The Annals of Statistics
1987, Vol. 15, No. 4, 1699–1708

STRONG CONVERGENCE OF DISTRIBUTIONS OF ESTIMATORS[1]

BY P. JEGANATHAN

University of Michigan

It is shown that the convergence in law of estimators entails convergence uniformly over all Borel sets whenever the estimators are asymptotically equivariant in a suitable sense and the likelihood ratios of the sample are appropriately smooth. This result generalizes a recent result of Boos in many directions.

1. Introduction. In a recent paper, Boos (1985) has obtained the interesting result that the densities of estimators of a one-dimensional location parameter converge uniformly to the density of a Gaussian distribution whenever the estimators converge weakly (in law) to a Gaussian distribution. (He also obtained a similar result for the estimators of a scale parameter.) He assumed that the estimators are location equivariant and that the Fisher information is finite. The usual general method of establishing such convergence of densities is through the method of characteristic functions and such a method seems to work well in this context only when the estimators are smooth functions, independent of the sample size, of sums of i.i.d. observations. On the other hand, Boos' method works for any location equivariant estimator. The main purpose of this paper is to introduce a method which results in improvements and generalizations of Boos' result in many directions. However, we shall obtain only the convergence of distributions of estimators uniformly over all Borel sets and uniformly over suitable neighborhoods of the parameter; this convergence will be referred to as *strong convergence* and is weaker than the convergence of densities but sufficient for many statistical purposes. Specifically, the present method gives the joint strong convergence of estimators of any finite number of parameters, whereas Boos' method does not seem to generalize well to situations with more than one parameter (see Section 3). Second, his condition of finiteness of Fisher information is replaced by the more general condition of *local asymptotic normality* [cf. Le Cam (1960)] of the likelihoods of the sample. Third, his method depends on the exact equivariance of the estimators, whereas the present method uses only a certain kind of asymptotic equivariance.

The crucial idea of the present method is that if one has the approximation for the expected values of a suitable class of even only extremely smooth functions of the estimators and if the likelihood of the sample is sufficiently smooth, then one can get the approximation for the expected value of any Borel measurable function, provided the estimators are asymptotically equivariant in a

Received October 1985; revised March 1987.

[1] Research supported by National Science Foundation Grant No. DMS-85-09837.

AMS 1980 *subject classifications*. Primary 62F12; secondary 62G20.

Key words and phrases. Strong convergence, local asymptotic normality, asymptotic equivariance.

suitable sense. One purpose of the present proof (see Section 2) is to make explicit the preceding idea, which we hope to develop further to obtain higher order approximations for nonparametric and robust estimators uniformly over all Borel sets. Another advantage of the present proof is that it almost immediately extends to the situation in which the likelihood ratios are *locally asymptotically mixed normal* [see, e.g., Davies (1985)], a situation which arises if, for example, one wants to prove the strong convergence of least squares estimators of an explosive autoregressive process; this and other applications of the present method can be found in Jeganathan (1986, 1987).

In Section 2, we state and prove our main result. In Section 3, we first recall the result of Boos (1985) in order to compare it with our result. Incidentally, we also point out in Section 3 that Boos' result holds also for certain nonregular location families treated, e.g., in Klaassen (1984), in which the Fisher information is infinite. We then give a straightforward application of our result to obtain the strong convergence of the estimators of location and scatter of an elliptical family of distributions. Our result can also be used, for example, to obtain the strong convergence of general M-estimators of the parameters of an autoregressive process, but the verification of our conditions in such a case requires more elaborate and involved arguments, especially the verification of asymptotic equivariance, since we require such equivariance to hold (uniformly) for all *measurable* functions: See Jeganathan (1986) for the details.

The motivation for considering such a strong convergence is the following. As explained by Le Cam in several papers [see, e.g., Le Cam (1975)], the information contained in the estimators is asymptotically reflected in the limiting distributions only when the likelihood ratios of the estimators converge in law appropriately to the likelihoods of the limiting distributions. In the present asymptotic equivariant situation, such a convergence becomes equivalent to the strong convergence of the estimators, as can be deduced from our main result when the limits are Gaussian. In the case of exact equivariance, this equivalence also follows as a very special case of a general result of Torgersen (1972). Further, such a convergence of likelihood ratios entails that the estimators compare asymptotically in the decision theoretic sense as described e.g., in Le Cam (1975), in the same manner as the covariance matrices of their Gaussian limits.

2. Likelihood of the sample and the strong convergence of estimators. For each $n \geq 1$, let $\{P_{\vartheta, n}\colon \vartheta \in \Theta\}$ be a family of probability measures (defined on some measurable space), where the parameter space Θ is an open subset of the k-dimensional euclidean space R^k, $k \geq 1$. $P_{\vartheta, n}$ usually stands for the joint probability distribution of the sample.

The family $\{P_{\vartheta, n}\colon \vartheta \in \Theta\}$ is said to be *locally asymptotically normal* (LAN) at $\vartheta_0 \in \Theta$ if there exists a sequence $\{W_n(\vartheta_0)\}$ of random k-vectors and a p.d. matrix $B(\vartheta_0)$ such that the differences

$$\log \frac{dP_{\vartheta_0 + \delta_n h_n, n}}{dP_{\vartheta_0, n}} - \left[h_n' W_n(\vartheta_0) - \frac{1}{2} h_n' B(\vartheta_0) h_n \right]$$

converge to zero in $P_{\vartheta_0, n}$-probability for every bounded $\{h_n\}$ of R^k, where δ_n, which may depend on ϑ_0, $n \geq 1$, are suitable p.d. matrices and the sequence $\{W_n(\vartheta_0)\}$ converges in law, under $P_{\vartheta_0, n}$, to the k-variate Gaussian distribution $\mathscr{N}(0, B(\vartheta_0))$.

Let $\{T_n\}$ be a sequence of estimators of interest. Since ϑ_0 is fixed throughout, it will be convenient to suppress ϑ_0 and set $T_n^* = \delta_n^{-1}(T_n - \vartheta_0)$ and $P_{\vartheta_0 + \delta_n h, n} = Q_{h, n}$.

We now state our main result, where $\sup_{|f| \leq 1}$ means supremum taken over all Borel measurable functions bounded in absolute value by 1.

THEOREM 1. *Assume that the LAN condition is satisfied at $\vartheta_0 \in \Theta$. Let $\{T_n\}$ be a sequence of estimators of ϑ. Assume that the following two conditions are satisfied*:

(A) *There is a sequence $\{M_n\}$ of almost surely p.d. (random) matrices converging in $Q_{0, n}$-probability to the unit matrix such that the quantities*

$$\sup_{|h| \leq \alpha} \sup_{|f| \leq 1} \left| E\left[f(T_n^* - h) | Q_{h, n}\right] - E\left[f(T_n^* - h - M_n u) | Q_{h+u, n}\right] \right|$$

converge to zero as $n \to \infty$ for all $\alpha > 0$ and for all $u \in R^k$.

(B) *The random vectors T_n^* converge in law, under $Q_{0, n}$, to a probability distribution $\mathscr{L}$.*

Then the quantities

$$\sup_{|h| \leq \alpha} \sup_{|f| \leq 1} \left| E\left[f(T_n^* - h) | Q_{h, n}\right] - \int f(x) \mathscr{L}(dx) \right|$$

converge to zero for all $\alpha > 0$.

REMARK. Before going into the details of the proof it may be noted that condition (A) is immediately satisfied with $M_n \equiv I$, the unit matrix, when ϑ is the location parameter and T_n is location equivariant; similarly for the equivariant estimators of regression parameters of a linear model. The same is the case for scale parameters but with random M_n. It seems that condition (A) is inadequate to deal with the joint convergence of location and scale and, therefore, we introduce the following modification. An illustration of the use of this modified condition is given in Section 3.

(A′). Let p, $0 \leq p < k$, be fixed. There are sequences $\{M_{n1}\}$ and $\{M_{n2}\}$ of almost surely p.d. (random) matrices, respectively, of orders $p \times p$ and $(k - p) \times (k - p)$, converging in $Q_{0, n}$-probability to unit matrices. There is also a sequence $\{T_n^*(u)\}$, $u \in R^k$, of random k-vectors such that the quantities, with $M_n = \operatorname{diag}\{M_{n1}, M_{n2}\}$,

$$\sup_{|h| \leq \alpha} \sup_{|f| \leq 1} \left| E\left[f(T_n^* - h) | Q_{h, n}\right] - E\left[f(T_n^*(u) - h - M_n u) | Q_{h+u, n}\right] \right|$$

converge to zero as $n \to \infty$ for all $\alpha > 0$ and for all $u \in R^k$. Here $T_n^*(u)$ is such that the differences $T_n^*(u) - T_n^*$ converge to zero in $Q_{0, n}$-probability for all

$u \in R^k$ and such that $T_n^*(u) = (T_{n1}^*(u), T_{n2}^*)$ with T_{n2}^* being the $(k-p)$-vector consisting of the last $(k-p)$ variables of T_n^* and with $T_{n1}^*(u)$ depending on u only through the last $(k-p)$ elements of u.

Note that the case $p = 0$ of condition (A′) means to coincide with condition (A). Theorem 1 will be proved under conditions (A′) and (B).

The following lemma will be used in the proof of Theorem 1. It can be found, at least implicitly, in Le Cam [(1974), Chapter 13], Hájek (1970) and Strasser (1978). For the sake of completeness, we present a proof in the Appendix. To state it, in addition to the notations introduced earlier, let μ be the Lebesgue measure in R^k and let $D_\alpha = \{h \in R^k: |h| \le \alpha\}$, $\alpha > 0$. Further, we set $W_n(\theta) = W_n$ and, for simplicity only $B(\theta) = I$, where the variables $W_n(\theta)$ and the matrix $B(\theta)$ are the ones occurring in the definition of the LAN condition. Also, let $S(x) = (2\pi)^{-k/2}\exp(-|x|^2/2)$.

LEMMA 2. *Assume that LAN condition holds at $\vartheta_0 \in \Theta$. Then the differences*

$$\frac{1}{\mu(D_\alpha)}\int_{D_\alpha} E\left[g(u)|Q_{h_n+u,n}\right] du \tag{1}$$
$$- \frac{1}{\mu(D_\alpha)}\int_{D_\alpha}\int\left\{\int_{R^k} g(v)S(v+h_n-W_n)\,dv\right\} dQ_{h_n+u,n}\,du$$

converge to zero, first by letting $n \to \infty$ and then $\alpha \to \infty$, for every bounded sequence $\{h_n\}$ of R^k and uniformly for all jointly measurable variables $g(u)$ such that $|g| \le 1$, $n \ge 1$.

In the following proof, "converge uniformly" means "converge uniformly with respect to all Borel measurable f such that $|f| \le 1$."

PROOF OF THEOREM 1. By conditions (A′) and (B), it follows that $\mathscr{L}(T_n^* - u|Q_{u,n})$, the distribution of $(T_n^* - u)$ under $Q_{u,n}$, converges in law to the distribution $\mathscr{L}$. Hence, since the sequence $\{h_n\}$ is bounded, it follows by contiguity that $\mathscr{L}(T_n^* - h_n|Q_{h_n,n})$ converges in law to $\mathscr{L}$ for every bounded $\{h_n\}$. Therefore, it is enough to show that for every subsequence $\{\gamma\} \subseteq \{n\}$, there is a further subsequence $\{p\} \subseteq \{\gamma\}$ such that $E[f(T_p^* - h_p)|Q_{h_p,p}]$ converges uniformly. Therefore, we assume without loss of generality that $h_n \to h \in R^k$ and that (T_n^*, W_n, M_n) converge in law under $Q_{0,n}$ to (T, W, I) and prove that $E[f(T_n^* - h_n)|Q_{h_n,n}]$ converges uniformly. According to condition (A′), the differences between $E[f(T_n^* - h_n)|Q_{h_n,n}]$ and the l.h.s. of (1) with $g(u) = f(T_n^*(u) - h_n - M_n u)$ converge uniformly to zero. Hence, by Lemma 2, it is enough to show that the r.h.s. of (1) converges uniformly, with the preceding $g(u)$, for every fixed $\alpha > 0$ to some limits as $n \to \infty$ and that these limits converge uniformly as $\alpha \to \infty$.

First, consider the case $p = 0$, that is, $T_n^*(u) = T_n^*$, so that the quantity inside the brackets of (1) is given by

$$\begin{aligned} &\int_{R^k} f(T_n^* - h_n - M_n v) S(v + h_n - W_n)\, dv \\ (2) \qquad &= |M_n|^{-1} \int_{R^k} f(v) S\big(-M_n^{-1}(v - T_n^* + h_n) + h_n - W_n\big)\, dv \\ &= J_f(M_n, T_n^*, W_n), \quad \text{say.} \end{aligned}$$

Using the fact that $S(x)$ is a Gaussian density, it follows readily that the family of functions $\{J_f(m, t, w)\colon |f| \le 1\}$ is equicontinuous at all $t \in R^k$, $w \in R^k$ and m positive definite; that is, the quantities

$$\sup\Big\{\big|J_f(m_1, t_1, w_1) - J_f(m_2, t_2, w_2)\big|\colon |m_1 - m| + |t_1 - t| + |w_1 - w| + |m_2 - m| + |t_2 - t| + |w_2 - w| < \varepsilon\Big\}$$

converge uniformly to zero as $\varepsilon \to 0$ for all t, w and m as before. Also the functions J_f are bounded in absolute value by 1, uniformly in f, w, t and m positive definite. Further, since $\mathscr{L}(T_n^*, W_n, M_n | Q_{0,n})$ converge in law to $\mathscr{L}(T, W, I) = Q_0$, say, contiguity implies that the laws $\mathscr{L}(T_n^*, W_n, M_n | Q_{h_n+u,n})$ converge in law to $\mathscr{L}(T, W, I | Q_{h+u})$, where Q_h is such that

$$\frac{dQ_h}{dQ_0} = \exp\left(h'W - \frac{|h|^2}{2}\right).$$

Hence, in view of the preceding three facts, it follows from well known results [see, e.g., Billingsley (1968), page 17, Problem 8] that the quantities $E[J_f(M_n, T_n^*, W_n) | Q_{h_n+u,n}]$ converge uniformly to $E[J_f(I, T, W) | Q_{h+u}]$ and, hence, the r.h.s. of (1) converges uniformly as $n \to \infty$ for every $\alpha > 0$ to the quantity

$$(3) \qquad \frac{1}{\mu(D_\alpha)} \int_{D_\alpha} E\big[J_f(I, T, W) | Q_{h+u}\big]\, du.$$

But this is the r.h.s. of (1) with $Q_{h_n+u,n} \equiv Q_{h+u}$ and $(T_n^*, W_n, M_n) \equiv (T, W, I)$, $n \ge 1$, so that the difference between (3) and the quantity [l.h.s. of (1)]

$$(4) \qquad \frac{1}{\mu(D_\alpha)} \int_{D_\alpha} E\big[f(T - h - u) | Q_{h+u}\big]\, du$$

converges uniformly in absolute value to zero as $\alpha \to \infty$. But since $\mathscr{L}(T - h - u | Q_{h+u})$, which is the limit of $\mathscr{L}(T_n^* - h - u | Q_{h+u,n})$, does not depend on h and u, the integrand of (4) does not depend on h and u. This completes the proof for the case $p = 0$.

Now consider the case $p > 0$. Let v_1 (resp. v_2) be the p-vector consisting of the first p [resp. last $(k - p)$] elements of $v \in R^k$ so that $v = (v_1, v_2)$. Also, write $S(v + h_n - W_n) = S_{n1}(v_1, v_2) S_{n2}(v_2)$, where, suppressing h_n and W_n, S_{n1} is the conditional Gaussian density of v_1 given v_2 and S_{n2} is the marginal density of v_2. Also, take for simplicity that $h_n = 0$. Then the quantity inside the brackets on the r.h.s. of (1) can be written as

$$(5) \quad \int_{R^{k-p}} \left\{ \int_{R^p} f(T_{n1}^*(v) - M_{n1}v_1, T_{n2}^* - M_{n2}v_2) S_{n1}(v_1, v_2)\, dv_1 \right\} S_{n2}(v_2)\, dv_2.$$

Now observe that by condition (A′), $T_{n1}^*(v)$ depends on v only through v_2 so that by making a change of variable it can be brought into the exponent of the density $S_{n1}(v_1, v_2)$. Now since the differences $T_n^*(v) - T_n^*$ converge to zero in $Q_{0,n}$-probability, it follows easily from the preceding observation that the differences between (2) and (5) converge to zero in $Q_{0,n}$-probability and, hence, in $Q_{h_n+u,n}$-probability for all $u \in R^k$, by contiguity. Hence, this case is reduced to the case $p = 0$ and the proof of Theorem 1 is complete. □

3. Estimators of location and scatter. We first recall Boos' (1985) result. Let $X_1, X_2, \dots$ be i.i.d. variables with common density, with respect to the Lebesgue measure, $f(x, \vartheta) = f(x - \vartheta)$, $x, \vartheta \in R$. Let T_n be a sequence of translation equivariant estimators of ϑ, that is,

$$T_n(X_1 + a, \dots, X_n + a) = T_n(X_1, \dots, X_n) + a$$

for all real a. Using the translation property,

$$P_\vartheta(T_n \le y + t) = \int I(T_n \le y) \prod_{j=1}^n f(x_j + t - \vartheta)\, dx_1 \cdots dx_n$$
$$= 1 - \int I(T_n > y) \prod_{j=1}^n f(x_j + t - \vartheta)\, dx_1 \cdots dx_n$$

for all real ϑ and t. Then if $f(x)$ is differentiable and if the differentiation can be taken inside, the densities of $\delta_n^{-1} T_n$ when $\vartheta = 0$ are given by

$$(6) \quad \begin{aligned} g_n(y) &= -\int I(\delta_n^{-1} T_n > y) S_n \prod_{j=1}^n f(x_j)\, dx_1 \cdots dx_n \\ &= \int I(\delta_n^{-1} T_n \le y) S_n \prod_{j=1}^n f(x_j)\, dx_1 \cdots dx_n, \end{aligned}$$

where $S_n = \delta_n \Sigma_{j=1}^n \dot{f}/f(x)$ with $\dot{f}(x)$ the derivative of $f(x)$. This identity was obtained by Klaassen (1984) under the condition $\int |\dot{f}(x)|\, dx < \infty$, which allows one to take the differentiation inside the integral. Now assume the stronger condition that f is absolutely continuous with respect to Lebesgue measure and that

$$(7) \quad 0 < \int \left[\frac{\dot{f}}{f}(x) \right]^2 f(x)\, dx < \infty.$$

Then under this condition Boos (1985) showed, using his lemma on the L_∞-compactness of densities, that $g_n(y)$ converge to the Gaussian density (with $\delta_n = n^{-1/2}$) uniformly in y whenever $\sqrt{n}\, T_n$ converge in law to a Gaussian variable.

Our first remark is that, using uniform integrability of S_n which follows from (7), one can also obtain the preceding result directly. First note that $g_n(y)$

converges pointwise, as can be seen from the r.h.s. of Klaassen's identity (6), whenever $(\delta_n^{-1}T_n, S_n)$ converge in law jointly, which can be assumed by passing to a subsequence if necessary. Then one can show, by considering again the r.h.s. of (6), that this pointwise convergence entails the uniform convergence; the required arguments are essentially the same as the ones used to prove the familiar fact that if a sequence of probability distribution functions converge weakly to a continuous probability distribution function, then the convergence will be uniform over the real line.

The reason for mentioning this remark is that according to Lemma 3.1 of Klaassen (1984), the uniform integrability of S_n holds also for many nonregular families treated there and, therefore, Boos' result extends to those cases also. It may also be noted that, apart from the equivariance condition, since the only requirement of Theorem 1 is the appropriate smoothness of the likelihood of the sample, our main result also extends to the aforementioned and many other nonregular families treated in Chapters 5 and 6 of Ibragimov and Hasminskii (1981), since the required smoothness is studied in those chapters.

Now, the arguments leading up to Klaassen's identity (6) do not extend well, unfortunately, for the estimators of more than one location parameter since one has to take more than one partial derivative and the required restriction on the density f becomes severe as the number of parameters increase. Furthermore, the resulting expression inside the integral of (6) becomes difficult to deal with. We refer to Boos [(1983), Section 3] for further details on the difficulties mentioned here. On the other hand, if one is interested, for example, in applying our main result to obtain the strong convergence of equivariant estimators of k, say, regression parameters of a linear model, then it is sufficient to assume that the common density of the error variables satisfies (7) together with suitable conditions on the design matrix to ensure the LAN condition.

We now discuss another straightforward, but important, example to illustrate the need for introducing condition (A′). Let $f(x) = f(|x|)$ be a spherically symmetric density in R^q. [A detailed treatment of spherical and elliptical distributions can be found in Muirhead (1982)]. Let $X_1, X_2, \dots$ be i.i.d. q-vectors with a common density $f(x, \mu, V) = |V| f(|V(x-\mu)|)$, where $\mu = (\mu_1, \dots, \mu_q)$ is a q-vector and $V = (V_{ij})$ is a p.d. symmetric matrix. Let $[V]$ be the lexicographically written row vector of the lower $q(q+1)/2$ entries of the matrix V, so that $\Theta = \{(\mu, [V]) \colon \mu \in R^q \text{ and } V \text{ is symmetric p.d.}\}$; this clearly is an open subset of R^k with $k = q(q+3)/2$. Let $\{U_n\}$ be a sequence of q-vector valued estimators of μ such that

$$U_n(X_1 + a, \dots, X_n + a) = U_n(X_1, \dots, X_n) + a \tag{8}$$

for all $a \in R^q$ and let $\{S_n\}$ be a sequence of symmetric matrix valued estimators of V^{-2}, such that

$$S_n(B(X_1 - a), \dots, B(X_n - a)) = BS_n(X_1, \dots, X_n)B' \tag{9}$$

for all p.d. B and $a \in R^q$.

Now fix μ and V, and let $\mu_{ng} = \mu + g/\sqrt{n}$, $g \in R^q$ and $V_{nH} = V + VHV/\sqrt{n}$, $H = (h_{ij})$ is symmetric. Without loss of generality, one can assume that V_{nH} is p.d. for all H and n. Define $G_{nH} = (I + HV/\sqrt{n})$. Now introduce the following two conditions:

(C1) The differences

$$\sqrt{n}\,[U_n(G_{nH}(X_1-\mu),\dots,G_{nH}(X_n-\mu)) - U_n((X_1-\mu),\dots,(X_n-\mu))]$$

converge to zero in probability for all H.

(C2) The density f is absolutely continuous with respect to Lebesgue measure and

$$0 < \int\left[\frac{f^{\cdot}}{\sqrt{f}}(|x|)\right]^2 dx < \infty \quad \text{and} \quad 0 < \int\left[|x|^2\frac{f^{\cdot}}{\sqrt{f}}(|x|)\right]^2 dx < \infty.$$

COROLLARY 3. *Let $P_{\mu, V, n}$ be the distribution of the sample with the common elliptical density $f(x,\mu,V)$. Assume that conditions* (8), (9), (C1) *and* (C2) *are satisfied. Further, assume that the distributions $\mathscr{L}(\sqrt{n}(U_n-\mu), \sqrt{n}(S_n - V^{-2})|P_{\mu,V,n})$ converge in law to a probability distribution $\mathscr{L}$. Then the differences*

$$E\left[f\left(\sqrt{n}(U_n-\mu_{ng_n}), \sqrt{n}\left(S_n - V_{nH_n}^{-2}\right)\right)\Big|P_{\mu_{ng_n}, V_{nH_n}, n}\right] - \int f(x)\mathscr{L}(dx)$$

converge to zero uniformly for all Borel measurable f such that $|f| \le 1$ and for all bounded sequences $\{(g_n, [H_n])\}$.

PROOF. First, note that (C2) entails the LAN condition as was noted in Bickel (1982). Second, with the notation $V_{nH}^Z = V_{nH} + V_{nH}ZV/\sqrt{n}$, Z is symmetric, so that $V_{nH}^{-1}V_{nH}^Z = G_{nZ}$, we have by (8) and (9) that

$$\mathscr{L}_{\mu_{ng}, V_{nH}}\left(U_n(X_1,\dots,X_n) - \mu, S_n^{1/2}\right)$$
$$= \mathscr{L}_{\mu_{n(1_{g+e_n})}, V_{nH}^Z}\left(U_n\left(G_{nZ}(X_1-\mu) - \frac{e}{\sqrt{n}},\dots,G_{nZ}(X_n-\mu) - \frac{e}{\sqrt{n}}\right), G_{nZ}S_n^{1/2}\right),$$

for all g, e, H and Z, where $e_n = G_{nZ}^{-1}e$. Note that $G_{nZ}S_n^{1/2} = S^{1/2} + ZVS_n^{1/2}/\sqrt{n}$. Further, $G_{nZ} \to I$ and $\sqrt{n}(V_{nH}^Z - V_{n(H+Z)}) \to 0$, so that, e.g., by the exponential approximation result of Le Cam [(1960), Theorem 3.1], condition (A′) is verified with $T_n = (U_n, S_n^{1/2})$. This gives the conclusion Corollary 3 for $(U_n, S_n^{1/2})$ and, hence, the required result for (U_n, S_n) can be obtained. □

REMARK. Note that in many situations, one has, in addition to (8), $U_n(BX_1,\dots,BX_n) = BU_n(X_1,\dots,X_2)$ for every p.d. B, so that (C1) is satisfied. If one is interested in the convergence of U_n alone, then (C1) is not needed. Similarly, if the convergence of S_n alone is of interest, then the location invariance in (9) is not needed.

APPENDIX

PROOF OF LEMMA 2. We present the proof only for $h_n = 0$, $n \ge 1$; for arbitrary h_n only notational changes are required. In view of the exponential approximation result of Le Cam [(1960), Theorem 3.1], one can assume without loss of generality that

$$\frac{dQ_{u,n}}{dQ_{0,n}} = C_n(u)\exp\left(u'W_n^* - \frac{|u|^2}{2}\right), \tag{10}$$

where the functions $u \to C_n(u)$ are Borel measurable such that

$$\sup_{|u| \le \alpha} |C_n(u) - 1| \to 0 \tag{11}$$

for all $\alpha > 0$ and the differences $W_n - W_n^*$ converge to zero in $Q_{u,n}$-probability for all $u \in R^k$. (This result also entails that the measurability of the functions $u \to Q_{u,n}$ involved in the statement of Lemma 2 is not a restriction.) One can also easily check that, since $S(x)$ is just a Gaussian density, in the r.h.s. of the difference (1) the variables W_n can be replaced by W_n^*. Therefore, one can, without loss of generality, take W_n and W_n^* to be the same. With this observation and by (10) and (11), the l.h.s. of (1) of Lemma 2 can be approximated uniformly in g by (for $h_n = 0$)

$$\frac{1}{\mu(D_\alpha)} \int_{D_\alpha} \int \left\{ \frac{\int_{D_\alpha} g(v)\exp(v'W_n - |v|^2/2)\, dv}{\int_{D_\alpha} \exp(v'W_n - |v|^2/2)\, dv} \right\} dQ_{u,n}\, du. \tag{12}$$

Now rewrite the ratio inside the brackets of (12) as

$$\int_{D_\alpha} g(v) S(v - W_n)\, dv \Big/ \int_{D_\alpha} S(v - W_n)\, dv = a/b, \quad \text{say.}$$

Let

$$\int_{R^k} g(v) S(v - W_n)\, dv = a + c, \quad \text{say.}$$

Then, since $|g| \le 1$, $|a| \le b$ and

$$|c| \le 1 - b = \int_{D_\alpha^c} S(v - W_n)\, dv = d, \quad \text{say,}$$

so that

$$|(a/b) - (a + c)| = |(a/b) - (a + c)/(b + d)| \le 2d.$$

Thus the difference between (12) and the r.h.s. of (1) is bounded in absolute value by

$$\frac{2}{\mu(D_\alpha)} \int_{D_\alpha} \int \int_{D_\alpha^c} S(v - W_n)\, dv\, dQ_{u,n}\, du.$$

Using the LAN condition and the resulting contiguity [cf. Theorem 2.1 of Le Cam (1960)], this quantity converges, as $n \to \infty$, to

$$\begin{aligned} &\frac{2}{\mu(D_\alpha)} \int_{D_\alpha} \int_{R^k} \int_{D_\alpha^c} S(v - w)\, dv\, S(u - w)\, dw\, du \\ &= \frac{2}{\mu(D_\alpha)} \int_{D_\alpha} \mathcal{N} * \mathcal{N}(D_\alpha^c - u)\, du, \end{aligned} \tag{13}$$

where $\mathscr{N}$ denotes the Gaussian measure with mean 0 and covariance matrix I, $*$ denotes the convolution and $D_\alpha^c - u = \{h - u\colon\ h \in D_\alpha^c\}$. Now if $u \in D_{\alpha - \sqrt{\alpha}}$, then $D_{\sqrt{\alpha}} \subseteq D_\alpha - u$ so that for any probability measure P on R^k,

$$\frac{1}{\mu(D_\alpha)} \int_{D_\alpha} P(D_\alpha^c - u)\, du = 1 - \frac{1}{\mu(D_\alpha)} \int_{D_\alpha} P(D_\alpha - u)\, du$$

$$\leq 1 - \frac{P(D_{\sqrt{\alpha}})\mu(D_{\alpha - \sqrt{\alpha}})}{\mu(D_\alpha)} \to 0 \quad \text{as } \alpha \to \infty.$$

Thus, by taking $P = \mathscr{N} * \mathscr{N}$, (13) converges to zero as $\alpha \to \infty$. This completes the proof of the lemma. □

REFERENCES

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

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

BOOS, D. D. (1983). On convergence of densities of translation and scale statistics. Institute of Statistics Mimeo Series No. 1625, North Carolina State Univ.

BOOS, D. D. (1985). A converse to Scheffé's theorem. *Ann. Statist.* **13** 423–427.

DAVIES, R. B. (1985). Asymptotic inference when the amount of information is random. In *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 841–864. Wadsworth, Monterey, Calif.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–330.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

JEGANATHAN, P. (1986). On the strong approximation of the distributions of estimators in linear stochastic models. I and II: Stationary and explosive AR models. Unpublished.

JEGANATHAN, P. (1987). Strong convergence of distributions of L.S. estimators in AR time series with roots near the unit circle. Unpublished.

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

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

LE CAM, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory*. Univ. Montreal Press, Montreal.

LE CAM, L. (1975). Distances between experiments. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 383–396. North-Holland, Amsterdam.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

STRASSER, H. L. (1978). Global asymptotic properties of risk functions in estimation. *Z. Wahrsch. verw. Gebiete* **45** 35–48.

TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.

DEPARTMENT OF STATISTICS
1444 MASON HALL
UNIVERSITY OF MICHIGAN
419 SOUTH STATE STREET
ANN ARBOR, MICHIGAN 48109-1027