The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1429–1433

IDENTIFYING PROBABILITY LIMITS[1]

BY GORDON SIMONS

University of North Carolina

1. Introduction and summary. The literature of mathematical statistics is filled with theorems on (weakly) consistent estimators.[2] Even though most statisticians want stronger evidence of an estimator's worth, these theorems have provided some comfort for the applied statistician. In this paper, we begin an investigation into the concept of consistency and, more specifically, investigate the extent to which a consistent sequence of estimators identifies the parameter they estimate.

It will be recalled that for any sequence of random variables which converge in probability to a limit, there is a subsequence which converges almost surely to that limit. This would seem to suggest that if one is given a consistent sequence of estimators $\hat{\phi}_1, \hat{\phi}_2, \cdots$ converging to $\phi(\theta)$ $(\theta \in \Theta)$, say, then one can find a subsequence which converges almost surely. That is, whenever there exists a weakly consistent sequence of estimators there exists a strongly consistent sequence as well. Unfortunately, the specific subsequence may depend upon the unknown parameter value θ. Still, an applied statistician might be able to choose sequentially which observed estimators to include in a (random) subsequence. This is easily seen to be equivalent to postulating the existence of functions $g_n(x_1, \cdots, x_n)$ $(n \geqq 1)$ such that $g_n(\hat{\phi}_1, \cdots, \hat{\phi}_n)$ converges to $\phi(\theta)$ with probability one $(\theta \in \Theta)$. In Section 2, we will show that such functions do not always exist.

It seems appropriate, therefore, to question whether the values of the *entire* sequence $\hat{\phi}_1, \hat{\phi}_2, \cdots$ (always) allow one to determine the value of $\phi(\theta)$ with probability one. We prefer the following mathematical reformulation of this question: Let $(\Omega, \mathscr{A}, P)$ be a probability space and $\mathscr{E}$ denote the set of infinite dimensional random vectors $\mathscr{X} = (X_1, X_2, \cdots)$ defined on this space whose coordinates converge *in probability* to a random variable (which we shall continue to denote as) $p\mathscr{X}$. The question becomes: Does there always exist a function f which maps R^∞ (infinite-dimensional Euclidean space) into R (the reals) such that for every $\mathscr{X} \in \mathscr{E}$, the set

$$(1) \qquad [f(\mathscr{X}) \neq p\mathscr{X}] \quad \text{is contained in a null set of } \mathscr{A}?^{3}$$

Received June 1, 1970; revised November 9, 1970.

[1] This research was partially supported by the National Science Foundation under Grant GU-2059.

[2] We contrast *weak* consistency (or simply consistency) with *strong* consistency. The former requires a sequence of estimators to converge in probability; the latter requires them to converge almost surely. This is common terminology.

[3] The notation $\mathscr{X}$ is being used in two distinct ways. $p\mathscr{X}$ is a function on the *random vectors* $\mathscr{X} \in \mathscr{E}$. $f(\mathscr{X})$ is a function on the *values* of $\mathscr{X}$, the values being points in R^∞. Our choice of notation is shorter than the possibly more appropriate notation $p(\mathscr{X}(\cdot))$ and $f(\mathscr{X}(\omega))$.

We shall refer to any function f which satisfies (1) for all $\mathscr{X} \in \mathscr{F} \subset \mathscr{E}$ as a *probability limit identification function* (PLIF) *on* $\mathscr{F}$. We *partially* justify the reformulation as follows: In Section 3, we will show that the vector of estimators $(\hat{\phi}_1, \hat{\phi}_2, \cdots)$ can be interpreted as defined on the *same* probability space for *every* $\theta \in \Theta$. As such, consistent estimators are equivalent to a family of vectors $\mathscr{F} = \{\mathscr{X}_\theta, \theta \in \Theta\} \subset \mathscr{E}$. Identifying $\phi(\theta)$ becomes equivalent to showing that there exists a PLIF on $\mathscr{F}$.

In Section 4, we show that there exists a PLIF on $\mathscr{E}$ if there exists a PLIF on $\mathscr{E}^*$, the set of $\mathscr{X} \in \mathscr{E}$ whose coordinates are Bernoulli variables and whose probability limit $p\mathscr{X}$ is almost surely a constant (necessarily zero or one). With values of θ corresponding to vectors $\mathscr{X} \in \mathscr{E}^*$ and $\phi(\theta)$ corresponding to $p\mathscr{X}$, the reformulation becomes complete.

We do not know whether a PLIF always exists on $\mathscr{E}^*$ except for certain elementary probability spaces. It is hoped that the current paper will stimulate further research into this question. If they do not always exist, this will cast further doubt on the importance of consistency.

Breiman, LeCam and Schwartz [2] have discussed an interesting problem whose formulation is closely related to the current one. They assume that they have a family of probability measures $\{P_\theta(\cdot), \theta \in \Theta\}$ each defined on the same measurable space $(\Omega, \mathscr{A})$ (with points $\omega \in \Omega$). They assume that $\phi(\theta)$ is measurable with respect to a σ-field defined on Θ and find necessary and sufficient conditions for the existence of an $\mathscr{A}$-measurable estimator $\hat{\phi}(\omega)$ for which

$$P_\theta\{\hat{\phi}(\omega) = \phi(\theta)\} = 1 \qquad \text{for all } \theta \in \Theta$$

(and also for a closely related condition). The question of the existence of a measurable PLIF on $\mathscr{E}^*$ translates into the question of the existence of a certain "zero-one set" in their context. Skibinski [3] has connected their work on zero-one sets with some work of Bahadur [1].

2. Weak and strong consistency. In this section, we demonstrate an estimation problem in which weakly consistent estimators exist but strongly consistent estimators do not. Let $Y_1, Y_2, \cdots$ be independent Bernoulli variables with arbitrary means $p_1, p_2, \cdots$ and $\{n_k, k \geqq 1\}$ be an arbitrary sequence of increasing positive integers. Set $n_0 = 0$ and

$$X_m = Y_k \quad \text{for } n_{k-1} < m \leqq n_k, \qquad k \geqq 1. \tag{2}$$

Finally, let $\mathscr{C}$ denote the set of random vectors $\mathscr{X} = (X_1, X_2, \cdots)$ which arise from (2) and the condition

$$p_k \to 0 \quad \text{or} \quad 1 \qquad \text{as } k \to \infty. \tag{3}$$

Let $\theta = (p_1, p_2, \cdots, n_1, n_2, \cdots)$ and $\phi(\theta) = \lim_{n\to\infty} p_n$. Then $\phi(\theta) = p\mathscr{X}$ and the sequence of estimators $\hat{\phi}_n = X_n$ $(n \geqq 1)$ is consistent for $\phi(\theta)$. Now assume that a strongly consistent estimating sequence exists. Then, there exists a sequence of functions $g_n(x_1, \cdots, x_n)$ taking values zero or one such that

$$\lim_{n\to\infty} g_n(X_1, \cdots, X_n) = \lim_{k\to\infty} p_k \tag{4}$$

almost surely for all $\mathscr{X} \in \mathscr{C}$.

Consider the situation $p_k = 0$ for all $k > j \geqq 1$. (4) requires that there exists a positive integer m_j such that for $m \geqq m_j$ and any partial sequence $x_1, \cdots, x_j$ of zeros and ones, $g_{j+m}(x_1, \cdots, x_j, 0, 0, \cdots, 0) = 0$. That is, independent of past history, a sufficiently long string of zeros will force the current g function to be zero. A similar thing is true for long strings of ones. Next consider a situation in which $\sum_{k=1}^{\infty} p_k = \infty$ and $\sum_{k=1}^{\infty}(1-p_k) = \infty$. It follows from the Borel-Cantelli lemma that the sequence $Y_1, Y_2, \cdots$ has infinitely many zeros and infinitely many ones. Clearly, by allowing the sequence $\{n_k\}$ to grow rapidly enough, we can insure that the sequence $X_1, X_2, \cdots$ will have sufficiently large strings of zeros and large strings of ones to bring about the nonexistence of the first limit in (4). This is a contradiction.

We do not know if a PLIF exists on $\mathscr{C}$. (Note that $\mathscr{C} \subset \mathscr{E}^*$.) We suspect that the class $\mathscr{C}$ would be a good class to work with in further research. There is a somewhat contrived subset $\mathscr{C}^* \subset \mathscr{C}$ on which a PLIF exists but for which no strongly consistent estimator exists (for the same reason as given above). Let $\mathscr{C}^*$ be the set of $\mathscr{X} \in \mathscr{C}$ for which the sequence $\{n_k\}$ has finitely many even (odd) integers when $\lim_{k\to\infty} p_k$ equals zero (one). The details are left to the reader.

3. Interpreting estimators as subsets of $\mathscr{E}$. Let $X_1', X_2', \cdots$ be a sequence of random variables defined on some probability space $(\Omega', \mathscr{A}', P')$ and suppose that the probability space $(\Omega, \mathscr{A}, P)$ (referred to in Section 1) admits a uniformly distributed variable U. One can show that $(\Omega, \mathscr{A}. P)$ also admits a sequence of random variables $X_1, X_2, \cdots$ such that for each $n \geqq 1$, the laws of $X_1', \cdots, X_n'$ and $X_1, \cdots, X_n$ agree. Briefly, one begins by showing that there exists a (one-to-one) measurable mapping of U into a sequence $U_1, U_2, \cdots$ of independent uniformly distributed random variables (since there exists a one-to-one mapping between a uniform variable and a sequence of i.i.d. Bernoulli variables with mean $\frac{1}{2}$). One defines $X_1 = \inf\{x\colon P'\{X_1' \leqq x\} \geqq U_1\}$ and then defines X_n $(n > 1)$ recursively using conditional distribution functions.

Now let $X_1', X_2', \cdots$ represent a sequence of consistent estimators of $\phi(\theta)$, $\theta \in \Theta$. Whereas the statistician often finds it convenient to view this sequence as one sequence of random variables and prefers to think of the distributions of finite sets of them as depending upon the parameter θ, the probabilist is inclined to say that the sequence of estimators represents a different sequence of random variables—one for each θ—since he views these random variable sequences as defined on probability spaces which can depend on θ. Taking this latter view, we can identify the sequence $X_1', X_2', \cdots$ *for each* $\theta \in \Theta$ with a random vector $\mathscr{X}_\theta = (X_1, X_2, \cdots) \in \mathscr{E}$. As such, a consistent sequence of estimators becomes identified with a subset of $\mathscr{E}$ as θ ranges over Θ.

4. Probability limit identification functions. The main result of this section is the following:

THEOREM. *There exists a* PLIF *on* $\mathscr{E}$, *if, and only if, there exists a* PLIF *on* $\mathscr{E}^*$.

PROOF. Suppose that f is a PLIF on $\mathscr{E}^*$. Let $\mathscr{X} = (X_1, X_2, \cdots) \in \mathscr{E}$ be arbitrary. Set

$$X_n^{(a)} = I_{[X_n < a]} \quad \text{and} \quad \mathscr{X}^{(a)} = (X_1^{(a)}, X_2^{(a)}, \cdots) \tag{5}$$

For each real a[4].

For fixed a and $\varepsilon > 0$, set

$$Y_n = X_n^{(a)} I_{[p\mathscr{X} > a+\varepsilon]} \quad \text{and} \quad \mathscr{Y} = (Y_1, Y_2, \cdots). \tag{6}$$

Then

$$P\{Y_n \neq 0\} \leqq P\{|X_n - p\mathscr{X}| > \varepsilon\} \to 0 \qquad \text{as } n \to \infty,$$

and hence, $\mathscr{Y} \in \mathscr{E}^*$ with $p\mathscr{Y} = 0$ almost surely. It follows that within the set $[p\mathscr{X} > a+\varepsilon]$,

$$\mathscr{X}^{(a)} = \mathscr{Y} \quad \text{and} \quad f(\mathscr{X}^{(a)}) = f(\mathscr{Y}) = p\mathscr{Y} = 0 \tag{7}$$

except for a subset of a null set·

Letting $\varepsilon \downarrow 0$ along a (countable) sequence, we obtain (7) for the set $[p\mathscr{X} > a]$. Similarly, (7) holds with 0 replaced by 1 for the set $[p\mathscr{X} < a]$. Therefore, for arbitrary $x = (x_1, x_2, \cdots) \in R^\infty$ and $x^{(a)}$ defined analogously to $\mathscr{X}^{(a)}$, the function

$$g(x) = \inf\{a : f(x^{(a)}) = 0\},$$

with a ranging over some *countably* dense set of the reals, is a PLIF on $\mathscr{E}$. The converse is immediate.

We conclude this section with some elementary observations about PLIF's. Let $\mathscr{F}_n$ be a non-decreasing sequence of subsets of $\mathscr{E}$ and f_n be a PLIF on $\mathscr{F}_n$ $(n \geqq 1)$. Then

$$\begin{aligned} f(x) &= \operatorname{limsup}_{n\to\infty} f_n(x) \qquad \text{if finite} \qquad (x \in R^\infty) \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{8}$$

is a PLIF on $\mathscr{F} = \bigcup_{n=1}^\infty \mathscr{F}_n$.

Let T be a measurable transformation of R^∞ into R^∞. We shall call T a *limit preserving transformation* if $T\mathscr{X} \in \mathscr{E}$ and $pT\mathscr{X} = p\mathscr{X}$ almost surely whenever $\mathscr{X} \in \mathscr{E}$. Common examples include the simple *shift transformation* $Tx = (x_2, x_3, \cdots)$ and, more generally, the *subsequence transformations* $Tx = (x_{n_1}, x_{n_2}, \cdots)$ for some increasing subsequence of positive integers $\{n_k\}$ $(x = (x_1, x_2, \cdots) \in R^\infty)$. Suppose that $\mathscr{F}$ is closed under T. That is, $T\mathscr{X} \in \mathscr{F}$ for every $\mathscr{X} \in \mathscr{F}$. Let $f_0(x)$ be a PLIF on $\mathscr{F}$. Then $f_n(x) = f_0(T^n x)$ is also a PLIF on $\mathscr{F}$ $(n \geqq 1)$ and the PLIF f, defined by (8), is such that $f(x) = f(Tx)$ $(x \in R^\infty)$.

[4] For $A \in \mathscr{A}$, $I_A = 1$ or 0 as A occurs or fails to occur.

(8) can be used to show that a PLIF exists on any countable set $\mathscr{F} \subset \mathscr{E}$. One simply orders the elements in $\mathscr{F}$ and lets $\mathscr{F}_n$ be the (finite) set composed of the first n elements of $\mathscr{F}$. (There exists a subsequence transformation T_n for which $T_n\mathscr{X}$ has coordinates that converge almost surely for every $\mathscr{X} \in \mathscr{F}_n$ ($n \geqq 1$). Thus, it is easy to define a PLIF on each $\mathscr{F}_n$.)

We shall say that a set $\mathscr{F} \subset \mathscr{E}$ is *closed under subsequencing* if for every subsequence transformation T, $T\mathscr{X} \in \mathscr{F}$ for every $\mathscr{X} \in \mathscr{F}$. Many times a PLIF f on a subset $\mathscr{F}^* \subset \mathscr{E}$ naturally extends to become a PLIF on the smallest set $\mathscr{F}$ containing $\mathscr{F}^*$ which is closed under subsequencing. Typical examples of such sets $\mathscr{F}$ are the set of $\mathscr{X} \in \mathscr{E}$ whose coordinates converge almost surely, and the set of $\mathscr{X} \in \mathscr{E}^*$ whose coordinates are independent. In the latter case, the function $f(x) = \limsup_{n\to\infty} n^{-1}\sum_{j=1}^{n} x_j$ is a PLIF on $\mathscr{F}$ because of (3) and the strong law of large numbers.

One might conceivably be able to use (8) (to find upper bounds of linearly ordered sets) in an effort to demonstrate the existence of a PLIF on $\mathscr{E}^*$ (by using Zorn's lemma [4] page 39). One can show that any maximal element in the class $\mathscr{D} = \{\mathscr{F} \subset \mathscr{E}^*: \mathscr{F}$ has a PLIF and $\mathscr{F}$ is closed under subsequencing$\}$ must be $\mathscr{E}^*$. ($\mathscr{D}$ is partially ordered by set inclusion.) The only difficulty the author has found in developing such an argument has been an inability to find upper bounds for *uncountable* linearly ordered subsets of $\mathscr{D}$.

It seems reasonable to ask whether one should require a PLIF to be a *measurable* mapping of R^∞ into R. This does not seem necessary and adding such a restriction might prevent their existence.

REFERENCES

[1] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[2] BREIMAN, L., LECAM, L. and SCHWARTZ, L. (1964). Consistent extimates and zero-one sets. *Ann. Math. Statist.* **35** 157–161.

[3] SKIBINSKI, MORRIS (1969). Some known results concerning zero-one sets. *Ann. Inst. Statist. Math.* **21** 541–545.

[4] TAYLOR, ANGUS E. (1964). *Introduction to Functional Analysis.* Wiley, New York.