MINIMAX ESTIMATION OF THE MEAN OF A NORMAL DISTRIBUTION SUBJECT TO DOING WELL AT A POINT

BY

P. J. BICKEL

TECHNICAL REPORT NO. 12
APRIL 1982

RESEARCH PARTIALLY SUPPORTED BY
ADOLPH C. AND MARY SPRAGUE MILLER FOUNDATION
FOR BASIC RESEARCH IN SCIENCE AND
OFFICE OF NAVAL RESEARCH CONTRACTS
NO. N00014-75-C-0444 AND N00014-80-C-0163.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY

MINIMAX ESTIMATION OF THE MEAN OF A NORMAL DISTRIBUTION SUBJECT TO DOING WELL AT A POINT

P. J. Bickel[1]

Department of Statistics
University of California
Berkeley, California

SUMMARY

We study the problem: Minimize $\max_\theta E_\theta(\delta(X)-\theta)^2$ subject to $E_0\delta^2(X) \leq 1-t$, $t > 0$, when $X \sim N(0,1)$. This problem arises in robustness questions in parametric models (Bickel (1982)). We

(1) Partially characterize optimum procedures.

(2) Show the relation of the problem to Huber's (1964) minimax robust estimation of location and its equivalence to a problem of Mallows on robust smoothing.

(3) Give the behaviour of the optimum risk for $t \to 0, 1$

and (4) Study some reasonable suboptimal solutions.

[1] *This research was supported in part by the Adolph C. and Mary Sprague Miller Foundation for Basic Research in Science and the U.S. Office of Naval Research Contract No. N00014-75-C-0444 and N00014-80-C-0163.*

The results of this paper constituted a portion of the 1980 Wald Lectures.

ISBN 0-12-589320-5

I. THE PROBLEM

Let $X \sim N(\theta, \sigma^2)$ where we assume σ^2 is known and without loss of generality equal to 1. Let δ denote estimates of θ (measurable functions of X), and

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

$$M(\delta) = \sup_\theta M(\theta,\delta)$$

For $0 \leq t \leq 1$, let

$$\mathcal{D}_t = \{\delta: M(0,\delta) \leq 1-t\}$$

and

$$\mu(t) = \inf\{M(\delta): \delta \varepsilon \mathcal{D}_t\}.$$

By weak compactness an estimate achieving $\mu(t)$ exists. Call it δ_t^*. Of course, $\mu(0) = 1$ and

$$\delta_0^* = X$$

while $\mu(1) = \infty$ and

$$\delta_1^* = 0.$$

Our purpose in this paper is to study δ_t^* and μ_t and approximations to them, based on a relation between the problem of characterizing μ and δ^* and Huber's classical (1964) minimax problem.

The study of δ^* and μ can be viewed as a special case, when the prior distribution is degenerate, of the class of restricted Bayes problems studied by Hodges and Lehmann (1952) and the subclass of normal estimation problems studied by Efron and Morris (1971).

Our seemingly artificial problem is fundamental to the study of the question: In the large sample estimation of a parameter η in the presence of a nuisance parameter θ, which we believe to be 0, how can we do well when $\eta = 0$ at little expense if we are wrong about η? This question is discussed in Bickel (1982).

The paper is organized as follows. In section II we sketch the nature of the optimal procedures, establish the connection to robust estimation and introduce and discuss reasonable suboptimal procedures. In sections III and IV we give asymptotic approximations to $\mu(t)$ and δ_t^* for t close to 0 and 1. Proofs here are sketched with technical details reserved for an appendix labeled (A) which is available only in the technical report version of this paper.

II. OPTIMAL AND SUBOPTIMAL PROCEDURES AND THE CONNECTION TO ROBUST ESTIMATION OF LOCATION

For $0 \leq \lambda \leq 1$ let,

$$M_\lambda(\delta) = (1-\lambda)m(\delta) + \lambda M(0,\delta)$$

$$\rho(\lambda) = \inf_\delta M_\lambda(\delta)$$

and let δ_λ be the estimate which by weak compactness achieves the inf. By standard arguments (see e.g. Neustadt (1976)) $\forall\ 0 < t < 1$ there exists $0 < \lambda(t) < 1$ such that

$$\delta_t^* = \delta_{\lambda(t)}, \qquad \mu(t) = \rho(\lambda(t)). \tag{2.1}$$

Given a prior distribution P on R define the Bayes risk of δ by

$$M(P,\delta) = \int M(\theta,\delta)\ P(d\theta)$$

and its risk,

$$R(P) = \inf_\delta M(P,\delta).$$

By arguing as in Hodges and Lehmann (1952) Thms. 1, 2 and using standard decision theoretic considerations,

$$\begin{aligned}\rho(\lambda) &= \inf_\delta \sup\{M(P,\delta) : P \in \mathcal{P}_\lambda\} \\ &= \sup\{R(P) : P \in \mathcal{P}_\lambda\}\end{aligned} \tag{2.2}$$

where $\mathcal{P}_\lambda$ is the set of all prior distributions P on $[-\infty, \infty]$ such that $P = (1-\lambda)K+\lambda I$ where K is arbitrary and I is point mass at 0. In fact, there exists a proper least favorable distribution $P_\lambda \varepsilon \mathcal{P}_\lambda$ against which δ_λ is necessarily Bayes. The distribution P_λ is unique and symmetric about 0. Unfortunately it concentrates on a denumerable set of isolated points. This fact as well as the approximation theorems which represent the only analytic information we have so far on $\mu(t)$, δ_t^* are related to the "robustness connection" which we now describe.

If ψ denotes functions from R to R, P, F, K probability distributions on $[-\infty, \infty]$ and * convolution let

$$\mathcal{F}_{\lambda o} = \{F : F = P * \Phi,\ P \varepsilon \mathcal{P}\}$$
$$= \{F : F = (1-\lambda)K*\Phi + \lambda\Phi,\ K \text{ arbitrary}\}$$

$$I(F) = \int \frac{[f'(x)]^2}{f(x)}\ dx$$

if F has an absolutely continuous density f with derivative f'.

$$= \infty \quad \text{otherwise.}$$

If $I(F) < \infty$ let

$$V(F,\psi) = \int (\psi^2(x)\ f(x) + 2\psi(x)\ f'(x))dx \tag{2.3}$$

if $\int \psi^2(x)\ F(dx) < \infty$

$$= \infty \quad \text{otherwise.}$$

By integration by parts if ψ is absolutely continuous and $\int|\psi'(x)|F(dx) < \infty$

$$V(F,\psi) = \int \psi^2(x)F(dx) - 2\int \psi'(x)F(dx). \tag{2.4}$$

Given δ define

$$\psi(x) = x - \delta(x). \tag{2.5}$$

Then, it is easy to show by direct computation

$$M(P,\delta) = 1 + V(P*\Phi,\ \psi) \tag{2.6}$$

a formula due to Stein (Hudson (1978)) if P is a point mass.

By minimizing (2.6) we get

$$R(P) = 1 - I(P*\Phi) \tag{2.7}$$

achieved when

$$\psi = \frac{-f'}{f}$$

where f is the density of $P*\Phi$, a special case of an identity of Brown (1971).

Standard minimax arguments yield that if $\mathcal{F}$ is any convex weakly closed set of distributions on $[-\infty, \infty]$ with finite Fisher information then

$$V(F_o, \psi_o) = \sup_{\mathcal{F}} V(F, \psi_o) = \inf_{\psi} V(F_o, \psi) = I(F_o)$$

where F_o minimizes I(F) over $\mathcal{F}$ and

$$\psi_o = \frac{-f'_o}{f_o} \tag{2.8}$$

and f_o is the density of F_o. Specializing to $\mathcal{F}_{\lambda o}$ we obtain

$$\begin{aligned} \rho(\lambda) - 1 &= - I(F_{\lambda o}), \\ \delta_\lambda(x) &= x + \frac{f'_{\lambda o}}{f_{\lambda o}}(x) \end{aligned} \tag{2.9}$$

where $F_{\lambda o}$ is the least favorable distribution in $\mathcal{F}_{\lambda o}$ and $f_{\lambda o}$ is its density. The characterization of P_λ we mentioned follows immediately from (2.9) and theorem 2 of Bickel and Collins (1982).

For $\mathcal{F}$ as above, Huber (1964) essentially considered the game (with "Nature" as player I) and payoff (to I),

$$W(F,\psi) = \int \psi^2(x)f(x)dx/(\int\psi(x)\ f'(x)dx)^2.$$

Here ψ is the score function of an (M) estimate and W its asymptotic variance under F. (Huber restricted ψ, for instance to continuously differentiable functions with compact support, redefining the denominator of W to be $\int \psi'(x)\ F(dx)$ and permitting $I(F) = \infty$. But this seems inessential.) Here again the game has a value,

$$W(F_o,\psi_o) = \mathrm{Sup}_F\, W(F,\psi_o) = \inf_\psi W(F_o,\psi) = I^{-1}(F_o)$$

where F_o, ψ_o are the same strategies as for the payoff V.

$\mathcal{F}_{\lambda o}$ arose in the context of Huber's game in connection with robust smoothing, Mallows (1978), (1980). He posed the problem of minimizing I(F) for $F \varepsilon \mathcal{F}_{\lambda o}$ and conjectured that K_λ corresponding to the optimal P_λ concentrates on $\{kh: k = \pm 1, \pm 2, \ldots\}$, for some $h > 0$, and assigns mass

$$K_\lambda\{kh\} = \frac{1}{2}\lambda(1-\lambda)^{|k|-1} \quad \forall k.$$

As of this writing it appears that this conjecture is false although a modification of D. Donoho in which the support is of the form $\{\pm(a+kh): a,h > 0,\ k = 0,1,\ldots\}$ may be true.

The Efron-Morris Estimates

Let

$$\mathcal{F}_{\lambda 1} = \{F: F = \lambda\Phi + (1-\lambda)G,\ G \text{ arbitrary}\}. \tag{2.10}$$

$\mathcal{F}_{\lambda 1}$ is Huber's (1964) contamination model. As Huber showed, the optimal $F_{\lambda 1}$ has $-\frac{f'}{f}$ of the form

$$\begin{aligned}\psi_m(x) &= x \qquad\quad , \quad |x| \le m \\ &= m \text{ sgn } x, \quad |x| > m.\end{aligned} \tag{2.11}$$

The estimate corresponding to ψ_m in the sense of (2.5) is given by

$$\begin{aligned}\overline{\delta}_m(x) &= 0 \qquad\qquad\quad , \quad |x| \le m \\ &= x - m \text{ sgn } x, \quad |x| > m.\end{aligned} \tag{2.12}$$

This is a special case of the limited translation estimates proposed by Efron and Morris (1971) as reasonable compromises between Bayes and minimax estimates in the problem of estimating θ when θ has a normal prior distribution. We will call $\overline{\delta}_m$ the E-M estimate. Since $\overline{\delta}_m$ is not analytic it cannot be optimal. Nevertheless it has some attractive features.

The M.S.E. of δ_m is given by

$$M(\theta,\overline{\delta}_m) = 1 + m^2 + (\theta^2-(1+m^2))(\Phi(m+\theta)+\Phi(m-\theta)-1)$$
$$- ((m-\theta)\Phi(m+\theta) + (m+\theta)\Phi(m-\theta)).$$

Since $-2\psi_m' + \psi_m^2$ is an increasing function of $|x|$ we remark, as did Efron and Morris, that $M(\theta,\overline{\delta}_m)$ is an increasing function of $|\theta|$ with

$$M(\overline{\delta}_m) = M(\infty,\overline{\delta}_m) = 1 + m^2.$$

For fixed λ the $m(\lambda)$ which minimizes $M_\lambda(\overline{\delta}_m)$ is the unique solution of the equation

$$2\Phi(m) - 1 + \frac{2\Phi(m)}{m} = \lambda^{-1}. \tag{2.13}$$

This is also the value of m which corresponds to F_{λ_1}. We deduce the following weak optimality property: Let ψ correspond to δ by (2.5) in the following.

$$\mathcal{D}_\infty = \{\delta: E_\theta|\delta'(X)| < \infty, \forall\theta; \lim_{|x|\to\infty}[\psi^2(x)-2\psi'(x)]$$
$$= \sup_x[\psi^2(x)-2\psi'(x)]\}$$

$\mathcal{D}_\infty$ is a subclass of estimates which achieve their maximum risk at $\pm\infty$.

Theorem 2.1. If $m(\lambda)$ is given by (2.13) then $\overline{\delta}_{m(\lambda)}$ is optimal in $\mathcal{D}_\infty$, i.e.

$$M_\lambda(\overline{\delta}_{m(\lambda)}) = \min\{M_\lambda(\delta): \delta \in \mathcal{D}_\infty\}.$$

Proof. By (2.6) and (2.4) if $E_\theta|\delta'(X)| < \infty, \forall\theta$,

$$M_\lambda(\delta) - 1 = \sup\{V(F,\psi) : F \in F_{\lambda o}\} \le \sup\{V(F,\psi):$$
$$F \in F_{\lambda 1}\} \le \sup_x\{\psi^2(x) - 2\psi'(x)\}.$$

These inequalities become equalities for $\delta \in \mathcal{D}_\infty$ by letting $|\theta| \to \infty$ in $M(\theta,\delta)$. The result follows from the optimality property of $F_{\lambda 1}$.

The Pretesting Estimates

There is a natural class of procedures which are not in $\mathcal{D}_\infty$ and are natural competitors to the E-M estimates. A typical member of this class is given by

$$\begin{aligned} \tilde{\delta}_m(x) &= 0, \quad |x| \leq m \\ &= x, \quad |x| > m. \end{aligned}$$

Implicitly, in using $\tilde{\delta}_m$ we test H: $\theta=0$ at level $2(1-\Phi(m))$. If we accept we estimate 0, otherwise we use the minmax estimate of X. We call these pretesting estimates. The ψ function corresponding to $\tilde{\delta}_m$ is of the type known as "hard rejection" in the robustness literature.

Comparison of E-M and Pretesting Estimates

Hard rejection does not work very well--nor do pretesting estimates. Both the E-M and pretesting procedures have members which are approximately optimal for λ close to 1 or what amounts to the same, t close to 1. However, the pretesting procedures behave poorly for λ (or t) close to 0. This is discussed further in sections III and IV. The following table gives the maximum M.S.E. of $\bar{\delta}$ and $\tilde{\delta}$ which have M.S.E. equal to 1-t at 0 as a function of t. The E-M rules always do better for the values tabled, spetacularly better in the ranges of interest. This is consistent with results of Morris et al. (1972) who show that Stein type rules render pretesting type rules inadmissible in dimension 3 or higher.

Notes:

(1) The connection between restricted minmax and more generally restricted Bayes and robust estimation was independently discovered by A. Marazzi (1980).

(2) Related results also appear in Berger (1982). His approach seems related to that of Hampel in the same way as ours is to that of Huber.

TABLE I. Maximum M.S.E. and Change Point m as a Function of M.S.E. at 0 for $\bar{\delta}_m$ *and* $\tilde{\delta}_m$

$M(0,\delta)$	$M(\delta)$		m	
	E-M	Pretest	E-M	Pretest
.1	2.393	3.626	1.180	2.500
.2	1.756	2.839	.869	2.154
.3	1.452	2.383	.672	1.914
.4	1.275	2.058	.525	1.716
.5	1.164	1.805	.405	1.538
.6	1.092	1.597	.303	1.367
.7	1.046	1.418	.215	1.193
.8	1.018	1.262	.137	1.002
.9	1.004	1.124	.065	.728
1.0	1.000	1.000	.000	.000

III. THE BEHAVIOUR OF $\mu(t)$ FOR SMALL t

Let

$$\Delta(t) = \mu(t) - 1$$

Theorem 3.1. As $t \to 0$

$$\Delta(t) = o(t^2) \tag{3.1}$$

but

$$t^{-(2+\varepsilon)}\Delta(t) \to \infty \tag{3.2}$$

for every $\varepsilon \to 0$.

Notes:

(1) The E-M rule $\overline{\delta}_m$ with $M(0,\overline{\delta}_m) = 1-t$ has $M(\overline{\delta}_m) =$ $= 1 + \frac{\pi}{2} t^2 + o(t^2)$. However our proof of (3.2) suggests that the asymptotic improvement over this rule is attainable only by very close mimicking of the optimal rule. This does not seem worthwhile because of the oscillatory nature of the optimal rule.

(2) The pretesting rule $\tilde{\delta}_m$ with $M(0,\tilde{\delta}_m) = 1-t$ has $M(\tilde{\delta}_m) = 1 + \Omega(t)$. This unsatisfactory behaviour is reflected in Table I.

We need

Lemma 3.1. Let $\lambda(t)$ be as in (2.1). Then λ is continuous.

Proof. By the unicity of δ_λ and weak compactness, $M(0,\delta_\lambda)$ is continuous and strictly decreasing in λ on $[0,1]$. The lemma follows.

Lemma 3.2. As $\lambda \to 0$

$$\lambda^{-2}(1-\rho(\lambda)) \to \infty. \tag{3.3}$$

Lemma 3.3. As $\lambda \to 0$

$$\lambda^{-2+\varepsilon}(1-\rho(\lambda)) \to 0 \tag{3.4}$$

for every $\varepsilon > 0$.

Proof of Theorem 3.1 from Lemmas 3.1-3.3

Claim 3.1. For any sequence $t_k \to 0$ let $\lambda_k = \lambda(t_k)$ so that,

$$1 - \rho(\lambda_k) = \lambda_k t_k - (1-\lambda_k)\Delta(t_k). \tag{3.5}$$

Then, by lemma 3.1, $\lambda_k \to 0$ and

$$\begin{aligned}\lambda_k^{-2}(1-\rho(\lambda_k)) &\le \lambda_k^{-2} \max_x\{\lambda_k x - (1-\lambda_k)\frac{\Delta(t_k)}{t_k^2} x^2\} \\ &= O(t_k^2/\Delta(t_k)).\end{aligned} \tag{3.6}$$

By (3.3), $\dfrac{t_k^2}{\Delta(t_k)} \to \infty$ and (3.1) follows.

Claim 3.2. Note that

$$1 - \rho(\lambda) \ge \max_t[\lambda t - (1-\lambda)\Delta(t)] \tag{3.7}$$

If $\Delta(t_k) \le C\, t_k^{2+\varepsilon}$ for some $C \le \frac{1}{2}$, $\varepsilon > 0$, $t_k \to 0$ put $\lambda_k = t_k^{1+\varepsilon}$ to get

$$1 - \rho(\lambda_k) \ge t_k^{2+\varepsilon}(1+o(1)) \ge \lambda_k^{2-\frac{\varepsilon}{1+\varepsilon}}(1 + o(1))$$

a contradiction to (3.4).

Proof of Lemmas 3.1-3.3

The proof proceeds via several sublemmas.

Lemma 3.4. Let $\{\nu_n\}$ be a sequence of Bayes prior distributions on R and let δ_n be the corresponding Bayes estimates. Suppose that, as $n \to \infty$,

$$M(\delta_n) \to 1. \tag{3.8}$$

Then

$$\delta_n(x) \to x \quad \text{a.e.} \tag{3.9}$$

$$\frac{\nu_n(I_1)}{\nu_n(I_2)} \to 1 \tag{3.10}$$

for any pair of intervals I_1, I_2 of equal length.

Proof. By Sacks' (1963) theorem, there exists a subsequence $\{n_k\}$ such that $\{\delta_{n_k}\}$ converge regularly to δ and

$$\delta_{n_k}(x) \to \delta(x) \quad \text{a.e.}$$

where δ is generalized Bayes with respect to ν (σ finite) such that for some sequence $\{a_k\}$

$$\nu_{n_k}/\nu_{n_k}(-a_k, a_k) \to \nu \tag{3.11}$$

weakly. But, by (3.8) and regular convergence $M(\delta) \leq 1$ and hence δ is minmax. Therefore, $\delta(x) = x$ a.e. and (3.9) follows by Sacks' theorem. Since δ is generalized Bayes with respect to ν, ν must be proportional to Lebesgue measure and (3.10) follows from (3.11).

Lemma 3.5. If $\lambda \to 0$, and $P_\lambda = (1-\lambda)K_\lambda + \lambda I$

$$\frac{1}{\lambda} \int_{-\infty}^{\infty} \phi(\theta) K_\lambda(d\theta) \to \infty.$$

Proof. Since $M(\tilde{\delta}_\lambda) \to 1$

$\{P_\lambda\}$ satisfies (3.10) as $\lambda \to 0$.

Therefore for all $a > 0$

$$\frac{P_\lambda[0,a)}{P_\lambda(0,a)} \to 1.$$

Hence,

$$\lambda = o(K_\lambda(0,a)). \tag{3.12}$$

By the same argument

$$\frac{K_\lambda(0,a)}{K_\lambda(0,1)} \to a, \ a \leq 1$$

and hence

$$[K_\lambda(0,1)]^{-1} \int_0^1 \phi(\theta) K_\lambda(d\theta) \to \int_0^1 \phi(\theta) d\theta. \tag{3.13}$$

The lemma follows from (3.12) and (3.13).

Proof of Lemma 3.2. We compute the Bayes risk of a reasonable E-M estimate, viz. $\bar{\delta}_\lambda$. We claim that

$$\lambda^{-2}(M(P_\lambda,\bar{\delta}_\lambda) - 1) \to -\infty \tag{3.14}$$

Since $\rho(\lambda) = M(P_\lambda) \le M(P_\lambda,\bar{\lambda})$ the lemma will follow. To prove (3.14) apply Stein's formula to get

$$M(P_\lambda,\bar{\delta}_\lambda) - 1 = \int_{-\infty}^{\infty} E_\theta(\bar{\delta}_\lambda(X)-X)^2 P_\lambda(d\theta)$$
$$+ 2\int_{-\infty}^{\infty} E_\theta(\bar{\delta}_\lambda(X) - 1)P_\lambda(d\theta).$$

The expression in (3.15) is bounded by

$$\lambda^2 - 2\int_{-\infty}^{\infty} [\Phi(\lambda-\theta) - \Phi(-\lambda-\theta)]P_\lambda(d\theta)$$
$$\le \lambda^2 - 2(1-\lambda)\int_{-\infty}^{\infty} [\Phi(\lambda-\theta) - \Phi(-\lambda-\theta)]K_\lambda(d\theta).$$

Since $\Phi(\lambda-\theta) - \Phi(-\lambda-\theta) \ge \lambda\phi(\theta)$ for $\lambda \le \sqrt{2 \log 2}$ we can apply Lemma 3.5 to conclude that

$$\lambda^{-2}\int_{-\infty}^{\infty} [\Phi(\lambda-\theta) - \Phi(-\lambda-\theta)]K_\lambda(d\theta) \to \infty$$

and claim (3.14) and the lemma follow.

We sketch the proof of Lemma 3.3. Details are available in (A).

Proof of Lemma 3.3. It suffices for each $\varepsilon > 0$ to exhibit a sequence of prior distributions $\bar{P}_\lambda$ such that

$$\lambda^{-2+\varepsilon}(1 - R(\bar{P}_\lambda)) \to 0. \tag{3.16}$$

By Brown's identity, claim (3.16) is equivalent to

$$\int \frac{[f_\lambda']^2}{f_\lambda} = o(\lambda^{2-\varepsilon}) \tag{3.17}$$

for f_λ the density of $\bar{P}_\lambda * \Phi$. Here is the definition of $\bar{P}_\lambda$. Let

$$e_\tau(x) = \frac{1}{\pi\tau}(1 + (\frac{x}{\tau})^2)^{-1}.$$

Write ϕ_σ, (Φ_σ) for the normal $(0,\sigma^2)$ density (d.f.). Given $k \geq 1$, let h be a (C^∞) function from R to R such that

$$|h(x)| \leq c_r(1 + |x|^r)^{-1} \text{ for all } r > 0,\ x, \text{ some } c_r. \tag{3.18}$$

$$\int_{-\infty}^{\infty} h(x)\,dx = 1$$

$$\int_{-\infty}^{\infty} x^j h(x)\,dx = 0, \quad 1 \leq j \leq 2k-1.$$

An example of h satisfying these conditions is

$$h(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \exp\{-itx-t^{2k}\}dt.$$

Let

$$h_\sigma(x) = \frac{1}{\sigma} h(\frac{x}{\sigma}).$$

Set, for c_k in (3.18),

$$\sigma = 2\pi\lambda m c_k, \tag{3.19}$$

and

$$m = \lambda^{-(2k+2)/(2k+3)}. \tag{3.20}$$

If $\overline{P}_\lambda = (1-\lambda)\, L_\lambda + \lambda I$ define $\overline{P}_\lambda$ by the density of L_λ,

$$\ell_\lambda = (e_m - \lambda h_\sigma)(1-\lambda)^{-1}$$

for $\lambda \leq 1$ and $\sigma \leq 1$. By construction L_λ is a probability measure (see (A)).

Let

$$g_\lambda = e_m * \phi$$

$$q_\sigma = \phi - h_\sigma * \phi$$

Then,

$$\int \frac{[f'_\lambda]^2}{f_\lambda} = \int \frac{[g'_\lambda]^2}{f_\lambda} + 2\lambda \int \frac{g'_\lambda q'_\sigma}{f_\lambda} + \lambda^2 \int \frac{[q'_\sigma]^2}{f_\lambda} \ .$$

It is shown in (A) that,

$$\int \frac{[g'_\lambda]^2}{f_\lambda} = O(m^{-2}) \tag{3.21}$$

$$\int \frac{g'_\lambda q'_\sigma}{f_\lambda} = O(m^{-1}) \tag{3.22}$$

$$\int \frac{[q'_\sigma]^2}{f_\lambda} = O(m\sigma^{2k}). \tag{3.23}$$

We combine (3.21) - (3.23) to get

$$\int \frac{[f'_\lambda]^2}{f_\lambda} = O(m^{-2}+\lambda m^{-1}+m\lambda^2\sigma^{2k}) = O(\lambda^{2-2(k+3)^{-1}}).$$

Since k is arbitrary we have proved (3.17) and the lemma.

The theorem is proved.

IV. THE BEHAVIOUR OF $\mu(t)$ FOR t CLOSE TO 1

We sketch the proof of,

Theorem 4.1. As $t \to 1$

$$\mu(t) = 2|\log(1-t)|(1+o(1)). \tag{4.1}$$

If $\delta_t \in \mathcal{D}_t$ and

$$\delta_t(x) = 0, \quad |x| \le g(t)$$

$$\sup\{|\delta_t(x)-x| : \quad |x|>g(t)\} = o(g(t)) \tag{4.2}$$

$$g(t) = \sqrt{2|\log(1-t)|}(1 + o(1)).$$

then

$$M(\delta_t) = \mu(t)(1 + o(1)). \tag{4.3}$$

Note. It is easy to see that both E-M and pretest estimates which are members of $\mathcal{D}_t$ satisfy (4.2) and are optimal in this sense. The approximation (4.1) is thus crude and not practically useful.

Lemma 4.1. As $\lambda \to 1$

$$\rho(\lambda) = 2(1-\lambda)|\log(1-\lambda)|(1+o(1)). \tag{4.4}$$

Moreover if $\{\delta_\lambda\}$ is any sequence of estimates such that

$$\begin{aligned} \delta_\lambda(x) &= 0, \quad |x| \le c(\lambda) \\ |\delta_\lambda(x)-x| &\le b(\lambda), \quad |x| > c(\lambda) \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} c(\lambda) &= [2|\log(1-\lambda)|]^{1/2}(1+o(1)) \\ b(\lambda) &= o(c(\lambda)) \end{aligned} \tag{4.6}$$

then

$$M(0,\delta_\lambda) = \frac{2}{\sqrt{\pi}}(1-\lambda)|\log(1-\lambda)|^{1/2}(1+o(1)) \tag{4.8}$$

$$M(\delta_\lambda) = 2|\log(1-\lambda)|(1+o(1)) \tag{4.9}$$

and hence

$$M_\lambda(\delta_\lambda) = \rho(\lambda)(1+o(1)).$$

Proof. We establish the lemma by

(i) For every $\gamma > 0$ exhibiting $\tilde{P}_\lambda$ such that

$$R(\tilde{P}_\lambda) \ge 2(1-\lambda)|\log(1-\lambda)|(1-\gamma)(1+o(1)).$$

(ii) Showing that δ_λ given in (4.5) satisfy (4.8) and

$$M(\delta_\lambda) \le 2|\log(1-\lambda)|(1+o(1)).$$

Since, by (4.8),

$$M(0,\delta_\lambda) = o((1-\lambda)|\log(1-\lambda)|)$$

and

$$R(\tilde{P}_\lambda) \le \rho(\lambda) \le (1-\lambda)M(\delta_\lambda) + \lambda M(0,\delta_\lambda)$$

the lemma will follow. Here is $\tilde{P}_\lambda$. Let,

$$\begin{aligned} \varepsilon &= 1 - \lambda \\ a &= a(\varepsilon) = \sqrt{|2\log\varepsilon|(1-\gamma)}, \quad \gamma > 0 . \end{aligned} \tag{4.10}$$

Let $\tilde{P}_\lambda$ put mass $\frac{\varepsilon}{2}$ at $\pm a$, and λ at 0. The calculations establishing (i) and (ii) are in (A).

Proof of Theorem 4.1. Putting $\lambda = t$ we must have,

$$\rho(t) \leq (1-t)(\mu(t)+t).$$

Therefore, by Lemmas 3.1 and 4.1, as $t \to 1$,

$$\mu(t) \geq |2 \log(1-t)|(1+o(1)).$$

By (4.8) and (4.9) we can find members of $\mathcal{D}_t$ with maximum risk $|2 \log(1-t)|(1+o(1))$ and the theorem follows.

V. ACKNOWLEDGMENT

A brief but stimulating conversation with P. J. Huber was very helpful.

VI. REFERENCES

Berger, J. (1982). Estimation in Continuous Exponential Families: Bayesian Estimation Subject to Risk Restrictions and Inadmissibility Results. Statistical Decision Theory and Related Topics III. S. Gupta and J.O. Berger Eds. Academic Press, New York.

Bickel, P.J. (1982). Parametric Robustness and Prestesting. Submitted to *J. Amer. Statist. Assoc.*

Bickel, P.J. and Collins, J. (1982). "Minimizing Fisher Information Over Mixtures of Distributions." *Sankhya*, to appear.

Brown, L.D. (1971). "Admissible Estimators, Recurrent Diffusions and Insoluble Boundary Value Problems." *Ann. Math. Statist. 42*, 855.

Efron, B. and Morris, C. (1971). "Limiting the Risk of Bayes and Empirical Bayes Estimates: Part I. The Bayes Case." *J. Amer. Statist. Assoc. 66*, 807.

Hodges, J.L. and Lehmann, E.L. (1952). "The Use of Previous Experience in Reaching Statistical Decisions." *Ann. Math. Statist. 23*, 396-407.

Huber, P.J. (1964). "Robust Estimation of a Location Parameter." *Ann. Math. Statist. 35*, 73.

Hudson, H.M. (1978). "A Natural Identity for Exponential Families with Applications in Multiparameter Estimation." *Ann. Statist. 6*, 473.

Mallows, C.L. (1978). Problem 78.4. *S.I.A.M. Review*.

Marazzi A. (1980). Robust Bayesian Estimation for the Linear Model. Tech. Report E.T.H. Zürich.

Morris, C.N., Radhadrishnan, R. and Sclove, S.L. (1972). "Non-optimality of Preliminary Test Estimators for the Mean of a Multivariate Normal Distribution." *Ann. Math. Statist. 43,* 1481.

Neustadt, L.W. (1976). *Optimization.* Princeton University Press.

Port, S. and Stone, C. (1974). "Fisher Information and the Pitman Estimator of a Location Parameter." *Ann. Statist. 2,* 25.

Sacks, J. (1963). "Generalized Bayes Solutions in Estimation Problems." *Ann. Math. Statist. 34,* 751.