ANOTHER PROOF OF A SLOW CONVERGENCE RESULT OF BIRGE

Luc Devroye
School of Computer Science
McGill University

ABSTRACT. We give a short proof of the following result. Let f_n be any density estimate based upon an i.i.d. sample drawn from a density f. For any monotone decreasing sequence $\{a_n\}$ of positive numbers converging to zero with $a_1 \leq 1/32$, a density f may be found such that

$$\mathbb{E}\left\{\int |f_n(x) - f(x)|\,dx\right\} \geq a_n$$

for all n. This density may be picked from the class of densities on $[0,1]$ that are bounded by two. The proof of this fact simplifies an earlier proof by Birgé (1986) and extends a weaker lower bound by the author (1983).

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: Primary: 62G07. Secondary: 62G05, 62F12, 60F25.

KEYWORDS AND PHRASES. Density estimation. Nonparametric estimation. Lower bounds. Minimax theory. Rate of convergence.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

We are given data $X_1, \ldots, X_n$, an i.i.d. sample drawn from an unknown density f on the real line. An estimate f_n of f is a mapping from $\mathbb{R}^{n+1}$ to $\mathbb{R}$. Given the data, $f(x)$ is estimated by $f_n(x; X_1, \ldots, X_n)$. The following "slow rate of convergence" result was shown in Devroye (1983).

THEOREM. *Let $\{f_n\}$ be a given sequence of estimates and let $a_n \downarrow 0$ be a sequence of real numbers. Then there exists a density f such that*

$$\mathbb{E} \int |f_n - f| > a_n$$

infinitely often. The density f may also be taken from the class of unimodal densities with infinitely many continuous derivatives.

The theorem states that to study rates of convergence in density estimation, we nned at least some combination of a tail condition and a smoothness condition. Nevertheless, the fact that the result referred to some unknown subsequence prompted Birgé (1986) to improve the above theorem as follows.

THEOREM (BIRGÉ, 1986). *Let $a_n \to 0$ such that* $\sup a_n \in (2/39, 2/13)$. *For any sequence $\{f_n\}$, there exists a density f on $[0,1]$ bounded by two such that*

$$\mathbb{E} \int |f_n - f| \geq a_n$$

for all n.

Birgé's Theorem 5.4 contains quite a bit more than what is stated above. We feel that it may be helpful to give a simple proof, that is strongly based on an embedding argument with coupled random variables. Hopefully, the proof method given here may find other applications in lower bounds that are valid for sequences of estimates. No attempt is made to get the best possible bound on a_1. With a different bound on a_1, the Theorem remains valid when f is restricted to all unimodal densities that are infinitely many times continuously differentiable.

MAIN THEOREM. *Let $a_n \downarrow 0$ such that $a_1 \leq 1/32$. For any sequence $\{f_n\}$, there exists a density f on $[0,1]$ bounded by two such that*

$$\mathbb{E} \int |f_n - f| \geq a_n$$

for all n.

LEMMA. *For any monotone decreasing sequence $\{a_n\}$ of positive numbers converging to zero with $a_1 \le 1/16$, a discrete probability distribution $(p_1, p_2, \ldots)$ may be found such that $p_1 \ge p_2 \ge \ldots$, and for all n*

$$\sum_{i=n+1}^{\infty} p_i \ge \max\left(8a_n, 32np_{n+1}\right).$$

PROOF. It suffices to look for p_i's such that

$$\sum_{i=n+1}^{\infty} p_i \ge \max\left(8a_n, 32np_n\right).$$

These conditions are easily satisfied. For positive integers $u < v$, define the function $H(v,u) = \sum_{i=u}^{v-1} 1/i$. First we find a sequence $1 = n_1 < n_2 < \ldots$ of integers with the following properties:

A. $H(n_{k+1}, n_k)$ is monotonically increasing;
B. $H(n_2, n_1) \ge 32$;
C. $8a_{n_k} \le 1/2^k$ for all $k \ge 1$.

Note that (C) may only be satisfied if $a_{n_1} = a_1 \le 1/16$. To this end, define constants $c_1, c_2, \ldots$ by

$$c_k = \frac{32}{2^k H(n_{k+1}, n_k)} \quad (k \ge 1),$$

so that the c_k's are decreasing in k, and

$$\frac{1}{32}\sum_{k=1}^{\infty} c_k H(n_{k+1}, n_k) = \sum_{k=1}^{\infty} 2^{-k} = 1.$$

For $n \in [n_k, n_{k+1})$, we define $p_n = c_k/(32n)$. We claim that these numbers have the required properties. Indeed, $\{p_n\}$ is decreasing, also,

$$\sum_{n=1}^{\infty} p_n = \sum_{k=1}^{\infty} \frac{c_k}{32} H(n_{k+1}, n_k) = 1.$$

Finally, if $n \in [n_k, n_{k+1})$, then

$$\sum_{i=n+1}^{\infty} p_i \ge \sum_{j=k+1}^{\infty} \frac{c_j}{32} H(n_{j+1}, n_j) = \sum_{j=k+1}^{\infty} 2^{-j} = 2^{-k}.$$

Clearly, on the one hand, by the monotonicity of $H(n_{k+1}, n_k)$, $1/2^k \ge c_k = 32np_n$. On the other hand, $1/2^k \ge 8a_{n_k} \ge 8a_n$. This concludes the proof. □

PROOF OF THE MAIN THEOREM. Define $\epsilon_n = 2a_n$. First we construct a family of densities f. Let $b = 0.b_1b_2b_3\ldots$ be a real number on $[0,1]$ with the shown binary expansion, and let

B be a random variable uniformly distributed on $[0,1]$ with expansion $B = 0.B_1B_2B_3\ldots$. Let us define a random variable W with

$$\mathbb{P}\{W = i\} = p_i \ (i \geq 1) \ ,$$

where $p_1 \geq p_2 \geq \ldots > 0$, and

$$\sum_{i=n+1}^{\infty} p_i \geq \max\left(8\epsilon_n, 32np_{n+1}\right)$$

for every n. That such p_i's exist follows from the Lemma.

Define an i.i.d. sequence of uniform $[0,1]$ random variables $U_1, U_2, \ldots$. Define another i.i.d. sequence $W_1, W_2, \ldots$ drawn from the distribution of W. These sequences are used to construct coupled data sequences. Each $b \in [0,1)$ describes a different distribution. With b replaced by B we have a random distribution. Define the random variables

$$X_i = F_{W_i} - U_i p_{W_i} + b_{W_i} \ ,$$

where $F_j = p_1 + \cdots p_j$. It is not difficult to see that if $W_i = j$ and $b_j = 0$, then X_i is uniformly distributed on $[F_{j-1}, F_j]$ where $F_0 = 0$ by definition. If $b_j = 1$, then it is uniformly distributed on $[F_{j-1}+1, F_j+1]$. In any case, the density of X_i is supported on $[0,2]$ and at every x, it takes the value zero or one. Introduce $A_j = [F_{j-1}+1, F_j+1] \cup [F_{j-1}, F_j]$.

We write f_b to denote the density of X_1 for parameter b. Introduce the shorthand notation $\mathcal{U}_n = (U_1, U_2, \ldots, U_n)$, $\mathcal{W}_n = (W_1, W_2, \ldots, W_n)$, and $\mathcal{X}_n = (X_1, X_2, \ldots, X_n)$. We define the error

$$J_n(b) = \int |f_n(x, \mathcal{X}_n) - f(x)| \, dx = \sum_{i=1}^{\infty} \int_{A_i} |f_n(x, \mathcal{X}_n) - f(x)| \, dx \ .$$

Thus,

$$\sup_b \inf_n \frac{\mathbb{E} J_n(b)}{\epsilon_n} \geq \sup_b \mathbb{E}\left\{\inf_n \frac{J_n(b)}{\epsilon_n}\right\} \geq \mathbb{E}\left\{\inf_n \frac{J_n(B)}{\epsilon_n}\right\} \ .$$

Consider now the following conditional expectation:

$$\begin{aligned}
\mathbb{E}\left\{\inf_n \frac{J_n(B)}{\epsilon_n} \middle| \mathcal{U}_\infty, \mathcal{W}_\infty\right\} &\geq \mathbb{P}\left\{\bigcap_{n=1}^{\infty} [J_n(B) \geq \epsilon_n] \middle| \mathcal{U}_\infty, \mathcal{W}_\infty\right\} \\
&\geq 1 - \sum_{n=1}^{\infty} \mathbb{P}\left\{J_n(B) < \epsilon_n | \mathcal{U}_\infty, \mathcal{W}_\infty\right\} \\
&= 1 - \sum_{n=1}^{\infty} \mathbb{P}\left\{J_n(B) < \epsilon_n | \mathcal{U}_n, \mathcal{W}_n\right\} \ .
\end{aligned}$$

We bound the conditional probabilities inside the sum: let $\mathcal{D}_n$ denote $\mathcal{U}_n, \mathcal{W}_n, B_{W_1}, \ldots, B_{W_n}$, so that $\mathcal{X}_n$ is fixed for every fixed element of $\mathcal{D}_n$. Condition on $\mathcal{D}_n$ and fix i. Observe that

A_i consists of two disjoint intervals —called A_i' and A_i'' below— of length p_i each, and that f_n is a fixed function, since $\mathcal{X}_n$ is fixed. Define $\alpha_i = \int_{A_i'} |f_n - 0|$, $\beta_i = \int_{A_i''} |f_n - 1|$, $\gamma_i = \int_{A_i'} |f_n - 1|$, and $\delta_i = \int_{A_i''} |f_n - 0|$. Observe that $\alpha_i + \gamma_i \geq p_i/2$, and $\beta_i + \delta_i \geq p_i/2$. We note that given $\mathcal{D}_n$, if $i \notin S \stackrel{\text{def}}{=} \{W_1, \ldots, W_n\}$, then

$$
\begin{aligned}
\int_{A_i} |f_n - f_B| &\geq (\alpha_i + \beta_i) I_{B_i=1} + (\gamma_i + \delta_i) I_{B_i=0} \\
&\geq \max\left(\alpha_i + \beta_i, \gamma_i + \delta_i\right) I_{B_i=m} \\
&\qquad \text{(where } m = 1 \text{ if the maximum occurs for } \alpha_i + \beta_i \text{ and } 0 \text{ otherwise)} \\
&\geq (p_i/2) I_{B_i=m} \\
&\qquad \text{(since } \alpha_i + \beta_i + \gamma_i + \delta_i \geq p_i) \\
&\stackrel{\mathcal{L}}{=} (p_i/2) I_{B_i=1} \ .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left\{ J_n(B) < \epsilon_n \middle| \mathcal{D}_n \right\} \\
&\leq \mathbb{P}\left\{ \sum_{i \notin \{W_1, \ldots, W_n\}} \int_{A_i} |f_n(x, \mathcal{X}_n) - f_B(x)|\, dx < \epsilon_n \middle| \mathcal{D}_n \right\} \\
&\leq \mathbb{P}\left\{ \sum_{i \notin \{W_1, \ldots, W_n\}} p_i I_{B_i=1} < 2\epsilon_n \middle| \mathcal{D}_n \right\} \\
&\leq \mathbb{P}\left\{ \sum_{i=n+1}^{\infty} p_i I_{B_i=1} < 2\epsilon_n \right\} \\
&\qquad \text{(since the } p_i\text{'s are decreasing; by stochastic dominance)} \\
&= \mathbb{P}\left\{ \sum_{i=n+1}^{\infty} p_i B_i < 2\epsilon_n \right\} \ .
\end{aligned}
$$

We bound the last probability by Chernoff's method.

$$
\begin{aligned}
\mathbb{P}\left\{ \sum_{i=n+1}^{\infty} p_i B_i < 2\epsilon_n \right\} &\leq \mathbb{E}\left\{ e^{2s\epsilon_n - s\sum_{i=n+1}^{\infty} p_i B_i} \right\} \\
&\qquad \text{(where } s > 0 \text{ is to be picked later)} \\
&= e^{2s\epsilon_n} \prod_{i=n+1}^{\infty} \left(\frac{1}{2} + \frac{1}{2} e^{-sp_i} \right) \\
&\leq e^{2s\epsilon_n} \prod_{i=n+1}^{\infty} \frac{1}{2} \left(2 - sp_i + s^2 p_i^2/2 \right) \\
&\qquad \text{(since } e^{-x} \leq 1 - x + x^2/2)
\end{aligned}
$$

$$\le \exp\left(2s\epsilon_n + \sum_{i=n+1}^{\infty} \left(-sp_i/2 + s^2p_i^2/4\right)\right)$$

(since $1 - x \le e^{-x}$)

$$\le \exp\left(2s\epsilon_n - s\Sigma/2 + s^2 p_{n+1}\Sigma/4\right)$$

(where $\Sigma = \sum_{i=n+1}^{\infty} p_i$)

$$= \exp\left(-\frac{1}{4}\frac{(4\epsilon_n - \Sigma)^2}{\Sigma p_{n+1}}\right)$$

(by taking $t = (\Sigma - 4\epsilon_n)/(p_{n+1}\Sigma)$, and the fact that $\Sigma \ge 4\epsilon_n$)

$$\le e^{-\frac{\Sigma}{16p_{n+1}}}$$

(since $\Sigma \ge 8\epsilon_n$)

$$\le e^{-2n}$$

(since $\Sigma \ge 32p_{n+1}n$) .

Thus, we conclude that

$$\sup_b \inf_n \frac{\mathbb{E}J_n(b)}{\epsilon_n} \ge 1 - \sum_{n=1}^{\infty} e^{-2n} = \frac{e^2-2}{e^2-1} > \frac{1}{2} ,$$

so that there exists a b for which $\mathbb{E}J_n(b) \ge a_n$ for all n. □

References.

L. Birgé, "On estimating a density using Hellinger distance and some other strange facts," *Probability Theory and Related Fields*, vol. 71, pp. 271–291, 1986.

L. Devroye, "On arbitrarily slow rates of global convergence in density estimation," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 62, pp. 475–483, 1983.