Прохорова критерий сходимости 206
Пуассонизация 24, 187
Путеводитель 248

Разбиения 19, 20, 30, 105—107
— последовательность вложенная 19
— — кубическая 19
Распознавание образов *см.* Классификация
Регуляризация 365, 381, 382
— стохастическая 382—384
Решение 261
— байесовское 261
Решета метод 207, 372
Решето 207
— сверточное 207
Робастная оценка параметра масштаба 118, 158, 253
— — — сдвига 253

Свертка плотностей 280
Семимонотонная последовательность 156, 157, 179
Сильно дельтаобразная последовательность 326
Симметризация оценок плотности 288—291
Сингулярно-интегральная оценка 325—342, 346
— — границы для смещения 335, 336
— — неограниченной мощности 338, 342
— — несмещенность 342
— — состоятельность 327
Сингулярный интеграл 325
— *Фейера* 338
Складного ножа метод 216
— — оценка 217
Сковгарда границы 318
Скорость сходимости адаптивной ядерной оценки 192—194
— — *Бартлета* оценки 213—216
— — *Гренандера* оценки 219, 220
— — кубической гистограммной оценки 105—107
— — проекционной оценки с тригонометрическим рядом 310—317
— — сингулярно-интегральной оценки 335—342
— — ядерной оценки 54, 60, 85—105, 128—140, 158, 220—221
Смесь плотностей 275
Смещение 87, 95, 98, 100, 129, 167
— методы понижения 211—219
Согласование моментов 239
Состоятельность адаптивной ядерной оценки 155, 156
— *Бартлета* оценки 213
— *Волвертона — Вагнера* оценки 205
— гистограммной оценки 31, 32
— детекторов 287
— кросс-проверочной ядерной оценки 161
— оценки *Гренандера* 220, 221
— — максимума правдоподобия 208, 209, 373—375
— оценок плотности, полученных методом стохастической регуляризации 384, 391
— переменной гистограммной оценки 211
— преобразованной ядерной оценки 258, 259
— проекционной оценки с рядом
— — — — — *Лежандра* 323
— — — — — *Фурье* 307, 308
— — — — — *Хаара* 298
— — — — — *Эрмита* 321
— рекуррентной ядерной оценки 200, 205
— *Террела — Скотта* оценки 218
— ядерной оценки 23, 156, 157
— — — с переменным параметром сглаживания 198
Спейсинги 187, 237
Структурной минимизации риска метод *см.* Метод решета
Сужение плотности 276, 277
Сходимость вполне 23, 155—157

Теорема *Берри — Эссеена* 99, 104
— *Джексона* вторая 315, 340
— — первая 339
— *Карлесона — Ханта* 306, 307, 322
— *Котельникова* 141
— *Лебега* о плотностях 17
— локальная предельная 280
— *Найквиста* 141
— о медленной сходимости 46, 54, 263
— *Фейера — Лебега* 303
— *Хинчина* 281

— центральная предельная 72, 73, 99, 104
— *Шеффе* 11, 21

Феррера функции 297
Фишеровская информация 354
Функция *C*-липшицева 52
— медленно меняющаяся 177, 256
— правильно меняющаяся 91, 255, 256
— регрессии 261
— характеристическая 140—143, 146—152, 203, 204
Фурье ряд 293
— — поточечная сходимость 306
— — частичная сумма 296

Хеллингера расстояние 231, 278

Чисара φ-дивергенция 354

Экспонентное семейство 357
Экспоненциальная сходимость 23, 265—267
Энтропия относительная 354, 358
— *Реньи* 283
— *Шеннона* 363

Ядерная оценка 13, 106, 107, 121—124, 129, 132, 135, 142, 143, 151, 152, 172, 193, 200, 207, 209, 210, 212, 217, 237—239, 251, 254, 257, 258, 289, 295, 309—311, 315, 325, 328, 335, 352, 356, 388, 389, 391
— — адаптивная *см.* Адаптивная ядерная оценка
— — асимптотическое распределение
— — — — L_1-ошибки 87
— — — — L_2-ошибки 40, 43
— — выбор параметра сглаживания 114—120, 125—128
— — генерирование случайных величин 242—247
— — инвариантность 190—192
— — кросс-проверочная 160—164
— — нижняя граница L_1-ошибки 88
— — относительная устойчивость 39, 43
— — преобразованная 88, 251—256
— — рекуррентная 199—206
— — скорость сходимости 54, 60, 85—105, 128—140, 158, 220—221
— — смещение и дисперсия 98
— — состоятельность 23, 156, 157
— — с переменным параметром сглаживания 197
— — — пониженным смещением 211—219
— — трапецеидальная 150—152, 339, 342
Ядро ассоциированное 129
— *Бартлета* см. Ядро *Епанечникова*
— *Валле-Пуссена* 328, 330, 332, 335, 337
— — второе 328, 331, 335, 339, 340
— *Джексона* 328, 334, 337, 339
— *Джексона — Валле-Пуссена* 328
— *Дирихле* 296, 301, 302, 325, 339
— *Епанечникова* 89, 114, 115, 124, 132, 133, 139, 239—242, 245, 252, 254, 258, 259
— неограниченной мощности 339
— равнобедренное треугольное 125, 127, 132, 239
— *Рогозинского* 328, 330, 335, 339
— *Фейера* 302, 327, 330, 334, 337
— *Фейера — Коровкина* 328, 330, 334, 337, 340

Оглавление

Предисловие редактора перевода 5
Предисловие к русскому изданию 7
Предисловие 9

Глава 1. Введение 10
Литература 14
Комментарий переводчика 15

Глава 2. Дифференцирование интегралов 16
Литература 21

Глава 3. Состоятельность 23

1. Ядерная оценка 23
2. Доказательство теоремы 1 24
3. Гистограммная оценка 30
4. Доказательство теоремы 2 32
5. Относительная устойчивость 34
Литература 42
Комментарий переводчика 43

Глава 4. Нижние границы для скоростей сходимости 45

1. Введение 45
2. Лемма Ассуа 50
3. Некоторые исторические замечания 56
4. Доказательство теоремы 1 60
5. Доказательства теорем 2 и 11 67
6. Доказательство теоремы 3 68
7. Доказательство теоремы 4 71
8. Доказательства теорем 5—9 74
Литература 82
Комментарий переводчика 84

Глава 5. Скорости сходимости в L_1 85

1. Введение 85
2. Коэффициент $B^*(f)$ 89
3. Доказательства теорем 1 и 2 98
4. Гистограммная оценка 105
5. Доказательства теорем 5 и 6 108
6. Выбор параметра сглаживания 114

7. Равномерная плотность 121
8. Минимаксная стратегия выбора параметра сглаживания . . . 125
9. Классы Липшица, классы Бретаньоля—Юбер и равномерные верхние границы 128
10. Плотности с неограниченным носителем 136
11. Несмещенность и достижимость скорости убывания ошибки $1/\sqrt{n}$ 140
Литература 152

Глава 6. Адаптивная ядерная оценка, ее сходимость в L_1 и поточечная сходимость 155

1. Основной результат 155
2. Поточечная сходимость адаптивной ядерной оценки 155
3. Поточечная сходимость обычной ядерной оценки 156
4. Примеры адаптивных ядерных оценок 157
5. Доказательства 164
6. Инвариантное оценивание плотности 190
7. Скорость сходимости адаптивных ядерных оценок 192
Литература 194

Глава 7. Оценки, близкие к ядерной и гистограммной оценкам 197

1. Введение 197
2. Ядерные оценки с переменным параметром сглаживания . . . 197
3. Рекуррентные ядерные оценки 199
4. Оценки максимума правдоподобия 206
5. Переменная гистограммная оценка 210
6. Ядерные оценки с пониженным смещением 211
7. Оценка Гренандера для монотонных плотностей 219
Литература 222
Комментарий переводчика 226

Глава 8. Моделирование, неравенства и генерирование случайных величин 227

1. Выбор критерия 227
2. Неравенства 228
3. Обобщение выборки для генерирования случайных величин 233
Литература 248

Глава 9. Преобразованная ядерная оценка 251

1. Введение 251
2. Выбор преобразования 253
3. Оценивание плотностей с тяжелыми хвостами 254
4. Состоятельность 258
Литература 259

Глава 10. Применения к задаче классификации 261

1. Задача классификации 261
2. Медленные скорости сходимости 263
3. Ядерный метод классификации 265
4. Гистограммный метод классификации 266
5. Метод ближайших соседей 267
Литература 273

Глава 11. Операции над оценками плотности 275

1. Маргинальные плотности 275
2. Композиция (смесь) плотностей 276
3. Сужения плотностей 276
4. Неотрицательные проекции 277
5. Произведения плотностей 278
6. Радиально-симметричные плотности 279
7. Свертки 280
8. Унимодальные плотности 281
9. Применения в задаче обнаружения 281
10. Симметризация и перестановочная инвариантность 288

Литература 291

Глава 12. Проекционные оценки 293

1. Определения 293
2. Примеры ортонормальных систем 296
3. Общие свойства 298
4. Состоятельность оценки с тригонометрическим рядом 301
5. Скорость сходимости оценки с тригонометрическим рядом . . 310
6. Оценка с рядом Эрмита 317
7. Оценка с рядом Лежандра 322
8. Сингулярно-интегральные оценки 325

Литература 342
Комментарий переводчика 346

Дополнение 1. Почему L_1-подход и что за горизонтом? *Н. Н. Ченцов* . . 348
Литература 360

Дополнение 2. Принципы оценивания плотности распределения вероятностей. *В. Н. Вапник* 362
Литература 396

Именной указатель 397
Предметный указатель 400

УВАЖАЕМЫЙ ЧИТАТЕЛЬ

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу:

129820, Москва, И-110, ГСП,
1-й Рижский пер., д. 2, издательство «Мир».

Монография

Люк Деврой, Ласло Дьёрфи

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ПЛОТНОСТИ

Заведующий редакцией профессор Б. В. Шабат
Зам. зав. редакцией А. С. Попов
Ст. научн. редактор И. А. Маховая
Мл. научн. редактор Т. А. Денисова
Художник А. В. Шипов
Художественный редактор В. И. Шаповалов
Технический редактор Е. В. Алехина
Корректор Н. А. Гиря

ИБ № 6311

Сдано в набор 18.05.87. Подписано к печати 09.11.87.
Формат 60×88¹/₁₆. Бумага книжно-журнальная. Печать офсетная.
Гарнитура литературная. Объем 12,75 бум. л. Усл. печ. л. 25,50.
Усл. кр.-отт. 25,50. Уч.-изд. л. 24,94. Изд. № 1/5225. Тираж 4500 экз.
Зак. 176. Цена 3 р. 20 к.

ИЗДАТЕЛЬСТВО «МИР» 129820,
ГСП, Москва, И-110, 1-й Рижский пер., 2

Набрано в Ленинградской типографии № 6 ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 193144, г. Ленинград, ул. Моисеенко, 10. Отпечатано по офсету в Ленинградской типографии № 4 ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 191126, Ленинград, Социалистическая ул., 14.

ЗБС
Д256