Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

M.L. Eaton

Concentration inequalities for Gauss Markov estimators

Department of Mathematical Statistics Report MS-R8706 April

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Concentration Inequalities for Gauss Markov Estimators

Morris L. Eaton
University of Minnesota,
School of Statistics

Let M be the regression subspace and Υ the set of possible covariances for a random vector Y. The linear model determined by M and Υ is <u>regular</u> if the identity is in Υ and if $\Sigma(M) \subseteq M$ for all $\Sigma \varepsilon \Upsilon$. For such models, concentration inequalities are given for the Gauss Markov estimator of the mean vector under various distributional and invariance assumptions on the error vector. Also, invariance is used to establish monotonicity results relative to a natural group induced partial ordering.

1980 Mathematics Subject Classification: Primary: 62H12, Secondary: 62J05, 62F10.
Key Words & phrases: Gauss Markov estimators, concentration inequalities, elliptical densities, log concave densities, Majorization, group induced orderings, reflection groups.
Note: This work was supported in part by National Science Foundation Grant No. DMS83-19924. Portions of this work were completed while the author was on sabbatical leave visiting the Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

Report MS-R8706
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Section 1. Introduction and Summary

In this paper we give conditions under which the distribution of the Gauss-Markov estimator is more concentrated about the unknown mean vector than the distribution of any other linear unbiased estimator. In order to describe our results more precisely, let $(V,(\cdot,\cdot))$ be a finite dimensional inner product space. By a linear model for a random vector Y in V, we mean the specification of

(i) a known linear subspace $M \subseteq V$ in which the mean vector of Y is assumed to lie.

(ii) a known set Υ of positive definite linear transformations in which the covariance of Y is assumed to lie.

Thus, the subspace M specifies the mean structure of Y in that $\mu = \mathcal{E}Y$ is in M. Similarly, Υ specifies the covariance structure of Y so $\text{Cov}(Y)\varepsilon\Upsilon$. The use of $\text{Cov}(Y)$ to denote the covariance of Y in $(V,(\cdot,\cdot))$ is consistent with Eaton (1983, Chapter 2). Throughout this paper it is assumed that the identity covariance is an element of Υ.

Definition 1: The linear model (M,Υ) for Y is regular if $\Sigma(M)\subseteq M$ for all $\Sigma\varepsilon\Upsilon$.

Under the assumption that $\text{Cov}(Y)$ is non-singular, regularity of the linear model for Y is the necessary and sufficient condition so that a best linear unbiased estimator of $\mu\varepsilon M$ exists (see Eaton (1983, Chapter 4) for a discussion).

To state one form of the Gauss-Markov Theorem, let $\underset{\sim}{A}$ be the class of linear transformations A on V to V which satisfy

(i) $Ax = x$ for $x\varepsilon M$

$$\text{(ii) } Ay \in M \text{ for } y \in V \qquad (1.1)$$

The elements of $\underset{\sim}{A}$ are the linear transformations which provide the linear unbiased estimates of μ subject to the condition that the estimator take values in M.

Theorem (Gauss-Markov).

Assume the linear model (M,Υ) is regular. Let $A_0 \varepsilon \underset{\sim}{A}$ be the orthogonal projection onto the subspace M, and let $\Sigma = \text{Cov}(Y) \varepsilon \Upsilon$. Then for all $A \varepsilon \underset{\sim}{A}$ and all $\Sigma \varepsilon \Upsilon$,

$$\text{Cov}(AY) = A\Sigma A' \geq A_0 \Sigma A_0' = \text{Cov}(A_0 Y) \qquad (1.2)$$

where $\geq$ means that

$$A\Sigma A' - A_0 \Sigma A_0'$$

is positive semi-definite.

The intuitive content of (1.2) is that $\text{Cov}(\cdot)$ is a multivariate measure of size, and for all $\Sigma \varepsilon \Upsilon$, the element of $\underset{\sim}{A}$ which minimizes $\text{Cov}(AY)$ is A_0. A possible alterative criterion for the selection of $A \varepsilon \underset{\sim}{A}$ is to ask that the distribution of the estimator AY be "most concentrated" about μ. One way to make this precise is to look at how concentrated the distribution of $AY-\mu$ is about $0 \varepsilon M$ - that is, look at

$$\psi(A) = P\{AY-\mu \varepsilon C\} \tag{1.3}$$

where C is a symmetric (about 0) convex set in M.

Of course, we would like to pick $A \varepsilon \underset{\sim}{A}$ so that $\psi(A)$ is as large as possible no matter what convex symmetric set C happens to be. Because $Ax=x$ for $x \varepsilon M$, (1.3) can be written

$$\psi(A) = P\{A(Y-\mu)\varepsilon C\} \tag{1.4}.$$

Essentially, the results in this paper give conditions, expressed in terms of the distribution of the error vector $Y-\mu$, so that A_0 maximizes ψ in (1.4) for all convex symmetric sets C contained in M.

Here is an outline of the paper. Section 2 contains background material on peakedness of distributions, log concave distributions, and elliptical distributions. Also, Anderson's Theorem (Anderson (1955)) and a result from Das Gupta et al. (1972) are reviewed.

Our main results are given in Section 3. For example, it is shown that if the distribution of the error vector $Z = Y-\mu$ is elliptical (as defined in Section 2), then, with ψ given by (1.4),

$$\psi(A_0) \geq \psi(A), \; A \varepsilon \underset{\sim}{A} \tag{1.5}$$

for all convex symmetric subsets $C \subseteq M$. With additional assumptions on the distribution of Z, the above result is extended to the case where the convex sets are allowed to depend on the data Y. This result is applicable to confidence set problems. The results in this section are generalizations of results in Berk and Hwang (1984) who established inequality (1.5) for the classical univariate regression model. In addition to allowing a wider class of error distributions, our results are applicable to all regular linear models which include the MANOVA model as well as certain structured covariance linear models.

Utilizing some invariance assumptions, the results in Section 4 establish monotonicity of the function ψ in (1.4). This monotonicity is expressed in terms of a partial ordering on $\underset{\sim}{A}$ which is induced by a group of transformations. These ideas lead to a strengthening of a majorization result due to Proschan (1965).

Section 2. Concentration and Probability Inequalities

The notion of peakedness (concentration) of a distribution on the real line was introduced in Birnbaum (1948). Sherman (1955) extended the notion to euclidean spaces. The vector space version of concentration runs as follows.

For a finite dimensional real vector space W, let $\zeta(W)$ be all the nonempty convex subsets of W which are symmetric about 0- that is, subsets C   W which are convex and satisfy C = -C.

<u>Definition 2.1</u>: Given two random vectors Y_1 and Y_2 in W, Y_1 is <u>more concentrated about 0</u> than Y_2 if

$$P\{Y_1 \varepsilon C\} \geq P\{Y_2 \varepsilon C\} \qquad (2.1)$$

for all $C \in \zeta(W)$.

In what follows, when Y_1 is more concentrated about 0 than Y_2, we will simply say Y_1 is more concentrated than Y_2.

Now, consider the vector space W with a given inner product $(\cdot,\cdot)$. In what follows, the word density means a probability density with respect to Lebesque measure on W.

<u>Definition 2.2</u>: A random vector X in $(W,(\cdot,\cdot))$ has an <u>elliptical distribution</u> if X has a density f of the form

$$f(w) = |B|^{-1/2} k[(w,B^{-1}w)] \tag{2.2}$$

where B is some positive definite transformation on W to W and k is a non-negative function defined on $[0,\infty)$ which satisfies

$$\int_W k[(w,w)]dw = 1 \tag{2.3}$$

Here is a theorem due to Das Gupta et al. (1972) which is needed in the next section. In what follows, $\underset{\sim}{L}(\cdot)$ denotes the probability law of "$\cdot$".

<u>Theorem 2.1</u>: Fix the function k in (2.2) and let P_B denote the probability measure defined on $(W,(\cdot,\cdot))$ by the density in (2.2). For random vectors

X_i, i = 1,2, assume that $\underset{\sim}{L}(X_i) = P_{B_i}$ where $B_2 - B_1$ is non-negative definite. Then X_1 is more concentrated than X_2.

Corollary 2.1: Let X in W have the density (2.2) and suppose α_i, i = 1,2 are full rank linear transformations on W to $(\underset{\sim}{U}, (\cdot,\cdot)_1)$. Set $X_i = \alpha_i X$, i = 1,2 and assume that $\alpha_2 B \alpha_2' - \alpha_1 B \alpha_1'$ is non-negative definite where B is given in (2.2). Then X_1 is more concentrated than X_2.

Proof: Because α_i has full rank, an easy argument shows that X_i has a density on $\underset{\sim}{U}$ of the form

$$f_i(u) = |B_i|^{-1/2} k_0[(u, B_i u)_1]$$

where

$$B_i = \alpha_i B \alpha_i'$$

Since $B_2 - B_1$ is assumed to be non-negative definite, Theorem 2.1 gives the result. □

The final topic of this section concerns log concave functions and Anderson's Theorem on $(W, (\cdot,\cdot))$.

Theorem (Anderson (1955)): Suppose f is a non-negative integrable function defined on W (integrable with respect to Lebesque measure). Also, suppose that for each $u > 0$,

$$\{w | f(w) \geq u\} \qquad (2.4)$$

is a convex symmetric subset of W. Then for each $C\varepsilon\zeta(W)$ and each $\theta\varepsilon W$, the function

$$\alpha \rightarrow \int I_C(w)\ f(W-\alpha\theta)dw \qquad (2.5)$$

is non-increasing for $\alpha\varepsilon[0,\infty)$.

Recall that a non-negative function f defined on W is log concave if for all $\alpha\varepsilon(0,1)$,

$$f(\alpha x + (1-\alpha)y) \geq f^{\alpha}(x)\ f^{1-\alpha}(y) \qquad (2.6)$$

for all x and $y\varepsilon W$. Observe that if f_1 defined on W satisfies

(i) $f_1(w) = f_1(-w)$, $w\varepsilon W$

(ii) f_1 is log concave on W

then (2.4) is a convex symmetric set, so Anderson's Theorem holds for such an f_1 when f_1 is integrable.

Now, suppose f is a log concave density function of a random vector X with values in W. Write $W = M \oplus N$ where M and N are perpendicular subspaces of W whose sum is W. Thus, X can be written uniquely as $X = Y+Z$ with $Y\varepsilon M$ and $Z\varepsilon N$. The marginal density of Z on the vector space N is

$$f_2(z) = \int_M f(y+z)dy$$

where dy means Lebesque measure on M. Thus, one version of the conditional density of Y given Z is

$$f_1(y|z) \doteq \begin{cases} \dfrac{f(y+z)}{f_2(z)} & \text{if } f_2(z) > 0 \\ \phi(y) & \text{if } f_2(z) = 0 \end{cases}$$

where $\phi(y)$ is the density of a standard normal distribution on M. Because f is log concave, a routine verification shows that for each fixed z, $f_1(\cdot|z)$ is log concave on the vector space M. This observation is used in the next section.

Section 3. Concentration of the Gauss-Markov Estimator

This section contains three results all of which deal with concentration of the Gauss-Markov estimator. Theorem 3.1 establishes inequality (1.5) for all regular linear models under the assumption that the error vector $Z = Y-\mu$ has an elliptical density. Using stronger assumptions, Theorem 3.1 is extended to cover some cases involving confidence statements about the unknown mean vector. The section closes with an example from the MANOVA model.

Throughout this section, it is assumed that (M,γ) is a regular linear model for a random vector Y taking values in the inner product space $(V,(\cdot,\cdot))$. As defined in Section 1, $\underset{\sim}{A}$ is the class of linear transformations defined on V which satisfy (1.1). Further, $A_0 \varepsilon \underset{\sim}{A}$ is the orthogonal projection onto M.

Theorem 3.1: Assume the error vector $Z = Y-\mu$ has an elliptical distribution on V. Then for each $C\varepsilon\zeta(M)$,

$$\psi(A) = P\{AY - \mu \varepsilon C\} \tag{3.1}$$

is maximized by taking $A = A_0$. That is, for each $C \varepsilon \zeta(M)$, the inequality

$$\psi(A) \leq \psi(A_0) \tag{3.2}$$

holds for all $A \varepsilon \underset{\sim}{A}$. Thus the distribution of $A_0 Y - \mu$ is more concentrated than the distribution of $AY - \mu$ for all $A \varepsilon \underset{\sim}{A}$.

Proof: Because $Ax = x$ for all $x \varepsilon M$, $AY - \mu = A(Y - \mu)$ so that

$$\psi(A) = P\{AZ \varepsilon C\} \tag{3.3}$$

Let $\Sigma = \text{Cov}(Y) \varepsilon \Upsilon$. Since $Z = Y - \mu$, it follows that

$$\text{Cov}(Z) = \text{Cov}(Y) = \Sigma \tag{3.4}$$

But, Z has an elliptical distribution with a density given by (2.2) for some positive definite B. It follows easily that

$$B = \beta \Sigma \tag{3.5}$$

for some real number $\beta > 0$.

Now, the regularity of the linear model and the Gauss-Markov Theorem imply that

$$A_0 \Sigma A_0' \leq A \Sigma A' \qquad (3.6)$$

for all $A \varepsilon \underset{\sim}{A}$. Thus (3.5) and (3.6) entail

$$A_0 B A_0' \leq A B A' \qquad (3.7)$$

for all $A \varepsilon \underset{\sim}{A}$. Each $A \varepsilon \underset{\sim}{A}$ is a linear transformation on V to M and each A is of full rank since each A is an onto linear transformation. The claimed result now follows immediately from (3.7) and Corollary 2.1 applied to $X_1 = A_0 Z$ and $X_2 = AZ$. □

It is possible to strengthen Theorem 3.1 by letting the symmetric convex set C in (3.1) depend on Y in certain ways, but this strengthening requires some modified assumptions on the distribution of Z. To specify how the set C is allowed to depend on Y, we have

Definition 3.1: For each $y \varepsilon V$, let $C(y) \varepsilon \zeta(M)$. Then, C(y) depends residually on y if

$$C(y) = C(y+x) \qquad y \varepsilon V,\ x \varepsilon M \qquad (3.8).$$

Theorem 3.2: Let C(Y) depend residually on Y and suppose the error vector $Z = Y - \mu$ has an elliptical density given by (2.2) where the function k is non-decreasing on $[0,\infty)$. Then for $A \varepsilon \underset{\sim}{A}$,

$$\psi_1(A) = P\{AY - \mu \ \varepsilon\ C(Y)\} \qquad (3.9)$$

is maximized at $A=A_0$.

Proof: Because $C(\cdot)$ depends residually on Y,

$$C(Y) = C(Y-\mu) = C(Z).$$

As in the proof of Theorem 3.1, $AY-\mu = AZ$ so (3.9) can be written

$$\psi_1(A) = P\{AZ\varepsilon C(Z)\} \qquad (3.10).$$

With $\bar{A}_0 = I-A_0$, the equation

$$A = A_0 + A\bar{A}_0 \qquad (3.11).$$

holds since $A\varepsilon \underset{\sim}{A}$. Also,

$$C(Z) = C(Z-A_0Z) = C(\bar{A}_0Z)$$

since A_0Z is in M. Hence, (3.10) can be written

$$\psi_1(A) = P\{A_0Z + A\bar{A}_0Z \ \varepsilon\ C(\bar{A}_0Z)\} =$$

$$\underset{\sim}{E}P\{A_0Z + A\bar{A}_0Z \ \varepsilon\ C(\bar{A}_0Z)|\bar{A}_0Z\} \qquad (3.12)$$

With $w = \bar{A}_0 Z$, the theorem will hold if we can verify the inequality

$$P\{A_0 Z + Aw \in C(w) | \bar{A}_0 Z = w\} \le$$
$$P\{A_0 Z \in C(w) | \bar{A}_0 Z = w\} \qquad (3.13)$$

for each w in the orthogonal complement of M.

To establish (3.13), argue as follows. As in the proof of Theorem 3.1, the linear transformation F in (2.2) occuring in the density of Z is some positive multiple of Cov(Z) = Σ, say

$$B = \beta \Sigma \qquad (3.14)$$

with β>0. Since M is invariant under Σ, M is also invariant under B. Thus, for any $x \in V$,

$$(x, B^{-1}x) = (A_0 x, B^{-1} A_0 x) + (\bar{A}_0 x, B^{-1} \bar{A}_0 x) \quad .$$

With $\underset{\sim}{U} = A_0 Z$ and $W = \bar{A}_0 Z$, the marginal density of W is

$$f_2(w) = \int_M |B|^{-1/2} k[(u, B^{-1}u) + (w, B^{-1}w)]du.$$

Thus a version of the conditional density of $\underset{\sim}{U}$ given W=w is

$$f_1(u|w) = \begin{cases} [f_2(w)]^{-1} |B|^{-1/2} k[(u, B^{-1}u) + (w, B^{-1}w)] & \text{if } f_2(w) > 0 \\ \phi(u) & \text{if } f_2(w) = 0 \end{cases}$$

where φ is the density of a standard normal distribution on M. For each w, it follows immediately that

(i) $f_1(u|w) = f_1(-u|w)$, $u \varepsilon M$

(ii) $\{u|f_1(u|w) \geq \alpha\}$ is convex.

Thus, for each w, Anderson's Theorem yields (3.13) so the proof is complete. □

The conclusion of Theorem 3.2 is also valid under log concavity and certain invariance assumptions on the density of Z. To state this result, let H be the group of two elements defined by

$$H = \{I, \bar{A}_0 - A_0\}.$$

Theorem 3.3: For each $\Sigma \varepsilon \Upsilon$, assume that the density f of the error vector Z satisfies

(i) f is log concave

(ii) $f(x) = f(hx)$ for $h \varepsilon H$, $x \varepsilon V$.

If $C(Y) \varepsilon \zeta(M)$ depends residually on Y, then ψ_1 defined in (3.9) is maximized at $A=A_0$.

Proof: The argument given in the first part of the proof of Theorem 3.2 shows that the verification of inequality (3.13) suffices to establish the present result. This verification involves the conditional density of $\underset{\sim}{U}=A_0Z$ given $W = \bar{A}_0 Z$. As argued at the end of Section 2, the marginal distribution of W is

$$f_2(w) = \int_M f(u+w)du,$$

for w in the orthogonal complement of M. Further, one version of the conditional density is

$$f_1(u|w) = \begin{cases} \dfrac{f(u+w)}{f_2(w)} & \text{if } f_2(w) > 0 \\ \phi(u) & \text{if } f(w) = 0 \end{cases}$$

where ϕ is the density of a standard normal distribution on M. That $f_1(\cdot|w)$ is log concave was noted earlier. We now claim that

$$f_1(u|w) = f_1(-u|w) \qquad ,u\varepsilon M \tag{3.15}$$

for each w. Obviously (3.15) holds if $f_2(w) = 0$ so assume $f_2(w) > 0$. Then

$$f_1(-u|w) = \frac{f(-u+w)}{f_2(w)} =$$

$$\frac{f((\bar{A}_0-A_0)(u+w))}{f_2(w)} = \frac{f(u+w)}{f_2(w)} = f_1(u|w)$$

Since f is invariant under the orthogonal transformation $\bar{A}_0-A_0$. The log concavity of $f_1(\cdot|w)$ together with (3.15) show that for each $\alpha>0$,

$$\{u|\ f_1(u|w) \geq \alpha\}$$

is a symmetric convex set. Anderson's Theorem shows that (3.13) holds so the proof is complete. □

Remark 3.1: Given $C(Y)\varepsilon\zeta(M)$ which depends residually on Y, again consider

$$\psi_1(A) = P\{AY-\mu \ \varepsilon \ C(Y)\} \tag{3.16}$$

for $A\varepsilon\underset{\sim}{A}$. As noted earlier, ψ_1 can be written

$$\psi_1(A) = P\{AZ \ \varepsilon \ C(Z)\} \tag{3.17},$$

where $Z = Y-\mu$ is the error vector. Let $\underset{\sim}{F}$ be the class of densities of Z for which $\psi_1(A) \leq \psi_1(A_0)$ no matter what choices are made for C(Y). Theorems 3.2 and 3.3 give examples of densities $f\varepsilon\underset{\sim}{F}$. But it is clear that $\underset{\sim}{F}$ is a convex set. This convexity can be used to extend Theorems 3.2 and 3.3 in an obvious way--namely by taking convex combinations and limits. In particular, suppose f is a density of Z which satisfies

(i) $\{x|f(x) \geq \alpha\}$ is convex for each $\alpha>0$ (3.18)

(ii) $f(x) = f(hx)$ for $h\varepsilon H$ where H is the group in Theorem 3.3.

For such an f, ψ_1 defined in (3.17) is maximized for $A=A_0$. To see this, observe that

$$f(x) = \int_0^\infty H(u,x)du$$

where

$$H(u,x) = \begin{cases} 1 & \text{if } f(x) \geq u \\ 0 & \text{otherwise.} \end{cases}$$

For $u \varepsilon (0,\infty)$ fixed such that $\int H(u,x)dx > 0$,

$$f_1(x|u) = \frac{H(u,x)}{\int_V H(u,x)dx}$$

is a log concave density on V to which Theorem 3.3 applies. Since f is an average (over u) of $f_1(\cdot|u)$, we see that ψ_1 is maximized at $A=A_0$ when the error vector Z has density f. □

Example 3.1 (MANOVA). For this example, the vector space V is the space of all real nxp matrices with the inner product given by the trace--that is, for two nxp matrices x and y, the inner product between x and y is

$$(x,y) = \text{tr}xy' \quad .$$

The regression subspace is $M = \{\mu|\mu = T\beta,\ \beta \text{ is a kxp real matrix}\}$ where T: nxk is a fixed known rank k real matrix. The set Υ of covariances of this model is

$$\Upsilon = \{I_n \otimes C | C \text{ is pxp and positive definite}\}.$$

Here, $\otimes$ denotes the usual Kronecker product as defined in Eaton (1983). Clearly the identity is in Υ and M is invariant under each element of Υ. Thus, the linear model is regular.

To apply the concentration results, it is necessary to add some distributional assumptions for the error vector $Z = Y-\mu$. Since $\mathrm{Cov}(Z)\varepsilon\Upsilon$, say $\mathrm{Cov}(Z) = I_n \otimes C$, when Z has an elliptical distribution with a density, then the density of Z has the form

$$f(z) = |C|^{-n/2} k_0(\mathrm{tr} z C^{-1} z') \ , \ z\varepsilon V \qquad (3.19)$$

In this case, Theorem 3.1 holds, and when k_0 is non-increasing on $(0,\infty)$, Theorem 3.2 holds.

An interesting case where Remark 3.1 applies is when Z has the density

$$f_1(z) = c_0 |\Lambda|^{-n/2} |I_n + z\Lambda^{-1} z'|^{-\alpha/2}, \ z\varepsilon V \qquad (3.20)$$

where $\alpha > n+p-1$ and Λ is a pxp positive definite matrix. Here, c_0 is a normalizing constant. When Cov(Z) exists and Z has the density (3.20), it is easy to check that $\mathrm{Cov}(Z)\ \varepsilon\Upsilon$. Now, observe that for each pxp positive definite matrix β, Theorem 3.3 applies directly when the density of Z is

$$f_1(z|\beta) = (\sqrt{2\pi})^{-np} \; |\beta|^{n/2} \exp[-1/2 \mathrm{tr}\; z\beta z'] \qquad (3.21)$$

since $f_1(\cdot|\beta)$ is log concave and satisfies assumption (ii) in Theorem (3.3). Thus, by Remark 3.1, averages over β of $f_1(\cdot|\beta)$, also yield densities for which the inequality

$$\psi_1(A) \leq \psi_1(A_0) \qquad (3.22)$$

holds, where ψ_1 is given by (3.17). For β positive definite choose the density

$$\psi(\beta) = c(\delta) \; |\beta|^{(\delta-p-1)/2} \exp[-1/2 \; \mathrm{tr} \; \beta]$$

where $\delta > p-1$ and $c(\delta)$ is a normalizing constant. Now, as easy integration gives

$$\int f_1(z|\beta) \; \psi(\beta)d\beta = c_0|I_n+zz'|^{-\alpha/2} \qquad (3.23)$$

where $\alpha = n+\delta$. Since $\delta > p-1$, $\alpha > n+p-1$ so inequality (3.2) holds for the density (3.23). However, the density (3.20) is obtained from (3.23) via a simple linear transformation and so (3.2) holds for the density (3.20). This completes Example 3.1. □

Section 4. Extensions

In this section, we establish some extensions of results in the previous section. In particular, a multivariate extension of a result due to Proschan (1965) is given which strengthens the multivariate extension of Olkin and Tong (1984, Theorem 3.2). The formulation of these extensions is expressed in terms of a partial ordering on the set $\underset{\sim}{A}$ defined in Section 1. This partial ordering is defined by a group and a discussion of this ordering follows.

Consider a finite dimensional inner product space $(V,(\cdot,\cdot))$ and let M be fixed subspace of V. As usual, $\underset{\sim}{A}$ is the set of all linear transformations on V to V which satisfy $Ax=x$ for $x\varepsilon M$ and $A(V) \subseteq M$. Also, let G be a closed group of orthogonal transformations on V to V which satisfies

$$gx = x \quad \text{for all } x\varepsilon M,\ g\varepsilon G \qquad (4.1).$$

Now, define G acting on $\underset{\sim}{A}$ by

$$g(A) \equiv Ag^{-1}, \quad A\varepsilon\underset{\sim}{A},\ g\varepsilon G, \qquad (4.2)$$

where Ag^{-1} means the composition of the two linear transformations A and g^{-1}. It is easily verified that (4.2) defines a left group action on $\underset{\sim}{A}$. The group action on $\underset{\sim}{A}$ defines a partial ordering on $\underset{\sim}{A}$ as follows. For $A\varepsilon\underset{\sim}{A}$, let $\rho(A)$ denote the convex hull of the set $\{Ag^{-1}|g\varepsilon G\} = \{Ag|g\varepsilon G\}$. Since $\underset{\sim}{A}$ is a convex set and is invariant under G, it follows that $\rho(A)\subseteq\underset{\sim}{A}$.

<u>Definition 4.1</u> For A_1, $A_2\varepsilon\underset{\sim}{A}$, write $A_1 \leq A_2$ iff $A_1\varepsilon\rho(A_2)$.

Partial orderings of the sort given in Definition 4.1 have arisen in a number of contexts. For example, see Rado (1952), Eaton and Perlman (1977), Marshall and Olkin (1979), Alberti and Uhlmann (1981) Eaton, (1984) and Jensen

(1984). That the above ordering is appropriate for linear models is suggested by the following result.

Lemma 4.1: Let A_0 denote the orthogonal projection onto M. Assume that

$$g_0 \equiv A_0 - \bar{A}_0 \varepsilon G$$

where

$$\bar{A}_0 = I - A_0 \quad .$$

Then, for each $A \varepsilon \underset{\sim}{A}$, $A_0 \leq A$.

Proof: Let ν denote the unique invariant probability measure on the compact group G. For $A \varepsilon \underset{\sim}{A}$, set

$$A^* = \int_G Ag \;\; \nu(dg).$$

We claim that $A^* = A_0$. To see this, consider $x \varepsilon M$. Then

$$A^*x = \int_G (Ag)x \;\; \nu(dg) = \int_G Ax \;\; \nu(dg) = x$$

since $gx = x$ for $x \varepsilon M$ and $g \varepsilon G$. For $x \varepsilon M^{\perp}$, note that $g_0 x = -x$. Using the invariance of ν, we have for $x \varepsilon M^{\perp}$,

$$y \equiv A^*x = \int Agx\ \nu(dg) = \int (Agg_0^{-1})g_0x\ \nu(dg) = -\int Agx\ \nu(dg) = -y.$$

Thus $y = A^*x = 0$. Hence A^* is the identity on M, zero on $M^\perp$ and is linear. Thus $A^* = A_0$. But A^* is an average of elements in the set $\{Ag|g\varepsilon G\}$ so $A^*\varepsilon\rho(A)$-- in other words, $A_0 = A^* \leq A$. This completes the proof. □

The above lemma shows that A_0 is always the minimal element of $\underset{\sim}{A}$ when $g_0\varepsilon G$, and of course it is A_0 which yields the Gauss-Markov estimator for regular linear models. This suggests that to study concentration inequalities for linear models, one should look at

$$\psi(A) = P\{AZ\ \varepsilon\ C\} \qquad (4.3)$$

where $C\varepsilon\zeta(M)$, $A\varepsilon\underset{\sim}{A}$ and Z is the error vector of the linear model. Conditions on Z which imply that ψ is decreasing in the ordering defined on A would automatically imply (3.2). (The statement that ψ is decreasing means: $A_1 \leq A_2$ implies $\psi(A_1) \geq \psi(A_2)$.)

We now give our first result. With $(V,(\cdot,\cdot))$, M, $\underset{\sim}{A}$ and G as above, let Z be a random vector in V. Rather than assuming Z has moments, it is more convenient in this section to express some assumptions concerning $\underset{\sim}{L}(Z)$ in terms of invariance of $\underset{\sim}{L}(Z)$.

<u>Theorem 4.1</u>

Assume that $\underset{\sim}{L}(Z) = \underset{\sim}{L}(gZ)$ for $g\varepsilon G$ and assume that Z has a density given by (2.2). Then $A_1 \leq A_2$ implies that $\psi(A_1) \geq \psi(A_2)$ where ψ is defined in (4.3). In particular, if $g_0 \varepsilon G$ (g_0 as defined in Lemma 4.1), then $\psi(A_0) \geq \psi(A)$ for all $A \varepsilon \underset{\sim}{A}$.

Proof: Because Z has a density given by (2.2) and $\underset{\sim}{L}(Z) = \underset{\sim}{L}(gZ)$ for $g \varepsilon G$, it follows that

$$gBg' = B, \quad g \varepsilon G \tag{4.4}$$

where B is given in (2.2). Recall that the function ϕ defined on $\underset{\sim}{A}$ by

$$\phi(A) = ABA' \tag{4.5}$$

is convex in the Loewner ordering--that is,

$$\phi(\alpha A + (1-\alpha)\tilde{A}) \leq \alpha\phi(A) + (1-\alpha)\phi(\tilde{A})$$

where "≤" is in the sense of positive definiteness, $\alpha \varepsilon [0,1]$, and A, $\tilde{A} \varepsilon \underset{\sim}{A}$. For a proof of this, see Marshall and Olkin (1979, p. 468).

Now, since $A_1 \leq A_2$, A_1 is a convex combination of $A_2 g$, $g \varepsilon G$ so A_1 can be written

$$A_1 = \int_G A_2 g \; \xi(dg)$$

where ξ is some probability measure on G. Applying the convexity of ϕ in (4.5), we have

$$A_1BA_1' = \phi(\int_G A_2 g\xi(dg)) \leq \int_G \phi(A_2 g)\xi(dg) \qquad (4.6)$$

But $\phi(A_2g) = A_2gBg'A_2' = A_2BA_2'$ by (4.4). Hence $A_1BA_1' \leq A_2BA_2'$. A direct application of Corollary 2.1 yields $\psi(A_1) \geq \psi(A_2)$. When $g_0 \varepsilon G$, then Lemma 4.1 shows that $A_0 \leq A$ for all $A \varepsilon \underset{\sim}{A}$ which yields the second assertion. This completes the proof. □

An immediate corollary of Theorem 4.1 which is useful in some applications is

Corollary 4.1 Let $\underset{\sim}{A}_0 \subseteq \underset{\sim}{A}$ be convex and G invarient. Then Ψ is decreasing when restricted to $\underset{\sim}{A}_0$.

Example 4.1: As in Example 3.1, take V to be the vector space of nxp matrices with the trace inner product. Let

$$M = \{\mu | \mu = e\theta' \ , \ \theta \varepsilon R^p\}$$

where e is the vector of ones in R^n.

Consider the group

$$G = \{g | g = P \otimes I_p, \ P \varepsilon P_n\}$$

where P_n is the group of nxn permutation matrices. The group G acts on V in the obvious way: $(P \otimes I_p)x = Px$ for $x \varepsilon V$. Suppose Z is a random vector in V which has an elliptical density and satisfies $\underset{\sim}{L}(Z) = \underset{\sim}{L}(PZ)$ for $P \varepsilon P_n$. For example, if Z has a density of the form (3.19), these two assumptions hold. Under these assumptions Theorem 4.1 applies directly, but it is interesting to consider $\underset{\sim}{A}_0 \subseteq \underset{\sim}{A}$ given by

$$\underset{\sim}{A}_0 = \{(eu')\otimes I_p | u \varepsilon R^n, u'e = 1\}.$$

Then, an element of $\underset{\sim}{A}_0$ evaluated at Z is

$$(eu'\otimes I_p)Z = e(\sum_1^n u_i Z_i)' \qquad (4.7)$$

where $Z_1', \ldots, Z_n'$ are the rows of Z.

The action of the group G on $\underset{\sim}{A}_0$ is

$$(P\otimes I_p)[(eu'\otimes I_p)] = (eu'\otimes I_p)(P\otimes I_p)^{-1} =$$
$$(eu'\otimes I_p)\ (P'\otimes I_p) = eu'P'\otimes I_p = e(Pu)'\otimes I_p.$$

Thus, this group action induces the obvious group action of P_n on

$$U = \{u | u \varepsilon R^n, u'e = 1\},$$

namely $u \longrightarrow Pu$, $P \varepsilon P_n$. For a convex symmetric set $C \subseteq R^n$, let

$$\xi(u) \equiv P\{\sum_1^n u_i Z_i \in C\}. \tag{4.8}$$

Theorem 4.1 shows that $\xi(u) \geq \xi(v)$ when u is in the convex hull of $\{Pv | P \in P_n\}$--in other words, ξ is a Schur concave function of $u \in U$. (See Marshall and Olkin (1979), p. 131 for a discussion of the equivalence of the usual definition of majorization and the one used above.) Since $u'e=1$ is just a normalization, this implies that ξ is Schur concave on all of R^n. In Application 4.1 of Olkin and Tong (1984), this result was proved for the case p=1 when the function k in (2.2) (defining the elliplical distribution) is decreasing. Paraphrased, the above result says that if Z is elliplical and its distribution is invariant under permutation of the rows of Z, then $\xi(u)$ in (4.8) is Schur concave. In particular, for all $u \in U$,

$$P\{\frac{1}{n}\sum_1^n Z_i \in C\} \geq P\{\sum_1^n u_i Z_i \in C\}$$

This completes Example 4.1. □.

Our final result extends a Theorem in Proschan (1965). Here is a statement of that theorem.

<u>Theorem (Proschan (1965))</u>. Let $Y_1, \ldots, Y_n$ be iid symmetric random variables with a common density which is log concave on R^1. For a>0 and non-negative real numbers $u_1, \ldots, u_n$, let

$$\xi(u) = P\{|\sum_1^n u_i \gamma_i| \leq a\} \tag{4.9}$$

where u is the vector with coordinates $u_1,\ldots,u_n$. Then $\xi(\cdot)$ is a Schur concave function.

Olkin and Tong (1984) extended this theorem to the case where $\gamma_1,\ldots,\gamma_n$ are iid symmetric random vectors in R^p with a common log concave density. In this case ξ is defined as

$$\xi(u) = P\{\sum_1^n u_i \gamma_i \in C\} \tag{4.10}$$

where C is a symmetric convex subset of R^p. The Olkin-Tong conclusion is that $\xi(\cdot)$ is a Schur concave function of u, $u \in R^n$.

To formulate our extension of the above results, let Z:nxp be a random matrix with rows $Z_1',\ldots,Z_n'$. Let V be the vector space of nxp matrices. For a given symmetric convex set $C \subseteq R^p$ and vector $u \in R^n$, let

$$\phi(u) = P\{\sum_1^n u_i Z_i \in C\} = P\{Z'u \in C\} \tag{4.11}$$

Our result below is most conveniently expressed in terms of a special group of nxn matrices G_0. This group consists of all nxn permutation matrices and all nxn diagonal matrices with ± 1's on the diagonal. The group G_0 defines a partial ordering on R^n as follows. For each $v \in R^n$, let $\rho(v)$ denote the convex hull of $\{gv | g \in G_0\}$ and write $u \leq v$ to mean $u \in \rho(v)$. This ordering is discussed at length in Eaton and Perlman (1977). A real valued function τ defined on R^n is decreasing relative to the above ordering if $u \leq v$ implies $\tau(u) \geq \tau(v)$.

Theorem 4: Suppose that the density of Z, say f, satisfies

(i) $f(gz) = f(z)$ for all $g \varepsilon G_0$, $z \varepsilon V$ (4.12)

(ii) f is log concave.

Then the function ϕ defined by (4.11) is decreasing for each convex symmetric set $C \subseteq R^p$.

Remark 4.1

Before proving Theorem 4.2, it is useful to see how this result implies those of Olkin and Tong (1984) and Proschan (1965). First observe that if the rows of Z are iid symmetric random vectors in R^p (as in the Olkin and Tong case) with a common log concave density, then the density of Z is easily shown to satisfy (4.12). Thus by Theorem 4.2 ϕ is decreasing. Now, if u, v are in R^n and v majorizes u, then u is an element of the convex hull of the set of all vectors of the form hv where h is an nxn permutation matrix. Hence, $u \varepsilon \rho(v)$ so $\phi(u) \geq \phi(v)$ which shows that ϕ is Schur concave. Thus $\xi(\cdot)$ given in (4.10) is Schur concave. □

Proof of Theorem 4.2: The proof is based on the theory developed in Eaton and Perlman (1977). Let t be either of the vectors

$$\begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad \text{or} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

in R^n. In order to show that ϕ is decreasing it is sufficient to show that for each vector u_0 perpendicular to t, the map

$$\beta \longrightarrow \phi(u_0 + \beta t) \tag{4.13}$$

is non-increasing for $\beta \varepsilon (0,\infty)$. (See Eaton and Perlman (1977); also see Eaton (1984), Section 3).

Now

$$\phi(u_0 + \beta t) = P\{Z'u_0 + \beta Z't \;\varepsilon\; C\} \tag{4.14}$$

If $u_0 = 0$, (4.14) is obviously non-increasing in $\beta \varepsilon (0,\infty)$. For $u_0 \neq 0$ and t one of the vectors above, the joint density of

$$(Z'u_0, Z't): px2$$

in R^{2p} is log concave. This follows from a result due to Prekopa (1973) which asserts that marginal distributions of log concave distributions are log concave. With

$$W_1 = Z'u_0, \quad W_2 = Z't,$$

there is a log concave version of the conditional density of W_1 given W_2 (see the remarks at the end of Section 2). Thus

$$\phi(u_0 + \beta t) = \underset{\sim}{E}P\{W_1 + \beta W_2 \in C | W_2 = w\} =$$

$$\underset{\sim}{E}P\{W_1 + \beta w \in C | W_2 = w\} \qquad (4.15)$$

Let $f_0(w_1|w)$ denote the version of the conditional density of W_1 given W_2 described in Section 2. For the moment, assume that

$$f_0(-w_1|w) = f_0(w_1|w) \qquad (4.16).$$

This identity is verified below. Under this assumption, the log concavity of $f_0(\cdot|w_1)$ implies that

$$\{w_1|f_0(w_1|w) \geq a\}$$

is a convex symmetric set for each $a > 0$. Thus, Anderson's Theorem shows that

$$\beta \longrightarrow P\{W_1 + \beta w \varepsilon C | W_2 = w\}$$

is non-increasing for $\beta \varepsilon [0,\infty)$. Thus, averaging over W_2 shows that (4.13) is non-increasing. Tnis completes the proof modulo the verification of (4.16).

The verification of (4.16) goes as follows. The joint density of (W_1, W_2), say $f_1(w_1, w_2)$ is log concave. Because of assumption 4.12(i),

$$\underset{\sim}{L}(Z) = \underset{\sim}{L}(gZ) \ , \ g \varepsilon G_0$$

so

$$\underset{\sim}{L}(W_1, W_2) = \underset{\sim}{L}(Z'u_0, Z't) =$$

$$\underset{\sim}{L}((gZ)'u_0,(gZ)'t) = \underset{\sim}{L}(Z'g'u_0,Z'g't) \quad (4.17)$$

for all $g \varepsilon G_0$. Picking $g' = -I_n$ in (4.17) shows that

$$\underset{\sim}{L}(W_1,W_2) = \underset{\sim}{L}(-W_1,-W_2) \quad (4.18).$$

Picking

$$g' = I_n - 2tt$$

which is in G_0 for the two possible values of t shows that

$$\underset{\sim}{L}(W_1,W_2) = \underset{\sim}{L}(W_1,-W_2) \quad (4.19)$$

and thus

$$\underset{\sim}{L}(W_1,W_2) = \underset{\sim}{L}(-W_1,W_2) \quad (4.20).$$

The relations (4.18), (4.19) and (4.20) show that the joint density of (W_1,W_2) can be chosen so that

$$f_1(w_1,w_2) = f_1(-w_1,-w_2) = f_1(w_1,-w_2) = f_1(-w_1,w_2) \quad (4.21).$$

The relations (4.21) together with the discussion at the end of Section 2 show that (4.16) holds. The proof is complete. □

Remark 4.2

Theorem 4.2 can be extended via a convex combination argument in much the same way that Theorems 3.2 and 3.3 were extended in Remark 3.1. For example, let $\underset{\sim}{F}_1$ denote the class of densities f such that

(i) $f(gz) = f(z)$ for all $g \varepsilon G_0$, $z \varepsilon V$

(ii) the function ϕ defined in (4.11) is decreasing.

Obviously $\underset{\sim}{F}_1$ is a convex set. By Theorem 4.2, $\underset{\sim}{F}_1$ contains the log concave f's. Hence, $\underset{\sim}{F}_1$ contains convex combinations of the log concave f's which satisfy (i). In particular, here is useful corollary.

Corollary 4.2: Suppose that the density of Z, say f, satisfies

(i) $f(gz) = f(z)$ for all $g \varepsilon G_0$, $z \varepsilon V$

(ii) $\{z | f(z) \geq \alpha\}$ is a convex set for all $\alpha > 0$.

Then ϕ defined in (4.11) is decreasing.

Proof: The argument is the same as that used in Remark 3.1. □

REFERENCES

Alberti, P.M. and Uhlmann, A. (1982). Stochasticity and Partial Order., Reidel, Holland.

Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170-176.

Berk, R. and Hwang, J.T. (1984). Optimality of the least squares estimator. Unpublished.

Birnbaum, Z.W. (1948). On random variables with comparable peakedness. Ann. Math. Statist. 19, 76-81.

Das Gupta, S., Eaton, M.L., Olkin, I., Perlman, M.D., Savage, L.J. and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. Proc. Sixth Berkeley Symp. on Prob. and Statist. Vol. II, 241-265.

Eaton, M.L. (1982). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

Eaton, M.L. (1984). On group induced orderings. In Inequalities in Statistics and Probability, Y.L. Tong ed. Institute of Mathematical Statistics, Hayward, CA, 13-25.

Eaton, M.L. and Perlman, M.D. (1977). Reflection groups, generalized Schur functions and the geometry of majorization. Ann. Probab. 5, 829-860.

Jensen, D. (1984). Invariant ordering and order preservation. In Inequalities in Statistics and Probability, Y.L. Tong ed. Institute of Mathematical Statistics, Hayward, CA, 26-34.

Marshall, A. and Olkin, I. (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

Olkin, I. and Tong, Y.L. (1984). Peakedness in multivariate distributions. Stanford Technical Report No. 200, Stanford, CA.

Prekopa, A. (1973). On logarithmic concave measures and functions. Acta. Scient. Mat. (Szeged) 34, 335-343.

Proschan, F. (1965). Peakedness of distributions of convex combinations. Ann. Math. Statist. 36, 1703-1706.

Rado, R. (1952). An inequality. J. London Math. Soc. 27, 1-6.

Sherman, S. (1955). A theorem on convex sets with applications. Ann. Math. Statist. 26, 763-766.