The Annals of Statistics
1980, Vol. 8, No. 6, 1252–1264

ASYMPTOTIC LOWER BOUNDS FOR RISK IN ROBUST ESTIMATION[1]

BY RUDOLF BERAN

University of California, Berkeley

Robustness and efficiency of a parameter estimate T can be assessed by comparing the fitted parametric distribution P_T with the actual distribution, which is assumed to lie near the parametric family $\{P_\theta : \theta \in \Theta\}$. Asymptotic lower bounds are established for the minimax risk over distributions near the parametric model, taking as loss function a monotone increasing function of the Hellinger distance between the actual distribution of the sample and the fitted distribution determined by T. The set of marginal distributions considered in the minimax calculation is a subset of the Hellinger ball of radius $O(n^{-1/2})$ centered at P_θ, n being the sample size. When the loss function is bounded, the lower bound on maximum risk can be attained asymptotically. However, an estimator of θ which is asymptotically minimax for bounded loss functions may be far from optimal when the loss function is unbounded. Such divergent behavior is exhibited, for instance, by the sample mean in nearly normal models.

1. Introduction. A basic question for the theory of robust estimation in parametric models is this: what is being estimated robustly when the postulated parametric model does not contain the actual distribution of the sample? In a few special cases, such as estimation of location in symmetrically contaminated symmetric distributions, an obvious answer exists. More generally, we may assert that the parameter estimate itself determines what is being estimated outside the parametric model (cf. Huber (1972), Hampel (1974), Bickel and Lehmann (1975)); however, it is not entirely clear then how robust estimates are to be compared or interpreted outside the parametric model.

An alternative view of estimation takes as a fundamental goal the estimation of the actual distribution of the sample. A parameter estimate determines a fitted parametric distribution, that member of the parametric family which is identified by the parameter estimate. The fitted parametric distribution is regarded as an estimator of the actual distribution. In this view:

(i) Parameter estimates are interpreted primarily as the parameter values which identify the fitted distribution;

(ii) Questions of efficiency and robustness are addressed by comparing the fitted parametric distribution with the actual distribution.

Some consequences of this attitude toward robust estimation are explored in this paper.

Let $\{P_\theta^n : \theta \in \Theta\}$ denote the parametric model for a sample of size n; let T_n be an estimator of θ; and let Q^n be the actual distribution of the sample. In principle, the adequacy of $P_{T_n}^n$ as an estimator of Q^n can be measured by the risk $E_{Q^n} w[d(P_{T_n}^n, Q^n)]$, where d is a metric on the set of all probabilities and w is nonnegative, monotone increasing. While exact calculation of such general risks is usually not feasible, it is possible to study their asymptotic behavior under sequences of distributions $\{Q^n;\ n \geq 1\}$ which are not too far from the sequence $\{P_\theta^n;\ n \geq 1\}$. Such a study is carried out in this paper, assuming independent identically distributed observations and taking Hellinger metric as d.

The main result obtained is an asymptotic lower bound on the minimax risk which generalizes the Fisher information bound of classical asymptotic estimation theory. (For the latter bound, see Chernoff (1956), page 12, who states a form attributed to Stein and to Rubin;

Received October 1978; revised August 1979.

[1] This research was supported in part by National Science Foundation Grant MCS 75-10376.

AMS 1970 *subject classification*. Primary 62G35; secondary 62F10.

Key words and phrases. Robust estimation, parametric models, risk, asymptotic minimax bounds, asymptotic minimax estimators.

also Hájek (1972) and Le Cam (1972). Koshevnik and Levit (1976) extend the information bound to certain nonparametric estimation problems.) As in the classical problem, the lower bound established in this paper sets an absolute standard against which practical robust estimators can be compared. When the function w is bounded, the lower bound on maximum risk can be attained asymptotically. However, an estimator of θ which is asymptotically minimax for bounded w may be far from optimal when w is unbounded. Such divergent behavior is exhibited, for instance, by the sample mean in nearly normal models. A discussion of what this result means is given at the end of Section 2.

2. The main results. The following notation provides a useful language in which to express the assumptions and results of this paper. Let Π be the set of all probabilities on a space $\mathscr{X}$ with σ-algebra $\mathscr{A}$. Define a set H as follows (cf. Neveu (1965), page 112, Koshevnik and Levit (1976)): A typical element of H is a pair (ξ, P), usually written $\xi(dP)^{1/2}$, such that $P \in \Pi$ and ξ is a random variable in $L_2(P)$. For simplicity, the element $1(dP)^{1/2}$ is written as $(dP)^{1/2}$. Suppose that $\xi(dP)^{1/2}$, $\eta(dQ)^{1/2}$ are elements of H and that $\mu = 2^{-1}(P + Q)$. Define the inner product

$$\langle \xi(dP)^{1/2}, \eta(dQ)^{1/2} \rangle = \int \xi\eta \left(\frac{dP}{d\mu}\right)^{1/2} \left(\frac{dQ}{d\mu}\right)^{1/2} d\mu \tag{2.1}$$

and, for arbitrary real a, b, the linear combination

$$a\xi(dP)^{1/2} + b\eta(dQ)^{1/2} = \left(a\xi\left(\frac{dP}{d\mu}\right)^{1/2} + b\eta\left(\frac{dQ}{d\mu}\right)^{1/2}\right)(d\mu)^{1/2}. \tag{2.2}$$

The corresponding norm $\|\cdot\|$ on H is given by

$$\begin{aligned} \|\xi(dP)^{1/2}\|^2 &= \langle \xi(dP)^{1/2}, \xi(dP)^{1/2} \rangle \\ &= \int \xi^2\, dP. \end{aligned} \tag{2.3}$$

In particular, $\|(dP)^{1/2} - (dQ)^{1/2}\|$ is the Hellinger distance between the probabilities $P, Q \in \Pi$. The elements $\xi(dP)^{1/2}$ and $\eta(dQ)^{1/2}$ of H are said to be equivalent if $\|\xi(dP)^{1/2} - \eta(dQ)^{1/2}\| = 0$. The set $\bar{H}$ of equivalence classes in H forms a Hilbert space with the above inner product and norm.

Let $\{P_\theta^n = P_\theta \times P_\theta \times \cdots \times P_\theta$ n-times: $\theta \in \Theta\}$ be the parametric model for a random sample of size n. Suppose that $Q_n^n = Q_n \times Q_n \times \cdots \times Q_n$ n-times is the actual distribution of the sample. Both $\{P_\theta\colon \theta \in \Theta\}$ and $\{Q_n\colon n \geq 1\}$ are probabilities on $(\mathscr{X}, \mathscr{A})$. Let T_n be any estimator of θ based upon a sample of size n. The risk of T_n is

$$R_n(T_n, Q_n) = E_{Q_n^n} w[\|(dP_{T_n}^n)^{1/2} - (dQ_n^n)^{1/2}\|^2], \tag{2.4}$$

where w is real-valued, nonnegative, and monotone increasing. Of primary interest here is the asymptotic behavior of the minimax risk $\inf_{T_n}\sup_{Q_n} R_n(T_n, Q_n)$, computed over sets of Q_n which shrink to some P_θ in a suitable manner as sample size n increases. The main assumptions to be made concern the parametric family $\{P_\theta\colon \theta \in \Theta\}$, the alternative distributions $\{Q_n\colon n \geq 1\}$, the function w, and the estimators $\{T_n\colon n \geq 1\}$.

Local contamination models, under which Q_n shrinks to the parametric model P_θ at rate $n^{-1/2}$, have been used in robustness studies by several authors (cf. Huber-Carol (1970), Jaeckel (1971), Beran (1978), Rieder (1980), Bickel (1978)). Over such shrinking neighborhoods of P_θ, both bias and variance of an estimate of θ remain comparable in magnitude; neither explodes relative to the other as in the fixed neighborhood asymptotics due to Huber. Thus, it becomes possible to analyze bias and variance simultaneously. For further discussion and history, see the review paper by Bickel (1978). Differences between this paper and the work cited above include the following:

(i) Both the estimate space and the parameter space here are sets of probability measures, rather than subsets of a Euclidean space. The risk (2.4) is an abstraction of mean squared error which expresses the view that the parameter of interest is the actual sample distribution Q_n^n and the estimate of this parameter is $P_{T_n}^n$.

(ii) This paper studies limiting behavior of minimax risk, rather than minimax behavior of asymptotic risk. The latter seems meaningless unless uniform convergence to the asymptotics can be established.

(iii) The particular local contamination model that will be used in this paper is weaker than those in the papers cited above.

In principle, there exists a hierarchy of possible local contamination models, loss functions, and corresponding asymptotic minimax estimates. This paper investigates one extreme—a light contamination model. Work carried out since this paper was submitted shows that within a very similar framework, but with heavier local contamination models, we can derive as asymptotically minimax estimates many of the familiar robust estimates (Millar (1979)) and also some adaptive robust estimates (Beran(1979)).

Assumption 1. The parameter space Θ is an open subset of R^k. Let $L_2^k(P_\theta)$ be the set of all k-dimensional column vectors whose components belong to $L_2(P_\theta)$. The mapping $\theta \to P_\theta$ has the following properties:

(i) for every $\theta \in \Theta$, there exists $\eta_\theta \in L_2^k(P_\theta)$ such that, for every ϕ in a Euclidean neighborhood of θ,

$$\|(dP_\phi)^{1/2} - (dP_\theta)^{1/2} - (\phi - \theta)'\eta_\theta(dP_\theta)^{1/2}\| = o(|\phi - \theta|); \tag{2.5}$$

(ii) for every $\theta \in \Theta$, the Fisher information matrix

$$I(\theta) = 4 \int \eta_\theta\eta_\theta' \, dP_\theta \tag{2.6}$$

is nonsingular.

From (2.5), it follows easily that $\int \eta_\theta \, dP_\theta = 0$ for every $\theta \in \Theta$.

Assumption 2. Let B be any subset of $\{\xi \in L_2(P_\theta): \int \xi \, dP_\theta = 0\}$ which is strongly compact in $L_2(P_\theta)$. $\{Q_n(h, \xi): h \in R^k, \xi \in B, n \geq 1\}$ is a family of probabilities on $(\mathscr{X}, \mathscr{A})$ with the following property: for every sequence $\{(h_n, \xi_n) \in R^k \times B; n \geq 1\}$ which converges strongly to some $(h, \xi) \in R^k \times B$,

$$\lim_{n\to\infty} \| n^{1/2}[(dQ_n(h_n, \xi_n))^{1/2} - (dP_{\theta_n})^{1/2}] - \xi(dP_\theta)^{1/2}\| = 0, \tag{2.7}$$

where $\theta_n = \theta + n^{-1/2}h_n$.

A construction for such a family of probabilities $\{Q_n(h, \xi)\}$ is described in Proposition 3 of Section 3. An important implication of (2.7) is $\lim_{n\to\infty} \|(dQ_n^n(h_n, \xi_n))^{1/2} - (dP_{\theta_n}^n)^{1/2}\|^2 < 2$; that is, the two probabilities do not separate as n increases.

Assumption 3. The function w: $[0, 2] \to \bar{R}^+$, the extended nonnegative half-line, is monotone increasing with $w(0) = 0$. Moreover, for every $b \geq 0$,

$$K_0(b) = \int w[2 - 2\exp(-|z|^2/8 - b^2/2)]\phi_k(z) \, dz < \infty, \tag{2.8}$$

where ϕ_k is the density of the standard k-dimensional normal distribution.

Assumption 4. The distributions of $\{n^{1/2}(T_n - \theta); n \geq 1\}$ under $\{P_\theta^n\}$ are tight.

This assumption is introduced because, for the loss functions considered in this paper (unlike ordinary quadratic loss), it may not be clear that other kinds of estimators have larger limiting risk. An alternative approach requires only that the $\{T_n\}$ be random vectors with values in Θ, but strengthens assumption 1 as follows:

Assumption 1′. The parameter space Θ is an open subset of R^k. The mapping $\theta \to P_\theta$ is a diffeomorphism in the following sense: the mapping is one-to-one and

(i) part (i) of assumption 1 holds;

(ii) for every $\theta \in \Theta$, there exists $\sigma_\theta \in L_2^k(P_\theta)$ such that, for every probability P_ϕ in a Hellinger neighborhood of P_θ and every $c \in R^k$,

$$|c'(\phi - \theta) - \langle c'\sigma_\theta(dP_\theta)^{1/2}, (dP_\phi)^{1/2} - (dP_\theta)^{1/2}\rangle| = o(\|(dP_\phi)^{1/2} - (dP_\theta)^{1/2}\|). \tag{2.9}$$

Assumption 1′ implies assumption 1 and that it is always possible to set

$$\sigma_\theta = \left[\int \eta_\theta\eta_\theta'\, dP_\theta\right]^{-1} \eta_\theta \tag{2.10}$$

in (2.9). To verify this claim, observe that, by projection

$$\sigma_\theta = A\eta_\theta + \rho, \tag{2.11}$$

where A is a $k \times k$ matrix, $\rho \in L_2^k(P_\theta)$, and the components of ρ are orthogonal in $L_2(P_\theta)$ to the components of η_θ. Substituting (2.11) and (2.5) into (2.9) gives

$$c'(\phi - \theta) = \left[c'A \int \eta_\theta\eta_\theta'\, dP_\theta\right](\phi - \theta) + o(|\phi - \theta|) \tag{2.12}$$

for every ϕ in a neighborhood of θ and every $c \in R^k$. Hence $A \int \eta_\theta\eta_\theta'\, dP_\theta$ equals the identity matrix for every $\theta \in \Theta$. Nonsingularity of A, hence of $I(\theta)$, follows; the choice $\rho = 0$ in (2.11) gives (2.10).

The use of quadratic mean differentiability (property (2.5)) as a statistical regularity condition is largely due to Hájek and to Le Cam; see for instance Hájek and Šidák (1967), Hájek (1972), Le Cam (1970). Simple sufficient conditions for Assumption 1 may be found, for example, in Lemma 1 of Beran (1977). In particular, these conditions are satisfied when $\{P_\theta: \theta \in \Theta\}$ is a canonical k-parameter exponential family and Θ is an open subset of the natural parameter space (cf. Berk (1972) for relevant properties of exponential families). In location models, Assumption 1 holds if P_θ has a density with respect to Lebesgue measure which is absolutely continuous with finite Fisher information (Hájek and Šidák (1967), Chapter 6).

Sufficient conditions for Assumption 1′ are: the satisfaction of Assumption 1, one-to-oneness of the mapping $\theta \to P_\theta$, and strong continuity in θ of $\eta_\theta(dP_\theta)^{1/2}$. For then, the mapping $\phi \to \langle (dP_\phi)^{1/2}, \eta_\theta(dP_\theta)^{1/2}\rangle$ is continuously differentiable in ϕ and the validity of (2.9) is a consequence of the inverse function theorem. In particular, the canonical exponentially family described in the previous paragraph satisfies Assumption 1′.

The main results of the paper are Theorems 1 and 2, stated below. Proofs of all results described in this section are deferred to Section 4. For every $\xi \in B$, the strongly compact set appearing in Assumption 2, let

$$d(\xi) = \left[\int \eta_\theta\eta_\theta'\, dP_\theta\right]^{-1} \int \xi\eta_\theta\, dP_\theta \tag{2.13}$$

and let

$$b^2 = \max_{\xi \in B} \|(\xi - \eta_\theta' d(\xi))(dP_\theta)^{1/2}\|^2 < \infty. \tag{2.14}$$

For every $c > 0$, let $S(c) = \{h \in R^k: h'I(\theta)h \leq c\}$.

THEOREM 1. *Suppose Assumptions* 1 *to* 4 *are satisfied or that Assumptions* 1′, 2, 3 *are satisfied. Then*

$$\lim_{c\to\infty} \liminf_{n\to\infty} \inf_{T_n} \sup_{(h,\xi)\in S(c)\times B} R_n(T_n, Q_n(h, \xi)) \geq K_0(b) \tag{2.15}$$

for b and $K_0(b)$ defined by (2.14) *and* (2.8) *respectively.*

The apparent complexity of the left side of (2.15) is not without purpose. Restricting (h, ξ) to a compact $S(c) \times B$ makes it possible, in some cases, to check attainability of the lower bound; letting $c \to \infty$ leads to the explicit, relatively simple, expression (2.8) for the bound. The first point is illustrated by the next theorem.

THEOREM 2. *Suppose Assumptions* 1, 2, *and* 3 *are satisfied, the function* w *is bounded, and the estimator sequence* $\{T_n;\, n \geq 1\}$ *is such that, under* P_θ^n,

$$n^{1/2}(T_n - \theta) = I^{-1}(\theta)n^{-1/2}2 \textstyle\sum_{i=1}^n \eta_\theta(x_i) + o_p(1). \tag{2.16}$$

Then, for every $c > 0$,

$$\lim_{n\to\infty} \sup_{(h,\xi)\in S(c)\times B} R_n(T_n, Q_n(h, \xi)) = K_0(b). \tag{2.17}$$

Since (2.17) obviously implies that the limit of the left side, as c tends to infinity, must equal $K_0(b)$, Theorem 2 provides conditions under which the asymptotic minimax bound of Theorem 1 is attained. Note that an estimator sequence $\{T_n;\, n \geq 1\}$ which satisfies (2.16) is asymptotically efficient, in the familiar senses, for the parameter θ under the model $\{P_\theta\colon \theta \in \Theta\}$. Implications of the two theorems for robust estimation are illustrated by the following examples; calculations underlying the examples are deferred to Section 4.

Example 1. Suppose $w(z) = z$ on $[0, 2]$ and P_θ is $N_k(\theta, I)$, i.e., k-dimensional standard normal translated to have mean θ. This normal model satisfies Assumption 1′ with $\eta_\theta(x) = 2^{-1}(x - \theta)$ and Fisher information matrix equal to the identity matrix. Theorem 1 applies with lower bound

$$K_0(b) = 2[1 - (4/5)^{k/2}\exp(-b^2/2)]. \tag{2.18}$$

Note that $K_0(b)$ approaches its maximum value of 2 if either parameter dimension k increases or if contamination neighborhood "diameter" b increases. Since w is bounded and the sample average $T_n = n^{-1} \sum_{i=1}^n x_i$ satisifes (2.16), the sample mean is an asymptotic minimax estimator of θ for the loss function determined by w, whatever the choice of B.

Example 2. Suppose w and P_θ are as in Example 1. Let $\{T_n;\, n \geq 1\}$ be a sequence of location invariant estimators of θ such that, under P_θ^n,

$$n^{1/2}(T_n - \theta) = n^{-1/2} \textstyle\sum_{i=1}^n \psi(x_i - \theta) + o_p(1), \tag{2.19}$$

where $\int |\psi|^2\phi_k\, dx < \infty$ and $\int \psi\phi_k\, dx = 0$, ϕ_k being the standard k-dimensional normal density. This class of estimators includes the usual M, L, and R estimators of location. Location invariance of T_n entails the property $\int \psi(x)x'\phi_k(x)\, dx = I$, the identity matrix. (To verify this, consider the effect of contiguous location shift alternatives upon the distribution of T_n). Hence, for every unit vector $a \in R^k$, $\int (a'\psi)^2\phi_k\, dx \geq 1$, with equality for every a if and only if $\psi(x) = x$ a.e. Equivalently, the eigenvalues $\{\lambda_j;\, 1 \leq j \leq k\}$ of $\int \psi\psi'\phi_k\, dx$ are such that $\min_j \lambda_j \geq 1$, $\max_j \lambda_j > 1$ unless $\psi(x) = x$ a.e. By calculation,

$$\lim_{n\to\infty} \sup_{(h,\xi)\in S(c)\times B} R_n(T_n, Q_n(h, \xi)) = 2[1 - \exp(-b^2/2)\cdot\textstyle\prod_{j=1}^k (1 + \lambda_j/4)^{-1/2}] \tag{2.20}$$

for every $c > 0$. The right side of (2.20) is strictly larger than (2.18) unless $\psi(x) = x$ a.e.

Example 3. Let P_θ be $N_k(\theta, I)$, let $B = \{0\}$, and define $Q_n(h, 0) = (1 - n^{-2})P_{\theta+n^{-1/2}h} + n^{-2}A(n^2e)$, where $A(z)$ is the unit mass located at $z \in R^k$ and e is a fixed nonnull vector in R^k. By calculation, Assumption 2 is satisfied: for every convergent sequence $\{h_n \in R^k;\, n \geq 1\}$,

$$\lim_{n\to\infty} \| n^{1/2}[(dQ_n(h_n, 0))^{1/2} - (dP_{\theta_n})^{1/2}] \| = 0, \tag{2.21}$$

where $\theta_n = \theta + n^{-1/2}h_n$. This implies that

$$\lim_{n\to\infty} \| (dQ_n^n(h_n, 0))^{1/2} - (dP_{\theta_n}^n)^{1/2} \| = 0 \tag{2.22}$$

or, equivalently, that the variation norm of $Q_n^n(h_n, 0) - P_{\theta_n}^n$ tends to zero as n increases. Hence, the limiting distribution of an estimator under $Q_n^n(h_n, 0)$ is the same as its limiting distribution under $P_{\theta_n}^n$.

Example 4. Suppose P_θ is $N_k(\theta, I)$ and $w(z) = -2\log(1 - z/2)$ on [0, 2]. Then the risk becomes

(2.23) $$R_n(T_n, Q_n) = -2nE_{Q_n^n}\log[\langle (dP_{T_n})^{1/2}, (dQ_n)^{1/2}\rangle]$$

and Theorem 1 is applicable with

(2.24) $$K_0(b) = 4^{-1}k + b^2.$$

Since w is not bounded, Theorem 2 cannot be used here. Indeed, let B and $Q_n(h, \xi)$ be chosen as in Example 3, so that $b = 0$, and let T_n be the sample mean. Then, by calculation,

(2.25) $$\lim_{n\to\infty}\sup_{(h,\xi)\in S(c)\times B} R_n(T_n, Q_n(h, \xi)) = \infty$$

for every $c > 0$.

Thus, for one natural loss function, the sample mean is an asymptotic minimax estimator in nearly normal distributions (Example 1) and strictly dominates, asymptotically, other M, L or R estimators of location (Example 2). For another loss function, however, the sample mean's maximum risk can tend to infinity as sample size increases (Example 4). Underlying this divergent performance is the boundedness or unboundedness of the loss function. Under a nearly normal distribution, the sample mean has small probability of being a wild estimate and large probability of being a relatively good estimate. Hence, its risk is relatively small when the loss function is sufficiently bounded, but can be large when the loss function is unbounded.

Interpretation of these results must take into account the relative smallness of the contamination neighborhoods used in this study. More pessimistic local contamination models, such as the Hellinger ball, or Kolmogorov-Smirnov ball, or Cramér-von Mises ball centered at P_θ and of radius $O(n^{-1/2})$ would be expected to identify different, more pessimistic, estimates as asymptotically minimax. (See Millar (1979), Beran (1979) for results along these lines obtained since this paper was submitted.) Apart from suggesting questions that need further investigation, the present results provide a theoretical background for Stigler's (1977) finding that the sample mean compares favorably with other robust estimates when applied to some real, roughly normal sets of data.

3. Ancillary results. The first two propositions developed in this section are needed for the proofs of Theorems 1 and 2. The third proposition constructs a family of probabilities $\{Q_n(h, \xi)\}$ which satisfies Assumption 2 of the previous section.

Consider the problem of estimating the parameter $t \in R^k$ from one observation on the $N_k(t, I) \times \nu$ distribution, where ν is a probability on R which does not depend on t. Let $\bar{R}$ and $\bar{R}^+$ denote, respectively, the extended real line and the extended nonnegative half-line. Suppose that the loss incurred when t is estimated by $s \in R^k$ is $u(|s - t|^2)$ and that the function u satisfies the following requirements: u: $\bar{R}^+ \to \bar{R}^+$ is monotone increasing with $u(0) = 0$ and

(3.1) $$\int u(|z|^2)\exp(-|z|^2/2)\, dz < \infty.$$

Let $\bar{\mathscr{B}}^k$ be the Borel sets of $\bar{R}^k$. A randomized estimator of t is described by a Markov kernel $D: \bar{\mathscr{B}}^k \times R^{k+1} \to [0, 1]$. Given an observation $x \in R^k$, $y \in R$ from the $N_k(t, I) \times \nu$ distribution, $D(A; x, y)$ is the probability that the estimate of t lies in the set $A \in \bar{\mathscr{B}}^k$. The risk of the randomized estimator under the loss function and model described above is

(3.2) $$r(D, t) = \int \left[\int u(|s - t|^2)D(ds; x, y)\right]\phi_k(x - t)\, dx\nu(dy).$$

For $c > 0$ and $a \in R^k$, let $M(c, a)$ be the set of all probabilities on $(R^k, \mathscr{B}^k)$ which are supported on $\{t \in R^k: |t - a|^2 \le c\}$. The following result is a modification of the usual minimax theorem for estimation in the normal location model.

PROPOSITION 1. *Under the assumptions of the preceding two paragraphs, for every* $a \in R^k$,

$$\lim_{c\to\infty} \sup_{\pi\in M(c,a)} \inf_D \int r(D, t)\pi(dt) = \lim_{c\to\infty} \inf_D \sup_{\pi\in M(c,a)} \int r(D, t)\pi(dt) = r_0(u), \tag{3.3}$$

where $r_0(u) = \int u(|z|^2)\phi_k(z)\,dz$. *Thus,*

$$\lim_{c\to\infty} \inf_D \sup_{|t-a|^2\le c} r(D, t) = r_0(u) \tag{3.4}$$

for every $a \in R^k$.

PROOF. Suppose, initially, that u is bounded. Let π be any probability on $(R^k, \mathscr{B}^k)$, let

$$\begin{aligned} h(x) &= \int \phi_k(x - t)\pi(dt) \\ \pi(dt, x) &= h^{-1}(x)\phi_k(x - t)\pi(dt), \end{aligned} \tag{3.5}$$

and let

$$\rho(s, x, \pi) = \int u(|s - t|^2)\pi(dt, x). \tag{3.6}$$

Note that ρ is the posterior risk of the estimate s given an observation (x, y) from the $N_k(t, I) \times \nu$ distribution. Rearranging the order of integration in (3.2) yields

$$\int r(D, t)\pi(dt) = \int \rho(s, x, \pi)D(ds; x, y)h\,dx\nu(dy). \tag{3.7}$$

The nonrandomized estimator x is ϵ-Bayes for t among all randomized estimators (cf. Wolfowitz (1950), Blyth (1951)); thus, for every $\epsilon > 0$ there exists a probability λ on $(R^k, \mathscr{B}^k)$ such that

$$\inf_D \int r(D, t)\lambda(dt) > r_0(u) - \epsilon/2. \tag{3.8}$$

Let $\lambda_c \in M(c, a)$ denote the restriction of λ to the set $\{t \in R^k: |t - a|^2 \le c\}$. It is readily verified that

$$\lim_{c\to\infty} \sup_s |\rho(s, x, \lambda_c) - \rho(s, x, \lambda)| = 0 \tag{3.9}$$

for every $x \in R^k$. Equations (3.7) and (3.9) yield

$$\lim_{c\to\infty} \sup_D \left| \int r(D, t)\lambda_c(dt) - \int r(D, t)\lambda(dt) \right| = 0. \tag{3.10}$$

Thus, by (3.8) and (3.10), for every $\epsilon > 0$ there exists $c(\epsilon) > 0$ such that for every $c \ge c(\epsilon)$,

$$\begin{aligned} \sup_{\pi\in M(c,a)} \inf_D \int r(D, t)\pi(dt) &\ge \inf_D \int r(D, t)\lambda_c(dt) \\ &> r_0(u) - \epsilon. \end{aligned} \tag{3.11}$$

Consequently,

$$\liminf_{c\to\infty} \sup_{\pi\in M(c,a)} \inf_D \int r(D, t)\pi(dt) \ge r_0(u) \tag{3.12}$$

for every $a \in R^k$.

Observe that (3.12) remains valid when u is unbounded. For in this case, (3.12) holds for $\hat{u}_d = u \wedge d$, where $0 < d < \infty$, and the left side of (3.12) is then monotone increasing in d while

$\lim_{d\to\infty} r_0(\hat{u}_d) = r_0(u)$. Hereafter, the boundedness assumption on u made at the start of the proof will be dropped.

Let D_0 be the Markov kernel describing the nonrandomized estimator x. Evidently, for every $c > 0$ and every $a \in R^k$,

$$\begin{aligned}\sup_{\pi\in M(c,a)} \inf_D \int r(D, t)\pi(dt) &\le \inf_D \sup_{\pi\in M(c,a)} \int r(D, t)\pi(dt) \\ &\le \inf_D \sup_{|t-a|^2\le c} r(D, t) \\ &\le \sup_{|t-a|^2\le c} r(D_0, t) = r_0(u).\end{aligned} \tag{3.13}$$

The inequalities (3.12) and (3.13) imply the assertions of the proposition.

Let $\{Q_n(h, \xi)\}$ be a family of probabilities which satisfies Assumption 2. The log-likelihood ratio of $Q_n^n(h, \xi)$ with respect to P_θ^n is defined, up to a P_θ^n null-set, by

$$L_n(h, \xi) = \sum_{i=1}^n \log\left[\frac{dQ_{n,c}(h, \xi)}{dP_\theta}(x_i)\right], \tag{3.14}$$

where $Q_{n,c}$ is the part of Q_n which is absolutely continuous with respect to P_θ. Asymptotic behavior of $L_n(h, \xi)$ under P_θ^n is described in the following result.

PROPOSITION 2. *Suppose* (2.5) *and Assumption* 2 *are satisfied. Let* $\{(h_n, \xi_n) \in R^k \times B;\ n \ge 1\}$ *be a sequence which converges strongly to* $(h, \xi) \in R^k \times B$. *Then, under* P_θ^n,

$$L_n(h_n, \xi_n) = 2n^{-1/2} \sum_{i=1}^n [h'\eta_\theta(x_i) + \xi(x_i)] - 2\, \| (h'\eta_\theta + \xi)(dP_\theta)^{1/2} \|^2 + o_p(1). \tag{3.15}$$

PROOF. Assumption 2 and (2.5) imply

$$\lim_{n\to\infty} \| n^{1/2}[(dQ_n(h_n, \xi_n))^{1/2} - (dP_\theta)^{1/2}] - (h'\eta_\theta + \xi)(dP_\theta)^{1/2} \| = 0. \tag{3.16}$$

The result follows from (3.16) by a familiar argument (cf. Hájek and Šidák (1967), page 205).

A family of probabilities $\{Q_n(h, \xi)]$ which satisfies Assumption 2 can be constructed in the following simple manner. Let $a_n;\ n \ge 1\}$ be a sequence of nonnegative constants such that

$$\lim_{n\to\infty} a_n = \infty, \qquad \lim_{n\to\infty} n^{-1/2} a_n = 0 \tag{3.17}$$

and, for some $\delta > 0$,

$$\sup_n (4n^{-1/2} a_n) \le 1 - \delta. \tag{3.18}$$

For every $(h, \xi) \in R^k \times L_2(P_\theta)$ such that $\int \xi\, dP_\theta = 0$, let

$$w_n(\xi) \begin{array}{ll} = \xi & \text{if } |\xi| \le a_n \\ = a_n \operatorname{sgn}(\xi) & \text{otherwise} \end{array} \tag{3.19}$$

and let

$$q_n(h, \xi) = w_n(\xi) - \int w_n(\xi)\, dP_{\theta+n^{-1/2}h}. \tag{3.20}$$

Define $Q_n(h, \xi)$ by

$$\frac{dQ_n(h, \xi)}{dP_{\theta+n^{-1/2}h}} = 1 + 2n^{-1/2} q_n(h, \xi). \tag{3.21}$$

It is easily checked that the $\{Q_n(h, \xi)]$ are probabilities on $(\mathscr{X}, \mathscr{A})$.

PROPOSITION 3. *Suppose that* (2.5) *is satisfied. Then the family of probabilities* $\{Q_n(h, \xi)\}$ *defined above satisfies Assumption* 2.

PROOF. It suffices to show that for every sequence $\{(h_n, \xi_n) \in R^k \times B\}$ converging strongly to $(h, \xi) \in R^k \times B$,

$$\lim_{n\to\infty} \int \left[n^{1/2}\left(\left(\frac{dQ_n(h_n, \xi_n)}{dP_{\theta_n}}\right)^{1/2} - 1\right) - q_n(h_n, \xi_n)\right]^2 dP_{\theta_n} = 0, \tag{3.22}$$

where $\theta_n = \theta + n^{-1/2}h_n$,

$$\lim_{n\to\infty} \| q_n(h_n, \xi_n)(dP_{\theta_n})^{1/2} - q_n(h_n, \xi_n)(dP_\theta)^{1/2} \| = 0, \tag{3.23}$$

and

$$\lim_{n\to\infty} \int [q_n(h_n, \xi_n) - \xi]^2 \, dP_\theta = 0. \tag{3.24}$$

By (2.5) and the definition of q_n, the term on the left side of (3.23) is $0(n^{-1/2}a_n)$, which tends to zero as asserted because of (3.17).

Next, observe that

$$\begin{aligned} \int [w_n(\xi_n) - \xi_n]^2 \, dP_\theta &\le 4 \int_{|\xi_n| > a_n} \xi_n^2 \, dP_\theta \\ &\le 4 \sup_k \int_{|\xi_k| > a_n} \xi_k^2 \, dP_\theta, \end{aligned} \tag{3.25}$$

which tends to zero as n increases because of (3.17) and uniform integrability of the $\{\xi_n\}$. Hence

$$\lim_{n\to\infty} \int [w_n(\xi_n) - \xi]^2 \, dP_\theta = 0. \tag{3.26}$$

Since $\int \xi \, dP_\theta = 0$, (3.26) implies that $\lim_{n\to\infty} \int w_n(\xi_n) \, dP_\theta = 0$. Moreover

$$\begin{aligned} \left| \int w_n(\xi_n) \, d(P_{\theta_n} - P_\theta) \right| &\le a_n \| P_{\theta_n} - P_\theta \|_{\text{Var}} \\ &= O(n^{-1/2}a_n), \end{aligned} \tag{3.27}$$

the last step using (2.5). Thus, in view of (3.17),

$$\lim_{n\to\infty} \int w_n(\xi_n) \, dP_{\theta_n} = 0. \tag{3.28}$$

The limits (3.26) and (3.28) imply (3.24).

Verification of (3.22) rests upon the Taylor expansion of $(1 + z)^{1/2}$ for $z \ge -1$. When $z = 2n^{-1/2}q_n(h_n, \xi_n)$, this expansion gives

$$n^{1/2}\left[\left(\frac{dQ_n(h_n, \xi_n)}{dP_{\theta_n}}\right)^{1/2} - 1\right] = q_n(h_n, \xi_n) + r_n \tag{3.29}$$

where

$$\begin{aligned} |r_n| &\le 2^{-1}n^{-1/2}q_n^2(h_n, \xi_n)\delta^{-3/2} \\ &\le n^{-1/2}a_n |q_n(h_n, \xi_n)| \delta^{-3/2} \end{aligned} \tag{3.30}$$

because of (3.18). Applying (3.17) and (3.23), (3.24) to the right side of (3.30) yields (3.22).

4. Proofs of the main results. This section contains proofs for Theorems 1 and 2 and for the examples of Section 2. Some of the arguments were suggested by the work of Hájek (1972) and of Le Cam (1972).

PROOF OF THEOREM 1. It suffices to prove the result for bounded and continuous w, since the general case then follows readily (cf. discussion after (3.12)). Suppose that the theorem is false under Assumptions 1 to 4. Then, there exists $\epsilon > 0$ and a sequence of nonnegative constants $\{c_j; j \geq 1\}$ such that $\lim_{j\to\infty} c_j = \infty$ and

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{(h,\xi)\in S(c_j)\times B} R_n(T_n, Q_n(h, \xi)) \leq K_0(b) - \epsilon \tag{4.1}$$

for every $j \geq 1$. Fix j. By going to a subsequence (depending on j), we may assume, without loss of generality, that

$$\inf_{T_n} \sup_{(h,\xi)\in S(c_j)\times B} R_n(T_n, Q_n(h, \xi)) \leq K_0(b) - \epsilon/2 \tag{4.2}$$

for every $n \geq 1$. Hence there exists a sequence of estimators $\{T_n; n \geq 1\}$ such that

$$R_n(T_n, Q_n(h, \xi)) \leq K_0(b) - \epsilon/4 \tag{4.3}$$

for every $(h, \xi) \in S(c_j) \times B$ and every $n \geq 1$. Observe that

$$\begin{aligned} R_n(T_n, Q_n) &= E_{Q_n^n} w[\| (dP_{T_n}^n)^{1/2} - (dQ_n^n)^{1/2} \|^2] \\ &= E_{Q_n^n} w[2 - 2\{\langle (dP_{T_n})^{1/2}, (dQ_n)^{1/2} \rangle\}^n] \\ &= E_{Q_n^n} w[2 - 2\{1 - 2^{-1} \| (dP_{T_n})^{1/2} - (dQ_n)^{1/2} \|^2\}^n] \\ &\geq E_{Q_n^n} w[2 - 2\exp\{-2^{-1} n \| (dP_{T_n})^{1/2} - (dQ_n)^{1/2} \|^2\}] \\ &\geq E_{P_\theta^n} w[2 - 2\exp\{-2^{-1} n \| (dP_{T_n})^{1/2} - (dQ_n)^{1/2} \|^2\}]\exp[L_n(h, \xi)], \end{aligned} \tag{4.4}$$

where $L_n(h, \xi)$ is the log-likelihood ratio defined in (3.14).

Recall the definitions (2.13) and (2.14) of $d(\xi)$ and b^2. Let $\xi_0 \in B$ be such that

$$b^2 = \| (\xi_0 - \eta_\theta' d(\xi_0))(dP_\theta)^{1/2} \|^2; \tag{4.5}$$

existence of ξ_0 is assured by the strong compactness of B. Write $d_0 = d(\xi_0)$ and $\tau_0 = \xi_0 - \eta_\theta' d_0$. The latter equation gives an orthogonal decomposition for ξ_0 in $L_2(P_\theta)$. By Proposition 2, under P_θ^n,

$$\begin{aligned} L_n(h, \xi_0) &= 2n^{-1/2} \textstyle\sum_{i=1}^n [h'\eta_\theta(x_i) + \xi_0(x_i)] - 2\| (h'\eta_\theta + \xi_0)(dP_\theta)^{1/2} \|^2 + o_p(1) \\ &= 2n^{-1/2} \textstyle\sum_{i=1}^n (h + d_0)'\eta_\theta(x_i) - 2^{-1}(h + d_0)'I(\theta)(h + d_0) \\ &\quad + 2n^{-1/2} \textstyle\sum_{i=1}^n \tau_0(x_i) - 2b^2 + o_p(1) \end{aligned} \tag{4.6}$$

for every $h \in S(c_j)$.

Let $Y_n = I^{1/2}(\theta)n^{1/2}(T_n - \theta)$, $t = I^{1/2}(\theta)(h + d_0)$, $a = I^{1/2}(\theta)d_0$, $Z_n = I^{-1/2}(\theta)2n^{-1/2} \sum_{i=1}^n \eta_\theta(x_i)$, and $V_n = b^{-1}n^{-1/2} \sum_{i=1}^n \tau_0(x_i)$. Under $\{P_\theta^n\}$, the distributions of $\{(Z_n, V_n)\}$ converge weakly to the standard $(k + 1)$-dimensional normal distribution, while the distributions of $\{Y_n\}$ are tight on $(R^k, \mathscr{B}^k)$ because of Assumption 4. By going to a subsequence, we may assume without loss of generality that $(Y_n, Z_n, V_n) \Rightarrow (Y, Z, V)$ under $\{P_\theta^n\}$, where (Z, V) has the standard $(k + 1)$-dimensional normal distribution and Y has a distribution on $(R^k, \mathscr{B}^k)$. Then, by (4.6),

$$L_n(h, \xi_0) \Rightarrow_{P_\theta^n} t'Z - 2^{-1}|t|^2 + 2bV - 2b^2 \tag{4.7}$$

for every $h \in S(c_j)$. Moreover, by (2.5) and (2.7),

$$\begin{aligned} &n \| (dP_{T_n})^{1/2} - (dQ_n(h, \xi_0))^{1/2} \|^2 \\ &\quad = \| \{n^{1/2}(T_n - \theta) - (h + d_0)\}'\eta_\theta(dP_\theta)^{1/2} - \tau_0(dP_\theta)^{1/2} \|^2 + o_p(1) \\ &\quad \Rightarrow_{P_\theta^n} 4^{-1} | Y - t |^2 + b^2. \end{aligned} \tag{4.8}$$

Applying (4.7) and (4.8) to (4.4) yields, for every $h \in S(c_j)$,

$$\liminf_{n\to\infty} R_n(T_n, Q_n(h, \xi_0))$$

$$\geq E[u_b(|Y-t|^2)\exp(t'Z-2^{-1}|t|^2+2bV-2b^2)] \tag{4.9}$$
$$= \int E[u_b(|Y-t|^2)\,|\,Z=z,\ V=v]\phi_k(z-t)\phi(v-2b)\,dz\,dv,$$

where

$$u_b(x)=w[2-2\exp(-8^{-1}x-2^{-1}b^2)]. \tag{4.10}$$

Let $D_1(\cdot;z,v)$ be a Markov kernel which represents the conditional distribution of Y given $Z=z$, $V=v$. In the risk (3.2), take the $N(2b,1)$ distribution for the probability ν and the loss function u_b in place of u. Then, (4.9) and (4.3) assert that

$$r(D_1,t)=\int\left[\int u_b(|y-t|^2)D_1(dy;z,v)\right]\phi_k(z-t)\phi(v-2b)\,dz\,dv \tag{4.11}$$
$$\leq K_0(b)-\epsilon/4$$

for every t such that $|t-a|^2\leq c_j$. Hence

$$\inf_D \sup_{|t-a|^2\leq c_j} r(D,t)\leq K_0(b)-\epsilon/4. \tag{4.12}$$

Set $u=u_b-u_b(0)$ in Proposition 1, then add $u_b(0)$ to each side of (3.4). The result conflicts with (4.12) because j was fixed arbitrarily and $\lim_{j\to\infty} c_j=\infty$. The contradiction completes the proof of Theorem 1 under Assumptions 1 to 4.

The argument under the second set of assumptions is the same through (4.6). Since $\{n^{1/2}(T_n-\theta)\}$ need not be tight under $\{P_\theta^n\}$, define a new estimator T_n^* by the rule $T_n^*=T_n$ if $n\|(dP_{T_n})^{1/2}-(dP_\theta)^{1/2}\|^2\leq C$ and $T_n^*=\theta$ otherwise. For sufficiently large finite C, the assumptions imply that

$$n\|(dP_{T_n^*})^{1/2}-(dP_\theta)^{1/2}\|^2\leq C \tag{4.13}$$

and

$$R_n(T_n^*,Q_n(h,\xi))\leq R_n(T_n,Q_n(h,\xi)) \tag{4.14}$$

for every $n\geq 1$ and every $(h,\xi)\in S(c_j)\times B$. Moreover, (2.9) and (4.13) imply that the sequence $\{n^{1/2}(T_n^*-\theta)\}$ is tight under $\{P_\theta^n\}$. Replace T_n by T_n^* in (4.3), (4.4) and argue thereafter as in the previous case, with T_n^* in place of T_n.

PROOF OF THEOREM 2. For every $\xi\in L_2(P_\theta)$, let $\tau(\xi)=\xi-\eta_\theta' d(\xi)$. It suffices to show that for every $c>0$,

$$\lim_{n\to\infty}\sup_{(h,\xi)\in S(c)\times B}|R_n(T_n,Q_n(h,\xi))-K_0(\|\tau(\xi)(dP_\theta)^{1/2}\|)|=0; \tag{4.15}$$

for then, by monotonicity of w,

$$\lim_{n\to\infty}\sup_{(h,\xi)\in S(c)\times B} R_n(T_n,Q_n(h,\xi))=\sup_{\xi\in B}K_0(\|\tau(\xi)(dP_\theta)^{1/2}\|)$$
$$=K_0(\sup_{\xi\in B}\|\tau(\xi)(dP_\theta)^{1/2}\|) \tag{4.16}$$
$$=K_0(b),$$

as asserted in the theorem.

Suppose (4.15) is false. Then, there exists a $c>0$ and a sequence $\{(h_n,\xi_n)\in S(c)\times B;\ n\geq 1\}$ such that the differences $\{A_n=R_n(T_n,Q_n(h_n,\xi_n))-K_0(\|\tau(\xi_n)(dP_\theta)^{1/2}\|)\}$ do not converge to zero. Consequently, there exists a subsequence $\{m\}\subset\{n\}$ such that $\{A_m\}$ does not converge to zero and $\{(h_m,\xi_m)\}$ converges strongly to some $(h,\xi)\in S(c)\times B$. But this is not possible.

To verify the last claim, suppose that $\{(h_m,\xi_m)\}$ converges strongly to (h,ξ). Then

$$\lim_m K_0(\|\tau(\xi_m)(dP_\theta)^{1/2}\|)=K_0(\|\tau(\xi)(dP_\theta)^{1/2}\|). \tag{4.17}$$

Moreover, by Proposition 2 and contiguity, the distribution of $m^{1/2}(T_m-\theta)$ under $Q_m^m(h_m,$

ξ_m) converges weakly to the $N_k(h + d(\xi), I^{-1}(\theta))$ distribution. Hence,

$$\begin{aligned} m \| (dP_{T_m})^{1/2} - (dQ_m(h_m, \xi_m))^{1/2} \|^2 \\ = \| \{m^{1/2}(T_m - \theta) - (h + d(\xi))\}'\eta_\theta(dP_\theta)^{1/2} - \tau(\xi)(dP_\theta)^{1/2} \|^2 + o_p(1) \\ \Rightarrow_{Q_m^m(h_m,\xi_m)} 4^{-1} | Z |^2 + \| \tau(\xi)(dP_\theta)^{1/2} \|^2, \end{aligned} \tag{4.18}$$

where Z has the standard k-dimensional normal distribution. Since w is bounded and has at most a countable number of discontinuities,

$$\begin{aligned} \lim_m R_m(T_m, Q_m(h_m, \xi_m)) &= \lim_m E_{Q_m^m} w[2 - 2\{1 - 2^{-1} \| (dP_{T_m})^{1/2} - (dQ_m)^{1/2} \|^2\}^m] \\ &= K_0(\| \tau(\xi)(dP_\theta)^{1/2} \|). \end{aligned} \tag{4.19}$$

Comparing (4.17) with (4.19) shows that the differences A_m converge to zero. This completes the proof.

PROOF FOR EXAMPLE 1. In this case,

$$K_0(b) = 2[1 - \exp(-b^2/2) \int \exp(-| z |^2/8)\phi_k(z)\, dz] \tag{4.20}$$

which reduces to (2.18) because the Laplace transform of the chi-square distribution with k degrees-of-freedom is $(1 + 2s)^{-k/2}$.

PROOF FOR EXAMPLE 2. The argument for (2.20) is similar to the proof of Theorem 2, the essential difference being that, under $Q_m^m(h_m, \xi_m)$, the distribution of $m^{1/2}(T_m - \theta)$ converges weakly to the $N_k(h + d(\xi), \Sigma)$ distribution, where $\Sigma = \int \psi\psi'\phi_k\, dx$. Consequently, (4.18) is replaced by

$$m \| (dP_{T_m})^{1/2} - (dQ_m(h_m, \xi_m))^{1/2} \|^2 \Rightarrow_{Q_m^m(h_m,\xi_m)} 4^{-1}Z'\Sigma Z + \| \tau(\xi)(dP_\theta)^{1/2} \|^2, \tag{4.21}$$

where Z has the standard k-dimensional normal distribution; moreover $Z'\Sigma Z$ has the same distribution as $Z' \operatorname{diag}\{\lambda_j\} Z$.

PROOF FOR EXAMPLE 3. Since P_θ and $A(z)$ are mutually singular,

$$(dQ_n(h_n, 0))^{1/2} = (1 - n^{-2})^{1/2}(dP_{\theta_n})^{1/2} + n^{-1}(dA(n^2 e))^{1/2} \tag{4.22}$$

and

$$\langle (dP_{\theta_n})^{1/2}, (dA(n^2 e))^{1/2} \rangle = 0. \tag{4.23}$$

Thus

$$\| n^{1/2}[(dQ_n(h_n, 0))^{1/2} - (dP_{\theta_n})^{1/2}] \|^2 = n[(1 - n^{-2})^{1/2} - 1]^2 + n^{-1}, \tag{4.24}$$

which tends to zero as n increases, proving (2.21).

PROOF FOR EXAMPLE 4. Because of (4.22), (4.23), and normality of P_θ,

$$\begin{aligned} \langle (dP_{T_n})^{1/2}, (dQ_n(0, 0))^{1/2} \rangle &= (1 - n^{-2})^{1/2} \langle (dP_{T_n})^{1/2}, (dP_\theta)^{1/2} \rangle \\ &= (1 - n^{-2})^{1/2} \exp[-8^{-1} | T_n - \theta |^2]. \end{aligned} \tag{4.25}$$

Thus,

$$\begin{aligned} R_n(T_n, Q_n(0, 0)) &= 4^{-1} n E_{Q_n^n} | T_n - \theta |^2 - n \log(1 - n^{-2}) \\ &\geq 4^{-1} n | E_{Q_n^n}(T_n) - \theta |^2 - n \log(1 - n^{-2}) \\ &= O(n), \end{aligned} \tag{4.26}$$

which implies (2.25).

REFERENCES

BERAN, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, R. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313.

BERAN, R. (1979). Efficient robust estimates in parametric models. Unpublished manuscript.

BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

BICKEL, P. J. (1978). Some recent developments in robust statistics. Unpublished manuscript.

BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models. II. Location. *Ann Statist.* **3** 1045–1069.

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

CHERNOFF, H. (1956). Large sample theory: parametric case. *Ann Math. Statist.* **27** 1–22.

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

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HUBER, P. (1972). Robust statistics: a review. *Ann. Math. Statist.* **43** 1041–1067.

HUBER-CAROL, C. (1970). Etude asymptotique de tests robustes. Thesis, ETH, Zurich.

JAECKEL, L. (1971). Robust estimates of location; symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020–1034.

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

LE CAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802–828.

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

MILLAR, P. W. (1979). Robust estimation via minimum distance methods. Unpublished manuscript.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

RIEDER, H. (1980). Estimates derived from robust tests. *Ann. Statist.* **8** 106–115.

STIGLER, S. M. (1977). Do robust estimators work with *real* data? *Ann. Statist.* **5** 1055–1078.

WOLFOWITZ, J. (1950). Minimax estimates of the mean of a normal distribution with known variance. *Ann. Math. Statist.* **17** 489–493.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720