The Annals of Statistics
1977, Vol. 5, No. 2, 400–404

ESTIMATING A DISTRIBUTION FUNCTION[1]

BY RUDOLF BERAN

University of California, Berkeley

It is shown that the limiting distribution of any regular estimator of a continuous cdf on [0, 1] can be represented as a convolution of the Brownian bridge process with another distribution on $C[0, 1]$. The result is related to Hájek's representation for limiting distributions of regular parametric estimators.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables, each having continuous cdf F on $[0, 1]$. Based upon $(X_1, X_2, \cdots, X_n)$, there exists a well-known piecewise linear estimator $\tilde{F}_n$ of F such that the distribution of $n^{\frac{1}{2}}[\tilde{F}_n(x) - F(x)]$ converges weakly in $C[0, 1]$ as $n \to \infty$ to the distribution of the Brownian bridge process $B(x)$, which is Gaussian with mean 0 and covariance kernel $\min[F(x), F(y)] - F(x)F(y)$. (For further details regarding $\tilde{F}_n$, see Billingsley (1968), page 104.) An interesting question is: to what extent is $\tilde{F}_n$ a good estimator of F, at least in large samples?

Complicating the issue is the existence of cdf estimators which are superefficient at a specified distribution. For instance, consider the estimator $\hat{F}_n$ defined by

$$(1.1)\qquad \begin{aligned} \hat{F}_n(x) &= H(x) \qquad \text{if} \quad \sup_x |\tilde{F}_n(x) - H(x)| \leqq n^{-\frac{1}{4}} \\ &= \tilde{F}_n(x) \qquad \text{otherwise,} \end{aligned}$$

where H is a cdf supported on the interval $[0, 1]$. If the data distribution is H, then $\lim_{n\to\infty} P[\hat{F}_n \equiv H] = 1$; otherwise $n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)]$ converges in distribution in $C[0, 1]$ to $B(x)$.

Dvoretzky, Kiefer and Wolfowitz (1956) showed that the sample distribution function, and hence $\tilde{F}_n$, is an asymptotically minimax estimator of F under a variety of risk structures. The theorem presented in Section 2 of this note permits direct comparisons among asymptotic distributions of estimators of F; however, the class of estimators is restricted to exclude superefficient estimators. The theorem established is an analogue of Hájek's (1970) theorem on parametric estimation. It neither implies, nor is it implied by, the results of Dvoretzky et al. (1956).

2. The main result. Let μ be a measure on $[0, 1]$ with respect to which the continuous cdf F has density f; if necessary μ can be the measure induced by F. Let $\mathscr{F}(\mu)$ denote the set of all densities with respect to μ. Let $\mathscr{C}(f, \delta)$ be the

Received May 1975; revised August 1976.

[1] This research was partially supported by National Science Foundation Grant GP-31091X.

AMS 1970 *subject classifications.* Primary 62G05; Secondary 62E20.

Key words and phrases. Distribution function, regular estimation, convolution representation.

set of all sequences of densities $\{f_m \in \mathscr{F}(\mu)\}$ such that

$$\lim_{m\to\infty} \|m^{\frac{1}{2}}(f_m^{\frac{1}{2}} - f^{\frac{1}{2}}) - \delta\| = 0 , \tag{2.1}$$

where $\delta \in L_2(\mu)$ and $\|\cdot\|$ is the $L_2(\mu)$ metric. Existence of the limit in (2.1) requires that δ be orthogonal to $f^{\frac{1}{2}}$ in $L_2(\mu)$. Let $\mathscr{C}(f)$ denote the union of all sets $\{\mathscr{C}(f, \delta) : \delta \in L_2, \delta \perp f^{\frac{1}{2}}\}$.

Pick an arbitrary sequence of densities $\{f_m\} \in \mathscr{C}(f)$ and for each m, let F_m represent the cdf of f_m. Consider a corresponding sequence of sampling experiments wherein the nth experiment realizes n independent random variables $\{X_{i,n}; 1 \leqq i \leqq n\}$ whose joint density is $\prod_{i=1}^n f_n(x_{i,n})$. Let $\{\hat{F}_n\}$ be any sequence of cdf estimators such that $\hat{F}_n$ is a function of the $\{X_{i,n}\}$ and takes its values in $C[0, 1]$.

DEFINITION. The sequence of cdf estimators $\{\hat{F}_n\}$ is said to be regular at f if the distributions of the centered estimators $\{n^{\frac{1}{2}}[\hat{F}_n(x) - F_n(x)]\}$ converge weakly in $C[0, 1]$ to a distribution $\mathscr{D}$ which depends only upon f and not upon the choice of the sequence $\{f_m\} \in \mathscr{C}(f)$ that determines the sampling scheme.

For the sake of brevity, we will speak of a regular estimator $\hat{F}_n$ rather than of a regular sequence of estimators $\{\hat{F}_n\}$. Note that the sequence of densities $\{f_m : f_m = f \text{ for every } m\}$ belongs to $\mathscr{C}(f)$ trivially. Thus, the property of regularity defined above can be viewed as a robustness or stability property: a cdf estimator which is not regular at f is one whose centered distributions change too severely when the data density is perturbed slightly from f in the Hellinger metric. Such unstable behaviour is undesirable when it is necessary to approximate the distribution of a cdf estimator with the help of asymptotic theory and the data. In particular, the property of regularity is not possessed by superefficient estimators such as (1.1).

Indeed, without loss of generality, we may assume that the cdf H in (1.1) has density h with respect to the measure μ. Let $\{h_m\}$ be a density sequence in $\mathscr{C}(h, \delta)$ and let $\{H_m\}$ be the corresponding cdf sequence. By (2.8) below, the sequence of joint distributions $\{\prod_{i=1}^n h_n(x_{i,n})\}$ is contiguous to the sequence $\{\prod_{i=1}^n h(x_{i,n})\}$. Thus, for the cdf estimator $\hat{F}_n$ defined in (1.1), and under the sampling scheme associated with the density sequence $\{h_m\}$, the process $n^{\frac{1}{2}}[\hat{F}_n(x) - H(x)]$ converges in probability in $C[0, 1]$ to the trivial zero process. Consequently, $n^{\frac{1}{2}}[\hat{F}_n(x) - H_n(x)]$ converges in probability to the trivial process $2 \int_0^x \delta(t)h^{\frac{1}{2}}(t)\, d\mu$. Since this process has distribution depending upon the choice of δ, the estimator (1.1) is not regular at h.

The following theorem provides a useful representation for the limiting distribution of a regular cdf estimator; a proof is given at the end of this paper. Mathematically, this theorem is a strict nonparametric analogue of Hájek's (1970) representation for the limiting distribution of a regular (weaker sense) parametric estimator.

THEOREM. *For any regular estimator $\hat{F}_n$ of the continuous cdf F, $\mathscr{D}$ can be represented as a convolution $\mathscr{D}_B * \mathscr{D}_1$, where $\mathscr{D}_B$ is the distribution of the Brownian bridge process B and $\mathscr{D}_1$ is a distribution on* $C[0, 1]$ *depending only upon f.*

Evidently the piecewise linear estimator $\tilde{F}_n$ is regular (a contiguity argument shows it) and has the property that its $\mathscr{D}$ coincides with $\mathscr{D}_B$ for all f. In view of the theorem, this property is an optimality property: roughly speaking, no regular estimator of F can have a centered limiting distribution which is less dispersed than that of $\tilde{F}_n$.

This last assertion can be made precise by introducing the Lévy concentration function $K_{\mathscr{D}}(t)$, defined for $0 < t < \infty$ as

$$K_{\mathscr{D}}(t) = \sup_{x \in C[0,1]} \mathscr{D}[u : ||u - x||_c \leqq t] , \tag{2.2}$$

where $||\cdot||_c$ denotes the sup-norm in $C[0, 1]$. Evidently, $K_{\mathscr{D}}(t)$ measures the maximum probability assignable by the distribution $\mathscr{D}$ to a ball of radius t in $C[0, 1]$. The representation of $\mathscr{D}$ as the convolution $\mathscr{D}_B * \mathscr{D}_1$ has the following consequence: for even $t > 0$,

$$K_{\mathscr{D}}(t) \leqq K_{\mathscr{D}_B}(t) . \tag{2.3}$$

Indeed, if $A_t = \{u : ||u||_c \leqq t\}$, then

$$\begin{aligned} K_{\mathscr{D}}(t) &= \sup_{x \in C[0,1]} \mathscr{D}[A_t + x] \\ &= \sup_{x \in C[0,1]} \int \mathscr{D}_B[A_t + x - y]\, d\mathscr{D}_1(y) , \end{aligned} \tag{2.4}$$

which implies (2.3). The argument remains valid if A_t is replaced by any other measurable subset of $C[0, 1]$.

Another comparison between $\mathscr{D}_B$ and $\mathscr{D}$ can be made using risk functions. Let $w : C[0, 1] \rightarrow R^1$ be convex and symmetric and let B and T denote random functions in $C[0, 1]$, independently distributed according to $\mathscr{D}_B$ and $\mathscr{D}_1$ respectively. The random function $S = B + T$ is distributed according to $\mathscr{D}$. Then

$$E\{w(B)\} \leqq E\{w(S)\} . \tag{2.5}$$

Indeed, by the properties of w (cf. Section 7 of Dvoretzky et al. (1956)),

$$2w(B) \leqq w(B + T) + w(-B + T) . \tag{2.6}$$

Since the distribution of $B + T$ is the same as that of $-B + T$, (2.6) implies the inequality (2.5). If w is also continuous on $C[0, 1]$, it follows that $\lim_{n\rightarrow\infty} \inf Ew\{n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)]\} \geqq E\{w(B)\}$ for every regular estimator $\hat{F}_n$. An interesting question is whether this argument can be extended to provide a proof for the heuristics in Section 7 of Dvoretzky et al. (1956).

We turn now to establishing the convolution representation for $\mathscr{D}$. Let

$$L_n = 2 \log \prod_{i=1}^n [f_n^{\frac{1}{2}}(X_{i,n})/f^{\frac{1}{2}}(X_{i,n})] . \tag{2.7}$$

The following lemma, whose proof can be inferred from Le Cam (1969), will be used.

LEMMA. *Let* $\{f_m\}$ *be a sequence of densities such that* (2.1) *holds for some* $\delta \in L_2$. *Then for every* $\varepsilon > 0$,

$$\lim_{n\rightarrow\infty} P_f[|L_n - 2n^{-\frac{1}{2}} \sum_{i=1}^n \delta(X_{i,n})/f^{\frac{1}{2}}(X_{i,n}) + 2||\delta||^2| > \varepsilon] = 0 . \tag{2.8}$$

This lemma implies that for every density sequence $\{f_m\} \in \mathscr{C}(f, \delta)$, the sequence of joint distributions $\{\prod_{i=1}^n f_n(x_{i,n})\}$ is contiguous to the sequence $\{\prod_{i=1}^n f(x_{i,n})\}$ (see, for instance, Le Cam (1969)).

PROOF OF THE THEOREM. Under (2.1),

$$\lim_{n\to\infty} \sup_{0\le x\le 1} |n^{\frac{1}{2}}[F_n(x) - F(x)] - 2\int_0^x \delta(t) f^{\frac{1}{2}}(t)\, d\mu| = 0\,, \tag{2.9}$$

where F_n is the cdf of f_n. Therefore, the characteristic functional of $n^{\frac{1}{2}}[\hat{F}_n - F_n]$ under f_n is

$$\begin{aligned} E_{f_n} \exp[i \textstyle\int_0^1 n^{\frac{1}{2}}[\hat{F}_n(x) - F_n(x)]\, dv(x)] \\ = E_f \exp[i \textstyle\int_0^1 n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)]\, dv(x) \\ - i \textstyle\int_0^1 dv(x) \int_0^x 2\delta(t) f^{\frac{1}{2}}(t)\, d\mu + L_n] + o(1)\,, \end{aligned} \tag{2.10}$$

v being any function of bounded variation on $[0, 1]$.

Choose $\delta(x) = \sigma^{-1} h[v(x) - \int_0^1 v(x) f(x)\, d\mu] f^{\frac{1}{2}}(x)$, where $\sigma^2 = \int_0^1 v^2(x) f(x)\, d\mu - [\int_0^1 v(x) f(x)\, d\mu]^2$ and h is real. Note that $\delta \perp f^{\frac{1}{2}}$ and $||\delta||^2 = h^2$. By considering only a subsequence if necessary, we can assume that under f, the processes $\{(n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)], n^{-\frac{1}{2}} \sum_{i=1}^n \delta(X_{i,n}) f^{-\frac{1}{2}}(X_{i,n}))\}$ converge weakly in $C[0, 1] \times R^1$ to a process $(S(x), Z)$, depending on the chosen δ, such that Z has a $N(0, 1)$ distribution. It follows from (2.8) and the choice of δ that the random processes $\{n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)], L_n\}$ converge weakly to the random process $(S(x), 2hZ - 2h^2)$ for every real h.

Evidently,

$$\begin{aligned} E_f |\exp[i \textstyle\int_0^1 n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)]\, dv(x) + L_n]| \\ = 1 = E|\exp[i \textstyle\int_0^1 S(x)\, dv(x) + 2hZ - 2h^2]|\,. \end{aligned} \tag{2.11}$$

Moreover, there exists a probability space and versions of $\{(n^{\frac{1}{2}}[\hat{F}_n(x) - F(x), L_n)\}$ and $(S(x), 2hZ - 2h^2)$ defined on that space such that convergence with probability one holds as well as weak convergence (cf. Skorokhod (1956)). Since (2.11) remains true for these versions, Vitali's theorem gives

$$\begin{aligned} \lim_{n\to\infty} E_f \exp[i \textstyle\int_0^1 n^{\frac{1}{2}}[\hat{F}_n(x) - F(x)]\, dv(x) + L_n] \\ = E \exp[i \textstyle\int_0^1 S(x)\, dv(x) + 2hZ - 2h^2]\,. \end{aligned} \tag{2.12}$$

Hence from (2.10), (2.12) and regularity of $\hat{F}_n$,

$$\begin{aligned} E \exp[i \textstyle\int_0^1 S(x)\, dv(x)] \\ = E \exp[i \textstyle\int_0^1 S(x)\, dv(x) + 2hZ] \\ \times \exp[-2i \textstyle\int_0^1 dv(x) \int_0^x \delta(t) f^{\frac{1}{2}}(t)\, d\mu - 2h^2] \\ = E \exp[i \textstyle\int_0^1 S(x)\, dv(x) + 2hZ] \exp[2i \int_0^1 \delta(x) f^{\frac{1}{2}}(x) v(x)\, d\mu - 2h^2]\,. \end{aligned} \tag{2.13}$$

For v a function of bounded variation and t real, let $\varphi(v, t) = E \exp[i \int_0^1 S(x)\, dv(x) + itZ]$. Equation (2.13) becomes

$$\varphi(v, 0) = E \exp[i \textstyle\int_0^1 S(x)\, dv(x) + 2hZ] \exp[2ih\sigma - 2h^2]\,. \tag{2.14}$$

The right side of (2.14) is analytic in h, constant for all real h, hence constant for all complex h. In particular, the choice $h = it/2$ yields

$$\varphi(v, 0) = \varphi(v, t)\exp[-\sigma t + t^2/2] = \varphi(v, t)\exp[(t-\sigma)^2/2]\exp[-\sigma^2/2] \tag{2.15}$$

for all real t and all functions of bounded variation v on $[0, 1]$. The special choice $t = \sigma$ gives

$$\varphi(v, 0) = \varphi(v, \sigma)\exp(-\sigma^2/2)\,. \tag{2.16}$$

Evidently, $\varphi(v, 0)$ is the characteristic functional of $\mathscr{D}$ while $\exp(-\sigma^2/2)$ is the characteristic functional of the Brownian bridge process $B(x)$ and $\varphi(v, \sigma)$ is the characteristic functional of $S(x) - B(x)$. Thus (2.16) is equivalent to the theorem conclusion.

We remark that the proof above is related to Bickel's proof for Hájek's representation theorem (see Roussas (1972) for a published version) rather than to Hájek's (1970) original argument.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 323–339.

LE CAM, L. (1969). *Théorie Asymptotique de la Décision Statistique.* Presses de l'Université de Montréal.

ROUSSAS, G. (1972). *Contiguity of Probability Measures.* Cambridge Univ. Press.

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 157–214.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720