ON THE CONVERGENCE OF SAMPLE PROBABILITY DISTRIBUTIONS

By V. S. VARADARAJAN

Indian Statistical Institute, Calcutta

0. Summary

Let $\xi_1, \xi_2, \ldots,$ be a sequence of independent and identically distributed, real-valued random variables with common distribution function $F(x)$. Let $S_n(x)$ be the sample distribution function which corresponds to the distribution with masses $\frac{1}{n}$ at each of the points $\xi_1, \ldots \xi_n$. It is known that (theorem of Glivenko-Cantelli) $P(\lim_{n\to\infty} \sup_{-\infty<x<+\infty} |F(x)-S_n(x)| = 0) = 1$ (Loeve, 1955); in particular $S_n \to F$ (in the sense of P. Levy) with probability 1. In this paper we discuss the analogue of this last result in the case when the random variables take values in an arbitrary metric space. The following is the main conclusion (the pertinent definitions are given in section 1).

Let $\xi_1, \xi_2 \ldots,$ be independent, identically distributed random variables whose values lie in a metric space $\mathfrak{M}$ and μ be their common distribution on $\mathfrak{M}$. Let μ_n be the sample probability distribution, i.e., the distribution with masses $\frac{1}{n}$ at the n points $\xi_1, \xi_2 \ldots, \xi_n$. Then $P(\mu_n \Longrightarrow \mu) = 1$ if and only if the random variables are strongly measurable. (In any case the probability is 0 or 1).

1. Definitions and remarks

In what follows, $(\Omega, \mathcal{S}, P)$ is a fixed probability space. $\mathfrak{M}$ is a metric space and $\mathcal{B}$ is the class of Borel sets of $\mathfrak{M}$, i.e., the minimal σ—field generated by the open sets of $\mathfrak{M}$. By a *measure on* $\mathfrak{M}$ we mean a finite, non-negative and countably additive set function on $\mathcal{B}$. A mapping ξ of Ω into $\mathfrak{M}$ is called *random* or *measurable* if for each $A \in \mathcal{B}$, $\xi^{-1}(A) \in \mathcal{S}$. The measure μ on $\mathcal{B}$ defined by setting $\mu(A) = P\{\xi^{-1}(A)\}$ for all $A \in \mathcal{B}$ is called the measure induced by ξ and is denoted by $P\xi^{-1}$. A measurable mapping is called *simple* if its range is a finite subset of $\mathfrak{M}$. It is called *strongly measurable* if there exists a set $N \in \mathcal{S}$ and a sequence $\{\xi_n\}$ of simple mappings such that $P(N) = 0$ and $\xi_n(w) \to \xi(w)$ for each $w \in \Omega - N$. It can be shown that if $\mathfrak{M}$ is separable any measurable mapping is strongly measurable. A proof of this fact is given in section 2.

A family $\{\xi_\alpha\}_{\alpha \epsilon I}$ of random mappings is called *independent* if

$$P(\xi_{\alpha_1} \epsilon A_1, \ldots, \xi_{\alpha_n} \epsilon A_n) = \prod_{i=1}^{n} P(\xi_{\alpha_i} \epsilon A_i)$$

for each choice of the integer n, indices α_i and sets A_i. The mappings are called *identically distributed* if $P\xi_\alpha^{-1}$ is the same for all $\alpha \epsilon I$. If g is a fixed real-valued

$\mathcal{B}$-measurable function defined on X and $\{\xi_\alpha\}$ is a family of independent identically distributed random mappings, then $\{g(\xi_\alpha)\}$ is a family of independent identically distributed real-valued random variables. Also for any random mapping ξ and any real-valued $\mathcal{B}$-measurable g defined on $\mathfrak{M}$, we have $\int_{\mathfrak{M}} g dP\xi^{-1} = \int_{\Omega} g(\xi) dP$ in the sense that if either integral exists, so does the other and the two are equal (Halmos, 1950).

We need a concept of convergence for measures on $\mathfrak{M}$. If $\{\mu_n\}$ is a sequence of measures on $\mathfrak{M}$, we say that μ_n converges weakly to $\mu_0 (\mu_n \Longrightarrow \mu_0)$ if and only if $\int_{\mathfrak{M}} g d\mu_n \rightarrow \int_{\mathfrak{M}} g d\mu_0$ for each bounded continuous g defined on $\mathfrak{M}$. We require the following theorem (*) on weak convergence (Varadarajan, 1958) : Let $\mathfrak{M}$ be separable. Then there exists a fixed countable set G of bounded continuous functions on $\mathfrak{M}$ such that for any sequence $\{\mu_n\}_{n=0,1,\ldots}$ of measures on $\mathfrak{M}$, $\mu_n \Longrightarrow \mu_0$ if and only if $\int_{\mathfrak{M}} g d\mu_n \rightarrow \int_{\mathfrak{M}} g d\mu_0$ for each $g \in G$.

We also introduce the notion of support of a measure on $\mathfrak{M}$. A Borel set E is a *support* of μ if $\mu(\mathfrak{M}-E) = 0$. The measure has *separable* support if there exists some separable set E which is a support.

2. Theorems

Theorem 1 : *If $\mathfrak{M}$ is separable and ξ a measurable mapping of Ω into $\mathfrak{M}$ then ξ is strongly measurable.*

Proof : Let $\mu = P\xi^{-1}$. Since $\mathfrak{M}$ is separable, there is a covering of $\mathfrak{M}$ by a countable number of spheres of radius $\frac{1}{2n}$, n being a positive integer. Consequently for each integer n, we can write $\mathfrak{M} = \bigcup_j A_{nj}$ where $A_{nj} \in \mathcal{B}$ for each $A_{nj} \cap A_{nk} = \phi$ whenever $j \neq k$ and diameter $(A_{nj}) \leqslant \frac{1}{n}$ for all j. Let $B_n = \bigcup_{j > k_n+1} A_{nj}$ where k_n is so chosen that $\mu(B_n) \leqslant \frac{1}{n^2}$. Setting $E_{nj} = \xi^{-1}(A_{nj})$ for $1 < j \leqslant k_n$ and $F_n = \xi^{-1}(B_n)$ and choosing arbitrary points t_n in B_n and t_{nj} in A_{nj} (if any set concerned is empty, we ignore it) we define a sequence $\{\xi_n\}$ of simple mappings as follows :

$$\xi_n(w) = \begin{cases} t_{nj} & \text{if } w \in E_{nj} \\ t_n & \text{if } w \in F_n \end{cases} \qquad 1 \leqslant j \leqslant k_n.$$

Then $P(F_n) < \frac{1}{n^2}$ and for $w \in \Omega - F_n$, $d(\xi_n(w), \xi(w)) \leqslant \frac{1}{n}$. If now we define F as lim sup F_n, $P(F) = 0$ and for $w \in \Omega - F$, $\xi_n(w) \rightarrow \xi(w)$. This completes the proof.

Theorem 2 : *Let ξ be a measurable mapping of Ω into $\mathfrak{M}$. Then ξ is strongly measurable if and only if $P\xi^{-1}$ has separable support.*

Proof : If ξ is strongly measurable, there is a sequence $\{\xi_n\}$ of simple mappings and a set $N \in \mathcal{S}$ such that $P(N) = 0$ and $\xi_n(w) \rightarrow \xi(w)$ for $w \in \Omega - N$. Since the range

A_n of each ξ_n is finite, $A = \bigcup_n A_n$ is countable and hence $\bar{A}$ is a closed, separable subset of $\mathfrak{M}$. Further $w \in \Omega - N$ implies $\xi_n(w) \to \xi(w)$ and since $\xi_n(w) \in A$, $\xi(w) \in \bar{A}$. In other words, $\xi^{-1}(\bar{A}) \supset \Omega - N$. Consequently, $P\xi^{-1}(\bar{A}) = 1$ and hence $P\xi^{-1}$ has separable support.

To prove the converse, let ξ be measurable and let $P\,\xi^{-1}$ have separable support E. Let $N = \xi^{-1}(\mathfrak{M} - E)$. Then $P(N) = 0$. The restriction $\hat{\xi}$ of ξ to $\Omega - N$ can now be considered a measurable mapping of $\Omega - N$ into the separable metric space E (relative topology) and hence by Theorem 1, $\hat{\xi}$ is strongly measurable. Since $P(N) = 0$ this implies that ξ itself is strongly measurable.

We need two lemmas on weak convergence of measures. The first is known (see for instance Billingsley, 1956).

Lemma 1 : *Let* $\{\mu_n\}_{n=0,1,\ldots}$ *be a sequence of measures on* $\mathfrak{M}$. *Then* $\mu_n \Longrightarrow \mu_0$ *if and only if* $\mu_n(\mathfrak{M}) \to \mu_0(\mathfrak{M})$ *and for each closed set* C, $\lim\sup_{n\to\infty} \mu_n(C) \leqslant \mu_0(C)$.

The proof is omitted.

The next lemma deals with convergence of measures on subspaces.

Lemma 2 : *A be a Borel set of* $\mathfrak{M}$ *and* $\{\mu_n\}$ $\{n = 0, 1, \ldots$ *measures on* $\mathfrak{M}$ *such that* $\mu_n(\mathfrak{M} - A) = 0$ *for* $n = 0, 1, 2, \ldots$. *Then* $\mu_n \Longrightarrow \mu_0$ *if and only if* $\hat{\mu}_n \Longrightarrow \hat{\mu}_0$ *where* $\hat{\mu}_n (n = 0, 1, \ldots)$ *is the restriction of* μ_n *to* A *and the convergence is with respect to functions on* A *continuous in the relative topology of* A.

Proof: Since the topology of A is the relative topology, a set $K \subset A$ is closed in A if and only if $K = C \cap A$ for some C closed in $\mathfrak{M}$. Then $\hat{\mu}_n(K) = \mu_n(C)$ and the lemma follows immediately from lemma 1.

We can now state and prove the main theorem as follows.

Theorem 3 : *Let* $\xi_1, \xi_2, \ldots,$ *be independent identically distributed random mappings of* Ω *into* $\mathfrak{M}$ *and let* μ *be the common induced measure. For each* $w \in \Omega$ *let* μ_n^w *denote the measure with masses* $\frac{1}{n}$ *at each of the n-points* $\xi_1(w), \ldots \xi_n(w)$. *Then* $P(\mu_n^w \Longrightarrow \mu) = 1$ *if and only if the mappings are strongly measurable.*

In any case the probability is zero or one.

Proof: To prove the 'only if' part let $E = \{w : \mu_n^w \Longrightarrow \mu\}$. Since $P(E) = 1$, E is non-empty. Let w_0 be some point of E. If A is the closure of the set $\{y;\ y = \xi_n(w_0)$ for some $n\}$ then A is closed and separable. Further $\mu_n^{w_0}(A) = 1$ for all n and hence as $\mu_n^{w_0} \Longrightarrow \mu$, $\mu(A) \geqslant \lim\sup_{n\to\infty} \{\mu_n^{w_0}(A)\}$. Consequently it follows that $\mu(A) = 1$ and hence that μ has a separable support. It now follows from Theorem 2 that the mappings are strongly measurable.

To prove the 'if' part, we can in view of Theorem 2, choose a separable Borel set E such that $\mu(E) = 1$. Thus $P(\xi_n^{-1}(E)) = 1$ for all n. If we write $N = \bigcup_n(\Omega - \xi_n^{-1}(E))$, then $N \in S$ and $P(N) = 0$. Further $w \in \Omega - N$ implies that $\xi_n(w) \in E$ and hence $\mu_n^w(\mathfrak{M} - E) = 0$ for all n. By virtue of Lemma 2 we can therefore replace $\mathfrak{M}$ by E. In other words we can and do suppose that $\mathfrak{M}$ is separable.

For any $w \in \Omega$ and any bounded continuous g on $\mathfrak{M}$ we have

$$\int_{\mathfrak{M}} g d\mu_n^w = \frac{1}{n} \sum_{i=1}^{n} g\{\xi_i(w)\}.$$

Since $\xi_1, \xi_2, \ldots,$ are independent and identically distributed, so are $g(\xi_1), g(\xi_2), \ldots,$ and hence by Kolmogorov's strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} g\{\xi_i(w)\} \rightarrow \int_{\Omega} g(\xi) dP = \int_{\mathfrak{M}} g d\mu$$

almost surely (Loeve, 1955). In other words, there is a set $N_g \epsilon \mathcal{S}$ such that $P(N_g) = 0$ and $w \epsilon \Omega - N$ implies that $\int_{\mathfrak{M}} g d\mu_n^w \rightarrow \int_{\mathfrak{M}} g d\mu$. We now obtain such a set N_{g_r} for each integer r, where $G = \{g_1, g_2, \ldots\}$ is the countable class of bounded continuous functions whose existence is ensured in the theorem (*) stated in section 1. If $N = \bigcup_r N_{g_r}$, then $P(N) = 0$ and $w \epsilon \Omega - N$ implies $\int g_r d\mu_n^w \rightarrow \int g_r d\mu$ for each r. This however implies by the same theorem (*) that $\mu_n^w \Longrightarrow \mu$ and completes the proof of the theorem.

Finally, to see that the probability $p = P\{\mu_n^w \Longrightarrow \mu\}$ is 0 or 1, we note from the preceeding analysis that $p = 1$ whenever μ has separable support. In the contrary case however, the 'only if' part of the above theorem shows that $E = \{w \,;\, \mu_n^w \Longrightarrow \mu\}$ is empty. Consequently $p = 0$.

3. CONCLUDING REMARKS AND AN UNSOLVED PROBLEM

A natural sequel to the problem discussed above is the question whether there exists a measurable mapping which is not strongly measurable. The measure μ induced by such a mapping will not have a separable support. Conversely any such measure gives rise to a measurable mapping which is not strongly measurable, since we can then take $\Omega = \mathfrak{M}$, $\mathcal{S} = \mathcal{B}$, $P = \mu$ and $\xi(w) = w$ for all $w \epsilon \Omega$. If we now write (for such a μ) $\lambda = \sup[\{\mu(E)\}, E$ separable and $\epsilon \mathcal{B}]$, $\lambda \geqslant 0$ and we can find a sequence E_n of separable Borel sets such that $\mu(E_n) \rightarrow \lambda$. Writing $E_0 = \bigcup_n E_n$ we have E_0 a separable Borel set and $\lambda \geqslant \mu(E_0) \geqslant \mu(E_n)$ for all n. Consequently $\mu(E_0) = \lambda$. Since μ does not have a separable support, $\lambda < 1$. The restriction of μ to $\mathfrak{M} - E_0$ is now a non-trivial measure vanishing for every separable Borel subset of $\mathfrak{M} - E_0$. We are thus led to the question : Does there exist a metric space $\mathfrak{M}$ and a measure μ on it such that $\mu(\mathfrak{M}) > 0$ and $\mu(E) = 0$ whenever E is a separable Borel set ? The existence of such a measure is necessary and sufficient for the existence of a measurable mapping which is not strongly measurable.

As far as we are aware, this question does not seem to have been solved.

REFERENCES

BILLINGSLEY, P. (1956) : Invariance principle for dependent random variables. *Trans. Amer. Math. Soc.*, 83, 1956.

HALMOS, P. R. (1950) : *Measure Theory*, D. Van Nostrand Company, New York, 163.

LOEVE, M. (1955) : *Probability Theory*, D. Van Nostrand Company, New York.

VARADARAJAN, V. S. (1958) : Weak convergence of measures in separable metric spaces. *Sankhyā*, 19, 15.

Paper received : October, 1957.