Minimax Estimators Incorporating Vague Prior Knowledge in Spherically Symmetric Location Problems
by
M.E. Bock* and Paul Klembeck
Technical Report #82-29

Department of Statistics
Purdue University
August 1982

*This research was supported by NSF grant #MCS-81-01670.
Introduction

The problem considered is the estimation of the location vector $\theta$ of a spherically symmetric distribution based on an observation $X$ from the distribution. It is assumed that the vector's dimension $p$ is greater than or equal to three. Initially, we consider squared error loss, later generalizing to loss functions concave and nondecreasing in squared error. The estimators presented are min-max and dominate the estimator $\hat{\theta}_0(X) = X$. They employ "vague" prior information in the following sense. The "vague" prior information is that $\theta$ is "likely" to lie in a certain convex region $G$. The estimate of $\theta$ is found by shrinking values of $X$ outside $G$ toward $G$. The amount of shrinkage depends on how far $X$ is from $G$. The closer $X$ is to $G$, the greater the fraction of the distance from $X$ to $G$ is reduced by the shrinkage estimator.

The estimators may be defined to depend on the loss function and the density $f(||X-\theta||)$ only through a single constant called the shrinkage factor.

In the case of squared error loss, the shrinkage factor is

$$2(p-2)\inf_{t>0} q(t)$$

where
\[ q(t) = \int_{t}^{\infty} uf(u)du. \]

This same factor was used by Berger (1975) to exhibit minimax estimators which shrink to a point. There and here, attention is restricted to the class of densities \( f \) for which

\[ \inf_{t>0} q(t) > 0. \]

Although the shrinkage factor \( 2(p-2)\inf_{t>0} q(t) \) is the best possible for the normal distribution, it is conservative for many densities \( f \). For instance, the factor satisfies

\[ 2(p-2)\inf_{t>0} q(t) \leq 2/E_{\theta=0}[||X||^{-2}] \]

In the case of shrinkage to a ball of fixed radius centered at some known vector the shrinkage factor is \( 2/E_{\theta=0}[||X||^{-2}] \) if it is known that \( q(t) \) is nondecreasing. (See Bock (1981).)

**Estimators**

Let \( G \) be a p-dimensional convex region in \( \mathbb{R}^p \) with twice-
differentiable boundary hypersurface $M = \partial G$. Let $N(P)$ be the number of principal curvatures at the point $P$ on $M$ which are zero. Let $ar{\rho}(P)$ be the average of the nonzero radii of curvature of $M$ at $P$.

**Theorem 1.** Let $X$ be a $p$-dimensional random vector with spherically symmetric distribution about $\theta$ such that $E[||X-\theta||^2] < \infty$ and $E[||X-\theta||^{-2}] < \infty$. For $p \geq 3$, under the loss

$$L(\theta, \hat{\theta}) = ||\theta - \hat{\theta}||^2,$$

the following minimax estimator is at least as good as $X$:

$$\hat{\theta}(X) = X - r(||X-P||^2,P)/[\bar{\rho}(P)||X-P||](X - P)$$

where

(a) $P = P(X)$ is the projection of $X$ to $G$, i.e.,

$$||X-P||^2 = \inf_{Q \in G} ||X-Q||^2;$$

(b) for each $P$ on $M$, $r(t,P)$ is nondecreasing and differentiable
in $t$ on $[0, \infty)$ such that for $t > 0$,

$$0 \leq r(t, P) \leq a(p - N(P) - 2) \inf_{s \geq 0} q(s)$$

where the density of $X$ is $f(||X - \theta||)$ and

$$q(s) = \int_{S} u f(u) du / f(s).$$

**Note:** If $X$ is in $G$ or $\bar{G}$, then $\delta(X) = X$.

**Remark.** Of course the result is not meaningful unless $\min_{s \geq 0} q(s)$ is positive for the spherically symmetric distribution considered. In the case that the distribution is a mixture of normals, then

$$\{\min_{s \geq 0} q(s)\} = q(0)$$

since $q$ is increasing. For the case of the standard normal distribution this quantity is one. Berger [1975] has considered this class of spherically symmetric distributions and shows that

$$\{\inf_{s \geq 0} q(s)\}$$

is positive if there exist $\alpha > 0$ and $K > 0$ for which

$$h(s^2) = f(s)e^{\alpha s^2}$$

is nonzero and nondecreasing if $s^2 > K$; that is, $f$ is not too light tailed. For example, he considers the density $f(s) = Cs^m e^{-s^2/2}$ and shows that $\{\inf_{s \geq 0} q(s)\}$ is one for $m > 0$.

**Proof of Theorem I.** Because the estimator $X$ is minimax with constant risk for all values of $\theta$, it suffices to show that $\Delta$ is nonpositive
for all values of $\theta$ where the difference in risks is

$$\Delta = E[||\delta(X)-\theta||^2] - E[||X-\theta||^2].$$

Using the definition of $\delta$, we may write

$$\Delta = E[r^2(||X-P||^2, P)/||X-P|| + \bar{c}(P))^2I_{G^C}(X)]$$

$$- 2E[r(||X-P||^2, P)(X-\theta)^t(X-P)/||X-P||(||X-P|| + \bar{c}))I_{G^C}(X)]$$

since $P = P(X) = X$ for $X$ in $G$. Thus

$$\Delta = \int_{G^C} [r^2(||X-P(X)||^2, P(X))/||X-P(X)|| + \bar{c}(P(X)))^2$$

$$- 2r(||X-P(X)||^2, P(X))(X-\theta)^t(X-P(X))/||X-P(X)||$$

$$(||X-P(X)|| + \bar{c}(P(X)))]f(||X-\theta||)dV(X)$$

where $dV$ is the volume element in $G^C$. Let $N_p$ denote the outward unit normal vector to $M$, the boundary hypersurface of $G$, at the point $P$ on $M$. Letting $M$ be oriented with this normal, denote by $K_i(P)$, $i=1,...,p-1$, the principal curvatures of $M$ at $P$. Let $dA$ be the element of surface area on $M$. With $P=P(X)$, reparameterize $X$ in $G^C$ by the map $X = P + tN_p$ where $t \geq 0$ and $P$ is in $M$. Thus for $N(P)$ equal to the number of the $K_i(P)$'s which are zero,
\[ \tilde{\rho}(P) = \sum_{1 < i < p-1 \atop K_i(P) > 0} (K_i(P))^{-1/(p - N(P) - 1)}. \]

By the theorem of the Appendix, the volume element on \( S^C \) is

\[ dV = \prod_{i=1}^{p-1} (K_i(P)t + 1) dA(P) dt. \]

Thus

\[ \Delta = \int \int \left[ \frac{r^2(t^2, P)}{(t + \tilde{\rho}(P))^2} \right. \\
- 2r(t^2, P)(P + tN_P - \theta) \left. \frac{tN_P}{(t + \tilde{\rho}(P))} \right] \\
f(||P + tN_P - \theta||)^{p-1} \prod_{i=1}^{p-1} (K_i(P)t + 1) dt dA(P). \]

Define

\[ q(s) = \begin{cases} \int uf(u) du / f(s) & \text{for } f(s) > 0 \\ S & f(s) = 0 \end{cases}. \]
Because
\[
\frac{\partial}{\partial t} \left( \int_{||P + tN_p - \theta||}^{\infty} uf(u) du \right) = -f(||P + tN_p - \theta||).
\]

\[
\cdot (P + tN_p - \theta) tN_p,
\]

integration by parts implies that

\[
(*) \quad - \int_0^{\infty} 2r(t^2, P)(P + tN_p - \theta) tN_p / (t + \overline{\rho}(P)) \sum_{i=1}^{p-1} \left( K_i(P) t + 1 \right) dt
\]

\[
\cdot f(||P + tN_p - \theta||) \sum_{i=1}^{p-1} \left( K_i(P) t + 1 \right)
\]

\[
= -2 \int_0^{\infty} \frac{\partial}{\partial t} \left[ r(t^2, P)/(t + \overline{\rho}(P)) \sum_{i=1}^{p-1} \left( K_i(P) t + 1 \right) \right] dt
\]

\[
q(||P + tN_p - \theta||) f(||P + tN_p - \theta||) dt
\]

\[
- 2r(0, P)/\overline{\rho}(P) \int_{||P - \theta||}^{\infty} uf(u) du.
\]

The fact that \( r \) is nondecreasing in its first argument implies

\[
\frac{\partial}{\partial t} \left[ r(t^2, P)/(t + \overline{\rho}(P)) \sum_{i=1}^{p-1} \left( K_i(P) t + 1 \right) \right]
\]

\[
\geq r(t^2, P) [-1 + \{t + \overline{\rho}(P) \cdot \sum_{i=1}^{p-1} (K_i(P)/(K_i(P) t + 1))]}
\]

\[
\cdot \sum_{i=1}^{p-1} (K_i(P) t + 1)/(t + \overline{\rho}(P))^2.
\]
By the lemma in the Appendix

\[ \{t + \tilde{\rho}(P)\} \sum_{i=1}^{p-1} (K_i(P)/(K_i(P)t + 1)) \geq (p - N(P) - 1). \]

Combining this inequality with the last inequality we have that

\[ \frac{\partial}{\partial t} [r(t^2,P)/(t + \tilde{\rho}(P)) \prod_{i=1}^{p-1} (K_i(P)t + 1)] \]

\[ \geq r(t^2,P)(p - N(P) - 2) \prod_{i=1}^{p-1} (K_i(P)t + 1)/(t + \tilde{\rho}(P))^2. \]

Using this in (\(\star\)), it is clear that

\[ - \int_{0}^{\infty} 2r(t^2,P)(P + tN_p - \theta)^{tN_p}/(t + \tilde{\rho}(P)) \cdot f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1) dt \]

\[ \leq -2 \int_{0}^{\infty} \frac{(p - N(P) - 2)r(t^2,P)/(t + \tilde{\rho}(P))^2}{(P + tN_p - \theta)^{tN_p}} \cdot q(||P + tN_p - \theta||) f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1) dt \]

since \( r(0,p)/\tilde{\rho}(P) \int_{||P-\theta||}^{\infty} uf(u)du \) is nonnegative. Thus
\[ \Delta \leq \int \int_{P \in M \ t \geq 0} \frac{r^2(t^2, P)}{(t + \tilde{\rho}(P))^2} \]

\[ -2r(t^2, P) \frac{(p - N(P) - 2)}{(t + \tilde{\rho}(P))^2} q(||P + tN_p - \theta||) \]

\[ \times f(||P + tN_p - \theta||)^{p-1} \prod_{i=1}^{p-1} (K_i(P)t + 1) \, dt \, dA(P). \]

\[ = \int \int_{P \in M \ t \geq 0} [r(t^2, P) - 2(p - N(P) - 2) q(||P + tN_p - \theta||)] \]

\[ \times \frac{r(t^2, P)}{(t + \tilde{\rho}(P))^2} f(||P + tN_p - \theta||)^{p-1} \prod_{i=1}^{p-1} (K_i(P)t + 1) \, dt \, dA(P). \]

Because assumption (b) of the theorem implies that

\[ [r(t^2, P) - 2(p - N(P) - 2) q(||P + tN_p - \theta||)] \leq 0, \]

we have

\[ \Delta \leq 0. \]

q.e.d.

**Remark:** Consider the situation where \( \{r(t, P)/t\} \leq 1 \) for \( t \geq 0 \). For values of \( X \) not in \( G \), if \( \delta(X) \neq X \), then \( \delta(X) \) lies on a line between
$X$ and $P(X)$. Thus $\delta(X)$ is closer to $\mathcal{G}$ than $X$, i.e., $\delta(X)$ shrinks $X$ towards $\mathcal{G}$. For these values of $X$, if $\theta$ is anywhere in $\mathcal{G}$, then the actual loss (rather than the expected loss or risk) of $\delta(X)$ is less than that of $X$, i.e.,

$$||\delta(X) - \theta||^2 < ||X - \theta||^2.$$

**Theorem 2.** Let $X$ be a spherically symmetric random vector about $\theta$ which is $p$-dimensional and assume that $f(||X-\theta||^2)$ is the density of $X$. Let $c$ be a nondecreasing nonnegative concave function and let the loss for estimation of $\theta$ be

$$L(\theta, \hat{\theta}) = c(||\hat{\theta} - \theta||^2).$$

Assume that $E[||X-\theta||^2c'(||X-\theta||^2)] < \infty$ and $E[||X-\theta||^{-2}c'(||X-\theta||^2)] < \infty$.

For $p \geq 3$, the estimator $\delta$ given in Theorem 1 is minimax provided

$$r(t, P) \leq 2(p - N(P) - 2)\inf_{s \geq 0} Q(s)$$

where

$$Q(s) = \int_{S} \frac{uc'(u^2)f(u)du}{c'(s^2)f(s)}.$$
Remark: $Q(t) \leq q(t)$ and so

$$\inf_{t>0} \{Q(t)\} \leq \inf_{t>0} \{q(t)\}.$$ 

Proof of Remark

$$Q(t) = \int_t^\infty u c'(u^2)f(u)\,du/(c'(t^2)f(t)).$$

Because $c$ is concave, $c'$ is nonincreasing, and

$$Q(t) \leq \int_t^\infty u[c'(t^2)]f(u)\,du/\{c'(t^2)f(t)\} = q(t).$$

q.e.d.

Proof of Theorem 2:

$$\Delta_\theta(X) = ||X-\theta||^2 - ||\delta(X) - \theta||^2.$$ 

The difference in risks for $X$ and $\delta$ under the concave loss $c(\|\delta(X) - \theta\|^2)$ is

$$(**E[c(||X-\theta||^2)] - E[c(||\delta(X)-\theta||^2)])$$

$$= E[c(||X-\theta||^2)] - E[c(||X-\theta||^2 - \Delta_\theta(X))].$$
Because \( c \) is a nondecreasing concave function, for any values \( u \) and \( v \),

\[
c(u) < c(v) + c'(u)(u-v).
\]

Thus

\[
c(||X-\theta||^2 - \Delta_\theta(X)) < c(||X-\theta||^2) + c'(||X-\theta||^2)(-\Delta_\theta(X))
\]

Therefore,

\[
(**) \geq E_\theta[c'(||X-\theta||^2)\Delta_\theta(X)].
\]

Let \( Y \) be a spherically symmetric random vector about \( \theta \) with density

\[
Kc'(||Y-\theta||^2)f(||Y-\theta||^2).
\]

Then

\[
E_\theta[c'(||X-\theta||^2)\Delta_\theta(X)] = K^{-1}E_\theta[\Delta_\theta(Y)].
\]

According to Theorem 1, \( E_\theta[\Delta_\theta(Y)] > 0 \). Thus \((**) \geq 0\).

\( q.e.d. \)
Remark: The argument of the above proof is like that of the proof of a theorem of Brandwein and Strawderman [1980].

Example: Let $X$ have the density $K|X-\theta|^m \exp(-|X-\theta|^2/2)$. Let $c(s^2) = s$. Then $c'(s^2) = (2s)^{-1}$ and $\left\{ \inf_{s>0} Q(s) \right\}$ is one for $m \geq 1$. 

Appendix

Lemma. Let \( t \) and \( K_i, i=1,\ldots,p-1, \) be nonnegative numbers. Let \( N \) be the number of \( K_i \) values equal to zero. Then

\[
\sum_{i=1}^{p-1} \left( \frac{K_i}{K_i t + 1} \right) (t + \sum_{1 \leq j \leq p-1 \atop K_j > 0} \frac{K_j^{-1}}{(p - N - 1)}) \geq p - N - 1.
\]

Proof:

Set \( \tilde{\rho} = \sum_{1 \leq j \leq p-1 \atop K_j > 0} \frac{K_j^{-1}}{(p - N - 1)}. \) Then

\[
W = \sum_{i=1}^{p-1} \left( \frac{K_i}{K_i t + 1} \right) (t + \sum_{1 \leq j \leq p-1 \atop K_j > 0} \frac{K_j^{-1}}{(p - N - 1)})
\]

\[
= \sum_{1 \leq i \leq p-1 \atop K_i > 0} \frac{1}{K_i} (t + K_i^{-1}) \left( [t + K_i^{-1}] + [\tilde{\rho} - K_i^{-1}] \right)
\]

\[
= \sum_{1 \leq i \leq p-1 \atop K_i > 0} \frac{[t + K_i^{-1}]/(t + K_i^{-1}) + [\tilde{\rho} - K_i^{-1}]/(t + K_i^{-1})}{K_i}
\]

\[
= p - N - 1 + \sum_{1 \leq i \leq p-1 \atop K_i > 0} \frac{[\tilde{\rho} - K_i^{-1}]/(t + K_i^{-1})}{K_i}
\]
Observe that if $\bar{\rho} \leq K_i^{-1}$, then

$$t + K_i^{-1} \geq t + \bar{\rho}$$

implies

$$\frac{1}{(t+\bar{\rho})} \geq \frac{1}{(t+K_i^{-1})}.$$ 

This implies

$$\frac{(\bar{\rho}-K_i^{-1})}{(t+\bar{\rho})} \leq \frac{(\bar{\rho}-K_i^{-1})}{(t+K_i^{-1})}$$

since $(\bar{\rho}-K_i^{-1}) \leq 0$. Also, if $\bar{\rho} > K_i^{-1}$, then $t + K_i^{-1} \leq t + \bar{\rho}$ implies

$$\frac{1}{(t+\bar{\rho})} \leq \frac{1}{(t+K_i^{-1})},$$

which implies

$$\frac{(\bar{\rho}-K_i^{-1})}{(t+\bar{\rho})} \leq \frac{(\bar{\rho}-K_i^{-1})}{(t+K_i^{-1})}.$$ 

Thus

$$\sum_{1 \leq j \leq p-1} \frac{[\bar{\rho} - K_i^{-1}]/(t + K_i^{-1})}{K_i > 0} \geq \sum_{1 \leq j \leq p-1} \frac{[\bar{\rho} - K_i^{-1}]/(t + \bar{\rho})}{K_i > 0}$$

= 0
from the definition of $\rho$. Thus $W$ is greater than or equal to $p-N-1$.

\[ \text{q.e.d.} \]

**Theorem.** Let $D$ be a $p$-dimensional convex region in $\mathbb{R}^D$ with twice-differentiable boundary hypersurface $M=\partial D$. For the point $Q$ in $M$, define $N_Q$ to be the outward unit normal to $M$ at $Q^Q$ and let $M$ be oriented with this normal. Denote by $K_i(Q)$, $i=1,\ldots,p-1$, the principal curvatures of $M$ at $Q$ and by $dA(Q)$ the element of surface area on $M$. For $X=(X_1,\ldots,X_{p-1})$ in $D^C$ define $P(X)$ to be the nearest point of $M$ to $X$, i.e.

\[ \|X-P(X)\| = \inf_{Q \text{ in } M} \|X-Q\|. \]

Reparameterize a neighborhood $W$ of $D^C$ by the map

\[ X = P + tN_p \]

where $P=P(X)$ and $t=\|X-P\|$. Then the volume element on $D^C$ is given by

\[ dV = \prod_{i=1}^{p-1} (K_i(P)t + 1)dA(P)dt. \]
Proof. Fix $x^0$ in $D^C$ and let $P_0$ in $M$ be the nearest point of $M$ to $x^0$ so that

$$x^0 = P_0 + t_0 N_{P_0}.$$ 

Let $(u_1, \ldots, u_{p-1})$ be a coordinate system for points $Q$ in $M$ which are in a neighborhood of $P_0$ such that

$$\left( \frac{\partial Q}{\partial u_i} \right) \left( \frac{\partial Q}{\partial u_j} \right)_{Q=P_0} = \delta_{ij}$$

where $\delta_{ij}$ is zero if $i \neq j$ and one otherwise. Then $(t, u_1, \ldots, u_{p-1})$ forms a coordinate system in a neighborhood of $x^0$ which is orthonormal at $x^0$. Note that the neighborhood can be enlarged to include $P_0$. It suffices to prove the theorem for the chosen coordinate system at $P_0$ because the formula for $dV$ is independent of the choice of $u_1, \ldots, u_{p-1}$.

The change of variables formula implies

$$dV = |\det X'| \, du_1 \ldots du_{p-1} \, dt$$
where $X'$ is the Jacobean matrix of $X$, i.e.

\[
X' = \begin{bmatrix}
\frac{\partial x_1}{\partial t} & \cdots & \frac{\partial x_p}{\partial t} \\
\frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_p}{\partial u_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial x_1}{\partial u_{p-1}} & \cdots & \frac{\partial x_p}{\partial u_{p-1}}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\frac{\partial P}{\partial u_1} t & \cdots & \frac{\partial P}{\partial u_p} t \\
\frac{\partial P}{\partial u_{p-1}} t + t \frac{\partial N}{\partial u_{p-1}} t
\end{bmatrix}
\]
The rules for expanding multilinear expressions imply that \( \det X' \) is a polynomial in \( t \) of degree at most \( p-1 \) and will be completely determined when we find its roots and its value at \( t=0 \). Following a similar derivation in Milnor [1969], p. 34, one may write the product of \( X' \) and the matrix

\[
Z = \left( N_{p} \frac{\partial}{\partial u_{1}}, \ldots, \frac{\partial}{\partial u_{p-1}} \right)
\]

as

\[
X'Z = \\
\begin{bmatrix}
1 & 0 \\
\frac{\partial N_{p}}{\partial u_{1}} & t N_{p} \\
t(\frac{\partial}{\partial u_{p-1}}) & [t(\frac{\partial}{\partial u_{1}})^{t} \frac{\partial}{\partial u_{1}} + t(\frac{\partial}{\partial u_{1}})^{t} \frac{\partial}{\partial u_{j}}]
\end{bmatrix}
\]

For \( p=p_{0} \), the vectors in \( Z \) are orthonormal and \( \det X'|_{p=p_{0}} \) equals
\[ \det X'Z_{|p=p_0}. \] But \( \det X'Z_{|p=p_0} \) equals the determinant of the lower right block of \( X'Z \) evaluated at \( p=p_0 \):

\[ [(\frac{\partial p}{\partial u_i})^t \frac{\partial p}{\partial u_j} + t(\frac{\partial p}{\partial u_i} \frac{\partial p}{\partial u_j})]_{p=p_0} = \left[ s_{ij} + t(\frac{\partial p}{\partial u_i} \frac{\partial p}{\partial u_j}) \right]_{p=p_0}. \]

The identity

\[ 0 = \frac{\partial}{\partial u_i} \left( N^t \frac{\partial p}{\partial u_j} \right) = \left( \frac{\partial N}{\partial u_i} \right)^t \frac{\partial p}{\partial u_j} + N^t \frac{\partial}{\partial u_i} \left( \frac{\partial p}{\partial u_j} \right) \]

implies that the lower right block of \( X'Z \) is

\[ \left[ \delta_{ij} - tN^t \frac{\partial}{\partial u_i} \left( \frac{\partial p}{\partial u_j} \right) \right]_{p=p_0} \]

which is singular when \( t^{-1} \) is an eigenvalue of \( [N^t \frac{\partial}{\partial u_i} \left( \frac{\partial p}{\partial u_j} \right)]_{p=p_0} \).

Note that the eigenvalues of \( [N^t \frac{\partial}{\partial u_i} \left( \frac{\partial p}{\partial u_j} \right)]_{p=p_0} \) are the negatives of the principal curvatures of \( M \) evaluated at \( p_0 \) by definition.

The multiplicity of \( t^{-1} \) as an eigenvalue equals the multiplicity of the corresponding root. So

\[ \det X'|_{p=p_0} = c(P_0)^{p-1} \prod_{i=1}^{p-1} (1+tK_i(P_0)). \]

Thus \( dV = c(P_0)^{p-1} \sum_{i=1}^{p-1} (tK_i(P_0)+1)du_1 \cdots du_{p-1}dt. \) Since this formula is valid on a neighborhood of \( M \) we may set \( t=0 \) and restrict to \( M \) to obtain
\[ dV|_M = dA(P_0) = c(P_0)du_1 \ldots du_{p-1} \]

Thus

\[ dV = \prod_{i=1}^{p-1} (K_i(P)t + 1)dA(P)dt. \]

q.e.d.
References


