The Annals of Statistics
1989, Vol. 17, No. 4, 1487–1500

ON THE ASYMPTOTIC INFORMATION BOUND[1]

BY AAD VAN DER VAART

Free University, Amsterdam

This paper discusses several lower bound results for the asymptotic performance of estimators of smooth functionals in i.i.d. models. The key idea is to look at a set of local limiting distributions of an estimator sequence, rather than to impose regularity conditions, or to consider limits of maximum risk. Special attention is paid to situations where the tangent cone is not a linear space. As an example, the local asymptotic minimax risk in mixture models is computed.

1. Introduction. This paper is concerned with bounds on the asymptotic performance of sequences of estimators for smooth functionals in i.i.d. models. The best known results in this field are the convolution theorem and the local asymptotic minimax (LAM) theorem. Both theorems originate from work by Hájek (1970, 1972) and Le Cam (1972) and have been generalized to nonparametric, semiparametric and general models by—among others—Koshevnik and Levit (1976), Pfanzagl and Wefelmeyer (1982) and Begun, Hall, Huang and Wellner (1983). The convolution theorem shows that the limiting distribution of *regular* estimator sequences is the convolution of a certain normal and another distribution. The LAM theorem applies to every estimator sequence and gives a lower bound for the limit of the maximum risk over a shrinking neighbourhood of a fixed probability distribution. An attraction of the convolution theorem over the LAM theory is that it is concerned with limiting distributions, rather than limits of expectations. However, the convolution theorem applies only to a subclass of estimators. Now "reasonable" estimator sequences have "local limiting distributions." Though these need not all be equal (as required for the convolution theorem), several results can be obtained concerning the set of local limiting distributions, including a (generalized) convolution theorem.

The main results in Section 2 are formulated in a terminology adapted from Pfanzagl and Wefelmeyer (1982). In particular, bounds are expressed in terms of a tangent cone (or tangent set). Since in semiparametric theory convex, but nonlinear, tangent cones may arise naturally, special attention is given to this case. It turns out that the technique based on finding a most difficult one-dimensional subproblem now yields a too optimistic bound. As an application we compute the LAM risk at finite mixtures in Section 3. Proofs have been put in Section 4.

We use both $N_m(\tau, \Lambda)$ and $N_{\tau, \Lambda}$ to denote the m-variate normal distribution.

Received December 1986; revised January 1989.

[1]Research partially supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and ONR Contract N00014-80-C-0163.

AMS 1980 *subject classifications*. 62G20, 62F12, 62C20.

Key words and phrases. Convolution theorem, asymptotic efficiency, local asymptotic minimax risk, local asymptotic normality, tangent cone, mixture model.

2. Main results. Let $\mathscr{P}$ be a set of probability measures on a measurable space $(\mathscr{X}, \mathscr{B})$. The statistical problem is to estimate a functional $\kappa(P) \in \mathbb{R}^k$, when an i.i.d. sample $X_1, \dots, X_n$ from some $P \in \mathscr{P}$ is given.

Fix $P \in \mathscr{P}$ and let $\mathscr{P}(P)$ be a set of maps $t \to P_t$ from some interval $(0, \varepsilon) \subset \mathbb{R}$ into $\mathscr{P}$ such that

$$\int \left[t^{-1}\left(dP_t^{1/2} - dP^{1/2} \right) - \tfrac{1}{2} g \, dP^{1/2} \right]^2 \to 0 \quad \text{as } t \downarrow 0, \tag{2.1}$$

for some $g \in L_2(P)$ (the set of P-square integrable, measurable functions from $(\mathscr{X}, \mathscr{B})$ into $\mathbb{R}$; the integral means $\int [t^{-1}(p_{tt}^{1/2} - p_t^{1/2}) - \frac{1}{2} g p_t^{1/2}]^2 \, d\mu_t$, where p_{tt} and p_t are densities of P_t and P with respect to an arbitrary σ-finite measure μ_t that dominates $P_t + P$). A direct consequence of (2.1) is that $\int g \, dP = 0$.

The set of all g's (often called *scores*) thus obtained is denoted $T(P)$ and called a *tangent set*. It is customary to let $t \to P_{th}$ be in $\mathscr{P}(P)$ for every $h \geq 0$, whenever $t \to P_t$ is. Then $T(P)$ is a *tangent cone*: $hg \in T(P)$ for every $h \geq 0$, whenever $g \in T(P)$.

It will be assumed that κ is differentiable in the sense of existence of a vector-valued function $\dot{\kappa}_P \in L_2(P)^k$ such that

$$t^{-1}\left(\kappa(P_t) - \kappa(P) \right) \to \int \dot{\kappa}_P g \, dP, \quad \text{every } g \in T(P), \tag{2.2}$$

when $t \to P_t$ is the path in (2.1). The vector $\dot{\kappa}_P$ is called a gradient or influence function. Since (2.2) only specifies inner products of $\dot{\kappa}_P$ with elements of $T(P)$, a gradient is unique only up to components orthogonal to $T(P)$ (orthogonal with respect to the inner product $\langle g_1, g_2 \rangle = \int g_1 g_2 \, dP$). The (unique) gradient which is contained in $\overline{\text{lin}\, T(P)}^k$ will be denoted $\tilde{\kappa}_P$. Its $(k \times k)$ covariance matrix is written $\tilde{J}_P = \int \tilde{\kappa}_P \tilde{\kappa}_P' \, dP$.

An estimator $T_n = t_n(X_1, \dots, X_n)$ corresponds, as usual, to a measurable map t_n: $(\mathscr{X}^n, \mathscr{B}^n) \to \mathbb{R}^k$. Our interest will be in estimator sequences satisfying

$$\mathscr{L}_{P_{1/\sqrt{n}}}\left(\sqrt{n} \left(T_n - \kappa\left(P_{1/\sqrt{n}} \right) \right) \right) \to L_g, \quad \text{every } g \in T(P), \tag{2.3}$$

where $t \to P_t$ is the path in (2.1) and L_g is a probability distribution on $\mathbb{R}^k$.

This restricts the class of estimator sequences under consideration a little, but in a sense (2.3) is only slightly stronger than tightness of the sequence of laws of $\sqrt{n}(T_n - \kappa(P))$ under the fixed P. Indeed, it can be shown under this condition (and for finite-dimensional tangent sets) that any subsequence of $\{n\}$ has a further subsequence such that (2.3) holds along the subsequence (cf. the first part of the proof of Theorem 2.1). The results below next go through for the set of limiting distributions thus obtained. Moreover, if one is willing to read *vague* instead of weak convergence in (2.3) (so that L_g may be defective), most of the results can even be formulated for arbitrary estimator sequences. However, to avoid technicalities, Assumption (2.3) is made throughout the paper.

Our first result says that, given a finite-dimensional subset G of $T(P)$, the set of limiting distributions $\{L_g\colon g \in G\}$ equals the set of distributions of a randomized estimator in a normal experiment. Let V be a random element in $\mathbb{R}^m$. By a randomized estimator based on V we mean here a statistic $t(V, U)$, where U is uniform on $[0,1]$ and independent of V and t: $\mathbb{R}^m \times [0,1] \to \mathbb{R}^k$ is measurable.

THEOREM 2.1. *Let* (2.2) *and* (2.3) *hold and let* $\{g_1, \dots, g_m\}$ *be a linearly independent subset of* $T(P)$. *Then there exists a randomized estimator for* $D_{\mathbf{g}}h$ *based on* $V \sim N_m(h, \Sigma^{-1})$ (*with* Σ *known*) *such that*

$$\mathscr{L}_h\big(t(V,U) - D_{\mathbf{g}}h\big) = L_{h'\mathbf{g}}, \quad \text{every } h'\mathbf{g} \in T(P). \tag{2.4}$$

Here $\mathbf{g} = (g_1, \dots, g_m)'$, $h'\mathbf{g} = \sum_{i=1}^m h_i g_i$, $\Sigma = \int \mathbf{g}\mathbf{g}'\,dP$ *and* $D_{\mathbf{g}} = \int \tilde{\kappa}_P \mathbf{g}'\,dP$.

A consequence for the theory of lower bounds is that the set of limiting distributions $\{L_{h'\mathbf{g}}\colon h'\mathbf{g} \in T(P)\}$ cannot be "better" than the set of distributions of the "best" randomized estimator for $D_{\mathbf{g}}h$ based on $V \sim N_m(h, \Sigma^{-1})$ where h is known to lie in $\{h \in \mathbb{R}^m\colon h'\mathbf{g} \in T(P)\}$. Theorem 2.1 is closely related to Le Cam's general theory of limiting experiments [cf. Le Cam (1972, 1986)] and can also be deduced from a combination of his results. Our proof in Section 4 is direct and easy.

It would be nicest to give a (nontrivial) lower bound for the quality of every L_g separately. This, however, is impossible. As is well known, given a $g \in T(P)$, there exist estimator sequences which perform particularly well in the direction of g, and L_g may even be the perfect limiting distribution: point mass at zero. To overcome this *superefficiency* problem Hájek (1970, 1972), within the context of parametric models, introduced two devices. Either assume that $\{T_n\}$ is *regular* (i.e., all L_g are equal) or consider the maximum risk over all g. This leads to the convolution and LAM theorem, respectively. Here we proceed by considering (normal) averages of the L_g.

THEOREM 2.2 (Generalized convolution theorem). *Let* $t(V,U)$ *be as in Theorem* 2.1 *and for* $h \in \mathbb{R}^m$ *define* $L_{h'\mathbf{g}}$ *by* (2.4). *Then for every* $\tau \in \mathbb{R}^m$ *and positive definite matrix* Λ,

$$\int L_{h'\mathbf{g}}\,dN_{\tau,\Lambda}(h) = N_{0,\,D_{\mathbf{g}}(\Sigma+\Lambda^{-1})^{-1}D'_{\mathbf{g}}} * M_{\tau,\Lambda}, \tag{2.5}$$

for some probability measure $M_{\tau,\Lambda}$ *on* $\mathbb{R}^k$.

The interpretation of Theorem 2.2 is that on the average the L_g's are more spread out than a $N_{0,\,D_{\mathbf{g}}(\Sigma+\Lambda^{-1})^{-1}D'_{\mathbf{g}}}$ distribution.

To see the role of Λ, note that

$$D_{\mathbf{g}}\Sigma^{-1}D'_{\mathbf{g}} - D_{\mathbf{g}}(\Sigma + \Lambda^{-1})^{-1}D'_{\mathbf{g}} \tag{2.6}$$

$[= D_{\mathbf{g}}(\Sigma + \Sigma\Lambda\Sigma)^{-1}D'_{\mathbf{g}}]$ is nonnegative definite for all Λ and converges to zero if, e.g., $\Lambda = \lambda I$ and $\lambda \to \infty$. Thus the improper, equivariant prior on $\mathbb{R}^m$ ($\Lambda \sim \infty$) gives the largest lower bound $N_{0,\,D_{\mathbf{g}}\Sigma^{-1}D'_{\mathbf{g}}}$.

Here $D_{\mathbf{g}}\Sigma^{-1}D'_{\mathbf{g}}$ has an interesting interpretation: It is the covariance matrix $\int \tilde{\kappa}_{P,G}\tilde{\kappa}'_{P,G}\,dP$ of the vector of orthogonal projections $\tilde{\kappa}_{P,G}$ of a gradient onto $\text{lin}\{g_1, \dots, g_m\}$ [i.e., $(\tilde{\kappa}_{P,G})_i$ is the unique linear combination $\alpha'\mathbf{g}$ satisfying

$\langle (\tilde{\kappa}_{P,G})_i - \alpha' \mathbf{g}, g_j \rangle = 0,\ j = 1, \dots, m]$. Thus, since $\tilde{\kappa}_P - \tilde{\kappa}_{P,G} \perp \tilde{\kappa}_{P,G}$,

$$\tilde{J}_P - D_{\mathbf{g}} \Sigma^{-1} D'_{\mathbf{g}} \tag{2.7}$$

is nonnegative definite for every choice of $\{g_1, \dots, g_m\}$. Moreover, by the definition of $\tilde{\kappa}_P$ this difference can be made arbitrarily small by appropriate choice of $\{g_1, \dots, g_m\}$.

A convolution and LAM theorem can be obtained as corollaries of Theorems 2.1 and 2.2.

THEOREM 2.3 (Convolution theorem). *Let $T(P)$ contain an inner point as a subset of lin $T(P)$ and let* (2.2) *and* (2.3) *hold with $L_g = L$ for every $g \in T(P)$. Then*

$$L = N_{0, \tilde{J}_P} * M \tag{2.8}$$

for some probability distribution M on $\mathbb{R}^k$.

Let l: $\mathbb{R}^k \to [0, \infty)$ satisfy

$$\begin{cases} l(0) = 0, \\ l(x) = l(-x), \\ \{x\colon l(x) \le c\} \text{ is convex and closed for every } c \in \mathbb{R}. \end{cases} \tag{2.9}$$

THEOREM 2.4 (LAM theorem). *Let $T(P)$ be a convex cone and let* (2.2), (2.3) *and* (2.9) *hold. Then*

$$\begin{aligned} &\sup_G \liminf_{n\to\infty} \sup_{g\in G} E_{P_{1/\sqrt{n},g}} l\Big(\sqrt{n}\big(T_n - \kappa(P_{1/\sqrt{n},g})\big)\Big) \\ &\quad \ge \sup_{g \in T(P)} \int l\, dL_g \ge \int l\, dN_{0, \tilde{J}_P}. \end{aligned} \tag{2.10}$$

Here for every $g \in T(P)$, $t \to P_{t,g}$ is the path in (2.1) *and the supremum on the left side of* (2.10) *is taken over all finite subsets G of $T(P)$.*

The assumptions made on $T(P)$ are weaker than the usual assumption that $T(P)$ is a linear space. For parametric models the strengthening of the convolution theorem was already obtained in Droste and Wefelmeyer (1984).

As regards the LAM theorem: The first inequality in (2.10) is of course true for any $T(P)$ (and any lower semicontinuous l). Convexity of $T(P)$ is crucial for the validity of the second inequality [though of course it suffices that $T(P)$ contains a convex cone, the closed linear span of which contains lin $\tilde{\kappa}_P$]. It turns out to be difficult to compute a sharp lower bound for $\sup_{g \in T(P)} \int l\, dL_g$ for other shapes of $T(P)$, in general. In view of Theorem 2.1 the quantity

$$\sup_{\mathbf{g}} \inf_t \sup_{h:\, h'\mathbf{g} \in T(P)} E_h l\big(t(V, U) - D_{\mathbf{g}} h\big) \tag{2.11}$$

qualifies. However, even for the simple normal experiment $V \sim N(h, 1)$, $h \in [-a, a] \subset \mathbb{R}$, the minimax risk is known for small a [Casella and Strawderman (1981)], but only approximately known for $a \to \infty$ [Bickel (1981)].

For completeness, we remark that the far left side of (2.10) is larger than the far right side for *any* estimator sequence. A proof of this is (for $k > 1$) somewhat more technical due to the fact that one has to keep account of mass escaping to "various infinities." A proof in the spirit of this paper is given in van der Vaart (1988a), page 29.

The next result is an "asymptotic Cramér–Rao bound."

THEOREM 2.5. *Let $T(P)$ be a cone and let* (2.2) *and* (2.3) *hold with $\int x \, dL_g = \mu \in \mathbb{R}^k$ for every $g \in T(P)$. Suppose that the covariance matrix $\Sigma(L_0)$ of L_0 exists. Then $\Sigma(L_0) - \tilde{J}_P$ is nonnegative definite.*

Clearly, if $T(P)$ is convex, Theorems 2.3 and 2.4 yield the same lower bound, loosely speaking a $N_{0, \tilde{J}_P}$ distribution. Hence, in particular, in this case a "best regular" estimator is a LAM estimator, at least in the sense that it attains equality in the second inequality of (2.10). Note that in this case the lower bound corresponds to a "direction" $\tilde{\kappa}_P$ which may not be contained in the closure of the tangent cone. This is different from the situation considered in Begun, Hall, Huang and Wellner (1983), where the lower bound corresponds to a "hardest one-dimensional subproblem." In the language of the latter paper one might say that the *effective score* for θ is obtained by subtracting from the score for θ its projection on the linear space spanned by the nuisance scores, rather than its projection onto this set itself.

On the other hand, if $T(P)$ fails to be convex the second inequality of (2.10) may fail to hold too and, in general, a best regular estimator may not be LAM. That this may easily happen follows for the case that $k = 1$ and $l(x) = x^2$ from Theorem 2.5, which implies that the minimax quadratic risk of a regular estimator (which is larger than its asymptotic variance) is larger than $\tilde{J}_P$, under only the condition that $T(P)$ is a cone. An example where $T(P)$ is a cone, $k = 1$ and the LAM risk is only half the asymptotic variance of a best regular estimator sequence is given at the end of this section.

In many applications $T(P)$ will indeed be convex. Then the LAM property of a best regular estimator ensures that not overly much is lost by considering regular estimators only. A best regular estimator may easily be (locally) asymptotically inadmissible, though. [Here we mean to compare the asymptotic risks $\int l \, dL_g$, $g \in T(P)$, for different sequences of estimators which satisfy (2.3).] The following theorem gives sufficient conditions for it to be the only LAM estimator sequence, in which case it is clearly admissible.

THEOREM 2.6. *Let $T(P)$ contain* lin $\tilde{\kappa}_P$, *let $\tilde{J}_P$ be nonsingular and let* (2.2) *hold. Suppose that V is unique as a minimax randomized estimator for h based on $V \sim N_k(h, \tilde{J}_P)$, $h \in \mathbb{R}^k$, with respect to the loss function l. Let* (2.3) *hold.*

Then

$$\sup_{g \in T(P)} \int l \, dL_g \leq \int l \, dN_{0, \tilde{J}_P} \tag{2.12}$$

if and only if

$$\sqrt{n}\,(T_n - \kappa(P)) = n^{-1/2} \sum_{j=1}^{n} \tilde{\kappa}_P(X_j) + o_P(1). \tag{2.13}$$

Moreover, if (2.13) *holds, then* $L_g = N(0, \tilde{J}_P)$ *for every* $g \in T(P)$.

More precisely, the condition on the normal experiment is: If $t(V, U)$ is a randomized estimator with $\sup_{h \in \mathbb{R}^k} E_h l(t(V, U) - h) \leq \int l \, dN_{0, \tilde{J}_P}$, then $t(V, U) = V$ a.e. Whether this condition holds depends on the loss function. It is well known that it generally fails if the loss function depends on three coordinates or more. In this case shrinkage estimation yields examples of inadmissibility of a best regular estimator. On the other hand, the condition is usually satisfied if the loss function depends on two coordinates or less [see Blyth (1951), Stein (1956), James and Stein (1960), Brown (1966), Hájek (1972) and Brown and Fox (1974)].

Finally, if $T(P)$ is not a linear space (which typically is the case when P is—in some sense—on the boundary of $\mathscr{P}$), then there may exist estimator sequences that improve the performance of a best regular estimator for *every* loss function. For parametric models this is obvious, because one can truncate a best regular estimator into the parameter set. For semiparametric models, such as the one in the next section, this is an unresolved matter.

EXAMPLE. Let $\mathscr{P}$ be the set of $N(\theta, 1 + \sqrt{2}\,\tau)$ distributions, where (θ, τ) is known to be in $B = \{(u, v)\colon u = 0,\ v \geq 0 \text{ or } u \geq 0,\ v = 0\}$, the boundary of the positive quadrant in $\mathbb{R}^2$. The paths $t \to N(t, 1)$ and $t \to N(0, 1 + \sqrt{2}\,t)$ lead to scores $\dot{l}_\theta(x) = x$ and $\dot{l}_\tau(x) = 2^{-1/2}(x^2 - 1)$ at $N(0,1)$. Thus set $T(N(0,1)) = \{u\dot{l}_\theta + v\dot{l}_\tau\colon (u, v) \in B\}$. The functional $N(\theta, 1 + \sqrt{2}\,\tau) \to \theta + \tau$ is differentiable at $N(0,1)$. Here $\tilde{\kappa}_{N(0,1)} = \dot{l}_\theta + \dot{l}_\tau$ and $\tilde{J}_{N(0,1)} = 2$. By Theorem 2.5, the LAM quadratic risk of a regular estimator is bounded from below by 2. Application of Theorem 2.4 to either of the convex subcones corresponding to the half lines in B yields a lower bound for the LAM quadratic risk of 1. Both bounds are sharp. The first is attained by $T_n = \overline{X}_n + 2^{-1/2}(n^{-1}\sum_{j=1}^n (X_j - \overline{X}_n)^2 - 1)$ and the second by $T_n = \max\{\overline{X}_n, 2^{-1/2}(n^{-1}\sum_{j=1}^n (X_j - \overline{X}_n)^2 - 1), 0\}$ [see van der Vaart (1988a), page 42, for a proof of the latter].

Note that if B would be slightly enlarged to $\tilde{B} = \{(u, v)\colon 0 \leq u \leq \varepsilon v\}$, $\varepsilon > 0$, fixed, then the LAM risk would jump to 2, though for $\varepsilon < 1$, $\tilde{\kappa}_{N(0,1)}$ is still not contained in $T(N(0,1))$.

3. Mixtures. Let Θ be an open subset of $\mathbb{R}^k$ and let $\mathscr{H}$ be a collection of probability measures on a measurable space $(\mathscr{Z}, \mathscr{A})$. For each $(\theta, z) \in \Theta \times \mathscr{Z}$ let $h_\theta(\cdot) g_\theta(\psi_\theta(\cdot), z)$ be a probability density with respect to a σ-finite measure μ on $(\mathscr{X}, \mathscr{B})$. Assume that h_θ and ψ_θ are a measurable map in $[0, \infty)$ and some measurable space, respectively. Moreover, let $g_\theta(y, z)$ be measurable as a function of (y, z). Now let $\mathscr{P} = \{P_{\theta,\eta}: (\theta, \eta) \in \Theta \times \mathscr{H}\}$, where $P_{\theta,\eta}$ is the probability distribution with density

$$p_{\theta,\eta}(x) = h_\theta(x)\int g_\theta(\psi_\theta(x), z)\, d\eta(z). \tag{3.1}$$

Suppose that $p_{\theta,\eta}$ is smooth in θ in the sense of existence of $\dot{l}_{\theta,\eta} \in L_2(P_{\theta,\eta})^k$ such that

$$\int\left[t^{-1}\left(p_{\theta+th,\eta}^{1/2} - p_{\theta,\eta}^{1/2}\right) - \tfrac{1}{2}h'\dot{l}_{\theta,\eta}p_{\theta,\eta}^{1/2}\right]^2 d\mu \to 0, \quad \text{as } t \downarrow 0, \tag{3.2}$$

for every $h \in \mathbb{R}^k$. Thus paths of the form $t \to P_{\theta+th,\eta}$ generate scores $h'\dot{l}_{\theta,\eta}$. It is clear from (3.1) that any score resulting from a path of the form $t \to P_{\theta,\eta_t}$ will be a function of ψ_θ. Then one expects joint paths $t \to P_{\theta+th,\eta_t}$ to yield scores of the type $h'\dot{l}_{\theta+\eta} + b(\psi_\theta)$. The main result of this section is that, under completeness of ψ_θ, there exists a *convex* tangent cone, which has the set of all functions of this form as its closed linear span.

For given (θ, η) let $G_{\theta,\eta}$ be the law of $\psi_\theta(X)$ when $X \sim P_{\theta,\eta}$ and set

$$\mathscr{H}_{\theta,\eta} = \left\{\eta' \in \mathscr{H}: P_{\theta,\eta'} \ll P_{\theta,\eta} \text{ and } \int p_{\theta,\eta'}^2/p_{\theta,\eta}\, d\mu < \infty\right\}.$$

THEOREM 3.1. *Let* (3.1) *and* (3.2) *hold for every* $(\theta, \eta) \in \Theta \times \mathscr{H}$, *let* $\mathscr{H}$ *be convex and suppose that* $\{G_{\theta,\eta'}: \eta' \in \mathscr{H}_{\theta,\eta}\}$ *is complete. Then there exists a set of paths* $t \to P_{\theta+th,\eta_t}$ *such that the resulting tangent set* $T(P_{\theta,\eta})$ *is a convex cone and satisfies*

$$\overline{\text{lim}\, T(P_{\theta,\eta})} = \left\{h'\dot{l}_{\theta,\eta} + b(\psi_\theta): h \in \mathbb{R}^k,\ b \in L_2(G_{\theta,\eta}),\ \int b\, dG_{\theta,\eta} = 0\right\}. \tag{3.3}$$

Together Theorems 2.4 and 3.1 lead to lower bounds for the LAM risk. We give two examples. Let

$$\tilde{l}_{\theta,\eta} = \dot{l}_{\theta,\eta} - E_\theta\left(\dot{l}_{\theta,\eta}(X)|\psi_\theta(X) = \psi_\theta(\ )\right), \tag{3.4}$$

$$\tilde{I}_{\theta,\eta} = E_{\theta,\eta}\tilde{l}_{\theta,\eta}(X)\tilde{l}_{\theta,\eta}(X)'. \tag{3.5}$$

COROLLARY 3.2. *Assume that* $\tilde{I}_{\theta,\eta}$ *is nonsingular. Then*:

(i) *The LAM risk for estimating* $P_{\theta,\eta} \to \theta$ *is bounded from below by* $\int l\, dN_{0,\tilde{I}_{\theta,\eta}^{-1}}$.

(ii) *The LAM risk for estimating* $P_{\theta,\eta} \to \int f\, dP_{\theta,\eta}$ (*where* f *is a known bounded function*) *is bounded from below by* $\int l\, dN_{0,\tilde{J}_{\theta,\eta}}$, *where*

$$\tilde{J}_{\theta,\eta} = \langle f, \tilde{l}_{\theta,\eta}\rangle'\tilde{I}_{\theta,\eta}^{-1}\langle f, \tilde{l}_{\theta,\eta}\rangle + \sigma_{\theta,\eta}^2\left(E_\theta(f(X)|\psi_\theta(X))\right).$$

The question of whether the bounds of Corollary 3.2 are sharp has been answered positively in two cases [Pfanzagl and Wefelmeyer (1982), Chapter 14 and van der Vaart (1988a), Chapter 5 and (1988b)]. If ψ_θ is independent of θ or $g_\theta(\cdot, z)$ is a smooth density of $\psi_\theta(X)$ with respect to Lebesgue measure on a convex open subset of $\mathbb{R}^m$, then there exist regular estimator sequences with the normal distribution in (i) or (ii) as the limiting distribution.

The above-mentioned work also discusses conditions under which Theorem 3.1 can be strengthened to the assertion of existence of a set of paths $t \to P_{\theta+th,\eta_t}$ having the linear space on the right side of (3.3) as a tangent cone. Sufficient for this is that $\mathscr{H}$ is the set of (essentially) all probability distributions on $\mathscr{Z}$ and

$$\{G_{\theta,z}: z \in A\} \text{ is complete for every } A \in \mathscr{A} \text{ with } \int_A d\eta = 1. \tag{3.6}$$

For instance, let $\mathscr{H}$ be the set of all probability distributions on $(\mathscr{Z}, \mathscr{A})$ and let

$$g_\theta(y, z) = c_\theta(z) d_\theta(y) e^{y'zq(\theta)}$$

be an exponential family. Then the weaker assertion (3.3) is typically true at every (θ, η), whereas (3.6) and its implication is true at (θ, η) where η is absolutely continuous, but may fail if η has a support without a limit point.

The difference between the two cases is important, e.g., in view of admissibility of a best regular estimator (see the discussion at the end of Section 2).

Example. Let $\mathscr{H}$ be all probability measures on $\mathbb{R}, \Theta = (0, \infty)$ and $h_\theta(\cdot) g_\theta(\psi_\theta(\cdot), z)$ the density of a $N_2(\binom{z}{z}, \theta I)$ distribution. Denote the observations by $(X_1, Y_1), \dots, (X_n, Y_n)$ and set $\psi_\theta(x, y) = x + y$. Then $G_{\theta,z}$ is a $N(2z, 2\theta)$ distribution and the conditions of Theorem 3.1 are satisfied. The obvious estimator for θ, $T_n = \frac{1}{2} n^{-1} \sum_{j=1}^n (X_j - Y_j)^2$, satisfies $\sqrt{n}(T_n - \theta) = n^{-1/2} \sum_{j=1}^n \tilde{l}_{\theta,\eta}(X_j, Y_j)$. Hence it is LAM and best regular at every (θ, η).

Example. Let Θ be empty, $\mathscr{H}$ be all probability measures on $(0, \infty)$ and $p(x, z) = e^{-z} z^x / x!$, $x \in \{0, 1, 2, \dots\}$, be the density of a Poisson distribution. Set $\psi(x) = x$. By Theorem 3.1 there exists a convex tangent cone $T(P_\eta)$ with closed linear span equal to all mean zero functions in $L_2(P_\eta)$. The functional $P_\eta \to P_\eta(X = a)$, where $a \in \{0, 1, 2, \dots\}$, fixed, satisfies (2.2) with gradient $\tilde{\kappa}_{P_\eta}(x) = 1_a(x) - P_\eta(X = a)$. Hence the empirical estimator $T_n = n^{-1} \sum_{j=1}^n 1_a(X_j)$ is LAM and best regular at every P_η, irrespective of its support. This strengthens results of Tierney and Lambert (1984).

Under some conditions on the support of η, Lambert and Tierney (1984) prove equivalence of $\{T_n\}$ and the maximum likelihood estimator. As Tierney and Lambert (1984) point out, the MLE may be preferable over $\{T_n\}$ for its finite sample performance. Perhaps it is also preferable for its asymptotic behaviour at P_η for which the support of η is finitely discrete. Indeed, for such η the stronger condition (3.6) is violated and the best regular estimator may be inadmissible for all loss functions. Unfortunately, it appears hard to calculate the limiting behavior of the MLE in this case. One clearly does not expect *normal* limiting distributions.

4. Proofs.

PROOF OF THEOREM 2.1. Since marginal tightness implies joint tightness, Prohorov's theorem can be applied to see the existence of a subsequence of $\{n\}$ (abusing notation denoted $\{n\}$) such that

$$\mathscr{L}_P\left(\sqrt{n}\,(T_n - \kappa(P)), n^{-1/2} \sum_{j=1}^{n} \mathbf{g}(X_j)\right) \to \mathscr{L}(T, S), \tag{4.1}$$

for some random vector (T, S) with $\mathscr{L}(S) = N_m(0, \Sigma)$. For $g \in T(P)$ let $t \to P_{t,g}$ be the path in (2.1). It is well known that (2.1) implies local asymptotic normality

$$\Lambda\left(\bigotimes_{j=1}^{n} P_{1/\sqrt{n},g}, \bigotimes_{j=1}^{n} P\right) - n^{-1/2} \sum_{j=1}^{n} g(X_j) + \tfrac{1}{2}E_P g^2(X_1) \to_P 0, \tag{4.2}$$

where $\Lambda(Q, P)$ is the log-likelihood ratio of Q and P (the logarithm of the quotient of their densities with respect to an arbitrary σ-finite dominating measure). Combination of (4.1), (4.2) and (2.2) yields

$$\begin{aligned}&\mathscr{L}_P\left(\sqrt{n}\left(T_n - \kappa\left(P_{1/\sqrt{n},\,h'\mathbf{g}}\right)\right), \Lambda\left(\bigotimes_{j=1}^{n} P_{1/\sqrt{n},\,h'\mathbf{g}}, \bigotimes_{j=1}^{n} P\right)\right)\\ &\qquad \to \mathscr{L}\left(T - D_{\mathbf{g}}h, h'S - \tfrac{1}{2}h'\Sigma h\right).\end{aligned} \tag{4.3}$$

This determines the limiting law of $\mathscr{L}_{P_{1/\sqrt{n},\,h'\mathbf{g}}}(\sqrt{n}\,(T_n - \kappa(P_{1/\sqrt{n},\,h'\mathbf{g}})))$ by Le Cam's third lemma. It is given by $B \to \int_{B\times\mathbb{R}} e^{\lambda}\, d\mathscr{L}(T - D_{\mathbf{g}}h, h'S - \frac{1}{2}h'\Sigma h)(y, \lambda)$. Thus [cf. (2.3)]

$$L_{h'\mathbf{g}}(B) = E 1_B\left(T - D_{\mathbf{g}}h\right)e^{h'S - 1/2h'\Sigma h}, \qquad h'\mathbf{g} \in T(P). \tag{4.4}$$

Now let $V \sim N_m(h, \Sigma^{-1})$. It is fairly straightforward to construct a randomized estimator such that

$$\mathscr{L}_0(t(V, U), \Sigma V) = \mathscr{L}(T, S) \tag{4.5}$$

(cf. Lemma 4.1). Then

$$\begin{aligned}&P_h\left(t(V, U) - D_{\mathbf{g}}h \in B\right)\\ &\quad = \int P\left(t(v, U) - D_{\mathbf{g}}h \in B\right)e^{-1/2(v-h)'\Sigma(v-h)}\sqrt{\det \Sigma}\, dv(2\pi)^{-m/2}\\ &\quad = E_0 1_B\left(t(V, U) - D_{\mathbf{g}}h\right)e^{h'\Sigma V - 1/2h'\Sigma h}.\end{aligned} \tag{4.6}$$

Combination of (4.4)–(4.6) yields (2.4). □

LEMMA 4.1. *Given a random vector* (T, S) *in* $\mathbb{R}^k \times \mathbb{R}^m$, *there exists a measurable function* t *such that* $\mathscr{L}(t(S, U), S) = \mathscr{L}(T, S)$, *where* U *has a uniform distribution on* $[0, 1]$ *and is independent of* S.

PROOF. To simplify notation we give a construction for the case that $m = 2$. It is easy to produce two independent uniform $[0,1]$ variables U_1 and U_2 from one given uniform $[0,1]$ variable. Therefore it suffices to construct a randomized estimator based on U_1 and U_2. It suffices to find a statistic $t(S, U_1, U_2)$ such that $\mathscr{L}(t(s, U_1, U_2)) = \mathscr{L}(T|S = s)$ for every s. Take versions of the cumulative distribution functions of T_1 given $S = s$ and T_2 given $S = s$ and $T_1 = t_1$. Let $F^{-1}_{T_1|S=v}$ and $F^{-1}_{T_2|S=v, T_1=t_1}$ be their quantile functions. Standard arguments show that $F^{-1}_{T_1|S=v}(u_1)$ and $F^{-1}_{T_2|S=v, T_1=t_1}(u_2)$ are measurable functions of their two and three arguments, respectively. (They are right continuous in the u's and for fixed u measurable in the remaining arguments.) Now $\mathscr{L}(F^{-1}_{T_1|S=v}(U_1)) = \mathscr{L}(T_1|S = s)$ and $\mathscr{L}(F^{-1}_{T_2|S=v, T_1=t_1}(U_2)) = \mathscr{L}(T_2|S = s, T_1 = t_1)$. This implies that $t(s, u_1, u_2) = (F^{-1}_{T_1|S=v}(u_1), F^{-1}_{T_2|S=s, T_1=t_1(s, u_1)}(u_2))$, where $t_1(s, u_1) = F^{-1}_{T_1|S=v}(u_1)$, satisfies the requirements. □

PROOF OF THEOREM 2.2. Note that

$$e^{h's - 1/2h'\Sigma h}\, dN_{\tau, \Lambda}(h) = c(s)\, dN_{\mu(s), (\Sigma + \Lambda^{-1})^{-1}}(h),$$

where

$$\mu(s) = (\Sigma + \Lambda^{-1})^{-1}(\Lambda^{-1}\tau + s),$$

$$c(s) = \det(\Sigma + \Lambda^{-1})^{-1/2} \det \Lambda^{-1/2} e^{-1/2[\tau'\Lambda^{-1}\tau - (\Lambda^{-1}\tau + s)'\mu(s)]}.$$

Set $T = t(V, U)$ and $S = \Sigma V$. Then by (2.4) and (4.6),

$$\int L_{h'\mathbf{g}}(B)\, dN_{\tau, \Lambda}(h) = \int E_0 1_B\big(T - D_{\mathbf{g}}\mu(S) - D_{\mathbf{g}}h\big) c(S)\, dN_{0, (\Sigma + \Lambda^{-1})^{-1}}(h).$$

Thus (2.5) holds with $M_{\tau, \Lambda}$ given by

$$M_{\tau, \Lambda}(B) = E_0 1_B\big(T - D_{\mathbf{g}}\mu(S)\big) c(S).$$ □

PROOF OF THEOREM 2.3. Let $\{g_1, \dots, g_m\} \subset T(P)$ be linearly independent. By Theorem 2.1, $L = \mathscr{L}_h(t(V, U) - D_{\mathbf{g}}h)$ for all h in an open ball in $\mathbb{R}^m$. The map $h \to E_h e^{is(t(V,U) - D_{\mathbf{g}}h)}$ makes sense for $h \in \mathbb{C}^m$ and is analytic. Since it equals the constant $\int e^{isy}\, dL(y)$ for all h in an open ball, it must be constant on $\mathbb{C}^m$. Thus $L = \mathscr{L}_h(t(V, U) - D_{\mathbf{g}}h)$ for all $h \in \mathbb{R}^m$. Take a $N_{0, \lambda I}$ average on both sides and apply Theorem 2.2 to conclude that $L = N_{0, D_{\mathbf{g}}(\Sigma + \lambda^{-1}I)^{-1}D_{\mathbf{g}}'} * M_\lambda$. Using characteristic functions it is easily seen that $M_\lambda \to_w M$ as $\lambda \to \infty$ and that $L = N_{0, D_{\mathbf{g}}\Sigma^{-1}D_{\mathbf{g}}'} * M$. Finally, choose a sequence of subsets $\{g_1, \dots, g_m\}$ of $T(P)$ such that the difference in (2.7) converges to zero. □

PROOF OF THEOREM 2.4. Since l is lower semicontinuous, the first inequality follows easily. For the second let $\{g_1, \dots, g_m\} \subset T(P)$ be linearly independent. Since $T(P)$ is a convex cone, $h'\mathbf{g} \in T(P)$ for every $h \geq 0$. The middle term of

(2.11) is therefore not smaller than (with $R \in \mathbb{R}^+$, arbitrary)

$$\sup_{h'\mathbf{g}\in T(P)} \int l \wedge R\, dL_{h'\mathbf{g}} \geq \int\int l \wedge R\, dL_{h'\mathbf{g}}\, dN_{\tau,\Lambda}(h) - RN_{\tau,\Lambda}(\mathbb{R}^m - \{h\colon h \geq 0\}),$$

$$\geq \int l \wedge R d\left[N_{0, D_\mathbf{g}(\Sigma+\Lambda^{-1})^{-1}D_\mathbf{g}'} * M_{\tau,\Lambda}\right]$$

$$- RN_{\tau,\Lambda}(\mathbb{R}^m - \{h\colon h \geq 0\}),$$

by Theorem 2.2, which by Anderson's lemma [see, e.g., Pfanzagl and Wefelmeyer (1985), page 454 and Ibragimov and Has'minskii (1981)] is greater than or equal to

$$\int l \wedge R\, dN_{0, D_\mathbf{g}(\Sigma+\Lambda^{-1})^{-1}D_\mathbf{g}'} - RN_{\tau,\Lambda}(\mathbb{R}^m - \{h\colon h \geq 0\}).$$

Now choose $\tau = (d, d, \dots, d)$ and let $d \to \infty$. Then the second term in the last expression converges to zero. Next set $\Lambda = \lambda I$ and let $\lambda \to \infty$. After that choose a sequence of finite subsets $\{g_1, \dots, g_m\}$ of $T(P)$ such that the difference in (2.7) converges to zero. Finally let $R \to \infty$. □

Proof of Theorem 2.5. It suffices to prove that for all $\beta \in \mathbb{R}^k$,

$$\beta'\Sigma(L_0)\beta - \beta'\tilde{J}_P\beta \geq 0.$$

Since this concerns the asymptotic variance of the estimators $\beta' T_n$ of the functional $\beta'\kappa$ (which has gradient $\beta'\tilde{\kappa}_P$), it suffices to prove the theorem for $k = 1$.

Let $\{g_1, \dots, g_m\} \subset T(P)$ be linearly independent. By Theorem 2.1, with $T = t(V, U)$,

$$\int x\, dL_{h'\mathbf{g}}(x) = E_h(T - D_\mathbf{g} h) = E_0(T - D_\mathbf{g} h)e^{h'V - 1/2h'\Sigma h} \tag{4.7}$$

whenever $h'\mathbf{g} \in T(P)$. Since $T(P)$ is a cone, (4.7) holds in particular for $h = \alpha e_i$, $0 \leq \alpha \leq 1$, $i = 1, \dots, m$. Taking partial derivatives with respect to h from the right at $h = 0$ yields

$$0 = ET V - D_\mathbf{g}. \tag{4.8}$$

Let $\tilde{V} = D_\mathbf{g}\Sigma^{-1}V$. By the Cauchy–Schwarz inequality

$$\sigma^2(S) \geq \sigma^{-2}(\tilde{V})\mathrm{Cov}^2(S, \tilde{V}) = D_\mathbf{g}\Sigma^{-1}D_\mathbf{g}',$$

by (4.8). Complete the proof by choosing appropriate $\{g_1, \dots, g_m\}$. □

Proof of Theorem 2.6. The last assertion of the theorem is an immediate consequence of Le Cam's third lemma and (2.2). Next (2.12) is obviously satisfied. We prove that (2.12) implies (2.13). By Theorem 2.1, applied with $\mathbf{g} = \tilde{J}_P^{-1}\tilde{\kappa}_P$, there exists a randomized estimator t based on $V \sim N_k(h, \tilde{J}_P)$ such that $\mathscr{L}_0(t(V,U) - h) = L_{h'g}$ for every $h \in \mathbb{R}^k$. By (2.12),

$$\sup_{h \in \mathbb{R}^k} E_h l(t(V,U) - h) \le \int l \, dN_{0, \tilde{J}_P}.$$

Thus $t(V,U)$ is a minimax randomized estimator for h. By assumption $t(V,U) = V$ a.e. Now refer back to the construction of t in the proof of Theorem 2.1. By (4.5) $T - \tilde{J}_P S = 0$ a.e. Next the definition of (T, S) in (4.1) yields (2.15). □

PROOF OF THEOREM 3.1. Fix (θ, η) in $\Theta \times \mathscr{H}$ and η' in $\mathscr{H}_{\theta,\eta}$. Set

$$b_{\eta'} = g_{\theta,\eta}^{-1}(g_{\theta,\eta'} - g_{\theta,\eta}) 1\{g_{\theta,\eta} > 0\}$$

and set $\eta_t = t\eta' + (1-t)\eta$. It will first be shown that

$$\int \Big[t^{-1}\big(p_{\theta+th,\eta_t}^{1/2} - p_{\theta,\eta}^{1/2} \big) - \tfrac{1}{2}\big(h'\dot{l}_{\theta,\eta} + b_{\eta'}(\psi_\theta) \big) p_{\theta,\eta}^{1/2} \Big]^2 \, d\mu \to 0, \tag{4.9}$$

as $t \downarrow 0$. The idea here is that because of convexity, differences of the form

$$p_{\theta,\eta}^{-1}(p_{\theta,\eta'} - p_{\theta,\eta}) = \partial/\partial t|_{t=0} \log p_{\theta, t\eta' + (1-t)\eta}$$

are η-scores. However, showing joint differentiability in (θ, η) as in (4.9) requires some tedious work.

By (3.2) and absolute continuity of $P_{\theta,\eta'}$ with respect to $P_{\theta,\eta}$, it suffices to show convergence to zero of the integral in (4.9) over the set $A = \{p_{\theta,\eta} > 0\}$. Let $M_t \uparrow \infty$ such that $tM_t \to 0$ as $t \downarrow 0$. Write $p_t = p_{\theta+th,\eta}$, $q_t = p_{\theta+th,\eta'}$, $l_t = h'\dot{l}_{\theta,\eta} 1_{\{|\dot{l}_{\theta,\eta}| \le M_t\}}$ and $k_t = h'\dot{l}_{\theta,\eta'} 1_{\{|\dot{l}_{\theta,\eta'}| \le M_t\}}$. Set

$$u_t = t^{-1}\big(p_t^{1/2} - p_0^{1/2} \big) - \tfrac{1}{2} l_t p_0^{1/2}, \qquad v_t = t^{-1}\big(q_t^{1/2} - q_0^{1/2} \big) - \tfrac{1}{2} k_t q_0^{1/2}.$$

Then write $p_{\theta+th,\eta_t} = p_0 + b_t + c_t + d_t + e_t$, where

$$\begin{aligned}
b_t &= t l_t p_0 + t(q_0 - p_0) 1_{\{|q_0 - p_0| \le M_t p_0\}},\\
c_t &= (1-t) t^2 u_t^2 + t^3 v_t^2 + 1/4 t^3 k_t^2 q_0,\\
d_t &= (1-t)\big[2 t u_t p_0^{1/2} + t^2 u_t l_t p_0^{1/2} + 1/4 t^2 l_t^2 p_0 \big]\\
&\quad + t^2 k_t q_0 + t(q_0 - p_0) 1_{\{|q_0 - p_0| > M_t p_0\}} - t^2 l_t p_0,\\
e_t &= 2 t^2 v_t q_0^{1/2} + t^3 v_t k_t q_0^{1/2}.
\end{aligned}$$

It is straightforward to check that

$$t^{-2} \int_A \big[p_0^{-1} d_t^2 + c_t + |e_t| \big] \, d\mu \to 0, \qquad t^{-2} \int_A p_0^{-1} b_t^2 \, d\mu = O(1).$$

Now obtain (4.9) by applying the inequality

$$\begin{aligned}
&\big| (a + b + c + d + e)^{1/2} - a^{1/2} - \tfrac{1}{2} b a^{-1/2} \big|^2\\
&\quad \le 4|e| + 4d^2/[a(1-\varepsilon)] + 4c + \big(1/\sqrt{1-\varepsilon} - 1 \big)^2 b^2/a
\end{aligned}$$

(valid for real numbers satisfying $a > 0$, $|ba^{-1}| \le \varepsilon < 1$, $c \ge 0$, $a + b + c + d \ge 0$ and $a + b + c + d + e \ge 0$).

To complete the proof, set

$$T(P_{\theta,\eta}) = \left\{h'\dot{l}_{\theta,\eta} + \alpha b_{\eta'}(\psi_\theta):\ h \in \mathbb{R}^k,\ \alpha \ge 0,\ \eta' \in \mathscr{H}_{\theta,\eta}\right\}.$$

This is a convex cone and corresponds to the paths $t \to P_{\theta+th,\eta_t}$. We only have to show (3.3). Let $b \in L_2(G_{\theta,\eta})$ be such that $b \perp \{\alpha b_{\eta'}(\psi_\theta):\ \alpha \ge 0,\ \eta' \in \mathscr{H}_{\theta,\eta}\}$. Then for all $\eta' \in \mathscr{H}_{\theta,\eta}$,

$$0 = \int b b_{\eta'}\, dG_{\theta,\eta} = \int b\, dG_{\theta,\eta'} - \int b\, dG_{\theta,\eta}.$$

By completeness, b must be constant. □

PROOF OF COROLLARY 3.2. It is easily checked that the gradients $\tilde{\kappa}_P$ are given by:

(i) $\tilde{I}_{\theta,\eta}^{-1}\tilde{l}_{\theta,\eta}$.

(ii) $\langle f, \tilde{l}_{\theta,\eta}\rangle'\tilde{I}_{\theta,\eta}^{-1}\tilde{l}_{\theta,\eta} + E_\theta(f(X)|\psi_\theta(X) = \psi_\theta(\ )) - \int f\, dP_{\theta,\eta}$.

Now apply Theorem 2.4. □

Acknowledgments. I thank Chris Klaassen and the referees for helpful remarks.

REFERENCES

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309.

BLYTH, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22** 22–42.

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

BROWN, L. D. and FOX, M. (1974). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248–266.

CASELLA, G. and STRAWDERMAN, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–878.

DROSTE, W. and WEFELMEYER, W. (1984). On Hájek's convolution theorem. *Statist. Decisions* **2** 131–144.

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

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press.

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

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2** 361–379. Univ. California Press.

KOSHEVNIK, YU. A. and LEVIT, B. YA. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

LAMBERT, D. and TIERNEY, L. (1984). Asymptotic properties of maximum likelihood estimates in the mixed Poisson case. *Ann. Statist.* **12** 1388–1399.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

PFANZAGL, J. and WEFELMEYER, W. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, New York.

PFANZAGL, J. and WEFELMEYER, W. (1985). *Asymptotic Expansions for General Statistical Models. Lecture Notes in Statist.* **31**. Springer, New York.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–208. Univ. California Press.

TIERNEY, L. and LAMBERT, D. (1984). Asymptotic efficiency of estimators of functionals of mixed distribution. *Ann. Statist.* **12** 1380–1387.

VAN DER VAART, A. W. (1988a). *Statistical Estimation in Large Parameter Spaces.* CWI Tracts **44**. Centrum voor Wiskunde en Informatica, Amsterdam.

VAN DER VAART, A. W. (1988b). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* **16** 1450–1474.

DEPARTMENT OF MATHEMATICS
FREE UNIVERSITY
DE BOELELAAN 1081
1081 HV AMSTERDAM
THE NETHERLANDS