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 reverse direction and the role that Gaussian symmetrization plays in the modern theory.

Professor Pyke seems to regret that I omitted the full statement of the CLT for the empirical process ν_n. That was one of the topics sacrificed in order to simplify the general presentation. My experience has been that very often one does not need the full force of a CLT. The approximation property represented by stochastic equicontinuity (or uniform tightness), which is the main ingredient in the empirical CLT, is often all that one needs. My Theorem 4.7 (when applied to classes of differences of functions from a fixed $\mathscr{F}$) can be reinterpreted as an assertion of stochastic equicontinuity; it is a much streamlined form of the argument I used to prove my 1982 empirical CLT.

One could handle the applications by setting up a formal empirical CLT as a functional limit theorem for stochastic processes (interpreted in the Hoffmann-Jørgensen sense mentioned by Dudley). One could then appeal, for example, to Dudley's almost uniform representation to approximate a version of ν_n by a version of the Gaussian limit process. Then the uniform approximation arguments in the illustrative examples would be replaced by continuity arguments for the sample paths of the Gaussian process. This approach was discussed in more detail in Pollard (1989b).

Pyke recognizes the intent of my second example to illustrate how off-the-shelf empirical process methods make short work of a typical sort of multidimensional estimation problem. Nevertheless he can't resist the temptation of trying to handle the same example using more traditional methods. I approve fully, since my instincts also push me towards the method of minimum machinery. I would suggest, however, that the contribution from the annulus B_3 could prove troublesome when one tries to establish bounds uniformly over a range of vector-valued parameters b.

To find the asymptotic distribution of the $\hat{\theta}_n$ that minimizes Pyke's $D_n(g, P_\theta)$ one can use empirical process methods (Pollard, 1980, Theorem 7.2). I do not think that it has the same normal limit distribution as $\hat{\tau}_n$.

Professors Csörgő and Horváth advertise an approximation technology that has much to recommend it. As I have already noted in my response to Pyke, the empirical process oscillation argument in my paper can be recast into the form of an almost uniform representation of an abstract empirical CLT. The paragraph following their equation (27) summarizes only the particular approach that I took in this particular paper; it is not a complete description of the abstract empirical process theory that has grown from Dudley's 1978 paper.

I regret that Csörgő and Horváth chose to illustrate their approximation methods with the one-dimensional form of the first example from my paper. For me, at least, the application to vector-valued t, as treated briefly at the end of Example 5.5, would have been more instructive. In higher dimensions the classical empirical distribution function—the empirical process indexed by orthants—is not as useful as its one-dimensional analog. It is not as easy to reduce multiparameter processes via an integration by parts to this classical process. The very fine almost sure approximations for multidimensional empirical distribution functions are not the right tools for many interesting multiparameter problems; the constructions of Dudley and Philipp (1983) or Massart (1989) are more appropriate.

I thank all the discussants for their comments.

ADDITIONAL REFERENCES

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 141–178. Springer, New York.

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete* **62** 509–552.

GILL, R. D. (1987). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). *Scand. J. Statist.* To appear.

GINÉ, E. and ZINN, J. (1988). Bootstrapping general empirical measures. *Ann. Probab.* To appear.

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

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