The Annals of Statistics
1985, Vol. 13, No. 1, 423–427

A CONVERSE TO SCHEFFÉ'S THEOREM

BY DENNIS D. BOOS

North Carolina State University

Convergence of densities implies convergence of their distribution functions via Scheffé's theorem. This paper is concerned with the converse: what are sufficient conditions to obtain convergence of densities from convergence of distribution functions? A general lemma is given and local limit results are obtained for translation and scale statistics.

1. Introduction. When does convergence in distribution lead to convergence of the associated densities? Specifically, let $T_n = (a_{1n}(T_{1n} - b_{1n}), \cdots, a_{kn}(T_{kn} - b_{kn}))$ be a standardized random vector in R^k which converges weakly to a distribution having density (Radon-Nikodym derivative) g with respect to (wrt) Lebesgue measure μ on R^k. If T_n has density g_n wrt μ, what are sufficient conditions for g_n to converge to g, say pointwise a.e. μ? Lemma 1 below gives one set of such conditions and Theorems 1 and 2 of Section 3 verify these conditions for certain translation and scale statistics. The statistical motivation for these new local limit theorems arose in a Bayesian context and is discussed briefly at the end of Section 3.

2. The main lemma. Let G_n and G denote the distribution functions (dfs) of g_n and g and let "$\Rightarrow$" stand for weak convergence. If $G_n \Rightarrow G$, a sufficient condition for $g_n(x) \to g(x)$ a.e. μ is that each subsequence $g_{n'}$ have a further subsequence $g_{n''}$ converging to some density g^*. In that case, Scheffé's theorem (Serfling, 1980, page 17) yields $G_{n''} \Rightarrow G^*$ which would contradict $G_n \Rightarrow G$ if $g^* \neq g$ on a set of positive μ measure. However, the existence of convergent subsequences is not easy to verify in the absence of a metric space. (In the topology of pointwise convergence g_n is compact if $\{g_n\}$ is closed and $\{g_n(x)\}$ is bounded for each x. Unfortunately, this topology is not metrizable here, e.g., Dugundji, 1966, page 273). Thus, it is convenient to use uniform convergence on compacts for which the Ascoli theory is available (e.g., Royden, 1968, page 179). This leads to Lemma 1 and (2.3). If the equicontinuity is uniform, then we can extend the result to uniform convergence on R^k, i.e., wrt the norm $\sup_x |g_n(x) - g(x)| \equiv \|g_n - g\|_\infty$.

LEMMA 1. *Suppose that G_n and G have continuous densities g_n and g wrt μ on R^k. If $G_n \Rightarrow G$ and*

$$\sup_n |g_n(x)| \le M(x) < \infty, \quad \textit{each} \quad x \in R^k, \tag{2.1}$$

Received March 1983; revised July 1984.
AMS 1980 *subject classifications.* Primary 62E20; secondary 62G05.
Key words and phrases. Convergence of densities, local limit theorems, translation statistics, scale statistics.

and

$$\text{(2.2)}\quad \begin{array}{l}\{g_n\} \textit{ is equicontinuous, i.e., for each } x \textit{ and } \varepsilon > 0 \textit{ there exists} \\ \delta(x, \varepsilon) \textit{ and } n(x, \varepsilon) \textit{ such that } |x - y| < \delta(x, \varepsilon) \textit{ implies that} \\ |g_n(x) - g_n(y)| < \varepsilon \textit{ for all } n \geq n(x, \varepsilon),\end{array}$$

then for any compact subset C of R^k

$$\sup_{x \in C} |g_n(x) - g(x)| \to 0 \quad \textit{as} \quad n \to \infty. \tag{2.3}$$

If $\{g_n\}$ is uniformly equicontinuous, i.e., $\delta(x, \varepsilon) = \delta(\varepsilon)$ and $n(x, \varepsilon) = n(\varepsilon)$ in (2.2) do not depend on x, and $g(x_n) \to 0$ whenever $|x_n| \to \infty$, then

$$\sup_{x \in R^k} |g_n(x) - g(x)| = \|g_n - g\|_\infty \to 0 \quad \textit{as} \quad n \to \infty. \tag{2.4}$$

PROOF. Conditions (2.1) and (2.2) are exactly what the classical Ascoli theorem requires for $\{g_n\}$ to be compact wrt the topology of uniform convergence on compacts. Thus, if $g_{n'}$ is a subsequence of g_n, then there is a further subsequence $g_{n''}$ which converges uniformly on each compact subset of R^k to some g^*. Scheffé's theorem then shows that $g^* = g$ a.e. μ, and since they are both continuous $g^*(x) = g(x)$, each $x \in R^k$. So, for each compact set C

$$\sup_{x \in C} |g_{n''}(x) - g(x)| \to 0, \tag{2.5}$$

and (2.3) follows by the usual argument (Royden, 1968, page 37, problem 11). Now, suppose that (2.4) is false. Then, there must exist an $\varepsilon > 0$ and a subsequence n' of n and a sequence $x_{n'}$ such that

$$|g_{n'}(x_{n'}) - g(x_{n'})| > \varepsilon \quad \text{for all} \quad n'. \tag{2.6}$$

If the $x_{n'}$ are bounded, then there exists some C which contain all the $x_{n'}$ and (2.6) contradicts (2.5). If the $x_{n'}$ are not bounded, then there is a subsequence $x_{n''}$ such that at least one coordinate of $x_{n''}$ is tending to $\pm\infty$. Since $g(x_{n''}) \to 0$ as $n'' \to \infty$ we have by (2.6) that $g_{n''}(x_{n''}) \geq \varepsilon/2$ for all n'' sufficiently large, say all $n'' \geq n_0$. Let $\delta = \delta(\varepsilon/4)$ be such that $|x_{n''} - y| < \delta$ implies $|g_{n''}(x_{n''}) - g_{n''}(y)| < \varepsilon/4$ for all $n \geq n(\varepsilon/4)$. Then

$$G_{n''}(x_{n''} + \delta) - G_{n''}(x_{n''} - \delta) = \int_{x_{n''}-\delta}^{x_{n''}+\delta} g_{n''}(y)\, dy \geq (2\delta)^k \varepsilon/4 \tag{2.7}$$

for all $n'' \geq \max(n_0, n(\varepsilon/4))$ since $g_{n''}(y) \geq \varepsilon/4$ for all y in the δ neighborhood of $x_{n''}$ (let $|\cdot|$ on R^k be the maximum of the coordinates). But $G_{n''}(x_{n''} + \delta) - G_{n''}(x_{n''} - \delta) \leq 2\|G_{n''} - G\|_\infty + G(x_{n''} + \delta) - G(x_{n''} - \delta)$, and $\|G_{n''} - G\|_\infty \to 0$ by Polya's theorem in R^k and $G(x_{n''} + \delta) - G(x_{n''} - \delta) = g(x_{n''}^*)(2\delta)^k$, where $x_{n''}^* \in (x_{n''} - \delta, x_{n''} + \delta)$. Since at least one coordinate of $x_{n''} \to \pm\infty$, the same is true for $x_{n''}^*$. Thus, $g(x_{n''}^*) \to 0$ as $n'' \to \infty$ and (2.7) is contradicted. □

3. Application to translation and scale statistics. A statistic $T(X) = T(X_1, \cdots, X_n)$ is a *translation* statistic if $T(x + a) = T(x) + a$ for all real a and vectors $x = (x_1, \cdots, x_n)$. Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) observations with common density $f(x)$. Then, using the translation property we have

$$
\begin{aligned}
G_n(y+t) &= P(T(X) \le y+t) \\
&= \int I(T(x) \le y+t)\ \prod_{i=1}^n f(x_i)\ dx_1 \cdots dx_n \\
&= 1 - \int I(T(x) > y)\ \prod_{i=1}^n f(x_i+t)\ dx_1 \cdots dx_n \\
&= 1 - \int I(T(x) > y) \exp\{\textstyle\sum_{i=1}^n \log f(x_i+t)\}\ dx_1 \cdots dx_n.
\end{aligned}
$$

Now, to get derivatives of $G_n(y)$ we need only justify the interchange of operations

$$
\begin{aligned}
\frac{d^k}{dy^k} G_n(y) &= \frac{d^k}{dt^k} G_n(y+t)\ |_{t=0} \\
&= -\int I(T(x) > y)\ \frac{d^k}{dt^k} \exp\{\textstyle\sum_{i=1}^n \log f(x_i+t)\}\,|_{t=0}\ dx_1 \cdots dx_n.
\end{aligned}
\tag{3.1}
$$

Klaassen (1984) introduced this approach and showed that if f is absolutely continuous with $\int |f'(x)|\,dx < \infty$, then

$$
g_n(y) = \int I(T(x) > y)[\textstyle\sum_{i=1}^n -(f'/f)(x_i)]\ \prod_{i=1}^n f(x_i)\ dx_1 \cdots dx_n. \tag{3.2}
$$

For notational convenience, the first derivative of f is often written f', and if G_n is a df then g_n is its first derivative.

We can get similar representations for the logarithm of positive *scale* statistics $S(X)$, i.e., those satisfying $S(ax) = aS(x) > 0$ for all $a > 0$ and $x = (x_1, \cdots, x_n)$. Then,

$$
\begin{aligned}
H_n(y+t) &= P(\log S(x) \le y+t) \\
&= 1 - \int I(\log S(x) > y) \exp\{\textstyle\sum_{i=1}^n \log(e^t f(e^t x_i))\}\ dx_1 \cdots dx_n.
\end{aligned}
\tag{3.3}
$$

If f is absolutely continuous with $\int |xf'(x)|\,dx < \infty$, then

$$
h_n(y) = \int I(\log S(x) > y)\ \sum_{i=1}^n \left(-1 - x_i \frac{f'}{f}(x_i)\right) \prod_{i=1}^n f(x_i)\ dx_1 \cdots dx_n. \tag{3.4}
$$

Obviously, higher order derivatives may be taken under sufficient regularity conditions. A similar approach is also possible for two-sample shift statistics and ratios of positive scale statistics.

Two local limit results can now be given. Let $I(f) = \int_{-\infty}^{\infty} [f'(x)/f(x)]^2 f(x)\,dx$, $I_1(f) = \int_{-\infty}^{\infty} [-1 - xf'(x)/f(x)]^2 f(x)\,dx$, and let $\Phi(x)$ and $\phi(x)$ be the standard normal df and density. F is the df associated with f and $T(F)$ is the target "parameter" of T_n.

THEOREM 1. *Let $X_1, \cdots, X_n$ be independent with common density $f(x)$ on R^1. Let T_n be a translation statistic such that $G_n(y) = P(n^{1/2}[T_n - T(F)] \le y) \to \Phi(y/\sigma)$, each y. If f is absolutely continuous and $I(f) < \infty$, then with $g(x) = \sigma^{-1}\phi(x/\sigma)$ we have*

$$\| g_n - g \|_\infty \to 0 \quad as \quad n \to \infty.$$

THEOREM 2. *Let $X_1, \cdots, X_n$ be independent with common density $f(x)$ on R^1. Let S_n be a positive scale statistic such that $H_n(y) = P(n^{1/2}[\log S - \log S(F)] \le y) \to \Phi(y/\sigma)$, each y. If f is absolutely continuous and $I_1(f) < \infty$, then with $h(x) = \sigma^{-1}\phi(x/\sigma)$ we have*

$$\| h_n - h \|_\infty \to 0 \quad as \quad n \to \infty.$$

PROOFS. Since both proofs are virtually the same, only the first will be given. From $I(f) < \infty$ via Cauchy-Schwarz we get $\int | f'(x) | \, dx < \infty$ and thus (3.2) with $g_n(y)$ replaced by $n^{-1/2}g_n(n^{-1/2}y + T(F))$ can be written as

$$g_n(y) = E[n^{-1/2} \textstyle\sum_{i=1}^n -(f'/f)\,(X_i)]I(n^{1/2}(T_n - T(F)) > y).$$

One application of the Cauchy-Schwarz inequality yields $g_n(y) \le [I(f)]^{1/2}$ and another gives

$$| g_n(x) - g_n(y) | \le [I(f)]^{1/2} | G_n(x) - G_n(y) |^{1/2} \le I(f) | x - y |^{1/2}.$$

Hence, $\{g_n\}$ is uniformly bounded and uniformly equicontinuous so that Lemma 1 applies. □

Relation to local limit theorems. Theorems 1 and 2 yield local limit theorems for virtually all location and scale estimators in use as long as $I(f) < \infty$ or $I_1(f) < \infty$. On the other hand, when $T_n(X) = \bar{X}$ the sample mean, then characteristic function arguments yield better results. From Feller (1966, page 515, 516) we know that if X_1 has finite 2nd moment and characteristic function ω such that $| \omega |^r$ is integrable for some integer r, then $\| g_n - g \|_\infty \to 0$. Moreover, if $| \omega |^r$ is not integrable for some r, then g_n must be unbounded. Since $I(f) < \infty$ leads to bounded g_n, it must be a stronger condition. For example, if X_1 has the uniform density $f(x) = I(-\frac{1}{2} \le x \le \frac{1}{2})$, then $I(f) = \infty$ but $| \omega(t) |^2 = 4t^{-2}\sin^2(t/2)$ is integrable.

Extensions. Analogous results for two-sample shift statistics and ratios of positive scale statistics should be clear. Boos (1983) also gives results for convergence of g_n' to g' and for the convergence of certain bivariate densities. However, the methods of this section are not ideally suited for joint densities, and there are no "density" versions of Slutsky's theorem or of the Cramér-Wold device to help extend results from 1 to k dimensions.

Statistical application. The motivation for considering convergence of densities of statistics arose in Boos and Monahan (1983). They consider Bayesian analysis based on the sampling density $\ell_{\hat\theta}(x \mid \theta)$ of an estimator $\hat\theta$ rather than on the full likelihood of the sample. Convergence of the resulting posteriors then depends on convergence of $\ell_{\hat{\theta}}(x \mid \theta)$ and Theorems 1 and 2 are relevant. For example, $\hat{\theta}$ might be the sample median or a trimmed mean. (We know that the density of the sample median converges pointwise (Serfling, 1981, page 85), but Theorem 1 adds the uniformity.) However, the approach is most useful in a semiparametric framework where $\ell_{\hat{\theta}}(x \mid \theta)$ is unknown but can be estimated by the bootstrap. In that case analogues to Theorems 1 and 2 can be proved for the bootstrap density estimator $\hat{\ell}_{\hat{\theta}}(x \mid \theta)$ as long as $I(f_n)$ or $I_1(f_n)$ stay bounded, where f_n is the density generating the bootstrap samples.

Acknowledgement. This paper was written during a sabbatical visit to the Department of Statistics, University of California, Berkeley. I would like to thank Peter Bickel, Harold Goldstein, Chris Klaassen, and Lucien LeCam for helpful discussions during this time.

REFERENCES

BOOS, D. D. (1983). On convergence of densities of translation and scale statistics. Institute of Statistics Mimeo Series #1625, North Carolina State University, Raleigh.

BOOS, D. D. and MONAHAN, J. F. (1983). A robust Bayesian approach using bootstrapped likelihoods. Preprint.

DUGUNDJI, J. (1966). *Topology.* Allyn and Bacon, Boston.

FELLER, W. (1966). *Introduction to Probability Theory and Its Applications II.* Wiley, New York.

KLAASSEN, C. A. J. (1984). Location estimators and spread. *Ann. Statist.* **12** 311–321.

ROYDEN, H. L. (1968). *Real Analysis.* Macmillan, New York.

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

DEPARTMENT OF STATISTICS
BOX 8203
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8203