The Annals of Statistics
1981, Vol. 9, No. 6, 1301–1309

MINIMAX ESTIMATION OF THE MEAN OF A NORMAL DISTRIBUTION WHEN THE PARAMETER SPACE IS RESTRICTED[1]

BY P. J. BICKEL

University of California, Berkeley

If X is a $N(\theta, 1)$ random variable, let $\rho(m)$ be the minimax risk for estimation with quadratic loss subject to $|\theta| \le m$. Then $\rho(m) = 1 - \pi^2/m^2 + o(m^{-2})$. We exhibit estimates which are asymptotically minimax to this order as well as approximations to the least favorable prior distributions. The approximate least favorable distributions (correct to order m^{-2}) have density $m^{-1} \cos^2 \left(\frac{\pi}{2m} s\right)$, $|s| \le m$ rather than the naively expected uniform density on $[-m, m]$. We also show how our results extend to estimation of a vector mean and give some explicit solutions.

1. Introduction If we want to estimate the completely unknown mean of a normal distribution with known variance using quadratic loss, then the sample mean is, of course, minimax. If, however, we have prior knowledge that the mean lies in a known interval, say $[-m, m]$, then the sample mean is inadmissible and it is well known that the minimax estimate is Bayes with respect to a least favorable prior distribution concentrating on a finite number of points. For m small (≤ 1.05) Casella and Strawderman (1980) show that this distribution concentrates on the end points. As m increases, the number of points increases, their location and the masses assigned to them vary in an as yet unknown fashion so that as $m \to \infty$, the prior distributions approximate Lebesgue measure (conditionally) and the minimax risk tends to the variance of the sample mean.

In Section 2, we ascertain a little more precisely what the behavior of the minimax risk is for large m. We do this by rescaling the least favorable prior distributions to the interval $[-1, 1]$ and finding the limit of these rescaled distributions as the solution of the variational problem of minimizing Fisher information among distributions concentrating on $[-1, 1]$. We show that this limit has density $\cos^2 (\pi/2)\, x$, $|x| \le 1$ and deduce that (for sample size 1, variance 1) the minimax risk is $1 - \pi^2/m^2 + o(m^{-2})$ as $m \to \infty$.

The key idea in obtaining this result is an identity relating Bayes risk with respect to any prior distribution to Fisher information. This identity is implicit in Brown (1971) and is related to an identity of Stein (Hudson, 1978). This relation is also used in Bickel (1980) and was independently discovered and used by Marazzi (1980) as well as Levit (1980).

In Section 3 we extend these results to estimation in p dimensions. We obtain the expected qualitative break in the shape of the limits of the rescaled prior for $p \ge 3$ and, parenthetically, can deduce the inadmissibility of the sample mean for $p \ge 3$.

2. The one dimensional case. Let X be a random $N(\theta, 1)$ variable. Let $\mathscr{D} = \{$all estimates of $\theta\}$ and for $\delta \in \mathscr{D}$, define

$$R(\theta, \delta) = E_\theta(\delta - \theta)^2.$$

Received July, 1980; revised January, 1981.

[1] Research supported by Office of Naval Research Grant Number N00014 80 C 0163 and the Adolph and Mary Sprague Miller Foundation.

AMS 1980 *subject classifications*: 62F10, 62C99.

Key words and phrases. Minimax, estimation, Fisher information, James-Stein estimate.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

If G is a Bayes prior probability distribution on R let $\delta(\cdot, G)$ denote the Bayes estimate and

$$r(G) = \int R(\theta, \delta(\cdot, G))\, G(d\theta)$$

denote its Bayes risk. The minimax risk for estimating θ, given that $|\theta| \le m$, is defined by

$$\rho(m) = \min \max\{R(\theta, \delta): |\theta| \le m, \delta \in \mathscr{D}\}.$$

It is well known by convexity and analyticity considerations that there is a unique symmetric Bayes prior distribution G_m^0 concentrating on a finite number of points such that $\delta(\cdot, G_m^0)$ is unique minimax and G_m^0 is least favorable. That is,

$$R(\theta, \delta(\cdot, G_m^0)) = \max\{R(\theta, \delta(\cdot, G_m^0)): |\theta| \le m\} = r(G_m^0) = \rho(m)$$

with G_m^0 probability 1.

The structure of G_m^0 has been studied for small m by Casella and Strawderman (1980) who showed that for $|m| \le 1.05$, G_m^0 assigns mass ½ each to $\pm m$. We proceed with our study of G_m^0 for m large.

Let G_1 be the distribution on $[-1, 1]$ with density

$$g_1(s) = \cos^2\left(\frac{\pi}{2} s\right), \quad |s| \le 1$$
$$= 0 \text{ otherwise,}$$

and let G_m be the corresponding distribution scaled up to $[-m, m]$ with density given by,

$$g_m(s) = m^{-1} g_1(sm^{-1}).$$

Then $\{G_m\}$ are approximately least favorable in the following sense.

THEOREM 2.1: *As $m \to \infty$*

$$\rho(m) = r(G_m) + o(m^{-2}) \tag{2.1}$$

$$r(G_m) = 1 - \frac{\pi^2}{m^2} + o(m^{-2}). \tag{2.2}$$

Moreover, let $G_1^{(m)}$ be the distribution obtained by scaling G_m^0 down to $[-1, 1]$, i.e.,

$$G_1^{(m)}(s) = G_m^0(ms)$$

then

$$G_1^{(m)} \to G_1 \tag{2.3}$$

in the sense of weak convergence.

It is *not* true that $\delta(\cdot, G_m)$ are asymptotically minimax. In fact, $\limsup_m R(m, \delta(\cdot, G_m)) > 1$. However, asymptotically minimax estimates can be constructed as follows. Let

$$\bar\psi(x) = -\frac{g_1'}{g_1}(x) = \pi \tan\left(\frac{\pi}{2} x\right), \quad |x| < 1. \tag{2.4}$$

Suppose $\{\psi_m\}$ is a sequence of functions and that $\{a_m\}$, $\{b_m\}$, $\{c_m\}$ are sequences of positive numbers with the following properties:

(a) $1 > a_m \downarrow 0$, $ma_m \to \infty$

(b) $\sup\{|\psi_m(x) - \bar\psi(x)| : |x| \le 1 - a_m^2\} \to 0$

(c) $\sup\{|\psi_m'(x) - \bar\psi'(x)| : |x| \le 1 - a_m^2\} \to 0$

(d) for $|x| \ge 1 - a_m^2$, $2|\psi_m'(x)| + \psi_m^2(x) \le b_m + c_m x^2$

(e) $b_m\{1 - \Phi(ma_m)\} \to 0$

(f) $c_m\{1 - \Phi(ma_m)\} \to 0$,

where Φ is the standard normal c.d.f. Let

$$n = m(1 - a_m)^{-1},$$

and define

$$\delta_m(x) = x - n^{-1}\psi_m(xn^{-1}).$$

THEOREM 2.2. *If properties (a) – (f) hold, the estimates $\{\delta_m\}$ are asymptotically optimal and have asymptotically constant risk on $[-m, m]$ in the sense that*

$$\max\left\{\left|R(\theta, \delta_m) - 1 + \frac{\pi^2}{m^2}\right| : |\theta| \le m\right\} = o(m^{-2}). \tag{2.5}$$

Estimates δ_m can readily be constructed. For example, let

$$\begin{aligned} \psi_m(x) &= \bar{\psi}(x), \ |x| \le 1 - a_m^2 \\ &= [\bar{\psi}(1 - a_m^2) + \bar{\psi}'(1 - a_m^2)\{x - (1 - a_m^2)\}] \operatorname{sgn} x, \ |x| > 1 - a_m^2. \end{aligned} \tag{2.6}$$

It is easy to see that we can then take $b_m \sim a_m^{-4}$, $c_m \sim a_m^{-8}$ and conditions (e), (f) reduce to

$$a_m^{-8}\{1 - \Phi(ma_m)\} \to 0.$$

It is also possible to establish

COROLLARY 2.1. *The estimates $\delta(\cdot, G_n)$ are optimal for n as above if*

$$ma_m^6 \to \infty.$$

PRELIMINARIES. The key to these theorems are two identities. The first is a special case of (13.4) of Brown (1971). For any prior distribution G let f_G be the density of the marginal distribution of X, ϕ the standard normal density.

$$f_G(x) = \phi * G(x) = \int_{-\infty}^{\infty} \phi(x - \theta) G(d\theta).$$

(Here and in the sequel * denotes convolution.) Brown's identity for the Bayes risk of G is

$$r(G) = 1 - \int_{-\infty}^{\infty} \frac{\{f_G'(x)\}^2}{f_G(x)} dx. \tag{2.7}$$

The second identity is due to Stein, see Hudson (1978). Let δ be an estimate differentiable in x and such that

$$E_\theta |\delta'(X)| < \infty.$$

Let

$$\psi(x) = x - \delta(x).$$

Then Stein's identity is

$$R(\theta, \delta) = 1 - E_\theta\{2\psi'(X) - \psi^2(X)\}. \tag{2.8}$$

Stein's identity is obtained by an integration by parts while Brown's follows from Stein's by putting

$$\delta(x) = \delta(x, G) = x + \frac{f_G'(x)}{f_G(x)}$$

and integrating with respect to G. Brown's identity can be written in terms of the more familiar Fisher information defined (Huber, 1964) for any distribution F by

$$I(F) = \int_{-\infty}^{\infty} \frac{\{f'(x)\}^2}{f(x)}\,dx$$

if F has an absolutely continuous density f, being infinite otherwise. Evidently (2.7) is just

$$r(G) = 1 - I(\Phi * G).$$

We need four properties of $I(\cdot)$ which may be found in Port and Stone (1974).

(i) If $H(x) = F\left(\frac{x-\mu}{\sigma}\right)$, all x and $\sigma > 0$, then $I(H) = \sigma^{-2} I(F)$;

(ii) If $F_n \to F$ weakly, then $I(F) \le \liminf_n I(F_n)$;

(iii) If $H_n \to \delta_0$ (point mass at 0) weakly then $I(F * H_n) \to I(F)$;

(iv) $I(F_1 * F_2) \le \max\{I(F_1), I(F_2)\}$.

Finally, we require a special case of a theorem of Huber (1974).

LEMMA 2.1. *The distribution G_1 minimizes $I(F)$ uniquely among all F concentrating on* $[-1, 1]$. *Moreover,*

$$I(G_1) = \pi^2. \tag{2.9}$$

This follows (after an obvious typographical correction) from Huber's work since G_1 does concentrate on $[-1, 1]$ and is of the right form, i.e.,

$$\frac{(g_1^{1/2})''}{g_1^{1/2}} = \frac{1}{4}\left\{\frac{2g_1''}{g_1} - \left(\frac{g_1'}{g_1}\right)^2\right\} = \frac{-\pi^2}{4}. \tag{2.10}$$

We can now see where Theorem 2.1 comes from. By Brown's identity (2.7) we have

$$\begin{aligned}\rho(m) &= \sup\{r(G) : G \text{ concentrating on } [-m, m]\}\\ &= 1 - \inf\{I(\Phi * G) : G \text{ concentrating on } [-m, m]\}\end{aligned}$$

which by property (i) of I is then equal to

$$1 - m^{-2} \inf\{I(\Phi_{1/m} * G) : G \text{ concentrating on } [-1, 1]\}$$

where Φ_σ is the $N(0, \sigma^2)$ c.d.f. By Lemma 2.1, the coefficient of m^{-2} should be approximately $I(G_1)$ for m large. Here is a formal proof.

PROOF. Since

$$r(G_m) = 1 - I(\Phi * G_m) = 1 - m^{-2} I(\Phi_{1/m} * G_1)$$

(2.2) follows from property (iii) of I. Since G_m^0 is least favorable

$$\rho(m) = r(G_m^0) = 1 - I(\Phi * G_m^0) = 1 - m^{-2} I(\Phi_{1/m} * G_1^{(m)}). \tag{2.11}$$

Suppose (without loss of generality, since $[-1, 1]$ is compact) that $G_1^{(m)} \to G$ weakly. Then so does $\Phi_{1/m} * G_1^{(m)}$ and by property (ii) of I and (2.11) we must have

$$I(G) \le \liminf_m m^2\{1 - \rho(m)\}.$$

On the other hand, by property (iv) of I,

$$m^2\{1 - \rho(m)\} \le m^2\{1 - \rho(G_m)\} = I(\Phi_{1/m} * G_1) \le I(G_1).$$

But by Lemma 2.1, $I(G) \geq I(G_1)$ with equality holding if and only if $G = G_1$. Claims (2.1) and (2.3) now follow from these inequalities. □

The proof of Theorem 2.2 uses Stein's identity in a similar way. By definition

$$R(\theta, \delta_m) = 1 - n^{-2} \int_{-\infty}^{\infty} \left\{ 2\psi_m'\left(\frac{z}{n}\right) - \psi_m^2\left(\frac{z}{n}\right) \right\} \phi(z - \theta)\, dz.$$

The theorem is clearly equivalent to

$$\sup\left\{ \left| \int_{-\infty}^{\infty} \left[\left\{ 2\psi_m'\left(\frac{z}{n}\right) - \psi_m^2\left(\frac{z}{n}\right) \right\} - \pi^2 \right] \phi(z - \theta)\, dz \right| : |\theta| \leq m \right\} = o(1). \tag{2.12}$$

But by (2.10)

$$2\,\bar{\psi}'\left(\frac{z}{n}\right) - \bar{\psi}^2\left(\frac{z}{n}\right) = \pi^2$$

for $|z| < m(1 + a_m) = n(1 - a_m^2)$. If we apply properties (b), (c) and (d) we can bound the sup in (2.12) by

$$\sup\left\{ \int_{|z| > m(1+a_m)} (c_m n^{-2} z^2 + b_m + \pi^2) \phi(z - \theta)\, dz : |\theta| \leq m \right\} + o(1)$$

The estimates in properties (e) and (f) complete the proof.

Proof of Corollary 2.1: We can write,

$$\delta(x, G_n) = x + n^{-1} \frac{h_n'\left(\frac{x}{n}\right)}{h_n\left(\frac{x}{n}\right)}, \tag{2.13}$$

where h_n is the density of $\Phi_{1/n} * G_1$, i.e.

$$h_n(x) = \int_{-\infty}^{\infty} g_1\left(x - \frac{y}{n}\right) \phi(y)\, dy. \tag{2.14}$$

Then we verify conditions (b) – (e). First note that

$$|h_n(x) - g_1(x)| \leq (a_m^2 m)^{-1} g_1(x)$$

if $|x| \leq 1 - a_m^2$, since $g_1(x) \geq (1 - |x|)^2$, $|x| \leq 1$. Next represent

$$h_n'(x) = \int g_1'\left(x - \frac{y}{n}\right) I_n(x, y) \phi(y)\, dy$$

$$h_n''(x) = \int g_1''\left(x - \frac{y}{n}\right) I_n(x, y) \phi(y)\, dy - n g_1'(1 -)\{\phi(n(1 + x)) + \phi(n(x - 1))\},$$

where $I_n(x, y)$ is the indicator of $\{n(x - 1) \leq y \leq n(x + 1)\}$.

By standard arguments we obtain, for universal C_1, C_2,

$$|h_n'(x) - g_1'(x)| \leq C_1\{n^{-1} + 1 - \Phi(m a_m^2)\} a_m^{-2} g_1(x)$$

$$|h_n''(x) - g_1''(x)| \leq C_2[m^{-1} + m\phi(m a_m^2) + \{1 - \Phi(m a_m^2)\}] a_m^{-2} g_1(x)$$

if $|x| \leq 1 - a_m^2$.

From these estimates it is easy to show that if $m a_m^6 \to \infty$ then conditions (b) and (c) of Theorem 2.2 hold for $\psi_m = -h_n'/h_n$. Next note that $\delta(x, G_n)$ is the mean of a distribution concentrating on $[-n, n]$ and $(\partial/\partial x)\, \delta(x, G_n)$ is its variance. Hence, by (2.14)

$$\left|\frac{h_n'(x)}{h_n(x)}\right| \le n^2(x+1) \qquad \left|\frac{h_n''(x)}{h_n(x)} - \left\{\frac{h_n'(x)}{h_n(x)}\right\}^2\right| \le n^2(n^2+1).$$

Therefore condition (d) holds, and from these estimates it is easy to see that conditions (e) and (f) are also satisfied if $ma_m^6 \to \infty$. □

3. The *p* variate case. Suppose X is $N_p(\theta, I)$ and that we want to estimate θ with quadratic loss. Then, the risk of an estimate δ is

$$R(\theta, \delta) = E_\theta \| \delta - \theta \|^2,$$

where $\| \ \|$ is the Euclidean distance. Conserving our previous notation, we consider the minimax risk of estimation given that $\|\theta\| \le m$, defined by $\rho_p(m) = \min \max\{R(\theta, \delta) : \|\theta\| \le m, \delta \in \mathscr{D}\}$.

By invariance the minimax estimate is Bayes with respect to a unique spherically symmetric least favorable prior distribution G_{mp}^0 concentrating (by analyticity considerations) on a finite number of spherical shells. We can again approximate G_{mp}^0 for large m.

Let J_t be the Bessel function of the first kind of order t, see Erdelyi et al. (1953), and let γ_t be its first positive zero. Let G_{1p} be the spherically symmetric distribution on the unit sphere $\{\theta : \|\theta\| \le 1\}$ with density given by

$$\begin{aligned} g_{1p}(\|x\|) &= C_p \|x\|^{-2t} J_t^2(\|x\|\gamma_t), \qquad && \|x\| \le 1, \\ &= 0, && \|x\| > 1, \end{aligned}$$

where

$$\begin{aligned} t &= \frac{p}{2} - 1 \qquad && \text{if } p \text{ is odd or divisible by 4} \\ &= -\left(\frac{p}{2} - 1\right) && \text{if } p \text{ is even and not divisible by 4} \end{aligned}$$

and c_p normalizes the density. It is well known that J_t has positive zeros (ibid, page 59) and by the standard representations (ibid, pages 2, 6) that $g_{1p}(0) > 0$. Moreover, $g_{1p}(r)$ is twice continuously differentiable on $[0, 1]$ and

(3.1) $$g_{1p}'(0) = g_{1p}'(1) = 0.$$

Let $$G_{mp}(s) = G_{1p}\left(\frac{s}{m}\right).$$

The generalization of Theorem 2.1 is then as follows.

THEOREM 3.1. *As* $m \to \infty$,

(3.2) $$\rho_p(m) = r(G_{mp}) + o(m^{-2})$$

(3.3) $$r(G_{mp}) = p - 4\gamma_t^2 m^{-2} + o(m^{-2});$$

and if $G_{1p}^{(m)}(s) = G_{mp}^{(0)}(ms)$ *then*

(3.4) $$G_{1p}^{(m)} \to G_{1p}$$

weakly as $m \to \infty$.

An analogue of Theorem 2.2 also holds. For simplicity we give the simplest example of asymptotically optimal estimates. Let

$$\bar{\psi}_p(r) = -\frac{g_{1p}'(r)}{g_{1p}(r)} = -\left\{2\gamma_t \frac{J_t'(\gamma_t r)}{J_t(\gamma_t r)} - \frac{(p-2)}{r}\right\}, \qquad r \ge 0.$$

$$\delta_m(x) = \left\{1 - n^{-1}\psi_{mp}\left(\frac{\|x\|}{n}\right)\|x\|^{-1}\right\}x,$$

where

$$\begin{aligned}\psi_{mp}(r) &= \bar{\psi}_p(r), & 0 \le r \le 1 - a_t^2\\ &= \bar{\psi}_p(1 - a_t^2) + \bar{\psi}_p'(1 - a_t^2)\{r - (1 - a_t^2)\}, & r > 1 - a_t^2\end{aligned}$$

and $n = m(1 + a_m)$. Then

$$\sup\left\{\left|R(\theta, \delta_m) - p + \frac{4\gamma_t^2}{m^2}\right| : \|\theta\| < m\right\} = o(m^{-2}) \tag{3.5}$$

if, for instance, $a_m \sim m^{-\epsilon}$, $0 < \epsilon < 1$.

NOTES:

1) As $m \to \infty$, $\forall x$, $\delta_{mp}(x) \to \left(1 - \frac{(p-2)}{\|x\|^2}\right)x$, Stein's (inadmissible) improvement for $p > 2$. Minimaxity of Stein's estimate follows.
2) These solutions have, for odd p, representations in terms of trigonometric and rational functions (Whittaker and Watson, 1927, page 364). In particular, for $p = 3$,

$$\begin{aligned}g_{13}(r) &= \frac{1}{2\pi}\frac{\sin^2(\pi r)}{r^2}, & 0 \le r \le 1\\ &= 0 & \text{otherwise,}\end{aligned}$$

and correspondingly

$$\gamma_{1/2} = \pi$$

which can be contrasted with the value $\gamma_{-1/2} = \frac{1}{2}\pi$ for $p = 1$.

These results are based on the general forms of Brown's and Stein's identities, which we give in the following form. Let $I(F)$ be the Fisher information for the p-variate location problem as defined for instance in Port and Stone (1974). If F has a density f with continuous partial derivatives

$$I(F) = \int_{R^p}\left\{\Sigma_{j=1}^p\left(\frac{\partial f}{\partial x_j}\right)^2(x)\right\}f^{-1}(x)\,dx.$$

Let

$$\delta(x) = x - \psi(x), \qquad \psi = (\psi_1, \cdots, \psi_p)$$

where $E_\theta\left|\frac{\partial}{\partial x_j}\psi_j(X)\right| < \infty$, $j = 1, \cdots, p$. Then

Brown's identity: For any prior distribution G,

$$r(G) = p - I(G * \Phi). \tag{3.6}$$

Stein's identity:

$$R(\theta, \delta) = p - E_\theta\left\{2\Sigma_{j=1}^p \frac{\partial}{\partial x_j}\psi_j(X) - \Sigma_{j=1}^p\psi_j^2(X)\right\}. \tag{3.7}$$

The generalization of Lemma 2.1 needed is

LEMMA 3.1. *The distribution G_{1p} uniquely minimizes $I(F)$ among all spherically symmetric F concentrating on the unit sphere and*

$$I(G_{1p}) = 4\gamma_t^2. \tag{3.8}$$

Moreover, $\sqrt{g_{1p}}$ *on* (0, 1) *satisfies the equation*

$$u''(r) + \frac{(p-1)}{r} u'(r) = -\gamma_t^2 u(r), \tag{3.9}$$

or equivalently

$$2\frac{g''_{1p}}{g_{1p}} - \left(\frac{g'_{1p}}{g_{1p}}\right)^2 + 2\frac{(p-1)}{r}\frac{g'_{1p}}{g_{1p}} = -4\gamma_t^2. \tag{3.10}$$

We can prove this lemma as in Huber (1974) (see also Huber, 1977) by considering the equivalent variational problem of minimizing $\int_0^\infty r^{p-1}\frac{\{f'(r)\}^2}{f(r)}\,dr$ subject to $\int_0^\infty r^{p-1}f(r)\,dr$ = constant. Equation (3.10) is equivalent to the Euler equation for the associated Lagrange problem of minimizing

$$\int_0^1 r^{p-1}\frac{\{f'(r)\}^2}{f(r)}\,dr - 4\gamma_t^2\int_0^1 r^{p-1}f(r)\,dr.$$

Convexity of the functional guarantees that a smooth solution of the Euler equation which satisfies the side conditions achieves the minimum. Unicity of a solution which is positive on (0, 1) is argued as in Huber (1974). Relation (3.8) follows by integrating (3.9) with respect to g and applying the identity

$$g''_{1p}(\|x\|) + \frac{(p-1)}{\|x\|}g'_{1p}(\|x\|) = \Sigma_{j=1}^p \frac{\partial^2}{\partial x_j^2} g_{1p}(\|x\|),$$

Gauss' theorem (e.g., Courant, 1937, page 401–402) and (3.1).

Claim (3.5) and similar results follow as in the one dimensional case when we note that if $\psi(x) = xw(\|x\|)/\|x\|$, where w is a smooth scalar function, then Brown's identity becomes

$$R(\theta,\delta) = p - E_\theta\left[2\left\{w'(\|x\|) + \frac{(p-1)}{\|x\|}w(\|x\|) - w^2(\|x\|)\right\}\right].$$

Generalizations in a variety of directions are possible. For example:
(1) to loss functions $l(\theta, \delta) = \Sigma_{j=1}^p \lambda_j(\delta_j - \theta_j)^2$ or equivalently to the case where X is $N(\theta, D)$ with D diagonal, known. Unfortunately Euler's equation now becomes a general elliptic partial differential equation.
(2) to study of the minimax risk over other sequences of growing regions. In general, this problem also seems very difficult. However, we note an interesting special case. The minimax risk over $\{\theta : \max_j |\theta_j| \le m\}$ is $p - p(\pi^2/m^2) + o(m^{-2})$ and is obtainable by using an asymptotically minimax estimate for each coordinate separately.

Acknowledgment. After submission of this paper I learned that B. Ya. Levit had just published and previously announced more extensive results of the same type (in Russian) in Levit (1980a, b). Further work is announced in Levit (1980c, d). The methods in this paper are somewhat different and perhaps simpler than Levit's.

REFERENCES

BICKEL, P. J. (1980). Minimax estimation of the mean of a normal distribution subject to doing well at a point. Unpublished technical report, Univ. of California, Berkeley.

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

CASELLA, G. and STRAWDERMAN, W. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 868–876.

COURANT, R. (1937). *Differential and Integral Calculus*, V. II. Interscience, New York.

ERDELYI, A. (1953). *Higher Transcendental Functions*, V. II. McGraw-Hill, New York.

GHOSH, M. N. (1964). Uniform approximation of minimax point estimates. *Ann. Math. Statist.* **35** 1031–1047.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1974). Fisher information and spline interpolation. *Ann. Statist.* **2** 1029–1034.

HUBER, P. J. (1977). Robust covariances. *Statistical Decision Theory and Related Topics* II. Academic, New York.

HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

LEVIT, B. YA. (1980a). On the second order asymptotically minimax estimates. *Theory Probab. Appl.* **25** 561–576.

LEVIT, B. YA. and BERHIN, P. E. (1980b). Second order asymptotically minimax estimates of the mean of a normal distribution. *Problems Inform. Transmission* **16** 60–79.

LEVIT, B. YA. (1980c). On the second order optimality in estimation theory. *Theory Probab. Appl.* **25** 653–654.

LEVIT, B. YA. (1980d). On some new results in the theory of second order optimality. *Theory Probab. Appl.* **25** 669–670.

MARAZZI, A. (1980). Robust Bayesian estimation for the linear model. Unpublished technical report, Fachgruppe fur Statistik, E. T. H. Zürich.

PORT, S. and STONE, C. (1974). Fisher information and the Pitman estimation of a location parameter. *Ann. Statist.* **2** 225–247.

WHITTAKER, E. T. and WATSON, G. N. (1927). *A Course of Modern Analysis.* Cambridge University Press, Cambridge.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720