The Annals of Statistics
1982, Vol. 10, No. 1, 11–43

SPECIAL INVITED PAPER

A REVIEW OF SELECTED TOPICS IN MULTIVARIATE PROBABILITY INEQUALITIES[1]

BY MORRIS L. EATON

University of Minnesota

This paper contains a review of certain multivariate probability inequalities. The inequalities discussed include the FKG inequalities and the related association inequalities, inequalities resulting from Schur convexity and its extension to reflection groups, and inequalities for probabilities of certain convex symmetric sets.

1. Overview. This paper has been written with the following dual purpose: (i) to provide the uninitiated reader with enough history and motivation to appreciate some current formulations and descriptions of several classes of multivariate probability inequalities, and (ii) to provide the more experienced reader with a review of some research techniques and problems of current interest.

In particular, we have chosen three areas around which to center our discussion. The description of each of these areas begins with a concrete problem which is used to motivate more abstract problems.

Once the three general problems have been described in Section 2, the technical discussion is given in Sections 3, 4 and 5. Section 6 contains general structural information concerning some of the particular results in the previous three sections.

2. Introduction. In this section we describe three problems which are treated more completely in later sections. The set of probability measures on measurable space $(\mathscr{X}, \mathscr{B})$ is denoted by $\mathscr{M}$, or sometimes $\mathscr{M}(\mathscr{X})$. In most cases $\mathscr{X}$ is a Borel subset of a Euclidean space, R^n, and in this case, elements of $\mathscr{X}$ are represented as column vectors with x' denoting the transpose of x.

DEFINITION 2.1. A *pre-order* on $\mathscr{X}$, $\leq$, is a relation defined on $\mathscr{X} \times \mathscr{X}$ such that (i) $x \leq x$, (ii) $x \leq y$ and $y \leq z$ implies that $x \leq z$ (transitivity). If, in addition $x \leq y$ and $y \leq x$ implies that $x = y$, then $\leq$ is a *partial order*. Both pre-orders and partial orders arise naturally below.

DEFINITION 2.2 If $\leq$ is a pre-order on $\mathscr{X}$, a real valued function is *increasing* (*decreasing*) if $x \leq y$ implies $f(x) \leq f(y)$ ($f(x) \geq f(y)$). A set $B \subseteq \mathscr{X}$ is increasing (decreasing) if the indicator function of B, denoted by I_B, is increasing (decreasing).

A common description of a multivariate probability inequality is:

$$P_1(A) \leq P_2(A), \qquad A \in \mathscr{A} \tag{2.1}$$

where $P_i \in \mathscr{M}(X)$, $i = 1, 2$, and $\mathscr{A}$ is some interesting class of sets; for example, $\mathscr{A}$ might be

Received October 1980; revised September 1981.

[1] This paper is based on a Special Invited Lecture presented at the Western Regional Meeting of the IMS held at Stanford University, June, 1977. This work was supported in part by National Science Foundation Grant MCS77-25112.

AMS 1970 *subject classifications* Primary 62H99; secondary 62E99.

Key words and phrases. multivariate probability inequalities, partial orderings, FKG inequality, associated random variables, majorization, reflection groups, elliptically contoured distributions.

the convex symmetric sets in a vector space, or perhaps a class of sets which are increasing relative to a pre-order. The above inequality is

$$\int I_A(x)P_1(dx) \le \int I_A(x)P_2(dx), \qquad A \in \mathscr{A}. \tag{2.2}$$

Now, let $\mathscr{F}$ be the class of functions which can be written in the form

$$f(x) = \sum_1^m a_i I_{A_i}(x) \tag{2.3}$$

where m is some positive integer, $A_i \in \mathscr{A}$ and $a_i \ge 0$, $i = 1, \cdots, m$. Then $\mathscr{F}$ is a convex cone and each $f \in \mathscr{F}$ is non-negative and bounded. Clearly (2.2) is equivalent to

$$\int f(x)P_1(dx) \le \int f(x)P_2(dx), \qquad f \in \mathscr{F}. \tag{2.4}$$

Many of the problems in this paper can be described as "Given $\mathscr{F}$, find conditions on P_1 and P_2 so that (2.4) holds."

REMARK 2.1. In many cases, inequalities of the form (2.4) are valid without the non-negativity or boundedness assumption on the elements of $\mathscr{F}$. The proof for these extensions is ordinarily a matter of truncating, translating, and taking limits. These details will be omitted.

Here is the first class of problems.

PROBLEM 2.1. Consider a two dimensional random vector X with coordinates X_1 and X_2. An intuitively appealing condition which expresses "positive dependence" between X_1 and X_2 is

$$P(X_1 \ge c_1, X_2 \ge c_2) \ge P(X_1 \ge c_1)P(X_2 \ge c_2) \tag{2.5}$$

for all c_1 and c_2; or, equivalently,

$$P(X_1 \ge c_1 \mid X_2 \ge c_2) \ge P(X_1 \ge c_1) \tag{2.6}$$

for all c_1 and c_2. Such variables were called "positively quadrant dependent" by Lehmann (1966). One of the most useful sufficient conditions which implies (2.5) is provided by the notion of association introduced in Esary et al. (1967).

DEFINITION 2.3. The coordinates of a random vector $Z \in R^n$ are *associated* if for each pair of non-negative bounded functions f and g on R^n to R^1 which are increasing in each coordinate variable (with the remaining ones held fixed), we have the inequality

$$\operatorname{Cov}\{f(Z), g(Z)\} \ge 0. \tag{2.7}$$

When $n = 2$, if X_1 and X_2 are associated, if f is the indicator of $[c_1, \infty)$, and if g is the indicator of $[c_2, \infty)$, then (2.7) implies (2.5). More generally, if Z is associated (i.e., the coordinates of Z are associated), then the inequality

$$P(Z_i \ge c_i, i = 1, \cdots, n) \ge \prod_{i=1}^n P(Z_i \ge c_i)$$

holds for all $c_1, \cdots, c_n$; see Section 3 for a discussion. Thus sufficient conditions that Z be associated would be useful.

For any two vectors $x, y \in R^n$, write $x \le y$ if $x_i \le y_i$ for $i = 1, \cdots, n$ so $\le$ is a partial order on R^n, and is usually called the coordinatewise partial order. Let $\mathscr{F}$ be the convex cone of non-negative bounded increasing (with respect to $\le$) functions on R^n. The class of functions for which (2.7) must hold in order that Z be associated is just the class $\mathscr{F}$. Hence Z is associated iff

$$\operatorname{Cov}\{f(Z), g(Z)\} \ge 0, \qquad f, g \in \mathscr{F}. \tag{2.8}$$

If P is the probability measure of Z, write $P(dx) = p(x)\ \mu(dx)$ where μ is some σ-finite measure on R^n and p is the density of P. Then (2.8) can be written

$$\int f(x)g(x)p(x)\mu(dx) \geq \int f(x)p(x)\mu(dx) \int g(x)p(x)\mu(dx) \tag{2.9}$$

for $f, g \in \mathscr{F}$. If $\delta = \int fpd\mu = 0$, (2.9) obviously holds. If $\delta > 0$, set $p_1(x) = p(x)$ and $p_2(x) = \delta^{-1}f(x)p(x)$ so p_2 is a density and (2.9) becomes

$$\int g(x)p_1(x)\mu(dx) \leq \int g(x)p_2(x)\mu(dx) \tag{2.10}$$

for $g \in \mathscr{F}$ and $f \in \mathscr{F}$. Thus, sufficient conditions for (2.10) could be useful for providing sufficient conditions for (2.9). Indeed, a main result in Section 3 provides conditions for (2.10) and hence for (2.9) to hold. To describe this condition, let $P_i(dx) = p_i(x)\mu(dx)$ so (2.10) is

$$\int g(x)P_1(dx) \leq \int g(x)P_2(dx), \qquad g \in \mathscr{F}. \tag{2.11}$$

which is inequality (2.4). When $n = 1$, then (2.11) means that the distribution P_1 is stochastically smaller than P_2. In this case, a sufficient condition for (2.11) is that $p_2(x)/p_1(x)$ be non-decreasing in x; that is, the pair (p_1, p_2) has a monotone likelihood ratio (MLR) (see Karlin and Rubin, 1956). An equivalent way to express the MLR property on R^1 is

$$p_1(x)p_2(y) \leq p_1(x \wedge y)p_2(x \vee y), \qquad x, y \in R^1,$$

where $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. This form of the definition MLR can now be expressed for densities defined on R^n. For x and $y \in R^n$, we define $x \vee y$ and $x \wedge y$ to be the vectors in R^n whose coordinates are respectively, $\max\{x_i, y_i\}$, $i = 1, \cdots, n$; and $\min\{x_i, y_i\}$, $i = 1, \cdots, n$. It will be shown in Section 3 that when μ is a product measure on R^n, if p_1 and p_2 satisfy

$$p_1(x)p_2(y) \leq p_1(x \wedge y)p_2(x \vee y), \qquad x, y \in R^n, \tag{2.12}$$

then (2.11) holds (see Holley, 1974; Preston, 1974; and Kemperman, 1977). This result will then provide useful conditions for (2.8) and (2.9) to hold. Inequality (2.9) is often called the FKG-inequality because of the work of Fortuin, Ginibre and Kasteleyn (1971). Our discussion of Problem 2.1 is complete.

The inequality (2.11) suggests that the distribution P_2 has more mass toward the "upper right hand corner" of R^n than does P_1. Our next problem involves the idea of concentration of a distribution about $0 \in R^n$.

PROBLEM 2.2. A set $A \subseteq R^n$ is *symmetric* if $A = -A$. If A is both convex and symmetric, then certainly 0 is the center of A, and such sets represent a possible generalization of intervals symmetric about 0 in R^1. With this in mind, we say that P_1 is *more concentrated about* 0 than P_2 if $P_1(C) \geq P_2(C)$ for all convex symmetric sets $C \subseteq R^n$. This relation will be written $P_1 <_\circ P_2$. If $\mathscr{F}$ is the convex cone generated by the indicators of convex symmetric sets, then $P_1 <_\circ P_2$ iff

$$\int f(x)P_1(dx) \geq \int f(x)P_2(dx), \qquad f \in \mathscr{F}. \tag{2.13}$$

Now, the problem is to give some useful conditions so that (2.13) holds. A reasonable starting place is to assume P_1 is the uniform distribution on a given bounded convex symmetric set of positive Lebesgue measure - say C_1 with $\ell(C_1) > 0$. Then P_1 has a density with respect to ℓ given by

$$p_1(x) = \begin{cases} \{\ell(C_1)\}^{-1} & \text{if } x \in C_1 \\ 0 & \text{if } x \notin C_1. \end{cases}$$

Suppose that P_2 is the uniform distribution on a translate $C_1 + x_0$ of C_1, so P_2 has density $p_1(x - x_0)$. That $P_1 <_\circ P_2$ seems plausible in this case and a verification of this entails showing that

$$\ell(C \cap C_1) \geq \ell(C \cap (C_1 + x_0)) \tag{2.14}$$

for each convex symmetric set C. The inequality (2.14) is a consequence of Anderson's Theorem (1955) which, among other things, asserts that for convex symmetric sets C and C_1, and $x_0 \in R^n$, the function

$$\Psi(\alpha) = \ell(C \cap (C_1 + \alpha x_0)), \qquad \alpha \in R^1 \tag{2.15}$$

satisfies $\Psi(\alpha) = \Psi(-\alpha)$ and Ψ is decreasing on $[0, \infty)$. The usual statement of Anderson's Theorem is

THEOREM (Anderson, 1955). *Suppose f is a density on R^n such that $f(x) = f(-x)$ and $\{x \mid f(x) \geq v\}$ is convex for each $v > 0$. If C is a convex symmetric set and $\theta \in R^n$ then*

$$\Psi(\alpha) = \int_C f(x - \alpha\theta)\, dx, \qquad \alpha \in R^1 \tag{2.16}$$

satisfies $\Psi(\alpha) = \Psi(-\alpha)$ and Ψ is decreasing on $[0, \infty)$.

It is this result and a variety of associated ideas and results that constitute the first portion of Section 5.

Although Anderson's result involves a translation parameter, the result also has a variety of consequences for probability inequalities involving covariance matrices. For example, if P_i is the normal distribution $N(0, \Sigma_i)$ on R^n with mean 0 and covariance matrix Σ_i, $i = 1, 2$, and if $\Sigma_2 - \Sigma_1 = \Delta$ is positive semi-definite, then it seems plausible that $P_1 <_\circ P_2$. That this holds can be proved by observing that $\Sigma_2 = \Sigma_1 + \Delta$ so P_2 is the convolution of P_1 with a $N(0, \Delta)$ distribution Q, say. Thus, for each convex symmetric set C,

$$P_2(C) = \int P_1(C - y)Q(dy) \leq P_1(C),$$

since (2.16) implies that $P_1(C - y) \leq P_1(C)$ for each y. A much more subtle application of Anderson's Theorem yields information about the behavior of certain probabilities as functions of correlations. In particular, if $X \in R^n$ is $N(0, \Sigma)$ on R^n, then

$$P(|X_i| \leq a_i, i = 1, \cdots, n) \geq \textstyle\prod_{i=1}^n P(|X_i| \leq a_i),$$

where $X_1, \cdots, X_n$ are the coordinates of X. This inequality was first established by Dunn (1958, 1959) for some special Σ's to provide a conservative solution to a confidence set problem. This and other inequalities involving correlations are the subject of the latter portion of Section 5.

The discussion of Section 4 has been postponed until now since Anderson's Theorem provides one possible way to motivate the material there. However, we will begin our discussion of Problem 2.3 with a problem concerning the number of empty cells in a multinomial distribution.

PROBLEM 2.3. Suppose $X \in R^k$ has a multinomial $\mathscr{M}(p, k, n)$ distribution so X has non-negative integer valued coordinates $X_1, \cdots, X_k$ such that $\Sigma X_i = n$. The probability vector $p \in R^k$ satisfies $0 \leq p_i \leq 1$ and $\Sigma p_i = 1$. Let Z be the number of X_i which are zero so Z is the number of empty cells in a multinomial, and the possible values of Z are 0, 1,

$\cdots, k-1$. For $m = 1, 2, \cdots k$, define probability vectors $p^{(m)}$ by

$$p_i^{(m)} = \begin{cases} \frac{1}{m}, & i = 1, \cdots, m \\ 0, & i = m+1, \cdots, k. \end{cases}$$

Thus, $p^{(k)}$ has all its coordinates equal to $1/k$. Intuitively, we expect Z to take on smaller values with higher probability the closer p is to $p^{(k)}$. One way to try to make this precise is to consider

$$\phi(p) = P_p(Z \geq a) \tag{2.17}$$

for a fixed number $a \in (0, k-1)$, and attempt to describe the behavior of ϕ as p varies. In particular, it seems reasonable to conjecture that $\phi(p^{(m)})$ is decreasing in m. More generally, it may be possible to define a pre-order on p's so that ϕ is monotone in this ordering. Although far from obvious, the order defined by majòrization, which we will now describe, is appropriate for the current problem. For $n \in R^k$, let $u_{(1)} \geq u_{(2)} \geq \cdots \geq u_{(k)}$ denote the ordered coordinates of u. For $y, z \in R^k$, write $y \leq z$ if

$$\textstyle\sum_1^m z_{(i)} \geq \sum_1^m y_{(i)}, \quad m = 1, \cdots, k-1, \quad \text{and} \quad \sum_1^k z_i = \sum_1^k y_i. \tag{2.18}$$

When $y \leq z$, we say *z majorizes y* or *y* is *majorized* by z. It was shown by Wong and Yue (1973) that ϕ is increasing in $\leq$ on probability vectors, that is, if $p \leq q$, then $\phi(p) \leq \phi(q)$. In particular,

$$p^{(k)} \leq p^{(k-1)} \leq \cdots \leq p^{(1)},$$

so $\phi(p^{(m)})$ is decreasing in m. The fact that ϕ is increasing can be expressed in another useful and interesting way. Let $P(\cdot \mid p)$ be the probability measure defined on R^k by the $\mathscr{M}(p, k, n)$ distribution. Thus, if B is a Borel set,

$$P(B \mid p) = \int_B h(x \mid p)\lambda(dx)$$

where λ is counting measure on the points in R^k which have integer coordinates and $h(\cdot \mid p)$ is the density of $P(\cdot \mid p)$ with respect to λ. If f is the indicator function of the set $\{Z \geq a\}$, then

$$\phi(p) = \int f(x)h(x \mid p)\lambda(dx). \tag{2.19}$$

Now, it is not hard to show that f is increasing in the majorization pre-ordering. Thus, it is plausible that the right hand side of (2.19) is increasing in p for *any* f which is increasing. Indeed this was proved by Rinott (1973) and will be one of the topics considered in Section 4.

REMARK 2.2. Functions which are increasing (decreasing) in the majorization pre-ordering are often called Schur convex (concave).

The above example is a special case of a class of problems to be treated in Section 4. To introduce these, we first reformulate Anderson's theorem. Let G_0 be the two element group consisting of I_k and $-I_k$ where I_k is the $k \times k$ identity matrix. For each $x \in R^k$, the convex set generated by $I_k x = x$ and $-I_k x = -x$, say $C(x)$, is just the line segment between x and $-x$. A function f on R^k which satisfies $f(x) = f(gx)$, $g \in G_0$, will be called G_0-invariant; that is, $f(x) = f(-x)$ for this particular group. Here is an alternative way to state Anderson's Theorem.

THEOREM. *Suppose f is a G_0-invariant density such that $\{x \mid f(x) \geq v\}$ is convex for* *each $v > 0$. Suppose C_0 is a convex set such that I_{C_0} is a G_0-invariant function; that is, $C_0 = -C_0$. Let*

$$\phi(\theta) = \int I_{C_0}(x) f(x-\theta)\,dx, \qquad \theta \in R^k. \tag{2.20}$$

Then ϕ is G_0-invariant and $\eta \in C(\theta)$ implies that $\phi(\eta) \geq \phi(\theta)$.

The conclusion here suggests we define a pre-order on R^k as follows:

$$\eta \leq \theta \quad \text{iff} \quad \eta \in C(\theta).$$

Thus, ϕ is G_0-invariant and decreasing in the pre-order $\leq$. The validity of the above result for other groups is the obvious question. It was answered by Mudholkar (1966). Let G be any subgroup of the group $k \times k$ orthogonal matrices. Typical groups of interest include

$\mathscr{P}_k$ – the group of all $k \times k$ permutation matrices
$\mathscr{D}_k$ – the matrix group of coordinate sign changes on R^k
$\mathscr{O}_k$ – the group of all $k \times k$ orthogonal matrices.

For each $y \in R^k$, let $C(y)$ be the convex set generated by $\{gy \mid g \in G\}$. Write $x \leq y$ if $x \in C(y)$. It is easy to verify that $\leq$ is a pre-order on R^k.

THEOREM (Mudholkar, 1966). *Suppose f is a G-invariant density such that $\{x \mid f(x) \geq v\}$ is convex for each $v > 0$. Suppose C_0 is a convex set such that I_{C_0} is a G-invariant function ($gC_0 = C_0$ for $g \in G$). Let*

$$\phi(\theta) = \int I_{C_0}(x) f(x-\theta)\,dx, \qquad \theta \in R^k.$$

Then ϕ is G-invariant and G-decreasing, that is, decreasing in the preorder defined by G.

REMARK 2.3. At this point the reader may want to draw some pictures. A simple example for which the geometry is easy is for $k = 2$, $G = \mathscr{D}_2$, $f(x) = (2\pi)^{-1}\exp(-\frac{1}{2}x'x)$ and $C_0 = \{x \mid \lambda_1 x_2^2 + \lambda_2 x_2^2 \leq 1\}$ where λ_1 and λ_2 are positive.

In the above theorem, the convexity assumptions coupled with the G-invariance assumptions imply that both f and I_{C_0} are decreasing. Note that any decreasing function must be G-invariant since $x \leq gx$ and $gx \leq x$ for all $x \in R^k$ and $g \in G$. This, together with Mudholkar's results, suggests the following question:

(Q.1) For which groups G is it true that if f_1 and f_2 are non-negative and G-decreasing then

$$\phi(\theta) = \int f_1(x) f_2(x-\theta)\,dx \tag{2.21}$$

is G-decreasing?

If f_2 is a density and f_1 is the indicator of a set C, then $\phi(\theta)$ is the probability of C computed when the value of the translation parameter is θ. A partial answer to (Q.1) is

THEOREM (Marshall and Olkin, 1974). *If $G = \mathscr{P}_k$, then ϕ is G-decreasing for each non-negative and G-decreasing f_1 and f_2.*

To connect the current discussion to the results of Wong and Yue (1973) and Rinott (1973), we first mention a result of Rado (1952) to the effect that $z \in R^k$ majorizes $y \in R^k$ iff y is in the convex hull of $\{gz \mid g \in \mathscr{P}_k\}$. In other words, the majorization pre-order defined earlier is the same as the pre-order defined by the group $\mathscr{P}_k$. One way to state the

Rinott (1973) result is:

> If $X \in R^k$ has a $\mathscr{M}(p, k, n)$ distribution and if f is $\mathscr{P}_k$-decreasing, then $\mathscr{E}_p f(X)$ is a $\mathscr{P}_k$-decreasing function of p.

Similarly, the Marshall-Olkin (1974) Theorem is:

> If $x \in R^k$ has a density $f_2(x - \theta)$ where f_2 is $\mathscr{P}_k$-decreasing and if f is $\mathscr{P}_k$-decreasing, then $\mathscr{E}_\theta f(X)$ is a $\mathscr{P}_k$-decreasing function of θ.

The parallel between the two results is now obvious. These results were extended to other parametric families, including the gamma and Dirichlet families, in Proschan and Sethuraman (1977) and Nevius, Proschan, and Sethuraman (1977).

For translation families, Eaton and Perlman (1977) extended the Marshall-Olkin Theorem to other groups, called reflection groups, which include $\mathscr{D}_k$ and the group generated by $\mathscr{P}_k \cup \mathscr{D}_k$. Based on this work, Conlon, León, Proschan and Sethuraman (1977a, 1977b) then extended the work in Proschan and Sethuraman (1977) and Nevius et al. (1977) to include the reflection group case. In Section 4, we describe the results concerning reflection groups together with some examples. The primary examples of reflection groups are $\mathscr{P}_k$, $\mathscr{D}_k$ and the group generated by $\mathscr{P}_k \cup \mathscr{D}_k$. This completes our discussion of Problem 2.3.

In Section 6, we discuss a general structure theorem given in Kamae et al. (1977) and Strassen (1965). In many cases, this result provides necessary and sufficient conditions so that inequality (2.4) holds. However, the conditions are of little practical use in deciding whether or not (2.4) holds for particular examples.

The recent book of Tong (1980) is a comprehensive general text on multivariate probability inequalities. A number of topics not discussed here can be found there, together with many applications and an extensive bibliography. The most important example of the results in Section 4 concern majorization and the permutation group. In fact, majorization has played an important role in the development of many probability inequalities. The recent volume by Marshall and Olkin (1979) is an encyclopedic work which, among other things, covers the theory of majorization and its applications to statistics, probability and many other branches of mathematics.

3. Inequalities on Coordinatewise Ordered Spaces. Throughout this section, the space $\mathscr{X}$ will be a product, $\mathscr{X} = \prod_1^n \mathscr{X}_i$ where each $\mathscr{X}_i$ is a Borel subset of R^1. For $x, y \in \mathscr{X}$, write $x \le y$ to mean $x_i \le y_i$ for $i = 1, \cdots, n$, so $\le$ is a partial order on $\mathscr{X}$. Let $\mathscr{F}$ be the convex cone of all non-negative bounded increasing functions on $\mathscr{X}$. given $P_1, P_2 \in \mathscr{M}$, we now consider Problem 2.1 described earlier—namely, find conditions on P_1 and P_2 so that

$$\int f(x) P_1(dx) \le \int f(x) P_2(x), \qquad f \in \mathscr{F}. \tag{3.1}$$

There are a couple of cases when (3.1) is easily verified.

PROPOSITION 3.1. *Take $\mathscr{X} = R^n$. For $P_0 \in \mathscr{M}$, and $\theta \in R^n$, define P_θ by $P_\theta(B) = P_0(B - \theta)$ for Borel sets B. If $\eta \le \theta$, then*

$$\int f(x) P_\eta(dx) \le \int f(x) P_\theta(dx), \qquad f \in \mathscr{F}. \tag{3.2}$$

PROOF. For any $\theta \in R^n$, $\int f(x)P_\theta(dx) = \int f(x + \theta)P_0(dx)$. Since $\eta \le \theta$, $x + \eta \le x + \theta$ for all $x \in R^n$ so $f(x + \eta) \le f(x + \theta)$ for all $f \in \mathscr{F}$. Thus, for $f \in \mathscr{F}$,

$$\int f(x)P_\eta(dx) = \int f(x + \eta)P_0(dx) \le \int f(x + \theta)P_0(dx) = \int f(x)P_\theta(dx),$$

which completes the proof.

The above result treats the case of a translational parameter family of distributions on R^n. There are other types of parametric families which are not translation families but for which a similar argument will establish (3.2). Rather than formulate a general result which would lead us astray from the theme of this section, we will give an example that well illustrates the idea.

EXAMPLE 3.1. Take $\mathscr{X} = R^n$, $\Theta = (0, \infty)^n$ and let dx denote Lebesgue measure on R^n. Consider the family of densities on R^n given by

$$p(x \mid \theta) = \prod_{i=1}^{n} \frac{x_i^{\theta_i - 1}}{\Gamma(\theta_i)} \exp(-x_i) I_{(0,\infty)}(x_i).$$

This family has the property

$$p(x \mid \theta + \eta) = \int_{R^n} p(x - y \mid \theta) p(y \mid \eta)\, dy$$

for $x \in \mathscr{X}$, $\theta, \eta \in \Theta$. Such families are called *convolution families* and will come up again in Section 4. Let P_θ be the probability measure defined by the density $p(\cdot \mid \theta)$. For $\eta \le \theta$, we claim that

$$\int f(x) P_\eta(dx) \le \int f(x) P_\theta(dx).$$

Let $\delta = \theta - \eta$ so $\delta_i \ge 0$, $i = 1, \cdots, n$. Then,

$$\begin{aligned}
\int f(x) P_\theta(dx) &= \int f(x) p(x \mid \eta + \delta)\, dx \\
&= \int \int f(x) p(x - y \mid \eta) p(y \mid \delta)\, dy\, dx \\
&= \int \left\{ \int f(x + y) p(x \mid \eta)\, dx \right\} p(y \mid \delta)\, dy \\
&= \int_{(0,\infty)^n} \left\{ \int f(x + y) p(x \mid \eta)\, dx \right\} p(y \mid \delta)\, dy \\
&\ge \int_{(0,\infty)^n} \left\{ \int f(x) p(x \mid \eta)\, dx \right\} p(y \mid \delta)\, dy \\
&= \int f(x) P_\eta(dx).
\end{aligned}$$

The inequality above follows from the fact that $f(x + y) \ge f(x)$ for $y \in (0, \infty)^n$ since $f \in \mathscr{F}$. The argument given is valid when $\delta_i > 0$ for $i = 1, \cdots, n$ but a simple continuity argument establishes the general case of $\delta_i \ge 0$. The above argument can also be used in the Poisson case (see Example 4.2).

We now return to the general case and assume that μ is a product measure on $\mathscr{X}$ - that is, $\mu = \mu_1 \times \cdots \times \mu_n$ where each μ_i is a σ-finite measure on $\mathscr{X}_i$. The following result, which underlies much of the remainder of this section, is basically due to Ahlswede and Daykin (1979). However, the proof which we outline below is due to Karlin and Rinott (1980). For $x, y \in \mathscr{X}$, the notation $x \vee y$ and $x \wedge y$ will be as defined in Problem 2.1.

THEOREM 3.1 (Ahlswede and Daykin, 1979; Karlin and Rinott, 1980) *Suppose p_i, $i = 1, \cdots, 4$ are non-negative functions on $\mathscr{X}$ which satisfy*

$$p_1(x) p_2(y) \le p_3(x \vee y) p_4(x \wedge y). \tag{3.3}$$

Then

$$\int p_1\, d\mu \int p_2\, d\mu \le \int p_3\, d\mu \int p_4\, d\mu. \tag{3.4}$$

PROOF. The proof is by induction on n. The important observation is that when (3.3) holds for p_i, $i = 1, \cdots, 4$, then (3.3) also holds for the marginals

$$\phi_i(u) \equiv \int_{\mathscr{X}_n} p_i(u, s)\mu_n(ds), \qquad i = 1, \cdots, 4$$

where $u \in \prod_1^{n-1} \mathscr{X}_i$ and $s \in \mathscr{X}_n$. A bit of algebra and the use of (3.3) with $x = y$ shows that the inequality $\phi_1(u)\phi_2(v) \le \phi_3(u \vee v)\phi_4(u \wedge v)$ is equivalent to

$$\int\int_{s<t} \{p_1(u, s)p_2(v, t) + p_1(u, t)p_2(v, s)\}\mu_n(ds)\mu_n(dt) \tag{3.5}$$
$$\le \int\int_{s<t} \{p_3(u \vee v, s)p_4(u \wedge v, t) + p_3(u \vee v, t)p_4(u \wedge v, s)\}\mu_n(ds)\mu_n(dt).$$

With $a = p_1(u, s)p_2(v, t)$, $b = p_1(u, t)p_2(v, s)$, $c = p_3(u \vee v, s)p_4(u \wedge v, t)$ and $d = p_3(u \vee v, t)p_4(u \wedge v, s)$ it suffices to show that $a + b \le c + d$ to verify (3.5). Since $s < t$, (3.3) implies that $a \le d$, $b \le d$ and $ab \le cd$. But $c + d - (a + b) = (1/d)\{(d - a)(d - b) + (cd - ab)\} \ge 0$, so the ϕ_i, $i = 1, \cdots, 4$, satisfy (3.3). However, exactly the same argument, suppressing u and v in (3.5), shows that (3.4) holds for $n = 1$. Since μ is a product measure, the induction step is now easily completed.

The first application of Theorem 3.1 yields a useful sufficient condition for (3.1) to hold.

THEOREM 3.2 (Preston, 1974; Holley, 1974; Kemperman, 1977; Edwards, 1978). *Let P_1 and P_2 be probability measures on $\mathscr{X}$. Assume there exists a product measure μ on $\mathscr{X}$ such that P_i has a density p_i with respect to μ and*

$$p_1(x)p_2(y) \le p_1(x \wedge y)p_2(x \vee y). \tag{3.6}$$

Then

$$\int f(x)p_1(x)\mu(dx) \le \int f(x)p_2(x)\mu(dx), \qquad f \in \mathscr{F}. \tag{3.7}$$

PROOF. Let $\tilde{p}_1 = p_1 f$, $\tilde{p}_2 = p_2$, $\tilde{p}_3 = p_2 f$ and $\tilde{p}_4 = p_1$. Then (3.3) holds for $\tilde{p}_i$, $i = 1, \cdots, 4$. Since p_i is a density, (3.4) implies (3.7) and the proof is complete.

REMARK 3.1. The proofs of Theorem 3.2 due to Kemperman and Edwards are similar and use induction. These proofs use a property of conditional distributions which is of independent interest and is connected with the work of Daley (1969) and of Kamae et al. (1977).

THEOREM 3.3 (Sarkar, 1969; Fortuin, Ginibre and Kasteleyn, 1971). *Suppose p is a density on $\mathscr{X}$ with respect to a product measure μ. If p satisfies*

$$p(x)p(y) \le p(x \wedge y)p(x \vee y), \tag{3.8}$$

then for f_1, $f_2 \in \mathscr{F}$,

$$\int f_1(x)f_2(x)p(x)\mu(dx) \ge \int f_1(x)p(x)\mu(dx) \int f_2(x)p(x)\mu(dx). \tag{3.9}$$

That is, $\operatorname{Cov}\{f_1(X), f_2(X)\} \ge 0$ *if $X \in \mathscr{X}$ has density p.*

PROOF. Let $\delta = \int f_1 p\, d\mu$, so (3.9) is obvious if $\delta = 0$. For $\delta > 0$, let $p_2 = \delta^{-1} f_1 p$ and $p_1 = p$. Then p_1 and p_2 are densities which satisfy (3.6). Applying (3.7) with $f = f_2$ yields (3.9) and the proof is complete.

REMARK 3.2. When $n = 2$, then (3.8) is just the assumption that p is totally positive of order 2 (TP_2); that is, p has a monotone likelihood ratio. For general n, a non-negative function p satisfying (3.8) is MTP_2 (multivariate TP_2); see Karlin and Rinott (1980). Condition (3.8) is difficult to check in most cases. However, under certain conditions, if p is TP_2 in each pair of variables with the others held fixed, then p is MTP_2. For a more complete discussion of this issue, see Kemperman (1977, Assertion (i) in Section 6) and Karlin and Rinott (1980, Proposition 2.1). Densities which are MTP_2 have a number of interesting properties. For example, any marginal of a MTP_2 density is again MTP_2. Also, a basic composition theorem for MTP_2 kernels extends the well known composition result for TP_2 kernels (Karlin, 1956). These as well as many other facts about MTP_2 densities are established in Karlin and Rinott (1980).

As mentioned in the discussion of Problem 2.1, a random vector $X \in R^n$ is associated if

$$\text{Cov}\{f(X), g(X)\} \geq 0, \qquad f, g \in \mathscr{F}, \tag{3.10}$$

and (3.8) is a sufficient condition for (3.10) to hold. For example, if the coordinates of X are independent, choose μ to be the probability measure of X and choose $p = 1$. Thus, for independent variables, (3.10) holds (Esary et al., 1967). If X is multivariate normal, $N(\mu, \Sigma)$, then (3.8) holds iff the matrix $B = \Sigma^{-1}$ satisfied $b_{ij} \leq 0$ for $i \neq j$ (Sarkar, 1969). For other examples and a more thorough discussion, see Tong (1980) and Karlin and Rinott (1980).

When X is associated, (3.10) yields useful probability inequalities. For example if

$$f(x_1, \cdots, x_n) = \begin{cases} 1 & \text{if } \ x_i \geq c_i, \quad i = 1, \cdots, k \\ 0 & \text{otherwise} \end{cases}$$

and

$$g(x_1, \cdots, x_n) = \begin{cases} 1 & \text{if } \ x_i \geq c_i, \quad i = k+1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

then (3.10) yields

$$P(X_i \geq c_i, i = 1, \cdots, n) \geq P(X_i \geq c_i, i = 1, \cdots, k)P(X_i \geq c_i, i = k+1, \cdots, n).$$

An easy induction argument then gives

$$P(X_i \geq c_i, i = 1, \cdots, n) \geq \textstyle\prod_{i=1}^n P(X_i \geq c_i).$$

Another interesting application of the results in this section has been given recently by Perlman and Olkin (1980).

A random vector $X \in R^n$ is called *associated in absolute value* if $|X_1|, \cdots, |X_n|$ are associated in the sense of Definition 2.2. Jogdeo (1977) has given some sufficient condition that X be associated in absolute value. In the case that $X \in R^n$ is $N(0, \Sigma)$, Karlin and Rinott (1981) have shown that X is associated in absolute value iff there exists an element $D \in \mathscr{D}_n$, the group of coordinate sign changes, such that $B \equiv (D\Sigma D)^{-1} = \{b_{ij}\}$ satisfies $b_{ij} \leq 0$ for $i \neq j$. A recent result of Pitt (1981) asserts that if X is normal, then X is associated iff all the covariances are nonnegative.

4. Majorization and reflection groups. Much of the material in this section is motivated by the vast literature concerning majorization and Schur functions as introduced in Section 1. With the results established below, it will be demonstrated that a number of common families of multivariate distributions (e.g., the multivariate normal with varying mean vector, the multinomial with varying cell probabilities) determine increasing families of distributions relative to pre-orders defined by special groups of orthogonal transformations. Many of the examples below concern the permutation group $\mathscr{P}_n$ and the majorization pre-ordering. In fact, these applications were known prior to the development of the general theory of this section. However, the application of the general theory given in Proposition 4.2 does show that problems of practical interest can be attacked by the results in this section. We now proceed with the technical details.

Consider n-dimensional space R^n and let G be a closed subgroup of $\mathcal{O}_n$, the group of $n \times n$ orthogonal matrices. Vectors in R^n are column vectors and gx, $x \in R^n$, $g \in G$ means the matrix g times the vector x. For $y \in R^n$, $C(y)$ will denote the convex hull of the set $\{gy \mid g \in G\}$. The set $C(y)$ is compact as G is compact. We will write $x \leq y$ to mean that $x \in C(y)$ so $\leq$ is a pre-order. A function f on R^n is *G-increasing* (*G-decreasing*) if $x \leq y$ implies that $f(x) \leq f(y)$ ($f(x) \geq f(y)$). Let $\mathcal{F}$ be the convex cone of non-negative, bounded, Borel measurable G-increasing functions. Observe that any f which is increasing or decreasing necessarily satisfies $f(x) = f(gx)$ for all $x \in R^n$, $g \in G$ as $x \leq gx$ and $gx \leq x$. In general terms, the problem to be discussed in this section is the following: Given two probability measures, P_1 and P_2, under what conditions will we have

$$\int f(x)P_1(dx) \leq \int f(x)P_2(dx), \qquad f \in \mathcal{F}\ ? \tag{4.1}$$

One case of particular interest is when G is the permutation group so the pre-order is that of majorization.

Very little is known for the general problem posed above; for a discussion, see Eaton (1975). However, a fair amount is known for a class of groups called reflection groups (see Benson and Grove, 1971). To define these, let $r \in R^n$, $\| r \| = 1$ and set $S_r = I - 2rr'$. Then S_r is an $n \times n$ orthogonal matrix such that $S_r r = -r$ and S_r is the identity on $H_r = \{x \mid x \in R^n, r'x = 0\}$. Thus, S_r is a *reflection* through the hyperplane H_r, or more simply, S_r is a reflection.

DEFINITION 4.1. Let G be a closed subgroup of $\mathcal{O}_n$. Then, G is called a *reflection group* if there exists a set $\Delta \subseteq \{x \mid x \in R^n, \| x \| = 1\}$ such that G is the smallest closed subgroup of $\mathcal{O}_n$ which contains $\{S_r \mid r \in \Delta\}$.

In what follows, it will always be assumed that G *is a finite reflection group.*

A decomposition result for reflection groups (see Proposition 3.2 in Eaton and Perlman, 1977) shows that there is no loss of generality with this assumption. For statistical applications, the three most important examples of reflection groups are:

(i) $\mathcal{P}_n$ – the group of $n \times n$ permutation matrices,
(ii) $\mathcal{D}_n$ – the group of coordinate sign changes acting on R^n – $\mathcal{D}_n$ has 2^n elements,
(iii) $\mathcal{P}_n \cup \mathcal{D}_n$ – the group generated by $\mathcal{P}_n$ and $\mathcal{D}_n$.

If G is a reflection group, then the set

$$\Delta(G) = \{r \mid \|r\| = 1, \qquad S_r \in G\}$$

is called the *root system* of G. It is clear that if $r \in \Delta(G)$, then $gr \in \Delta(G)$ since $gS_r g^{-1} = S_{gr}$ for $g \in G$. A set Δ_0 of vectors of length one is called a *generating set for G* if G is the smallest closed subgroup of $\mathcal{O}_n$ which contains $\{S_r \mid r \in \Delta_0\}$. Generating sets for $\mathcal{P}_n$ and $\mathcal{D}_n$ provide useful examples.

EXAMPLE 4.1. Let $\varepsilon_1, \cdots, \varepsilon_n$ denote the standard unit vectors in R^n. For $\mathcal{P}_n$, a minimal generating set is

$$\{(\varepsilon_i - \varepsilon_{i+1})/\sqrt{2} : i = 1, \cdots, n-1\}.$$

For $\mathcal{D}_n$, a minimal generating set is

$$\{\varepsilon_i : i = 1, \cdots, n\}.$$

If Δ_0 is a generating set for G, then $\Delta(G) = \{gr \mid r \in \Delta_0, g \in G\}$; see Theorem 4.2.5 in Benson and Grove (1971).

The basic property of reflection groups, $G \subseteq \mathcal{O}_n$, which simplifies a characterization of G-increasing functions, will now be discussed. Let $\mathcal{X} \subseteq R^n$ be a G-invariant subset of R^n. Given $r \in \Delta(G)$, let $u \in R^n$ be any vector such that $r'u = 0$ and define $B_{u,r} \subseteq R^1$ by

$$B_{u,r} = \{\beta \mid u + \beta r \in \mathcal{X}\}.$$

Note that $B_{u,r}$ may be empty for certain vectors u and $r \in \Delta(G)$. Since $S_r(u + \beta r) = u - \beta r$, it is clear that $B_{u,r} = -B_{u,r}$.

THEOREM 4.1. *Let f be a real valued function defined on $\mathcal{X}$, and let Δ_0 be a generating set for G. The following are equivalent:*

(i) *f is a G-increasing function,*
(ii) *f is G-invariant and for each $r \in \Delta_0$ and u such that $r'u = 0$, the function φ defined on $B_{u,r}$ by $\varphi(\beta) = f(u + \beta r)$ is non-decreasing for $\beta \geq 0$, $\beta \in B_{u,r}$.*

PROOF. For (i) implies (ii), observe that if $0 < \beta_1 < \beta_2$ then $u + \beta_1 r$ is in the line segment connecting $u + \beta_2 r$ and $S_r(u + \beta_2 r) = u - \beta_2 r$ and thus $u + \beta_1 r \in C(u + \beta_2 r)$. Hence $\varphi(\beta_1) \leq \varphi(\beta_2)$ since f is G-increasing.

The converse is the more difficult and useful assertion. It depends on the following basic geometric fact concerning reflection groups.

Fact 4.1. Suppose $x \in C(z)$. Then there exists $g_1, g_2 \in G$ and a sequence of vectors $y_0, y_1, \cdots, y_k$, and $r_0, \cdots, r_{k-1} \in \Delta_0$, such that $y_0 = g_1 x$, $y_k = g_2 z$, $r_i' y_i \geq 0$ and $y_{i+1} = y_i + \gamma_i r_i$ where $\gamma_i \geq 0$, for $i = 0, \cdots, k - 1$.

Geometrically, Fact 4.1 means that there is a polygonal path from $g_1 x$ to $g_2 z$, which we write $y_0 \to y_1 \to \cdots \to y_k$, and on each affine segment $\Gamma_i = \{y_i + \delta r_i, 0 \leq \delta \leq \gamma_i\}$, $\delta \leq \tilde{\delta}$ implies $y_i + \delta r_i \leq y_i + \tilde{\delta} r_i$. To show $f(x) \leq f(z)$, it suffices to show that $f(y_i) \leq f(y_{i+1})$ and for this we write

$$y_i + \delta r_i = u_i + \alpha r_i \in \Gamma_i,$$

where $u_i' r_i = 0$. Hence $\alpha = r_i' y_i + \delta \geq 0$ and as δ moves from 0 to γ_i, α moves from $r_i' y_i$ to $r_i' y_i + \gamma_i$. Since $u_i + \alpha r_i \leq u_i + \tilde{\alpha} r_i$ for $\alpha \leq \tilde{\alpha}$, we see that

$$f(y_i) = f(u_i + (r_i' y_i) r_i) \leq f(u_i + (r_i' y_i + \gamma_i) r_i) = f(y_{i+1}),$$

so $f(x) \leq f(z)$. The proof of Fact 4.1 is given in Eaton and Perlman (1979) for the general reflection group. However, this fact can be established quite easily for $G = \mathcal{D}_n$. For $G = \mathcal{P}_n$, a related result due to Hardy, Littlewood and Polya (1934) concerns the existence of T-transforms; see Marshall and Olkin (1979, page 21) for a discussion.

The content of Theorem 4.1 is that $f \in \mathcal{F}$ iff f is G-invariant and behaves properly on certain lines in R^n, namely f must be symmetric and non-decreasing as $\beta \in B_{u,r}$ moves away from zero on the line $\{u + \beta r \mid \beta \in R^1\}$. Of course, a similar characterization holds for G-decreasing functions. That is, f is G-decreasing iff f is G-invariant and $\phi(\beta) = f(u + \beta r)$ is symmetric and unimodal on $B_{u,r}$, that is, symmetric on $B_{u,r}$ and decreasing on $[0, \infty) \cap B_{u,r}$.

Theorem 4.1 can be used to give an easy proof of a basic convolution result. Take $\mathcal{X} = R^n$ and suppose f_0 is a density with respect to Lebesgue measure on $\mathcal{X}$.

THEOREM 4.2. *Suppose f_0 is a G-decreasing function on R^n. For any bounded G-decreasing function f,*

$$\Psi(\theta) \equiv \int_{R^n} f(x) f_0(x - \theta)\, dx$$

is also a G-decreasing function.

PROOF. Let $r \in \Delta(G)$ and let u be such that $r'u = 0$. It must be shown that

$$\Psi(u + \beta r) = \int_{R^n} f(x) f_0(x - u - \beta r)\, dx$$

is a decreasing function on $[0, \infty)$. The G-invariance of Ψ is easily verified as f, f_0 and Lebesgue measure are all G-invariant. Write $x = v + \nu r$ with $v \in H_r$ and $\nu \in R^1$. First integrating on ν and then integrating over H_r, the above integral is

$$\Psi(u + \beta r) = \int_{H_r} \int_{-\infty}^{\infty} f(v + \nu r) f_0(v - u - (\beta - \nu))\, d\nu\, dv.$$

But, for fixed v, $f(v + \nu r)$ is a symmetric function of ν and is decreasing for $\nu \in [0, \infty)$. Also, $f(v - u - \delta r)$ is a symmetric function of δ and decreases for $\delta \in [0, \infty)$. Now, for fixed u and v,

$$\xi(\beta) = \int_{-\infty}^{\infty} f(v + \nu r) f_0(v - u - (\beta - \nu) r)\, d\nu$$

is the convolution of two symmetric unimodal functions on R^1, and hence ξ is symmetric and unimodal (see Wintner, 1939). Then, integrating over H_r also results in a symmetric unimodal function of β. By Theorem 4.1, Ψ is G-decreasing.

For $G = \mathscr{P}_n$, Theorem 4.2 was first proved by Marshall and Olkin (1974). The general reflection group case was established by Eaton and Perlman (1977) using the proof above. For $G = \mathscr{D}_n$, this result was proved independently by Jogdeo (1977). Of course, Theorem 4.2 implies that if f is G-increasing, then Ψ is G-increasing. Examples of G-decreasing functions are provided by any G-invariant function f such that $\{x \mid f(x) \geq v\}$ is convex for each $v \in R^1$. In particular, if f is log concave and G-invariant, then f is G-decreasing. When $f_0(x \mid \sigma) = (\sqrt{2\pi}\sigma)^{-n} \exp(-\frac{1}{2}\sigma^{-2}x'x)$, then f_0 is G-decreasing for every reflection group G so Theorem 4.2 holds for f_0. This implies that any mixture (on σ) of $f_0(\cdot \mid \sigma)$ also is G-decreasing for any reflection group G. For $G = \mathscr{P}_n$, $G = \mathscr{D}_n$ or $G = \mathscr{D}_n \cup \mathscr{P}_n$, the density $f_1(x) = 2^{-n} \exp(-\Sigma \mid x_i \mid)$ is G-decreasing and Theorem 4.2 holds. Further examples will be given later.

Since there are many useful and interesting parametric families which are not translation families, it is natural to ask for sufficient conditions on a density $p(x \mid \theta)$ so that

$$\Psi(\theta) = \int_{\mathscr{X}} f(x) p(x \mid \theta) \lambda(dx)$$

is G-decreasing (or G-increasing) when f is G-decreasing (or G-increasing). When $G = \mathscr{P}_n$, Proschan and Sethuraman (1977), Nevius et al. (1977) and Hollander et al. (1977) have obtained some useful sufficient conditions in order that Ψ be monotone. These results have been extended to reflection groups in Conlon et al. (1977a, 1977b). We will discuss general reflection groups but will use a slightly different approach than in Conlon et al. In the discussion that follows, the reader is urged to keep the following examples in mind:

(i) the translation family example with G taken as $\mathscr{P}_n$, $\mathscr{D}_n$ or $\mathscr{P}_n \cup \mathscr{D}_n$,
(ii) the multinomial distribution with G taken as $\mathscr{P}_n$,
(iii) the n-variable Poisson distribution with independent coordinates, a different parameter for each coordinate and $G = \mathscr{P}_n$.

To set the stage for our general discussion, let G be a finite reflection group acting on R^n and let $\mathscr{X}$ and Θ be G-invariant Borel subsets of R^n. A useful property of real valued functions defined on $\mathscr{X} \times \Theta$ is the following.

DEFINITION 4.2. A function k on $\mathscr{X} \times \Theta$ to R^1 is a *decreasing reflection* (DR) *function* if

(i) $k(x, \theta) = k(gx, g\theta)$, $x \in \mathscr{X}$, $\theta \in \Theta$, $g \in G$, and
(ii) for $r \in \Delta(G)$, if $r'xr'\theta \geq 0$, then $k(x, \theta) \geq k(x, S_r\theta)$.

REMARK 4.1. In the case that $G = \mathscr{P}_n$, DR functions have been studied in a number of contexts. Savage (1957) considered this property in his study of rank order statistics. DR functions were said to have "property M" by Eaton (1966) in a paper concerned with ranking problems. Later, Hollander et al. (1977) introduced the term "decreasing in transposition" for such functions. See Marshall and Olkin (1979, 6.F) for some related material.

REMARK 4.2. If we think of $k(x, \theta)$ as a likelihood function, the DR property has an interesting geometric interpretation. For $r \in \Delta(G)$, the condition $r'xr'\theta \geq 0$ means that x and θ are on the same side of the hyperplane $H_r = \{x \mid r'x = 0\}$. Thus, condition (ii) of Definition 4.2 means that: given x, between the two possible parameter points θ and $S_r\theta$, the one that is more likely is the one on the same side of H_r as x.

REMARK 4.3. In some special cases, DR functions are related to functions with a monotone likelihood ratio. Let $\mathscr{X}_1$ be a subset of R^1 and let Θ_1 be a subset of R^1. Assume that $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_1 \subseteq R^n$ and $\Theta = \Theta_1 \times \cdots \times \Theta_1 \subseteq R^n$. Let ξ be defined on $\mathscr{X}_1 \times \Theta_1$ to $[0, \infty)$, and consider k on $\mathscr{X} \times \Theta$ given by

$$k(x, \theta) = \textstyle\prod_{i=1}^n \xi(x_i, \theta_i)$$

for $x \in \mathscr{X}$ and $\theta \in \Theta$. For $G = \mathscr{P}_n$, clearly $k(gx, g\theta) = k(x, \theta)$. It is not difficult to show that k is a DR function ($G = \mathscr{P}_n$) iff ξ has a monotone likelihood ratio (see Eaton, 1966).

REMARK 4.4. When k on $\mathscr{X} \times \Theta$ satisfies (i) of Definition 4.2, the verification of (ii) can often be restricted to a much smaller set of reflections than $\{S_r \mid r \in \Delta(G)\}$. For example, if (ii) holds for S_r, $r \in \Delta_0$ where Δ_0 is a set of generators, then (ii) holds for all S_r, $r \in \Delta(G)$. More generally, if Δ_1 is a set of roots such that

$$\Delta(G) = \{r \mid r = gr_1, \quad g \in G, \quad r_1 \in \Delta_1\},$$

and if (ii) holds for S_r, $r \in \Delta_1$, then (ii) holds for all S_r. When $G = \mathscr{P}_n$, it is sufficient to verify (ii) for one particular r, namely $(\varepsilon_1 - \varepsilon_2)/\sqrt{2}$.

REMARK 4.5. A definition equivalent to Definition 4.2 for functions k has been given in Conlon et al. (1977a, b). We have chosen the definition above because of the interpretation in Remark 4.2 and because of technical considerations.

A useful connection between G-increasing (G-decreasing) functions and DR functions follows.

PROPOSITION 4.1. *Assume $\mathscr{X} \subseteq R^n$ is a group under addition, and that G acts on $\mathscr{X}$. Given f defined on $\mathscr{X}$ to R, define k_1 and k_2 on $\mathscr{X} \times \mathscr{X}$ by*

$$k_1(x, y) = f(x - y), \qquad k_2(x, y) = f(x + y).$$

Then, f is G-decreasing iff *k_1 is a* DR *function. Also, f is G-increasing* iff *k_2 is a* DR *function.*

PROOF. Assume f is decreasing. Since f is G-invariant, obviously $k_1(x, y) = f(x - y)$ satisfies $k_1(gx, gy) = k_1(x, y)$. For $r \in \Delta$ and x, y such that $r'xr'y \geq 0$, it must be verified that $k_1(x, y) \geq k_1(x, S_r y)$. Write $x = u + \beta r$ and $y = v + \gamma r$ with $\beta, \gamma \in R^1$ and $u'r = v'r$

$= 0$. Thus, $r'xr'y = \beta\gamma \geq 0$. This implies that $|\beta - \gamma| \leq |\beta + \gamma|$. Since $f(w + \alpha r)$ is a unimodal function of α for each w with $w'r = 0$, we have

$$k_1(x, y) = f(u - v + (\beta - \gamma)r) \geq f(u - v + (\beta + \gamma)r) = k_1(x, S_r y).$$

Conversely, assume that $k_1(x, y) = f(x - y)$ is a DR function. For $r \in \Delta$, u with $u'r = 0$, and $0 < \alpha_1 < \alpha_2$, we must show that $f(u + \alpha_1 r) \geq f(u + \alpha_2 r)$. Now, set $x = u + \frac{1}{2}(\alpha_1 + \alpha_2)r$ and $y = \frac{1}{2}(\alpha_2 - \alpha_1)r$. Then $r'xr'y \geq 0$ and

$$f(u + \alpha_1 r) = k_1(x, y) \geq k_1(x, S_r y) = f(u + \alpha_2 r).$$

The case for f increasing is similar.

For $G = \mathscr{P}_n$, Proposition 4.1 is due to Hollander et al. (1977) while the general case is in Conlon et al. (1977a, b). The following basic composition theorem was first established by Hollander et al. (1977) for $G = \mathscr{P}_n$ and was proved for general reflection groups in Conlon et al. (1977a, b).

THEOREM 4.3. *Consider Borel sets $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z} \subseteq R^n$ which are acted on by a reflection group G. Let $k_1(x, y)$ and $k_2(x, y)$ be* DR *functions on $\mathscr{X} \times \mathscr{Y}$ and $\mathscr{Y} \times \mathscr{Z}$. Let λ be a σ-finite G-invariant measure such that*

$$k_3(x, z) = \int_{\mathscr{Y}} k_1(x, y)k_2(y, z)\lambda(dy)$$

is well defined and finite for all $x \in \mathscr{X}$ and $z \in \mathscr{Z}$. Then k_3 is a DR *function on $\mathscr{X} \times \mathscr{Z}$.*

PROOF. That $k_3(gx, gz) = k_3(x, z)$ is easily verified from the G-invariance of k_1, k_2 and λ. For $r \in \Delta(G)$ and for x and z such that $r'xr'z \geq 0$, we must show that $\delta = k_3(x, z) - k_3(x, S_r z)$ is non-negative. Define $\mathscr{Y}_1 \subseteq \mathscr{Y}$ by $\mathscr{Y}_1 = \{y \mid y \in \mathscr{Y}, r'y > 0\}$. Using the assumed invariance properties of k_1, k_2 and λ, a bit of manipulation shows that

$$\delta = \int_{\mathscr{Y}_1} \{k_1(x, y) - k_1(x, S_r y)\}\{k_2(y, z) - k_2(y, S_r z)\}\lambda(dy). \tag{4.2}$$

However, when $r'xr'z \geq 0$ and $y \in \mathscr{Y}_1$, the assumption that k_1 and k_2 are DR functions implies that the integrand in (4.2) is always non-negative. Thus $\delta \geq 0$ and the result follows.

Theorem 4.3 has a number of interesting applications. This result together with Proposition 4.1 provides another proof of Theorem 4.2 (in fact, a slight extension of Theorem 4.2).

THEOREM 4.4. *Suppose $\mathscr{X} \subseteq R^n$ is a group under addition and $\mathscr{X}$ is acted on by the reflection group G. Let μ be a translation-invariant measure on $\mathscr{X}$. Suppose f_0 is a G-decreasing density with respect to μ and let f be a non-negative G-decreasing function. Then*

$$\Psi(\theta) \equiv \int_{\mathscr{X}} f(x)f_0(x - \theta)\mu(dx)$$

is a G-decreasing function of $\theta \in \mathscr{X}$.

PROOF. It suffices to show that $k_3(\theta, \eta) \equiv \Psi(\theta - \eta)$ is a DR function on $\mathscr{X} \times \mathscr{X}$. But,

$$\Psi(\theta - \eta) = \int f(x)f_0(x - (\theta - \eta))\mu(dx) = \int f(x - \eta)f_0(x - \theta)\mu(dx),$$

where the last equality follows from the translation invariance of μ. From Proposition 4.1, $k_1(\theta \mid x) = f_0(x - \theta)$ and $k_2(x \mid \eta) = f(x - \eta)$ are DR functions. The conclusion follows from Theorem 4.3.

REMARK 4.6. Theorem 4.4 applies to the case when $\mathscr{X}$ is the set of vectors in R^n with integer coordinates and μ is counting measure. The group of interest here would ordinarily be $\mathscr{P}_n$.

Now, we consider the general problem. As usual, $\mathscr{X}$ and Θ are G-invariant subsets of R^n, λ is a G-invariant measure on $\mathscr{X}$. Suppose $k(x, \theta)$ is a DR function on $\mathscr{X} \times \Theta$ such that $k(\cdot, \theta)$ is a density with respect to λ for each $\theta \in \Theta$. Here is the problem: Find further conditions on k so that

$$\Psi(\theta) \equiv \int_{\mathscr{X}} f(x)k(x, \theta)\lambda(dx) \tag{4.3}$$

is G-increasing on Θ whenever f is G-increasing on $\mathscr{X}$.

REMARK 4.7. The assumptions made on k are not sufficient to show Ψ in (4.3) is G-decreasing. For example let $\mathscr{X} = R^2$ and $\Theta = \{(\theta_1, \theta_2) \mid \theta_i > 0, i = 1, 2\}$. Consider

$$\xi(x_i, \theta_i) = \begin{cases} 1/\theta_i \exp(-x/\theta_i), & x > 0 \\ 0, & x \leq 0 \end{cases}$$

and let

$$k(x, \theta) = \textstyle\prod_1^2 \xi(x_i, \theta_i).$$

Take λ to be Lebesgue measure on R^2, and take $G = \mathscr{P}_2$. The results of Diaconis (see Marshall and Olkin, 1979, 12.K.3, page 377) show that Ψ is not G-decreasing for all G-decreasing f.

One possible approach is to consider $\Psi(\theta + \eta)$ and attempt to verify that this is a DR function. Now, assume Θ is closed under addition and assume that $\mathscr{X} \subseteq R^n$ is a group under addition. Then,

$$\Psi(\theta + \eta) = \int f(x)k(x, \theta + \eta)\lambda(dx).$$

Following Proschan and Sethuraman (1977), now assume that $\{k(\cdot \mid \theta) : \theta \in \Theta\}$ is a *convolution family* – that is, assume there exists a measure ν on $\mathscr{X}$ such that

$$k(x, \theta + \eta) = \int_{\mathscr{X}} k(x - y, \theta)k(y, \eta)\nu(dy) \tag{4.4}$$

for $x \in \mathscr{X}$, $\theta, \eta \in \Theta$ (see Example 3.1). With the further assumption that λ is translation invariant, we have

$$\begin{aligned} \Psi(\theta + \eta) &= \int_{\mathscr{X}} f(x)k(x, \theta + \eta)\lambda(dx) \\ &= \int_{\mathscr{X}}\int_{\mathscr{X}} f(x)k(x - y, \theta)k(y, \eta)\nu(dy)\lambda(dx) \\ &= \int_{\mathscr{X}} \left\{ \int_{\mathscr{X}} f(x + y)k(x, \theta)\lambda(dx) \right\} k(y, \eta)\nu(dy). \end{aligned} \tag{4.5}$$

When f is G-increasing, Theorem 4.3 shows that the inside integral is a DR function in θ and y. Assuming ν is G-invariant, a second application of Theorem 4.3 coupled with Proposition 4.1 yields that Ψ is G-increasing. Summarizing all of this yields

THEOREM 4.5. *Suppose $\mathscr{X} \subseteq R^n$ is a group under addition and $\Theta \subseteq R^n$ is closed under addition. Assume G is a reflection group acting on both $\mathscr{X}$ and Θ and assume that λ is a translation-invariant and G-invariant measure on $\mathscr{X}$. Let $k(\cdot, \cdot)$ be a* DR *function on $\mathscr{X} \times \Theta$ such that $\{k(\cdot, \theta) \mid \theta \in \Theta\}$ is a convolution family with respect to a G-invariant measure ν on $\mathscr{X}$. If f is G-increasing, then*

$$\Psi(\theta) \equiv \int_{\mathscr{X}} f(x)k(x, \theta)\lambda(dx)$$

is a G-increasing function on θ.

PROOF. The proof is given above.

Although the assumptions of Theorem 4.5 are rather restrictive, this result does have a number of important applications.

EXAMPLE 4.2. (Poisson; Nevius, et al., 1977). In this case, let Z be the set of integers in R^1, $\mathscr{X} = Z^n$ and let λ be counting measure on $\mathscr{X}$. With $\Theta = (0, \infty)^n$, consider

$$k(x, \theta) = \prod_{i=1}^n \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!} I_{[0, \infty)}(x_i)$$

The assumptions of Theorem 4.5 are readily checked with $G = \mathscr{P}_n$ and $\nu = \lambda$. If we let P_θ denote the probability measure on $\mathscr{X}$ defined by $k(\cdot, \cdot)$, then, in terms of the notation in Section 1, we have $P_\theta < P_{\theta'}$, whenever $\theta \le \theta'$. Such families $\{P_\theta \mid \theta \in \Theta\}$ were called *Schur families* in Nevius et al. (1977).

EXAMPLE 4.3. (Gamma shape family; Nevius et al., 1977). Take $\mathscr{X} = R^n$ and $\Theta = (0, \infty)^n$ with λ as Lebesgue measure. Define k by

$$k(x, \theta) = \prod_{i=1}^n \frac{x_i^{\theta_i - 1}e^{-x_i}}{\Gamma(\theta_i)} I_{(0, \infty)}(x_i).$$

With $G = \mathscr{P}_n$ and $\nu = \lambda$, the assumptions of Theorem 4.5 are easily checked.

In the case that $G = \mathscr{P}_n$, other examples of Schur families can be constructed by conditioning and by mixing.

EXAMPLE 4.4. (Multinomial; Rinott, 1973, Nevius et al., 1977). Consider the multinomial density

$$k(x \mid p) = N! \prod_{i=1}^n \left(\frac{p_i^{x_i}}{x_i!}\right) I_{[0, \infty)}(x_i) I_N(\Sigma_1^n x_i),$$

where each x_i is an integer, $0 < p_i < 1$, $\Sigma_1^n p_i = 1$ and $I_N(\Sigma_1^n x_i)$ is the indicator of $\{\Sigma_1^n x_i = N\}$. The sample space, $\mathscr{X}$, and the dominating measure are the same as in Example 4.2. Now suppose f is a $\mathscr{P}_n$-decreasing function and consider a random vector $X \in \mathscr{X}$ whose components are independent Poisson's with parameter p_i, $i = 1, \cdots, k$. Conditional on $\Sigma_1^n x_i = N$, X has the above multinomial distribution. But, the expectation of f under the multinomial distribution is proportional to the expectation of $\tilde{f}(X) = I_N(\Sigma_1^n X_i)f(X)$. However, $\tilde{f}$ is $\mathscr{P}_n$-decreasing so by Example 4.2, this expectation is $\mathscr{P}_n$-decreasing in p. Thus the multinomial family is a Schur family.

EXAMPLE 4.5. (Dirichlet; see Application 4.2 in Nevius et al., 1977). Let $X \in R^n$ have independent coordinates with the ith component having a Gamma distribution with density

$$\frac{x_i^{\theta_i - 1}e^{-x_i}}{\Gamma(\theta_i)} I_{(0, \infty)}(x_i), \qquad x_i \in R^1,$$

where $\theta_i > 0$, $i = 1, \cdots, n$. Let f be a $\mathscr{P}_n$-decreasing function defined on $(0, \infty)^n$ (the complement of this set in $\mathscr{X}$ has probability zero) and note that $\tilde{f}$ defined by

$$\tilde{f}(x) = f((\Sigma_1^n x_i)^{-1} x)$$

is also $\mathscr{P}_n$-decreasing. Thus, from Example 4.3, the expectation of $\tilde{f}(X)$ is $\mathscr{P}_n$-decreasing in the vector θ. But $(\Sigma_1^n X_i)^{-1} X$ has a Dirichlet distribution with parameter vector θ. This shows that the Dirichlet family is a Schur family.

EXAMPLE 4.6. (Negative Multivariate Hypergeometric; see Application 4.2 in Nevius et al., 1977). Let $K_1(\cdot \mid p)$ be the probability measure of the multinomial distribution and let $K_2(\cdot \mid \theta)$ be the probability measure of the Dirichlet distribution. Consider $K_3(\cdot \mid \cdot)$ given by

$$K_3(\cdot \mid \theta) = \int K_1(\cdot \mid p) K_2(dp \mid \theta).$$

From Examples 4.4 and 4.5, $K_i(\cdot \mid \cdot)$, $i = 1, 2$ maps $\mathscr{P}_n$-increasing functions into $\mathscr{P}_n$-increasing functions. It follows easily that $K_3(\cdot \mid \cdot)$ does the same. (This argument holds in much greater generality; see Section 6). But the Dirichlet mixture of a multinomial distribution yields the negative multivariate hypergeometric distribution, so $\{K_3(\cdot \mid \theta)\}$ is a Schur family.

For the case of $G = \mathscr{P}_n$, many other examples and applications are given in Marshall and Olkin (1979). Although Theorems 4.4 and 4.5 are essentially the only general results currently available, in some cases, a direct verification that

$$\Psi(\theta) = \int_{\mathscr{X}} f(x) k(x, \theta) \lambda(dx) \tag{4.6}$$

is G-increasing (or G-decreasing) is possible. What must be shown is that Ψ is G-invariant and for $r \in \Delta(G)$ and u such that $r'u = 0$, the function

$$\eta(\beta) \equiv \Psi(u + \beta r)$$

is an increasing (or decreasing) function of β for $\beta \geq 0$ (when $u + \beta r \in \Theta$). In fact, since Theorem 4.5 is relatively recent, it should not be surprising that many results were first established using this direct approach. Examples of the direct approach can be found in Eaton (1970, 1974), Rinott (1973) and Gleser (1975).

To illustrate the usefulness of the general results for reflection groups, we now will discuss a method for obtaining probability inequalities when covariance matrices change in special ways. First suppose that Σ_0 is a fixed $p \times p$ positive definite matrix and define $\Sigma(\theta)$ by $\Sigma(\theta) = \Sigma_0 + \theta\theta'$ for $\theta \in R^p$. Particular cases of such covariances have occurred in a number of contexts; for example, see Dunnett and Sobel (1955), Dunn (1958) and Das Gupta et al. (1972). When X is $N(0, \Sigma(\theta))$, the problem is to describe the behavior of

$$\Psi(\theta) = \mathscr{E}_\theta f(X)$$

for certain types of functions f.

REMARK 4.8. When Σ_0 is diagonal, say $\Sigma_0 = D$ and

$$f_0(x) = \begin{cases} 1 & \text{if } |x_i| \le a_i, \quad i = 1, \cdots, p, \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\Psi(\theta) = \mathscr{E}_\theta f_0(X) = P_\theta(|X_i| \le a_i, i = 1, \cdots, p).$$

If X is $N(0, D + \theta\theta')$, it will be shown below that $\Psi(\theta)$ is a $\mathscr{D}_n$-decreasing function so $\Psi(0) \ge \Psi(\theta)$. This inequality should not be confused with the inequality of Dunn (1959). Fix θ and let $D_1 = D + D_\theta$ where D_θ has diagonal elements $\theta_2^2, \cdots, \theta_p^2$. When X is normal, Dunn's result implies that

$$\Psi(\theta) \ge P_{D_1}(|X_i| \le a_i, i = 1, \cdots, p).$$

A generalization of this inequality is given in Das Gupta et al. (1972), see their Corollary 3.1 and the ensuing discussion.

To proceed with the development here, first write Σ_0 in a spectral decomposition as

$$\Sigma_0 = \Sigma_{i=1}^p \lambda_i x_i x_i',$$

where $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of Σ_0 and $\{x_1, \cdots, x_p\}$ is an orthonormal basis of R^p consisting of eigenvectors of Σ_0. Let $S_i = I_p - 2x_i x_i'$ so S_i is a reflection and note that

$$S_i \Sigma_0 S_i = \Sigma_0, \qquad i = 1, \cdots, p.$$

Now, let G be a reflection group acting on R^p such that $g\Sigma_0 g' = \Sigma_0$ for all $g \in G$. In particular, G could be the reflection group generated by the reflections $S_1, \cdots, S_p$.

PROPOSITION 4.2. *Suppose X is $N_p(0, \Sigma_0 + \theta\theta')$, and G is a reflection group such that $g\Sigma_0 g' = \Sigma_0$ for all $g \in G$. If f is a G-decreasing function, then*

$$\Psi(\theta) = \mathscr{E}_\theta f(X)$$

is a G-decreasing function.

PROOF. First observe that X has the same distribution as $Z + Y\theta$ where Z is $N_p(0, \Sigma_0)$ and is independent of $Y \in R^1$ which is $N(0, 1)$. Fix $Y = y$ and let

$$\chi(\theta, y) = \mathscr{E} f(Z + y\theta).$$

With $\eta = y\theta$, the density of $Z + \eta$ is

$$p(Z - \eta) = (\sqrt{2\pi})^{-P} |\Sigma_0|^{-1/2} \exp\{-\tfrac{1}{2}(Z - \eta)' \Sigma_0^{-1}(Z - \eta)\}.$$

Since $p(\cdot)$ is log concave and is G-invariant ($g\Sigma_0 g' = \Sigma_0$ for $g \in G$), it follows that $p(\cdot)$ is G-decreasing. Theorem 4.1 implies that $\mathscr{E} f(Z + \eta)$ is a G-decreasing function of η. It is then easy to show that for each y, $\chi(\theta, y)$ is a G-decreasing function of θ and hence that

$$\Psi(\theta) = \mathscr{E}\chi(\theta, Y)$$

is G-decreasing. The proof is complete.

EXAMPLE 4.7. Suppose $\Sigma_0 = D$ is diagonal and take $G = \mathscr{D}_n$, the group of coordinate sign changes. Also, take f to be

$$f(x) = \begin{cases} 1 & \text{if } |X_i| \le a_i, \ i = 1, \cdots, P \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Sigma(\theta) = D + \theta\theta'$ and

$$\Psi(\theta) = P_\theta(|X_i| \le a_i, i = 1, \cdots, p).$$

Proposition 4.2 implies that Ψ is $\mathscr{D}_n$-decreasing, that is, Ψ is a function of $|\theta_1|, \cdots, |\theta_p|$ and is decreasing in $|\theta_i|$ for $i = 1, \cdots, p$.

EXAMPLE 4.8. Suppose Σ_0 has intraclass correlation structure - that is, $\Sigma_0 = \sigma^2 R$ where the diagonal elements of R are all one and the off diagonal elements of R are all ρ, $-1/(p - 1) < \rho < 1$. Take $G = \mathscr{P}_n$ so $g\Sigma_0 g' = \Sigma_0$ for all $g \in G$. Then, if f is any $\mathscr{P}_n$-decreasing function (Schur concave function) and $\Sigma(\theta) = \Sigma_0 + \theta\theta'$, then

$$\Psi(\theta) = \mathscr{E}_\theta f(X)$$

is $\mathscr{P}_n$-decreasing when X is $N(0, \Sigma(\theta))$.

EXAMPLE 4.9. Suppose $\Sigma_0 = \sigma^2 I_p$ and let $G = \mathscr{P}_n \cap \mathscr{D}_n$. As in the previous examples, if f is G-decreasing then $\Psi(\theta) = \mathscr{E}_\theta f(X)$ is G-decreasing when X is normal.

The extension of Proposition 4.2 to multivariate distributions which are scale mixtures of the mean zero multivariate normal is quite easy. However, the validity of Proposition 4.2 for distributions with densities of the general form

$$p(x \mid \theta) = |\Sigma(\theta)|^{-1/2} q(x'\Sigma(\theta)^{-1}x), \qquad x \in R^P$$

is in question.

A natural question to ask is whether or not any of the results of this section are valid for non-reflection groups. We will focus our attention on Theorem 4.2 and show by counter-example that this result is false with a vengeance for some very simple rotation groups in R^2. Let G_4 be the group of counterclockwise rotations through $\pi/2$, so that G_4 has four elements. Also, let f_0 be defined by

$$f_0(x) = \begin{cases} \dfrac{1}{\pi} & \text{if } \|x\| \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

so f_0 is a density with respect of Lebesgue on R^2. Let $\theta_0 \in R^2$ have coordinates (10, 10) and let θ_1 have coordinates (10, 0) so θ_1 is in $C(\theta_0)$, the convex hull $\{g\theta_0 \mid g \in G_4\}$, that is, $\theta_1 \le \theta_2$. With $S = \{x \mid \|x\| \le 1\}$, define the set A by

$$A = \cup_{x \in S+\theta_0} C(x).$$

Then, $u \in A$ implies that $C(u) \subseteq A$ so I_A is a G_4-decreasing function. Obviously, f_0 is a G-decreasing function. Now, consider

$$\Psi(\theta) = \int_{R^2} I_A(x) f_0(x - \theta)\, dx.$$

We will now argue that Ψ is not a G_4-decreasing function although Ψ is the convolution of two G-decreasing functions. First note that $\Psi(\theta_0) = 1$ since $S + \theta_0 \subseteq A$ and the support of $f_0(x - \theta_0)$ is $S + \theta_0$. However, a careful analysis of the set A will convince the reader that $S + \theta_1$ is not contained in A since the boundary of the set A "caves in" where the boundary intersects the coordinate axes. Since the support of $f_0(x - \theta_1)$ is $S + \theta_1$, it follows that $\Psi(\theta_1) < 1$. But $\theta_1 \le \theta_0$ so Ψ cannot be G_4-decreasing. One possible attempt to salvage something in the current situation is to ask for conditions on a particular density, f_1, so that

$$\Psi_1(\theta) = \int_{R^2} f(x) f_1(x - \theta)\, dx$$

is G-decreasing whenever f is G-decreasing. However, even for $f_1(x) = (2\pi)-1 \exp(-\frac{1}{2}\|x\|^2)$, it is not known whether or not Ψ_1 is G-decreasing when f is G-decreasing. There is nothing special about rotations through $\pi/2$: a similar analysis provides a counter example when G, acting on R^2, is the group of rotations through $2\pi/k$ for $k = 3, 4, 5, \cdots$.

5. Inequalities for Special Convex Sets. The results of this section are related to the problem introduced in Problem 2.2. Our approach here is to proceed from first principles and establish a few of the important facts concerning log concave functions, the behavior of probabilities of convex symmetric sets under translation, and the behavior of probabilities of some special convex sets when correlations change.

We begin with a basic result concerning log concave functions.

DEFINITION 5.1. A function f defined on R^n to $[0, \infty)$ is log concave if for all $x, y \in R^n$ and $\lambda \in (0, 1)$, we have

$$f(\lambda x + (1 - \lambda)y) \geq f^{\lambda}(x)f^{1-\lambda}(y). \tag{5.0}$$

The following theorem first appeared in Prékopa (1973), but the proof given here is due to Brascamp and Leib (1974).

THEOREM 5.1. *Suppose* $f: R^m \times R^n \rightarrow [0, \infty)$ *is* log *concave on* $R^m \times R^n$ *and*

$$g(x) \equiv \int_{R^n} f(x, z)\, dz$$

is finite for each x. *Then* g *is also* log *concave.*

We first argue that it suffices to prove the theorem for $m = n = 1$ and when f has compact support. Given $k > 0$, let I_k denote the indicator of the set $\{(x, z) \mid \|x\|^2 + \|z\|^2 \leq k\}$ so I_k is log concave. The Monotone Convergence Theorem shows that

$$g_k(x) = \int_{R^n} f(x, z)I_k(x, z)\, dz$$

converges to $g(x)$ for each x. Since $f \cdot I_k$ is log concave and (5.0) is preserved under pointwise limits, it suffices to prove the Theorem for f with compact support. Now, g is log concave on R^m iff for each x_1 and $x_2 \in R^m$, the function $\tilde{g}(t) \equiv g(x_1 + tx_2)$ is log concave on R^1. Since $\tilde{f}(t, z) \equiv f(x_1 + tx_2, z)$ is log concave on $R^1 \times R^n$, it suffices to prove the theorem for $n = 1$. However, if the result is known to hold for $m = 1$, then an easy induction argument will establish the Theorem for general m.

We now proceed with the proof for $m = n = 1$ when f has compact support. First, a technical lemma.

LEMMA 5.1. *Let* C *be a bounded convex subset of* R^2 *and define* g_0 *by*

$$g_0(x) = \int_{-\infty}^{\infty} I_C(x, z)\, dz.$$

On the set $D = \{x \mid g_0(x) > 0\}$, g_0 *is a concave function.*

PROOF. For any x, the set

$$C_x = \{z \mid (x, z) \in C\}$$

is a convex subset of R^1. For $x_1, x_2 \in D$, $g_0(x_1) > 0$, $i = 1, 2$, so C_{x_i} must be an interval with a non-empty interior, say $\text{int}(C_{x_i}) = (a_i, b_i)$ with $a_i < b_i$, $i = 1, 2$. Thus

$$g_0(x_i) = \int_{-\infty}^{\infty} I_C(x, z)\, dz = b_i - a_i, \qquad i = 1, 2.$$

For any two sets A and B, define $A + B = \{a + b \mid a \in A, b \in B\}$, and $\lambda A = \{\lambda a \mid a \in A\}$

with $\lambda \in R^1$. For $\lambda \in (0, 1)$, it is easy to show that

$$C_{\lambda x_i + (1-\lambda)x_2} \supseteq \lambda C_{x_1} + (1-\lambda) C_{x_2}.$$

But, the interior of $\lambda C_{x_1} + (1-\lambda)C_{x_2}$ is $(\lambda a_1 + (1-\lambda)a_2, \lambda b_1 + (1-\lambda)b_2)$. With μ_1 denoting Lebesgue measure on R^1, we have

$$\begin{aligned} g_0(\lambda x_1 + (1-\lambda)x_2) &= \mu_1(C_{\lambda x_1 + (1-\lambda)x_2}) \geq \mu_1(\lambda C_{x_1} + (1-\lambda)C_{x_2}) \\ &= \lambda b_1 + (1-\lambda)b_2 - (\lambda a_1 + (1-\lambda)a_2) = \lambda g_0(x_1) + (1-\lambda)g_0(x_2). \end{aligned}$$

The proof of Lemma 5.1 is complete.

To complete the proof in the case of $m = n = 1$ when f has compact support, it suffices to show that

$$g(\lambda x_1 + (1-\lambda)x_2) \geq g^{\lambda}(x_1) g^{1-\lambda}(x_2) \tag{5.1}$$

for $\lambda \in (0, 1)$ and for x_i such that $g(x_i) > 0$, $i = 1, 2$. Clearly, f is bounded. Without loss of generality, assume

$$\sup_z f(x, z) = \sup_z f(x_2, z) = s_0. \tag{5.2}$$

(If this is not the case, replace f by $e^{bx} f(x, z)$ where b is chosen so (5.2) holds. Multiplication by e^{bx} does not affect (5.1).) Since $g(x_i) > 0$, $s_0 > 0$. For $\gamma > 0$, let

$$D(\gamma) = \{(x, z) \mid f(x, z) \geq \gamma\}.$$

Since

$$f(x, z) = \int_0^{s_0} I_{D(\gamma)}(x, z)\, d\gamma$$

we have

$$g(x) = \int_0^{s_0} \int_{-\infty}^{\infty} I_{D(\gamma)}(x, z)\, dz\, d\gamma.$$

For each $\gamma \in (0, s_0)$, the set

$$D_{x_i}(\gamma) \equiv \{z \mid f(x_i, z) \geq \gamma\} \subseteq R^1$$

is non-empty, as $\gamma < s_0$, and is convex, $i = 1, 2$. The log concavity of f and the condition $g(x_i) > 0$ imply that $D_{x_i}(\gamma)$ has a non-empty interior for each $\gamma \in (0, s_0)$. Now, we apply Lemma 5.1 to the function $x \to \int_{-\infty}^{\infty} I_{D(\gamma)}(x, z)\, dz$ for $0 < \gamma < s_0$ to obtain

$$\begin{aligned} g(\lambda x_1 + (1-\lambda)x_2) &= \int_0^{s_0} \int_{-\infty}^{\infty} I_{D(\gamma)}(\lambda x_1 + (1-\lambda)x_2, z)\, dz\, d\gamma \\ &\geq \int_0^{s_0} \left\{ \lambda \int_{-\infty}^{\infty} I_{D(\gamma)}(x_1, z)\, dz + (1-\lambda) \int_{-\infty}^{\infty} I_{D(\gamma)}(x_2, z)\, dz \right\} d\gamma \\ &= \lambda g(x_1) + (1-\lambda) g(x_2) \geq g^{\lambda}(x_1) g^{1-\lambda}(x_2). \end{aligned}$$

The final inequality follows from the arithmetic-geometric mean inequality. This completes the proof of Theorem 5.1.

Theorem 5.1 has a number of important consequences. The following result, due to Davidovic et al. (1969), is one. It shows that the convolution of two log concave functions is log concave.

THEOREM 5.2. *If h_1 and h_2 are* log *concave functions on R^n such that*

$$g(x) \equiv \int h_1(x - z) h_2(z)\, dz$$

exists for each $x \in R^n$, then g is log *concave.*

PROOF. The function

$$f(x, z) = h_1(x - z)h_2(z)$$

is log concave on $R^n \times R^n$. The result follows immediately from Theorem 5.1.

LEMMA 5.2. *If C_1 and C_2 are convex sets in R^n and μ_n is Lebesgue measure, then the set*

$$D(\gamma) \equiv \{x \mid \mu_n(C_1 \cap (C_2 + x)) \geq \gamma\}$$

is convex for each $\gamma \in [0, \infty)$.

PROOF. This result, proved originally by Sherman (1955), is deduced from Theorem 5.1 as follows. The function

$$g(x) \equiv \mu_n(C_1 \cap (C_2 + x)) = \int I_{C_1}(z)I_{(-C_2)}(x - z)\, dz$$

is log concave on R^n. Thus, the upper sections of g, namely $D(\gamma)$, are convex.

Now, we will establish a result due to Mudholkar (1966) which generalized an important theorem proved by Anderson (1955). First, we need to introduce one possible definition of a unimodal function on R^1. This definition is due to Anderson (1955); see Dharmadhikari and Jogdeo (1976) and Das Gupta (1980) for other possible definitions.

DEFINITION 5.2. A function f on R^n to R^1 is A-unimodal if $\{x \mid f(x) \geq v\}$ is convex for each $v \in R^1$.

Now, let G be a subgroup of $\mathcal{O}_n$, so each $g \in G$ preserves Lebesgue measure. Recall that G induces a pre-order on R^n - that is, $x \leq y$ iff $x \in C(y)$ where $C(y)$ is the convex hull of $\{gy \mid g \in G\}$.

LEMMA 5.3. *If $f: R^n \to R^1$ is A-unimodal and G-invariant, then f is G-decreasing.*

PROOF. For $x \in C(y)$, we must show that $f(x) \geq f(y)$. But, the set $D \equiv \{z \mid f(z) \geq f(y)\}$ is convex and invariant. Since $y \in D$, $gy \in D$ for each $g \in G$ so $C(y) \subseteq D$ as D is convex. Thus $x \in D$ and the proof is complete.

THEOREM 5.3 (Mudholkar, 1966). *Suppose f_1 and f_2 are non-negative, A-unimodal and G-invariant. If $h(x) = \int f_1(x - y)f_2(y)\, dy$ is finite for each x, then h is G-decreasing.*

PROOF. Clearly h is G-invariant. Let $D_i(\gamma) = \{x \mid f_i(x) \geq \gamma\}$, $0 \leq \gamma < +\infty$. Since $f_i(x) = \int_0^\infty I_{D_i(\gamma)}(x)\, d\gamma$, we have

$$h(x) = \int_0^\infty \int_0^\infty \int_{R^n} I_{D_1(\tau)}(x - y)I_{D_2(\gamma)}(y)\, dy\, d\gamma\, d\tau.$$

Since the set of non-negative G-decreasing functions is a convex cone, it suffices to show that

$$h_0(x) = \int_{R^n} I_{D_1(\tau)}(x - y)I_{D_2(\gamma)}(y)\, dy$$

is G-decreasing. Of course, $D_1(\tau)$ and $D_2(\gamma)$ are convex G-invariant sets. Hence h_0 is G-invariant. But, Lemma 5.2 implies that h_0 is also A-unimodal. By Lemma 5.3, h_0 is G-decreasing and the proof is complete.

Recall that f on R^n to R^1 is called *symmetric* if $f(x) = f(-x)$ for all $x \in R^n$.

THEOREM 5.4 (Anderson, 1955). *If f_1 and f_2 are non-negative, symmetric and A-unimodal and if*

$$h(x) \equiv \int_{R^n} f_1(x-y) f_2(y)\, dy$$

exists for each x, then $h(ax_0)$ is a symmetric unimodal function of $a \in R^1$ for each $x_0 \in R^n$.

PROOF. Consider $G = \{I_n, -I_n\}$ acting on R^n, where I_n is the $n \times n$ identity matrix. Clearly, $C(x) = \{\lambda x + (1-\lambda)(-x) \mid 0 \le \lambda \le 1\}$. A function h_0 defined on R^n is G-decreasing iff h_0 is symmetric and $a \to h_0(ax_0)$ is decreasing on $[0, \infty)$ for each $x_0 \in R^n$. The result now follows immediately from Theorem 5.3.

REMARK 5.1. Here, we briefly describe an extension of Anderson's Theorem due to Sherman (1955). For a real valued function on R^n, let $\|f\|_1 = \int |f(x)|\, dx$, let $\|f\| = \sup_x |f(x)|$, and let $\|f\|_3 = \max\{\|f\|_1, \|f\|\}$. Also, let $\mathscr{C}_3$ denote the closure under $\|\cdot\|_3$ of the convex cone of functions generated by the indicator functions of compact convex symmetric sets in R^n. Thus, functions of the form

$$(*) \qquad f(x) = \textstyle\sum_{i=1}^m a_i I_{C_i}(x)$$

for $a_i \ge 0$ and compact convex symmetric C_i, $i = 1, \cdots, m$ are all in $\mathscr{C}_3$. For compact convex symmetric sets C_1 and C_2, the function

$$\phi(y) = \int_{R^n} I_{C_1}(y-x) I_{C_2}(x)\, dx$$

is log concave, continuous, and has compact support. Thus, for each $u > 0$,

$$A(u) = \{y \mid \phi(y) \ge u\}$$

is a compact convex set. It is not difficult to show that

$$\lim_{\varepsilon \to 0} \| \varepsilon^{-1} \textstyle\sum_{j=1}^\infty I_{A(j\varepsilon)}(\cdot) - \phi(\cdot) \| = 0$$

and the same holds for $\|\cdot\|_3$. Thus $\phi \in \mathscr{C}_3$. The continuity of convolution in $\|\cdot\|_3$ shows that the convolution of the $\mathscr{C}_3$ functions is in $\mathscr{C}_3$ (Sherman, 1955). However, the convolution of two symmetric A-unimodal functions need not be A-unimodal (see Anderson, 1955). Using similar arguments, it can also be shown that the marginal of $\mathscr{C}_3$ function is again a $\mathscr{C}_3$ function, on a lower dimensional Euclidean space.

Theorems 5.1 and 5.2 have lead to a number of interesting and useful generalizations; for example, see Borell (1975) and Rinott (1976). In the statistical literature, the main applications of Theorems 5.3 and 5.4 have been to establish properties of testing procedures such as unbiasedness and monotonicity of power functions. For some typical applications, see Das Gupta et al. (1964) and Eaton and Perlman (1974). Ordinarily, Theorems 5.3 and 5.4 are stated as follows:

If f_0 is a G-invariant A-unimodal density on R^n and if A is a convex G-invariant set, then $h(\theta) \equiv \int_A f_0(x-\theta)\, dx$ is a G-decreasing function.

Here, θ is a translation parameter and $h(\theta)$ is the probability of A when the density is $f_0(x - \theta)$.

Although Anderson's Theorem is stated in terms of a translation parameter, it does have applications to other types of problems. In what follows, we will establish a result due to Das Gupta, et al. (1972) which shows that probabilities of certain convex sets are monotone functions of the multiple correlation coefficient. A formal statement of this result follows. Let Σ be an $n \times n$ positive definite matrix and for $\lambda \in [0, 1]$, define Σ_λ by

$$\Sigma_\lambda = \begin{pmatrix} \Sigma_{11} & \lambda\Sigma_{12} \\ \lambda\Sigma_{21} & \sigma_{22} \end{pmatrix}$$

where Σ_{11} is $(n-1) \times (n-1)$. Since $\lambda \in [0, 1]$, Σ_λ is positive definite. Consider $Y \in R^n$ with a density of the form

$$|\Sigma_\lambda|^{-1/2} f(y' \Sigma_\lambda^{-1} y), \qquad y \in R^n. \tag{5.3}$$

THEOREM 5.5 (Das Gupta et al., 1972). *Let $C \subseteq R^{n-1}$ be a symmetric convex set and suppose $h \geq 0$. Partition Y as*

$$Y = \begin{pmatrix} \dot{Y} \\ Y_{n-1} \\ Y_n \end{pmatrix}, \qquad \dot{Y} \in R^{n-2}.$$

For $\lambda \in [0, 1]$, let

$$\Psi_1(\lambda) = P_\lambda\{(\dot{Y}, Y_{n-1})' \in C, \qquad |Y_n| \leq h\}.$$

Then Ψ_1 is non-decreasing.

Before beginning the proof, a few remarks are in order. The usual reduction to canonical correlations shows that there exists an $n \times n$ non-singular matrix A of the form

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & a_{22} \end{pmatrix}$$

with A_{11} being $(n-1) \times (n-1)$ and $a_{22} > 0$ such that

$$A \Sigma_\lambda A' = \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{0} \\ \hline & 1 & \lambda\rho \\ 0 & \lambda\rho & 1 \end{array}\right) \tag{5.4}$$

where $0 < \rho < 1$; the trivial case of $\rho = 0$ is omitted. Of course, ρ is the multiple correlation coefficient between $(\dot{Y}, Y_{n-1})$ and Y_n (assuming the coordinates of Y have second moments) when the density of Y is (5.3) with $\lambda = 1$. Thus the effect of varying λ is to vary the multiple correlation coefficient. Setting $Z = AY$, and absorbing ρ into λ, it suffices to show that

$$\Psi_2(\lambda) = P_\lambda\{(\dot{Z}, Z_{n-1})' \in A_{11}(C), |Z_n| \leq a_{22} h\} \tag{5.5}$$

is non-decreasing when Z has the density

$$|Q_\lambda|^{-1/2} f(z' Q_\lambda^{-1} z), \qquad z \in R^n,$$

where

$$Q_\lambda = \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{0} \\ \hline & 1 & \lambda \\ 0 & \lambda & 1 \end{array}\right), \qquad 0 \leq \lambda < 1.$$

Now, write $Q_\lambda = T_\lambda T_\lambda'$ where

$$T_\lambda = \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{0} \\ \hline & 1 & 0 \\ 0 & \lambda & \sqrt{1-\lambda^2} \end{array}\right), \qquad 0 \leq \lambda < 1,$$

and set $Z = T_\lambda X$ where X has the density $f(x' x)$ for $x \in R^n$. Then (5.5) becomes

$$\Psi_2(\lambda) = P\{(\dot{X}, X_{n-1}) \in C_1, |\lambda X_{n-1} + \sqrt{1-\lambda^2}\, X_n| \leq h_1\} \tag{5.6}$$

where $h_1 = a_{22} h > 0$, and $C_1 = A_{11}(C)$ is again a symmetric convex subset of R^{n-1}. The importance of the representation (5.6) is that the distribution of X is orthogonally invariant on R^n since the density of X is $f(x' x)$ on R^n. In what follows, it will be shown that Ψ_2 is non-decreasing when X has the uniform distribution on $S_{n-1} = \{x \mid \|x\| = 1, x \in R\}$. This case easily implies that Ψ_2 is non-decreasing for all X which have $\mathcal{O}_n$-invariant distributions.

We now proceed to the technical details. In what follows, all convex sets are assumed to be closed. This assumption does not affect the generality of Theorem 5.5 since the

Lebesgue measure of the boundary of a convex set is zero. Let μ_k denote the uniform probability measure on $S_{k-1} \subseteq R^k$, and let $B_k = \{x \mid x \in R^k, \|x\| \le 1\}$. For

$$w = \begin{pmatrix} \dot{w} \\ w_k \end{pmatrix} \in R^k, \qquad \dot{w} \in R^{k-1},$$

recall that for any integrable function q on S_{k-1},

$$\int_{S_{k-1} \cap \{w_k > 0\}} q(\dot{w}, w_k) w_k \mu_k(dw) = c(k) \int_{B_{k-1}} q(\dot{w}, \sqrt{1 - \|\dot{w}\|^2})\, d\dot{w} \tag{5.7}$$

where $c(k)$ is a fixed positive constant.

LEMMA 5.4. Let f on R^n to $[0, \infty)$ be a symmetric A-unimodal function with continuous partial derivatives. For each $x \in R^n$,

$$\int_{S_{n-1}} f(x + w) x' w \mu_n(dw) \le 0. \tag{5.8}$$

PROOF. Without loss of generality, we can take $x = t\varepsilon_n$ where $t > 0$ and ε_n is the nth standard unit vector in R^n. This reduction uses the $\mathcal{O}_n$-invariance of μ_n. With

$$w = \begin{pmatrix} \dot{w} \\ w_n \end{pmatrix}, \qquad \dot{w} \in R^{n-1},$$

it must be shown that

$$\Psi_3(t) \equiv \int_{S_{n-1}} f(\dot{w}, t + w_n) w_n \mu_n(dw) \le 0$$

for $t \ge 0$. Using a change of variable and (5.7), we have

$$\begin{aligned}
\Psi_3(t) &= c(n) \int_{B_{n-1}} \{f(\dot{w}, t + \sqrt{1 - \|w\|^2}) - f(\dot{w}, t - \sqrt{1 - \|w\|^2})\}\, d\dot{w} \\
&= c(n) \int_{B_{n-1}} \left\{ \int_{-\sqrt{1-\|\dot{w}\|^2}}^{\sqrt{1-\|\dot{w}\|^2}} (D_n f)(\dot{w}, t + u)\, du \right\} d\dot{w},
\end{aligned}$$

where D_n denotes partial differential with respect to the nth argument. Interchanging integration and differentiation, we have

$$\Psi_3(t) = c(n) \frac{\partial}{\partial t} \left\{ \int_{R^n} I_{B_n}(\dot{w}, u) f(\dot{w}, u + t)\, du d\dot{w} \right\}.$$

However, the expression inside the square brackets is the convolution of two symmetric A-unimodal functions evaluated at $t\varepsilon_n$ and hence is a decreasing function of t, $t \ge 0$ (Theorem 5.4). Thus, $\Psi_3(t) \le 0$ for each $t \ge 0$.

REMARK 5.2. Inequality (5.8) does have a geometric explanation. If f were the indicator of a convex symmetric set, say $f = I_C$, then the left hand side of the inequality is

$$\operatorname{Cov}\{I_{C-x}(W), x' W\}$$

where W is uniform on S_{n-1}. Now, $x' W$ is the component of W in the x-direction, and C has been translated so its center is at $-x$, in a direction opposite to x. Thus, when $x' W$ is large $I_{C-x}(W)$ tends to be small, and conversely. Hence it is not surprising that this covariance should be non-positive.

LEMMA 5.5. *Suppose U has the uniform distribution on S_{n-1} and partition U as*

$$U = \begin{pmatrix} \dot{U} \\ U_{n-1} \\ U_n \end{pmatrix}, \qquad \dot{U} \in R^{n-2}.$$

Let C be a convex symmetric set in R^{n-1} and let $h > 0$. For $\lambda \in [0, 1]$,

$$\Psi(\lambda) = P\{(\dot{U}, U_{n-1})' \in C, |\lambda U_{n-1} + \sqrt{1-\lambda^2}\, U_n| \leq h\}$$

is a non-decreasing function of λ.

PROOF. Let $f: R^n \to [0, 1]$ be the indicator function of C and let $g: R^1 \to [0, 1]$ be the indicator of $[-h, h]$. Then,

$$\Psi(\lambda) = \mathscr{E}\{f(\dot{U}, U_{n-1})g(\lambda U_{n-1} + \sqrt{1-\lambda^2}\, U_n)\}.$$

For $\varepsilon > 0$, let

$$f_\varepsilon(x) = \int_{R^{n-1}} (\sqrt{2\pi}\,\varepsilon)^{(n-1)} \exp\left(-\frac{1}{2\varepsilon^2}\|x-y\|^2\right) f(y)\, dy$$

and let

$$g_\varepsilon(z) = \int_{R^1} (\sqrt{2\pi}\,\varepsilon)^{-1} \exp\left\{-\frac{1}{2\varepsilon^2}(z-y)^2\right\} g(y)\, dy.$$

Then both f_ε and g_ε are symmetric A-unimodal functions, bounded and continuous with continuous partials and $\lim_{\varepsilon\to 0} f_\varepsilon(x) = f(x)$ for $x \neq \partial C$, $\lim_{\varepsilon \to 0} g_\varepsilon(z) = g(z)$ for $z \neq \pm h$. This implies that

$$\Psi_\varepsilon(\lambda) = \mathscr{E}\{f_\varepsilon(\dot{U}, U_{n-1}) g_\varepsilon(\lambda U_{n-1} + \sqrt{1-\lambda^2}\, U_n)\}$$

converges to $\Psi(\lambda)$ as $\varepsilon \to 0$. Thus it suffices to show that Ψ_ε is non-decreasing. But

$$\begin{aligned}\Psi_\varepsilon'(\lambda) &= \mathscr{E}[f_\varepsilon(\dot{U}, U_{n-1})\{U_{n-1} - \lambda(1-\lambda^2)^{-1/2} U_n\} g_\varepsilon'(\lambda U_{n-1} + \sqrt{1-\lambda^2}\, U_n)] \\ &= (1-\lambda^2)^{-1/2} \mathscr{E}\{f_\varepsilon(\dot{V}, \sqrt{1-\lambda^2}\, V_{n-1} + \lambda V_n) V_{n-1} g_\varepsilon'(V_n)\},\end{aligned}$$

where

$$V = \begin{pmatrix} \dot{V} \\ V_{n-1} \\ V_n \end{pmatrix} = \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{0} \\ \hline 0 & \sqrt{1-\lambda^2} & -\lambda \\ & \lambda & \sqrt{1-\lambda^2} \end{array}\right) \begin{pmatrix} \dot{U} \\ U_{n-1} \\ U_n \end{pmatrix}.$$

Since the transformation from U to V is orthogonal, V has the uniform distribution on S_{n-1}. Now, we argue with $V_n \in (-1, 1)$ fixed. First, $V_n g_\varepsilon(V_n) \leq 0$ since g_ε is a symmetric A-unimodal function on R^1. Conditional on V_n, $(\dot{V}, V_{n-1})$ has a uniform distribution on $\{x \mid x \in R^{n-1}, \|x\|^2 = 1 - V_n^2\}$. A direct application of Lemma 5.4 shows that

$$V_n \mathscr{E}\{f_\varepsilon(\dot{V}, \sqrt{1-\lambda^2}\, V_{n-1} + \lambda V_n) V_{n-1} \mid V_n\} \leq 0.$$

This implies that $g_\varepsilon'(V_n)$ times the above conditional expectation is non-negative so Ψ_ε' is non-negative. This completes the proof.

To complete the proof of Theorem 5.5, we will show that $\Psi_2(\lambda)$ given in (5.6) is non-decreasing. Recall that $X \in R^n$ has an $\mathscr{O}_n$-*invariant* distribution if (i) $P(X = 0) = 0$ and (ii) $\mathscr{L}(X) = \mathscr{L}(\Gamma X)$ for all $\Gamma \in \mathscr{O}_n$. (Condition (i) is to void some annoying technical problems).

FACT 5.1. The random vector $X \in R^n$ has an $\mathscr{O}_n$-invariant distribution iff $\mathscr{L}(X) = \mathscr{L}(RU)$ where U is uniform on S_{n-1}, R is a positive random variable, and R is independent of U.

Since the X occurring in (5.2) has an $\mathcal{O}_n$-invariant distribution, we can use Fact 5.1 to write

$$\Psi_2(\lambda) = \mathscr{E}P\left\{(\dot{U}, U_{n-1}) \in \frac{1}{R} C, \left|\lambda U_{n-1} + \sqrt{1-\lambda^2}\, U_n\right| \le \frac{h}{R} \,\middle|\, R\right\}.$$

But, for each R, the above conditional probability is non-decreasing in λ, by Lemma 5.5, so Ψ_2 is non-decreasing. Thus Theorem 5.5 is proved.

REMARK 5.3. The above proof of Theorem 5.5 is a minor modification of the original proof given in Das Gupta et al. (1972).

A key step in the above proof is the representation of $\mathscr{L}(Y)$ as $\mathscr{L}(HX)$ where H is a non-singular $n \times n$ matrix and X has a spherical distribution. We will now discuss some possible generalizations of Theorem 5.5 when $\mathscr{L}(Y)$ has such a representation. Partition Y into $\dot{Y} \in R^p$ and $\ddot{Y} \in R^q$ so $p + q = n$. Let $C_1 \subseteq R^p$ and $C_2 \subseteq R^q$ be two convex symmetric sets, and assume without loss of generality, that $q \le p$. When $\mathscr{L}(Y) = \mathscr{L}(HX)$ where X has a spherical distribution, the distribution of Y depends on H only through HH' since $\mathscr{L}(X) = \mathscr{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}_n$. Now, write HH' in its canonical correlation form:

$$HH' = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I_p & \begin{pmatrix} 0 \\ D \end{pmatrix} \\ (0, D) & I_q \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix},$$

where A_{11} is $p \times p$ and A_{22} is $q \times q$ with both nonsingular and D is a $q \times q$ diagonal matrix with diagonal elements $1 > \theta_1 \ge \cdots \ge \theta_q \ge 0$. This implies that

$$H = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ (0, D) & (I_q - D^2)^{1/2} \end{pmatrix} \Gamma_0$$

where $\Gamma_0 \in \mathcal{O}_n$ and $(I_q - D^2)^{1/2}$ is a $q \times q$ diagonal matrix with diagonal elements $(1 - \theta_i^2)^{1/2}$, $i = 1, \cdots, q$. Since X has a spherical distribution, we have

$$(5.8) \qquad P(\dot{Y} \in C_1, \ddot{Y} \in C_2) = P\{\dot{X} \in \tilde{C}_1, (0, D)\dot{X} + (I_q - D^2)^{1/2}\ddot{X} \in \tilde{C}_2\} \equiv \Psi(\theta)$$

where $\tilde{C}_i = A_{ii}^{-1} C_i$, $i = 1, 2$, is a convex symmetric set. Here, θ is a q dimensional vector with elements $\theta_1, \cdots, \theta_q$ as defined above. The problem is to describe the behavior of Ψ. When $q = 1$, Fact 5.1 and Lemma 5.5 imply that Ψ is nondecreasing. This observation yields a minor improvement on Theorem 5.5 since a density for Y was not assumed in this discussion. When $q > 1$, virtually nothing is known about Ψ. If $q = 2$ and Y is multivariate normal, Pitt (1977) has proved $\Psi(\theta) \ge \Psi(0)$, but the method of proof does not extend. Of course, a result for the uniform distribution on S_{n-1} would carry over to all Y's, but even a result for the multivariate normal would be useful.

Although the distribution $\mathscr{L}(Y) = \mathscr{L}(HX)$ may not have mixed second moments, so Y will not have a covariance matrix, the numbers $\theta_1, \cdots, \theta_q$ do have a geometric interpretation. First, identify R^p with those vectors in R^n of the form $\begin{pmatrix} \dot{x} \\ 0 \end{pmatrix} \in R^n$ where $\dot{x} \in R^p$ and identify R^q with those vectors of the form $\begin{pmatrix} 0 \\ \ddot{x} \end{pmatrix}$ where $\ddot{x} \in R^q$. With this identification, regard R^p and R^q as subspaces of R^n. Then, $\theta_1, \cdots, \theta_q$ are the cosines of the angles between these two subspaces of R^n computed in the inner product $(\cdot, \cdot)$ given by

$$(x, y) = x'HH'y, \qquad x, y \in R^n.$$

Of course, when Y does have a covariance, $\theta_1, \cdots, \theta_q$ are the canonical correlations between $\dot{Y}$ and $\ddot{Y}$.

Although Theorem 5.5 is rather special, it does have a number of useful statistical applications. Some of these are given in Das Gupta et al. (1972). The particular parameterization chosen for Theorem 5.5 seems rather natural, but there are others of interest. For example, the generalization of Slepian's theorem given in Section 5 of Das Gupta et al. (1972) is expressed in terms of correlations as opposed to canonical correlations.

6. Representation Theorems. In some situations, it is possible to obtain a useful structural relationship between two probability measures P_1 and P_2 which satisfy

$$\int f(x)P_1(dx) \le \int f(x)P_2(dx), \qquad f \in \mathscr{F} \tag{6.1}$$

where $\mathscr{F}$ is some given convex cone of functions on a measurable space $(\mathscr{X}, \mathscr{B})$. For example, in Sections 3 and 4, $\mathscr{F}$ is the convex cone of non-negative bounded increasing (in a given pre-order) functions. The results of this section are ordinarily of little use in verifying (6.1), but do provide important theoretical information.

Before beginning the abstract discussion, we first look at an example. On R^1, a distribution P_1 is stochastically smaller than P_2 ($P_1 < P_2$) if (6.1) holds where $\mathscr{F}$ is all bounded non-negative increasing (in the usual order) functions. Let X_i have distribution P_i, $i = 1, 2$, so $P_1 < P_2$ iff

$$\mathscr{E}f(X_1) \le \mathscr{E}f(X_2), \qquad f \in \mathscr{F} \tag{6.2}$$

This condition involves only the marginal distributions of X_1 and X_2. However, suppose (X_1, X_2) had a joint distribution on R^2 such that the conditional distribution of X_1 given $X_2 = x_2$ had all its mass in $U(x_2) = \{x \mid x \le x_2\}$. Then, conditionally,

$$\mathscr{E}\{f(X_1) \mid X_2\} \le f(X_2) \quad \text{w.p.1} \tag{6.3}$$

since X_1 is in $U(X_2)$ w.p.1. Taking expectations of both sides of (6.3) yields (6.2). Thus, the above condition on the joint distribution is sufficient for $P_1 < P_2$. For simplicity, assume the distribution function of X_i, say F_i, is continuous, $i = 1, 2$. We will now argue that the above condition, i.e., the existence of a conditional distribution satisfying (6.3), is also necessary. First observe that $P_1 < P_2$ iff $F_1(x) \ge F_2(x)$ for all $x \in R^1$. Construct a joint distribution on R^2 as follows: (i) X_2 has distribution F_2, and (ii) given $X_2 = x_2$, $X_1 = F_1^{-1}(F_2(x_2))$ w.p.1.

When $F_1(x) \ge F_2(x)$, then $x \ge F_1^{-1}(F_2(x))$ so given $X_2 = x_2$, $X_1 \in U(x_2)$ w.p.1, and it is easily verified that X_1 has distribution F_1. For the case at hand we have shown that $P_1 < P_2$ iff there exists a distribution λ on R^2 with marginals P_1 and P_2 such that, given x_2, the conditional probability under λ of $U(x_2)$ is one; that is, $X_1 \le X_2$ w.p.1 under λ.

The obvious question is: To what extent can the above observation be carried over to other cases? In particular, can we obtain results similar to that above in the cases considered in Sections 3 and 4 of this paper. More generally, suppose that $\mathscr{X}$ is a complete separable metric space (a Polish space), $\mathscr{B}$ is the σ-algebra of open sets, and $\le$ is a pre-order on $\mathscr{X}$ such that

$$U = \{(x, y) \in \mathscr{X} \times \mathscr{X} \mid x \le y\}$$

is a closed subset of $\mathscr{X} \times \mathscr{X}$. In this case, the pre-order is called *closed*. Let $\mathscr{F}$ be the convex cone of measurable non-negative bounded increasing functions. For $P_1, P_2 \in \mathscr{M}$, write $P_1 < P_2$ if (6.1) holds. If λ is a probability on $\mathscr{X} \times \mathscr{X}$, then $Q_1(B) \equiv \lambda(B \times \mathscr{X})$ is the *first marginal* of λ and $Q_2(B) \equiv \lambda(\mathscr{X} \times B)$ is the *second marginal* of λ.

THEOREM 6.1. (Strassen, 1965; Kamae et al., 1977). *Under the above assumptions, the following are equivalent.* (i) $P_1 < P_2$, (ii) *There exists a probability* λ *on* $\mathscr{X} \times \mathscr{X}$ *with first marginal* P_1 *and second marginal* P_2 *such that* $\lambda(U) = 1$.

Essentially, this result follows from Theorem 11 in Strassen (1965). The above equivalence and other useful equivalences are given in Kamae et al. (1977) for the case that $\le$ is a partial order, but the same arguments apply to the pre-order case. Condition (ii) says that we can construct random variables $(X_1, X_2) \in \mathscr{X} \times \mathscr{X}$, with joint distribution λ, such that X_i has distribution P_i, $i = 1, 2$, and $X_1 \le X_2$ w.p.1.

REMARK 6.1. In the case of Proposition 3.1, the explicit construction of λ is easy once P_η and P_θ (with $\eta \le \theta$) are specified.

In Proposition 3.1, 3.2, and in Examples 4.2 to 4.6, the family of distributions increased, in the sense of (6.1), when the parameter vector increased in the specified pre-order. To discuss this type of situation in general, it is useful to introduce Markov kernels. Consider two Polish spaces $\mathscr{X}_1$ and $\mathscr{X}_2$ with σ-algebras $\mathscr{B}_1$ and $\mathscr{B}_2$. A *Markov kernel K* is a function defined on $\mathscr{B}_1 \times \mathscr{X}_2$ to $[0, 1]$ such that $K(\cdot \mid x_2) \in \mathscr{M}(\mathscr{X}_1)$ and $K(B \mid \cdot)$ is $\mathscr{B}_2$ measurable for each $B \in \mathscr{B}_1$. Given pre-orders $\leq_i$ on $\mathscr{X}_i$, $i = 1, 2$, a Markov kernel is called *increasing* if $x_2 \leq_2 y_2$ $(x_2, y_2 \in \mathscr{X}_2)$ implies that

$$\int_{\mathscr{X}_1} f(x)K(dx \mid x_2) \leq \int_{\mathscr{X}_1} f(x)K(dx \mid y_2), \qquad f \in \mathscr{F}_1, \tag{6.4}$$

where $\mathscr{F}_1$ is the convex cone of non-negative bounded increasing functions on $(\mathscr{X}_1, \leq_1)$. When (6.4) holds, we write $K(\cdot \mid x_2) < K(\cdot \mid y_2)$ in accordance with our previous usage of $<$ for elements of $\mathscr{M}(\mathscr{X}_1)$.

EXAMPLE 6.1. Take $\mathscr{X}_1 = R^n$, $\mathscr{X}_2 = (0, \infty)^n$ and $\leq_i$ to be the pre-order defined on $\mathscr{X}_i$ by the permutation group $\mathscr{P}_n$, $i = 1, 2$. Let $K(\cdot \mid \theta)$ be the probability measure of the Gamma shape family, $\theta \in \mathscr{X}_2$, defined in Example 4.3. The results of Section 4 show that $K(\cdot \mid \theta)$ is an increasing Markov kernel (also called a Schur family for the particular pre-order of this example). Most of the other examples in Section 4 can also be used to construct increasing Markov kernels.

REMARK 6.2. It is useful to interpret condition (ii) in Theorem 6.1 in terms of kernels. Suppose $(X_1, X_2) \in \mathscr{X} \times \mathscr{X}$ has the joint distribution λ such that $\lambda(U) = 1$ and λ has marginals P_1 and P_2. Then, the conditional distribution of X_1 given $X_2 = x_2$ (which exists since $\mathscr{X}$ is Polish), say $K(\cdot \mid x_2)$, can be chosen so that $K(U(x_2) \mid x_2) = 1$ where $U(x_2) = \{x_1 \mid x_1 \leq x_2\}$. By definition, we have

$$P_1(B) = \int K(B \mid x_2)P_2(dx_2), \qquad B \in \mathscr{B}.$$

Note that such a K satisfies

$$\int f(x_1)K(dx_1 \mid x_2) \leq f(x_2) \quad \text{w.p.1}, \qquad f \in \mathscr{F}$$

and integrating this (P_2) gives (i) of Theorem 6.1.

Now, suppose X_i, $i = 1, 2, 3$ are Polish spaces with pre-orders $\leq_i$, $i = 1, 2, 3$. Suppose $K_1(\cdot \mid \cdot)$ is an increasing kernel on $\mathscr{B}_1 \times \mathscr{X}_2$ and $K_2(\cdot \mid \cdot)$ is an increasing kernel on $\mathscr{B}_2 \times \mathscr{X}_3$. Then the kernel K_3 on $\mathscr{B}_1 \times \mathscr{X}_3$ defined by

$$K_3(B \mid x_3) = \int_{\mathscr{X}_2} K_1(B \mid x_2)K_2(dx_2 \mid x_3)$$

is easily shown to be an increasing kernel. This is another example of a composition result similar in content to Theorem 4.3 and to Proposition 3.4 in Karlin and Rinott (1980). An application of this is Example 4.6.

There are a number of important applications of and results pertaining to increasing kernels we will not discuss in detail here. For example, applications to stochastic processes includes the work of Daley (1969), Kamae et al. (1977), Harris (1977), Kamae and Krengel (1978), and Karlin and Rinott (1980). Alternative proofs of Theorem 3.2 which used properties of increasing kernels occur in Kemperman (1977) and Edwards (1978).

We now return to the general situation where $(\mathscr{X}, \mathscr{B})$ is a measurable space, $\mathscr{F}$ is an arbitrary convex cone of nonnegative bounded measurable functions. As usual, for P_1, P_2 $\in \mathscr{M}$, write $P_1 < P_2$ to mean

$$\int f(x)P_1(dx) \leq \int f(x)P_2(dx), \qquad f \in \mathscr{F}. \tag{6.5}$$

To provide a sufficient condition for (6.5) to hold, suppose there is a Markov kernel K which satisfies

$$P_1(B) = \int_{\mathcal{X}} K(B \mid x_2) P_2(dx_2), \qquad B \in \mathscr{B} \tag{6.6}$$

and

$$\int f(x_1) K(dx_1 \mid x_2) \le f(x_2) \quad \text{w.p.1 } (P_2), \qquad f \in \mathscr{F}. \tag{6.7}$$

Then, we have

$$\int f(x_1) P_1(dx_1) = \int\int f(x_1) K(dx_1 \mid x_2) P_2(dx_2) \le \int f(x_2) P_2(dx_2),$$

so $P_1 < P_2$. A kernel which satisfies (6.7) is called an *$\mathscr{F}$-dilation*. Thus, the existence of an $\mathscr{F}$-dilation so that (6.6) holds is sufficient for (6.5). In some instances not covered by Theorem 6.1, the converse is also true. That is, $P_1 < P_2$ implies the existence of a K so (6.6) and (6.7) hold. For example, see Strassen (1965), Meyer (1966, Chapter 11) and Alfsen (1971). Some recent results directly related to the theme of this section are given in Rüschendorf (1981).

Acknowledgment. I wish to thank J. H. B. Kemperman, David Lane and two referees for many useful comments.

REFERENCES

AHLSWEDE, R. and DAYKIN, D. E. (1979). An inequality for weights of two families of sets, their unions and intersections. *Z. Wahrsch. verw. Gebiete* **93** 183–185.

ALFSEN, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. Springer, New York.

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.

BENSON, C. T. and GROVE, L. C. (1971). *Finite Reflection Groups*. Bogden and Quigley, Tarrytown on Hudson, New York.

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.

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn-Minkowski and Prekopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equations. *J. Functional Anal.* **22** 366–389.

CONLON, J. C., LEÓN, R., PROSCHAN, F. and SETHURAMAN, J. (1977a). G-ordered functions with applications in statistics, I. Theory. Tech. Report M432, Dept. of Statist., Florida State Univ.

CONLON, J. C., LEÓN, R., PROSCHAN, F. and SETHURAMAN, J. (1977b). G-ordered functions with applications in statistics, II. Applications. Tech. Report M433, Dept. of Statist., Florida State Univ.

DALEY, D. J. (1968). Stochastically monotone Markov chains. *Z. Wahrsch. verw. Gebiete* **10** 307–317.

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

DAS GUPTA, S., ANDERSON, T. W. and MUDHOLKAR, G. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. *Ann. Math. Statist.* **35** 200–205.

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.

DAVIDOVIC, JU. S., KORENBLJUM, B. I. and HACET, B. I. (1969). A property of logarithmically concave functions. *Soviet Math. Doklady* **10**, No. 2, 477–480 (English Translation).

DHARMADHIKARI, S. and JOGDEO, K. (1976). Multivariate unimodality. *Ann. Statist.* **4** 607–613.

DUNN, O. J. (1958). Estimation of means of dependent variables. *Ann. Math. Statist.* **29** 1095–1111.

DUNN, O. J. (1959). Confidence intervals for the means of dependent, normally distributed variables. *J. Amer. Statist. Assoc.* **54** 613–621.

DUNNETT, C. W. and SOBEL, M. (1955). Approximations to the probability integral and certain percentage points to a multivariate analogue of Student's t-distribution. *Biometrika* **42** 258–260.

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

EATON, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41** 1223–1226.

EATON, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *Ann. Statist.* **2** 609–614.

EATON, M. L. (1975). Orderings induced on R^n by compact groups of linear transformations with applications to probability inequalities. (Preliminary Report). Tech. Report No. 251, School of Statist., Univ. of Minnesota.

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

EDWARDS, D. A. (1978). On the Holley-Preston inequalities. *Proc. Royal Soc. Edinburgh 78A*, 265–272.

ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38** 1466–1474.

FORTUIN, C. M., GINIBRE, J. and KASTELEYN, P. N. (1971). Correlation inequalities on some partially ordered sets. *Commun. Math. Phys.* **22** 89–103.

GLESER, L. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* **3** 182–188.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934, 1952). *Inequalities*, 1st ed., 2nd ed. Cambridge University Press, London.

HARRIS, T. E. (1977). A correlation inequality for Markov-processes in partially ordered state spaces. *Ann. Probab.* **5** 451–454.

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

HOLLEY, R. (1974). Remarks on the FKG inequalities. *Comm. Math. Phys.* **36** 227–231.

JOGDEO, K. (1977). Association and probability inequalities. *Ann. Statist.* **5** 495–504.

KAMAE, T. and KRENGEL, U. (1978). Stochastic partial ordering. *Ann. Probab.* **6** 1044–1049.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

KARLIN, S. (1956). Decision theory for Polya-type distributions. Case of two actions. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 115–129. Univ. of California Press, Berkeley.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498.

KARLIN, S. and RINOTT, Y. (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Ann. Statist.* **9** 1035–1049.

KARLIN, S. and RUBIN, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272–299.

KEMPERMAN, J. H. B. (1977). On the FKG-inequality for measures on a partially ordered space. *Indag. Math.* **39** 313–331.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1152.

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, New York.

MEYER, P. A. (1966). *Probability and Potentials.* Ginn (Blaisdell), Boston.

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.

NEVIUS, E., PROSCHAN, F. and SETHURAMAN, J. (1977). Schur functions in statistics. II. Stochastic majorization. *Ann. Statist.* **5** 263–273.

PERLMAN, M. (1980). Unbiasedness of multivariate tests: Recent results. In *Multivariate Analysis V*, 413–432. Ed. P. R. Krishnaiah, North-Holland, New York.

PERLMAN, M. and OLKIN, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist.* **8** 1326–1341.

PITT, L. (1977). A Gaussian correlation inequality for symmetric convex sets. *Ann. Probab.* **5** 470–474.

PITT, L. (1981). Positively correlated normal variables are associated. To appear in *Ann. Probab.*

PRÉKOPA, A. (1973). On logarithmic concave measures and functions. *Acta. Scient. Mat.* (Szeged) **34** 335–343.

PRESTON, C. J. (1974). A generalization of the KFG inequalities. *Comm. Math. Phys.* **36** 233–241.

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

RADO, R. (1952). An inequality. *J. London Math. Soc.* **27** 1–6.

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

RINOTT, Y. (1976). On convexity of measures. *Ann. Probab.* **4** 1020–1026.

RÜSCHENDORF, L. (1981). Ordering of distributions and rearrangement of functions. To appear in *Ann. Prob.*

SARKAR, T. K. (1969). Some lower bounds of reliability. Tech. Report #124. Dept. of Oper. Res. and Statist., Stanford Univ.

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.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

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

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

WONG, C. K. and YUE, P. C. (1973). A majorization theorem for the number of distinct outcomes in N-independent trials. *Discrete Math.* **6** 391–398.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
206 CHURCH STREET, S.E.
MINNEAPOLIS, MN 55455