Comment

Ron Pyke

I have greatly enjoyed reading this paper by David Pollard. It is a further example of his fine expositing skills. By focusing on two particular problems, he underlines for statisticians the practical values that are intrinsic to the subject of weak convergence of empirical processes.

It is slightly more than 60 years since Harald Cramér introduced the idea of an empirical distribution function for real random variables and suggested its use in statistics (Cramér, 1928). Shortly thereafter, the Glivenko-Cantelli-Kolmogorov result of 1931 showed that the empirical distribution function was a strongly consistent estimator of the population distribution function. This was followed for about 25 years by a virtual explosion of applied and theoretical activity on distribution-free nonparametric procedures: Kolmogorov-Smirnov and Cramér-von Mises type statistics: one-sided and two-sided; one-sample and two-sample; weighted and unweighted; asymptotic and exact results; with tables of critical values provided for most.

Forty years ago, in the midst of this activity, J. L. Doob proposed a result that would enable one to obtain the asymptotic behavior of most of the procedures that had been, or would ever be, proposed from an appropriate limiting Gaussian process, a tied-down Brownian motion. (Cf. Doob, 1949). Out of this heuristic beginning, a vast literature has emerged concerning the asymptotic behavior of empirical processes. Throughout this research, the central and enabling property of the empirical measure of a sample of iid observations $X_1, X_2, \cdots, X_n$ has been its basic structure as a *sample average* of iid objects, namely,

$$P_n = n^{-1}(\delta_{X_1} + \delta_{X_2} + \cdots + \delta_{X_n})$$

where δ_x is the degenerate probability measure that puts probability 1 at x. Because of this structure, it is natural that the asymptotic distributional, or weak convergence, results are referred to as central limit theorems (CLT) for empirical processes. (It also suggests interest in other sample average limit laws for P_n, such as the SLLN and LIL.) Shortly before (1), Pollard states that, "In some asymptotic sense, the process $\nu_n f(\cdot, t)$ is approximately Gaussian." This is as close as the author gets to mentioning a CLT for empirical processes. This brief sentence encompasses an enormous literature that pertains to the asymptotic distribution of empirical processes.

Although the setting for these CLT's is much more general than for the classical CLT and the technical aspects are accordingly more complex, their much greater applicability to statistics makes their study worthwhile. In fact, I view any CLT for empirical processes as a conveniently packaged collection of many individual limit theorems of importance to statisticians. For exposition purposes to statisticians, I prefer to define convergence in law of empirical processes (i.e., when a CLT holds) to mean precisely the convergence in law of all statistics that are continuous functions of the empirical process. (Cf. Pyke and Shorack, 1968). From this viewpoint, CLT's for empirical processes are powerful tools that statisticians, or their consulting probabilists, can check out of our Asymptotic Methods' Toolroom. Often, however, these tools need to be individually customized to handle statistics that are only approximately continuous, and this is the situation with the two examples presented here by David Pollard; the simple substitution of $\bar{X}$ for t in the first problem is unfortunately one complication that requires technical care to justify the natural Taylor-expansion heuristics, while the question of asking for the location of a min or max as in the second problem is in and of itself another complication. Regardless of the excellent quality of exposition, the level of this complexity cannot be hidden. The important message, however, is that applications of this type *can* be handled by the theory, regardless of whether or not the particular methodology is understood or even fully appreciated by the user.

Major advances in the theory of statistics are driven ultimately by applications. In 1978, I felt that rather complete results about all three major types of limit theorems for empirical processes were available; namely, for the CLT or weak convergence, Dudley (1978); for the SLLN or Glivenko-Cantelli result, Steele (1978); and for the LIL, Kuelbs and Dudley (1980; a preprint was available in 1978). I therefore used my 1978 IMS Special Invited Lecture to survey the 50 years since Cramér (1928) and to encourage that the rather complete theory then available be brought to bear on applications of empirical process involving multidimensional data. The considerable theoretical advances of the last decade clearly indicate that the subject's theory and methodologies were far from complete in 1978. Many major advances have occurred since then, and along with these have come

Ron Pyke is Professor, Department of Mathematics, University of Washington, Seattle, Washington 98195.