A NECESSARY AND SUFFICIENT CONDITION FOR THE EXISTENCE OF CONSISTENT ESTIMATES

By LUCIEN LECAM[1] AND LORRAINE SCHWARTZ[2]

University of California, Berkeley

1. Introduction. Let $\mathfrak{X}$ be an arbitrary set and let $\mathfrak{A}$ be a σ-field of subsets of $\mathfrak{X}$. Let $\mathcal{P}$ be the family of all probability measures on $\mathfrak{A}$. Let Θ be a topological space which is homeomorphic to a subset of the cube $K = J^{N_0}$, the product of a countable family of copies of the interval $J = [0, 1]$.

Let $\mathfrak{D}$ be a subset of $\mathcal{P}$ and let $\varphi: P \rightarrow \varphi(P)$ be a function defined on $\mathfrak{D}$ and taking its values in Θ.

Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of independent identically distributed variables taking their values in $\mathfrak{X}$ and distributed according to some $P \varepsilon \mathfrak{D}$. Our purpose is to give a necessary and sufficient condition for the existence of consistent estimates of the function $\varphi(P)$.

More precisely, the problem can be described as follows. For each integer n let $\mathfrak{X}^n$ be the product of n copies of $\mathfrak{X}$, let $\mathfrak{A}^n$ be the σ-field product of n copies of $\mathfrak{A}$ and let P^n be the measure defined on $\mathfrak{A}^n$ by the product of n copies of P.

Let $\mathfrak{F}$ be an arbitrary family of subsets of $\mathfrak{D}$. If θ and θ' are elements of the cube K let θ_i and θ_1' be their ith coordinates in K and let $\rho(\theta, \theta')$ be the distance

$$(1) \qquad \rho(\theta, \theta') = \Sigma \frac{1}{2^i} \mid \theta_i - \theta_i' \mid .$$

By assumption the distance ρ defines on $\Theta \subset K$ its original topology.

Let $\mathcal{B}$ denote the σ-field of Borel subsets of Θ (or K). We shall say that φ is $\mathfrak{F}$-consistently estimable if there is a sequence $\{T_n\}$ with the following properties:

(1) The function T_n is a measurable map from $\{\mathfrak{X}^n, \mathfrak{A}^n\}$ to $\{\Theta, \mathcal{B}\}$.

(2) For every $\epsilon > 0$ and $P \varepsilon \mathfrak{D}$ let $V(P \epsilon)$ be the sphere set of elements of Θ whose distance to $\varphi(P)$ is not larger than ϵ. Then for every $\epsilon > 0$ and every $F \varepsilon \mathfrak{F}$ the quantity

$$\sup_{P \varepsilon F} P^n[T_n \not\varepsilon V(P, \epsilon)]$$

tends to zero as n tends to infinity.

The explicit purpose of the present paper is to give a characterization of the functions φ which are $\mathfrak{F}$-consistently estimable.

The terminology and results of a topological nature used in this paper can be found in either [1] or [2]. The concept of a precompact uniform structure, neces-

---

Received March 23, 1959; revised August 3, 1959.

[1] Fellow of the A. P. Sloan Foundation.

[2] This paper was prepared with the partial support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171, Task Order 3.

sary to the main result of the paper, corresponds to the notion of proximity introduced by Efremovicz [3] and may be replaced by it (see also I. S. Gál [4]). For a comparison of this to the more usual topologies and distances used the reader is referred to section 3.

2. A characterization of $\mathfrak{F}$-consistently estimable functions. On the space $\mathcal{P}$ of probability measures on $\mathcal{A}$ define a uniform structure $\mathcal{U}_n$ by the vicinities of the diagonal of $\mathcal{P} \times \mathcal{P}$ which are of the form

$$W = W\{f_1, f_2, \cdots, f_k\}$$

$$(2) \qquad = \left\{ (P, Q); \left| \int f_j\, dP - \int f_j\, dQ \right| < 1; \quad j = 1, 2, \cdots, k \right\},$$

where the f_j's are $\mathcal{A}^n$-measurable bounded numerical functions defined on $\mathfrak{X}^n$.

Let $\mathcal{U}$ be the uniform structure obtained by taking all vicinities of the preceding type, for all values of n.

THEOREM 1: *The function φ is $\mathfrak{F}$-consistently estimable if and only if there is a sequence $\{\varphi_k\}$ of functions from $\mathfrak{D}$ to K such that*

(a) *Each φ_k is uniformly continuous for the structure $\mathcal{U}$ on $\mathfrak{D}$ and the structure defined by ρ on K.*

(b) *The sequence $\{\varphi_k\}$ converges to φ uniformly on the elements of $\mathfrak{F}$.*

PROOF: Suppose that $\{T_n\}$ is a consistent sequence of estimates of φ converging uniformly on the subsets of $\mathfrak{F}$. For each n let $\varphi_n(P) = E[T_n \mid P]$ be the point of K whose coordinates are the expectations for P^n of the corresponding coordinates of T_n. Clearly φ_n is $\{\mathcal{U}_n, \rho\}$ uniformly continuous on $\mathcal{P}$. In addition, the coordinates of T_n converge in probability to those of $\varphi(P)$ so that $\varphi_n(P)$ converges to $\varphi(P)$. It is also clear that the convergence is uniform on the elements of $\mathfrak{F}$. Conversely, let $\{\beta_k\}$ be a sequence of $\{\mathcal{U}, \rho\}$ uniformly continuous functions from $\mathfrak{D}$ to K such that $\beta_k(P)$ converges to $\varphi(P)$ for each $P \in \mathfrak{D}$.

For each integer m, one can find integers $N(m)$ and $k(m)$ and functions $f_{m,j}$; $j = 1, 2, \cdots, k(m)$, which are $\mathcal{A}^{N(m)}$-measurable and bounded and such that $P \in \mathfrak{D}$ and $Q \in \mathfrak{D}$ and

$$(3) \qquad \sup_j \left| \int f_{m,j}\, dP^{N(m)} - \int f_{m,j}\, dQ^{N(m)} \right| \leqq 1$$

implies

$$(4) \qquad \rho[\beta_r(P), \beta_r(Q)] < 1/m$$

for every $r \leqq m$. For a pair (m, j) let

$$(5) \qquad \| f_{m,j} \| = \sup_x \{ | f_{m,j}(x) |; x \in \mathfrak{X}^{N(m)} \}.$$

Let $\{Z_s; s = 1, 2, \cdots, S(m)\}$ be a sequence of independent variables taking their values in $\{\mathfrak{X}^{N(m)}, \mathcal{A}^{N(m)}\}$. Chebyshev's inequality implies that

$$\sum_{j=1}^{k(m)} \text{Prob}\left\{ \left| \frac{1}{S(m)} \sum_{s=1}^{S(m)} [f_{m,j}(Z_s) - Ef_{m,j}(Z_s)] \right| \geqq \tfrac{1}{4} \right\} \leqq \frac{16}{S(m)} \sum_{j=1}^{k(m)} \| f_{m,j} \|^2. \tag{6}$$

Therefore, there exists an integer $S(m)$ such that the left-hand side of the foregoing inequality is inferior to m^{-1} whatever may be the distribution of Z_s.

Without loss of generality one can assume $N(m + 1) \geqq N(m)$ and

$$S(m + 1) \geqq 1 + S(m).$$

Let then $\nu(m) = N(m)S(m)$ and let $m(n)$ be the integer m for which $\nu(m) \leqq n < \nu(m + 1)$. Let Z_s be the $N(m)$-tuple defined by

$$Z_s = \{X_{(s-1)N(m)+1}, X_{(s-1)N(m)+2}, \cdots, X_{sN(m)}\}. \tag{7}$$

Note that for $\mathfrak{U}$ the space $\mathcal{P}$ is precompact (= totally bounded). Hence it is possible to find a finite subset $\mathfrak{D}_m = \{P_{m,l}\,; l = 1, 2, \cdots, L(m)\}$ of $\mathfrak{D}$ such that if $P \in \mathfrak{D}$ there is a $P_{m,l} \in \mathfrak{D}_m$ for which

$$\sup_{j=1,2,\cdots,k(m)} | E[f_{m,j} \mid P_{m,l}] - E[f_{m,j} \mid P] | < \tfrac{1}{4}. \tag{8}$$

Consider the quantity

$$\gamma(P_{m,l}) = \sup_{j=1,2,\cdots,k(m)} \left| \frac{1}{S(m)} \sum_{s=1}^{S(m)} f_{m,j}(Z_s) - E[f_{m,j} \mid P_{m,l}] \right| \tag{9}$$

and let $\hat{P}_n$ be the first element $P_{m,l}$ of $\mathfrak{D}_m$ which is such that

$$\gamma(\hat{P}_n) = \min_l \gamma(P_{m,l}).$$

In this fashion to each point n of $\mathfrak{X}^{\nu(m)}$ one has associated an element $\hat{P}_n(x)$ of $\mathfrak{D}_m$. The function so defined takes only a finite number of values and the sets of constancy of $\hat{P}_n$ are $\mathfrak{A}^{\nu(m)}$-measurable.

In addition, if P is the distribution from which the sequence $\{X_i\,; i = 1, 2, \cdots, \nu(m)\}$ is obtained, then

$$\sup_j \left| \frac{1}{S(m)} \sum_{s=1}^{S(m)} f_{m,j}(Z_s) - E[f_{m,j}(Z_s) \mid P] \right| < \tfrac{1}{4} \tag{10}$$

except for cases of probability inferior to m^{-1}.

There is a $P_{m,l_0} \in \mathfrak{D}_m$ such that

$$\sup_j | E(f_{m,j} \mid P) - E(f_{m,j} \mid P_{m,l_0}) | < \tfrac{1}{4}. \tag{11}$$

Consequently,

$$\sup_j \left| \frac{1}{S(m)} \sum_{s=1}^{S(m)} f_{m,j}(Z_s) - E(f_{m,j} \mid \hat{P}_n) \right| < \tfrac{1}{2} \tag{12}$$

and finally

$$\sup_j | E[f_{m,j} | \hat{P}_n] - E[f_{m,j} | P] | < \tfrac{3}{4} \tag{13}$$

except for cases having a probability inferior to m^{-1}.

By definition of the functions $f_{m,j}$ this inequality implies

$$P^n \left\{ \rho[\beta_{m(n)}(\hat{P}_n), \beta_{m(n)}(P)] > \frac{1}{m} \right\} < \frac{1}{m} \tag{14}$$

for every $P \, \varepsilon \, \mathfrak{D}$.

Take $T_n' = \beta_{m(n)}(\hat{P}_n)$ and let T_n be any point of Θ such that

$$\rho[T_n, T_n'] \leqq \inf \{\rho(T_n', \theta); \theta \epsilon \Theta \} + \frac{1}{n}. \tag{15}$$

Since T_n' takes only a finite number of values, the function T_n will also be $\mathfrak{A}^n$-measurable provided that to any given value of T_n' one always associates the same value of T_n.

By construction we have:

$$P^n \left\{ \rho[T_n', \varphi(P)] > \frac{1}{m(n)} + \rho[\beta_{m(n)}(P), \varphi(P)] \right\} < \frac{1}{m} \tag{16}$$

for every $P \, \varepsilon \, \mathfrak{D}$. Therefore,

$$\rho(T_n', T_n) \leqq \frac{1}{n} + \frac{1}{m(n)} + \rho[\beta_{m(n)}(P), \varphi(P)] \tag{17}$$

except for the case of probability inferior to m^{-1} where the inequality holds within the brackets of the preceding expression. Finally

$$P \left\{ \rho[T_n, \varphi(P)] > \frac{1}{n} + \frac{2}{m(n)} + 2\rho[\beta_{m(n)}(P), \varphi(P)] \right\} < \frac{1}{m} \tag{18}$$

for every $P \, \varepsilon \, \mathfrak{D}$.

Since $\rho[\beta_{m(n)}(P), \varphi(P)]$ converges to zero uniformly on the sets $F \, \varepsilon \, \mathfrak{F}$ this completes the proof of the theorem.

Remark 1. Suppose that $\mathfrak{D}$ is the union of an increasing sequence $\{\Delta_\nu\}$ of subsets such that

(1) Each element of $\mathfrak{F}$ is contained in a set Δ_ν.

(2) There is a sequence of functions φ_k such that φ_k is defined and uniformly continuous on $\Delta_{\nu(k)}$, and φ_k converges to φ uniformly on the elements of $\mathfrak{F}$.

(3) $\nu(k) \to \infty$ as $k \to \infty$.

Then the function φ is $\mathfrak{F}$-consistently estimable. To prove it, note that $\mathfrak{U}$ is precompact. Hence $\mathfrak{D}$ can be completed to a compact space $\hat{\mathfrak{D}}$. If φ_k is defined and uniformly continuous on $\Delta_{\nu(k)}$ then φ_k can be extended by continuity to the closure $\hat{\Delta}_{\nu(k)}$ of $\Delta_{\nu(k)}$ in $\hat{\mathfrak{D}}$. However, since $\hat{\mathfrak{D}}$ is compact, hence normal, one can then extend φ_k to a function $\hat{\varphi}_k$ which is defined and continuous on $\hat{\mathfrak{D}}$ hence *a fortiori* on the whole of $\mathfrak{D}$.

It is clear that $\{\hat{\varphi}_k\}$ converges to φ uniformly on the elements of $\mathfrak{F}$.

REMARK 2. The structure $\mathfrak{U}$ enters in Theorem 1 only by the space of uniformly continuous bounded numerical functions it determines on $\mathfrak{D}$. Any other structure giving rise to the same space of uniformly continuous functions could be substituted for $\mathfrak{U}$.

3. Relation between various types of continuity. The preceding theorem involves the uniform continuity of functions φ_n with respect to a uniform structure $\mathfrak{U}$ which is not very easily accessible. For this reason some remarks on the structure $\mathfrak{U}$ and its relation to other structures are in order.

One can define on $\mathcal{P}$ a norm, called the L_1-norm by the expression $\| P - Q \| = \sup | \int f\, dP - \int f\, dQ |$ the supremum being taken over the set of $\mathcal{A}$-measurable functions f which are bounded by (-1) and $(+1)$. If $\lambda = P + Q$ this can also be written

$$\| P - Q \| = \int \left| \frac{dP}{d\lambda} - \frac{dQ}{d\lambda} \right| d\lambda.$$

It is easily seen that the structure $\mathfrak{U}(\mathfrak{N})$ defined by this norm is finer than $\mathfrak{U}$. This gives the following corollary.

COROLLARY 1. *For the function φ to be $\mathfrak{F}$-consistently estimable it is necessary that there be a sequence $\{\varphi_k\}$ of functions from $\mathfrak{D}$ to K with the following properties:*

(1) *$\{\varphi_k\}$ converges to φ uniformly on the elements of $\mathfrak{F}$.*

(2) *Each φ_k is uniformly continuous for the structure $\mathfrak{U}(\mathfrak{N})$ on $\mathfrak{D}$ and the structure defined by ρ on K.*

For the $\mathfrak{F}$-consistent estimability of φ it is sufficient that the φ_k's be uniformly continuous with respect to one of the structures $\mathfrak{U}_n$.

The above corollary may be used to show that, for certain hypotheses, consistent tests do not exist. For instance, let $\mathfrak{D}$ be the family of distributions having densities with respect to the Lebesgue measure on the real line. There do not exist consistent tests of the hypothesis that the expectation of the distribution is finite. The function to be estimated is the indicator of the set representing the hypothesis tested in $\mathfrak{D}$. It can easily be seen that this function is not a pointwise limit of a sequence of functions which are continuous for the norm $\mathfrak{N}$.

When the space $\mathfrak{X}$ is the real line (or a Euclidean space) it is customary to define distances, and consequently uniform structures on $\mathcal{P}$ by taking either

$$\delta(P, Q) = \sup_x | P[X \leqq x] - Q[X \leqq x] | \tag{19}$$

or

$$\lambda(P, Q) = \inf_\alpha \{ -\alpha + P(X \leqq x - \alpha) \leqq Q(X \leqq x) \leqq P(X \leqq x + \alpha) + \alpha \}. \tag{20}$$

The distance δ is referred to as the Kolmogorov-Smirnov distance and λ as the Paul Lévy distance.

Denote by $\mathcal{U}(\delta)$ and $\mathcal{U}(\lambda)$ the corresponding uniform structures. Further, let $\mathfrak{I}$ be the topology associated with $\mathcal{U}$ and let $\mathfrak{I}(\mathfrak{N})$, $\mathfrak{I}_n$, $T(\delta)$, $\mathfrak{I}(\lambda)$ be the topologies associated with the other structures just defined. Finally, let Δ(resp. $\Delta(\mathfrak{N})$, Δ_n, $\Delta(\delta)$, $\Delta(\lambda)$) be the spaces of bounded uniformly continuous numerical functions for the structure $\mathcal{U}$(resp. $\mathcal{U}(\mathfrak{N})$, etc.).

It is well known that the following inclusions hold and are usually strict.

$$(21)\qquad \begin{cases} \mathfrak{I}(\mathfrak{N}) \supset \mathfrak{I} \supset \mathfrak{I}_n \supset \mathfrak{I}(\delta) \supset \mathfrak{I}(\lambda) \\ \mathcal{U}(\mathfrak{N}) \supset \mathcal{U} \supset \mathcal{U}_n \\ \mathcal{U}(\mathfrak{N}) \supset \mathcal{U}(\delta) \supset \mathcal{U}(\lambda) \end{cases}$$

The structures $\mathcal{U}_n$ and $\mathcal{U}(\delta)$ are not comparable. Similarly the structures $\mathcal{U}_n$ and $\mathcal{U}(\lambda)$ are not comparable. However, it will be shown further on that

$$(22)\qquad \Delta(\mathfrak{N}) \supset \Delta \supset \Delta(\delta).$$

It does not seem to be generally true that $\Delta_n \supset \Delta(\delta)$ or that $\Delta_n \supset \Delta(\lambda)$. Let us show this for Δ_1 and $\Delta(\delta)$, for the sake of completeness.

Consider the family $\mathfrak{D}$ of all distributions P on the real line R which are such that there is some point x for which $P\{x\} \geqq 2/3$. Let $\varphi(P)$ be defined by $\varphi(P) = \sup_x P\{x\}$. Then φ is uniformly continuous on $\mathfrak{D}$ for the structure $\mathcal{U}(\delta)$. Let $\{f_j, j = 1, 2, \cdots, m\}$ be a finite family of bounded $\mathfrak{A}$-measurable numerical functions. For every $x \in R$ let $F(x)$ be the point $F(x) = \{f_1(x), f_2(x), \cdots, f_m(x)\}$ in the m-dimensional Euclidean space $\mathcal{E}_m$. For every $\epsilon > 0$ there exist two points x and y of R such that

$$(23)\qquad | F(x) - F(y) | = \max_j | f_j(x) - f_j(y) | < \epsilon.$$

Indeed $F(R)$ is either finite or an infinite set having at least one accumulation point. Let P be a measure giving mass 5/6 to x and let Q be the measure obtained from P by removing a mass 1/6 at x and placing it at y. Then

$$(24)\qquad | \varphi(P) - \varphi(Q) | \geqq \tfrac{1}{6} \quad\text{and}\quad \left| \int f_j\, dP - \int f_j\, dQ \right| \leqq \tfrac{1}{6}\, \epsilon.$$

To show that the conditions of uniform continuity with respect to $\mathcal{U}$ cannot usually be replaced by mere continuity with respect to the topology generated by $\mathcal{U}$, consider the following example.

Let $\mathfrak{X}$ be the interval [0, 1] and let $\mathfrak{A}$ be the σ-field of Borel sets of $\mathfrak{X}$. Let $\mathfrak{D}$ be the class of probability measures $\mathfrak{D} = \{\delta_x ; x \in \mathfrak{X}\}$ where δ_x is the measure giving mass unity to the point x. The uniform structures $\mathcal{U}$ and $\mathcal{U}_1$ coincide on $\mathfrak{D}$. Further, identifying $\mathfrak{X}$ and $\mathfrak{D}$ one can easily verify that the $\mathcal{U}$ or $\mathcal{U}_1$ uniformly continuous bounded numerical functions on $\mathfrak{D}$ are precisely the $\mathfrak{A}$-measurable bounded numerical functions on $\mathfrak{X}$. In particular, the pointwise limit of a sequence of uniformly continuous functions is uniformly continuous if it is bounded. However, the topology associated to $\mathcal{U}$ is discrete, so that *every* numerical function on $\mathfrak{X}$ is continuous.

In other words, one should expect that there will exist continuous functions which are not pointwise limits of sequences of uniformly continuous functions. It seems also plausible that, in general, there will be functions which are $\mathfrak{U}$-uniformly continuous but not limits of sequences of $\mathfrak{U}_1$-uniformly continuous functions.

However, it is easily seen that if all the elements of $\mathfrak{D}$ are absolutely continuous with respect to a given probability measure μ, then every $\mathfrak{U}$-uniformly continuous function on $\mathfrak{D}$ is the pointwise limit of a sequence of $\mathfrak{U}_1$-uniformly continuous functions.

Since $\mathfrak{U}_1$ is much more manageable than $\mathfrak{U}$ the following theorem is of interest.

PROPOSITION 1. *Let Δ be a subset of $\mathfrak{D}$ which is relatively compact in $\mathcal{P}$ for the topology induced by $\mathfrak{U}_1$. Then on Δ the structures $\mathfrak{U}$ and $\mathfrak{U}_1$ coincide.*

PROOF. The proof depends on a well-known theorem of Dunford and Pettis (see [5] and also [6]) which states that Δ is relatively compact in $\mathcal{P}$ if and only if one of the following two equivalent conditions is satisfied:

(1) There is a finite measure μ such that for every $\epsilon > 0$ there is a $\delta > 0$ for which $\mu(A) < \delta$ implies $P(A) < \epsilon$ for every $P \varepsilon \Delta$.

(2) Every sequence $\{P_k ; k = 1, 2, \cdots\}$ of elements of Δ contains a subsequence which converges to an element of $\mathcal{P}$.

Let $\bar{\Delta}$ be the closure of Δ in $\mathcal{P}$ for the structure $\mathfrak{U}_1$. If $\{P_k\}$ is a sequence of elements of $\mathcal{P}$ which converges to P_0 for $\mathfrak{U}_1$ then P_k^n converges to P_0^n for the structure $\mathfrak{U}_n$ because of the equicontinuity described in Condition (1) above. Hence $\bar{\Delta}$ is also compact for $\mathfrak{U}_n$ and therefore for $\mathfrak{U}$ itself. But $\mathfrak{U}$ being compact and finer than $\mathfrak{U}_1$ and $\mathfrak{U}_1$ being separated, $\mathfrak{U}$ and $\mathfrak{U}_1$ must coincide on the set $\bar{\Delta}$ hence on Δ.

From this result we can deduce the following.

PROPOSITION 2. *Let $\{\Delta_k ; k = 1, 2, \cdots\}$ be a sequence of subsets of $\mathfrak{D}$ such that*

$$\mathfrak{D} = \bigcup_k \Delta_k .$$

Assume that each element of $\mathfrak{F}$ is contained in a finite union of sets Δ_k.

If each one of the sets Δ_k is relatively compact for $\mathfrak{U}_1$ in $\mathcal{P}$ then uniform continuity with respect to $\mathfrak{U}$ can be replaced by uniform continuity with respect to $\mathfrak{U}_1$ in the statement of Theorem 1.

If each one of the sets Δ_k is compact for $\mathfrak{U}_1$ then uniform continuity with respect to $\mathfrak{U}$ can be replaced by continuity with respect to $\mathfrak{U}_1$ in the statement of Theorem 1.

PROOF. First one can assume that $\Delta_k \subset \Delta_{k+1}$. To prove the second statement let H be the space of functions from $\mathfrak{D}$ to K which are $\mathfrak{U}$-uniformly continuous on $\mathfrak{D}$. If β_k is a continuous function from $\mathfrak{D}$ to K there is an element α_k of H such that $\rho[\alpha_k(P), \beta_k(P)] < k^{-1}$ for every $P \varepsilon \Delta_k$. This is easily seen by application of the Stone-Weierstrass theorem ([1], chap. 10, p. 55 or [7], p. 9) to the coordinates of β_k in K.

Consequently, if $\{\beta_k\}$ converges to φ uniformly on the element F of $\mathfrak{F}$ and if $F \subset \Delta_k$ then α_k converges to φ uniformly on the elements of $\mathfrak{F}$.

The first statement is a consequence of Proposition 1 and of the remark made after the proof of Theorem 1.

To show that the families $\mathfrak{D}$ which satisfy the conditions of Proposition 2 are not exceedingly rare let us mention the following. If Δ is a set of probability measures which are all absolutely continuous with respect to a given finite measure μ then Δ is relatively compact if all the densities $dP/d\mu$ are bounded by the same number M or more generally if they are bounded by a given μ integrable function. Hence, if $\mathfrak{D}$ consists of probabilities whose densities with respect to a finite measure μ are bounded (not uniformly) then $\mathfrak{D}$ is a union $\mathfrak{D} = \bigcup \Delta_k$ of $\mathfrak{U}_1$ relatively compact sets.

Another example is the following. Suppose that ξ is a parameter taking its values in a subset S of a Euclidean space $\mathcal{E}$. Assume that S is the intersection of an open set of $\mathcal{E}$ with a closed set of $\mathcal{E}$. To each $\xi \varepsilon S$ make correspond a probability measure P_ξ on the real line. Assume that:

(1) If $\xi_n \rightarrow \xi_0$ then the distribution functions of the P_{ξ_n} converge to the distribution function of P_{ξ_0} at all points of continuity of the latter.

(2) For each $\xi \varepsilon S$ there is a neighborhood $V(\xi)$ of ξ in S a finite measure μ_ξ and a μ_ξ-integrable function f_ξ such that $dP_{\xi'}/d\mu_\xi \leqq f_\xi$ for every $\xi' \varepsilon V(\xi)$.

Let C be a compact subset of S and let $\Delta(C) = \{P_\xi : \xi \varepsilon C\}$ then $\Delta(C)$ is compact in $\mathcal{P}$ for $\mathfrak{U}_1$. Since S is a union of a sequence of compact sets C_k the set $\mathfrak{D} = \{P_\xi ; \xi \varepsilon S\}$ is a union of a sequence of compact sets.

As an example of a different phenomenon, suppose that all the P_ξ defined above, instead of satisfying (1) and (2) satisfy

(3) If $\xi_n \rightarrow \xi_0$ and if μ is a σ-finite measure with respect to which all the $\{P_{\xi_n} ; n = 0, 1, 2, \cdots\}$ are absolutely continuous, then $dP_{\xi_n}/d\mu$ tends to $dP_{\xi_0}/d\mu$ in μ measure. Under such a stringent restriction, it follows from Scheffé's theorem [8] that if C is compact then $\Delta(C) = \{P_\xi ; \xi \varepsilon C\}$ is compact in the sense of the L_1-norm.

From these considerations one can deduce the following result:

Proposition 3. *Let Θ be a subset of a Euclidean space $\mathcal{E}$. Assume that to each $\theta \varepsilon \Theta$ there corresponds a probability measure P_θ on the real line, in such a way that $P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$. Furthermore, assume that the following conditions hold:*

(1) If θ_n converges to θ_0 then the distribution functions of the P_{θ_n} converge to the distribution of P_{θ_0} at all points of continuity of the latter.

(2) For each $\theta \varepsilon \Theta$ there is a neighborhood $V(\theta)$ of θ and a finite measure μ_θ such that for every $\epsilon > 0$, there is a $\delta > 0$ for which the inequality $\mu_\theta(A) < \delta$ implies $P_\xi(A) < \epsilon$ for $\xi \varepsilon V(\theta)$.

(3) Θ is the intersection of an open set of $\mathcal{E}$ with a closed subset of $\mathcal{E}$.

(4) Each element of $\mathfrak{F}$ is contained in a compact subset of Θ.

Let $\theta \rightarrow \varphi(\theta)$ be a numerical function defined on Θ.

In order that there exist an $\mathfrak{F}$-consistent sequence of estimates of φ it is necessary and sufficient that φ be the limit $\mathfrak{F}$-uniformly of a sequence of continuous functions of θ.

In particular, if $\mathfrak{F}$ is the family of all points of Θ, there exists a consistent estimate of φ if and only if φ is of the first Baire class on Θ.

Proof. Since the correspondence $\theta \leftrightarrow P_\theta$ is one to one the function $\varphi(\theta)$ can also be considered as a function defined on $\mathfrak{D}$. Let $\psi(P) = \varphi[\theta(P)]$. If $\{\psi_k\}$ is a sequence of continuous functions defined on $\mathfrak{D}$ and converging uniformly to $\psi(P)$ on the images of the elements of $\mathfrak{F}$ then $\{\varphi_k\}$ defined by $\varphi_k(\theta) = \psi_k(P_\theta)$ converges $\mathfrak{F}$-uniformly to φ. Hence the necessity of the condition.

To prove the sufficiency, let $\{C_n\}$ be a sequence of compact subsets of Θ such that $C_n \subset C_{n+1}$ and such that every compact subset of Θ be contained in a C_n. Let $\Delta_n = \{P_\theta ; \theta \in C_n\}$ be the image of C_n in $\mathfrak{D}$.

Since the function $\theta \rightarrow P_\theta$ is continuous and one to one, the inverse function $P \rightarrow P_\theta$ is continuous on each one of the compacts Δ_n.

Let $\{\varphi_k\}$ be a sequence of continuous functions of θ converging to φ uniformly on the elements of $\mathfrak{F}$. Define ψ_k by $\psi_k(\theta) = \varphi_k[\theta(P)]$. Then ψ_k converges to ψ uniformly on the images of the elements of $\mathfrak{F}$ and ψ_k is continuous, hence uniformly continuous on each Δ_n. The result follows by the remark made at the end of the proof of Theorem 1.

As an example of application consider the case where $\mathfrak{F}$ is either the family of compact subsets of Θ or more generally any family of compact sets such that each $\theta \varepsilon \Theta$ is interior to an element F_θ of $\mathfrak{F}$.

Then a function φ is $\mathfrak{F}$-consistently estimable if and only if it is continuous on Θ.

Another result obtainable directly from Proposition 1 is the following. Let $\mathfrak{D}$ be relatively compact in $\mathcal{P}$ for $\mathfrak{U}_1$. Let A and B be two disjoint subsets of $\mathfrak{D}$.

A sequence of tests of the hypothesis A against the alternative B is a sequence of measurable functions T_n from $\{\mathfrak{X}^n, \mathfrak{A}^n\}$ to the interval [0, 1]. The sequence is called uniformly consistent if $\varphi_n(P) = E[T_n \mid P]$ converges to zero on A and to one on B, the convergence being uniform in P. It is clear that the existence of a uniformly consistent sequence of tests is equivalent to the existence of a uniformly consistent estimate of the function φ equal to zero on A and to one on B. Therefore, such a sequence T_n will exist if and only if the indicator of A in $\mathfrak{D}$ is $\mathfrak{U}_1$-uniformly continuous, that is, if there is a finite family $\{f_j ; j = 1, 2, \cdots, m\}$ of $\mathfrak{A}$-measurable bounded functions on $\mathfrak{X}$ such that

$$\sup \left| \int f_j \, dP - \int f_j \, dQ \right| < 1$$

implies that either both P and Q are elements of A or both are elements of B.

Another type of restriction on $\mathfrak{D}$ under which the structure $\mathfrak{U}$ can be replaced by a somewhat more accessible structure is the restriction considered by W. Hoeffding and J. Wolfowitz in [9].

To simplify we shall present this condition only for the case of the line, although the argument is given by these authors for an arbitrary Euclidean space. Let P and Q be two probability measures on the line and let f and g be the densities of P and Q with respect to $\mu = P + Q$.

If there are intervals I_i, $i = 1, 2, \cdots, m$ such that f-g has a constant sign on each I_i, let $V^c = \bigcup_i I_i$. Let $J[P, Q; \epsilon]$ be the smallest value of m for which $\min [P(V), Q(V)] \leqq \epsilon$ if such a value exists. Otherwise let $J[P, Q; \epsilon] = \infty$.

A family $\mathfrak{D}$ satisfies the H-W condition if for every $\epsilon > 0$ the quantity

$$(25) \qquad \sup \{J[P, Q; \epsilon]; \quad P \varepsilon \mathfrak{D}, \quad Q \varepsilon \mathfrak{D}\}$$

is finite.

Hoeffding and Wolfowitz show that most of the usual parametric families of univariate distributions satisfy the H-W condition. Furthermore, these authors have shown that, δ representing the Kolmogorov-Smirnov distance, the inequality

$$(26) \qquad \| P - Q \| \leqq 4J[P, Q; \epsilon]\delta(P, Q) + 2\epsilon$$

always holds. This inequality implies that, on a set $\mathfrak{D}$ satisfying the H-W condition the distance δ and the norm $\| P - Q \|$ are equivalent. In other words, the H-W condition implies that $\mathfrak{U}(\mathfrak{N}) = \mathfrak{U}(\delta)$.

The classical result of Glivenko and Cantelli implies that, P_n denoting the empirical distribution of a sample of n independent variables having a distribution P, the distance $\delta(P_n, P)$ converges to zero in probability uniformly for $P \varepsilon \mathcal{P}$. This, in turn, by application of Theorem 1, shows that, whatever may be $\mathfrak{D}$ a function φ from $\mathfrak{D}$ to K which is $\mathfrak{U}(\delta)$ uniformly continuous on $\mathfrak{D}$ is also $\mathfrak{U}$-uniformly continuous on $\mathfrak{D}$.

Conversely, if $\mathfrak{D}$ *satisfies the H-W condition, the space of functions which are* $\mathfrak{U}(\mathfrak{N})$*-uniformly continuous coincides with the space of functions which are* $\mathfrak{U}$*-uniformly continuous and with the space of functions which are* $\mathfrak{U}(\delta)$*-uniformly continuous. Hence, under the H-W condition, the structure* $\mathfrak{U}$ *can be replaced by either* $\mathfrak{U}(\mathfrak{N})$ *or* $\mathfrak{U}(\delta)$ *in Theorem* 1.

4. Historical note. Several authors have obtained results on the existence of consistent estimates. Here are some incomplete references on the subject.

A study of the existence of consistent tests has been made for particular cases by Mrs. A. Berger in [10] and [11]. The subject was investigated further by C. Kraft in [12] without making assumptions of independence and identity of distributions. The recent paper [9] by W. Hoeffding and J. Wolfowitz contains a very deep study of the concept of distinguishability of sets of probability measures. These authors place themselves in a framework where the variables are independent and identically distributed. Their concept of finite distinguishability corresponds roughly to the concept of existence of uniformly consistent tests.

Theorems of the same nature as Theorem 1 have been obtained by J. L. Hodges, Jr. for the existence of consistent tests and by C. Stein for the existence of consistent estimates, and presented at a meeting of the Institute of Mathematical Statistics in Boston in 1952. These two authors restricted themselves to pointwise convergence. The result mentioned by C. Stein could be quoted as follows: "For a function φ on $\mathfrak{D}$ to be pointwise consistently estimable it is necessary that it be the limit of a sequence of $\mathfrak{U}(\mathfrak{N})$-uniformly continuous functions and it is sufficient that it be the limit of a sequence of $\mathcal{U}(\delta)$ uniformly continuous function." Thus, under the H-W condition C. Stein's result coincides with ours.

REFERENCES

[1] N. Bourbaki, *Elements de Mathématique—Première Partie. Les Structures Fondamentales de l'Analyse*, Livre 3, Topologie Générale, Hermann et Cie., Paris, 1940–1953.

[2] John L. Kelley, *General Topology*, Van Nostrand, New York, 1955.

[3] V. A. Efremovicz, "The geometry of proximity I," *Mat. Sbornik N. S.*, 31 (73) (1952), pp. 189–200.

[4] I. S. Gál, *Amer. Math. Soc. Notices*, Vol. 6 (1959), Abstract 554-23, p. 140.

[5] Nelson Dunford and Jacob T. Schwartz, *Linear Operators—Part I. General Theory*, Interscience Publishers, Inc., New York, N. Y., 1958.

[6] A. Grothendieck, "Sur les applications faiblement compactes d'espaces du type $C(K)$," *Canadian J. of Math.*, Vol. 5 (1953), pp. 129–173.

[7] Lynn H. Loomis, *Abstract Harmonic Analysis*, Van Nostrand, New York (1953).

[8] Henry Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.

[9] Wassily Hoeffding and J. Wolfowitz, "Distinguishability of sets of distributions," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 700–718.

[10] Agnes Berger, "On uniformly consistent tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 289–293.

[11] Agnes Berger, "On orthogonal probability measures," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 800–806.

[12] Charles Kraft, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. of Calif. Publ. in Stat.*, Univ. of California Press, Berkeley, Vol. 2 (1955), pp. 125–142.