Stochastic Orders and Decision under Risk
IMS Lecture Notes – Monograph Series (1991)

MULTIVARIATE PROBABILITY INEQUALITIES: CONVOLUTION THEOREMS, COMPOSITION THEOREMS, AND CONCENTRATION INEQUALITIES

BY MORRIS L. EATON* AND MICHAEL D. PERLMAN†

University of Minnesota and University of Washington

Several important multivariate probability inequalities can be formulated in terms of multivariate convolutions of the form $\int f_1(x)f_2(x-\theta)dx$, where usually $f_1 = I_C$ is the indicator of a region $C \subseteq \mathbb{R}^n$, f_2 is a probability density on $\mathbb{R}^n$, and θ is a translation parameter. Often f_1 and f_2 possess convexity, monotonicity, and/or symmetry properties. More general multivariate compositions of the form $\int h(x)f(x \mid \theta)\mu(dx)$ also arise. Here several important convolution and composition theorems will be reviewed; these provide comparisons of Prob(C) under differing multivariate distributions. The convolution theorems are then applied to obtain concentration inequalities for Prob(C) under Gaussian or elliptically contoured distributions with varying multivariate scale parameter Σ.

1. Introduction. In multivariate statistical analysis, the power function of a hypothesis test and the confidence coefficient of a confidence region are determined by the probability of a multivariate region $C \subseteq \mathbb{R}^n$. Frequently the region C possesses convexity, monotonicity, and/or symmetry properties inherited from corresponding properties of the multivariate distributions in the statistical model. In order to establish properties of the statistical procedure such as unbiasedness, it is necessary to compare the probabilities of C under different multivariate distributions in the model.

Several important multivariate probability inequalities can be formulated in terms of *convolutions* of the form

$$\psi(\theta) \equiv \int_{\mathbb{R}^n} f_1(x)f_2(x-\theta)dx. \tag{1.1}$$

* Research supported in part by National Science Foundation Grant No. DMS-88-00284

† Research supported in part by National Science Foundation Grant No. DMS-89-02211

AMS 1980 Subject Classification: primary 60E15; secondary 62H10

Key words and phrases: Multivariate convolutions, compositions, concentration inequalities, convex region, Schur convex, Gaussian distribution, elliptically contoured distribution.

In most applications $f_1 \equiv I_C$ is the indicator function of a region $C \subset \mathbb{R}^n$ and f_2 is a probability density function (pdf) on $\mathbb{R}^n$. For example, if C is the acceptance region of a test of a statistical hypothesis concerning a translation parameter θ, then $1 - \psi$ is the power function of the test. It is desired to find conditions on θ_1, θ_2 which guarantee that $\psi(\theta_1) \geq \psi(\theta_2)$.

In Section 2 four such convolution theorems are discussed and several applications are sketched. The first theorem, due to T. W. Anderson (1955), gives sufficient conditions that the convolution ψ be *ray decreasing* on $\mathbb{R}^n$. (A function ψ defined on $\mathbb{R}^n$ is ray decreasing if, for $\theta \in \mathbb{R}^n$ and $\beta \in [0, \infty)$, $\psi(\beta\theta)$ is a decreasing [= non-increasing] function of β). Anderson placed convexity and symmetry assumptions on f_1 and f_2. The symmetry assumption suggested that invariance under a group of orthogonal transformations may be playing a role. Mudholkar (1966) extended Anderson's original theorem by developing this group-theoretic theme. Marshall and Olkin (1974) showed that the convexity assumption in the Anderson-Mudholkar treatment could be weakened in the important special case that the symmetry group under consideration was the group $\mathcal{P}_n$ of $n \times n$ permutation matrices. The Marshall-Olkin work was then extended to all reflection groups by Eaton and Perlman (1977a). These four papers form the basis of the discussion in Section 2.

There are a number of important parametric multivariate statistical problems where the vector θ is not a translation parameter. In such cases the convolution (1.1) is often replaced by the more general *composition*

$$\psi(\theta) \equiv \int_{\mathbb{R}^n} h(x) f(x \mid \theta) \mu(dx), \tag{1.2}$$

where μ is either Lebesgue measure on $\mathbb{R}^n$ or counting measure on the integer lattice points of $\mathbb{R}^n$, and where $f(\cdot \mid \theta)$ is either a continuous or discrete pdf for every $\theta \in \mathbb{R}^m$. In general m and n need not be the same, although they will be so here. In Section 3 we discuss composition theorems of Hollander, Proschan, and Sethuraman (1977) and Proschan and Sethuraman(1977) which extend the Marshall-Olkin and Eaton-Perlman convolution theorems. Application is made to the case where $f(x \mid \theta)$ is the joint density of independent Poisson variates $X_1, \ldots, X_n$ with different intensity parameters $\theta_1, \ldots, \theta_n$.

If $X \sim N_1(0, \sigma^2)$ (the univariate normal distribution with mean 0 and variance σ^2) then $P_\sigma\{X \in [-a, a]\}$ is a decreasing function of σ for every symmetric interval $[-a, a]$. This remains true for any univariate distribution whose density is symmetric about 0: the probability of any symmetric interval decreases as the scale parameter increases. In Sections 4 and 5 we discuss extensions of these *concentration inequalities* to the multivariate case, first for the multivariate normal (Gaussian) distribution and then for multivariate distributions with elliptically contoured densities. In each case one is concerned with the behavior of the probability of symmetric multivariate sets as a function of the multivariate scale parameter matrix Σ, which is (proportional to) the covariance matrix for distributions with finite second moments. In both the Gaussian case and (surprisingly) the elliptically contoured case, the concentration inequalities are obtained as corollaries of the corresponding convolution inequalities in Section 2.

Statistical applications of both concentration inequalities and the convolution theorem for reflection groups can be found in Eaton (1988).

2. Multivariate Convolution Theorems. Anderson (1955) proved an important inequality for the convolution of two symmetric unimodal functions on $\mathbb{R}^n$. A function f on $\mathbb{R}^n$ is said to be *(centrally) symmetric* if $f(-x) = f(x)$ for every x, while f is *unimodal* if the set $\{x \mid f(x) \geq c\}$ is convex for every real number c. We shall deal only with nonnegative unimodal functions f, in which case only nonnegative c need be considered.

THEOREM 2.1 (Anderson (1955)). *Suppose that f_1 and f_2 are nonnegative symmetric unimodal functions on $\mathbb{R}^n$. Then*

$$\psi(\theta) \equiv \int f_1(x) f_2(x-\theta) dx \tag{2.1}$$

is symmetric and is ray decreasing. In particular, ψ is maximized at $\theta = 0$.

An important special case occurs when f_2 is a symmetric unimodal probability density and f_1 is the indicator function of a convex symmetric set C (hence $-C = C$). An immediate application of Theorem 2.1 yields that

$$\int_{C+\beta\theta} f_2(y) dy \equiv \mathrm{Prob}[C + \beta\theta]$$

is a symmetric unimodal function of the real variable β for each $\theta \in \mathbb{R}^n$. For this reason, Anderson's result is often described as a moving set inequality.

To prove Theorem 2.1, first observe that (2.1) is linear in f_1 and f_2. Further, an easy argument shows that if Theorem 2.1 holds when f_1, f_2 are indicators of convex symmetric sets, then it holds not only for symmetric unimodal functions, but also for functions which can be approximated (in an appropriate topology – see Sherman (1955)) by non-negative linear combinations of indicators of convex symmetric sets. When $f_i = I_{C_i}$, $i = 1, 2$, however, (2.1) reduces to

$$\psi(\theta) \equiv \mu(C_1 \cap (C_2 + \theta)) \tag{2.2}$$

where μ is Lebesgue measure. Obviously ψ is a symmetric function of θ and an application of the Brunn-Minkowski inequality shows that ψ is ray decreasing (see Anderson (1955) or Perlman (1990) for details).

An interesting alternative proof of Theorem 2.1 follows by first noting that $\psi(\theta)$ in (2.2) is a log concave function. This observation derives from Davidovic, Korenbljum and Hacét (1969) or from Prekopa (1973), who showed that the convolution of two log concave functions is log concave. [For a discussion and proof of Prekopa's theorem based on an unpublished argument of Brascamp and Lieb (1974), see Eaton (1987a). Other interesting treatments may be found in Borell (1975), Das Gupta (1980), and Dharmadhikari and Joag-dev (1988)]. The symmetry and log concavity of ψ show that for $\beta \in \mathbb{R}$ and fixed $x \in \mathbb{R}^n$, $\psi(\beta x)$ is log concave and symmetric in β, hence ψ is ray decreasing.

The central symmetry assumption in Theorem 2.1 can be expressed in terms of the two-element group $G = \{\pm I\}$ which acts on $\mathbb{R}^n$, where I is the $n \times n$ identity matrix. Symmetry of a function f is equivalent to its G-invariance, i.e., $f(gx) = f(x)$ for each $g \in G$. The ray decreasing condition (together with symmetry) can be expressed in terms of G by introducing a pre-ordering on $\mathbb{R}^n$. Write $x \prec_G y$ iff x is in the convex hull of $\{y, -y\}$, the orbit of y under the action of G. It is easy to check that a function f is symmetric and ray decreasing iff $x \prec_G y$ implies $f(x) \geq f(y)$ – such functions are called *G-decreasing*. Thus Theorem 2.1 asserts that the convolution of two G-invariant unimodal functions is G-decreasing, where $G = \{\pm I\}$.

Mudholkar (1966) extended Theorem 2.1 by considering more general groups G of orthogonal linear transformations. Let G be an arbitrary closed subgroup of the group $\mathcal{O}_n$ of $n \times n$ orthogonal matrices. Given $y \in \mathbb{R}^n$, let $C_G(y)$ denote the convex hull of $\{gy \mid g \in G\}$ (the G-orbit of y) and write $x \prec_G y$ to denote that $x \in C_G(y)$. The relation $\prec \equiv \prec_G$ is easily shown to be reflexive and transitive, and is sometimes called a pre-ordering (see the discussion on page 13 of Marshall and Olkin (1979)).

DEFINITION 2.1. A real-valued function f defined on $\mathbb{R}^n$ is called *G-decreasing* if $x \prec_G y$ implies that $f(x) \geq f(y)$. A subset $C \subset \mathbb{R}^n$ is called *G-decreasing* if its indicator function I_C is G-decreasing.

Note that if f is G-decreasing, then necessarily f is G-invariant because $x \prec gx \prec x$ for all $x \in \mathbb{R}^n$ and all $g \in G$. Also observe that if f is G-invariant and log concave, then f is G-decreasing. [If $x \prec y$, then $x = \sum_g \alpha_g gy$ where the weights α_g are nonnegative and add to one. Log concavity and G-invariance yield $f(x) = f\left(\sum \alpha_g gy\right) \geq \Pi[f(gy)]^{\alpha_g} = f(y)$.]

THEOREM 2.2 (Mudholkar (1966)). *Suppose that f_1 and f_2 are nonnegative G-invariant unimodal functions and define ψ as in (2.1). Then ψ is G-decreasing.*

Because (2.1) is linear in both f_1 and f_2, it is again sufficient to establish the theorem when $f_i = I_{C_i}$, $i = 1, 2$, where C_1 and C_2 are now convex G-invariant sets. In this case ψ is again given by (2.2). At this point Mudholkar used the Brunn-Minkowski inequality in much the same manner as Anderson to show that ψ is G-decreasing. An alternative argument is to observe that ψ in (2.2) is clearly G-invariant and is log concave, being the convolution of two log concave functions, hence ψ is G-decreasing. Note that when $G = \{\pm I\}$, Theorem 2.2 reduces to Theorem 2.1.

Now, suppose that f satisfies the conditions of Theorem 2.2, – that is, f is G-invariant and unimodal. Then it is not hard to show that f is G-decreasing. Thus it is natural to ask if the assumptions in Theorem 2.2 can be weakened to the simpler assumption that f_1 and f_2 are G-decreasing. In other words, is it true that the convolution of two nonnegative G-decreasing functions is G-decreasing? In general the answer is no, as it is easy to construct counterexamples when $G = \{\pm I\}$ as in Theorem 2.1. The first positive result in this direction was established by Marshall and Olkin (1974) when G is the group $\mathcal{P}_n$ of all $n \times n$ permutation matrices.

With $G = \mathcal{P}_n$, the pre-ordering $\prec \equiv \prec_G$ defined on $\mathbb{R}^n$ is the classical majorization pre-ordering (see Marshall and Olkin (1979) for a thorough treatment of majorization). In this case, the G-decreasing functions are often called *Schur-concave* functions.

THEOREM 2.3 (Marshall and Olkin (1974)). *Suppose that f_1 and f_2 are nonnegative $\mathcal{P}_n$-decreasing (Schur-concave) functions and define ψ as in (2.1). Then ψ is a $\mathcal{P}_n$-decreasing function.*

Rather than indicate a proof of Theorem 2.3, it is instructive to first give some background information which underlies both Theorem 2.3 and its generalization, Theorem 2.4. For any vector $t \in \mathbb{R}^n$ with $|t| = 1$, let

$$R_t = I - 2tt'. \tag{2.3}$$

It is easily seen that R_t is an orthogonal transformation which reflects vectors across the $(n-1)$-dimensional subspace perpendicular to t. Any such transformation is called a *reflection*. For $i = 1, \ldots, n-1$, let $t_i \in \mathbb{R}^n$ be the unit vector with ith coordinate $1/\sqrt{2}$, $(i+1)$-th coordinate $(-1)/\sqrt{2}$, and the remaining coordinates zero. Also, let $T_0 = \{t_1, t_2, \ldots, t_{n-1}\}$.

Here are two basic facts.

(F.1) $\mathcal{P}_n$ is generated (algebraically) by $\{R_t \mid t \in T_0\}$, i.e., every element of $\mathcal{P}_n$ can be written as a product of R_t's with $t \in T_0$.

(F.2) A function f on $\mathbb{R}^n$ is $\mathcal{P}_n$-decreasing (Schur concave) iff for each $t \in T_0$

and $u \in \mathbb{R}^n$ with $t'u = 0$, $f(u + \beta t)$ is a symmetric unimodal function of $\beta \in \mathbb{R}$.

Assertion F.1 simply states that every permutation can be written as a product of adjacent transpositions. However, F.2 is somewhat deeper and essentially relies on what is called a "path lemma" in Eaton and Perlman (1977a). A direct proof of F.2 from first principles can be found in Eaton (1987a), pages 15–30.

There are other well known groups where a similar analysis is valid. For example, let $\mathcal{D}_n$ be the group of $n \times n$ diagonal matrices where each diagonal element is ± 1. Obviously $\mathcal{D}_n$ has 2^n elements. Let $\epsilon_1, \ldots, \epsilon_n$ be the standard orthonormal basis for $\mathbb{R}^n$ and set $T_1 = \{\epsilon_1, \ldots, \epsilon_n\}$. Then

(F.3) $\mathcal{D}_n$ is generated algebraically by $\{R_t \mid t \in T_1\}$.

(F.4) A function f on $\mathbb{R}^n$ is $\mathcal{D}_n$-decreasing iff for each $t \in T_1$ and $u \in \mathbb{R}^n$ with $t'u = 0$, $f(u + \beta t)$ is a symmetric unimodal function of $\beta \in \mathbb{R}$.

These observations suggest the possibility of developing an approach which covers any group $G \subseteq \mathcal{O}_n$ that is generated by reflections. This was done in Eaton and Perlman (1977a). To outline the results there, consider an arbitrary set T of unit vectors in $\mathbb{R}^n$ and let $G(T)$ be the closure of the group generated algebraically by $\{R_t \mid t \in T\}$; $G(T)$ is called a *reflection group.* The group $G(T)$ is *reducible* if there exists a proper subspace $M \subsetneq \mathbb{R}^n$ such that $gM = M$ for all $g \in G(T)$; otherwise $G(T)$ is *irreducible.* The arguments in Eaton and Perlman (1977a) show that attention can be restricted to the irreducible case without loss of generality. There are then two cases. If $G(T)$ is infinite and irreducible, then $G(T) = \mathcal{O}_n$ (cf. Eaton and Perlman (1977b)), a rather trivial case since the $\mathcal{O}_n$-decreasing functions are just the radial decreasing functions. If $G(T)$ is finite and irreducible, a complete listing of the possibilities for $G(T)$ is known – see Grove and Benson (1985).

In order to extend Theorem 2.3 from $\mathcal{P}_n$ to a general reflection group $G(T)$, one requires the following result, proved in Eaton and Perlman (1977a) and Eaton (1987a):

PROPOSITION 2.1. *Suppose that $G(T)$ is an irreducible reflection group. Given a real-valued function f on $\mathbb{R}^n$, the following are equivalent:*

(i) f is $G(T)$-decreasing;

(ii) For each $t \in T$ and each $u \in \mathbb{R}^n$ with $t'u = 0$, $f(u + \beta t)$ is a symmetric unimodal function of $\beta \in \mathbb{R}$.

THEOREM 2.4. (Eaton and Perlman (1977a)). *Let $G(T)$ be a reflection group. Suppose that f_1, f_2 are nonnegative $G(T)$-decreasing functions and define ψ as in (2.1). Then ψ is a $G(T)$-decreasing function.*

PROOF: It suffices to treat the irreducible case. For $\beta \in \mathbb{R}, t \in T$, and $u \in \mathbb{R}^n$ with $t'u = 0$, consider the function

$$\varphi(\beta) \equiv \psi(u + \beta t) = \int f_1(x) f_2(x - u - \beta t) dx.$$

By Proposition 2.1, it suffices to show that ψ is a symmetric unimodal function of β. Define $M = \{x \mid x't = 0\}$. Then x may be uniquely decomposed as $x = v + \gamma t$ where $\gamma \in \mathbb{R}$ and $v't = 0$. Thus we have that

$$\varphi(\beta) = \int_M \int_{-\infty}^{\infty} f_1(v + \gamma t) f_2(v - u + (\gamma - \beta)t) d\gamma dv.$$

However, for u, v and t fixed, Proposition 2.1 implies that the inside integral is the convolution of two symmetric unimodal functions on $\mathbb{R}$, hence is a symmetric unimodal function of β (cf. Wintner (1938)). Thus, φ is symmetric and unimodal and the result follows. ∎

As mentioned in Section 1, a common application of such convolution theorems is to the study of power functions of tests in translation-parameter problems. A sampling of specific applications can be found in Perlman (1990) as well as Marshall and Olkin (1979), Tong (1980), Eaton (1987a), (1988), and Dharmadhikari and Joag-dev (1988).

Finally, there are a number of open problems regarding G-decreasing functions and the validity of Theorem 2.4 when $G \subseteq \mathcal{O}_n$ is not necessarily a reflection group. For example, suppose that f is G-decreasing on $\mathbb{R}^n$ and has a differential ∇f. Then for all $\alpha \in [0, 1]$, $g \in G$, and $x \in \mathbb{R}^n$,

$$f((1 - \alpha)x + \alpha g x) \geq f(x),$$

which yields the necessary condition

$$(gx)'(\nabla f)(x) \geq x'(\nabla f)(x) \tag{2.4}$$

that f be G-decreasing. When G is a reflection group, (2.4) is also sufficient, provided that f is G-invariant and has a differential. The sufficiency of this condition for non-reflection groups is an open question.

With regard to extending the convolution theorems, here is the appropriate question: *under what conditions on G is it true that the convolution of two nonnegative G-decreasing functions again be G-decreasing?* Theorem 2.4 provides one condition, namely that G be a reflection group. Of course when $G = \{\pm I\}$ as in Anderson's theorem, then the convolution of G-decreasing (= symmetric ray-decreasing) functions need not be G-decreasing. However, more delicate counterexamples are also available. Let G_k denote the cyclic group of

2×2 rotation matrices generated by the counterclockwise rotation through the angle $2\pi/k$, $k = 3, 4, \ldots$. Thus G_k is an Abelian group with k elements acting on $\mathbb{R}^2$. Eaton (1984) showed that the convolution of G_k-decreasing functions need not be G_k-decreasing.

There are groups of interest, however, for which the argument used in Eaton (1984) will not produce counterexamples. These groups, discussed in some detail in Eaton (1987a, Chapter 6, Examples 6.3, 6.4 and 6.6), provide interesting open questions concerning extensions of the convolution theorems in this section.

3. Multivariate Composition Theorems. The convolution theorems of the previous section give conditions under which various inequalities can be obtained for parametric functions of the form

$$\psi(\theta) = \int_{\mathbb{R}^n} h(x) f(x - \theta) dx. \tag{3.1}$$

In most applications f is a probability density on $\mathbb{R}^n$ and h is the indicator of a set. For example, Theorem 2.3 shows that if f is a Schur-concave density and h is Schur concave, then ψ is also Schur-concave. This establishes certain inequalities for ψ which are direct consequences of majorization results.

There are a number of important parametric families where the vector θ is not a translation parameter. An interesting example is the Poisson distribution. Suppose that $X_1, \ldots, X_n$ are independent Poisson variables with the density of X_i given by

$$p(x \mid \theta_i) = \begin{cases} \frac{e^{-\theta_i} \theta_i^x}{x!}, & x = 0, 1, 2, \cdots \\ 0, & x = -1, -2, \cdots \end{cases}$$

where $\theta_i > 0$. With $\mathbf{Z} = \{0, \pm 1, \ldots\}$, the joint density of $X \equiv (X_1, \ldots, X_n)$ is given by

$$f(x \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta_i) \tag{3.2}$$

for $x \in \mathbf{Z}^n$. Given a Schur convex function h defined on $\mathbf{Z}^n$ (i.e., $-h$ is Schur-concave on $\mathbf{Z}^n$), set

$$\psi(\theta) = \int h(x) f(x \mid \theta) \mu(dx) \tag{3.3}$$

where μ denotes counting measure on $\mathbf{Z}^n$. Rinott (1973) showed that ψ in (3.3) is Schur convex. His proof consisted of first establishing a corresponding result for the multinomial distribution and then averaging the multinomial to obtain the Poisson result. An alternative method was developed in Hollander, Proschan and Sethuraman (1977) and Proschan and Sethuraman (1977). It is this alternative method which is of interest here because it involves reflections and a composition theorem.

As in the previous section, let T be a set of unit vectors in $\mathbb{R}^n$ and let

$$\mathcal{R} \equiv \mathcal{R}(T) = \{R_t \mid t \in T\}$$

be the corresponding set of reflections (cf. (2.3)).

DEFINITION 3.1. A function $K(x, y)$ defined on $\mathbb{R}^n \times \mathbb{R}^n$ is a decreasing reflection (DR) kernel (relative to $\mathcal{R}$) if

(i) for all $x, y \in \mathbb{R}^n$ and $t \in T$, $K(x, y) = K(R_t x, R_t y)$

(ii) $(t'x)(t'y) \geq 0$ implies $K(x, y) \geq K(R_t x, y)$.

Condition (i) of Definition 3.1 implies that K is invariant under the group $G \subseteq \mathcal{O}_n$ of all orthogonal transformations generated algebraically by $\mathcal{R}$. That is, (i) is equivalent to the condition that $K(x, y) = K(gx, gy)$, for every $g \in G$. Condition (ii) has a geometric interpretation. The inequality $(t'x)(t'y) \geq 0$ is equivalent to the assertion that x and y lie on the same side of the hyperplane $H_t = \{u \mid t'u = 0\}$ while $R_t x$ and y are on opposite sides of H_t (at least when $(t'x)(t'y) > 0$). Thus condition (ii) compares K at (x, y) [on the same side of H_t] to K at $(R_t x, y)$ [on opposite sides of H_t].

REMARK 3.1. In some cases K is not defined on $\mathbb{R}^n \times \mathbb{R}^n$, but on $\mathcal{X} \times \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are both $\mathcal{R}$-invariant subsets of $\mathbb{R}^n$. There is no change in Definition 3.1 or its interpretation in this case. The example below illustrates the need for this generality.

EXAMPLE 3.1. Let T_0 be the set of reflections defined in Section 2 which generate the group $\mathcal{P}_n$ of $n \times n$ permutation matrices. Thus, $T_0 = \{t_1, t_2, \ldots, t_{n-1}\}$ where

$$t_i = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad i = 1, \cdots, n-1$$

and 1 occurs as the ith coordinate. Consider a parametric family of densities $p(v \mid \eta)$ defined for $v \in V \subseteq \mathbb{R}$ and $\eta \in \Gamma \subseteq \mathbb{R}$. With $\mathcal{X} = V \times \cdots \times V \subseteq \mathbb{R}^n$ and $\Theta = \Gamma \times \cdots \times \Gamma \subseteq \mathbb{R}^n$, define K on $\mathcal{X} \times \Theta$ by

$$K(x, \theta) = \prod_{i=1}^{n} p(x_i \mid \theta_i).$$

Obviously $K(gx, g\theta) = K(x, \theta)$ for all permutation matrices g so (i) of Definition 3.1 holds. It is not hard to show that (ii) of Definition 3.1 holds iff p has a monotone likelihood ratio (MLR) (Eaton (1967)). In particular, f in (3.2) defined by the Poisson distribution is a DR kernel for the set of reflections defined by T_0.

In the case of the permutation group $\mathcal{P}_n$ generated by reflections $R_t, t \in T_0$, the use of the decreasing reflection properties of certain parametric densities was present implicitly in the work of Savage (1957). Later Eaton (1967) isolated the DR property (in the case of $\mathcal{P}_n$) in work related to ranking problems. The systematic development of these ideas by Hollander, Proschan and Sethuraman (1977) showed the power and usefulness of such notions. Here are two results from that paper which provide an elegant proof of the Schur-convexity of ψ defined in (3.3) (the Poisson case). These results are now stated for $\mathcal{P}_n$ and the reflections generated by T_0.

PROPOSITION 3.1. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex cone which is $\mathcal{P}_n$-invariant. A nonnegative function f defined on $\mathcal{X}$ is Schur-convex iff $K(x, y) \equiv f(x + y)$ is DR on $\mathcal{X} \times \mathcal{X}$.*

THEOREM 3.1 (Hollander, Proschan, and Sethuraman (1977)). *Let $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ be $\mathcal{P}_n$-invariant subsets of $\mathbb{R}^n$. Suppose that $K_1(x, y)$ and $K_2(y, z)$ are DR on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$, respectively. If μ is a $\mathcal{P}_n$-invariant measure on $\mathcal{Y}$, then*

$$K_3(x, z) \equiv \int_{\mathcal{Y}} K_1(x, y) K_2(y, z) \mu(dy)$$

is DR on $\mathcal{X} \times \mathcal{Z}$ when the above integral exists.

The promised application to the Poisson case runs as follows. First observe that the Poisson family defined by the density (3.2) is a *convolution family*, that is, the density given in (3.2) satisfies

$$f(x \mid \theta + \xi) = \int_{\mathbf{Z}^n} f(x - u \mid \theta) f(u \mid \xi) \mu(du) \tag{3.4}$$

where μ is counting measure on $\mathbf{Z}^n$.

PROPOSITION 3.2 (Proschan and Sethuraman (1977)). *For the Poisson family, the function ψ defined by (3.3) is Schur-convex when h is Schur-convex.*

PROOF: By Proposition 3.1, it suffices to show that $\psi(\theta+\xi)$ is DR. But, from (3.4) we have that

$$\begin{aligned}\psi(\theta+\xi) &= \int h(x)f(x \mid \theta+\xi)\mu(dx)\\ &= \int\int h(x)f(x-u \mid \theta)f(u \mid \xi)\mu(du)\mu(dx)\\ &= \int\left[\int h(z+u)f(z \mid \theta)\mu(dz)\right] f(u \mid \xi)\mu(du)\end{aligned}$$

where the translation invariance of the measure μ has been used to obtain the final equality. But Theorem 3.1 shows that

$$K_1(\theta,u) \equiv \int h(z+u)f(z \mid \theta)\mu(dz)$$

is DR and a second application of Theorem 3.1 yields that

$$\psi(\theta+\xi) = \int K_1(\theta,u)f(u \mid \xi)\mu(du)$$

is also DR. Thus by Proposition 3.1, ψ is Schur-convex. ∎

The argument used to prove Proposition 3.2 is also valid for the Gamma shape-parameter family. Discussions of this and many other interesting cases can be found in Proschan and Sethuraman (1977), Marshall and Olkin (1979) and Eaton (1982). Extensions and application of these ideas for groups other than $\mathcal{P}_n$ are discussed in Eaton and Perlman (1977a) and in Eaton (1984, 1987a, 1987b, 1988). Extensions in another direction appear in Karlin and Rinott (1988).

An important aspect of the proof of Proposition 3.2 is the convolution property of the Poisson family. Without this, the conclusion of Proposition 3.2 may fail even though the density is DR.

EXAMPLE 3.2. Let X_1 and X_2 be independent exponential random variables with densities

$$p(x \mid \theta_i) = \begin{cases} \frac{1}{\theta_i}e^{-x/\theta_i}, & x>0\\ 0, & x \le 0\end{cases}$$

where $\theta_i > 0$, $i=1,2$. Then it is easy to show that

$$f(x \mid \theta) \equiv p(x_1 \mid \theta_1)p(x_2 \mid \theta_2)$$

is DR on $\mathcal{X}\times\mathcal{X}$, where $\mathcal{X} = \{(x_1,x_2) \mid x_1>0, x_2>0\}$. Let $\mathcal{A} = \{x \in \mathbb{R}^2 \mid x_1+x_2 \le 1\}$. Then $h(x) \equiv I_{\mathcal{A}}(x)$ is both Schur-convex and Schur-concave on

$\mathcal{X}$, but

$$\psi(\theta) \equiv \int_{\mathcal{X}} h(x) f(x \mid \theta) dx$$

is neither Schur-convex nor Schur-concave. However, a change of variables to $\eta_i = \log \theta_i$, $i = 1,2$, yields a function which is Schur-concave, i.e., $\psi_1(\eta) \equiv \psi(e^{\eta_1}, e^{\eta_2})$ is Schur-concave (cf. Marshall and Olkin (1979)). This example raises many questions concerning choice of coordinate systems and the validity of Proposition 3.2 for non-convolution families. This area of investigation appears not to have been explored.

Both Proposition 3.1 and Theorem 3.1 have extentions to reflection groups other than $\mathcal{P}_n$. These extensions are discussed in Chapter 6 of Eaton (1987). However, applications of these ideas to interesting statistical problems have been far less numerous than in the case of the permutation group. From the point of view of applications, a most intriguing problem is that of providing general conditions under which ψ in (3.1) is Schur-convex (concave) whenever h is Schur-convex (concave). Example 3.2 shows that f being a DR kernel (for the group $\mathcal{P}_n$) is not sufficient, but the DR assumption is perhaps a good starting place. The work of Rüschendorf (1981) provides some general theory, but seems not to be directly applicable to this problem.

4. Multivariate Concentration Inequalities: the Gaussian Case. Suppose that $X \sim N_n(0, \Sigma)$, the n-variate normal ($\equiv$ Gaussian) distribution with mean vector 0 and covariance matrix Σ. Write $\Sigma_2 \geq \Sigma_1$ to denote that $\Sigma_2 - \Sigma_1$ is positive semidefinite. For any subset $C \subseteq \mathbb{R}^n$ let

$$P_\Sigma(C) = \text{Prob}(X \in C).$$

We seek to determine classes $\mathcal{C}$ of subsets of $\mathbb{R}^n$ such that $\Sigma_2 \geq \Sigma_1$ implies that P_{Σ_1} is *more concentrated* than P_{Σ_2} relative to $\mathcal{C}$, i.e., such that

$$\Sigma_2 \geq \Sigma_1 \Rightarrow P_{\Sigma_2}(C) \leq P_{\Sigma_1}(C) \quad \forall C \in \mathcal{C}. \tag{4.1}$$

We shall study group symmetry (= invariance) conditions on C and on Σ_1, Σ_2, as well as convexity or monotonicity conditions on C, under which (4.1) is valid.

Let $\mathcal{C}_1$ denote the class of all convex, centrally symmetric (i.e., $-C = C$) subsets C of $\mathbb{R}^n$. Anderson (1955) showed that (4.1) is valid for $\mathcal{C} = \mathcal{C}_1$ with no additional symmetry conditions on Σ_1 or Σ_2. His approach uses the convolution (= additive) property of the multivariate normal distribution and a conditioning argument.

THEOREM 4.1 Anderson (1955)). *When $\mathcal{C} = \mathcal{C}_1$, (4.1) is valid.*

PROOF: Let $\Delta = \Sigma_2 - \Sigma_1$ and $X \sim N_n(0, \Sigma_2)$. Then we have the representation

$$X \overset{d}{=} Y + Z \tag{4.2}$$

where $Y \sim N_n(0, \Sigma_1)$, $Z \sim N_n(0, \Delta)$, and Y is independent of Z. Thus

$$\begin{aligned} P_{\Sigma_2}(C) &= E_\Delta\{P_{\Sigma_1}[Y \in C - Z \mid Z]\} \\ &\leq E_\Delta\{P_{\Sigma_1}[Y \in C]\} \\ &= P_{\Sigma_1}(C) \end{aligned} \tag{4.3}$$

by Theorem 2.1, Anderson's convolution theorem, since the multivariate normal density is symmetric and unimodal. Thus (3.1) is established. ∎

We shall now show that results complementing Theorem 4.1 may be obtained by this argument provided that Anderson's convolution theorem is replaced by those of Mudholkar (1966) or Eaton and Perlman (1977a). In each case the symmetry condition $-C = C$ is replaced by the invariance of C under a group G of $n \times n$ orthogonal matrices as in Section 2. Corresponding invariance conditions on Σ_1 (but not Σ_2!) are also required. When G is a *reflection group*, convexity of C can be replaced by a weaker monotonicity condition.

As in Section 2, let G be a finite (or, more generally, compact) subgroup of $\mathcal{O}_n$, the group of all $n \times n$ orthogonal matrices acting on $\mathbb{R}^n$. Let $\mathcal{M}_G$ denote the class of all G-decreasing subsets of $\mathbb{R}^n$ (cf. Definition 2.1), and let $\mathcal{C}_G$ denote the class of all convex and *G-invariant* subsets C, i.e., $gC = C$ $\forall g \in G$. Note that $\mathcal{M}_G \supset \mathcal{C}_G$: the inclusion is clear, while strict inclusion follows from the fact that $\mathcal{M}_G$ is closed under unions while $\mathcal{C}_G$ is not. Every $C \in \mathcal{M}_G$ is necessarily G-invariant. If $G = \{\pm I\}$, where I denotes the $n \times n$ identity matrix, then $\mathcal{C}_G = \mathcal{C}_1$, while $\mathcal{M}_G \equiv \mathcal{M}_1$ is the class of all centrally symmetric sets that are star-shaped with respect to the origin in $\mathbb{R}^n$.

DEFINITION 4.1. A real $n \times n$ symmetric matrix Σ is *G-invariant* if $g\Sigma g' = \Sigma$ $\forall g \in G$. The class of all positive definite G-invariant $n \times n$ matrices is denoted by $\mathcal{S}_G^+$.

In multivariate statistical analysis, $\mathcal{S}_G^+$ is called the *group symmetry covariance model* determined by G (cf. Andersson (1975), Eaton (1983), Perlman (1987)). If $G = \{I\}$ or $\{\pm I\}$ then $\mathcal{S}_G^+ = \mathcal{S}_n^+$, the class of *all* $n \times n$ positive definite symmetric matrices. If $G = \mathcal{D}_n$ then $\mathcal{S}_G^+$ is the class of all *diagonal* matrices $\Sigma = \mathrm{Diag}(\sigma_{11}, \ldots, \sigma_{nn})$ with each $\sigma_{ii} > 0$. If $G = \mathcal{P}_n$ then $\mathcal{S}_G^+$ is the set of all positive definite matrices $\Sigma \equiv (\sigma_{ij})$ having intraclass structure, i.e., $\sigma_{ii} = a$ $\forall i, \sigma_{ij} = b$ $\forall i \neq j$. It is a consequence of Schur's Lemma that if G acts irreducibly on $\mathbb{R}^n$ then $\mathcal{S}_G^+ = \{\lambda I \mid \lambda > 0\}$.

DEFINITION 4.2. The group G acts *effectively* on $\mathbb{R}^n$ if $0 \prec_G x$ $\forall x \in \mathbb{R}^n$, i.e., if 0 is the minimal element in $\mathbb{R}^n$ under $\prec_G$.

It may be seen that $G = \{\pm I\}$ acts effectively on $\mathbb{R}^n$, the group $\mathcal{D}_n$ of all sign changes of coordinates acts effectively on $\mathbb{R}^n$, every cyclic rotation group acts effectively on $\mathbb{R}^2$, while the group $\mathcal{P}_n$ of all $n \times n$ permutation matrices does not act effectively on $\mathbb{R}^n$.

We resume our discussion of concentration inequalities. As above, G denotes a finite (or compact) subgroup of $\mathcal{O}_n$. The next result contains Theorem 4.1 as a special case (set $G = \{\pm I\}$):

THEOREM 4.2. *When $\mathcal{C} = \mathcal{C}_G$, (4.1) is valid provided that Σ_1 is G-invariant and G acts effectively on $\mathbb{R}^n$.*

PROOF: The proof of Theorem 4.1 remains applicable here except that Anderson's Theorem 2.1 must be replaced by Mudholkar's Theorem 2.2 in order to obtain the inequality in (4.3). Here are the details. The assumption that Σ_1 is G-invariant implies that the probability density function of $Y \sim N_n(0, \Sigma_1)$ is G-invariant, and it is clearly unimodal. Since $C \in \mathcal{C}_G$ is convex and G-invariant, Theorem 2.2 shows that

$$f(z) \equiv P_{\Sigma_1}[Y \in C - Z \mid Z = z] \tag{4.4}$$

is a G-decreasing function of z. Hence the assumption that G acts effectively on $\mathbb{R}^n$ implies that $f(z) \le f(0)\ \forall z \in \mathbb{R}^n$ so the inequality in (4.3) holds. ∎

If G' is a subgroup of G that acts effectively on $\mathbb{R}^n$ then so does G, and clearly $\mathcal{C}_G \subseteq \mathcal{C}_{G'}$, so Theorem 4.2 is strongest when G is replaced by its smallest effective subgroup. In fact, if $-I \in G$ then $\mathcal{C}_G \subseteq \mathcal{C}_1$, so the conclusion of Theorem 4.2 is implied by Theorem 4.1 *without* the additional assumptions on Σ_1 and G. However, interesting examples of groups G that do not contain $-I$ but which act effectively on $\mathbb{R}^n$ are readily found, in which cases the additional assumptions appear to be required. For example, let G be any cyclic rotation group of *odd* order acting on $\mathbb{R}^2$. Such G acts irreducibly on $\mathbb{R}^2$, hence Σ_1 is G-invariant iff $\Sigma_1 = \lambda I$. Thus if C is a regular k-gon in $\mathbb{R}^2$ centered at 0 with k *odd*, then Theorem 4.2 implies that $P_{\Sigma_2}(C) \le P_{\Sigma_1}(C)$ whenever $\Sigma_2 \ge \Sigma_1 = \lambda I$. If k is even, however, then Theorem 4.1 applies *without* the restriction that $\Sigma_1 = \lambda I$. [This suggests, in fact, that the conclusion of Theorem 4.2 might remain valid under weaker hypotheses.]

Theorems 4.1 and 4.2 concern concentration inequalities for the classes $\mathcal{C}_1$ and $\mathcal{C}_G$, which consist of *convex* sets C. Theorem 4.3 below presents a concentration inequality for the class $\mathcal{M}_G$ of *G-decreasing* sets C, which need not be convex, when G is both effective and a *reflection group*, i.e., is generated by simple reflections in $\mathbb{R}^n$ (cf. Eaton and Perlman (1977a), Section 3). The *dihedral group* of all rotations and reflections that leave invariant a regular k-gon in $\mathbb{R}^2 (k \ge 2)$ is a reflection group of order $2k$, while its cyclic subgroup of rotations only (order k) is not a reflection group. Both $\mathcal{D}_n$ and P_n are reflection groups; however, only $\mathcal{D}_n$ acts effectively on $\mathbb{R}^n$. The first is generated by the simple reflections that change the sign of a single coordinate, while the second is generated by the elementary transpositions of two coordinates. The group $G = \{\pm I\}$ is not a reflection group in $\mathbb{R}^n$, $n \geq 2$.

THEOREM 4.3. *When $C = \mathcal{M}_G$, (4.1) is valid provided that G is a reflection group, Σ_1 is G-invariant, and G acts effectively on $\mathbb{R}^n$.*

PROOF: Proceed as in the proof of Theorem 4.1, except now apply Theorem 2.4, the convolution theorem of Eaton and Perlman (1977a), to obtain that f in (4.4) is a G-decreasing function. ∎

Note that Theorem 4.3 remains valid if it is only assumed that G *contains* a subgroup G' that is a reflection group and that acts effectively on $\mathbb{R}^n$, since $\mathcal{M}_G \subseteq \mathcal{M}_{G'}$.

If $G = \mathcal{D}_n$ then G is an effective reflection group acting on $\mathbb{R}^n$, so Theorem 4.3 applies. Here Σ_1 is G-invariant iff Σ_1 is diagonal. Some examples of $\mathcal{D}_n$-decreasing sets may be constructed as follows. For positive numbers $y_1, \ldots, y_n$ and $-\infty \leq r \leq \infty$ the r-th mean of $y_1, \ldots, y_n$ is defined by continuity as

$$m_r(y_1, \ldots y_n) = \left[\frac{1}{n}\sum_{i=1}^{n} y_i^r\right]^{1/r}, r \neq 0, \pm\infty,$$
$$m_\infty(y_1, \ldots, y_n) = \max(y_1, \ldots, y_n)$$
$$m_0(y_1, \ldots, y_n) = \left(\prod_{i=1}^{n} y_i\right)^{1/n}$$
$$m_{-\infty}(y_1, \ldots, y_n) = \min(y_1, \ldots, y_n).$$

Select positive numbers $\alpha_1, \ldots, \alpha_n, k$, and then define

$$C_r = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid m_r(|x_1|^{\alpha_1}, \ldots, |x_n|^{\alpha_n}) \leq k\}.$$

Then $C_r \in \mathcal{C}_{\mathcal{D}_n}$ if $r \geq 1$ while $C_r \in \mathcal{M}_{\mathcal{D}_n} \backslash \mathcal{C}_{\mathcal{D}_n}$ if $r < 1$, so in the latter case C_r is a non-convex but $\mathcal{D}_n$-decreasing set for which (4.1) holds whenever $\Sigma_2 \geq \Sigma_1 =$ Diagonal.

5. Multivariate Concentration Inequalities: the Elliptically Contoured Case. It is evident that the concentration inequalities in Theorems 4.1, 4.2, and 4.3 remain valid when P_Σ is taken to be a scale mixture of Gaussian distributions on $\mathbb{R}^n$ with mean 0 and covariance matrix Σ, e.g., a multivariate Student $-t$ distribution. Like the Gaussian distribution itself, such scale mixtures are both unimodal and elliptically contoured. It is perhaps surprising, however, that at least the first of these theorems remain valid for *all* elliptically contoured distributions without any assumption of unimodality.

DEFINITION 5.1. The random vector X has an *elliptically contoured* pdf on $\mathbb{R}^n$ if its pdf has the form

$$|\Sigma|^{-1/2} g(x'\Sigma^{-1}x), \tag{5.1}$$

where Σ is an $n \times n$ positive definite matrix. In this case we write $X \sim EC_n(\Sigma)$.

Clearly the Gaussian distribution $N_n(0,\Sigma)$ is a special case of (5.1). Fefferman, Jodeit, and Perlman (1972) substantially strengthened Theorem 4.1 by extending it to the elliptically contoured case. (See also Das Gupta *et al.* (1972), Theorem 3.3.) Surprisingly, the proof of Theorem 5.1, like that of Theorem 4.1, is based on Anderson's convolution theorem, Theorem 2.1.

THEOREM 5.1 (Fefferman, Jodeit, and Perlman (1972)). *Suppose that* $X \sim EC_n(\Sigma)$. *Then for every* $C \in \mathcal{C}_1$,

$$\Sigma_2 \geq \Sigma_1 \Rightarrow P_{\Sigma_2}(C) \leq P_{\Sigma_1}(C). \tag{5.2}$$

PROOF (sketch): The second inequality in (5.2) is equivalent to the inequality

$$P_I(D\tilde{C}) \leq P_I(\tilde{C}), \tag{5.3}$$

where I is the $n \times n$ identity matrix, $\tilde{C}$ is the image of C under an appropriate linear transformation (hence $\tilde{C} \in \mathcal{C}_1$), and $D = \mathrm{Diag}(d_1, \ldots, d_n)$ with $d_1^2, \ldots, d_n^2$ the characteristic roots of $\Sigma_1\Sigma_2^{-1}$ (hence $0 < d_i \leq 1$). Since the distribution $EC_n(\Sigma)$ is spherically symmetric when $\Sigma = I$, the conditional distribution of X given $|X| = r$ is uniform on the sphere S_r of radius r, so (5.3) will follow from the stronger inequality

$$\nu_r(D\tilde{C}) \leq v_r(\tilde{C}), \tag{5.4}$$

where ν_r denotes uniform surface measure on S_r. (Note, however, that $D\tilde{C} \cap S_r$ is not necessarily contained in $\tilde{C} \cap S_r$.) By the Divergence Theorem, however, it may be shown that for suitably smooth C,

$$\frac{\partial}{\partial d_i}[\nu_r(D\tilde{C})] = -\alpha \frac{d^2}{d\beta^2}[(I_{B_r} * I_{D\tilde{C}})(\beta e_i)]_{\beta=0}, \tag{5.5}$$

where α is a positive constant, B_r is the ball of radius r ($S_r = \partial B_r$), and e_i is the unit vector with ith component 1 and all other components 0. Because both B_r and $D\tilde{C} \in \mathcal{C}_1$, Theorem 2.1 implies that $I_{B_r} * I_{D\tilde{C}}$ is ray decreasing, hence the second derivative on the right side of (5.5) is non-positive. Therefore

$\nu_r(D\tilde{C})$ is nondecreasing in each d_i, $1 \leq i \leq n$, which establishes (5.4) and hence the result. ∎

It was seen in Section 4 that in the Gaussian case, the method of proof of Theorem 4.1 could be applied to establish Theorems 4.2 and 4.3 provided that Theorem 2.1 was replaced by Theorems 2.2 or 2.4. Unfortunately this is not so in the elliptically contoured case. The proof of Theorem 5.1 invokes Anderson's Theorem to show that the convolution

$$(I_{B_r} * I_{D\tilde{C}})(y) \equiv \mu[B_r \cap (D\tilde{C} + y)] \tag{5.6}$$

has a (global) maximum at $y = 0$ when C (and hence $D\tilde{C}$) $\in \mathcal{C}_1$, where μ denotes Lebesgue measure. In order to use this argument to extend Theorems 4.2 and 4.3 to the elliptically contoured case, it would be necessary to show that (5.6) has a (local) maximum at $y = 0$ when $C \in \mathcal{C}_G$ or $C \in \mathcal{M}_G$, where the group G is as in Section 4. Unfortunately, however, the transformation $C \to D\tilde{C}$ may not preserve G-invariance, so Theorems 4.2 and 4.3 cannot be applied. Since only the existence of a *local* maximum at $y = 0$ is required, however, this method of proof may succeed in extending Theorems 4.2 and 4.3, provided that suitable local versions of Theorems 2.2 and 2.4 can be found. Note too that one of the sets in (5.6), namely B_r, is a ball, so the full generality of Theorems 2.2 and 2.4 would not be needed.

REFERENCES

ANDERSON, T. W. (1958). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170–176.

ANDERSSON, S. A. (1975). Invariant normal models. *Ann. Statist.* **3**, 132–154.

BORELL, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6**, 111–136.

BRASCAMP, H. J., and LIEB, E.H. (1974). A logarithmic concavity theorem with some applications. Unpublished manuscript.

DAS GUPTA, S. (1980). Brunn-Minkowski inequality and its aftermath. *J. Multivariate Anal.* **10**, 296–318.

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. Probab. Statist. Vol. II,* 241–265.

DAVIDOVIC, JU. S., KORENBLJUM, B. I., and HACET, B. I. (1969). A property of logarithmically concave functions. *Soviet Math. Dokl.* **10**, 477–480.

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, New York.

EATON, M. L. (1967). Some optimum properties of ranking procedures. *Ann. Math. Statist.* **38**, 124–137.

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* **10**, 11–43.

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* John Wiley and Sons, New York.

EATON, M. L. (1984). On group induced orderings, monotone functions, and convolution theorems. In *Inequalities in Statistics and Probability* (Y. L. Tong, ed.), 13–25. Institute of Mathematical Statistics Lecture Notes-Monograph Series, Hayward, CA.

EATON, M. L. (1987a). *Lectures on Topics in Probability Inequalities.* Centrum voor Wiskunde en Informatica, Amsterdam.

EATON, M. L. (1987b). Group induced orderings with some applications in statistics. *CWI Newslett.*, Issue **16**, 3–31.

EATON, M. L. (1988). Concentration inequalities for Gauss-Markov estimators. *J. Multivariate Anal.* **25**, 119–137.

EATON, M. L. and PERLMAN, M. D. (1977a). Reflection groups, generalized Schur functions, and the geometry of majorization. *Ann. Probab.* **5**, 829–860.

EATON, M. L. and PERLMAN M. D. (1977b). Generating $O_{(n)}$ with reflections. *Pacific J. Math.* **73**, 73–80.

FEFFERMAN, C., JODEIT, M., and PERLMAN, M. D. (1972). A spherical surface measure inequality for convex sets. *Proc. Amer. Math. Soc.* **33**, 114–119.

GROVE, L. C. and BENSON, C. T. (1985). *Finite Reflection Groups* (2nd ed.). Springer-Verlag, New York.

HOLLANDER, M., PROSCHAN, F., and SETHURAMAN, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* **5**, 722–733.

KARLIN, S. and RINOTT, Y. (1988). A generalized Cauchy-Binet formula and applications to total positivity and majorization. *J. Multivariate Anal.* **27**, 284–299.

MARSHALL, A. W., and OLKIN, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* **2**, 1189–1200.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

MUDHOLKAR, G. (1966). The integral of an invariant unimodal function over an invariant convex set – an inequality and applications. *Proc. Amer. Math. Soc.* **17**, 1327–1333.

PERLMAN, M. D. (1987). Group symmetry covariance models. *Statist. Sci.* **2**, 421–425.

PERLMAN, M. D. (1990). Anderson's theorem on the integral of a symmetric unimodal function over a symmetric convex set, and its applications to probability and statistics. In *The Collected Papers of T. W. Anderson: 1943–1985* (G. Styan, ed.). John Wiley and Sons, New York.

PRÉKOPA, A. (1973). On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)* **34**, 335-343.

PROSCHAN, F. and SETHURAMAN, J. (1977). Schur functions in statistics, I: the preservation theorem. *Ann. Statist.* **5**, 256–262.

RINOTT, Y. (1973). Multivariate majorization and rearrangement inequalities with applications to probability and statistics. *Israel J. Math.* **15**, 60–77.

RÜSCHENDORF, L. (1981). Ordering of distributions and rearrangement of functions. *Ann. Probab.* **9**, 276–283.

SAVAGE, I. R. (1957). Contributions to the theory of rank order statistics – the "trend" case. *Ann. Math. Statist.* **28**, 968–977.

SHERMAN, S. (1955). A theorem on convex sets with applications. *Ann. Math. Statist.* **26**, 763–766.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic Press, New York.

WINTNER, A. (1938). *Asymptotic Distributions and Infinite Convolutions.* Edwards Brothers, Ann Arbor, Michigan.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
270 VINCENT HALL
206 CHURCH STREET, SE
MINNEAPOLIS, MINNESOTA 55455

DEPARTMENT OF STATISTICS, GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195