translation parameter θ, and make use of the Kolmogorov-type statistic $D_n(g, P)$ based on the shell function $g(x) = (1 - |x|^2)^+$. Let $\mathscr{P} = \{P^\theta: \theta \in \mathbb{R}^2\}$ be the translation family of probability measures defined by $P^\theta(A) = P(A - \theta)$. Suppose one wishes to estimate θ by the minimum distance estimator, $\hat{\theta}_n$, defined as that value of $\theta \in \mathbb{R}^d$ which minimizes the distance

$$D_n(g, P^\theta) = \sup_t | P_n g(\cdot - t) - Pg(\cdot - t - \theta) |.$$

Under the assumption that the true parameter is $\theta_0 = 0$, it appears that the asymptotic distribution of $\hat{\theta}_n$ may be the same as that for Pollard's estimate, $\hat{\tau}_n$, the value of t at which $P_n g(\cdot - t)$ is maximized, even though the minimization problems are different. Let me offer as a third test of the author's methodology the question of determining the limiting distribution of $n^{1/2}\hat{\theta}_n$. This type of problem is similar to one considered by Blackman (1955), except that he used a Cramér-von Mises distance rather than a Kolmogorov one; in Pyke (1970) this simpler problem was used to illustrate the applicability of the "weak implies strong" methodology mentioned above.

Although I have directed my comments on the paper towards statisticians as users of this theory, I would stress that the paper is also of great value to those doing research in the area. From both viewpoints I greatly appreciate the efforts of David Pollard for preparing this valuable exposition.

ADDITIONAL REFERENCES

BERAN, R. and MILLER, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14** 431–443.

BLACKMAN, J. (1955). On the approximation of a distribution function by an empiric distribution. *Ann. Math. Statist.* **26** 256–267.

CRAMÉR H. (1928). On the composition of elementary errors. II. *Skand. Aktuarietidskr.* **11** 141–180.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

KUELBS, J. and DUDLEY, R. M. (1980). Loglog laws for empirical measures. *Ann. Probab.* **8** 405–418.

PYKE, R. (1970). Asymptotic results for rank statistics. In *Proc. First Symp. on Non-Parametric Techniques* (M. Puri, ed.) 21–40. Cambridge Univ. Press, Cambridge.

PYKE, R. and SHORACK, G. (1968). Weak convergence of a two sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

PYKE, R. and WILBOUR, D. C. (1988). New approaches for goodness-of-fit tests for multidimensional data. In *Statistical Theory and Data Analysis II* (K. Matusita, ed.) 139–154. North-Holland, Amsterdam.

SAPOGOV, N. A. (1974). A uniqueness problem for finite measures in Euclidean spaces. Problems in the theory of probability distributions. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **41** 3–13. (In Russian.)

STEELE, J. M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probab.* **6** 118–127.

Comment

Miklós Csörgő and Lajos Horváth

It is a pleasure to congratulate David Pollard on his masterly glimpse into the theory of empirical processes. His artful development here of the technique of Gaussian symmetrization, of the resulting maximal inequalities for Gaussian processes and their application in the empirical process context leaves us no room for comment on his methods, which extend the concept of a Vapnik-Červonenkis class of sets. He demonstrates the efficiency of these methods by use of two motivating, nontrivial asymptotic problems and succeeds very well in conveying the look and feel of a powerful tool of contemporary mathematical statistics.

Miklós Csörgő is Professor of Mathematics and Statistics at Carleton University. His address is Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6. Lajos Horváth is Associate Professor, Department of Mathematics, University of Utah, Salt Lake City, Utah 84112.

There are also other powerful contemporary tools available for tackling asymptotic problems of mathematical statistics. The ones we have in mind are strong and weak approximations (almost sure and in probability invariance principles) for empirical and partial sum processes based on various forms of the Skorohod embedding scheme, or on various forms of the Hungarian construction. The quoted book of Shorack and Wellner (1986) is also an excellent source of information on these methods. For further references on the methods and their applications, we mention the books of Csörgő and Révész (1981), Csörgő (1983), and Csörgő, Csörgő and Horváth [CsCsH] (1986). For an insightful overview of strong and weak approximations we refer to Philipp (1986) (cf. also the review of Csörgő (1984)). Concerning Hungarian constructions, for those who are really interested, the papers of Bretagnolle and Massart (1989), and Einmahl (1989) are most recommended readings.

Here *we* make use of the first problem discussed by David Pollard to *illustrate* what we mean by *strong* and *weak approximation methods* and by their direct, straightforward application to this problem. Let $X_1, X_2, \cdots$ be independent, identically distributed random variables with distribution function F on the real line. Let

$$F_n(x) = n^{-1}\#\{1 \le i \le n: X_i \le x\}$$

and

$$\alpha_n(x) = n^{1/2}(F_n(x) - F(x)), \quad x \in \mathbf{R},$$

be the *empirical distribution function* and the *empirical process*, respectively, of the first n of these random variables. Let us first take it for granted that we have constructed a sequence of continuous Gaussian processes $\{B_n(t), 0 \le t \le 1\}$ on a probability space $(\Omega, \mathscr{A}, P)$ on which $X_1, X_2, \cdots$, and B_n live together so that $P\{X_1 \le x\} = F(x)$, B_n is a Brownian bridge for each $n = 1, 2 \cdots$, i.e., a real valued, mean zero Gaussian process with covariance function $EB_n(s)B_n(t) = \min(s, t) - st$ (thus, for each fixed $t \in (0, 1)$, $B_n(t)$ is a $N(0, t(1-t))$ random variable) and, as $n \to \infty$, we have

$$\sup_{-\infty<x<\infty} |\alpha_n(x) - B_n(F(x))| = o_P(1), \tag{1}$$

and, on assuming that $EX_1^2 = \int_{-\infty}^{\infty} x^2\, dF(x) < \infty$, we have also

$$\int_{-\infty}^{\infty} |\alpha_n(x) - B_n(F(x))|\, dx = o_P(1). \tag{2}$$

The statements (1) and (2) are examples of what we mean by *weak approximation* (*invariance principle in probability*). They are conceptually simpler than the notion of weak convergence (functional central limit theorem). Also, (1) actually is a stronger form of Donsker's theorem; it implies the corresponding classical Donsker functional theorem. From the point of view of this remark, which may very well coincide with that of the statistician in general, it is irrelevant how exactly the construction leading up to (1) and (2) is carried out. For details we may refer to Chapters 2 and 3 of CsCsH (1986), and for (2) in particular to Lemma 3.2 of the latter monograph. Here we will simply use them as if one were using a central limit theorem, only more directly however, as building blocks in the process of obtaining our weak approximations.

As in the Pollard exposition, for $t \in \mathbf{R}$ we let

$$G_n(t) = n^{-1} \sum_{i=1}^{n} |X_i - t| = \int_{-\infty}^{\infty} |x-t|\, dF_n(x),$$

$$G(t) = \int_{-\infty}^{\infty} |x-t|\, dF(x)$$

and define, what he calls *the standardized difference*, β_n by

$$\beta_n(t) = n^{1/2}(G_n(t) - G(t)) = \int_{-\infty}^{\infty} |x-t|\, d\alpha_n(x).$$

The statistic $G_n(\bar{X}_n) = n^{-1}\sum_{i=1}^{n} |X_i - \bar{X}_n|$ is the first of the two problems discussed in Section 2 of Pollard's work. Going at it from the weak approximation point of view, the definition of β_n and (1) immediately suggest that for large n the process $\beta_n(t)$ in $t \in \mathbf{R}$ should be close to the sequence of Gaussian processes

$$\Gamma_n(t) = \int_{-\infty}^{\infty} |x-t|\, dB_n(F(x)), \quad t \in \mathbf{R},$$

which have the same distribution for each n. Indeed, we have the following simple invariance principles in probability. With all due respect to many in statistics who cannot stand theorem-proof like presentations, we have found it most economical to summarize our view exactly that way.

Proposition 1. If $EX_1^2 < \infty$, then as $n \to \infty$ we have

$$\sup_{-\infty<t<\infty} |\beta_n(t) - \Gamma_n(t)| = o_P(1). \tag{3}$$

Proof. Integrating by parts, using the assumption $EX_1^2 < \infty$, we have, writing $\sup_t$ for $\sup_{-\infty<t<\infty}$,

$$\begin{aligned}
&\sup_t |\beta_n(t) - \Gamma_n(t)| \\
&= \sup_t \left| \int_{-\infty}^{\infty} |x-t|\, d(\alpha_n(x) - B_n(F(x))) \right| \\
&= \sup_t \left| \int_{-\infty}^{\infty} (\alpha_n(x) - B_n(F(x)))\, d|x-t| \right| \\
&= \sup_t \left| \int_{-\infty}^{\infty} (\alpha_n(u+t) - B_n(F(u+t)))\, d|u| \right| \\
&= \sup_t \left| -\int_{-\infty}^{0} (\alpha_n(u+t) - B_n(F(u+t)))\, du \right. \\
&\qquad \left. + \int_{0}^{\infty} (\alpha_n(u+t) - B_n(F(u+t)))\, du \right| \\
&\le \sup_t \left\{ \int_{-\infty}^{t} |\alpha_n(x) - B_n(F(x))|\, dx \right. \\
&\qquad \left. + \int_{t}^{\infty} |\alpha_n(x) - B_n(F(x))|\, dx \right\} \\
&= \int_{-\infty}^{\infty} |\alpha_n(x) - B_n(F(x))|\, dx = o_P(1),
\end{aligned}$$

on account of (2), where the equality obtained by integrating by parts holds with probability one for each n.

The result of (3) for each fixed t rhymes with Pollard saying that if F has a finite variance, the standardized difference β_n is asymptotically normal, for each *fixed* t, namely

$$\beta_n(t) \xrightarrow{\mathscr{D}} \Gamma(t) := \int_{-\infty}^{\infty} |x-t| \, dB(F(x)), \tag{4}$$

where B is a Brownian bridge. Of course (3) implies also a weak convergence version of the latter convergence in distribution result.

We can write $\bar{X}_n = \int_{-\infty}^{\infty} x \, dF_n(x)$, and similarly to Proposition 1, (2) implies

$$\left| n^{1/2}(\bar{X}_n - \mu) - \int_{-\infty}^{\infty} x \, dB_n(F(x)) \right| = o_P(1), \tag{5}$$

where $\mu = EX_1 = \int_{-\infty}^{\infty} x \, dF(x)$.

One of the points made in Pollard's work is, of course, the asymptotic normality of $n^{1/2}(G_n(\bar{X}_n) - G(\mu))$. In terms of our weak approximation language, the latter reads and easily established as follows.

Proposition 2. If $EX_1^2 < \infty$ and F is continuous in a neighborhood of μ, then as $n \to \infty$ we have

$$\left| n^{1/2}(G_n(\bar{X}_n) - G(\mu)) - \left\{ \Gamma_n(\mu) + (2F(\mu) - 1) \int_{-\infty}^{\infty} x \, dB_n(F(x)) \right\} \right| = o_P(1), \tag{6}$$

and hence also

$$n^{1/2}(G_n(\bar{X}_n) - G(\mu)) \xrightarrow{\mathscr{D}} \Gamma(\mu) + (2F(\mu) - 1) \int_{-\infty}^{\infty} x \, dB(F(x)). \tag{7}$$

Proof. An elementary calculation yields

$$G'(t) = 2F(t) - 1. \tag{8}$$

We write

$$n^{1/2}(G_n(\bar{X}_n) - G(\mu)) = \beta_n(\bar{X}_n) + n^{1/2}(G(\bar{X}_n) - G(\mu)). \tag{9}$$

It is easy to see that $\Gamma_n(t)$ is almost surely continuous at μ for each n (cf. (17)), and therefore (3) and (5) yield

$$\begin{aligned} &|\beta_n(\bar{X}_n) - \Gamma_n(\mu)| \\ &\le |\beta_n(\bar{X}_n) - \Gamma_n(\bar{X}_n)| + |\Gamma_n(\bar{X}_n) - \Gamma_n(\mu)| \\ &= o_P(1), \end{aligned} \tag{10}$$

i.e., $\beta_n(\bar{X}_n) = n^{1/2}(G_n(\bar{X}_n) - G(\bar{X}_n))$ has the same limiting normal distribution as $\beta_n(\mu) = n^{1/2}(G_n(\mu) - G(\mu))$, namely $\beta_n(\bar{X}_n) \xrightarrow{\mathscr{D}} \Gamma(\mu)$. Now the mean value theorem, the assumed continuity of F around μ, (5), and (8) result in

$$n^{1/2}(G(\bar{X}_n) - G(\mu)) - (2F(\mu) - 1) \int_{-\infty}^{\infty} x \, dB_n(F(x)) = o_P(1). \tag{11}$$

A combination of (9), (10), and (11) implies the result in (6).

We note that (11) spells out exactly what the contribution of $n^{1/2}(\bar{X}_n - \mu)$ is to the limiting distribution of $n^{1/2}(G_n(\bar{X}_n) - G(\mu))$ in (6). This asymptotic contribution of $n^{1/2}(\bar{X}_n - \mu)$ vanishes if $F(\mu) = ½$, i.e., if μ were also a median of F, for then $2F(\mu) - 1 = 0$. This brings us to the natural proposition of replacing $\bar{X}_n$ by the sample median m_n, or by any other consistent sequence of estimators of a *population median* m of F. This is also mentioned of course in Pollard's work, who notes also that the choice of m_n for the centering leads to another measure of spread, $\inf_t G_n(t)$, for the sample. In this particular example the solution is easy. Indeed, arguing as in the proof of Proposition 2 we have the next, obvious from the weak approximations point of view, result.

Proposition 3. If $EX_1^2 < \infty$ and F is continuous in a neighborhood of m and $m_n - m = o_P(1)$, i.e., m_n is a weakly consistent sequence of estimators for m, then as $n \to \infty$ we have

$$|n^{1/2}(G_n(m_n) - G(m)) - \Gamma_n(m)| = o_P(1), \tag{12}$$

and hence also

$$n^{1/2}(G_n(m_n) - G(m)) \xrightarrow{\mathscr{D}} \Gamma(m). \tag{13}$$

Next, in addition to (1) and (2), let us take it for granted that, on an appropriate probability space, as $n \to \infty$, we have already established

$$\sup_{-\infty<x<\infty} \frac{|\alpha_n(x) - B_n(F(x))|}{(F(x)(1 - F(x)))^{1/2-\nu}} = \begin{cases} O_P(n^{-1/2} \log n), & \text{if } \nu = ½, \\ O_P(n^{-\nu}), & \text{if } 0 < \nu < ½. \end{cases} \tag{14}$$

This is Lemma 7 in Csörgő and Horváth (1988a), and it is based on earlier versions in Csörgő, Csörgő, Horváth and Mason (1986), Csörgő and Horváth (1986) and Mason and van Zwet (1987). Given (14) and some slight conditions on F, we can restate our *invariance principles in probability so far with rates of convergence* attached this time around, and watch how new conditions present themselves in a most natural way for the job at hand.

*Proposition 1**. If

$$J\left(\frac{2}{1-2\nu}\right) := \int_{-\infty}^{\infty} (F(x)(1 - F(x)))^{1/2-\nu} dx < \infty$$

for some $\nu \in (0, \frac{1}{2})$, then as $n \to \infty$

(15) $$\sup_{-\infty<t<\infty} |\beta_n(t) - \Gamma_n(t)| = O_P(n^{-\nu}).$$

Proof. First we note that $J(2/(1-2\nu)) < \infty$ implies $E|X_1|^{2/(1-2\nu)} < \infty$ for some $\nu \in (0, \frac{1}{2})$, a stronger moment condition than that of Proposition 1. This follows from extending the discussion in the Appendix of Hoeffding (1973). Hence $EX_1^2 < \infty$ and, on integrating by parts we get, as in the proof of Proposition 1,

$$\begin{aligned}\sup_t |\beta_n(t) - \Gamma_n(t)| &= \sup_t \left| \int_{-\infty}^{\infty} (\alpha_n(x) - B_n(F(x)))\, d|x-t| \right| \\ &\le \int_{-\infty}^{\infty} |\alpha_n(x) - B_n(F(x))|\, dx \\ &\le \sup_x \frac{|\alpha_n(x) - B_n(F(x))|}{(F(x)(1-F(x)))^{1/2-\nu}} \\ &\quad \cdot \int_{-\infty}^{\infty} (F(x)(1-F(x)))^{1/2-\nu}\, dx,\end{aligned}$$

and hence (14) implies (15).

We note that by the Appendix of Hoeffding (1973) it is easy to give a sufficient moment condition for the finiteness of $J(2/(1-2\nu))$ of Proposition 1*. For example, if

$$E\{|X_1|^{2/(1-2\nu)}(\log(1+|X_1|))^2\} < \infty,$$

then $J(2/(1-2\nu)) < \infty$ (cf. also Section 3 of CsCsH (1986) for related material).

*Proposition 2**. If $J(2/(1-2\nu)) < \infty$ for some $\nu \in (0, \frac{1}{2})$ and F possesses a bounded density f in a neighborhood of μ, then as $n \to \infty$

(16) $$\left| n^{1/2}(G_n(\bar{X}_n) - G(\mu)) - \left\{\Gamma_n(\mu) + (2F(\mu)-1)\int_{-\infty}^{\infty} x\, dB_n(F(x))\right\} \right| = O_P(n^{-\nu}).$$

Proof. First, along the lines of the proof of Proposition 1, we observe that we have with probability one for each n and all t

$$\Gamma_n(t) = \int_{-\infty}^{t} B_n(F(x))\, dx - \int_{t}^{\infty} B_n(F(x))\, dx.$$

Hence we have with probability one for each n and all s, t

(17) $$|\Gamma_n(t) - \Gamma_n(s)| \le 2|t-s| \sup_x |B_n(F(x))|,$$

and note also that, instead of (5), we now have

(18) $$\left| n^{1/2}(\bar{X}_n - \mu) - \int_{-\infty}^{\infty} x\, dB_n(F(x)) \right| = O_P(n^{-\nu}), \quad \nu \in (0, \tfrac{1}{2}),$$

similarly to Proposition 1* by (14). With an eye on (9), from (15), (17) and (18) we conclude

(19) $$\begin{aligned}|\beta_n(\bar{X}_n) - \Gamma_n(\mu)| &\le |\beta_n(\bar{X}_n) - \Gamma_n(\bar{X}_n)| + |\Gamma_n(\bar{X}_n) - \Gamma_n(\mu)| \\ &\le |\beta_n(\bar{X}_n) - \Gamma_n(\bar{X}_n)| \\ &\quad + 2|\bar{X}_n - \mu| \sup_x |B_n(F(x))| \\ &= O_P(n^{-\nu}) + O_P(n^{-1/2}) = O_P(n^{-\nu}),\end{aligned}$$

while a two-term Taylor expansion gives

(20) $$\begin{aligned}&n^{1/2}(G(\bar{X}_n) - G(\mu)) \\ &\quad = (2F(\mu)-1)n^{1/2}(\bar{X}_n - \mu) + f(\xi_n)n^{1/2}(\bar{X}_n - \mu)^2,\end{aligned}$$

where $\min(\bar{X}_n, \mu) \le \xi_n \le \max(\bar{X}_n, \mu)$. Hence by (18) and the assumed boundedness of f around μ we get

(21) $$\begin{aligned}&\left| n^{1/2}(G(\bar{X}_n) - G(\mu)) - (2F(\mu)-1)\int_{-\infty}^{\infty} x\, dB_n(F(x)) \right| \\ &\quad = O_P(n^{-\nu}) + O_P(n^{-1/2})) \\ &\quad = O_P(n^{-\nu}).\end{aligned}$$

On account of (9), (19) and (21) we now have also (16).

For the sake of a similar version of Proposition 3 we estimate the median m of F by the sample median

$$m_n := \inf\{x: F_n(x) \ge \tfrac{1}{2}\}.$$

*Proposition 3**. If $J(2/(1-2\nu)) < \infty$ for some $\nu \in (0, \frac{1}{2})$, and F possesses a density f in a neighborhood of m and f is positive and continuous at m, then

(22) $$|n^{1/2}(G_n(m_n) - G(m)) - \Gamma_n(m)| = O_P(n^{-\nu}).$$

Proof. It is well known (cf., e.g., Csörgő (1983, Section 1.5) or Serfling (1980, Section 2.3.3)) that under the given conditions on f we have, as $n \to \infty$,

(23) $$n^{1/2}(m_n - m) \xrightarrow{\mathscr{D}} N(0, 1/(4f^2(m))).$$

As in (9)

(24) $$\begin{aligned}&n^{1/2}(G_n(m_n) - G(m)) \\ &\quad = \beta_n(m_n) + n^{1/2}(G(m_n) - G(m)),\end{aligned}$$

and a two-term Taylor expansion gives (compare with (20))

(25) $n^{1/2}(G(m_n) - G(m)) = f(\eta_n)n^{1/2}(m_n - m)^2$,

where $\min(m_n, m) \le \eta_n \le \max(m_n, m)$. By (15), (17) and (23) we get, as in (19),

(26) $|\beta_n(m_n) - \Gamma_n(m)| = O_P(n^{-\nu})$,

while (23) and (25) combined yield

(27) $n^{1/2}(G(m_n) - G(m)) = O_P(n^{-1/2})$.

Now (24), (26) and (27) result in (22).

This also concludes what we wish to say about quick and easy applications of weak approximation methods (invariance principles in probability) to the first problem discussed by Pollard. There are of course similarities between the two approaches taken. Pollard too uses approximation arguments, but to make them precise he puts the onus on probabilistic bounds on the oscillations of the empirical process in shrinking neighborhoods of a point (like μ or m above), while we work with approximating the whole empirical process instead (as in (1), (2) and (14)), and then use the approximating Gaussian sequences as building blocks in the process of replacing the empirical parts by Gaussian ones and thus piecing together asymptotic representations (identifications in the limit) for the sample processes at hand.

As to the nature of the results of the above propositions, they are similar to those obtained for the empirical process with parameters estimated (cf. Burke, Csörgő, Csörgő and Révész, 1979; Durbin, 1973a, b) in that they are also asymptotically distribution dependent. Hence computations for the desired asymptotic distribution functions are difficult to come by. These results, however, can be bootstrapped by resampling the data. For tools of bootstrapping empirical functionals, we refer to Bickel and Freedman (1981), CsCsH (1986, Chapter 17), and for a successful execution of bootstrapping the empirical process when the underlying parameters are estimated we refer to Burke and Gombay (1988).

We have also promised *to illustrate strong approximation methods* on the above discussed first problem of Pollard. Let $\{K(y, t); 0 \le y \le 1, 0 \le t < \infty\}$ be a Kiefer process, that is, a real valued, mean zero, two-parameter Gaussian process, with covariance function $EK(y, t)K(u, s) = \min(t, s)(\min(y, u) - yu)$ (thus in t, $K(y, t)$ is like a Brownian motion, and it is like a Brownian bridge in y), which we accept to have been constructed for α_n on an appropriate probability space so that, as $n \to \infty$,

(28)
$$\sup_{-\infty<x<\infty} |\alpha_n(x) - n^{-1/2}K(F(x), n)| \overset{\text{a.s.}}{=} O((\log n)^2/n^{1/2}).$$

This is an example of *strong approximation* (*strong invariance principle*) and a famous one at that (cf. Komlós, Major and Tusnády (1975) for the original result, or Theorem 4.4.3 in Csörgő and Révész (1981)). The empirical process $\alpha_n(x)$ is approximated almost surely (a.s.) as a two-parameter process in x and n by a *single* two-parameter Gaussian process $K(F(x), n)$. In addition to weak laws, via (28) $\alpha_n(\cdot)$ inherits also strong laws, like for example the law of the iterated logarithm (LIL), from $K(\cdot, \cdot)$. Here, using (28) along the lines of the proofs of Propositions 2* and 3*, combined with appropriate LIL laws like Theorem 1.15.1 of Csörgő and Révész (1981), under the respective conditions of Propositions 2* and 3* one easily obtains immediate respective LIL laws as follows:

$$0 < \limsup_{n\to\infty} \left(\frac{n}{\log\log n}\right)^{1/2} |G_n(\bar{X}_n) - G(\mu)| < \infty \quad \text{a.s.}$$

and

$$0 < \limsup_{n\to\infty} \left(\frac{n}{\log\log n}\right)^{1/2} |G_n(m_n) - G(m)| < \infty \quad \text{a.s.}$$

For multivariate generalizations of (1) and (28) with rates, and with an arbitrary distribution function F on $\mathbf{R}^d$, $d \ge 2$, we refer to Philipp and Pinzur (1980), Borisov (1982), and Csörgő and Horváth (1988b).

ADDITIONAL REFERENCES

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

Borisov, I. S. (1982). Approximation of empirical fields, constructed with respect to vector observations with dependent components. *Sibirsk. Mat. Zh.* **23** 31–41.

Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256.

Burke, M. D., Csörgő, M., Csörgő, S. and Révész, P. (1979). Approximations of the empirical process when parameters are estimated. *Ann. Probab.* **7** 790–810.

Burke, M. D. and Gombay, E. (1988). On goodness-of-fit and the bootstrap. *Statist. Probab. Lett.* **6** 287–293.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* SIAM, Philadelphia.

Csörgő, M. (1984). Invariance principles for empirical processes. In *Handbook of Statistics 4, Nonparametric Methods* (P. R. Krishnaiah and P. K. Sen, eds.) 431–462. North-Holland, Amsterdam.

Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes. Lecture Notes in Statist.* **33**. Springer, New York.

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

Csörgő, M. and Horváth, L. (1986). Approximations of weighted empirical and quantile processes. *Statist. Probab. Lett.* **4** 275–280.

Csörgő, M. and Horváth, L. (1988a). Central limit theorems for L_p-norms of density estimators. *Probab. Theory Related Fields* **80** 269–291.

Csörgő, M. and Horváth, L. (1988b). A note on strong approximations of multivariate empirical processes. *Stochastic Process. Appl.* **28** 101–109.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Akadémiai Kiadó, Budapest/Academic, New York.

Durbin, J. (1973a). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

Durbin, J. (1973b). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

Einmahl, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68.

Hoeffding, W. (1973). On the centering of a simple linear rank statistic. *Ann. Statist.* **1** 54–66.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

Mason, D. M. and van Zwet, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab.* **17** 266–291.

Philipp, W. (1986). Invariance principles for independent and weakly dependent random variables. In *Dependence in Probability and Statistics. A Survey of Recent Results* (E. Eberlein and M. S. Taqqu, eds.). *Progress in Probability and Statistics* **11** 227–268. Birkhäuser, Boston.

Philipp, W. and Pinzur, L. (1980). Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. verw. Gebiete* **54** 1–13.

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

Rejoinder

David Pollard

I find myself in the position of a man who has just pointed out how one can balance a checkbook using a high-powered graphics workstation. Professor Dudley responds by suggesting some further applications in the same spirit. Professors Giné and Zinn point out that one can also use the machine for high speed interactive graphics. Professor Pyke mentions other uses more suited for a piece of high technology, while suggesting (perhaps tongue in cheek) that my particular checkbook might also be balanced using a hand-held calculator. Professors Csörgő and Horváth demonstrate that their super parallel processor can also balance checkbooks.

In large part I agree with, and welcome, the comments of this distinguished group of discussants. But to maintain the correct atmosphere of contrariness and provocation, I will find some way to disagree with all of them.

Professor Dudley suggests that Fréchet differentiability, with the right choice of norm, should be used in preference to compact differentiability. As he has convincingly argued in his 1989 preprint, this new viewpoint does free Fréchet differentiability from the uncomfortable constraint of distribution functions on the real line. However, compact differentiability (with derivative Δ_x) of a functional T is enough to imply

$$\sqrt{n}\,[T(x + z_n/\sqrt{n}) - T(x)] = \Delta_x \cdot z_n + o(1)$$

for each convergent sequence $\{z_n\}$, a property that is ideally suited to application of Dudley's (1985) almost uniform representation theorem. Gill (1987) has explored this aspect of compact differentiability.

Dudley also suggests substitution of the smooth convex $\rho(x)$ for $|x|$, to eliminate the problems caused by nondifferentiability of $|x|$ at the origin. As a device to simplify the asymptotic theory this is unnecessary (Pollard 1989a); Tchebychev's inequality, the CLT for bounded (vector-valued) summands, and an elementary convexity argument can handle the estimator, even for $c = 0$.

Professors Giné and Zinn quite properly point out some of the beautiful general theory—in particular, the work of Talagrand—that I failed to mention. I feel that conditions expressed in terms of limiting Gaussian processes will not appeal to many potential users of empirical process theory, even though there are excellent theoretical reasons for preferring their approach. At this stage in the history of the world, I feel it is more important that potential users be enticed by small examples of empirical process ideas rather than be impressed and intimidated by the full force and elegance of the latest theory. Times will change. More papers along the lines of Giné and Zinn (1988) will convince us all that sample path properties of abstract Gaussian processes are relevant, even for popular topics such as the bootstrap.

Jain and Marcus (1975, inequality 2.30) did use the idea of dominating a process involving Rademachers by a related Gaussian process, but Giné and Zinn are right concerning the role of the inequality in the