UNIFORM CONVERGENCE OF RANDOM FUNCTIONS WITH APPLICATIONS TO STATISTICS[1]

By Herman Rubin

Stanford University[2]

0. Introduction and Summary. In many statistical problems, we obtain functions of both the random variable and the parameters involved, from whose asymptotic behavior we may deduce the asymptotic behavior of certain estimates. In many of these cases, it is sufficient to demonstrate uniform convergence with probability one of these functions. In this paper, a set of sufficient conditions for this is given, and we show how these results may be applied to some statistical problems.

1. Statement of the theorem.[3] Let $X_1, \cdots, X_n, \cdots$ be a sequence of independent and identically distributed variables with values in an arbitrary space X. Let T be a compact topological space, and let f be a complex-valued function on $T \times X$, measurable in x for each $t \varepsilon T$. Let P be the common distribution of the X_i.

Theorem 1. *If there is an integrable g such that $|f(t, x)| < g(x)$ for all $t \varepsilon T$ and $x \varepsilon X$, and if there is a sequence S_i of measurable sets such that*

$$P(X - \textstyle\bigcup_{i=1}^{\infty} S_i) = 0,$$

and for each i, $f(t, x)$ is equicontinuous in t for $x \varepsilon S_i$, then with probability one,

$$\frac{1}{n}\sum_{k=1}^{n} f(t, X_k) \to \int f(t, x)\, dP(x)$$

uniformly jor $t \varepsilon T$, and the limit function is continuous.

We may assume the S_i are monotonically increasing. Let $\epsilon > 0$ be given. Then for some i, $\int_{X-S_i} g(x)\, dP(x) < \epsilon/5$.

Since $f(t, x)$ is equicontinuous in t for $x \varepsilon S_i$ and T is compact, there exist $t_1, \cdots, t_q$ and open subsets $N_1, \cdots, N_q$ of T such that $\bigcup_{j=1}^{q} N_j = T$, $t_j \varepsilon N_j$, and for $t \varepsilon N_j$ and $x \varepsilon S_i$, $|f(t, x) - f(t_j, x)| < \epsilon/4$. Let $Y_{jk} = f(t_j, X_k)$; $Z_k = g(X_k)$, $X_k \not\varepsilon S_i$; $Z_k = 0$, $X_k \varepsilon S_i$.

By the strong law of large numbers, we may select an N such that, if $A_j = \mathcal{E}(Y_{jk})$ and $\delta > 0$,

$$P\left(\text{for some } n > N, \quad \left|\frac{1}{n}\sum_{k=1}^{n} Y_{jk} - A_j\right| \geqq \epsilon/4\right) < \delta/2q, \qquad j = 1, \cdots, q,$$

Received November 8, 1952.

[1] Research performed under an ONR contract. Presented to the Institute of Mathematical Statistics, June 15, 1951.

[2] Present address, University of Oregon.

[3] An essentially equivalent theorem was proved by Le Cam in [2].

and

$$P\left(\text{for some } n > N, \quad \left|\frac{1}{n}\sum_{k=1}^{n} Z_k\right| \geqq \epsilon/4\right) < \delta/2.$$

But if $t \varepsilon N_j$, $|f(t, X_k) - Y_{jk}| < \epsilon/4 + 2Z_k$. Hence,

$$P\left(\text{for some } n > N \text{ and some } t, t \varepsilon N_j \text{ and } \left|\frac{1}{n}\sum_{k=1}^{n} f(t, X_k) - A_j\right| \geqq \epsilon\right) < \delta.$$

Therefore, $1/n \sum_{k=1}^{n} f(t, X_k)$ converges uniformly to a continuous function with probability one. By the strong law of large numbers, that function is

$$\int f(t, x)\, dP(x).$$

2. Applications. As an application of this theorem, we see that the sample characteristic function converges to the population characteristic function uniformly with probability one in any bounded interval, since $f(t, x) = e^{itx}$ satisfies the conditions of the theorem.

It may happen that $\log L(x \mid \theta) = f(x, \theta)$ satisfies the conditions of the theorem. For example, for the multivariate normal, the Poisson, Cauchy, χ^2, double exponential, and many other distributions, we are led to the almost certain convergence of maximum likelihood estimates to the true values if the parameter is restricted to a compact set.

More difficult estimation procedures can also be shown to be consistent. For example, consider a problem of Reiersøl [4]. The model is

$$x_i = \xi_i \cos \alpha + u_i,$$

$$y_i = \xi_i \sin \alpha + v_i,$$

where u_i and v_i have a joint normal distribution, ξ_i is not normal, and ξ_i, (u_i, v_i) are independent. Let $\rho = t \sin \beta$, $\sigma = -t \cos \beta$, and let

$$\varphi(t, \beta, X_j) = e^{i\rho x_j + i\sigma y_j}.$$

Then $1/n \sum_{j=1}^{n} \varphi(t, \beta, X_j) \to \psi(t, \beta)$ uniformly with probability one for t in any finite interval. Let

$$\chi(\beta) = \int_{-\infty}^{\infty} |\psi(2t, \beta) - \psi^3(t, \beta)\psi(-t, \beta)|^2\, d\lambda(t),$$

where λ is a bounded monotone function, such that for any $\epsilon > 0$, $\lambda(\epsilon) - \lambda(-\epsilon) > 0$. Then χ is a periodic function of period π, and $\chi(\beta) = 0$ only for $\beta = \alpha + k\pi$. Let

$$\psi_n(t, \beta) = \frac{1}{n}\sum_{j=1}^{n} \varphi(t, \beta, X_j),$$

$$\chi_n(\beta) = \int_{-\infty}^{\infty} |\psi_n(2t, \beta) - \psi_n^3(t, \beta)\psi_n(-t, \beta)|^2\, d\lambda(t).$$

Since ψ_n is bounded, it follows that $\chi_n(\beta) \to \chi(\beta)$ uniformly with probability one. Hence, if b_n minimizes $\chi_n(\beta)$, it follows that $b_n \to \beta$ with probability one, in the sense of convergence mod π.

This result is stronger than that obtained by Neyman about Reiersøl's problem. The method can also be extended to Neyman's extension of the problem [3].

We can, in fact, obtain some very strong results on the existence of consistent estimates.

THEOREM 2. *Let $\mathfrak{F}$ be a family of distributions on Euclidean n-space. Let π be a continuous function mapping $\mathfrak{F}$ into the topological space $\mathcal{P}$. Then there exists a sequence p_k of functions on Euclidean kn-space to $\mathcal{P}$ such that if $X_1, \cdots, X_k, \cdots$ are independently distributed with distribution function $F \varepsilon \mathfrak{F}$ then $\lim_{k\to\infty} p_k(X_1, \cdots, X_k) = \pi(F)$ with probability one.*

In other words, any continuous parameter is consistently estimable. The converse is not true, since moments, which are not continuous parameters, are consistently estimable. It is an unsolved problem, which functions π of a distribution in a family $\mathfrak{F}$ are consistently estimable—even whether the topological structure of the family $\mathfrak{F}$ and the topological properties of the function π are sufficient to characterize consistently estimable parameters π.

Let us proceed to the proof of the theorem.

Let λ be a non-negative finite measure on Euclidean n-space such that every open set has positive measure. For each $F \varepsilon \mathfrak{F}$, let

$$\psi(t_1, \cdots, t_n, F) = \mathcal{E}\left(\exp i \sum_{j=1}^{n} t_j x_j \mid F\right),$$

i.e., $\psi(t_1, \cdots, t_n, F)$ is the characteristic function of F evaluated at $t_1, \cdots, t_n$.

Similarly, let

$$\psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) = \frac{1}{k}\sum_{h=1}^{k} \exp\left(i \sum_{j=1}^{n} t_j X_{hj}\right).$$

Then define

$$\rho_k^2(X_1, \cdots, X_k) = \inf_{F \varepsilon \mathfrak{F}} \int \mid \psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F) \mid^2 \cdot d\lambda(t_1, \cdots, t_n)$$

and let p_k be any function such that for every $X_1, \cdots, X_k$ there is an $F_k \varepsilon \mathfrak{F}$ satisfying

$$\int \mid \psi_k(t_1, \cdots t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F_k) \mid^2 d\lambda(t_1, \cdots, t_n) < \rho_k^2(X_1, \cdots, X_k) + \frac{1}{k}$$

and

$$p_k(X_1, \cdots, X_k) = \pi(F_k).$$

From Theorem 1, we see that $\psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k)$ approaches $\psi(t_1, \cdots, t_n, F)$ uniformly with probability one for $t_1, \cdots, t_n$ in any bounded set. Therefore,

$$\int | \psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F) |^2 \, d\lambda(t_1, \cdots, t_n)$$

approaches zero with probability one. Hence,

$$\int | \psi(t_1, \cdots, t_n, F_k) - \psi(t_1, \cdots, t_n, F) |^2 \, d\lambda(t_1, \cdots, t_n)$$

approaches zero with probability one, and thus $F_k \rightarrow F$ almost certainly. The result then follows from the continuity of π.

Similar results can be obtained in the case of a continuously identified parameter. If a structure S generates the distribution F, we may ask whether a function φ defined on the space $\mathcal{S}$ of structures is determined by the distribution $\mathfrak{F}$. If so, we say [1] that φ is *identified* at F.

Let us formulate the preceding definition without regard to the structure S. We obtain for each F in a class $\mathfrak{F}$ of distributions, a non-null set $\Phi(F)$ in the parameter space. The condition that Φ is identified at F then becomes that $\Phi(F)$ has one element.

Let us say that Φ is *continuously identified* at F with respect to $\mathfrak{F}$ if for every sequence $F_1, \cdots, F_n, \cdots$ of distributions of $\mathfrak{F}$ such that $F_n \rightarrow F$, and for any sequence $\theta_1, \cdots, \theta_n, \cdots$ such that $\theta_k \in \Phi(F_k)$ for all k, θ_n converges to the one element of $\Phi(F)$.

Then by a method similar to that of Theorem 2 we obtain

THEOREM 3. *Let $\mathfrak{F}$ be a family of distributions on Euclidean n-space and let Φ map $\mathfrak{F}$ into the set of non-null subsets of $\mathcal{P}$. Then there exists a sequence p_k of functions on Euclidean kn-space to $\mathcal{P}$ such that for any $F \in \mathfrak{F}$, if $X_1, \cdots, X_k, \cdots$ are independently distributed with distribution function F and Φ is continuously identifiable at F with respect to $\mathfrak{F}$, then $p_k(X_1, \cdots, X_k)$ approaches the element of $\Phi(F)$ with probability one.*

REFERENCES

[1] L. HURWICZ, "Generalization of the concept of identification," *Statistical Inference in Dynamic Economic Models*, John Wiley and Sons, New York, 1950, pp. 245–257.

[2] L. LE CAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates", *University of California Publications in Statistics*, Vol. 1 (1953), pp. 277–300.

[3] J. NEYMAN, "Existence of consistent estimates of the directional parameter in a linear structural relation between two variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 497–512.

[4] O. REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–389.