CONSISTENT ESTIMATES AND ZERO-ONE SETS[1]

BY LEO BREIMAN, LUCIEN LECAM AND LORRAINE SCHWARTZ

University of California, Los Angeles and Berkeley, and University of British Columbia

1. Introduction. Let Ω be a space of points ω with a σ-field $\mathfrak{F}$ on Ω. Take $\{P_\theta(\cdot),\ \theta\ \varepsilon\ \Theta\}$ to be a family of probability measures on $\mathfrak{F}$ for which there is a σ-field $\mathfrak{B}$ on Θ with $P_\theta(A)$ $\mathfrak{B}$-measurable for fixed $A\ \varepsilon\ \mathfrak{F}$. The present paper answers the following two questions. Let the function f on Θ be real-valued and $\mathfrak{B}$-measurable. Then (a) when does there exist an $\mathfrak{F}$-measurable function $\hat{f}$ such that

$$P_\theta\{\omega:\hat{f}(\omega) = f(\theta)\} = 1, \qquad \text{all } \theta?$$

(b) If there is a probability measure Q on $\mathfrak{B}$, when does there exist an $\mathfrak{F}$-measurable $\hat{f}$ such that

$$P_\theta\{\omega:\hat{f}(\omega) = f(\theta)\} = 1, \qquad \text{a.s.} \quad \theta?$$

We deal with (a) in Section 2, with (b) in Section 3, and apply these results to sequences of independent random variables in Section 4.

The basic device is to look at those sets $A\ \varepsilon\ \mathfrak{F}$ which have the property $P_\theta^2(A) = P_\theta(A)$ for all θ or for almost all θ, and the essential condition is that, roughly, the field $\mathfrak{F}$ be rich enough in this type of set.

2. Question (a). We shall say that $A\ \varepsilon\ \mathfrak{F}$ is a *strong zero-one set* if for each $\theta\ \varepsilon\ \Theta$ the value $P_\theta(A)$ is either zero or one. Let $\mathfrak{C}$ be the class of all zero-one sets. Reflect $\mathfrak{C}$ into the parameter set Θ by defining the class $\mathfrak{S} \subset \mathfrak{B}$ in the following way: $S\ \varepsilon\ \mathfrak{S}$ if there is a set $A\ \varepsilon\ \mathfrak{C}$ such that

$$S = \{\theta:P_\theta(A) = 1\}.$$

PROPOSITION 1. *$\mathfrak{C}$ and $\mathfrak{S}$ are σ-fields.*

PROOF. Evidently, $\mathfrak{C}$ is closed under complementation. It is also immediate that $\mathfrak{C}$ is closed under countable unions since if $A_i\ \varepsilon\ \mathfrak{C}$, then $P_\theta(\bigcup_i A_i) = 1$ if one of $P_\theta(A_i) = 1$, and equals zero otherwise. Futhermore, $\mathfrak{S}$ is closed under complementation, and if $S_i\ \varepsilon\ \mathfrak{S}$ let $A_i\ \varepsilon\ \mathfrak{C}$ be such that $S_i = \{\theta:P_\theta(A_i) = 1\}$; then $\bigcup_i S_i = \{\theta:P_\theta(\bigcup_i A_i) = 1\}\ \varepsilon\ \mathfrak{S}$.

Let $\mathfrak{B}(f)$ be the minimal σ-field with respect to which f is measurable. Then

THEOREM 1. *There exists an $\mathfrak{F}$-measurable function $\hat{f}$ such that $P_\theta\{\hat{f} = f(\theta)\} = 1$ for all θ, if and only if $\mathfrak{B}(f) \subset \mathfrak{S}$.*

PROOF. One way is easy. Suppose there does exist such a function $\hat{f}$. Consider

Received 21 August 1962; revised 4 October 1963.

[1] This research was prepared with the partial support of the Information Systems Branch of the Office of Naval Research under Contract Nonr 222(53), and of the U. S. Army Research Office (Durham), Grant DA-ARO(D)-31-124-G83.

the set $S_\alpha = \{\theta : f(\theta) \geqq \alpha\}$. Then $P_\theta\{\hat{f} \geqq \alpha\} = 1$ if $\theta \varepsilon S_\alpha$ and equals zero if θ is in the complement S_α^c. Hence $S_\alpha \varepsilon \mathcal{S}$ which implies that $\mathcal{B}(f) \subset \mathcal{S}$. To go the other way, take $\mathcal{B}(f) \subset \mathcal{S}$. We may suppose f is bounded for if not, take g to be 1-1, Borel measurable, mapping the real line into [0, 1]. If we show that there is an $\hat{f}$ such that $P_\theta\{\hat{f} = g(f(\theta))\} = 1$, all θ, then $P_\theta\{g^{-1}(\hat{f}) = f(\theta)\} = 1$, all θ. By the boundedness of f, there is a sequence $\{f_n\}$ of $\mathcal{S}$-measurable simple functions which converge uniformly to f. Write $f_n = \sum_i \alpha_i I_{S_i}$, where I_{S_i} denotes the set indicator of S_i and let $A_i \varepsilon \mathcal{C}$ be any set such that $S_i = \{\theta : P_\theta(A_i) = 1\}$. Note that if S_i and S_j are disjoint then $P_\theta(A_i \cap A_j) = 0$, all θ. Therefore if we define $\hat{f}_n = \sum_i \alpha_i I_{A_i}$, then $\hat{f}_n$ is well except on a set B_n such that $P_\theta(B_n) = 0$, all θ. It is easily verified that $P_\theta\{\hat{f}_n = f_n(\theta)\} = 1$, all θ. By the uniform convergence of f_n to f, for every $\epsilon > 0$ and n sufficiently large we have

$$P_\theta\{f(\theta) - \epsilon \leqq \hat{f}_n \leqq f(\theta) + \epsilon\} = 1, \quad \text{all} \quad \theta.$$

The limit $\hat{f} = \lim_n \hat{f}_n$ is well defined except possibly for a set B with $P_\theta(B) = 0$ all θ, and set it equal to zero on B. Then for every $\epsilon > 0$,

$$P_\theta\{f(\theta) - \epsilon \leqq \hat{f} \leqq f(\theta) + \epsilon\} = 1, \quad \text{all} \quad \theta,$$

and, letting ϵ go to zero, this completes the proof of the theorem.

There is another way of looking at Theorem 1 in terms of the orthogonality of the measures involved. Recall that two probabilities P_1 and P_2 are said to be *orthogonal* if there is a measurable set A such that $P_1(A) = 1$ and $P_2(A) = 0$. Theorem 1 implies that to estimate a function f consistently it is not sufficient that the probabilities P_θ be pairwise orthogonal; i.e., that for $\theta_1 \neq \theta_2$, P_{θ_1} be orthogonal to P_{θ_2}. In other words, it is not enough that the probabilities P_θ "separate points." What is necessary and sufficient is that the P_θ "separate sets," that is, given any set $S \varepsilon \mathcal{B}(f)$ there is a set A such that $P_\theta(A) = 1$ for $\theta \varepsilon S$ and $P_\theta(A) = 0$, $\theta \epsilon S^c$. This becomes clearer in the following important special case.

Let Θ be the unit interval $I = [0, 1]$. Then

Corollary 1. *In order that for every Borel measurable function f there exist an $\mathcal{F}$-measurable function $\hat{f}$ such that $P_\theta\{\hat{f} = f(\theta)\} = 1$, all θ, it is necessary and sufficient that for every $\alpha \varepsilon I$ there exist an $A_\alpha \varepsilon \mathcal{F}$ such that*

$$P_\theta(A_\alpha) = 1 \quad \text{if} \quad \theta \geqq \alpha,$$
$$= 0 \quad \text{if} \quad \theta < \alpha.$$

Proof. Suppose we wish to estimate $f(\theta) = \theta$. Then $\mathcal{B}(f)$ is the Borel field and hence it is necessary that $\mathcal{S}$ be the Borel field, so that the condition of the corollary is necessary. If the condition is satisfied, then all intervals of the form $[\alpha, 1]$ are in $\mathcal{S}$, implying that $\mathcal{S}$ is the Borel field and hence contains $\mathcal{B}(f)$ for f Borel-measurable.

3. Question (b). Let there be a probability measure Q on $\mathcal{B}$. We will say that $A \varepsilon \mathcal{F}$ is a *weak zero-one set* with respect to Q if $P_\theta^2(A) = P_\theta(A)$ except for a set of θ values whose Q measure is zero. Let $\bar{\mathcal{C}}$ be the class of all weak zero-one sets with respect to Q. Reflect $\bar{\mathcal{C}}$ into the parameter space by defining a class of sets $\bar{\mathcal{S}} \subset \mathcal{B}$ such that $S \varepsilon \bar{\mathcal{S}}$ if there is a set $A \varepsilon \bar{\mathcal{C}}$ such that S differs by at most a set of Q measure zero from the set $\{\theta: P_\theta(A) = 1\}$.

PROPOSITION 2. *$\bar{\mathcal{C}}$ and $\bar{\mathcal{S}}$ are σ-fields.*

PROOF. Again, both $\bar{\mathcal{C}}$ and $\bar{\mathcal{S}}$ are closed under complementation. Let $A_i \varepsilon \bar{\mathcal{C}}$; then the A_i are zero-one sets on $(\bigcup_i B_i)^c$ where B_i is the set off which $P_\theta(A_i)$ equals zero or one. $\bigcup_i B_i$ has Q measure zero and for θ in its complement $P_\theta(\bigcup_i A_i) = 1$ if one of $P_\theta(A_i) = 1$ and equals zero if all of $P_\theta(A_i) = 0$ so that $\bigcup_i A_i \varepsilon \bar{\mathcal{C}}$. In a way similar to that in Proposition 1 it can be shown that $\bar{\mathcal{S}}$ is also closed under countable unions.

If f is $\mathcal{B}$-measurable, then

THEOREM 2. *There exists an $\mathcal{F}$-measurable function $\hat{f}$ such that $P_\theta\{\hat{f} = f(\theta)\} = 1$ for almost all θ, (Q) if and only if $\mathcal{B}(f) \subset \bar{\mathcal{S}}$.*

PROOF. The proof is very similar to that of Theorem 1. If there exists such a function $\hat{f}$, let S_α be the set $[\theta: f(\theta) \geqq \alpha\}$. Then a.s. $P_\theta\{\hat{f} \geqq \alpha\} = 1$ if $\theta \varepsilon S_\alpha$ and equals zero if $\theta \varepsilon S_\alpha^c$. Therefore $S_\alpha \varepsilon \bar{\mathcal{S}}$ and $\mathcal{B}(f) \subset \bar{\mathcal{S}}$. The argument in the other direction very nearly duplicates that of Theorem 1.

We may rephrase the above theorem in terms of orthogonality. Let $\mathcal{D} \subset \mathcal{B}(f)$ be the set of generators of $\mathcal{B}(f)$; that is $\mathcal{B}(f)$ is the minimal σ-field containing $\mathcal{D}$.

COROLLARY 2. *There exists an $\mathcal{F}$-measurable function $\hat{f}$ such that $P_\theta\{\hat{f} = f(\theta)\} = 1$* a.s. (Q) *if and only if for every $S \varepsilon \mathcal{D}$ the two probabilities*

$$\int_S P_\theta(\cdot)Q(d\theta), \qquad \int_{S^c} P_\theta(\cdot)Q(d\theta)$$

are orthogonal.

PROOF. If $\mathcal{B}(f) \subset \bar{\mathcal{S}}$, then for $S \varepsilon \mathcal{D}$ there is a set $A \varepsilon \mathcal{F}$ such that a.s.

$$\begin{aligned} P_\theta(A) &= 1, \qquad \theta \varepsilon S, \\ &= 0, \qquad \theta \varepsilon S^c, \end{aligned}$$

which establishes the orthogonality of the two probabilities of the corollary. Conversely, if they are orthogonal, then there is a set $A \varepsilon \mathcal{F}$ such that $P_\theta(A) = 1$ or 0 as θ is in S or S^c, a.s. (Q).

If Θ is the unit interval I and $\mathcal{B}$ the Borel field, we may state

COROLLARY 3. *In order that for every Borel-measurable f there exist an $\mathcal{F}$-measurable $\hat{f}$ such that $P_\theta\{\hat{f} = f(\theta)\} = 1$* a.s. (Q) *it is necessary and sufficient that for every $\alpha \varepsilon I$ the two probabilities*

$$\int_{[\alpha,1]} P_\theta(\cdot)Q(d\theta) \text{ and } \int_{[0,\alpha)} P_\theta(\cdot)Q(d\theta)$$

be orthogonal.

PROOF. The proof follows that of Corollary 1.

4. Application to sequences of independent random variables. In this section we start with a space $\hat{\Omega}$ carrying a σ-field $\hat{\mathcal{F}}$ and a family of probabilities $\hat{P}_\theta(\cdot)$

on $\hat{\mathfrak{F}}$ which for fixed $\hat{A}$ in $\hat{\mathfrak{F}}$ are $\mathcal{B}$-measurable functions. Let Ω be the product space $\prod_1^\infty \hat{\Omega}_i$ where each $\hat{\Omega}_i$ is a copy of $\hat{\Omega}$, with $\mathfrak{F}$ the usual product σ-field in Ω constructed starting with the copies $\hat{\mathfrak{F}}_i$ of $\hat{\mathfrak{F}}$ in $\hat{\Omega}_i$. Take $P_\theta(\cdot)$ on $\mathfrak{F}$ to be the product measure constructed from $\hat{P}_\theta(\cdot)$ on $\mathfrak{F}$. It is not difficult to prove that $P_\theta(A)$, for fixed $A \varepsilon \mathfrak{F}$, is again $\mathcal{B}$-measurable.

Let $\mathcal{B}_1$ be the minimal σ-field with respect to which $\hat{P}_\theta(\hat{A})$ is a measurable function on Θ for every $\hat{A} \varepsilon \hat{\mathfrak{F}}$ and let $\overline{\mathcal{B}}_1$ be the completion of $\mathcal{B}_1$ with respect to Q.

THEOREM 3. *Let f be $\mathcal{B}$-measurable. Then there exits $\hat{f}$, $\mathfrak{F}$-measurable, such that*

$$P_\theta\{\hat{f} = f(\theta)\} = 1 \quad \text{all} \quad \theta,$$

$$P_\theta\{\hat{f} = f(\theta)\} = 1 \quad \text{a.s.} \quad (Q),$$

if and only if, respectively,

$$\mathcal{B}(f) \subset \mathcal{B}_1, \qquad \mathcal{B}(f) \subset \overline{\mathcal{B}}_1.$$

PROOF. By definition Ω is the set of all infinite sequences $\omega = (x_1, x_2, \cdots)$, $x_i \varepsilon \hat{\Omega}_i$. For any given $\hat{A} \varepsilon \hat{\mathfrak{F}}$ and $\alpha \varepsilon [0, 1]$, let $A_\alpha \varepsilon \mathfrak{F}$ be the set

$$A_\alpha = \left\{\omega: \lim_n \frac{1}{n} \sum_{i=1}^n I_{\hat{A}}(x_i) \geqq \alpha\right\}.$$

By the law of large numbers,

$$\begin{aligned} P_\theta(A_\alpha) &= 1, \qquad \hat{P}_\theta(\hat{A}) \geqq \alpha, \\ &= 0, \qquad \hat{P}_\theta(\hat{A}) < \alpha. \end{aligned}$$

This implies that A_α is a strong zero-one set, and also that $\mathcal{B}_1 \subset \mathcal{S}$ since each $\mathcal{B}(\hat{P}_\theta(\hat{A}))$ is, $\hat{A}$ any set in $\hat{\mathfrak{F}}$. Further, $\mathcal{B}_1 \subset \overline{\mathcal{S}}$, so that by Theorems 1 and 2 the condition $\mathcal{B}(f) \subset \mathcal{B}_1$, $\mathcal{B}(f) \subset \overline{\mathcal{B}}_1$, respectively, of Theorem 3 is seen to be sufficient for the existence of $\hat{f}$. Let us go the other way. If $P_\theta\{\hat{f} = f(\theta)\} = 1$ all θ, then for every $B \varepsilon \mathcal{B}(f)$ there is an $\mathfrak{F}$-measurable function $\hat{g}$ such that $P_\theta\{\hat{g} = I_B(\theta)\} = 1$ for all θ. Thus, except possibly on a set A_0 such that $P_\theta(A_0) = 0$ for all θ, $\hat{g}$ is the set indicator of some $A \varepsilon \mathfrak{F}$, so that

$$P_\theta(A) = 1 \quad \text{for} \quad \theta \varepsilon B, \qquad P_\theta(A^c) = 1 \quad \text{for} \quad \theta \varepsilon B^c.$$

Since $P_\theta(A)$ is measurable with respect to $\mathcal{B}_1$, we conclude that $B \varepsilon \mathcal{B}_1$. A similar argument goes through for the condition involving $\overline{\mathcal{B}}_1$.

To give a more concrete application, let $X_1, X_2, \cdots$ be a sequence of independent random variables, each having the distribution function $F(\cdot; \theta)$, where $\theta \varepsilon [0, 1]$, and for any fixed x assume $F(x; \cdot)$ is Borel measurable.

COROLLARY 4. *The necessary and sufficient condition for the existence of a function $\hat{f}(X_1, X_2, \cdots)$, measurable in the appropriate sense, such that*

$$P_\theta\{\hat{f}(X_1, X_2, \cdots) = \theta\} = 1 \quad \textit{for all} \quad \theta \varepsilon [0, 1]$$

is that the Borel field be the smallest σ-field with respect to which all the functions $F(x; \cdot)$ on $\Theta = [0, 1]$ are measurable.

PROOF. This follows immediately from Theorem 3. The corresponding condition when "all θ" is replaced by "almost all $\theta(Q)$" is clear and its formulation is left to the reader.

The existence of the functions $\hat{f}$ discussed in the present paper is connected with problems of estimation in the following way. Assume that the σ-field $\mathfrak{F}$ is the smallest σ-field containing an increasing sequence $\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \cdots$ of subσ-fields. In the usual estimation problems it may be desired to find a sequence $\{\hat{f}_n\}$ of functions such that $\{\hat{f}_n\}$ is $\mathfrak{A}_n$-measurable and

$$P_\theta\{\lim \hat{f}_n = f(\theta)\} = 1 \quad \text{for every} \quad \theta \,\varepsilon\, \Theta.$$

It is easily seen that the existence of such a sequence implies the existence of a function $\hat{f}$ such that

$$P_\theta\{\hat{f} = f(\theta)\} = 1 \quad \text{for every} \quad \theta \,\varepsilon\, \Theta.$$

Conversely if such a function $\hat{f}$ exists and if Q is a probability measure on $\mathfrak{B}$ the (martingale) argument of Doob [1] shows that there is a sequence $\{\hat{f}_n\}$ such that $P_\theta\{\lim \hat{f}_n = f(\theta)\} = 1$ for almost all $\theta \,\varepsilon\, \Theta$. In this paper Doob shows that Bayes estimators of f when f is the identity map, $f(\theta) = \theta$, and Q is an arbitrary probability measure on the Borel sets of the line, are consistent for almost all θ. To establish the existence of the function $\hat{f}$ Doob relies on the theorem according to which, in complete separable metric spaces, the inverse of a one-to-one Borel map is a Borel map. Corollary 4 makes the recourse to such a theorem unnecessary. Further, Theorem 3 makes it possible to obtain similar results for Bayes estimators of functions f on Θ even if $\hat{\Omega}$ is arbitrary and if the σ-field $\hat{\mathfrak{F}}$ does not have a countable set of generators.

In some cases, for instance the independent, identically distributed case treated in [2] it has been found possible to give explicit necessary and sufficient conditions for the existence of a sequence $\{\hat{f}_n\}$ such that $\hat{f}_n \to f(\theta)$ in probability for each θ. The gap between the consistency for every θ and consistency for almost every θ is discussed further in [3].

REFERENCES

[1] DOOB, J. L. (1948). Application of the theory of martingales. *Coll. Int. du C.N.R.S. Paris*. 22–28.

[2] LECAM, L. and SCHWARTZ, LORRAINE (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31** 140–150.

[3] SCHWARTZ, LORRAINE (1960). Consistency of Bayes' procedures. Unpublished Ph.D. Thesis. Univ. of California, Berkeley.