SOME LIMIT THEOREMS FOR JOINT DISTRIBUTIONS

By J. SETHURAMAN

Indian Statistical Institute

SUMMARY. In this paper we discuss the convergence of a sequence of joint distributions when it is known that the associated sequences of marginal and conditional distributions converge. Illustrating by means of an example that the joint distributions need not converge weakly when the marginal and conditional distributions converge only weakly, several results are obtained by suitably strengthening the modes of convergence of the latter distributions. An illustrative application of these results is given.

1. Introduction

When Z is the product space $(X \times Y)$, a random variable ζ on (Z, U) is of the form (ξ, η) where ξ is a random variable on (X, S) and η on (Y, T). Let $\{\zeta_n\}$ be a sequence of random variables on Z and let $\lambda_n(\cdot)$, $\mu_n(\cdot)$ and $\nu_n(x, \cdot)$ denote the joint distribution of (ξ_n, η_n), the marginal distribution of ξ_n and the conditional distribution η_n given $\xi_n = x$, respectively. The theorems of the present work are concerned with the convergence of $\{\lambda_n\}$ when $\{\mu_n\}$ and $\{\nu_n\}$ are known to converge in some sense. In an earlier paper this problem has been mentioned by Sukhatme and Sethuraman (1959). Theorem 2 was referred to and used earlier in another paper by the author (Sethuraman, 1961).

2. Notations and preliminaries

Before embarking on the statement and the proofs of our theorems, we explain in this section our notations and mention some well-known results which form the basic tools of this paper.

Throughout this work we will be concerned with two measure spaces (X, S) and (Y, T), their product $(Z, U) = (X \times Y, S \times T)$ and a sequence of random variables (ξ_n, η_n) taking values in Z. The distribution of (ξ_n, η_n) on (Z, U) will be denoted by λ_n while the distribution of ξ_n on (X, S) will be denoted by μ_n. We assume that the conditional distribution of η_n given $\xi_n = x$ exists as a probability measure, i.e. there exists a function $\nu_n(x, B)$ which is a probability measure on T for each $x \epsilon X$ and is a measurable function on X for each $B \epsilon T$ and further satisfies the equation

$$\lambda_n(A \times B) = \int_A \nu_n(x, B) d\mu_n \text{ for all } A \epsilon S \text{ and } B \epsilon T.$$

In this case, if C is any set in U and C_x for each $x \epsilon X$ denotes the sub-set $\{y : y \epsilon Y, (x, y) \epsilon C\}$, then $\lambda_n(C) = \int_X \nu_n(x, C_x) d\mu_n$. (See Halmos, 1950).

If $p_1, p_2, \ldots,$ is a sequence of probability measures defined on a measurable space (M, V), we shall say p_n converges strongly to $p(p_n \to p$ in symbols) if $p_n(C) \to p(C)$ for each $C \epsilon V$. It is well known that for a given sequence p_n there exists a p such that $p_n \to p$ if and only if $\lim p_n(C)$ exists for all $C \epsilon V$; in this case the p_n's are equicontinuous, i.e. if C_k is any decreasing sequence of sets with $\bigcap_{k=1}^{\infty} C_k = \phi$, the null set, then $\sup_n p_n(C_k) \to 0$ as $k \to \infty$. (See Halmos, 1950). It is also known that $p_n \to p$ if and only if $\int g dp_n \to gdp$ for all bounded measurable functions g. (See Halmos, 1950). If the densities $f_n(m)$ of p_n with respect to some σ-finite measure p_0, converge in measure $[p_0]$ to a density function $f(m)$ then there is a p such that $p_n \to p$. (See Scheffe, 1947). The above is a sufficient condition for the strong convergence of a sequence of probability measures $\{p_n\}$ which is convenient in practice.

We shall also require the notion of weak convergence of probability measures. This requires that the basic space be topological, and that all continuous functions be measurable. If $p_1, p_2, \ldots,$ are probability measures on such a space M, we shall say p_n converges to p weakly ($p_n \Longrightarrow p$ in symbols) if $\int g dp_n \to \int g dp$ for all bounded continuous functions $g(m)$ on M. A set C is said to be a continuity set of p if $p(bd\ C) = 0$ where $bd\ C$ is the boundary of C. $p_n \Longrightarrow p$ if and only if $\lim p_n(C) = p(C)$ for each C that is a continuity set of p. (See Billingsley, 1956). Further, if M is separable complete metric, $\{p_n\}$ is compact under weak convergence if and only if, for each $\epsilon > 0$, there is a compact set $C \subset M$ with $p_n(C) \geqslant 1-\epsilon$ for all n. (See Prohorov, 1936 ; Varadarajan, 1958).

For the formulation of one of our theorems we require the notion of UC^* convergence (allied to that of Parzen (1954)) of a family of sequences of probability measures. Let $\nu_n(\theta, \cdot)$, $n = 0, 1, \ldots$ be a family of sequences of probability measures on $M = R_k$, the euclidean space of k dimensions. It is assumed that the index θ takes values in a compact metric space I. Let $\phi_n(t, \theta)$ denote the characteristic function of $\nu_n(\theta, \cdot)$, i.e.

$$\phi_n(t, \theta) = \int_M \exp(i\, \underline{t}\, \underline{m}')\nu_n(\theta, dm), \quad n = 0, 1, \ldots \qquad \ldots \quad (2.1)$$

Definition: $\nu_n(\theta, \cdot)$ is said to converge to $\nu_0(\theta, \cdot)$ in the UC^* sense relative to $\theta \epsilon I$ if

$$\text{(a)} \sup_{\theta \epsilon I} |\phi_n(t, \theta) - \phi(t, \theta)| \to 0 \text{ as } n \to \infty$$

and $\phi_0(t, \theta)$ is equicontinuous in θ at $t = 0$ ... (2.2)

and (b) $\phi_0(t, \theta)$ is a continuous function of θ for each t. ... (2.3)

We mention here that Parzen (1954) uses (2.2) alone as the definition of UC^* convergence. (2.2) and (2.3) imply that $\phi_0(t, \theta)$ is continuous in t and θ. (2.2) and (2.3) also imply the following:

$$\theta_n \to \theta_0 \quad \text{implies} \quad \phi_n(t, \theta_n) \to \phi_0(t, \theta_0) \qquad \dots \quad (2.4)$$

which means

$$\nu_n(\theta_n, \cdot) \Longrightarrow \nu_0(\theta_0, \cdot); \qquad \dots \quad (2.5)$$

$$\theta_n \to \theta_0 \text{ implies } \phi_0(t, \theta_n) \to \phi_0(t, \theta_0) \qquad \dots \quad (2.6)$$

which means

$$\nu_0(\theta_n, \cdot) \Longrightarrow \nu_0(\theta_0, \cdot). \qquad \dots \quad (2.7)$$

3. Main theorems

Before proceeding to state and prove the first of our main results we establish the following lemma.

Lemma 1: *Let $\{p_n\}$ be a sequence of probability measures on an arbitrary measure space (M, V) converging strongly to a measure p. Let $\{u_n\}$ be a sequence uniformly bounded V-measurable functions converging almost everywhere $[p]$ to a function u. Then* $\int_M u_n \, dp_n \to \int_M u dp$.

Proof: Let $|u(m)| < A$ for all n.

Define

$$D_n = \bigcup_{k=n}^{\infty} \{m : |u_k(m) - u(m)| > \epsilon\}.$$

We know that the D_n are decreasing and that if $\bigcap_{n=1}^{\infty} D_n = D$ then $\mu(D) = 0$. We have

$$\left|\int_M u_n dp_n - \int_M u dp\right| = \left|\int_M (u_n - u) dp_n + \int_M u dp_n - \int_M u dp\right|$$

$$\leq \left|\int_{D_n} (u_n - u) dp_n\right| + \left|\int_{M-D_n} (u_n - u) dp_n\right| + \left|\int_M u dp_n - \int_M u dp\right|$$

$$\leq 2A p_n(D_n) + \epsilon + \left|\int_M u dp_n - \int_M u dp\right| \dots \qquad \dots \quad (3.1)$$

The first term tends to zero since p_n are equicontinuous and the last term tends to zero since $p_n \to p$. Since $\epsilon > 0$ is arbitrary, the lemma is proved.

Suppose that, in some sense, the (marginal) distributions μ_n converge to μ and the (conditional) distributions $\nu_n(x, .)$ converge to $\nu(x, .)$. Further, let the joint distributions λ_n converge. It is plausible that λ_n then converges to the distribution λ_0 defined by μ and $\nu(\cdot, \cdot)$, i.e.

$$\lambda_0(A \times B) = \int_A \nu(x, B) d\mu \qquad \dots \quad (3.2)$$

over rectangle sets $A \times B$. This defines a distribution λ_0 uniquely on (Z, U). In what follows, by λ_0 we mean the distribution defined by the relation (3.2).

Theorem 1 : *If the sequence of (marginal) distributions $\{\mu_n\}$ converges strongly to μ and if for almost all $x[\mu]$ the sequence of (conditional) distributions $\{\nu_n(x, \cdot)\}$ converges strongly to $\nu(x, \cdot)$ then the sequence of (joint) distributions $\{\lambda_n\}$ converges strongly to λ_0.*

Proof: Let $g(x, y)$ be any U-measurable function on Z, bounded by K. We define sequence of S-measurable functions $v_n(x)$ as follows :

$$v_n(x) = \int_Y g(x, y)\ \nu_n(x, dy). \qquad \dots \quad (3.3)$$

It is plain that $|v_n(x)| \leqslant K$ and that

$$v_n(x) \to v(x) = \int_Y g(x, y)\, \nu(x, dy) \qquad \dots \quad (3.4)$$

for almost all $x[\mu]$.

On application of Lemma 1 we find that

$$\int g(x, y) d\lambda_n = \int_X d\mu_n \int_Y g(x, y)\ \nu_n(x, dy)$$

$$= \int_X v_n(x) d\mu_n \to \int_X v(x) d\mu$$

$$= \int_X d\mu \int_Y\ g(x, y) \nu(x,\ dy)$$

$$= \int g(x, y) d\lambda \qquad \dots \quad (3.5)$$

where $\lambda(C) = \int_x \nu(x,\ C_x) d\mu$ with $C_x = \{y : y \in\ Y,\ (x, y) \in C\}$. (See Halmos, 1950). This λ is the same as λ_0. Thus $\lambda_n \to \lambda_0$.

Theorem 2: *If the sequence of (marginal) distributions $\{\mu_n\}$ converges strongly to μ and if the sequence of (conditional) distributions $\{\nu_n (x, .)\}$ converges weakly to $\nu(x, .)$ for almost all $x[\mu]$, then the sequence of (joint) distributions $\{\lambda_n\}$ converges weakly to λ_0.*

Proof: The proof of this theorem is on the same lines as Theorem 1 and so is omitted.

In Theorems 1 and 2 we assumed that the marginal distributions converge strongly. We now ask ourselves what happens if the marginal distributions converge only weakly. Naturally, one expects that one shall have to strengthen the mode of convergence of the conditional distributions. We have given an example in section 4 to illustrate this fact, that, in general, such strengthening would be necessary. The difficulty in this situation is that the conditional distribution at the n-th stage is defined almost everywhere with respect to μ_n and the $[\mu_n]$ null x-sets, the sets of misbehaviour of $\nu_n (x, \cdot)$, vary with n. Thus we should introduce some smoothness restriction on the conditional distributions.

Before presenting the last of our theorems we prove several lemmas.

In the following lemmas, I is a compact metric space and $M = R_k$. $g(\theta, m)$ is a bounded continuous function on $I \times M$. J is any bounded interval in M. $\{\nu_n(\theta, \cdot)\}$ $n = 0, 1, \ldots$ is a family of sequences of probability measures on M indexed by θ in I. $\nu_n(\theta, \cdot)$ converges to $\nu_0(\theta, \cdot)$ the UC^* sense relative to $\theta \in I$. Let

$$v_n(\theta) = \int g(\theta, m) \quad \nu_n(\theta, dm), \quad n = 0, 1, \ldots \qquad \ldots \quad (3.6)$$

Lemma 2 : *$g(\theta, m)$ is equicontinuous in θ at each m, i.e. if $\theta_n \to \theta_0$, then $g(\theta_n, m) \to g(\theta_., m)$ uniformly in $m \in J$.*

Proof: This follows immediately from the uniform continuity of $g(\theta, m)$ on $I \times J$.

Lemma 3 : $\sup\limits_{\theta} | v_n(\theta) - v_0(\theta) | \to 0$ *as* $n \to \infty$. $\qquad \ldots \quad (3.7)$

Proof : This lemma is a simple corollary of the general results of Ranga Rao (1960), (1961). We however, give here a short proof for the sake of continuity.

If (3.7) were not true, then there would be a sequence $\{\theta_n\}$ and $\alpha > 0$, such that

$$| v_n(\theta_n) - v_0(\theta_n) | \geqslant \alpha \text{ for each } n. \qquad \ldots \quad (3.8)$$

Since I is compact, there is a subsequence $\{\theta_{n_r}\}$ such that $\theta_{n_r} \to \theta_0$ as $r \to \infty$. We then have

$$\begin{aligned} | v_{n_r}(\theta_{n_r}) - v_0(\theta_{n_r}) | & \\ &= \left| \int g(\theta_{n_r}, m)\, \nu_{n_r}(\theta_{n_r}, dm) - \int g(\theta_{n_r}, m)\, \nu_0(\theta_{n}, dm) \right| \\ &\leqslant \int | g(\theta_{n_r}, m) - g(\theta_0, m) |\, \nu_{n_r}(\theta_{n_r}, dm) \\ &+ \left| \int g(\theta_0, m)\, \nu_{n_r}(\theta_{n_r}, dm) - \int g(\theta_0, m) \nu_0(\theta_0, dm) \right| \\ &+ \left| \int g(\theta_0, m)\, \nu_0(\theta_0, dm) - \int g(\theta_0, m) \nu_0(\theta_{n_r}, dm) \right| \\ &+ \int | g(\theta_0, m) - g(\theta_{n_r}, m) | \nu_0(\theta_{n_r}, dm). \end{aligned}$$

Given any $\epsilon > 0$ the first, second, third and fourth terms on the right hand side can each be made $< \epsilon$ by using Lemma 2 and (2.5), (2.5), (2.7), and Lemma 2 and (2.7), respectively if $r \geqslant R$. Since $\epsilon > 0$ is arbitrary, this is a contradiction to (3.8). Hence the lemma.

Lemma 4 : *$v_0(\theta)$ is continuous in θ.*

Proof : This follows immediately from (2.7).

Let $u_n(m)$ be a sequence of functions on a separable complete metric space M converging to a bounded continuous function $u(m)$ uniformly on every compact set. Let μ_n be a sequence of probability measures on M converging weakly to μ. We then have the following lemma.

Lemma 5 : $\int u_n(m)d\mu_n \to \int u(m)d\mu$.

Proof : The proof is immediate.

In the following theorem $Y = R_k$ and X is any separable complete metric space. Here we impose certain conditions on the conditional distributions similar to those employed by Parzen (1954), Steck (1957) and others.

Theorem 3 : *Let the sequence of (marginal) distributions $\{\mu_n\}$ converge weakly to μ. Let the sequence of (conditional) distributions $\{\nu_n(x, \cdot)\}$ converge to $\nu(x, \cdot)$ in the UC^* sense relative to $x \in I$ for every compact subset I of X. Then the sequence $\{\lambda_n\}$ of (joint) distributions converges weakly to λ_0.*

Proof : Let $g(x, y)$ be any bounded continuous function on $X \times Y$.

$$\text{Let} \quad u_n(x) = \int_Y g(x, y)\, \nu_n(x, dy), \quad u(x) = \int_Y g(x, y)\, \nu(x, dy).$$

Lemma 3 shows that $u_n(x) \to u(x)$ uniformly in $x \in I$ for every compact $I \subset X$. Lemma 4 shows that $u(x)$ is continuous in x.

Now, using Lemma 5,

$$\int g(x, y) d\lambda_n = \int_X d\mu_n \int_Y g(x, y)\ \nu_n(x, x, dy)$$

$$= \int u_n(x) d\mu_n \to \int u(x) d\mu$$

$$= \int_X d\mu \int_Y g(x, y) \nu(x, dy)$$

$$= \int g(x, y) d\lambda$$

where $\lambda(C) = \int_X d\mu \int_{C_x} \nu(x, dy)$ with $C_x = \{y : (x, y) \in C\}$. (See Halmos, 1950). Thus $\lambda = \lambda_0$ and $\lambda_n \Longrightarrow \lambda_0$.

4. A counter example

We present below an example to show that some such conditions, as imposed in Theorems 1, 2 and 3 are, in general, necessary.

X and Y are the real line and S and T the usual field of Borel sets. The random variable (ξ_n, η_n) takes the values $(1/n, 1/n)$, $(1/n, 1+1/n)$, $(1+1/n, 1/n)$ and $(1+1/n, 1+1/n)$ with probabilities $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, and $\frac{1}{8}$ respectively if n is even and with probabilities $\frac{3}{8}$, $\frac{1}{8}$, $\frac{1}{8}$ and $\frac{3}{8}$ respectively if n is odd.

It is easy to see that the marginal distributions of ξ_n and η_n converge weakly to the same distribution with masses $\frac{1}{2}$ and $\frac{1}{2}$ at 0 and 1. The conditional distributions are trivially convergent. The joint distributions do not converge.

5. An illustrative application

In section 1 we have already mentioned some applications of the results of section 3 made in earlier works. As an illustration we can deduce the asymptotic distribution of several sample quantiles from that of a single quantile and Theorem 2. In particular, assuming the theorem for the asymptotic distribution of the sample median (see Smirnov, 1949; Cramer, 1946), we will show that the sample first quartile and the median are jointly asymptotically normally distributed.

Let $z_1, \ldots, z_{4n+3}$ be $4n+3$ independent observations on a random variable Z with distribution function $F(z)$ which possesses a density function $f(z)$. It is assumed that $f(z)$ is continuous and nonzero at θ, the population median and at δ, the population first quartile.

Let us denote $\sqrt{n}(z_{(2n+2)}-\theta)$, $\sqrt{n}(z_{(n+1)}-\delta)$ by (ξ_n, η_n) where $z_{(2n+2)}$ is the sample median and $z_{(n+1)}$ is the sample first quartile. When ξ_n is fixed at x, η_n is the normalised median of a sample of size $(2n+1)$ on the random variable Z truncated to the region $\left(-\infty, \theta+\frac{x}{\sqrt{n}}\right)$, and hence is asymptotically normal. Some algebraic computations show that the mean and variance of the limiting conditional distribution are $\frac{xf(\theta)}{2f(\delta)}$ and $\frac{1}{32f^2(\delta)}$. Since the densities converge pointwise (Cramer, 1946) ξ_n tends strongly to the normal distribution with mean zero and variance $\frac{1}{16f^2(\theta)}$. An application of Theorem 3 shows that the joint distribution of (ξ_n, η_n) converges to the bivariate normal distribution with means zero, zero, variances $\frac{1}{16f^2(\theta)}$, $\frac{1}{64f^2(\delta)}$ and correlation $\frac{1}{\sqrt{3}}$.

References

Billingsley, P. (1956): The invariance principle for dependent random variables. *Trans. Amer. Math. Soc.*, **83**, 250-268.

Cramér, H. (1946): *Mathematical Methods of Statistics*, Princeton University Press.

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

Parzen, E. (1954): *On Uniform Convergence of Families of Sequences of Random Variables*, University of California Publications in Statistics, **2**, No. 2, 23-54

Prohorov, Yu. V. (1956): *Convergence of Random Processes and Limit Theorems in the Theory of Probability*, Theorya Voroyat i. Primon, **1**, 177-238 (Russian).

RANGA RAO, R. (1960): Some limit theorems in probability theory. Unpublished D. Phil. thesis, submitted to the University of Calcutta.

——— (1961): Some theorems on weak convergence of measures and its applications, submitted to *Ann. Math. Stat.*

SCHEFFÉ, H. (1947): A useful convergence theorem for probability distributions. *Ann. Math. Stat.*, 18, 434-438.

SETHURAMAN, J. (1961): Some limit theorems connected with fractile graphical analysis. *Sankhyā*, Series A, 23, 79-90.

SETHURAMAN, J. and SUKHATME, B. V. (1959): Joint asymptotic distribution of U-statistics and order statistics. *Sankhyā*, 21, 289-298.

SMIRNOV, N. V. (1949): Limit distributions for the terms of a variational series. *Trudy. Mat. Inst. A. Steklov.*, 25, 1-61 (Russian).

STECK, G. P. (1957): *Limit Distributions for Conditional Distributions*, University of California Publications in Statistics, 2, No. 12, 237-284.

VARADARAJAN, V. S. (1958): Convergence of stochastic processes. Unpublished D. Phil. thesis, submitted to the University of Calcutta.

Paper received : June, 1961.