The Annals of Probability
1983, Vol. 11, No. 3, 745–751

CONVERGENCE OF A CLASS OF EMPIRICAL DISTRIBUTION FUNCTIONS OF DEPENDENT RANDOM VARIABLES[1]

BY B. W. SILVERMAN

University of Bath

A class of empirical processes having the structure of U-statistics is considered. The weak convergence of the processes to a continuous Gaussian process is proved in weighted sup-norm metrics stronger than the uniform topology. As an application, a central limit theorem is derived for a very general class of non-parametric statistics.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random elements of some space S and let h be a function (not necessarily symmetric) from S^m to $\mathbb{R}$. Consider the empirical distribution function H_n constructed from the set of $n(n-1) \cdots (n-m+1)$ random variables $h(X_j, \cdots, X_z)$ obtained by every possible choice of ordered set $(j, \cdots, z)$ of m distinct integers drawn from $\{1, \cdots, n\}$. Suppose that $h(X_1, \cdots, X_m)$ has distribution function H_F.

The main object of this paper is to study the weak convergence of the empirical process $n^{1/2}(H_n - H_F)$ to a Gaussian process, in various weighted metrics on $D[-\infty, \infty]$. The case $m = 1$ corresponds to the standard empirical process of independent random variables, for which weak convergence results in weighted sup-norm metrics have been obtained by Chibisov (1964), O'Reilly (1974) and Shorack (1979). In the case of general m, the function H_n is an empirical distribution function of identically distributed, but dependent, random variables. The weak convergence of $n^{1/2}(H_n - H_F)$ in the unweighted Skorohod topology was considered by Silverman (1976a), who considered a slightly more general class of random variables, to which the results of the present paper carry over directly.

Serfling (1981) has discussed how a very wide range of non-parametric statistics can be expressed as functionals of the empirical process H_n, for suitable choice of the kernel $h(x_1, \cdots, x_m)$. Given a real function J defined on (0, 1), and constants $a_1, \cdots, a_d, p_1, \cdots, p_d$ with p_i in (0, 1), the statistic

$$T_n = T(H_n) \equiv \int_0^1 H_n^{-1}(t) J(t)\, dt + \textstyle\sum_{j=1}^d a_j H_n^{-1}(p_j)$$

is called a *GL-statistic*. For a discussion of the extremely wide scope of GL-statistics and of the interest in considering them as a unified class, see Serfling (1981). In Section 3 below, the result of Section 2 will be used to derive a central limit theorem for GL-statistics under mild conditions, extending the central limit theorems given in Serfling's paper.

We close this section with some technical remarks concerning weak convergence, which may be omitted on a first reading. In contrast to the conventional treatment of empirical processes, the main result below is formulated (following Dudley, 1978) using the supremum metric rather than the Skorohod topology discussed by Billingsley (1968). To avoid measure-theoretic difficulties, we endow $D[-\infty, \infty]$ with the sigma-field (Dudley's $\mathscr{B}_b$) generated by the open spheres in the supremum metric, and we interpret weak convergence in Dudley's sense. A useful introduction to this notion of weak convergence is given by Pollard (1982). Readers who prefer to work in the Skorohod topology should substitute O'Reilly's (1974) d_q for ρ_q below and replace the supremum metric by the Prohorov metric

Received January 1982; revised December 1982.

[1] Research supported by United States Office of Naval Research Contract No. N00014-79-C-0801

AMS 1980 *subject classifications*. Primary 60F17; secondary 62E20, 62G80.

Key words and phrases. Weak convergence, sup-norm metrics, empirical process, dissociated random variables, U-statistics, order statistics, GL-statistics, asymptotic normality.

throughout. The proof then goes through in the same way, at least for H_F the uniform distribution function.

2. The main result. Given a function q on $[-\infty, \infty]$, define a distance ρ_q on $D[-\infty, \infty]$, following O'Reilly (1974), by

$$\rho_q(x, y) = \sup_t | \{x(t) - y(t)\}/q(t) |.$$

The main theorem of this paper is as follows:

THEOREM A. *Suppose $q(s) = q(1 - s)$ for $0 \le s \le ½$, and let $v(t) = t^{1/2}/q(t)$. Assume that*

(1) *q is increasing, continuous and non-negative on* $[0, ½]$;

(2) *v is increasing on* $(0, ½]$ *and $v(t) \to 0$ as $t \to 0^+$*;

(3) $\int_0^{1/2} \{\log(1/t)\}^{1/2}\, dv(t) < \infty$.

Define $\psi(x) = q\{H_F(x)\}$ for all x. Then, defining H_n and H_F as above, $n^{1/2}(H_n - H_F)$ converges weakly in the ρ_ψ metric on $D[-\infty, \infty]$ to a zero-mean Gaussian process W^*. The covariance function of W^* is given in (5) *below, and W^* is continuous at all continuity points of H_F.*

It is an immediate consequence of the tail condition (3) that, as $t \to 0^+$,

$$v(t) = o(\{\log(1/t)\}^{-1/2}),$$

while a sufficient condition for (3) is

$$v(t) = o(\{\log(1/t)\}^{-1/2-\delta}) \quad \text{for some} \quad \delta > 0.$$

Thus the condition (3) is very slightly stronger than the condition

$$v(t) = o(\{\log\log(1/t)\}^{-1/2}) \tag{4}$$

shown by Shorack (1979) to be necessary and sufficient for the i.i.d. case $m = 1$. It would be interesting to know whether the conclusion of Theorem A still holds with (3) replaced by (4); since the i.i.d. case is a special case for any m (set $h(x_1, \cdots, x_m) = x_1$) it will never be possible to improve (3) further than Shorack's condition without imposing additional conditions elsewhere.

PROOF OF THEOREM A. For each t, $H_n(t)$ is a U-statistic, and hence, by Hoeffding (1948), the finite dimensional distributions of $n^{1/2}(H_n - H_F)$ converge to those of a zero-mean Gaussian process W^* with

$$EW^*(s)W^*(t) = \textstyle\sum_J \sum_K P\{h(X_J) \le s, h(X_K) \le t\} - m^2 H_F(s)H_F(t), \tag{5}$$

where the sum is over all $J = (j_1, \cdots, j_m)$ which are cyclic rearrangements of $(1, 2, \cdots, m)$ and all $K = (k_1, \cdots, k_m)$ which are cyclic rearrangements of $(1, m + 1, m + 2, \cdots, 2m - 1)$. The notation $h(X_J)$ is shorthand for $h(X_{j_1}, \cdots, X_{j_m})$.

Thus it only remains to prove tightness. Given any permutation α of $\{1, \cdots, n\}$, follow (5.1.6) of Serfling (1980) and define $H_n^\alpha(t)$ to be the empirical distribution function of the $[n/m]$ (= integer part of n/m) random variables

$$h(X_{\alpha(mj+1)}, X_{\alpha(mj+2)}, \cdots, X_{\alpha(mj+m)})$$

for $j = 0, 1, \cdots, [n/m] - 1$. Write $u(t) = q(t)^{-1}$ here and subsequently and define

$$Z_n(t) = n^{1/2}u(H_F(t))\{H_n(t) - H_F(t)\}$$

and

$$Z_n^\alpha(t) = n^{1/2}u(H_F(t))\{H_n^\alpha(t) - H_F(t)\}.$$

For $0 < y < 1$, define generalized moduli of continuity Ω_n and Ω_n^α by

$$\Omega_n(y) = \sup_{A(y)} |Z_n(s) - Z_n(t)|$$

and

$$\Omega_n^\alpha(y) = \sup_{A(y)} |Z_n^\alpha(s) - Z_n^\alpha(t)|$$

where

$$A(y) = \{s, t: |H_F(s) - H_F(t)| \le y\}.$$

It is immediate that Ω_n is the ordinary modulus of continuity (defined as in (8.1) of Billingsley, 1968) of $Z_n \circ H_F^{-1}$ over the set $H_F(-\infty, \infty)$.

As in (5.1.6) of Serfling (1980), it is easy to see that, summing over all permutations α of $\{1, \cdots, n\}$,

$$Z_n = (n!)^{-1} \sum_\alpha Z_n^\alpha$$

and hence that, for all y,

$$\Omega_n(y) \le (n!)^{-1} \sum_\alpha \Omega_n^\alpha(y). \tag{6}$$

For any r, suppose F_r is the empirical distribution function of r independent random variables uniformly distributed on [0, 1]. Set $Y_r(t) = r^{1/2}u(t)\{F_r(t) - t\}$ and let w_r^Y be the (ordinary) modulus of continuity of Y_r over [0, 1]. The process H_n^α is constructed from $[n/m]$ independent random variables with distribution H_F and therefore the processes $Z_n^\alpha \circ H_F^{-1}$ and $[n/m]^{-1/2}n^{1/2}Y_{[n/m]}$ restricted to the set $H_F(-\infty, \infty)$ have the same distribution. From the definitions of Ω_n^α and w_r^Y it now follows easily that

$$E\Omega_n^\alpha(y) \le [n/m]^{-1/2}n^{1/2}Ew_{[n/m]}^Y(y)$$

with equality if H_F is continuous. Substituting the inequality (6) gives

$$E\Omega_n(y) \le [n/m]^{-1/2}n^{1/2}Ew_{[n/m]}^Y(y). \tag{7}$$

In Proposition 1 below, the asymptotic behavior of w_r^Y will be investigated.

Substituting the result of Proposition 1 into (7) gives

$$\lim_{x\to 0}\limsup_{n\to\infty} E\Omega_n(x) = 0. \tag{8}$$

By using the expectation to give a bound on the tail probability, this implies, given $\varepsilon > 0$, that there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} P\{\Omega_n(\delta) > \varepsilon\} < \varepsilon. \tag{9}$$

Tightness of the family $\{Z_n\}$ now follows by exactly the argument used to prove Theorem (1.2) of Dudley (1978); compare Billingsley's Theorem 15.5 and condition (4.1) of Pollard (1981). The continuity of the limit process at continuity points of H_F follows by the same argument as used in Billingsley's theorem; see also the remarks on page 47 of Pollard (1980) for a discussion of the continuity properties that the limit process will satisfy. (Readers working in the Skorohod topology with uniform H_F can apply Billingsley's theorem directly; condition (15.17) is trivially satisfied.) This argument completes the proof of Theorem A. □

The reason why our proof requires slightly more stringent conditions than those required by Shorack (1979) seems to be that the condition (8), which we establish, is sufficient but not necessary for condition (9), which is what is required for tightness.

It only remains to state and prove Proposition 1.

PROPOSITION 1. *Defining w_n^Y as above, and assuming that q obeys the conditions of Theorem* A,

$$\lim_{x\to 0}\limsup_{n\to\infty} Ew_n^Y(x) = 0.$$

PROOF. Define F_n and Y_n as above, and let w_1 and w_2 be the moduli of continuity of Y_n over [0, ½] and [½, 1] respectively. Since w_1 and w_2 are identically distributed and $w_n^Y \leq w_1 + w_2$, it suffices to prove the proposition with w_n^Y replaced by w_1. As in most papers dealing with $\| \cdot / q \|$ metrics, we shall consider the empirical process separately over intervals $[0, \theta]$ and $[\theta, \tfrac{1}{2}]$, for suitably chosen θ, depending on n. We first deal with the interval $[\theta, \tfrac{1}{2}]$.

Similarly to (7) of Shorack (1979) and (24) of Silverman (1976a), apply Theorem 3 of Komlos, Major and Tusnady (1975) to obtain suitable versions of F_n and the continuous Brownian bridge W^0 such that, for any θ in (0, ½),

$$E \sup_{\theta \leq t \leq 1/2} | Y_n(t) - u(t) W^0(t) | \leq C_0 u(\theta) n^{-1/2} \log n \tag{10}$$

where C_0 is an absolute constant for $n \geq 2$.

This bound is obtained by using Komlos, Major and Tusnady (1975) to give a bound on the tail probabilities of $\sup | Y_n - uW^0 |$. Expressing the expectation as an integral of tail probabilities then gives the result (10); in terms of Komlos, Major and Tusnady's constants, the value obtained for C_0 by performing the integration is $C + K/(\lambda \log 2)$.

The next lemma gives a limiting result for the modulus of continuity of $u(t) W^0(t)$.

LEMMA 1. *Let w_3 be the modulus of continuity of $u(t) W^0(t)$ over* [0, ½]. *Provided $q(t) = u(t)^{-1}$ satisfies the conditions of Theorem* A.

$$E w_3(x) \to 0 \quad as \quad x \to 0.$$

PROOF. The proof rests on a general result about the modulus of continuity of Gaussian processes. Let W be a standard continuous Brownian motion such that

$$u(t) W^0(t) = u(t) W(t) - t u(t) W(1). \tag{11}$$

Let w_3^* and $w_3^\dagger$ be the moduli of continuity over [0, ½] of $u(t) W(t)$ and $tu(t)$ respectively, so that, from (11),

$$E w_3(x) \leq E w_3^*(x) + w_3^\dagger(x) E | W(1) |.$$

By assumption $tu(t) = t^{1/2} v(t)$ is continuous on [0, ½], and hence $w_3^\dagger(x) \to 0$ as $x \to 0$. Thus it suffices to prove that $E w_3^*(x) \to 0$ as $x \to 0$.

Given s and t with $0 \leq s < t \leq \tfrac{1}{2}$, use standard properties of W, and the assumptions (1) and (2) that $u(x)$ is decreasing and $v(x) = x^{1/2} u(x)$ is increasing on [0, ½] to obtain

$$\begin{aligned} E\{u(s) W(s) - u(t) W(t)\}^2 &= s u(s)^2 + t u(t)^2 - 2 s u(s) u(t) \\ &= v(s)^2 + v(t)^2 - 2 s u(s) u(t) \leq 2 v(t)^2 - 2 s u(t)^2 \\ &= 2(t - s) u(t)^2 \leq 2(t - s) u(t - s)^2 \\ &= 2 v(t - s)^2 \leq 2 v(y)^2 \end{aligned}$$

if $t - s \leq y$. By the trivial generalization of Theorems 1 and 2 of Garsia (1972) given as Lemma 2 of Silverman (1976b), w_3^* satisfies

$$w_3^*(x) \leq B_0 v(x) + c_0 \int_0^x \{\log(1/t)\}^{1/2} \, dv(t),$$

where B_0 is a random variable with finite expectation and c_0 is a constant. The proof that $E w_3^*(x)$ tends to zero follows at once from the assumptions in Theorem A. □

The second lemma deals with the behavior of the standard empirical process on $[0, \theta]$.

LEMMA 2. *Defining F_n as the empirical distribution function constructed from n independent standard uniform random variables, and provided $u(x)$ is non-negative and* *decreasing and* $xu(x)$ *is increasing on* $(0, \frac{1}{2})$, *it follows that*

$$E \sup_{0 \le t \le \theta} n^{1/2} |u(t)\{F_n(t) - t\}| \le 2n^{1/2} \int_0^\theta u(t)\, dt$$

for θ *in* $(0, \frac{1}{2})$.

PROOF. Following page 475 of Wellner (1977), for $0 < t < \theta$,

$$|u(t)\{F_n(t) - t\}| \le tu(t) + u(t)F_n(t) \le \theta u(\theta) + \int_0^t u(s)\, dF_n(s)$$

$$\le \int_0^\theta u(s)\, ds + \int_0^\theta u(s)\, dF_n(s).$$

Taking the supremum over t and using the fact that

$$E \int_0^\theta u(s)\, dF_n(s) = \int_0^\theta u(s)\, ds$$

completes the proof. □

It is now possible to complete the proof of Proposition 1. Use (10) and Lemma 2 to write

$$Ew_1(x) \le 2E \sup_{\theta \le t \le 1/2} |Y_n(t) - u(t)W^0(t)| + Ew_3(x) + 2E \sup_{0 \le t \le \theta} |Y_n(t)|$$

$$\le O(n^{-1/2}\log n)u(\theta) + Ew_3(x) + 4n^{1/2} \int_0^\theta u(t)\, dt \tag{12}$$

$$\le Ew_3(x) + O(n^{-1/2}\log n + n^{1/2}\theta) \int_0^1 u(\theta s)\, ds$$

since u is decreasing. For any $\theta \le 1$, it will be the case that

$$\int_0^1 (\theta s)^{-1/2}\log(1/\theta s)^{-1/2}\, ds = \theta^{-1/2}\log(1/\theta)^{-1/2} \int_0^1 s^{-1/2}\{1 + \log(1/s)/\log(1/\theta)\}^{-1/2}\, ds$$

$$\le \theta^{-1/2}\log(1/\theta)^{-1/2} \int_0^1 s^{-1/2}\, ds = 2\theta^{-1/2}\log(1/\theta)^{-1/2}$$

and hence, since $u(t) = o\{t^{-1/2}\log(1/t)^{-1/2}\}$, it follows that

$$\int_0^1 u(\theta s)\, ds = o\{\theta^{-1/2}\log(1/\theta)^{-1/2}\} \quad \text{as} \quad \theta \to 0. \tag{13}$$

Now set $\theta = n^{-1}\log n$, so that $n^{1/2}\theta = n^{-1/2}\log n$ and

$$\theta^{-1/2}\log(1/\theta)^{-1/2} = n^{1/2}(\log n)^{-1/2}(\log n - \log\log n)^{-1/2} \sim n^{1/2}(\log n)^{-1}.$$

Substituting these results and (13) into (12) then gives

$$Ew_1(x) \le Ew_3(x) + o(1) \quad \text{as} \quad n \to \infty,$$

so that, for each x,

$$\limsup_{n\to\infty} Ew_1(x) \le Ew_3(x).$$

Now apply Lemma 1 to complete the proof of Proposition 1.

Notice that, in the case of constant q, it is possible to put $\theta = 0$ and deduce the result direct from (8) and Lemma 1. This gives a simpler proof of Theorem B of Silverman (1976a).

It can turn out in certain special cases that the covariance in (5) is identically zero and that the U-statistic H_n is, for all t, of the form discussed in Section 5.5.2 of Serfling (1980). Theorem A will then give a degenerate limit. An example of such behavior is given in Silverman (1978), where it is shown, for the case $m = 2$ and q constant, that $n(H_n - H_F)$ is tight. The extension of this result to general orders m and weight functions q is a subject for possible future investigation.

3. An application to *GL*-statistics. The GL-statistics defined by Serfling (1981) and in the introduction above are a generalization of L-statistics as discussed in Chapter 8 of Serfling (1980). In his 1981 paper, Serfling provides generalizations to GL-statistics of the asymptotic normality results given in Theorems 8.2.4A and 8.2.4C of his 1980 book, but no analog of his Theorem 8.2.4B. The following theorem fills this gap.

THEOREM B. *Suppose the GL-statistic T_n is as defined in the introduction above, and that the following conditions are satisfied*:

(i) *J is bounded and continuous* a.e. *Lebesgue and* a.e. H_F^{-1};

(ii) *H_F has positive derivatives at its p_j-quantiles, for* $1 \le j \le d$;

(iii) *for some q satisfying the assumptions of Theorem* A,

$$\int_{-\infty}^{\infty} q(H_F(x))\,dx < \infty.$$

Then

$$n^{1/2}(T_n - T(H_F)) \to N(0, \sigma^2(T, H_F))$$

in distribution, where $\sigma^2(T, H_F)$ is as defined in (3.3) *of Serfling* (1981).

To prove the theorem, extend the proof of Serfling's (1981) Theorem 3.1 in exactly the same way as he proves his (1980) Theorem 8.2.4B; instead of using the result of O'Reilly (1974) to deduce Serfling's (1980) Lemma 8.2.4C, Theorem A above is used to deduce that $\sup|(H_n - H_F)/q \circ H_F| = O_p(n^{-1/2})$.

The remaining details of the proof are omitted.

Acknowledgments. It is with great pleasure that the author acknowledges the helpful suggestions and encouragement of Robert Serfling, and the kind hospitality of the Department of Mathematical Sciences, The Johns Hopkins University. The author is most grateful to a referee for urging him to work in the supremum norm; a conversation with Adrian Baddeley on this subject was most valuable.

REFERENCES

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

CHIBISOV, D. M. (1964). Some theorems on the limiting behavior of empirical distribution functions. *Selected Translations in Math. Statist. and Probab.* **6** 147–156.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

GARSIA, A. M. (1972). Continuity properties of Gaussian processes with multidimensional time parameter. *Proc. 6th Berk. Symp. Math. Statist. Probab. II*, 369–374.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

KOMLOS, J., MAJOR, P., and TUSNADY, G. (1975). An approximation of partial sums of independent random variables, and the sample distribution function. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

O'REILLY, N. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probab.* **2** 642–651.

POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70.

POLLARD, D. (1981). Limit theorems for empirical processes. *Z. Wahrsch. verw. Gebiete* **57** 181–195.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SERFLING, R. J. (1981). Generalized *L*-, *M*- and *R*-statistics. Mathematical Sciences Technical Report 349, Johns Hopkins University, Baltimore.

SHORACK, G. R. (1979). Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. *Stochastic Process. Appl.* **9** 95–98.

SILVERMAN, B. W. (1976a). Limit theorems for dissociated random variables. *Adv. in Appl. Probab.* **8** 806–819.

SILVERMAN, B. W. (1976b). On a Gaussian process related to multivariate probability density estimation. *Math Proc. Cambridge Philos. Soc.* **80** 135–144.

SILVERMAN, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probab.* **15** 136–143.

WELLNER, J. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BATH
BATH BA2 7AY
UNITED KINGDOM