The Annals of Statistics
1975, Vol. 3, No. 3, 787–792

WEAK CONVERGENCE OF THE EMPIRIC PROCESS FOR INDEPENDENT RANDOM VARIABLES

BY R. RECHTSCHAFFEN

IBM Corporation

Introduction. This paper investigates the (weak) convergence properties of the Empiric Process in a more general framework of independent random variables (without common distribution). In this situation the Empiric Process converges to a Gaussian Process on the unit interval which is "dominated" by the Brownian Bridge.

The weak convergence of the Empirical Process, the "smoothed" Empirical Process and the Empiric Process for a sequence of independent random variables is derived without the restriction that they have a common distribution. The characteristics of the Gaussian Process which is their "weak" limit is discussed.

The author would like to acknowledge the guidance given him by his advisor, Professor Kei Takeuchi, and the interest and assistance of Professor E. Parzen and Professor W. Rosenkrantz.

1. Theoretical preliminaries. Let S be a metric space, $\mathscr{S}$ the σ-algebra generated by the open sets. If P_n and P are probability measures on $(S, \mathscr{S})$ such that $\int_S f\,dP_n \to \int_S f\,dP$ for every bounded continuous function f on S, we say that P_n converges weakly to P and write $P_n \Rightarrow P$. Further generic properties of weak convergence can be found in [2]; in particular, $P_n \Rightarrow P$ if and only if $P_n(A) \to P(A)$ for all subsets A such that $P(\partial A) = 0$ where ∂A is the boundary of A.

Let Y_n and Y be random elements of S. We say that Y_n converges in distribution to Y and write $Y_n \to_{\mathscr{D}} Y$ if and only if the probability distributions of Y_n converge weakly to the probability distribution of Y.

For the case of $\hat{S}(x, \omega)$, the empirical distribution function of independent identically distributed random variables with continuous distribution F, weak convergence can be used to prove the Glivenko–Cantelli Theorem ([3], page 20)

$$P[\omega \mid \sup_{-\infty<x<\infty} |\hat{S}_n(x, \omega) - F(x)| \to 0] = 1 \tag{1}$$

and the sharper result due to Kolmogorov ([2], page 104)

$$P[\omega \mid n^{\frac{1}{2}} \sup_{-\infty<x<\infty} |\hat{S}_n(x, \omega) - F(x)| \leqq \alpha] \to 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2\alpha^2} \tag{2}$$

for all $\alpha \geqq 0$. The method of proof involves weak convergence of random elements of the metric space D, of all real-valued functions on $[0, 1]$ which are right continuous and have left-hand limits. The metric used for D is defined by ([2], page 112)

$$d_0(x, y) = \inf\{\varepsilon > 0 \mid \|\lambda\| \leqq \varepsilon, \sup_{0\leqq t\leqq 1} |x(t) - y(\lambda(t))| \leqq \varepsilon\}$$

where $x, y \in D$, λ is a non-decreasing function on $[0, 1]$ such that $\lambda(0) = 0$ and

Received July 1970; revised August 1974.

$\lambda(1) = 1$ and $||\lambda|| = \sup_{s \neq t} |\log(\lambda(t) - \lambda(s))/(t - s)|$. The metric space (D, d_0) is separable and complete. A special subset C of D, the space of all continuous functions on $[0, 1]$, is a separable and complete metric space in the sup-norm metric $\rho(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$. The importance of identifying (C, ρ) and (D, d_0) as complete separable metric spaces is that weak convergence can be established by use of Prohorov's Theorem ([2], page 37).

Let $\{X_i\}$ be a sequence of independent random variables defined on $(\Omega, \mathscr{B})$ and let F_i be the distribution function of X_i for each i, and $\bar{F}_n(\cdot) = (1/n) \sum_1^n F_i(\cdot)$ for each n. Let P_X denote the probability measure of a random variable X with distribution function F. Let I_A denote the characteristic function of the set A and $\mathscr{R}$ be the σ-algebra generated by the open subsets of the real line. For each $A \in \mathscr{R}$ and each $\omega \in \Omega$, we define the empirical measure for $X_1(\omega), \cdots, X_n(\omega)$ by

$$P_{n,\omega}(A) = \frac{1}{n} \sum_{i=1}^n I_A(X_i(\omega)) .$$

THEOREM 1.1. *A necessary and sufficient condition that* $P[\omega \mid P_{n,\omega} \Rightarrow P_x] = 1$ *is that* $\bar{F}_n(x) \to F(x)$ *for all continuity points of* F.

PROOF. Let $A_x = (-\infty, x]$, where x is a continuity point of F. By the Strong Law of Large Numbers,

$$P[P_{n,\omega}(A_x) \to P_X(A_x)] = 1 .$$

Let V be any countable dense subset of $\{x \mid x \text{ a continuity point of } F\}$, then

$$P[P_{n,\omega}(A_x) \to P_X(A_x) \text{ for all } x \in V] = 1 ,$$

from which the result follows ([2], page 14). The converse follows from the bounded convergence theorem.

COROLLARY 1.1. (Glivenko-Cantelli Theorem). *If* $F(x)$ *is continuous for all* x, *and* $\lim \bar{F}_n(x) = F(x)$ *for all* x, *then*

$$P[\lim_{n\to\infty} \sup_{-\infty<x<\infty} |\hat{S}_n(x) - F(x)| = 0] = 1$$

where $\hat{S}_n(x)$ *is the empirical distribution function of the independent random variables* $\{X_i\}$ $i = 1, 2, \cdots, n$.

PROOF. Applying the previous theorem to all sets of the form $(-\infty, x]$ we have that $\hat{S}_n(x)$ converges with probability 1 to $F(x)$. Since both are monotonic bounded functions of x, it follows that the convergence is uniform.

2. Notation. Let

(N1) $$b_{n,i}(t) = P[F_n(X_i) \leq t] ,$$

(N2) $$h_n(s, t) = \frac{1}{n} \sum_{i=1}^n b_{n,i}(s)(1 - b_{n,i}(t)) \qquad 0 \leq s \leq t \leq 1 ,$$

and

(N3) $$h(s, t) = \lim h_n(s, t) \qquad \text{which is assumed to exist.}$$

The notation,

(N4) $$Y(\cdot, \omega) \propto \{(\alpha_i(\omega), \beta_i(\omega)) : i = 0, \cdots, k\}$$

means that $Y(t, \omega)$ is the polygonal arc connecting the $k+1$ points, $(\alpha_i(\omega), \beta_i(\omega))$, $i = 0, \cdots, k$, with straight lines,

$$Y(t, \omega) = \beta_{i-1}(\omega) + \frac{t - \alpha_{i-1}(\omega)}{\alpha_i(\omega) - \alpha_{i-1}(\omega)} \beta_i(\omega)$$

if

$$\alpha_{i-1}(\omega) \leqq t \leqq \alpha_i(\omega) \quad \text{and} \quad \alpha_0(\omega) \equiv 0 , \quad \alpha_k(\omega) \equiv 1 .$$

If $0 < \alpha_1(\omega) < \alpha_2(\omega) < \cdots < \alpha_{k-1}(\omega) < 1$, then $Y(t, \omega)$ is a well-defined element of C.

Let

(N5) $$\hat{\bar{F}}_n(t, \omega) = \frac{1}{n} \sum_{i=1}^n I_{[0,t]}(\bar{F}_n(X_i(\omega))) ,$$

(N6) $$\hat{F}_n(t, \omega) = \frac{1}{n} \sum_{i=1}^n I_{[0,t]}(F(X_i(\omega))) ,$$

(N7) $$\tilde{F}_n(\cdot, \omega) \propto \left\{\left(F(X_{(i)n}(\omega)), \frac{i}{n+1}\right) : i = 0, \cdots, n+1\right\} ,$$

and

(N8) $$F_n^*(\cdot, \omega) \propto \left\{\left(\frac{i}{n+1}, F(X_{(i)n}(\omega))\right) : i = 0, \cdots, n+1\right\} .$$

We now define the following stochastic processes on $[0, 1]$:

(N9) $$\hat{L}_n(t, \omega) = n^{\frac{1}{2}}(\hat{\bar{F}}_n(t, \omega) - t) ,$$

the empirical process

(N10) $$L_n(t, \omega) = n^{\frac{1}{2}}(\hat{F}_n(t, \omega) - t) ,$$

the smoothed empirical process

(N11) $$\tilde{L}_n(t, \omega) = n^{\frac{1}{2}}(\tilde{F}_n(t, \omega) - t) ,$$

and the empiric process [4]

(N12) $$L_n^*(t, \omega) = n^{\frac{1}{2}}(F_n^*(t, \omega) - t) .$$

3. Main theorems. Applying Theorem 1.1 to the sequence of random variables $\{F(X_i)\}$, we have the following corollary:

COROLLARY 1.2. *Under the hypothesis of Corollary* 1.1,

$$P[\lim_{n\to\infty} \sup_{0 \leqq t \leqq 1} |\hat{F}_n(t, \omega) - t| = 0] = 1 .$$

PROOF. Let $p_i(t) = P[F(X_i) \leqq t]$, and $\bar{p}_n(t) = (1/n) \sum_{i=1}^n p_i(t)$. We define $Z_n = X_{I_n}$ where $\{I_n\}$ is a sequence of random variables independent of $\{X_i\}$ and for each n, $P[I_n = k] = 1/n$, $1 \leqq k \leqq n$. Then $P[Z_n \leqq x] = \bar{F}_n(x)$. Since $\bar{F}_n(x) \to F(x)$, for all x, $Z_n \to_{\mathscr{D}} X$ where X has distribution function F. Since F is continuous, $F(Z_n) \to_{\mathscr{D}} F(X) \sim U$, where U is a uniformly distributed random variable on [0, 1]. Hence for all t, $P[F(Z_n) \leqq t] \to P[U \leqq t] = t$, and by the definition of Z_n, $\bar{p}_n(t) = P[F(Z_n) \leqq t] \to t$ for all t. The random variables $\{F(X_i)\}$ and their distribution functions $\{p_i(t)\}$ satisfy the hypothesis of Corollary 1.1 where the limiting distribution is $F(t) = t$, $0 \leqq t \leqq 1$, and the result follows.

We note that in the non-identically distributed case in order for (1) or (2) to hold it is necessary that $\bar{F}_n \to F$. A sufficient condition for (2) is the weak convergence of the empirical process (N10) to the Brownian Bridge. It is of interest to note that the convergence of the smoothed empirical (N11) implies the convergence of the empiric process (N12) directly as follows: Since $\tilde{F}_n$ is the inverse of F_n^* (see N7, N8) we have that $\tilde{L}_n(F_n^*(t, \omega), \omega) = -L_n^*(t, \omega)$. Further, $\rho(\tilde{F}_n, \hat{F}_n) \leqq 1/n$ so that Corollary 1.2 implies $\sup_{0 \leqq t \leqq 1} |\tilde{F}_n(t, \omega) - t| \to_p 0$; $\sup_{0 \leqq t \leqq 1} |F_n^*(t, \omega) - t| \to_p 0$. Hence ([2], page 145) if $\tilde{L}_n$ converges in distribution, so does L_n^*. The fact that $\rho(L_n, \tilde{L}_n) = 1/n^{\frac{1}{2}}$ establishes the following result.

LEMMA 1.1. *If any of the three processes L_n, $\tilde{L}_n$, and L_n^* converge so do the other two. Further, if the limiting process is symmetric (has zero expectation) they converge to a common limit.*

It is actually $\hat{L}_n$ (N9), which is a generalization of L_n when all the F_i are not assumed to be equal, which can be shown to converge. The proof parallels the proof of convergence in the identically distributed case and is therefore omitted ([2], pages 125, 128) and differs only in minor technicalities.

THEOREM 1.2. *Suppose $\{X_i\}$ is a sequence of continuous random variables. Let $h(s, t)$ (N3) exist for all $0 \leqq s \leqq t \leqq 1$. Then there exists a Gaussian process Z with support on C such that; $E(Z(t)) = 0$, $E(Z(s)Z(t)) = h(s, t)$ $0 \leqq s \leqq t \leqq 1$, and $\hat{L}_n \to_{\mathscr{D}} Z$.*

THEOREM 1.3. *The conclusion of Theorem 1.2 also follows from the existence of $h(s, t)$ and the assumption that $\{X_i\}$ are discrete random variables such that $P[X_i = X_j] = 0$ for $i \neq j$ and the existence of a uniform bound for the ratio of the maximum jump in $\bar{F}_n$ to the minimum jump in $\bar{F}_n$.*

To establish the convergence of L_n from the above we require the following additional assumption B.

ASSUMPTION A. Let $\{X_i\}_{i=1,2,\dots}$ either satisfy the hypothesis of Theorem 1.2 or 1.3.

ASSUMPTION B. Let $F(x)$ be an arbitrary distribution function such that

$$\lim_n n^{\frac{1}{2}} \sup_x |\bar{F}_n(x) - F(x)| = 0 ,$$
$$P[0 < F(X_i) \neq F(X_j) < 1] = 1 \qquad \text{for } i \neq j .$$

THEOREM 1.4. *Under Assumptions* A *and B,*

$$\tilde{L}_n \to_{\mathscr{D}} Z \quad in \quad C\,,$$
$$L_n \to_{\mathscr{D}} Z \quad in \quad D\,,$$
$$L_n^* \to_{\mathscr{D}} Z \quad in \quad C\,.$$

PROOF. Except for some minor complications which require careful handling when $\bar{F}_n(X_{(n)n}) = 1$ the proof is straightforward. Omitting these complications we define (see N4)

$$L_{1n}(\cdot, \omega) \propto \left\{\left(\bar{F}_n(X_{(i)n}(\omega)),\, n^{\frac{1}{2}}\left(\frac{i}{n+1} - \bar{F}_n(X_{(i)n}(\omega))\right)\right);\, i = 0, \cdots, n+1\right\}.$$

Theorem 1.2 or 1.3 (whichever is appropriate) implies that $L_{1n} \to_{\mathscr{D}} Z$ since $\rho(L_{1n}, \hat{L}_n) \leqq 1/n^{\frac{1}{2}}$, and $\hat{L}_n \to_{\mathscr{D}} Z$. Let

$$L_{2n}(\cdot, \omega) \propto \left\{\left(F(X_{(i)n}(\omega)),\, n^{\frac{1}{2}}\left(\frac{i}{n+1} - \bar{F}_n(X_{(i)n}(\omega))\right)\right);\, i = 0, \cdots, n+1\right\}$$

and

$$\lambda_n(\cdot, \omega) \propto \{(F(X_{(i)n}(\omega)),\, \bar{F}_n(X_{(i)n}(\omega)));\, i = 0, \cdots, n+1\}\,,$$

then $L_{1n}(\lambda_n(t, \omega), \omega) = L_{2n}(t, \omega)$. Since $\sup_{0\leqq t\leqq 1}|\lambda_n(t, \omega) - t| \leqq \sup_x |\bar{F}_n(x) - F(x)|$, which converges to zero by Assumpion B, we can apply the result used in Lemma 1.1 ([2], page 145) to show that $L_{2n} \to_{\mathscr{D}} Z$. Finally, $\rho(\tilde{L}_n, L_{2n}) \leqq n^{\frac{1}{2}} \sup_x |\bar{F}_n(x) - F(x)|$, and Assumption B establish the convergence of $\tilde{L}_n$. Lemma 1.1 can then be used to establish the remaining conclusions.

The relationship between the Gaussian process Z and the Brownian Bridge, as well as the Kolmogorov result (2) on the limiting distribution of $n^{\frac{1}{2}} \sup_{0\leqq t\leqq 1} |L_n(t)|$, can now be stated. Since $h(x) = \sup_{0\leqq t\leqq 1} |x(t)|$ is a continuous function on C, $|h(x) - h(y)| \leqq \rho(x, y)$ for all $x, y \in C$, we have that $\sup_{0\leqq t\leqq 1} n^{\frac{1}{2}}|L_n(t)| \to_{\mathscr{D}} \sup_{0\leqq t\leqq 1} |Z(t)|$ under Assumptions A and B.

THEOREM 1.5. *For all values of* $\alpha \geqq 0$,

$$P[\sup_{0\leqq t\leqq 1} |Z(t)| \leqq \alpha] \geqq P[\sup_{0\leqq t\leqq 1} |W^0(t)| \leqq \alpha]\,.$$

PROOF. This theorem follows from the fact that the quadratic form generated by the Brownian Bridge W^0 dominates the quadratic form generated by Z and [1]. To prove this dominance, let,

$0 \leqq t_1 \leqq t_2 \leqq \cdots \leqq t_k \leqq 1$:

- $\Sigma^0(t_1, \cdots, t_k)$ the dispersion matrix of $(W^0(t_1), \cdots, W^0(t_k))$
- $\Sigma(t_1, \cdots, t_k)$ the dispersion matrix of $(Z(t_1), \cdots, Z(t_k))$.

Then with $\bar{F}_n$ continuous, we have that

$$\frac{1}{n} \sum_{i=1}^n b_{n,i}(s) = s\,, \qquad \text{for all } s, \text{ and}$$

$$s(1-t) - h_n(s, t) = \frac{1}{n} \sum_{i=1}^n (b_{n,i}(s) - s)(b_{n,i}(t) - t) \qquad \text{(N1, N2)}\,.$$

Hence

$$\sum_{i,j=1}^{k} c_i c_j [\Sigma_{ij}^0(t_1, \cdots, t_k) - \Sigma_{ij}(t_1, \cdots, t_k)] = \lim_n \frac{1}{n} [\sum_{i=1}^{k} c_i(b_{n,i}(t_i) - t_i)]^2 \geqq 0 .$$

Since

$$\Sigma_{ij}^0(t_1, \cdots, t_k) = t_i(1 - t_j) , \qquad 1 \leqq i \leqq j \leqq k .$$

$$c^T \cdot \Sigma^0(t_1, \cdots, t_k) \cdot c \geqq c^T \cdot \Sigma(t_1, \cdots, t_k) \cdot c .$$

REFERENCES

[1] ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 173.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] LOÈVE, M. (1963). *Probability Theory.* Van Nostrand, Princeton.

[4] ROSENKRANTZ W. (1969). A rate of convergence for the von Mises statistic. *Trans. Amer. Math. Soc.* **139** 330.

IBM CORPORATION
2651 STRANG BOULEVARD
YORKTOWN HEIGHTS, NEW YORK 10598