LOCAL ASYMPTOTIC MINIMAX AND ADMISSIBILITY IN ESTIMATION

JAROSLAV HÁJEK
FLORIDA STATE UNIVERSITY
and
CHARLES UNIVERSITY, PRAGUE

1. Introduction

In their vigorous search for an adequate asymptotic theory of estimation, statisticians have tried almost all their methodological tools: prior densities, minimax, admissibility, large deviations, restricted classes of estimates (invariant, unbiased), contiguity, and so forth. The resulting body of knowledge is somewhat atomized and a certain synthetic work seems to be needed. In this section, let us try to single out several pieces of the jigsaw puzzle and combine them into a logically connected theory. Along with this we shall criticize some other approaches and make a few historical remarks.

Consider a fixed parametric space Θ and a sequence of experiments described by families of densities $p_n(x_n, \theta)$, say $p_n(x_n, \theta) = \prod_{i=1}^n f(y_i, \theta)$, where $x_n = (y_1, \cdots, y_n)$. First of all, it is necessary to single out "regular cases." This should not be done only formally, for example, only in terms of θ derivatives of $f(y, \theta)$. The statistical essence of regularity consists in the possibility of replacing the family of distributions by a normal family in a local asymptotic sense. Loosely speaking, given a point $t \in \Theta$ and a small vicinity V_t of t, the quantity

$$\Delta_{n,t} = n^{-1/2} \frac{\partial}{\partial \theta} \log p_n(x_n, \theta)\Big|_{\theta = t} \tag{1.1}$$

should be approximately sufficient and normal (with constant covariance and expectation linear in θ) for $\theta \in V_t$ and n large. See Section 3 for a precise definition. The idea of approximating a general family by a normal family was first formulated by A. Wald [19], and then sophisticatedly developed by L. LeCam [12], [13], [15]. In spite of its importance, the idea has not yet found its way into current textbooks.

The next step is to get rid of ill behaved estimates and to characterize optimum ones. This may be achieved by scrutinizing an arbitrary sequence of estimates T_n from the point of view of minimax and admissibility, again in a *local* asymptotic sense. Theorem 4.1 below entails that there is a lower bound for asymptotic local maximum risk and that this bound may be achieved only if

Research sponsored in part by National Science Foundation Grant # GU 2612 to the Florida State University and in part by the Army, Navy and Air Force under Office of Naval Research Contract Number NONR 988(08), Task Order NR 042-004.

$$\left[\sqrt{n}\,(T_n - t) - \frac{\Lambda_{n,t}}{\Gamma_t}\right] \to 0. \qquad [p_{n,t}] \tag{1.2}$$

where Γ_t does not depend on the observation nor on the choice of $\{T_n\}$.

Once the condition (1.2) has been established, we can develop particular methods providing estimates satisfying (1.2) for every $t \in \theta$. These methods would include Bayes estimates with respect to diffuse priors, maximum likelihood estimates, maximum probability estimates, RBAN estimates, "nonparametric" estimates, and so forth. The justification for a diffuse prior in Bayes estimation—and probably the only sound one—is that the resulting estimates satisfy (1.2), and, under additional conditions, have satisfactory local asymptotic minimax properties. Minimax scrutiny of Bayes estimates seems to be even more important in cases when we estimate a scalar parametric function $\tau = \tau(\theta)$ of a parameter whose dimensionality increases with n. Then it may be far from obvious what a "diffuse" prior means. Unfortunately local asymptotic minimax results are not systematically available for this case.

In maximum likelihood and Bayes estimation, we are troubled by the behavior of likelihood tails. This may be avoided by considering only some vicinity of a consistent estimate. This leads to consistent estimates considered as a starting point for arriving at better estimates.

If the families $p_n(\cdot, \theta)$ are well behaved, then $\mathscr{L}_\theta(\Delta_{n,\theta})$ will be smooth in θ. In turn, the distribution of any estimate T_n for which (1.2) is satisfied will be smooth in θ. Assuming that the laws $\mathscr{L}_\theta[\sqrt{n}(T_n - \theta)]$ converge continuously in θ to some laws L_θ, it was proved in [6] that L_t may be decomposed as a convolution

$$L_t = \phi_t * G_t \tag{1.3}$$

where ϕ_t is a normal distribution and G_t is a distribution depending on the choice of T_n. It also may be shown that the best possible G_t is degenerate at the point zero. If G_t is normal, the L_t is also normal with larger covariance matrix than ϕ_t. In such a case, the ratio of the generalized variances of ϕ_t and L_t may serve as a measure of the efficiency of T_n. A simple short proof of (1.3) has been recently suggested by P. Bickel (personal communication). LeCam [16] provided still another proof and extended the result to families which may not be locally asymptotically normal. Then, of course, ϕ_t will not be a normal distribution, but, for example, the one sided exponential distribution.

Some points in the above exposition need comment. First, let us explain what we understand by local asymptotic minimax and admissibility. Roughly speaking, by saying that $\{T_n\}$ is locally asymptotically minimax (admissible) we shall mean that for any open interval $V \subset \theta$ the estimate is approximately minimax (admissible) on V if n is large. Actually, for n large there is little justification for relating the minimax criterion and admissibility to the whole parametric space, since after having obtained our observations, we are able to locate the unknown parameter with considerable precision.

It is important to note that an estimate may be exactly minimax for every n, but may fail to be *locally* asymptotically minimax. For example, this is true about the minimax estimator of the binomial p from X successes in n trials:

$$t_n(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}. \tag{1.4}$$

This estimate is obviously not approximately minimax relative to any open interval V not containing $p = 1/2$, consequently, it is not locally asymptotically minimax. Actually, the maximum mean square error of t_n on V is $[4n(1 + n^{-1/2})^2]^{-1}$, whereas the minimax on V is smaller than $n^{-1} \max_{p \in V}\{p(1 - p)\}$.

Also admissibility for all n does not entail local asymptotic admissibility. Estimate (1.4) may illustrate this point as well. Another example may be given from the field of sample surveys, where Joshi and Godambe (see [11]) proved that the sample average is an admissible estimate for the population average no matter what the sample design is. Again this estimate fails to be locally asymptotically admissible if the concept is properly defined for this situation.

According to Chernoff [5] the idea of local asymptotic minimax is due to C. Stein and H. Rubin, who showed that Fisher's programs could be rescued by it. The proof that local asymptotic minimax implies local asymptotic admissibility was first given by LeCam ([12], Theorem 14). There, he proved that superefficiency at one point entails bad risk values in the vicinity of this point, that is, that superefficiency excludes the local asymptotic minimax property. Apparently not many people have studied LeCam's paper so far as to read this very last theorem, and the present author is indebted to Professor LeCam for giving him the reference. Theorems 4.1 and 4.2 below may be regarded as extensions of the above mentioned result by LeCam and a related theorem by Huber. Huber, [10] in addition to LeCam's statement, proves that a locally asymptotically minimax estimate must be asymptotically normal with a given variance. This is made more precise by (1.2), since (1.2) entails asymptotic normality of $\sqrt{n}(T_n - t)$ on the basis of assumed asymptotic normality of $\Delta_{n,t}$. From the methodological point of view, LeCam's and the present paper both use the Blyth [4] approach to admissibility based on a normal prior, whereas Huber used the Hodges-Lehmann [9] approach based on the Cramér-Rao inequality.

In his subsequent papers [13], [15] LeCam perfected the idea of a locally asymptotically normal family. We shall base our proof on these papers, rather than on his 1953 paper [12] which contains some omissions. When approximating a general family by a normal one, the basic point is that the distance should be expressed in terms of the L_1 norm (variation) given by (3.5) below. The Lévy or Prohorov distance would not do in proving (4.8) below. Of course, we may not expect that, for example, the distribution of $\Delta_{n,t}$ approaches its normal limit in the L_1 norm. However, as was shown by LeCam [13] and as is proved here under different assumptions in Lemmas 3.2 and 3.3, it approaches in the L_1 norm a slightly deformed normal law with slightly deformed likelihood function, which is sufficient for our purposes. Another important point provided in LeCam [15] was a general definition of locally asymptotically normal families, which is not restricted to partial sequences in an infinite sequence of independent replications of some basic experiment. We take this attitude also, and our local asymptotic normality (LAN) conditions of Section 3 represent a selection from LeCam's [15] conditions DN1–7.

Another comment is invited by (1.2) in connection with the approach suggested by C. R. Rao ([17], p. 285) in his admirable book. He uses property (1.2) directly as a definition of asymptotic efficiency, which seems to be justified by the fact that the ratio of Fisher's information for T_n and for the whole observation, respectively, approaches unity under (1.2) and some additional assumptions. However, since Fisher's information is invariant under one-to-one transformations, I doubt whether giving it a central position in estimation problems can be rationally explained, even if used jointly with consistency. For example, in estimating the location parameter from n independent observations if the density is otherwise known, the maximum likelihood estimator contains all of the Fisher information (the apparently lost part of it may be "recovered"), but it is not the best estimate, not even the best location invariant estimate. The extent to which the maximum likelihood estimate lags behind the best one depends on how much the likelihoods are irregularly shaped, asymmetric, for example. Theorem 4.1 below provides a minimax justification for (1.2), which is somewhat less mystical.

To be more precise, Rao calls an estimate efficient, if (1.2) holds with Γ_t^{-1} replaced by any function β_t not involving the observations, but possibly depending on the estimate T_n. However, Theorem 4.1 implies that T_n cannot be locally asymptotically minimax, if $\beta_t \neq \Gamma_t^{-1}$. Rao is aware of bad consequences of $\beta_t \neq \Gamma_t^{-1}$ and proves in [18] that then the law of $\sqrt{n}(T_n - \theta)$ does not converge uniformly in θ, so that the limiting laws cannot be used to provide approximate confidence intervals.

Continuing our comments on approaches not embodied in the above exposition, let us mention another paper by LeCam [14]. There he proved that under certain conditions Bayes estimates provide asymptotic risks that can be improved on a set of measure zero only. Thus, if the prior is diffuse and the asymptotic risk of a Bayes estimate is continuous in θ, as it usually is, then it dominates the asymptotic risk of any other estimate having continuous asymptotic risk. A disadvantage of this approach is that it defines lower bounds in terms of certain estimates—Bayes estimates—and not in terms of intrinsic properties of the families of distributions involved. Also, a verification of assumptions entails difficult consistency investigations. On the other hand, it is the first paper of a very general scope, for example, the estimates are not assumed to be necessarily asymptotically normal. In regular cases, on which the present paper is focused, Bayes estimates for diffuse priors will satisfy (1.2), so that we get a good agreement with our previous considerations.

Investigations of estimates which are not necessarily asymptotically normal were started from a different angle by J. Wolfowitz [22]. He again defined the best possible asymptotic behavior in terms of certain special estimates—generalized maximum likelihood estimates. In order to be comparable in his sense, the estimates must be "regular," so that his approach cannot be applied to arbitrary estimates. However, for regular estimates we can under LAN conditions establish the decomposition (1.3), which provides the conclusions obtained by J. Wolfowitz more directly. As was shown by LeCam [16] similar decomposition as (1.3) may be obtained also for the "nonregular case" investigated by Weiss and Wolfowitz [20].

The property of being a generalized maximum likelihood estimate is implied by (1.2) and in the usual situation also the converse is true. The same holds about maximum probability estimators.

All of the above approaches are based on "ordinary" deviations of the estimate from the parameter, which are typically of order $n^{-1/2}$. A concept of asymptotic efficiency based on large deviations has been suggested by D. Basu [3] and R. R. Bahadur [1] and [2]: Fixing θ and some $\varepsilon > 0$, we define $\tau_n(\varepsilon, \theta)$ as the standard deviation of a normal distribution under which $|T_n - \theta| \geqq \varepsilon$ would have the same probability as under $P_{n,\theta}$. Formally

$$(1.5)\qquad P_\theta(|T_n - \theta| \geqq \varepsilon) = 2\Phi\left[-\frac{\varepsilon}{\tau_n(\varepsilon, \theta)}\right].$$

Now, under regularity conditions somewhat stronger than LAN below it may be proved that for any consistent estimate

$$(1.6)\qquad \lim_{\varepsilon \to 0} \lim_{n \to \infty} \{n\tau_n^2(\varepsilon, \theta)\} \geqq \Gamma_\theta^{-1}$$

and that the equality holds for the maximum likelihood estimator. It will generally hold also for estimates satisfying (1.2). The only trouble with this approach is that (1.2) is no longer necessary and that too many estimates satisfy equality in (1.6), for example, the "superefficient" Hodges estimator, which must clearly be rejected by the minimax and ordinary deviations point of view. However, equality in (1.6) certainly may be used as an additional requirement for good estimates. Our main proposition is partially proved by a different method in [16].

2. Normal families of distributions

Consider the estimation of $\theta \in R$ by a random variable Z which has the normal distribution $N(\theta, 1)$ given θ. We shall assume that the loss function is $l(\hat{\theta} - \theta)$, where l satisfies the following conditions:

$$(2.1)\qquad l(y) = l(|y|),$$

$$(2.2)\qquad l(y) \leqq l(z) \qquad |y| \leqq |z|.$$

$$\int_{-\infty}^{\infty} \ell(y) \exp\left\{-\tfrac{1}{2}\lambda y^2\right\} dy < \infty \qquad \lambda > 0. \tag{2.3}$$

$$\ell(0) = 0. \tag{2.4}$$

Condition (2.3) entails

$$\int_{-\infty}^{\infty} \ell(y) y^2 \exp\left\{-\tfrac{1}{2}\lambda y^2\right\} dy < \infty, \qquad \lambda > 0. \tag{2.5}$$

We shall also introduce a truncated version of ℓ:

$$\ell_a(y) = \min(\ell(y), a), \qquad 0 < a \leq \infty. \tag{2.6}$$

$$r = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell(y) \exp\left\{-\tfrac{1}{2}y^2\right\} dy. \tag{2.7}$$

$$r_a = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell_a(y) \exp\left\{-\tfrac{1}{2}y^2\right\} dy \tag{2.8}$$

and

$$r_a(b) = (2\pi)^{-1/2} \int_{-\sqrt{b}}^{\sqrt{b}} \ell_a(y) \exp\left\{-\tfrac{1}{2}y^2\right\} dy. \tag{2.9}$$

The estimator will be considered randomized, that is,

$$\hat{\theta} = \xi(Z, U) \tag{2.10}$$

where U is a randomized variable. For convenience, we shall assume that U is uniformly distributed on $(0, 1)$. As always, U is independent of Z and θ. The introduction of randomized estimates is justified since our loss function $\ell(y)$ may not be convex.

The risk function corresponding to ℓ_a and ξ will be denoted as follows:

$$R_a(\theta; \xi) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \int_0^1 \ell_a[\xi(z, u) - \theta] \exp\left\{-\tfrac{1}{2}(z - \theta)^2\right\} du\, dz. \tag{2.11}$$

The following is an extension of a result of Blyth [4] to situations involving truncation and randomization. The important feature in the lemma is the independence of the numbers a, α, b on ξ if (2.12) holds. Since the lemma follows the pattern of Blyth [4], the proof will be condensed.

LEMMA 2.1. *Under the above notations and assumptions, for any $\varepsilon > 0$ there exist positive numbers, a, b, α and a prior density $\pi(\theta)$, all depending on ε only, with the following property:*

For any randomized estimator $\xi(Z, U)$ such that

$$P(|\xi(Z, U) - Z| > \varepsilon \mid \theta = 0) > \varepsilon \tag{2.12}$$

then

$$\int_{-b}^{b} \pi(\theta) R_a(\theta; \xi)\, d\theta > r + \alpha. \tag{2.13}$$

PROOF. It suffices to show that

$$\int_{-b}^{b} \pi(\theta) R_a(\theta;\, \xi)\, d\theta > r_a(b) + 2\alpha \tag{2.14}$$

for a, b sufficiently large and for an α which is independent of a, b. Note that

$$\begin{aligned}
&(2\pi)^{-1/2} \int_{-\sqrt{b}}^{\sqrt{b}} l_a(y - \beta) \exp\{-\tfrac{1}{2} y^2\}\, dy \\
&\geqq r_a(b) + (2\pi)^{-1/2} \int_0^{\sqrt{b}} \left[\int_{-x}^{x} - \int_{-x+\beta}^{x+\beta}\right] \exp\{-\tfrac{1}{2} y^2\}\, dy\, dl_a(x) \\
&\geqq r_a(b) + \delta, \qquad \text{if} \quad |\beta| > \tfrac{1}{2}\varepsilon
\end{aligned} \tag{2.15}$$

where $\delta > 0$ depends only on ε, but not on a, b when they are sufficiently large.

Next, following the idea of Blyth [4], we shall assume θ to be distributed with density

$$\pi(\theta) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left\{-\frac{\theta^2}{2\sigma^2}\right\} \tag{2.16}$$

where σ will be appropriately chosen to depend on ε later. Then the conditional density of θ given $Z = z$ is

$$\psi(\theta|z) = \frac{(1 + \sigma^2)^{1/2}}{\sigma(2\pi)^{1/2}} \exp\left\{-\frac{1 + \sigma^2}{2\sigma^2}\left(\theta - \frac{z\sigma^2}{1 + \sigma^2}\right)^2\right\} \tag{2.17}$$

and the overall density of Z will be

$$f(z) = \frac{1}{[(1 + \sigma^2)(2\pi)]^{1/2}} \exp\left\{-\frac{z^2}{2(1 + \sigma^2)}\right\}. \tag{2.18}$$

In what follows we shall assume that

$$|z| \leqq b - \sqrt{b}. \tag{2.19}$$

Then

$$\begin{aligned}
&\int_{-b}^{b} l_a[\xi(z, u) - \theta]\, \psi(\theta|z)\, d\theta \\
&\geqq (2\pi)^{-1/2} \int_{-\sqrt{b}}^{\sqrt{b}} l_a(y) \exp\left\{-\frac{y^2(1 + \sigma^2)}{2\sigma^2}\right\} dy \\
&\geqq r_a(b) - \frac{K}{\sigma^2}
\end{aligned} \tag{2.20}$$

where K does not depend on a, b, σ^2. We have used the inequality

$$\exp\left\{-\frac{y^2(1 + \sigma^2)}{2\sigma^2}\right\} > \left[1 - \frac{y^2}{\sigma^2}\right] \exp\{-\tfrac{1}{2} y^2\}. \tag{2.21}$$

On the other hand, if

$$(2.22)\qquad |\xi(z, u) - z| > \varepsilon, \qquad |z| < M$$

we shall have

$$(2.23)\qquad \left|\xi(z, u) - \frac{z\sigma^2}{1 + \sigma^2}\right| > \tfrac{1}{2}\varepsilon \qquad \text{for} \quad (1 + \sigma^2) > \frac{2M}{\varepsilon}.$$

Consequently, in view of (2.15), we have

$$(2.24)\qquad (2\pi)^{-1/2} \int_{-b}^{b} l_a[\xi(z, u) - \theta]\psi(\theta|z)\, d\theta$$

$$\geqq (2\pi)^{-1/2} \int_{-\sqrt{b}}^{\sqrt{b}} l_a(y + \tfrac{1}{2}\varepsilon) \exp\left\{-\frac{y^2(1 + \sigma^2)}{2\sigma^2}\right\} dy$$

$$\geqq (2\pi)^{-1/2} \int_{-\sqrt{b}}^{\sqrt{b}} l_a(y + \tfrac{1}{2}\varepsilon) \exp\{-\tfrac{1}{2}y^2\}\, dy - \frac{K}{\sigma^2}$$

$$\geqq r_a(b) + \delta - \frac{K}{\sigma^2}.$$

if (2.19) and (2.23) hold.

Altogether, we have

$$(2.25)\qquad \int_{-b}^{b} \pi(\theta) R_a(\theta: \xi)\, d\theta = \int_0^1 \int_{-\infty}^{\infty} \int_{-b}^{b} l_a[\xi(z, u) - \theta]\psi(\theta|z) f(z)\, d\theta\, dz\, du$$

$$\geqq r_a(b) P(|Z| < b - \sqrt{b}) - \frac{K}{\sigma^2} + \delta P(|\xi(Z, U) - Z| > \varepsilon, |Z| < M).$$

From (2.12) we see that

$$(2.26)\qquad P(|\xi(Z, U) - Z| > \varepsilon, |Z| < M | \theta = 0) > \tfrac{1}{2}\varepsilon$$

for M sufficiently large. Now under $\theta = 0$ the density of Z is $(2\pi)^{-1/2} \exp\{-\frac{1}{2}z^2\}$, whereas the overall density is given by $f(z)$ of (2.18). The likelihood ratio for the two densities is for $|z| < M$ greater than

$$(2.27)\qquad \frac{1}{(1 + \sigma^2)^{1/2}} \exp\{-\tfrac{1}{2}M^2\}.$$

Thus

$$(2.28)\qquad P(|\xi(Z, U) - Z| > \varepsilon, |Z| < M) > \frac{\varepsilon}{2(1 + \sigma^2)^{1/2}} \exp\{-\tfrac{1}{2}M^2\}.$$

Now it suffices to put

$$(2.29)\qquad 3\alpha = \frac{\delta\varepsilon}{2(1 + \sigma^2)^{1/2}} \exp\{-\tfrac{1}{2}M^2\} - \frac{K}{\sigma^2}$$

which is positive for σ sufficiently large. In the last step we choose a and b in such a way that

$$r_a(b) > r - \tfrac{1}{2}\alpha \tag{2.30}$$

and given the above chosen σ^2,

$$rP(|Z| > b - \sqrt{b}) < \tfrac{1}{2}\alpha. \tag{2.31}$$

Now (2.25) to (2.31) yield (2.14). *Q.E.D.*

Of course, usually there will be nuisance parameters, and then the following k-dimensional version of the preceding lemma is useful.

LEMMA 2.2. *Let $Z = (Z_1, \cdots, Z_k)$ be a normal random vector with expectation $\theta = (\theta_1, \cdots, \theta_k) \in R^k$ and fixed positive definite covariance matrix $\Sigma = \{\sigma_{i,j}\}_{i,j=1}^k$. Consider a function ℓ satisfying* (2.1) *to* (2.4). *Then for every $\varepsilon > 0$ there exist positive numbers a, b, α and a* one-dimensional *density $\pi(\cdot)$ with the following property:*

For any randomized estimator $\xi(Z_1, \cdots, Z_k, U)$ of θ_1 such that

$$P(|\xi(Z_1, \cdots, Z_k, U) - Z_1| > \varepsilon | \theta_1 = \cdots = \theta_k = 0) > \varepsilon \tag{2.32}$$

then

$$\int_{-b}^{b} \pi(\theta_1) E\{\ell_a[\xi(Z_1, \cdots, Z_k, U) - \theta_1] | \theta_i = \theta_1\, \sigma_{1,i}\sigma_{1,1}^{-1}, 1 \leqq i \leqq k\}\, d\theta_1 \tag{2.33}$$

$$\geqq (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell[y\sigma_1] \exp\{-\tfrac{1}{2}y^2\}\, dy + \alpha$$

where $\sigma_1 = (\sigma_{1,1})^{1/2}$

PROOF. For the submodel $\theta_i = \theta_1 \sigma_{1,i}\sigma_{1,1}^{-1}$, $1 \leqq i \leqq k$, $-\infty < \theta_1 < \infty$, Z_1 is a sufficient statistic, and the result follows from the preceeding lemma.

LEMMA 2.3. *Let Z be the same vector as in Lemma 2.2. Then for every $\delta > 0$ there exist positive numbers a, b and a density $\pi(\cdot)$ such that for any randomized estimator $\xi(Z_1, \cdots, Z_k, U)$*

$$\int_{-b}^{b} \pi(\theta_1) E\{\ell_a[\xi(Z_1, \cdots, Z_k, U) - \theta_1] | \theta_i = \theta_1 \sigma_{1,i}\sigma_{1,1}^{-1}, 1 \leqq i \leqq k\}\, d\theta_1 \tag{2.34}$$

$$\geqq (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell[y\sigma_1] \exp\{-\tfrac{1}{2}y^2\}\, dy - \delta$$

where σ_1 has the same meaning as in Lemma 2.2.

PROOF. The proof follows the same lines as the proofs of the previous lemmas.

3. Locally asymptotically normal families of distributions

The essence of the definition of a locally asymptotically normal family is that the log likelihood ratio is asymptotically normally distributed with a covariance matrix, which is locally constant, and with an expectation which is locally a linear function of θ. For our purpose—to obtain lower bounds for risks and necessary conditions for allowing equality in corresponding inequalities—we need the following version, employed already in [6]: consider a sequence of statistical experiments $(\mathscr{X}_n, \mathscr{A}_n, P_n(\cdot, \theta))$, $n \geqq 1$, where θ runs through an open subset Θ of R^k. Take a point $t \in \Theta$ and assume it to be the true value of the parameter θ. We shall abbreviate $P_n = P_n(\cdot, t)$ and $P_{n,h} = P_n(\cdot, t + n^{-1/2}h)$. If appropriate, we could use in place of $\sqrt{n}$ some more general norming numbers $k(n) \to \infty$, or even matrices as in [6].

The norm of a point $h = (h_1, \cdots, h_k)$ from R^k will be denoted by $|h| = \max_{1 \leqq i \leqq k} |h_i|$, and $h'v$ will denote the scalar product of two vectors.

Given two probability measures P and Q, denote by dQ/dP the Radon-Nikodym derivative of the *absolutely continuous* part of Q with respect to P. Introduce the family of likelihood ratios

$$r_n(h, x_n) = \frac{dP_{n,h}}{dP_n}(x_n), \qquad h \in R^k, n \geqq n_h, x_n \in \mathscr{X}_n, \tag{3.1}$$

where n_h denotes the smallest integer such that $n \geqq n_h$ entails $t + n^{-1/2} h \in \Theta$. In what follows the argument x_n will usually be omitted.

ASSUMPTION 3.1. LAN (*local asymptotic normality*) *at* $\theta = t$. *Assume that*

$$r_n(h) = \exp\{h'\Delta_{n,t} - \tfrac{1}{2}h'\Gamma_t h + Z_n(h, t)\}, \qquad h \in R^k, n \geqq n_h, \tag{3.2}$$

where the random vector $\Delta_{n,t}$ *satisfies* $\mathscr{L}(\Delta_{n,t}|P_n) \to N(0, \Gamma_t)$, *and* $Z_n(h, t) \to 0$ *in* P_n *probability for every* $h \in R^k$. *Further assume that* $\det \Gamma_t > 0$.

EXAMPLE. Consider the case $k = 1$, $x_n = (y_1, \cdots, y_n) \in R^n$ and $p_n(x_n, \theta) = \Pi_{i=1}^n f(y_i, \theta)$.

Then the existence of Fisher's information, its positivity and continuity at $\theta = t$, entails LAN. Of course, in order for Fisher's information to be well defined, $f(y, \theta)$ must be absolutely continuous in θ in a vicinity of t, and the derivative $\dot{f}(y, t) = (\partial/\partial\theta) f(y, \theta)|_{\theta=t}$ must exist for almost all y. Then the information equals

$$I_\theta = \int_{-\infty}^{\infty} \left\{\frac{[\dot{f}(y, \theta)]^2}{f(y, \theta)}\right\} dy. \tag{3.3}$$

Satisfaction of LAN for this case may be proved by methods developed in [8] as is shown in the Appendix below. Specifically (3.2) will be satisfied for

$$\Delta_{n,t} = n^{-1/2} \sum_{i=1}^n \frac{\dot{f}(y_i, t)}{f(y_i, t)}, \qquad \Gamma_t = I_t. \tag{3.4}$$

Let $\|P - Q\|$ be the L_1 norm of two probability measures. If p and q are densities of P and Q with respect to μ, then

$$\|P - Q\| = \int |p - q| \, d\mu. \tag{3.5}$$

The following lemmas are essentially contained in LeCam [13], and form a bridge between exactly normal models and locally asymptotically normal models.

LEMMA 3.1. *For any sequence of statistics* $\{S_n\}$ *put*

$$s_n(x, u) = \inf\{y: P_n[S_n \leqq y \mid \Delta_{n,t} = x] \geqq u\}, \qquad x \in R,\ 0 < u < 1, \tag{3.6}$$

and denote by $F_{n,h}$ *the distribution of* S_n *under* $P_{n,h}$ *and by* $F^*_{n,h}$ *the distribution of* $s_n(\Delta_{n,t}, U)$ *also under* $P_{n,h}$, *if* U *is uniformly distributed on* (0, 1) *and independent of* $\Delta_{n,t}$.

Then, under Assumptions LAN

$$\lim_{n\to\infty} \|F_{n,h} - F^*_{n,h}\| = 0, \qquad h \in R. \tag{3.7}$$

PROOF. We see from (3.6) that $F_{n,h} = F^*_{n,h}$ if $h = 0$. Now, as shown in the proof of the theorem in [6], $P_{n,h}$ may be approximated by $Q_{n,h}$ such that $\Delta_{n,t}$ is a sufficient statistic for the pair $(P_n, Q_{n,h})$ and $\|Q_{n,h} - P_{n,h}\| \to 0$. That yields (3.7).

In the next two lemmas we shall treat the cases $k = 1$ and $k > 1$ separately, because for $k = 1$ we are able to reach a greater degree of explicitness.

LEMMA 3.2. *Assume* $k = 1$. *Denote* $G_{n,h}(x) = P_{n,h}(\Delta_{n,t} \leqq x)$ *and* $\Phi(x) = (2\pi) \int_{-\infty}^{x} \exp\{-\frac{1}{2}y^2\}\, dy$. *Let* $G^*_{n,h}$ *be the distribution of* $G^{-1}_{n,0}\,\Phi(Z\Gamma_t^{-1/2})$, *if* Z *is normal* $(h\Gamma_t, \Gamma_t)$.

Then, under Assumption LAN,

$$\lim_{n\to\infty} \|G_{n,h} - G^*_{n,h}\| = 0, \qquad h \in R. \tag{3.8}$$

PROOF. We again have $G_{n,h} = G^*_{n,h}$ for $h = 0$. Since $G_{n,0}(x) \to \Phi(x\Gamma_t^{-1/2})$ uniformly under Assumptions LAN we have

$$G^{-1}_{n,0}\,\Phi(x\Gamma_t^{-1/2}) \to x \text{ uniformly on compacts.} \tag{3.9}$$

Further, referring again to the proof of the theorem of [6], we may approximate $G_{n,h}$ by $\bar{G}_{n,h}(x) = Q_{n,h}(\Delta_{n,t} \leqq x)$ satisfying $\|G_{n,h} - \bar{G}_{n,h}\| \to 0$. The rest follows by showing that

$$\frac{d\bar{G}_{n,h}}{dG_{n,0}}(x) \to \exp\{-hx - \tfrac{1}{2}h^2\Gamma_t\}, \tag{3.10}$$

$$\frac{dG^*_{n,h}}{dG_{n,0}}(x) \to \exp\{-hx - \tfrac{1}{2}h^2\Gamma_t\}. \tag{3.11}$$

LEMMA 3.3. *Assume* $k \geqq 1$. *Denote* $G_{n,h}(x) = P_{n,h}(\Delta_{n,t} \leqq x)$, $x \in R^k$, *and denote by* $Z = (Z_1, \cdots, Z_k)$ *a random vector such that* $\mathscr{L}(Z) = N(\Gamma_t h, \Gamma_t)$, *if* $\theta = t + n^{-1/2}h$. *Then there exists a sequence of functions* $\phi_n(x)$ *such that*

$$\lim_{n\to\infty} \sup_{x\in R^k} |\phi_n(x) - x| = 0 \tag{3.12}$$

and the distribution of $\phi_n(Z)$ *under* $\theta = t + n^{-1/2}h$, *say* $G^*_{n,h}$, *satisfies*

$$\lim_{n\to\infty} \|G^*_{n,h} - G_{n,h}\| = 0. \tag{3.13}$$

PROOF. Consider a sequence of cubes

$$C_j = \{(x_n, \cdots, x_k): |x_i| \leqq j, 1 \leqq i \leqq k\}, \qquad j = 1, 2, \cdots. \tag{3.14}$$

Partition each cube C_j into j^{2k} subcubes $C_{j,i}$, $1 \leqq i \leqq j^{2k}$, all of equal volume $(2/j)^k$. Since $G_{n,0} \to N(0, \Gamma_t)$, we have

$$\int_{C_{j,i}} dG_{n,0} \to \int_{C_{j,i}} dN(0, \Gamma_t). \tag{3.15}$$

Let $\phi_{n,j}$ be a function mapping $C_{j,i}$ into $C_{j,i}$,

$$\phi_{n,j}: C_{j,i} \to C_{j,i}, \qquad 1 \leqq j \leqq \infty, 1 \leqq i \leqq j^{2k} \tag{3.16}$$

and such that

$$\phi_{n,j}(x) = x \quad \text{for} \quad x \notin C_j. \tag{3.17}$$

Furthermore, we choose $\phi_{n,j}$ so that the difference between the $G^*_{n,j,0} = \mathscr{L}(\phi_{n,j}(Z)|N(0,\Gamma_t))$ and $G_{n,0}$ tends to zero in the L_1 norm, that is,

$$\| G_{n,j,0} - G_{n,0} \| < \frac{1}{j} \qquad \text{as} \quad n \geqq n_j. \tag{3.18}$$

This is possible in view of (3.15). Now define $j(n)$ by $n_{j(n)} \leqq n < n_{j(n)+1}$ and put

$$\phi_n(x) = \phi_{n,j(n)}(x). \tag{3.19}$$

Then $\phi_n(x)$ satisfies (3.12) because of (3.17) and (3.16) and the fact that diameters of the cubes $C_{j,i}$ converge to 0 as $j \to \infty$. Furthermore, (3.18) entails $\| G^*_{n,0} - G_{n,0} \| \to 0$ since $G^*_{n,0} = G^*_{n,j(n),0}$.

The proof may then be concluded by showing that (3.10) holds, referring again to the theorem of [6], and that (3.11) is true. The last statement follows from (3.12).

4. The main proposition

We shall generalize and modify theorems by LeCam [12] and P. Huber [10], treating the cases $k = 1$ and $k \geqq 1$ separately, again.

THEOREM 4.1. *Assume $k = 1$. Under Assumptions* LAN *of Section* 3, *any sequence of estimates $\{T_n\}$ for θ satisfies for ℓ of* (2.1) *to* (2.4)

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{|\theta - t| < \delta} E_\theta\{\ell[\sqrt{n}(T_n - \theta)]\} \geqq (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell(y\Gamma_t^{-1/2}) \exp\{-\tfrac{1}{2}y^2\}\, dy. \tag{4.1}$$

Furthermore, we can have for a nonconstant ℓ

$$\lim_{\delta - 0} \lim_{n\to\infty} \sup_{|\theta - t| < \delta} E_\theta\{\ell[\sqrt{n}(T_n - \theta)]\} = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell(y\Gamma_t^{-1/2}) \exp\{-\tfrac{1}{2}y^2\}\, dy \tag{4.2}$$

only if

(4.3) $$\sqrt{n}(T_n - t) - \Gamma_t^{-1}\Delta_{n,t} \to 0$$

in P_n probability.

PROOF. We shall first prove (4.1). Introducing local coordinates by $\theta = t + hn^{-1/2}$ and using $E_\theta(\cdot)$ and $E(\cdot|\theta)$ interchangeably, we may write for n sufficiently large

(4.4) $$\sup_{|\theta - t| < \delta} E_\theta\{\ell[\sqrt{n}(T_n - \theta)]\} \geqq \int_{-b}^{b} \pi(h) E\{\ell_a[\sqrt{n}(T_n - t) - h]|t + n^{-1/2}h\}\, dh,$$

whatever the constants a, b and density $\pi(\cdot)$ may be. We fix some $\bar{\delta} > 0$ and choose a, b and π in such a way that

(4.5) $$\int_{-b}^{b} \pi(h) E\{\ell_a[\xi(Z, U) - h]|t + n^{-1/2}h\}\, dh \geqq (2\pi)^{-1/2} \int_{-\infty}^{\infty} \ell(y\Gamma_t^{-1/2}) \exp\{-\tfrac{1}{2}y^2\}\, dy - \bar{\delta}$$

for any estimator $\xi(Z, U)$, provided that $\mathscr{L}(Z|t + n^{-1/2}h) = N(\Gamma_t h, \Gamma_t)$. This is possible according to Lemma 2.3, if applied to $\Gamma_t^{-1} Z$ which is normal (h, Γ_t^{-1}).

Next we identify $S_n = \lambda_n(T_n - t)$ in Lemma 3.1 and conclude (ℓ_a is bounded!) that for every $h \in R$

(4.6) $$|E\{\ell_a[\sqrt{n}(T_n - t) - h]|t + n^{-1/2}h\} - E\{\ell_a[s_n(\Delta_{n,t}, U) - h]|t + n^{-1/2}h\}| \to 0.$$

Furthermore, by Lemma 3.2, putting

(4.7) $$\xi_n(Z, U) = s_n(G_{n,0}^{-1}\phi(Z\Gamma_t^{-1/2}), U)$$

we obtain for every $h \in R$

(4.8) $$|E\{\ell_a[s_n(\Delta_{n,t}, U) - h]|t + n^{-1/2}h\} - E\{\ell_a[\xi_n(Z, U) - h]|t + n^{-1/2}h]| \to 0.$$

Consequently

(4.9) $$\int_{-b}^{b} \pi(h) E\{\ell_a[\sqrt{n}(T_n - t) - h]|t + n^{-1/2}h\}\, dh \geqq \int_{-b}^{b} \pi(h) E\{\ell_a[\xi_n(Z, U) - h]|t + n^{-1/2}h\}\, dh - \bar{\delta}, \quad n > n(a, b, \pi, \bar{\delta}).$$

Combining (4.4), (4.9), and (4.5), we obtain (4.1).

Now we shall prove the necessity of (4.3) for (4.2). Assuming that (4.3) does not hold we shall contradict (4.2). Again putting $S_n = \lambda_n(T_n - t)$, and recalling Lemma 3.1, we can see that $\lambda_n(T_n - t) - \Gamma_t^{-1}\Delta_{n,t}$ has the same distribution under P_n as

$$\xi_n(\Delta_{n,t}, U) - \Gamma_t^{-1}\Delta_{n,t}. \tag{4.10}$$

Furthermore, in view of (3.9), if expression (4.10) fails to converge to zero in probability, then also

$$\xi_n(Z, U) - \Gamma_t^{-1}Z \tag{4.11}$$

fails to do so. But then, there is an $\varepsilon > 0$ such that for every n there exists an $m > n$ such that

$$P_m(|\xi_m(Z, U) - \Gamma_t^{-1}Z| > \varepsilon) > \varepsilon. \tag{4.12}$$

Therefore, according to Lemma 2.2, applied to $\Gamma_t^{-1}Z$ and $k = 1$, we choose a, b, α, and $\pi(\cdot)$ such that

$$\int_{-b}^{b} \pi(h)E\{\ell_a[\xi_m(Z, U) - h]|t + n^{-1/2}h\}\, dh > (2\pi)^{-1/2}\int_{-\infty}^{\infty} \ell(y\Gamma_t^{-1/2}) \exp\{-\tfrac{1}{2}y^2\}\, dy + \alpha. \tag{4.13}$$

This in connection with (4.4) and (4.9) contradicts (4.2), since $\bar{\delta}$ can be made smaller than α. *Q.E.D.*

Without proof let us also formulate a k-dimensional version of the previous theorem. The parametric function of interest will be the first coordinate, θ_1, while the other coordinates will be regarded as nuisance parameters. The proof could be based on Lemma 2.2.

THEOREM 4.2. *Under Assumptions* LAN *above, any sequence of estimates T_n for the first coordinate θ_1 of θ satisfies*

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{|\theta - t| < \delta} E_\theta\{\ell[\sqrt{n}(T_n - \theta_1)]\} \geqq (2\pi)^{-1/2}\int_{-\infty}^{\infty} \ell(\sigma_1^t y) \exp\{-\tfrac{1}{2}y^2\}\, dy \tag{4.14}$$

where $\sigma_1^t = (\sigma_{t,1,1})^{1/2}$ and $\{\sigma_{t,i,j}\}_{i,j=1}^k = \Gamma_t^{-1}$.

We can have

$$\lim_{\delta\to 0} \lim_{n\to\infty} \sup_{|\theta - t| < \delta} E_\theta\{\ell[\sqrt{n}(T_n - \theta_1)]\} = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \ell(\sigma_1^t y) \exp\{-\tfrac{1}{2}y^2\}\, dy \tag{4.15}$$

only if

$$\{\sqrt{n}(T_n - t_1) - (\Gamma_t^{-1}\Delta_{n,t})_1\} \to 0 \tag{4.16}$$

in P_n probability, where $(\cdot)_1$ denotes the first coordinate in the corresponding vector.

REMARK 1. In (4.1) we could replace the left side by

$$\lim_{a\to\infty} \liminf_{n\to\infty} \sup_{\sqrt{n}|\theta - t| < a} E_\theta\{\ell[\sqrt{n}T_n - \theta)]\}. \tag{4.17}$$

REMARK 2. (4.16) in connection with asymptotic normality entails that $\sqrt{n}(T_n - t_1)$ is asymptotically normal $N(0, \sigma_{t,1,1})$, as is proved in Huber [9].

REMARK 3. Condition (4.16) is only necessary. In order that there exists an estimate for which (4.15) holds for all $t \in \Theta$, additional global as well as local conditions on the underlying family of distributions are necessary. If ℓ is not bounded, we also need to know something about how fast probabilities of moderate deviations of T_n from t approach zero.

REMARK 4. If we are interested in a general scalar function $\sigma = \tau(\theta)$, we may reduce the problem to one considered in Theorem 4.2 by introducing new local coordinates $(\tau_1, \cdots, \tau_k)$ such that $\tau_1 \equiv \tau$.

◇ ◇ ◇ ◇ ◇

APPENDIX

We shall here prove that the LAN conditions are satisfied under assumptions of Example in Section 3. To this aim it will be sufficient to adapt Theorem VI.2.1 of [8] for the present situation (see also Problem 7 of Chapter VI l.c.).

Let us restate our conditions carefully:

A.1. In some vicinity of $\theta = t$ the functions $f(y, \theta)$ are absolutely continuous in θ for all $y \in R$.

A.2. For every θ in some vicinity of t the θ derivative $\dot{f}(y, \theta) = (\partial/\partial\theta)f(y, \theta)$ exists for almost all (Lebesgue measure) $y \in R$.

A.3. The Fisher information

$$I_\theta = \int_{-\infty}^{\infty} \left\{\frac{[\dot{f}(y, \theta)]^2}{f(y, \theta)}\right\} dy \tag{A.1}$$

exists, is continuous at $\theta = t$ and $I_t > 0$. (The integrand in (A.1) is to be interpreted as zero if $f(y, \theta) = 0$).

Our preliminary goal is to show that

$$s(y, \theta) = [f(y, \theta)]^{1/2} \tag{A.2}$$

is also absolutely continuous and has a mean square derivative at $\theta = t$.

LEMMA A.1. *If $g(\theta) \geqq 0$ is absolutely continuous on (a, b) and its derivative $\dot{g}(\theta)$ satisfies*

$$\int_a^b \frac{|\dot{g}(\theta)|}{[g(\theta)]^{1/2}} d\theta < \infty, \tag{A.3}$$

then $[g(\theta)]^{1/2}$ is also absolutely continuous on (a, b).

PROOF. If $g(\theta) > 0$ and $\dot{g}(\theta)$ exists for some point θ, then it is well known from calculus that

$$\frac{\partial}{\partial \theta}[g(\theta)]^{1/2} = \frac{\dot{g}(\theta)}{2[g(\theta)]^{1/2}}. \tag{A.4}$$

Furthermore, if $a \leqq \alpha < \beta \leqq b$ and g is positive on $[\alpha, \beta]$, it is easy to see that $[g(\alpha)]^{1/2}$ is absolutely continuous on $[\alpha, \beta]$ and

$$[g(\beta)]^{1/2} - [g(\alpha)]^{1/2} = \frac{1}{2}\int_{\alpha}^{\beta} \frac{\dot{g}(u)}{[g(u)]^{1/2}}\, du. \tag{A.5}$$

In view of the continuity of $g(\theta)$ and in view of (A.3), (A.5) extends to $\alpha < \beta$ such that g is positive on (α, β), that is, even if possibly $g(\alpha) = 0$ or $g(\beta) = 0$. Now for any c, $a < c \leqq b$ the interval (a, c) may be decomposed as follows:

$$(a, c) = \left[\bigcup_{i=1}^{\infty} (\alpha_i, \beta_i)\right] \bigcup A \tag{A.6}$$

where (α_i, β_i) are disjoint intervals such that $g(\theta)$ is positive on them, $g(\alpha_i) = 0$ if $\alpha_i \neq a$ and $g(\beta_i) = 0$ if $\beta_i \neq c$, and $g(\theta) = 0$ for $\theta \in A$. Then, interpreting $\dot{g}(\theta)[g(\theta)]^{-1/2}$ as zero when $g(\theta) = 0$, we may write, in view of (A.3),

$$\int_a^c \frac{\dot{g}(\theta)}{[g(\theta)]^{1/2}}\, d\theta = \sum_{i=1}^{\infty} \int_{\alpha_i}^{\beta_i} \frac{\dot{g}(\theta)}{[g(\theta)]^{1/2}}\, d\theta = [g(c)]^{1/2} - [g(a)]^{1/2}, \tag{A.7}$$

because the summands with endpoints satisfying $0 = g(\alpha_i) = g(\beta_i)$ vanish according to (A.5). Relation (A.7) holding for all $c \in (a, b)$ proves absolute continuity.

LEMMA A.2. *Under Assumptions* A.1–A.3 *the functions $s(y, \theta)$ are absolutely continuous in some vicinity of $\theta = t$ for almost all y.*

PROOF. Continuity of I_θ and the Fubini theorem imply for some $\varepsilon > 0$

$$\infty > \int_{t-\varepsilon}^{t+\varepsilon} I_\theta d\theta = \int_{-\infty}^{\infty} \int_{t-\varepsilon}^{t+\varepsilon} \left(\frac{[\dot{f}(y, \theta)]^2}{f(y, \theta)}\right) d\theta dy. \tag{A.8}$$

Consequently, for almost all y

$$\int_{t-\varepsilon}^{t+\varepsilon} \left(\frac{[\dot{f}(y, \theta)]^2}{f(y, \theta)}\right) d\theta < \infty \tag{A.9}$$

and, in turn, for almost all y,

$$\int_{t-\varepsilon}^{t+\varepsilon} \frac{|\dot{f}(y, \theta)|}{[f(y, \theta)]^{1/2}}\, d\theta < \infty. \tag{A.10}$$

Thus it suffices to apply Lemma A.1 for $g(\theta) = f(y, \theta)$, for those y that satisfy (A.10).

LEMMA A.3. *Under A.1–A.3 the function $\dot{s}(y, t)$ defined by*

$$\dot{s}(y,t) = \begin{cases} \dfrac{\dot{f}(y,t)}{2[f(y,t)]^{1/2}} & \text{if } f(y,t) > 0 \text{ and } \dot{f}(y,t) \text{ exists} \\ 0 & \text{otherwise} \end{cases} \tag{A.11}$$

is the mean square derivative of $s(y, \theta)$ at $\theta = t$, that is,

$$\lim_{\Delta \to 0} \int_{-\infty}^{\infty} \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] - \dot{s}(y, t) \right\}^2 dy = 0. \tag{A.12}$$

PROOF. Lemma A.2 entails

$$\begin{aligned} \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] \right\}^2 &= \left(\frac{1}{\Delta}\right)^2 \left(\int_0^{\Delta} \dot{s}(y, t + \lambda)\, d\lambda \right)^2 \\ &\leqq \frac{1}{\Delta} \int_0^{\Delta} [\dot{s}(y, t + \lambda)]^2\, d\lambda. \end{aligned} \tag{A.13}$$

Consequently, in view of continuity of I_θ,

$$\begin{aligned} \int_{-\infty}^{\infty} \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] \right\}^2 dy &\leqq \frac{1}{\Delta} \int_{-\infty}^{\infty} \int_0^{\Delta} [\dot{s}(y, t + \lambda)]^2\, d\lambda dy \\ &= \frac{1}{4} \frac{1}{\Delta} \int_0^{\Delta} I_{t+\lambda}\, d\lambda \to \frac{1}{4} I_t \\ &= \int_{-\infty}^{\infty} [\dot{s}(y, t)]^2\, dy, \quad \text{as } \Delta \to 0. \end{aligned} \tag{A.14}$$

Put $M = \{y : s(y, t) > 0\}$. Then $(1/\Delta)[s(y, t + \Delta) - s(y, t)]$ converges to $\dot{s}(y, t)$ almost everywhere (Lebesgue measure) on M, and (A.14) entails

$$\begin{aligned} \limsup_{\Delta \to 0} \int_M \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] \right\}^2 dy &\leqq \int_{-\infty}^{\infty} [\dot{s}(y, t)]^2\, dy \\ &= \int_M [\dot{s}(y, t)]^2\, dy. \end{aligned} \tag{A.15}$$

Utilizing Theorem V.1.3 of [8], we conclude that

$$\lim_{\Delta \to 0} \int_M \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] - \dot{s}(y, t) \right\}^2 dy = 0 \tag{A.16}$$

and

$$\lim_{\Delta \to 0} \int_M \left\{ \frac{1}{\Delta} [s(y, t + \Delta) - s(y, t)] \right\}^2 dy = \int_{-\infty}^{\infty} [\dot{s}(y, t)]^2\, dy. \tag{A.17}$$

However (A.17) and (A.14) are compatible only if

$$\lim_{\Delta\to 0}\int_{M^c}\left\{\frac{1}{\Delta}[s(y, t+\Delta) - s(y, t)]\right\}^2 dy = 0 \tag{A.18}$$

where $M^c = R - M$. Now (A.17) and (A.18) together are equivalent to (A.12) if $\dot{s}(y, t)$ is defined by (A.11).

REMARK A.1. Since $s(y, t) = 0$ on M^c, (A.18) is equivalent to

$$\int_{\{y: f(y, t) = 0\}} f(y, \theta)\, dy = o[(\theta - t)^2] \tag{A.19}$$

which describes how large the singular part of $f(y, \theta)$ relative to $f(y, t)$ may be. Satisfaction of (A.18) is necessary for $q_n(x) = \prod_{i=1}^n f(y_i, t + n^{-1/2}h)$ to be contiguous with respect to $p_n(x) = \prod_{i=1}^n f(y_i, t)$.

THEOREM A.4. *Under* A.1–A.3 *the* LAN *conditions defined in Section* 3 *are satisfied at* $\theta = t$ *with* $\Gamma_t = I_t$ *and* $\Delta_{n,t}$ *of* (3.4).

PROOF. Introduce

$$\begin{aligned} L_{n,h} &= \sum_{i=1}^n \log \frac{f(Y_i, t + n^{-1/2}h)}{f(Y_i, t)} \\ W_{n,h} &= 2\sum_{i=1}^n \left\{\frac{s(Y_i, t + n^{-1/2}h)}{s(Y_i, t)} - 1\right\}. \end{aligned} \tag{A.20}$$

We need to prove that for every $h \in R$

$$(L_{n,h} - h\Delta_{n,t} + \tfrac{1}{2}h^2 I_t) \xrightarrow{P_n} 0 \tag{A.21}$$

in P_n probability, P_n referring to $\theta = t$. According to LeCam's second lemma in [8] (A.21) is equivalent to

$$(W_{n,h} - h\Delta_{n,t} + \tfrac{1}{4}h^2 I_t) \xrightarrow{P_n} 0 \tag{A.22}$$

if we show that $\mathscr{L}(W_{n,h}|P_n) \to N(-\frac{1}{4}h^2 I_t, h^2 I_t)$. All this will be accomplished, if we prove the following relations:

$$E(\Delta_{n,t}|P_n) = 0, \tag{A.23}$$

$$E(W_{n,h}|P_n) \to -\tfrac{1}{4}h^2 I_t, \tag{A.24}$$

$$\operatorname{Var}(W_{n,h} - h\Delta_{n,t}|P_n) \to 0, \tag{A.25}$$

$$\mathscr{L}(\Delta_{n,t}|P_n) \to N(0, I_t). \tag{A.26}$$

We shall start with (A.23). We have

$$0 = \int_{-\infty}^{\infty} \frac{1}{\Delta}[f(y, t+\Delta) - f(y, t)\, dy \tag{A.27}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\Delta}[s^2(y, t+\Delta) - s^2(y, t)]\,dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\Delta}[s(y, t+\Delta) - s(y, t)]^2\,dy$$

$$+ 2\int_{-\infty}^{\infty} \frac{1}{\Delta}[s(y, t+\Delta) - s(y, t)]s(y, t)\,dy$$

$$\to 0 + 2\int_{-\infty}^{\infty} s(y, t)\dot{s}(y, t)\,dy = \int_{-\infty}^{\infty} \dot{f}(y, t)\,dy$$

as $\Delta \to 0$.

The last statement follows from (A.11) and (A.12). Consequently

(A.28)
$$E(\Delta_{n,t}|P_n) = n^{-1/2}\prod_{i=1}^{n} E\left[\frac{\dot{f}(Y_i, t)}{f(Y_i, t)}\Big|P_n\right]$$

$$= n^{1/2}\int_{-\infty}^{\infty} \dot{f}(y, t)\,dy = 0.$$

In order to prove (A.24) and (A.25) we employ the same idea as in Lemmas VI.2.1a and VI.2.1b in [8]:

(A.29)
$$E(W_{n,h}|P_n) = 2\sum_{i=1}^{n} E\left[\frac{s(Y_i, t+n^{-1/2}h)}{s(y_i, t)} - 1\right]$$

$$= -h^2\int_{-\infty}^{\infty}\left[\frac{s(y, t+n^{-1/2}h) - s(y, t)}{n^{-1/2}h}\right]^2 dy$$

$$\to -h^2\int_{-\infty}^{\infty} [\dot{s}(y, t)]^2\,dy = -\tfrac{1}{4}h^2 I_t.$$

Furthermore,

(A.30) $\operatorname{Var}(W_{n,h} - h\Delta_{n,t})$

$$= 4\sum_{i=1}^{n}\operatorname{Var}\left[\frac{s(Y_i, t+n^{-1/2}h)}{s(Y_i, t)} - 1 - \tfrac{1}{2}n^{-1/2}h\frac{\dot{f}(Y_i, t)}{f(Y_i, t)}\right]$$

$$\leqq 4\sum_{i=1}^{n} E\left[\frac{s(Y_i, t+n^{-1/2}h)}{s(Y_i, t)} - 1 - \tfrac{1}{2}n^{-1/2}h\frac{\dot{f}(Y_i, t)}{f(Y_i, t)}\right]^2$$

$$\leqq 4h^2\int_{-\infty}^{\infty}\left[\frac{s(y, t+n^{-1/2}h) - s(y, t)}{n^{-1/2}h} - \dot{s}(y, t)\right]^2 dy \to 0.$$

Finally, (A.26) follows easily from (3.4) by the central limit theorem. *Q.E.D.*

REFERENCES

[1] R. R. BAHADUR, "On the asymptotic efficiency of tests and estimates," *Sankhyā*, Vol. 22 (1960), pp. 229–252.

[2] ———, "Rates of convergence of estimates and test statistics," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 303–324.

[3] D. BASU, "The concept of asymptotic efficiency," *Sankyā*, Vol. 17 (1956), pp. 193–196.

[4] C. R. BLYTH, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 22–42.

[5] H. CHERNOFF, "Large sample theory: Parametric case," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 1–22.

[6] J. HÁJEK, "A characterization of limiting distributions of regular estimates, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 14 (1970), pp. 323–330.

[7] ———, "Limiting properties of likelihoods and inference," *Foundations of Statistical Inference* (edited by V. P. Godambe and D. A. Sprott), Toronto, Holt, Rinehart and Winston, 1971.

[8] J. HÁJEK and Z. ŠIDÁK, *Theory of Rank Tests*, New York, Academic Press, 1967.

[9] J. L. HODGES and E. L. LEHMANN, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1950, pp. 13–22.

[10] P. HUBER, "Strict efficiency excludes superefficiency," (abstract), *Ann. Math. Statist.*, Vol. 37 (1966), p. 1425, proof unpublished but available.

[11] V. M. JOSHI, "Admissibility of estimates of the mean of a finite population," *New Developments in Survey Sampling*, New York, Wiley, 1969.

[12] L. LECAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *Univ. California Publ. Statist.*, Vol. 1 (1953), pp. 277–330.

[13] ———, "On the asymptotic theory of estimation and testing hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 129–156.

[14] ———, "Les propriétés asymptotiques des solutions de Bayes," *Publ. Inst. Statist. Univ. Paris*, Vol. 7, fasc. (3–4)(1958), pp. 17–35.

[15] ———, "Locally asymptotically normal families of distributions," *Univ. California Publ. Statist.* Vol. 3 (1960), pp. 27–98.

[16] ———, "Limits of experiments," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1971, pp. 245–261.

[17] C. R RAO, *Linear Statistical Inference and Its Applications*, New York, Wiley, 1965.

[18] ———, "Criteria of estimation in large samples," *Sankhyā Ser. A*, Vol. 25 (1963), pp. 189–206.

[19] A. WALD, "Tests of statistical hypothesis concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[20] L. WEISS and J. WOLFOWITZ, "Generalized maximum likelihood estimators, *Teor. Verojatnost. i Primenen.*, Vol. 11 (1966), pp. 68–93.

[21] ———, "Maximum probability estimators," *Ann. Inst. Statist. Math.*, Vol. 19 (1967), pp. 193–206.

[22] J. WOLFOWITZ, "Asymptotic efficiency of the maximum likelihood estimator," *Teor. Verojatnost. i Primenen.*, Vol. 10 (1965), pp. 267–281.