SOME LIMIT DISTRIBUTIONS CONNECTED WITH FIXED INTERVAL ANALYSIS*

By J. SETHURAMAN

Indian Statistical Institute

and

University of North Carolina

SUMMARY. The proofs of some theorems stated by the author (Sethuraman, 1963) on the limiting distributions of some statistics that enter in the method of Fixed Interval Analysis are presented.

1. Introduction

Let (Y, X) be a random variable taking values in $(\mathcal{Y} \times \mathcal{X})$ where $\mathcal{Y}$ is E_k the Euclidean space of k dimensions and $\mathcal{X}$ is a measurable space. Let $E_1, E_2, \ldots, E_g$ be g disjoint measurable sets in $\mathcal{X}$ whose union is the whole space $\mathcal{X}$.

$(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ are n independent observations on (Y, X). The number of x_i's that fall in E_j is n_j, $j = 1, \ldots, g$. u_j is defined by the relation

$$u_j = \Sigma^j y_i / n_j \qquad j = 1, \ldots, g$$

where Σ^j is the summation over all "i" such that x_i is in E_j.

Throughout this paper it is assumed that

$$V(Y) < \infty \qquad \ldots \quad (1.1)$$

and

$$\text{prob}\,(X \epsilon E_j) = \pi_j > 0 \qquad j = 1, \ldots, g \qquad \ldots \quad (1.2)$$

where for any random variable Z, $v(Z)$ denotes the variance covariance matrix of Z.

The following theorem is established in Section 3.

Theorem 1 : *The asymptotic distribution of* $(u_1, \ldots, u_g)$ *is the distribution of* g *independent normal distributions.*

This theorem plays a fundamental role in the method of Fixed Interval Analysis (for instance, see Sethuraman (1963)). Interpreted in Sample Survey language this theorem, among other things, states that the post-stratified stratum means are independently distributed in the limit.

2. Notations, definitions and preliminaries

Let $Y(E_j)$, called the conditional random variable of Y given that X is in E_j, denote a random variable on $\mathcal{Y}$ with the distribution defined by prob $(Y(E_j) \epsilon A)$ $=$ prob $(Y \epsilon A,\ X \epsilon E_j)$/prob $(X \epsilon E_j)$. For any random variable Z, $E(Z)$ denotes the vector of expectations of Z.

---

* This research work was done at the Indian Statistical Institute. The writing of this paper was supported by the Mathematics Division of the Air Force Office of Scientific Research under grant number AF-FOSR-62-169.

Define

$$E(Y(E_j)) = \mu_j \qquad \dots \quad (2.1)$$

$$V(Y(E_j)) = \Sigma_j \qquad \dots \quad (2.2)$$

$$p_j = n_j/n \qquad \dots \quad (2.3)$$

$$\sqrt{n}(u_j-\mu_j) = \eta_j(n) \qquad \dots \quad (2.4)$$

$$\sqrt{n}(p_j-\pi_j) = \zeta_j(n) \qquad \dots \quad (2.5)$$

$$j = 1, \dots, g.$$

Let $\{\xi_n(\cdot, \theta)\}$, $n = 0, 1, \dots$ be a sequence of families of probability distributions on the Borel-subsets of E_m (or more generally, of any topological space) and θ vary in a compact topological space K.

Definition: $\{\xi_n(\cdot, \theta)\}$ is said to converge weakly, uniformly and continuously (in other words, in the UC^* sense) to $\xi_0(\cdot, \theta)$ with respect to θ in K if for every bounded continuous function $h(y)$ on E_m

$$\int g(y)\ \xi_n(dy, \theta) \to \int g(y)\ \xi_0(dy, \theta) \text{ uniformly in } \theta$$

and

$$\int g(y)\ \xi_0(dy, \theta) \text{ is a continuous function of } \theta.$$

The following theorem given by the author (Sethuraman, 1961) will be used in Section 3.

Theorem 2 : *Let (Y_n, X_n) be a sequence of random variable on $(E_m \times S)$ where S is a complete separable metric space. Let the conditional probability measure of Y_n given that $X_n = x$ be denoted by $\xi_n(\cdot, x)$ and the marginal distribution of X_n be μ_n. Let $\xi_n(\cdot, x)$ converge in the UC^* sense to $\xi_0(\cdot, x)$ with respect to x in any compact subset of S and μ_n converge weakly to μ_0. Then the joint distribution of (Y_n, X_n) converges weakly to the distribution determined by $\xi_0(\cdot, x)$ and μ_0 or, more precisely, to the distribution of (Y_0, X_0) where*

$$\text{prob}\{Y_0 \in A, \quad X_0 \in B\} = \int_B \xi_0(A, x)\mu(dx).$$

Lemma which is immediate, is useful in establishing the UC^* convergence of a special sequence of families of distributions.

Let

$$Z_{11}, \dots, Z_{1k_1(\theta)}$$

$$Z_{21}, \dots, Z_{2k_2(\theta)}$$

$$\dots\dots\dots\dots\dots\dots$$

$$Z_{n1}, \dots, Z_{nk_n(\theta)}$$

be a triangular scheme of random variables in E_m where the variables in any row are identically and independently distributed. Assume that $E(Z_{n1}) = \nu_n$ and $V(Z_{n1}) = V_n$ are finite and that $V_n \to V$ as $n \to \infty$. Again let $\inf_\theta \ k_n(\theta) \to \infty$ as $n \to \infty$.

Let $MN(\alpha, L)$ stand for the multivariate normal distribution with mean vector α and variance covariance matrix L.

Lemma 1 : *The sequence of families of distributions of*

$$\{(Z_{n_1}+\ldots+Z_{nk_n(\theta)}-k_n(\theta)\nu_n)/\sqrt{k_n(\theta)}\}$$

converges in the UC^* sense to the distribution $MN(0, V)$ with respect to θ.

3. Main theorems

We first prove the following lemma.

Lemma 2 : *The distributions of $(\eta_1(n), \ldots, \eta_g(n))$ given that $\zeta(n) = z$, $\sum_1^g z_i = 0$ converges in the UC^* sense to the distribution $MN(0, \Lambda)$ with respect to z in any closed bounded subset of E_g, where*

$$\Lambda = \begin{pmatrix} \frac{1}{\pi_1}\Sigma_1 & 0 & 0 \\ 0 & \frac{1}{\pi_2}\Sigma_2 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \frac{1}{\pi_g}\Sigma_g \end{pmatrix} \quad \ldots \quad (3.1)$$

Proof: The event $\zeta(n) = z$ is equivalent with probability one to the event $n_i = [n\pi_i+\sqrt{n}\,z_i]$ $i = 1, \ldots, g$, since

$$\text{prob }\{n\pi_i+\sqrt{n}\zeta_i(n) = [n\pi_i+\sqrt{n}\zeta_i(n)],\quad i = 1, \ldots, g\} = 1.$$

The conditional distribution of $y_1, \ldots, y_n$ given that $n_i = [n\pi_i+\sqrt{n}\,z_i]$, $i = 1, \ldots, g$ is the distribution of g independent samples of size $n_1, \ldots, n_g$ on $Y(E_1), \ldots, Y(E_g)$, respectively. $\sqrt{\frac{n_1}{n}}\,\eta_1(n), \ldots, \sqrt{\frac{n_g}{n}}\,\eta_g(n)$ are the normalized means of these g independent samples. For z in a closed bounded subset of E_g we note that $\inf_{i,z}$ $[n\pi_i+\sqrt{n}\,z_i]\to\infty$ as $n\to\infty$. Thus all the conditions of Lemma 1 are satisfied. Further $[n\pi_i+\sqrt{n}\,z_i)/n$ tends to π_i, uniformly in z in any closed bounded subset of E_g. Hence the conditional distributions of $(\eta_1(n), \ldots, \eta_g(n))$ given that $\zeta(n) = z$ converges in the UC^* sense to the distribution $MN(0, \Lambda)$ with respect to z in any closed bounded subset of E_g.

Theorem 3 : *The joint distribution of $(\eta_1(n), \ldots, \eta_g(n), \zeta(n))$ converges weakly to the distribution $MN(0, B)$*

where

$$B = \begin{pmatrix} \Lambda & 0 \\ 0 & C \end{pmatrix} \quad \ldots \quad (3.2)$$

and where

$$C = \begin{vmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_g \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_g \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot \\ -\pi_1\pi_g & -\pi_2\pi_g & \cdots & \pi_g(1-\pi_g) \end{vmatrix} \quad \ldots \quad (3.3)$$

Proof: This theorem is an immediate consequence of Theorem 2, Lemma 2 and the observation that the distribution of $\zeta(n)$ converges weakly to the distribution $MN(0, C)$.

Proof of Theorem 1 : Theorem 1 is contained in Theorem 3.

References

SETHURAMAN, J. (1961) : Some limit theorems for joint distributions. *Sankhyā*, Series A, **23**, 379-386.

———— (1963) : Fixed interval analysis and fractile analysis. *Contributions to Statistics*, Presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday, Pergamon Press, Oxford and Statistical Publishing Society, Calcutta.

Paper received : February, 1963.