Proc. Camb. Phil. Soc. (1968), **64**, 155
PCPS 64-21
Printed in Great Britain

The weak convergence of the empirical process with random sample size†

BY RONALD PYKE
University of Washington

(*Received* 10 *March* 1967)

1. *Introduction.* In many applied probability models, one is concerned with a sequence $\{X_n: n \geqslant 1\}$ of independent random variables (r.v.'s) with a common distribution function (d.f.), F say. When making statistical inferences within such a model, one frequently must do so on the basis of observations $X_1, X_2, \ldots, X_N$ where the sample size N is a r.v. For example, N might be the number of observations that it was possible to take within a given period of time or within a fixed cost of experimentation. In cases such as these it is not uncommon for statisticians to use fixed-sample-size techniques, even though the random sample size, N, is not independent of the sample. It is therefore important to investigate the operating characteristics of these techniques under random sample sizes. Much work has been done since 1952 on this problem for techniques based on the sum, $X_1 + \ldots + X_N$ (see, for example, the references in (3)). Also, for techniques based on $\max(X_1, X_2, \ldots, X_N)$, results have been obtained independently by Barndorff-Nielsen (2) and Lamperti (9).

It is the purpose of this paper to study the asymptotic distribution, under random sample sizes, of statistics like those of Kolmgorov, Smirnov, Cramér and von Mises, which are distribution-free statistics based upon the empirical d.f. Previous work on this problem is limited to (8) and (1). In these references the random sample sizes are Poisson r.v.'s independent of the sample. Unlike the cases mentioned above of the sum with its linearity and the maximum with its monotoneity, statistics based upon the empirical d.f. do not have such simplifying properties. In fact, a study of any particular statistic, such as Kolmogorov's D_n, leads one quite naturally into a study of the weak convergence of the empirical process ((2·2) below). The method of proof seems to be of interest since it utilizes results about stochastic process with two-dimensional parameter sets.

2. *The main theorem.* Let $\{X_n: n \geqslant 1\}$ be a sequence of independent r.v.'s each with the uniform d.f. on (0, 1). Let $\{N_t: t \geqslant 0\}$ be a positive-integer-valued stochastic process satisfying

$$t^{-1}N_t \xrightarrow{p} 1 \quad \text{as} \quad t \to +\infty. \tag{2·1}$$

For $0 \leqslant u \leqslant 1$ and integral $n \geqslant 1$, set

$$W_n(u) = n^{\frac{1}{2}}[F_n(u) - u], \tag{2·2}$$

where F_n denotes the empirical d.f. based on $(X_1, X_2, \ldots, X_n)$. Although

$$\{W_n(u): 0 \leqslant u \leqslant 1\}$$

† This research was supported in part by the National Science Foundation under G-5719.

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press

is the standard empirical process, we will find it more convenient to introduce one whose sample paths are continuous. To this end, let $U_1 \leqslant \ldots \leqslant U_n$ denote the ordered values of $(X_1, X_2, \ldots, X_n)$, and let F_n^* be the function obtained by joining the points $\{(0,0), (U_i, i/n)$ for $i = 1, 2, \ldots, n, (1,1)\}$ with straight lines. Then set, for $0 \leqslant u \leqslant 1$,

$$W_n^*(u) = n^{\frac{1}{2}}[F_n^*(u) - u]. \tag{2·3}$$

(We will assume that the definition of F_n^* is slightly modified to ensure that F_n^* is always a continuous function even on the null event where there are ties among the X^i's.) Let $\{W(u): 0 \leqslant u \leqslant 1\}$ denote a tied-down Wiener process, i.e. a Gaussian process with $E[W(u)] = 0$ and $E[W(u)\,W(v)] = u(1-v)$ for all $0 \leqslant u \leqslant v \leqslant 1$.

Let C denote the set of all continuous real functions on $[0,1]$ and let D denote the set of real functions on $[0,1]$ with no discontinuities of the second kind. On each of these sets there exists a metric under which one obtains a complete separable metric space. For C, the metric is the uniform metric, $\rho(f,g) = \sup_{0 \leqslant x \leqslant 1} |f(x) - g(x)|$. For details of the definition of D and its metric, d say, the reader is referred to appendix 1 of Prokhorov (11).

Let P_n denote the probability measure on the Borel sets of D which corresponds to the W_n-process and let P_n^* and P denote the probability measures on the Borel sets of C corresponding to the W_n^*- and W-processes, respectively. (When convenient, we shall also let P_n^* and P denote their extensions to D, recalling from (11) that C is a closed subset in D under d.) It is known (see Donsker (5), Dudley (7) and Prokhorov (11)) that $\{P_n\}$ converges weakly to P, denoted $P_n \xrightarrow{w} P$, where weak convergence is defined by saying that $\int g\,dP_n \to \int g\,dP$ for all continuous bounded functions g defined on the appropriate metric space. (An equivalent definition of weak convergence for separable metric spaces is that $P_n(A) \to P(A)$ for all finite intersections of open spheres, A, whose boundaries, $\bar{A}$, are P-null sets.) As a consequence of this result, one obtains that the continuous processes $\{W_n^*\}$ have weakly-convergent measures on (C, ρ).

COROLLARY 2·1. *On* (C, ρ), $P_n^* \xrightarrow{w} P$ *as* $n \to \infty$.

Proof. This may be proved in several ways. Perhaps the shortest proof is to use the fact that one can construct a probability space and a sequence of processes $\{\overline{W}, \overline{W}_n: n \geqslant 1\}$ with sample paths in D having the same finite-dimensional d.f.'s as $\{W, W_n: n \geqslant 1\}$, but for which it is also true that $\rho(W_n, \overline{W}) \xrightarrow{p} 0$. This fact is a consequence of Skorokhod's Corollary 3·1·2 in (13) and the weak convergence, $P_n \xrightarrow{w} P$, and is given as Lemma 2·1 in (12). Since by definition $\rho(W_n, W_n^*) = n^{-\frac{1}{2}}$ (a.s.), it follows that $\rho(\overline{W}_n^*, \overline{W}) \xrightarrow{p} 0$ where $\overline{W}_n^*$ is the continuous process obtained from $\overline{W}_n$ in the same way as W_n^* was obtained from W_n. Thus if ϕ is any continuous bounded functional on (C, ρ) it follows that $\phi(\overline{W}_n^*) \xrightarrow{p} \phi(\overline{W})$. Hence

$$E\phi(W_n^*) = E\phi(\overline{W}_n^*) \to E\phi(\overline{W}) = E\phi(W),$$

which completes the proof.

For $t \geqslant 0$, define

$$V_t = W_{N_t} \quad \text{and} \quad V_t^* = W_{N_t}^*. \tag{2·4}$$

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press

Furthermore, let $Q_t(Q_t^*)$ denote the probability measure determined on $D(C)$ by the process $\{V_t(u)\colon 0 \leqslant u \leqslant 1\}(\{V_t^*(u)\colon 0 \leqslant u \leqslant 1\})$. The main purpose of this paper is to prove

THEOREM 2·1. *As* $t \to \infty$, $Q_t^* \xrightarrow{w} P$ *on* (C, ρ) *and* $Q_t \xrightarrow{w} P$ *on* (D, d).

Proof. We shall deal with $\{Q_t^*\}$ first. For a given $\epsilon > 0$, there exists t_ϵ by (2·1) such that

$$P[r_t \leqslant N_t \leqslant s_t] \geqslant 1-\epsilon \quad \text{for} \quad t \geqslant t_\epsilon, \tag{2·5}$$

where $r_t = [t(1-\epsilon)]$, $s_t = [t(1+\epsilon)]$ and $[x]$ denotes the greatest integer not exceeding x.

Let $S_1, S_2, \ldots, S_k$ be k open spheres in (C, ρ) whose boundaries have P measure zero. Set $A = \bigcap_{i=1}^{k} S_i$. By the remark preceding Corollary 2·1 it suffices to study the limit of Q_t^* on sets of this form. For $\delta > 0$, let A^δ, $A^{-\delta}$ denote the intersection of the corresponding k spheres with the same centres and radii increased, decreased, respectively by δ. It follows that if we write

$$M_{t,\epsilon}^* = \max_{r_t \leqslant k \leqslant s_t} \rho(W_{r_t}^*, W_k^*),$$

then
$$\begin{aligned} P[V_t^* \in A] &\leqslant P[V_t^* \in A, r_t \leqslant N_t \leqslant s_t] + \epsilon \\ &\leqslant P[W_{N_t}^* \in A, r_t \leqslant N_t \leqslant s_t, M_{t,\epsilon}^* \leqslant \delta] + P[M_{t,\epsilon}^* > \delta] + \epsilon \\ &\leqslant P[W_{r_t}^* \in A^\delta] + P[M_{t,\epsilon}^* > \delta] + \epsilon \end{aligned}$$

and
$$\begin{aligned} P[V_t^* \in A] &\geqslant P[V_t^* \in A, r_t \leqslant N_t \leqslant s_t] \\ &\geqslant P[W_{r_t}^* \in A^{-\delta}, r_t \leqslant N_t \leqslant s_t, M_{t,\epsilon}^* \leqslant \delta] \\ &\geqslant P[W_{r_t}^* \in A^{-\delta}] + 1 - \epsilon + P[M_{t,\epsilon}^* \leqslant \delta] - 2 \\ &= P[W_{r_t}^* \in A^{-\delta}] - P[M_{t,\epsilon}^* > \delta] - \epsilon; \end{aligned}$$

that is,
$$P_{r_t}^*(A^{-\delta}) - P[M_{t,\epsilon}^* > \delta] - \epsilon \leqslant Q_t^*(A) \leqslant P_{r_t}^*(A^\delta) + P[M_{t,\epsilon}^* > \delta] + \epsilon.$$

By using only values of δ for which the boundaries of A^δ and $A^{-\delta}$ are P-null, it follows from Corollary 2·1 that

$$\begin{aligned} P(A^{-\delta}) - \limsup_t P[M_{t,\epsilon} > \delta] - \epsilon &\leqslant \liminf Q_t^*(A) \\ &\leqslant \limsup Q_t^*(A) \leqslant P(A^\delta) + \limsup_t P[M_{t,\epsilon}^* > \delta] + \epsilon. \end{aligned}$$

The weak convergence, $Q_t^* \xrightarrow{w} P$, would then follow by letting $\delta \searrow 0$ if we could prove that for all $\delta > 0$,

$$\lim_{\epsilon \to 0} \limsup_{t \to \infty} P[M_{t,\epsilon}^* > \delta] = 0. \tag{2·6}$$

Set $\nu_n(u) = nF_n(u)$ for $0 \leqslant u \leqslant 1$. Then, for $m > k$

$$W_k(u) - W_m(u) = k^{-\frac{1}{2}}\nu_k(u) - m^{-\frac{1}{2}}\nu_m(u) - u(k^{\frac{1}{2}} - m^{\frac{1}{2}}).$$

If we let ν'_{m-k} and W'_{m-k} be defined just the same as ν_{m-k} and W_{m-k} but in terms of $(X_{k+1}, \ldots, X_m)$ instead of $(X_1, \ldots, X_{m-k})$, then we may write

$$\begin{aligned} W_k(u) - W_m(u) &= \nu_k(u)\, k^{-\frac{1}{2}}[1 - (k/m)^{\frac{1}{2}}] - uk^{\frac{1}{2}}[1 - (k/m)^{\frac{1}{2}}] \\ &\qquad - ukm^{-\frac{1}{2}} - \nu'_{m-k}(u)\,(m-k)^{-\frac{1}{2}}\,(1-k/m)^{\frac{1}{2}} + um^{\frac{1}{2}} \\ &= [1 - (k/m)^{\frac{1}{2}}]\, W_k(u) - (1-k/m)^{\frac{1}{2}} W'_{m-k}(u). \end{aligned} \tag{2·7}$$

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press

Since by construction $\rho(W_n, W_n^*) = n^{-\frac{1}{2}}$ a.s.,

$$\max_{k \leqslant m \leqslant n} \rho(W_k^*, W_m^*) \leqslant 2k^{-\frac{1}{2}} + \max_{k \leqslant m \leqslant n} \rho(W_k, W_m)$$
$$\leqslant 2k^{-\frac{1}{2}} + [1 - (k/n)^{\frac{1}{2}}]\rho(W_k, 0) + \max_{1 \leqslant j \leqslant n-k} (j/k)^{\frac{1}{2}} \rho(W'_j, 0).$$

From this inequality, it follows that (2·6) would be established if one could prove that for all $\delta > 0$,

$$\lim_{\epsilon \to 0} \limsup_{t \to \infty} P[\{1 - (r_t/s_t)^{\frac{1}{2}}\}\rho(W_{r_t}, 0) > \delta] = 0 \tag{2·8}$$

and

$$\lim_{\epsilon \to 0} \limsup_{t \to \infty} P[\max_{1 \leqslant j \leqslant s_t - r_t} (j/r_t)^{\frac{1}{2}} \rho(W'_j, 0) > \delta] = 0. \tag{2·9}$$

Since $\rho(.\,, 0)$ is a continuous functional a.s.$-P$ on (D, d), the weak convergence, $P_n \xrightarrow{w} P$, implies that $\rho(W_{r_t}, 0)$ converges in law as $t \to \infty$, so that (2·8) is true. (Actually, $\rho(W_n, 0)$ is simply Kolmogorov's statistic, D_n, so that the convergence in law of $\{\rho(W_n, 0)\}$ is well known (cf. (4)).) It remains to establish (2·9). To do this, we shall show that $\{M_n\}$ is $O_p(1)$, where

$$M_n = \max_{1 \leqslant k \leqslant n} (k/n)^{\frac{1}{2}} \rho(W_k, 0) \quad (n \geqslant 1). \tag{2·10}$$

We shall show that $\{M_n\}$ is $O_p(1)$ by studying the collection of processes

$$\{(W_1(u), \ldots, W_n(u)) \colon 0 \leqslant u \leqslant 1\},$$

for which we first introduce a specific representation.

Let $\{N_n(u, t) \colon 0 \leqslant u, t \leqslant 1\}$ be a two-dimensional homogeneous Poisson process with parameter n. (Here and in what follows, a 'two-dimensional process' will mean a real valued stochastic process with a two-dimensional index set.) That is, $N_n(u, t)$ is a Poisson r.v. with expectation $E[N_n(u, t)] = nut$. Set

$$N_n(t) = N_n(1, t), \quad V_n(u, t) = N_n(u, t) - nut \quad \text{and} \quad V_n(t) = N_n(t) - nt.$$

Notice that $\{N_n(t) \colon 0 \leqslant t \leqslant 1\}$ is an ordinary Poisson process on $[0, 1]$ with parameter n. Assume that the sample functions $N_n(u, .)$ and $N_n(.\,, t)$ are right continuous a.s., for each u and t.

It is well known that if $\{N_n(u), 0 \leqslant u \leqslant 1\}$ is a Poisson process with mean $E[N_n(u)] = nu$, the conditional probability distribution of the first n discontinuity points of $N_n(.\,)$ given $[N_n(1) = n]$ is the same as the distribution of n order statistics from a sample of n independent uniform r.v.'s on $(0, 1)$ (see (6), p. 400). In particular, this implies that conditionally given $[V_n(1) = 0]$ the process $\{n^{-\frac{1}{2}} V_n(u) \colon 0 \leqslant u \leqslant 1\}$ is equivalent to $\{W_n(u) \colon 0 \leqslant u \leqslant 1\}$. An analogous result is possible for the two-dimensional Poisson process introduced above. If on the event $[N_n(1, 1) = n]$ we define $T_1 \leqslant \ldots \leqslant T_n$ to be the first n points of discontinuity of $\{N_n(t) \colon 0 \leqslant t \leqslant 1\}$, it is a straightforward matter to check that conditionally given $[N_n(1, 1) = n]$, the processes

$$\{[N_n(1, T_i)]^{-\frac{1}{2}} [V_n(u, T_i) - uV_n(1, T_i)] \colon 0 \leqslant u \leqslant 1,\ i = 1, 2, \ldots, n\}$$

and $\{W_i(u) \colon 0 \leqslant u \leqslant 1,\ i = 1, 2, \ldots, n\}$ are equivalent. Therefore, since

$$V_n(u, .) - uV_n(1, .) = N_n(u, .) - uN_n(1, .)$$

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press

is constant on the intervals $[T_{i-1}, T_i)$ $(i = 1, 2, \ldots, n+1; T_0 = 0, T_{n+1} = 1)$, for each u, one may write

$$P[M_n > y] = P[\sup_{0 \leqslant u,t \leqslant 1} |V_n(u,t) - uV_n(1,t)|\, n^{-\frac{1}{2}} > y \,|V_n(1,1) = 0]$$
$$\leqslant P[\sup_{0 \leqslant u,t \leqslant 1} |n^{-\frac{1}{2}} V_n(u,t)| > \tfrac{1}{2}y \,|V_n(1,1) = 0]$$
$$+ P[\sup_{0 \leqslant t \leqslant 1} |n^{-\frac{1}{2}} V_n(t)| > \tfrac{1}{2}y \,|V_n(1,1) = 0]. \quad (2{\cdot}11)$$

By the analogy described above between the conditional Poisson process and Uniform order statistics, the second term on the right-hand side of (2·11) equals $P[D_n > \frac{1}{2}y]$, where $D_n = \rho(W_n, 0)$ is Kolmogorov's statistic. In particular, we therefore have

$$\lim_{y \to \infty} \lim_{n \to \infty} P[\sup_{0 \leqslant t \leqslant 1} |n^{-\frac{1}{2}} V_n(t)| > \tfrac{1}{2}y \,|V_n(1,1) = 0] = 0, \quad (2{\cdot}12)$$

in view of the convergence in law of $\{D_n\}$.

The same limiting result is true for the first term on the right side of (2·11) as may be seen as follows. Let $U_1, U_2, \ldots, U_n$ be n independent two-dimensional r.v.'s, each uniformly distributed over the unit square with opposite vertices $(0,0)$ and $(1,1)$. Let $F_n(u,t)$ denote the proportion of these r.v.'s which lie in the square with opposite vertices $(0,0)$ and (u,t). That is, F_n is the empirical distribution function of the sample. Consider the empirical process

$$\{W_n(u,t)\colon 0 \leqslant u, t \leqslant 1\} \quad \text{defined by} \quad W_n(u,t) = n^{\frac{1}{2}}[F_n(u,t) - ut].$$

Then as remarked earlier, this $W_n(.,.)$-process is equivalent to the conditional $n^{-\frac{1}{2}}V_n(.,.)$-process given $[V_n(1,1) = 0]$. Hence the first term on the right-hand side of (2·11) equals

$$P[\sup_{0 \leqslant u,t \leqslant 1} |W_n(u,t)| > \tfrac{1}{2}y].$$

However, Dudley (7) proves that the two-dimensional empirical process

$$\{W_n(u,t)\colon 0 \leqslant u, t \leqslant 1\}$$

converges weakly to a particular stochastic process, $\{W(u,t)\colon 0 \leqslant u, t \leqslant 1\}$ say. This $W(.,.)$-process is equivalent to two-dimensional Brownian Motion tied down at $(1,1)$; that is, a Gaussian process for which

$$E[W(u,t)] = 0 \quad \text{and} \quad E[W(u,t)\, W(r,s)] = \min(u,r)\min(t,s) - rust$$

for all $0 \leqslant r, u, s, t \leqslant 1$. Since the sample paths of the $W(.,.)$-process are continuous a.s. (cf. Lévy (10)) the r.v. $\sup_{0 \leqslant u,t \leqslant 1} |W(u,t|$ is finite a.s. Therefore, because of the a.s. continuity of the supremum functional, the above-mentioned weak convergence implies in particular that

$$\lim_{y \to \infty} \lim_{n \to \infty} P[\sup_{0 \leqslant u,t \leqslant 1} |W_n(u,t)| > \tfrac{1}{2}y] = 0. \quad (2{\cdot}13)$$

Upon applying (2·12) and (2·13) to (2·11) we obtain the desired conclusion

$$\lim_{y \to \infty} \lim_{n \to \infty} P[M_n > y] = 0. \quad (2{\cdot}14)$$

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press

Since the probability which appears in (2·9) is equal to $P[M_n > y]$ with $n = s_t - r_t$ and $y = \delta s_t^{\frac{1}{2}}(s_t - r_t)^{-\frac{1}{2}}$ it is immediate that (2·9) is a consequence of (2·14). This completes the proof that $Q_t^* \xrightarrow{w} P$ on (C, ρ).

In order to prove that $Q_t \xrightarrow{w} P$ on (D, d), one constructs, as in Corollary 2·1, processes $\{\overline{W}, \overline{V}_{t_k}^*: k \geqslant 1\}$ for any divergent sequence $\{t_k\}$ which have the same finite-dimensional d.f.'s as $\{W, V_{t_k}^*: k \geqslant 1\}$ and satisfy the condition, $\rho(\overline{V}_{t_k}^*, W) \xrightarrow{p} 0$. Since by (2·4), $\rho(V_t^*, V_t) \xrightarrow{p} 0$, it follows that $\rho(\overline{V}_{t_k}^*, V_{t_k}) \xrightarrow{p} 0$ and hence $\rho(V_{t_k}, \overline{W}) \xrightarrow{p} 0$. With probability one, $\overline{W} \in C$. Therefore, since d-convergence and ρ-convergence are equivalent on C (Theorem 1 of Appendix 1 in (11)) one obtains $d(V_{t_k}, \overline{W}) \xrightarrow{p} 0$. Since this implies that $Q_{t_k} \xrightarrow{w} P$, for every subsequence $\{t_k\}$, we have proved that $Q_t \xrightarrow{w} P$ on (D, d).

Remark. It is possible to consider the more general problem in which (2·1) is replaced by the condition: $t^{-1}N_t \xrightarrow{p} N$ as $t \to +\infty$ where N is a strictly positive r.v. When N is a discrete r.v. hardly any additional effort is required. However, for arbitrary N, the added effort and notation that is required is such as to justify its omission. For practical purposes (2·1) is all that is needed. The more desirable type of generalization would be to permit more general metrics than d and ρ. In particular, weighted metrics such as $d_w(f, g) = d(f/w, g/w)$ for some given function (e.g. $w(u) = [u(1-u)]^{\frac{1}{2}-\delta}$) are important in applications.

REFERENCES

(1) ALLEN, J. L. and BEEKMAN, J. A. A statistical test involving a random number of random variables. *Ann. Math. Statist.* **37** (1966), 1305–1311.

(2) BARNDORFF-NIELSEN, O. On the limit distribution of the maximum of a random number of independent random variables. *Acta. Math. Acad. Sci. Hungar.* **15** (1964), 399–403.

(3) BILLINGSLEY, PATRICK. Limit theorems for randomly selected partial sums. *Ann. Math. Statist.* **33** (1962), 85–92.

(4) DARLING, D. A. The Kolmogorov–Smirnov, Cramér–von Mises tests. *Ann. Math. Statist.* **28** (1957), 823–838.

(5) DONSKER, M. Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** (1952), 277–281.

(6) DOOB, J. L. *Stochastic processes* (Wiley; New York, 1953).

(7) DUDLEY, R. M. Weak convergence of probabilities on non-separable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** (1966), 109–126.

(8) KAC, M. On deviations between theoretical and empirical distributions. *Proc. Nat. Acad. Sci. U.S.A.* **35** (1949), 252–257.

(9) LAMPERTI, JOHN. On the maximum of a random number of random variables. (1964) (Unpublished).

(10) LÉVY, PAUL. *Processus Stochastiques et Mouvement Brownien* (Gauthier-Villars; Paris, 1948).

(11) PROKHOROV, YU. V. Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** (1956), 157–214.

(12) PYKE, RONALD and SHORACK, GALEN. Weak convergence of the two-sample empirical process and the Chernoff–Savage theorem. (Unpublished.)

(13) SKOROKHOD, A. V. Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** (1956), 261–290.

https://doi.org/10.1017/S0305004100042663 Published online by Cambridge University Press