The Annals of Statistics
1990, Vol. 18, No. 3, 1496–1499

NO EMPIRICAL PROBABILITY MEASURE CAN CONVERGE IN THE TOTAL VARIATION SENSE FOR ALL DISTRIBUTIONS[1]

BY LUC DEVROYE AND LÁSZLÓ GYŐRFI[2]

McGill University and Hungarian Academy of Sciences

For any sequence of empirical probability measures $\{\mu_n\}$ on the Borel sets of the real line and any $\delta > 0$, there exists a singular continuous probability measure μ such that

$$\inf_n \sup_A |\mu_n(A) - \mu(A)| \geq \tfrac{1}{2} - \delta \quad \text{almost surely.}$$

We consider a probability measure μ on the Borel sets of the real line, from which we draw an i.i.d. sample $X_1, \dots, X_n$. An empirical probability measure μ_n is a probability measure on the same Borel sets and for a fixed set A, $\mu_n(A)$ is a measurable function of the data $X_1, \dots, X_n$. In particular, we are interested in the *total variation*

$$T_n \triangleq \sup_A |\mu_n(A) - \mu(A)|,$$

where the supremum is over all the Borel sets. By considering suprema over left infinite intervals only, it is easy to see that $T_n \geq \sup_x |F_n(x) - F(x)|$, the Glivenko–Cantelli norm, where F_n and F are the distribution functions corresponding to μ_n and μ, respectively. The *standard empirical measure*, defined by

$$\mu_n(A) \triangleq \frac{1}{n} \sum_{i=1}^{n} I_{[X_i \in A]},$$

is atomic in nature. Hence, whenever μ is continuous, we have $T_n \equiv 1$ almost surely for all n. This is in stark contrast with the Glivenko–Cantelli norm, which is known to converge to zero almost surely as $n \to \infty$ (by the Glivenko–Cantelli theorem). If μ is atomic, it is quite obvious that $T_n \to 0$ almost surely as $n \to \infty$. In order for T_n to be small when μ is nonatomic, we should not use the standard empirical measure. For example, for absolutely continuous μ (with density f), Scheffé's lemma [Scheffé (1947) states that

$$T_n = \tfrac{1}{2} \int |f_n - f|,$$

Received September 1988; revised July 1989.
[1]Research supported by NSERC Grant A-3456 and by the Hungarian National Foundation for Scientific Research (Grant no. 1485).
[2]Written at the Department of Mathematics, University of Leuven, Belgium, while L. Győrfi was visiting Professor E. C. van der Meulen.
AMS 1980 *subject classifications*. Primary 62G05; secondary 60E05.
Key words and phrases. Empirical measure, total variation distance, singular continuous distributions.

when μ_n is an absolutely continuous empirical measure with density f_n. But it is very easy to construct density estimates f_n with the property that for all f, $\int |f_n - f| \to 0$ almost surely: It suffices to take for f_n the kernel estimate

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

[Parzen (1962); Rosenblatt (1956)], where K is an arbitrary fixed density and $h = h_n$ is any sequence of random variables, possibly dependent upon the data, for which $h_n \to 0$ almost surely and $nh_n \to \infty$ almost surely [see Devroye and Győrfi (1985), Chapter 6, and the references therein]. Other estimates, such as the histogram estimate, share the same universal consistency property. In summary, T_n is rather sensitive to the nature of the underlying probability measure μ and for discrete and absolutely continuous μ, it is possible to construct empirical measures for which $T_n \to 0$ almost surely. The same is obviously true for mixtures of discrete and absolutely continuous measures; it suffices to consider appropriate mixtures of the two empirical measures introduced above, where for the discrete part, we only take into account those X_i's for which $X_j \equiv X_i$ for some $j \neq i$. The question thus arises: Can we construct an empirical measure that is weakly or strongly consistent (in the total variation sense) for all μ?

We have to answer this question in the negative, simply because a universally consistent empirical measure does not even exist for all singular continuous μ. Indeed, in the space of all probability measures on the Borel sets of the real line, the atomic and absolutely continuous measures can be considered as two miniscule islands in a vast ocean of singular continuous measures. No finite sample can possibly be large enough to identify one of these singular continuous probability measures.

THEOREM. *Let $\{\mu_n\}$ be a sequence of empirical probability measures and let δ be a positive constant. Then there exists a probability measure μ such that*

$$\inf_n \sup_A |\mu_n(A) - \mu(A)| \geq \tfrac{1}{2} - \delta \quad \textit{almost surely}.$$

The theorem shows that for any sequence of empirical measures, there exists a singular continuous μ for which $T_n \geq \frac{1}{2} - \delta$ almost surely, for all n. In other words, consistent empirical measures can only be constructed for certain specific subclasses of measures μ.

If in the statement of the theorem, we omit $\inf_n$, a standard minimax statement is obtained. However, the bad probability measure that is singled out in $\sup_\mu$ is now allowed to vary with n, whereas in the theorem, the same μ is to be used for all n. In fact, in the minimax format, it is possible to replace the phrase "singular continuous" by "absolutely continuous" or "atomic" [Devroye (1983)]. For certain subclasses of absolutely continuous probability measures, lower bounds for individual μ and all n were obtained by Devroye (1983) and Birgé (1985, 1986).

Finally, the constant $\frac{1}{2}$ in the theorem can undoubtedly be replaced by the constant 1 at the expense of a more involved proof.

PROOF OF THE THEOREM. The proof borrows some arguments from Devroye (1983) and Rényi (1959). First, we need a rich family of singular continuous probability measures. The family of probability measures considered here is parametrized by a number $b \in [0,1]$ with binary expansion $b = 0.b_{(1)}b_{(2)}b_{(3)}\dots$, $b_{(i)} \in \{0,1\}$. Choose an integer $m > 1/(2\delta)$. Let the random variables $Y_{(1)}, Y_{(2)}, \dots$ be i.i.d. and uniformly distributed on $\{0, 1, \dots, m-1\}$. We define the random variable $X = X(Y, b)$ by setting $X = 0.X_{(1)}X_{(2)}X_{(3)}\dots$ in the m-ary radix system used for $Y = 0.Y_{(1)}Y_{(2)}\dots$, where

$$X_{(k)} \triangleq \begin{cases} 0, & \text{if } b_{(k)} = 0, \\ Y_{(k)}, & \text{if } b_{(k)} = 1. \end{cases}$$

Let μ_b denote the probability measure of $X = X(Y, b)$. If in the binary expansion of b there are finitely many (L) zeros, then μ_b is absolutely continuous and distributes its mass uniformly on a set of Lebesgue measure m^{-L}. If there are finitely many (L) ones, then μ_b is discrete and puts its mass uniformly on a set of cardinality m^L. In the other cases, μ_b is singular.

We write $X(Y_1, b), \dots, X(Y_n, b)$ to denote a sample drawn from the distribution of $X(Y, b)$. We will replace b at a crucial step in the argument by a uniform $[0,1]$ random variable B, which is independent of $Y_1, \dots, Y_n$. Let μ_n be the empirical measure based upon $X(Y_1, b), \dots, X(Y_n, b)$. Put

$$A_k = \{0.x_{(1)}x_{(2)} \cdots : x_{(i)} \in \{0, \dots, m-1\} \text{ for all } i;\ x_{(k)} = 0\}.$$

Then

$$\mu_b(A_k) = \begin{cases} 1, & \text{if } b_{(k)} = 0, \\ \dfrac{1}{m}, & \text{if } b_{(k)} = 1. \end{cases}$$

Let us now define $b_n = 0.b_{n1}b_{n2} \cdots$ by its binary expansion with bits

$$b_{nk} = \begin{cases} 0, & \text{if } \mu_n(A_k) > \dfrac{1 + 1/m}{2}, \\ 1, & \text{otherwise}. \end{cases}$$

Then,

$$|\mu_n(A_k) - \mu_b(A_k)| \geq \frac{1 - 1/m}{2} I_{[b_{nk} \neq b_{(k)}]}.$$

Therefore,

$$\begin{aligned} \sup_b \inf_n \sup_A |\mu_n(A) - \mu_b(A)| &\geq \sup_b \inf_n \sup_k |\mu_n(A_k) - \mu_b(A_k)| \\ &\geq \sup_b \inf_n \sup_k \frac{1 - 1/m}{2} I_{[b_{nk} \neq b_{(k)}]}. \end{aligned}$$

Replace b by B and the resulting b_{nk} by B_{nk}. Then

$$\begin{aligned} \sup_b \inf_n \sup_A |\mu_n(A) - \mu_b(A)| &\geq \inf_n \sup_k \frac{1 - 1/m}{2} I_{[B_{nk} \neq B_{(k)}]} \\ &\triangleq \frac{1 - 1/m}{2} \inf_n Z_n. \end{aligned}$$

Our theorem is proved if we can show that $Z_n = 1$ almost surely for all n. Put $Z_{Nn} = I_{[\bigcup_{k=1}^{N}[B_{nk} \neq B_{(k)}]]}$. Then $Z_{Nn} \uparrow Z_n = I_{[\bigcup_{k=1}^{\infty}[B_{nk} \neq B_{(k)}]]}$. Therefore, it suffices to show that

$$\lim_{N\to\infty} \mathbf{P}(Z_{Nn} = 1) = 1$$

or equivalently,

$$\lim_{N\to\infty} \mathbf{P}\left(\bigcup_{k=1}^{N}\left[B_{nk} \neq B_{(k)}\right]\right) = 1.$$

But $\mathbf{P}(\bigcup_{k=1}^{N}[B_{nk} \neq B_{(k)}])$ is the error probability of the decision $(B_{n1}, \dots, B_{nN})$ on $(B_{(1)}, \dots, B_{(N)})$ for the observations $X_1, \dots, X_n$. For this decision problem the Bayesian decision is

$$\tilde{B}_{nk} = \begin{cases} 0, & \text{if } X_{i(k)} = 0 \text{ for all } i = 1, \dots, n, \\ 1, & \text{otherwise}. \end{cases}$$

Thus,

$$\mathbf{P}(Z_{Nn} = 1) = \mathbf{P}\left(\bigcup_{k=1}^{N}\left[B_{nk} \neq B_{(k)}\right]\right) \geq \mathbf{P}\left(\bigcup_{k=1}^{N}\left[\tilde{B}_{nk} \neq B_{(k)}\right]\right)$$

$$= 1 - \left(1 - \frac{1}{2m^n}\right)^N \uparrow 1. \qquad \square$$

REFERENCES

BIRGÉ, L. (1985). Non-asymptotic minimax risk for Hellinger balls. *Probab. Math. Statist.* **5** 21–29.

BIRGÉ, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291.

DEVROYE, L. (1983). On arbitrarily slow rates of global convergence in density estimation. *Z. Wahrsch. Verw. Gebiete* **62** 475–483.

DEVROYE, L. and GYŐRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

PARZEN, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33** 1065–1076.

RÉNYI, A. (1959). On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hungar.* **10** 193–215.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–458.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
3480 UNIVERSITY STREET
MONTREAL
CANADA H3A 2A7

HUNGARIAN ACADEMY OF SCIENCES
TECHNICAL UNIVERSITY OF BUDAPEST
STOCZEK U.2
BUDAPEST
HUNGARY H-1111