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 θ 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 minimax and dominate the estimator $\hat{\theta}_0(X)=X$. They employ "vague" prior information in the following sense. The "vague" prior information is that θ is "likely" to lie in a certain convex region G. The estimate of θ 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).)

<u>Estimators</u>

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 $\bar{\rho}(P)$ 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 θ 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:

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

where

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

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

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

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

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

$$q(s) = \int_s^\infty uf(u)du/f(s).$$

Note: If X is in G or $\overline{G}$, 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 $f(s) = Cs^m \exp(-s^2/2)$ and shows that $\{\inf_{s\geq 0} q(s)\}$ is one for $m \geq 0$.

Proof of Theorem 1. Because the estimator X is minimax with constant risk for all values of θ, it suffices to show that Δ is nonpositive for all values of θ where the difference in risks is

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

Using the definition of δ, we may write

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

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

since $P = P(X) = X$ for X in $\bar{G}$. Thus

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

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

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

where dV is the volume element in $\bar{G}^c$. Let N_p denote the outward unit normal vector to M, the boundary hypersurface of $\bar{G}$, at the point P on M. Letting M be oriented with this normal, denote by $K_i(P)$, $i=1,\ldots,p-1$, the principal curvatures of M at P. Let dA be the element of surface area on M. With $P=P(X)$, reparameterize X in $\bar{G}^c$ 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,

$$\bar{\rho}(P) = \sum_{\substack{1 \le i \le p-1 \\ K_i(P) > 0}} (K_i(P))^{-1}/(p - N(P) - 1).$$

By the theorem of the Appendix, the volume element on $\bar{G}^c$ is

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

Thus

$$\Delta = \int_{P \text{ in } M} \int_{t \ge 0} [r^2(t^2,P)/\{t + \bar{\rho}(P)\}^2$$

$$- 2r(t^2,P)(P + tN_p - \theta)^t N_p/\{t + \bar{\rho}(P)\}]$$

$$f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1)dtdA(P).$$

Define

$$q(s) = \begin{cases} \int_s uf(u)du/f(s) & \text{for } f(s) > 0 \\ 0 & \text{for } f(s) = 0 \end{cases} .$$

Because

$$\frac{\partial}{\partial t}\{\int_{||P+tN_p-\theta||}^{\infty} uf(u)du\} = -f(||P + tN_p - \theta||).$$

$$\cdot\ (P + tN_p - \theta)^t N_p,$$

integration by parts implies that

$$(*) \qquad -\int_0^{\infty} 2r(t^2,P)(P + tN_p - \theta)^t N_p/\{t + \bar{\rho}(P)\}$$

$$\cdot\ f(||P + tN_p - \theta||)\prod_{i=1}^{p-1} (K_i(P)t + 1)dt$$

$$= -2\int_0^{\infty} \frac{\partial}{\partial t}[r(t^2,P)/\{t + \bar{\rho}(P)\}\prod_{i=1}^{p-1} (K_i(P)t + 1)]$$

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

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

The fact that r is nondecreasing in its first argument implies

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

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

$$\cdot \prod_{i=1}^{p-1} (K_i(P)t + 1)/\{t + \bar{\rho}(P)\}^2.$$

By the lemma in the Appendix

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

Combining this inequality with the last inequality we have that

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

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

Using this in (*), it is clear that

$$- \int_0^\infty 2r(t^2,P)(P + tN_p - \theta)^t N_p/\{t + \bar{\rho}(P)\}$$

$$\cdot f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1)dt$$

$$\leq -2 \int_0^\infty (p - N(P) - 2)r(t^2,P)/\{t + \bar{\rho}(P)\}^2$$

$$\cdot q(||P + tN_p - \theta||)f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1)dt$$

since $r(0,p)/\bar{\rho}(P) \int_{||P-\theta||}^{\infty} uf(u)du$ is nonnegative. Thus

$$\Delta \leq \int_{P \text{ in } M} \int_{t \geq 0} [r^2(t^2,P)/\{t + \bar{\rho}(P)\}^2$$

$$-2r(t^2,P)(p-N(P)-2)/\{t + \bar{\rho}(P)\}^2 q(||P + tN_p - \theta||)]$$

$$\cdot f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1)dtdA(P).$$

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

$$\cdot r(t^2,P)/\{t + \bar{\rho}(P)\}^2 f(||P + tN_p - \theta||) \prod_{i=1}^{p-1} (K_i(P)t + 1)dtdA(P).$$

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 $\bar{G}$, if $\delta(X) \neq X$, then $\delta(X)$ lies on a line between

X and P(X). Thus $\delta(X)$ is closer to $\bar{G}$ than X, i.e., $\delta(X)$ shrinks X towards $\bar{G}$. For these values of X, if θ is anywhere in $\bar{G}$, then the actual loss (rather than the expected loss or risk) of $\delta(X)$ is less than that of X, i.e.,

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

<u>Theorem 2.</u> Let X be a spherically symmetric random vector about θ 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 θ be

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

Assume that $E[||X-\theta||^2 c'(||X-\theta||^2)] < \infty$ and $E[||X-\theta||^{-2} c'(||X-\theta||^2)] < \infty$. For $p \geq 3$, the estimator δ given in Theorem 1 is minimax provided

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

where

$$Q(s) = \int_s^\infty uc'(u^2)f(u)du/\{c'(s^2)f(s)\}.$$

Remark: $Q(t) \leq q(t)$ and so

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

Proof of Remark

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

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 δ 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 θ 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, $E_\theta[\Delta_\theta(Y)] \geq 0$. 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 $\{\inf_{s\geq 0} Q(s)\}$ is one for $m \geq 1$.

Appendix

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

$$\sum_{i=1}^{p-1} (K_i/\{K_i t + 1\})(t + \sum_{\substack{1\le j\le p-1 \\ k_j>0}} K_j^{-1}/(p - N - 1)) \ge p - N - 1.$$

Proof:

Set $\bar{\rho} = \sum_{\substack{1\le j\le p-1 \\ K_j>0}} K_j^{-1}/(p - N - 1)$. Then

$$W = \sum_{i=1}^{p-1} (K_i/\{K_i t + 1\})(t + \sum_{\substack{1\le j\le p-1 \\ K_j>0}} K_j^{-1}/(p - N - 1))$$

$$= \sum_{\substack{1\le i\le p-1 \\ K_i>0}} (1/\{t + K_i^{-1}\})([t + K_i^{-1}] + [\bar{\rho} - K_i^{-1}])$$

$$= \sum_{\substack{1\le i\le p-1 \\ K_i>0}} [t + K_i^{-1}]/\{t + K_i^{-1}\} + \sum_{\substack{1\le i\le p-1 \\ K_i>0}} [\bar{\rho} - K_i^{-1}]/\{t + K_i^{-1}\}$$

$$= p - N - 1 + \sum_{\substack{1\le i\le p-1 \\ K_i>0}} [\bar{\rho} - K_i^{-1}]/\{t + K_i^{-1}\}$$

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

$$\sum_{\left\{\substack{1 \leq j \leq p-1 \\ K_i > 0}\right\}} [\bar{\rho} - K_i^{-1}]/\{t + K_i^{-1}\}$$

$$\geq \sum_{\left\{\substack{1 \leq i \leq p-1 \\ K_i > 0}\right\}} [\bar{\rho} - K_i^{-1}]/(t + \bar{\rho})$$

$$= 0$$

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

q.e.d.

<u>Theorem.</u> Let D be a p-dimensional convex region in $\mathbb{R}^p$ 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,...,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 $\bar{D}^c$ 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 $\bar{D}^c$ by the map

$$X = P + tN_p$$

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

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

Proof. Fix X^0 in $\bar{D}^c$ 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

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

where δ_{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.

$$X' = \begin{bmatrix} \frac{\partial X_1}{\partial t} \cdots \frac{\partial X_p}{\partial t} \\ \frac{\partial X_1}{\partial u_1} \cdots \frac{\partial X_p}{\partial u_1} \\ \vdots \qquad \vdots \\ \frac{\partial X_1}{\partial u_{p-1}} \cdots \frac{\partial X_p}{\partial u_{p-1}} \end{bmatrix}$$

$$= \begin{bmatrix} N_p^t \\ (\frac{\partial P}{\partial u_1})^t + t(\frac{\partial N_p}{\partial u_1})^t \\ \vdots \\ (\frac{\partial P}{\partial u_{p-1}})^t + t(\frac{\partial N_p}{\partial u_{p-1}})^t \end{bmatrix} .$$

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

$$Z = (N_p, \frac{\partial P}{\partial u_1}, \ldots, \frac{\partial P}{\partial u_{p-1}})$$

as

$$X'Z = \begin{bmatrix} 1 & \bar{0} \\ t(\frac{\partial N_p}{\partial u_{p-1}})^t N_p & \vdots \\ & [(\frac{\partial P}{\partial u_1})^t \frac{\partial P}{\partial u_j} + t(\frac{\partial N_p}{\partial u_i})^t \frac{\partial P}{\partial u_j}] \\ t(\frac{\partial N_p}{\partial u_{p-1}})^t N_p & \end{bmatrix}$$

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 $P=P_0$

$$\left[\left(\frac{\partial P}{\partial u_i}\right)^t \frac{\partial P}{\partial u_j} + t\left(\frac{\partial N_p}{\partial u_i}\right)^t \frac{\partial P}{\partial u_j}\right]_{P=P_0} = \left[\delta_{ij} + t\left(\frac{\partial N_p}{\partial u_i}\right)^t \frac{\partial P}{\partial u_j}\right]_{P=P_0}.$$

The identity

$$0 = \frac{\partial}{\partial u_i}\left(N_p^t \frac{\partial P}{\partial u_j}\right) = \left(\frac{\partial N_p}{\partial u_i}\right)^t \frac{\partial P}{\partial u_j} + N_p^t \frac{\partial}{\partial u_i}\left(\frac{\partial P}{\partial u_j}\right)$$

implies that the lower right block of $X'Z$ is

$$\left[\delta_{ij} - tN_p^t \frac{\partial}{\partial u_i}\left(\frac{\partial P}{\partial u_j}\right)\right]_{P=P_0}$$

which is singular when t^{-1} is an eigenvalue of $\left[N_p^t \frac{\partial}{\partial u_i}\left(\frac{\partial P}{\partial u_j}\right)\right]_{P=P_0}$. Note that the eigenvalues of $\left[N_p^t \frac{\partial}{\partial u_i}\left(\frac{\partial P}{\partial u_j}\right)\right]_{P=P_0}$ 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

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

Thus $dV = c(P_0)\prod_{i=1}^{p-1}(tK_i(P_0)+1)\,du_1\ldots du_{p-1}\,dt$. 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

Berger, James (1975) "Minimax estimation of location vectors for a wide class of densities". The Annals of Statistics, Vol. 3, 1318-1328.

Bock, M.E. (1982) "Employing vague inequality information in the estimation of normal mean vectors (estimators that shrink to closed convex polyhedra)". Statistical Decision theory and Related Topics III, Vol. 1, 169-193, Shanti Gupta and J.O. Berger, Editors, Academic Press, New York.

Bock, M.E. (1981) "Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions." Purdue University Statistics Department Technical Report #81-51.

Brandwein, Ann Cohen and Strawderman, William E. (1980) "Minimax estimation of location parameters for spherically symmetric distributions with concave loss". The Annals of Statistics, Vol. 8, 279-284.

Milnor, J. (1969) Morse Theory, Annals of Mathematical Studies, No. 51, Princeton University Press, Princeton, New Jersey.