CWI Tracts

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

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

CWI Tract 35

Lectures on topics in probability inequalities

M.L. Eaton

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: primary: 60E15, secondary: 62H05.

ISBN 90 6196 316 8

Printed in the Netherlands

Preface

The material in this book is based primarily on a set of lectures given at the University of Amsterdam in the first half of 1985. During my 1984-85 sabbatical leave from the University of Minnesota, I was fortunate enough to be visiting the Centrum voor Wiskunde en Informatica in Amsterdam. With the encouragement of Richard Gill and Piet Groeneboom and the interst of some others in Holland, it was agreed that I would give some lectures on topics in probability inequalities to be sponsored both by the University of Amsterdam and the Centrum voor Wiskunde en Informatica. After some discussion it was decided that the lectures would highlight the following topics:

(i) majorization results and their extensions to reflection groups

(ii) association and the FKG inequality

(iii) log concavity, Anderson's theorem, and related topics.

To a large extent the treatment of the material is self contained, although the examples sometime require a bit of specialized statistical knowledge. In particular the canonical form of the multivariate analysis of variance model is assumed known in two examples. However, most examples can be skipped without interrupting the general development.

This book consists of six chapters and an appendix devoted to some special topics in convexity. It is recommended that the reader begin with the appendix because the material there is assumed known in the six chapters. In the first and introductory chapter stochastic ordering, monotone likelihood ratio, and symmetric unimodality on the real line are reviewed as a preview to association, the FKG inequality and log concavity on R^n. In addition the Behrens-Fisher problem is discussed in detail as it provides a very natural setting in which majorization arises.

The basic facts about majorization are established in Chapter 2 using a geometric approach which was outlined in Eaton (1984). This approach was chosen because it generalizes naturally to other group induced orderings such as those induced by reflection groups which are discussed in Chapter 6. Also in Chapter 2 the Schur concave functions are characterized and the so-called Convolution Theorem of Marshall and Olkin (1974) is established.

Chapter 3 is devoted to a number of statistical applications of majorization and some related ideas. Decreasing reflection functions are introduced and applied to ranking problems. The Composition Theorem of Hollander et al. (1977) is proved and is used to show certain parametric families have the property that expectations of Schur concave functions are Schur concave in a parameter vector.

Log concavity and some of its implications are developed in Chapter 4. Applications in this chapter include a derivation of some new concentration inequalities for Gauss-Markov estimators and a discussion of the behavior of power functions of some invariant tests in the multivariate analysis of variance problem.

Association, the FKG inequality and a variety of connected ideas form the bulk of Chapter 5. Two examples involving the multivariate normal distribution are also given. The first concerns the unbiasedness of certain invariant tests in multivariate analysis of variance problems (Perlman and Olkin (1980)), and the second shows that the coordinates of a multivariate normal random vector are associated iff the elements of the covariance are all non-negative (Pitt (1982)).

In the final chapter, we present what appears to be a fruitful approach in trying to extend various majorization results to other orderings induced by compact groups. The theory is far from complete and in fact there are many interesting examples where important questions are unanswered.

The interdependence between the chapters is roughly this. Chapter 1 is background for the remaining chapters. Chapters 2, 3 and 6 form a unit based on the common theme of group induced orderings and related topics. Both Chapters 4 and 5 are pretty much self contained units and can be read independently. As stated earlier, the Appendix contains material which is assumed in all chapters.

I would like to thank Richard Gill and Piet Groeneboom for arranging the lectures that lead to this book. The opportunity to visit the Centrum voor Wiskunde en Informatica and to lecture at the University of Amsterdam is greatly appreciated. The sabbatical leave from the University of Minnesota and the supplemental salary support provided by the National

Science Foundation under NSF Grant DMS-83-19924 are here gratefully ackowledged.

Finally, I want to thank Ann Marie Ruggles whose skill at the word processor made the preparation of this manuscript far less painful than it might have been. It was a pleasure to work with her.

Morris L. Eaton
February, 1986

Notation

R^1	the real line
R^n	Euclidean coordinate space of all n-dimensional column vectors
x'	the transpose of a coordinate vector $x \in R^n$
$x \perp y$	the vectors x and y are perpendicular--that is, $x'y = 0$
$\mathcal{S}_p$	the vector space of all p×p real symmetric matrices
$\mathcal{O}_p$	the group of n×n orthogonal matrices
$\mathcal{P}_n$	the group of n×n permutation matrices
$\mathcal{D}_n$	the group of n×n diagonal matrices with 1 or -1 on the diagonal
F	a convex cone
F*	the dual cone of F
$\mathcal{R}$	a set of reflections
$\mathcal{F}$	the class of Schur concave functions
□	denotes end of proof, end of example, and end of remark
$\mathcal{E}$	expectation
i.i.d.	independent and identically distributed
$N(\mu,\sigma^2)$	the univariate normal distribution with mean μ and variance σ^2
$N_p(\mu,\Sigma)$	the p dimensional normal distribution with mean vector μ and covariance matrix Σ.
χ_k^2	a chi-squared random variable with k degress of freedom
$W(\Sigma,p,n)$	a Wishart distribution on $\mathcal{S}_p$ with n degrees of freedom and expectation of $n\Sigma$.
$\mathcal{L}(\cdot)$	the distributional law of "$\cdot$"

Contents

Chapter 1: Motivation

Many of the ideas associated with multivariate probability inequalities have their origins in related ideas on the real line, R^1. In this chapter, some of these ideas are reviewed with an emphasis on their extension to more than one dimension. A rather natural stochastic ordering exists on R^1 and is discussed in Section 1. That monotone likelihood ratio (MLR) implies this stochastic ordering is reviewed and an alternative interpretation of MLR, which has a multivariate analog, is given. Unimodality and related topics such as log concavity, are covered in Section 2. In Section 3, the Behrens-Fisher problem is used to provide one, of many, possible statistical motivations for the study of majorization.

Section 1: Stochastic Ordering on R^1

For random variables X and Y with distribution functions F and G, the following definition makes precise the rather intuitive idea that "X tends to be smaller than Y".

Definition 1.1: If $F(x) \geq G(x)$ for all $x \in R^1$, we say that X is stochastically smaller than Y, and write $X \leq_{st} Y$.

The condition

$$F(x) \geq G(x), \quad x \in R^1 \tag{1.1}$$

is easily seen to be equivalent to the condition

$$F(x-) = P\{X < x\} \geq P\{Y < x\} = G(x-), \quad x \in R^1 \tag{1.2}$$

since the continuity points of a distribution are dense in R^1. An alternative formulation of (1.1) which has natural extensions to R^n follows.

Proposition 1.1: The following are equivalent

(i) $X \leq_{st} Y$

(ii) $Ef(X) \le Ef(Y)$ for all non-decreasing f for which the expectations are defined.

Proof: To show (ii) implies (i), take f to be the indicator function of the open interval (x,∞). For (i) implies (ii), first consider a non-negative bounded f which is non-decreasing. For $x,u \in R^1$, define $H(u,x)$ by

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

Then,

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

Thus,

$$Ef(X) = \int_{-\infty}^{\infty} f(x)\,dF(x) = \int_0^\infty \int_{-\infty}^{\infty} H(u,x)\,dF(x)\,du.$$

Since f is non-decreasing, for each $u \in [0,\infty)$, $H(u,\cdot)$ is the indicator function of an interval (a,∞) or $[a,\infty)$. In either case, (1.1) and (1.2) show that

$$\int_{-\infty}^{\infty} H(u,x)\,dF(x) \le \int_{-\infty}^{\infty} H(u,x)\,dG(x) \tag{1.3}$$

Integrating (1.3) from 0 to ∞ establishes (ii). For other non-decreasing f's, (ii) follows by truncation, translation, and taking limits--the details of which are left to the reader. □

For translation families, say F_θ with $\theta \in R^1$, it is clear that $\theta_1 \le \theta_2$ implies $F_{\theta_1}(x) \ge F_{\theta_2}(x)$ for $x \in R^1$. (By a translation family, I mean $F_\theta(x) = F_0(x-\theta)$ for $x,\theta \in R^1$.). However, in more complicated situations, it is sometimes not so clear when (1.1) holds. One useful sufficient condition is directly related to monotone likelihood ratio (MLR).

Definition 1.2: For two non-negative functions (p_1,p_2) on R^1, the pair (p_1,p_2) has a MLR if $x \le y$ implies that

$$p_1(x)p_2(y) \ge p_1(y)p_2(x). \qquad (1.4)$$

Remark 1.1. The terminology monotone likelihood ratio arises from the observation that (1.4) is equivalent to the condition that the ratio $(p_2/p_1)(x)$ is non-decreasing--when there is no problem with p_1 vanishing. Of course, (1.4) is used in the definition so that points where p_1 vanishes do not cause a problem. □

Remark 1.2: It is more common for MLR to be defined for a non-negative function r of two real variables--say $r(x,\theta)$ for real x and θ. Then, r has a MLR if $x_1 \le x_2$ and $\theta_1 \le \theta_2$ implies

$$r(x_1,\theta_1)r(x_2,\theta_2) \ge r(x_1,\theta_2)r(x_2,\theta_1). \qquad (1.5)$$

However, it is more useful for our purposes to think of θ_1 and θ_2 fixed so (1.5) reduces to (1.4) if we set

$$p_i(x) = r(x,\theta_i), \quad i = 1,2. \quad □$$

Condition (1.4) has a likelihood interpretation which provides some motivation for a multivariate version of MLR to be discussed in Chapter 5. Assume that p_1 and p_2 are densities with respect to some dominating measure on R^1. Consider an observation pair (U,V) with U and V independent where one of the two following alternatives holds:

$$A(i) \begin{cases} U \text{ is from } p_1 \\ V \text{ is from } p_2 \end{cases}$$

$$A(ii) \begin{cases} U \text{ is from } p_2 \\ V \text{ is from } p_1 \end{cases}$$

To decide between A(i) and A(ii), look at the likelihood under each alternative. Condition (1.4) means that when $U \le V$, A(i) is more likely and when $V \le U$, A(ii) is more likely. In other words, (1.4) guarantees that the pairing "min(U,V) with p_1, max (U,V) with p_2" is more likely.

<u>Proposition 1.2</u>: Suppose p_1 and p_2 are densities with respect to a dominating measure μ defined on R^1. Let X (respectively Y) have the distribution determined by p_1 (respectively p_2). If (1.4) holds, then X $\le$st Y.

<u>Proof</u>: Condition (ii) of Proposition 1.1 will be verified for a non-decreasing function f. Define s on R^2 by

$$s(x,y) = f(y) - f(x)$$

so $s(x,x) = 0$ and $s(x,y) = -s(y,x)$. Then

$$\delta = Ef(Y) - Ef(X) = Es(X,Y) = \iint s(x,y)p_1(x)p_2(y)\mu(dx)\mu(dy) = \iint_{\{y>x\}} s(x,y)p_1(x)p_2(y)\mu(dx)\mu(dy) + \iint_{\{y<x\}} s(x,y)p_1(x)p_2(y)\mu(dx)\mu(dy).$$

In the second integral, interchange x and y, and use the relation $s(y,x) = -s(x,y)$ to obtain

$$\delta = \iint_{\{y>x\}} s(x,y)[p_1(x)p_2(y) - p_1(y)p_2(x)]\mu(dx)\mu(dy).$$

But on the set $\{y>x\}$, both $s(x,y)$ and the term in square brackets are non-negative. Thus, $\delta \ge 0$. □

Many examples of MLR arise in statistics in the form of parametric families as indicated in Remark 1.1. For example,

$$r(x,\theta) = \beta(\theta)h(x)\exp[\theta x], \quad x,\theta \in R^1$$

satisfies (1.5) for non-negative functions β and h. An alternative name for MLR is total positivity of order 2 (TP_2) which will arise in its multivariate version in Chapter 5. Further discussion of MLR, TP_2 and related topics can be found in Karlin (1968), Lehmann (1959), Marshall and Olkin (1979) and the references there in.

Section 2: Symmetric Unimodality on R^1

Some of the most common continuous distributions on R^1 such as the normal, Cauchy and double exponential, are symmetric and unimodal about their center. The following definition makes this notion precise.

Definition 1.3: A real valued function f defined on R^1 is unimodal about x_0 if the function $h(t) = f(x_0 + t)$, $t \in R^1$ is non-increasing for $t \in [0,\infty)$ and non-decreasing for $t \in (-\infty,0]$. If f is unimodal about x_0 and $h(t) = h(-t)$, $t \in R^1$, then f is a symmetric unimodal function about x_0. The function f is symmetric unimodal if f is symmetric unimodal about 0.

When a random variable $X \in R^1$ has a density p (with respect to Lebesque measure) which is symmetric unimodal, our intuition suggests that for fixed $a > 0$, the function

$$g(b) = P\{-a+b \le X \le a+b\} \tag{1.6}$$

should be decreasing for $b \in [0,\infty)$. That this is true is a special case of the following.

Proposition 1.3 (Wintner (1938)). Suppose f_1 and f_2 are two symmetric unimodal functions such that the convolution

$$g(y) = \int_{-\infty}^{\infty} f_1(y-x)f_2(x)dx$$

is well defined for each $y \in R^1$. Then g is symmetric unimodal.

<u>Proof</u>: The symmetry of g is easily checked. Thus, for $0 \le y_1 < y_2$, it must be shown that

$$\delta = g(y_2) - g(y_1)$$

is non-positive. With $c = \frac{1}{2}(y_1+y_2)$ and $b = \frac{1}{2}(y_2-y_1)$, write

$$\delta = \int_{c}^{\infty}[f_1(y_2-x) - f_1(y_1-x)]f_2(x)dx + \int_{-\infty}^{c}[f_1(y_2-x) - f_1(y_1-x)]f_2(x)dx.$$

Making the change of variable x to x-c, the first integral is

$$I_1 = \int_{0}^{\infty}[f_1(b-x) - f_1(-b-x)]f_2(x+c)dx.$$

With the change of variable x to -(x-c), the second integral is

$$I_2 = \int_{0}^{\infty}[f_1(b+x) - f_1(-b+x)]f_2(-x+c)dx.$$

Using the symmetry of f_1 and f_2, we then have

$$\delta = \int_{0}^{\infty}[f_1(b-x) - f_1(-b-x)][f_2(x+c) - f_2(c-x)]dx.$$

Because $b > 0$, the symmetric unimodality of f_1 implies that

$$f_1(b-x) - f_1(-b-x) \ge 0 \quad \text{for } x \in [0,\infty)$$

since $|b-x| \le |-b-x|$ for $x \in [0,\infty)$. Similarly,

$$f_2(x+c) - f_2(c-x) \le 0 \qquad \text{for } x \in [0,\infty)$$

and hence $\delta \le 0$. □

The above result is often paraphrased "the convolution of two symmetric unimodals is symmetric unimodal." It should be noted that this result is false without the symmetry assumption (see Gnedenko and Kolmogorov (1954), Appendix II). A direct application shows that g defined by (1.6) is symmetric unimodal. Just take f_2 to be the density p of X (assumed to be symmetric unimodal) and take f_1 to be the indicator function of the interval $[-a,a]$. Then, an easy calculation shows that g(b) is the convolution of f_1 and f_2 evaluated at b, so g is symmetric unimodal.

A particularly interesting class of symmetric unimodal functions on R^1 is the class of symmetric log concave functions.

Definition 1.4: A function f defined on R^1 to $[0,\infty)$ is log concave if for all $x,y \in R^1$ and $\alpha \in (0,1)$,

$$f(\alpha x + (1-\alpha)y) \ge f^{\alpha}(x)f^{1-\alpha}(y). \tag{1.7}$$

If (1.7) holds and if $f(x) = f(-x)$ for $x \in R^1$, then f is a symmetric log concave function.

Proposition 1.4: A symmetric log concave function, f, is symmetric unimodal.

Proof: For $0 \le y_1 < y_2$, it must be shown that $f(y_1) \ge f(y_2)$. But y_1 is in the interval $(-y_2,y_2)$ so $y_1 = \alpha(-y_2) + (1-\alpha)y_2$ for some $\alpha \in (0,1)$. Hence

$$f(y_1) = f(\alpha(-y_2) + (1-\alpha)y_2) \ \ge\ f^{\alpha}(-y_2)f^{1-\alpha}(y_2) = f(y_2)$$

where the last equality holds since $f(-y_2) = f(y_2)$. □

The reason for writing the log concavity condition in the form (1.7) is that f is allowed to take on the value 0. For example, the indicator function of any interval of R^1 is a log concave function. When f is strictly positive, (1.7) just means log f is a concave function.

There is a connection between MLR and log concavity. Schoenberg (1951) has shown that a non-negative function f on R^1 is log concave iff

$$r(x,\theta) = f(x-\theta), \quad x,\theta \in R^1$$

is TP_2 (that is, r satisfies (1.5)).

Attempts to generalize unimodality to higher dimensions have led to numerous results which have applications in statistics and probability. Some of these are discussed in Chapter 4. Of course, the extension to higher dimension of the definition of log concavity is immediate (just let x and y be vectors in Definition 1.4). This will be exploited at length in Chapter 4.

Section 3: The Behrens-Fisher Problem and Majorization

One version of the Behrens-Fisher problem goes as follows. Consider random variables $X_1, \ldots, X_{m+1}$ which are i.i.d. $N(\mu_1,\sigma_1^2)$ and $Y_1, \ldots, Y_{n+1}$ which are i.i.d. $N(\mu_2,\sigma_2^2)$, with the X's and Y's independent. Here, μ_i, σ_i^2, i = 1,2 are unknown parameters. The problem is to provide a confidence statement (perhaps approximate) about $\mu_1 - \mu_2$ with a specified confidence coefficient $1 - \alpha$. An intuitively appealing procedure is to look at $\bar{X} - \bar{Y}$ which is $N(\mu_1-\mu_2,\tau^2)$ where

$$\tau^2 = \frac{\sigma_1^2}{m+1} + \frac{\sigma_2^2}{n+1} .$$

To estimate τ^2, consider the sample variances s_1^2 and s_2^2 where

$$s_1^2 = \frac{1}{m} \sum_1^{m+1} (X_i - \bar{X})^2$$

with a corresponding expression for s_2^2. Then s_i^2 is an unbiased estimator of σ_i^2, i = 1,2 so

$$\hat{\tau}^2 = \frac{s_1^2}{m+1} + \frac{s_2^2}{n+1}$$

is an unbiased estimator of τ^2. Now, for fixed $c > 0$, if we could bound (above and below) the probability

$$\delta = P\{(\bar{X} - \bar{Y} - (\mu_1-\mu_2))^2 \le c\hat{\tau}^2\}, \tag{1.8}$$

then we would have bounds for the confidence interval

$$\bar{X}-\bar{Y}-\sqrt{c}\hat{\tau} \le \mu_1-\mu_2 \le \bar{X}-\bar{Y}+\sqrt{c}\hat{\tau}. \tag{1.9}$$

The random variable

$$W = \frac{\left(\bar{X}-\bar{Y}-(\mu_1-\mu_2)\right)^2}{\hat{\tau}^2} = \frac{\left[\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\tau}\right]^2}{\frac{\hat{\tau}^2}{\tau^2}} \tag{1.10}$$

is the ratio of the two independent random variables

$$Z = \frac{\left(\bar{X}-\bar{Y}-(\mu_1-\mu_2)\right)^2}{\tau^2} \tag{1.11}$$

and $\hat{\tau}^2/\tau^2$. Obviously Z has a chi-squared distribution with 1 degree of freedom. With

$$\lambda = \frac{(m+1)^{-1}\sigma_1^2}{(m+1)^{-1}\sigma_1^2 + (n+1)^{-1}\sigma_2^2},$$

The random variable $\hat{\tau}^2/\tau^2$ is

$$\hat{\tau}^2/\tau^2 = \lambda m^{-1}V_1 + (1-\lambda)n^{-1}V_2 \tag{1.12}$$

where $V_1 = ms_1^2/\sigma_1^2$ and $V_2 = ns_2^2/\sigma_2^2$. Clearly, V_1 has a χ_m^2 distribution and V_2 has a χ_n^2 distribution.

Proposition 1.5: Define weights w_i, $i = 1,\ldots, m+n$ by

$$w_i = \begin{cases} \frac{\lambda}{m}, & i = 1,\ldots, m \\ \frac{1-\lambda}{n}, & i = m+1,\ldots, n+m. \end{cases}$$

Then W given in (1.10) has the same distribution as

$$W_1 = \frac{Z}{\sum_1^{m+n} w_i U_i} \tag{1.13}$$

where $Z, U_1,\ldots, U_{m+n}$ are i.i.d. random variables each with a χ_1^2-distribution.

Proof: Since the χ_k^2 distribution can be represented as the k-fold convolution of the χ_1^2 distribution, (1.12) shows that $\hat{\tau}^2/\tau^2$ has the same distribution as $\Sigma_1^{m+n} w_i U_i$. The result of the proposition now follows from the expression (1.10) for W and (1.11) for Z. □

The above proposition shows that to bound (1.8), it is necessary to study

$$\psi(w) = P\left\{Z < c\left(\sum_1^{m+n} w_i U_i\right)\right\} \tag{1.14}$$

where $c > 0$ is a fixed constant and the vector w has coordinates $w_1,\ldots, w_{m+n}$ which satisfy $0 < w_i < 1$, $\Sigma w_i = 1$.

Proposition 1.6: The function ψ in (1.14) satisfies

(i) ψ is a symmetric function of its arguments

(ii) ψ is concave.

<u>Proof</u>: That (i) holds is clear since $U_1, \ldots, U_{m+n}$ are i.i.d. For (ii), first observe that for $t > 0$, the distribution function of Z, say

$$F(t) = P\{Z \leq t\},$$

is concave since the density of Z is decreasing on $(0,\infty)$. Therefore, for $\alpha \in (0,1)$, for fixed $U_1, \ldots, U_{m+n}$, and for weight vectors w and v in the domain of ψ,

$$F(\alpha c\Sigma w_i U_i + (1-\alpha)c\Sigma v_i U_i) \geq \alpha F(c\Sigma w_i U_i) + (1-\alpha)F(c\Sigma w_i U_i).$$

Taking the expectation of this inequality over $U_1, \ldots, U_{m+n}$ yields

$$\psi(\alpha w + (1-\alpha)v) \geq \alpha\psi(w) + (1-\alpha)\psi(v)$$

which is just the concavity of ψ. □

One consequence of Proposition 1.6 can be described as follows. Let P_{m+n} be the group of $(m+n)\times(m+n)$ permutation matrices (see the Appendix for the definition of a permutation matrix). For a weight vector w, let C(w) denote the convex set generated by all the gw where $g \in P_{m+n}$. For any vector $v \in C(w)$ the claim is that

$$\psi(v) \geq \psi(w). \tag{1.15}$$

To see this, observe that if $v \in C(w)$, then

$$v = \Sigma\alpha_g gw$$

where the sum extends over P_{m+n} and the non-negative weights α_g add up to 1. Since ψ is concave,

$$\psi(v) = \psi(\Sigma\alpha_g gw) \geq \Sigma\alpha_g \psi(gw).$$

But $\psi(gw) = \psi(w)$ by (i) of Proposition 1.6. Hence

$$\Sigma\alpha_g \psi(gw) = \psi(w)\Sigma\alpha_g = \psi(w)$$

so (1.14) holds. In particular, if

$$\alpha_g = \frac{1}{(m+n)!}, \quad g \in P_{m+n},$$

then v has all its coordinates equal to 1/(m+n) which shows that (1.14) is maximized when all the w_i are equal to 1/(m+n). Of course, this provides a tight upper bound for (1.8).

Lower bounds for (1.8) also follow from (1.15), but a discussion of this is postponed until Chapter 3. The important observation at this point is that $v \in C(w)$ implies (1.15). This suggests a monotonicity property of ψ relative to some partial ordering defined on the weight vectors. In fact, the above argument suggests that we define a relation among weight vectors given by $v \leq w$ iff $v \in C(w)$. This relation "≤" is exactly the definition of majorization adopted in the next chapter. Much of the material in Chapter 2 is devoted to characterizing and understanding the relation "≤". In Chapter 3, the lower bounds for (1.8) are given together with other applications and extensions of the results in Chapter 2.

A number of authors have written papers concerned with bounding (1.8) and related problems. Let $F_{p,q}$ denote a random variable with and F-distribution with (p,q) degrees of freedom. The argument above shows that

$$P\{F_{1,m+n} \leq c\} \geq \delta$$

where δ is given in (1.8)--in other words, the random variable W of (1.10) is stochastically larger than $F_{1,m+n}$. This result was originally obtained by Hsu (1938) along with the companion result that W is stochastically smaller than $F_{1,r}$ where $r = \min(m,n)$. (This result will be established in Chapter 3). Hajek (1962) extended Hsu's work using an argument very similar to the one given above. Mickey and Brown (1966) independently established Hajek's results with Hajek-type arguments. This work was generalized in Lawton (1965,1968) and later extended and modified in Eaton and Olshen (1972). The arguments in this last work use the full force of majorization.

Chapter 2: Majorization, Basic Results

Three basic topics in majorization are discussed in this chapter. First, majorization is defined geometrically and is then characterized in an analytically useful way. Next the functions which are decreasing in the majorization ordering are described. Finally, the so called convolution theorem (Marshall and Olkin (1974)), which has many applications, is proved.

Before beginning with the formal discussion, it should be noted that the development given here is somewhat different than in other treatments. A geometric definition (Definition 2.1) is used because it has very natural extensions to many other cases of interest (see Chapter 6). However, unlike the traditional analytic definition, the geometric definition is hard to check in practice. (The equivalence of the two definitions is established in this chapter). But, the most convincing argument for using the geometric definition is that the general theoretical development following from this geometric point of view also carries over to other important cases with only minor modifications. This geometric treatment of majorization is similar in spirit to the development in Rado (1952). For a discussion and history of majorization, the reader is referred to Chapter 1 of Marshall and Olkin (1979) where the traditional analytic definition of majorization (due to Hardy, Littlewood and Polya (1934)) is used.

Section 1: Majorization; Definition and Properties

The setting for our discussion is Euclidean n-dimensional space R^n whose elements are represented as column vectors. If $x \in R^n$, then x' denotes the transpose of x. Let P_n denote the group of n×n permutation matrices g (see the Appendix for a discussion of P_n). For $x \in R^n$, the set $\{gx | g \in P_n\}$ is the permutation orbit of x--often called the orbit of x when the context is clear. Thus, the orbit of x is just the set of vectors obtained by permuting the coordinates of x. When the coordinates of x are distinct, the orbit of x contains n! points.

For $x \in R^n$, C(x) denotes the convex hull of the orbit of x. Hence $u \in C(x)$ iff u has a representation

$$u = \sum_g \alpha_g gx$$

where the sum is over P_n and the real numbers α_g, $g \in P_n$, satisfy

$$0 \le \alpha_g, \quad \sum_g \alpha_g = 1.$$

The convex set C(x) is permutation invariant ($u \in C(x)$ iff $gu \in C(x)$) and satisfies

$$C(x) = C(gx), \quad x \in R^n, \quad g \in P_n \qquad (2.1)$$

since the orbit of x is the same as the orbit of gx for any $g \in P_n$.

Definition 2.1: A point x is majorized by y if $x \in C(y)$. Equivalently, y majorizes x when $x \in C(y)$.

When x is majorized by y, we write $x \le y$. Here are some basic properties of the relation $\le$.

Proposition 2.1: The following are equivalent:

(i) $x \le y$

(ii) $C(x) \subseteq C(y)$

(iii) $g_1 x \le g_2 y$ for some $g_1, g_2 \in P_n$.

Proof: That (ii) implies (i) is clear since $x \in C(x)$. For (i) implies (ii), observe that $x \le y$ means that $x \in C(y)$ so that $gx \in C(y)$ for $g \in P_n$. Thus, the convexity of C(y) implies that all convex combinations of the gx, $g \in P_n$ are in C(y)--that is, $C(x) \subseteq C(y)$. That (i) and (iii) are equivalent follows from (2.1) and the permutation invariance of C(u) for $u \in R^n$. □

Proposition 2.2: The relation $\le$ is transitive--that is, $x \le y$ and $y \le z$ implies $x \le z$. If $x \le y$ and $y \le x$, then x is in the orbit of y, and conversely.

Proof: By (ii) of Proposition (2.1), $x \le y$ and $y \le z$ implies $C(x) \subseteq C(y) \subseteq C(z)$ so $x \le z$. For the second assertion, we have $C(x) = C(y)$ from (ii) of Proposition 2.1. But the set of extreme points of $C(x)$ is just the orbit of x. Since $C(x) = C(y)$, x must be an extreme point of $C(y)$, so x is in the orbit of y. The converse is obvious. □

Now, we turn to the problem of giving an analytic description of the relation $\le$. In what follows, $\le$ is often called an ordering because of its geometric interpretation and the transitivity given in Proposition 2.2. The first observation is that $x \in C(y)$ iff

$$u'x \le \sup_{g \in P_n} u'gy \quad \text{for all } u \in R^n. \tag{2.2}$$

This equivalence follows directly from Proposition A.3 in the Appendix with A taken to be the orbit of y and $B = C(y)$. Thus the function

$$m[u,y] = \sup_{g \in P_n} u'gy \tag{2.3}$$

is important in understanding $\le$.

Proposition 2.3: For all $u,y \in R^n$, the function m in (2.3) satisfies

(i) $m[u,y] = m[y,u]$

(ii) $m[g_1u, g_2y] = m[u,y]$ for $g_1, g_2 \in P_n$

(iii) $m[u,\cdot]$ is convex for each u.

Proof: For (i), note that

$$u'gy = y'g'u. \tag{2.4}$$

Since P_n is a group and $g' = g^{-1}$, taking the sup over P_n of both sides of (2.4) yields (i). For (ii), we have

$$m[g_1u, g_2y] = \sup_g (g_1u)'gg_2y =$$
$$\sup_g u'g_1'gg_2y = \sup_g u'gy = m[u,y]$$

since as g ranges over P_n, so does $g_1'gg_2$ because P_n is a group. To prove (iii), first observe that for u and g fixed, the function

$$y \to u'gy, \qquad y \in R^n$$

is linear in y and hence convex. Since the supremum of convex functions is again convex, (iii) follows. □

That m completely characterizes ≤ is the content of the next result.

<u>Proposition 2.4</u>: The following are equivalent:

(i) $x \le y$

(ii) $m[u,x] \le m[u,y]$ for all $u \in R^n$

<u>Proof</u>: If (i) holds, then (2.2) yields

$$u'x \le m[u,y] \qquad \text{for all } u \in R^n.$$

Substituting g'u for u gives

$$u'gx \le m[g'u,y] = m[u,y] \tag{2.5}$$

where the last equality follows from (ii) of Proposition 2.3. Taking the sup over $g \in P_n$ of (2.5) yields (ii). Conversely, if (ii) holds, then clearly (2.2) holds so $x \in C(y)$. Hence $x \le y$ by definition. □

<u>Remark 1.1</u>: For y fixed, $m[\cdot,y]$ is easily seen to be the support function of C(y)--see Rockafeller (1970), p. 28 for the definition of the support function. That inequality (ii) in Proposition 2.4 is equivalent to $C(x) \subseteq C(y)$ is well known and a proof can be found in Rockafeller (1970) (Section 13). However, results in this chapter will be proved directly rather than relying on other sources. □

The result in Proposition 2.4 shows that m completely characterizes majorization. In addition, (ii) of Proposition 2.3 shows that m is determined by its values on the quotient space R^n/P_n (with points being identified iff they are in the same orbit). Thus, to describe majorization, it is sufficient to calculate m explicitly on some convenient representation of the quotient space R^n/P_n. It is this which is behind the technical development which follows.

Let $F \subseteq R^n$ be the set of x whose coordinates, say $\alpha_1, \ldots, \alpha_n$, satisfy $\alpha_1 \geq \ldots \geq \alpha_n$. Thus F consists of all vectors whose coordinates are ordered (from largest to smallest). Also, let τ be the function on R^n to F which maps any point u into the vector whose coordinates are the ordered coordinates of u. Since $\tau(u) = u$ for all $u \in F$, the map τ is onto. Of course, given $u \in R^n$, there is a $g \in P_n$ such that $gu = \tau(u)$. Hence the orbit of every point in R^n has a non-empty intersection with F.

<u>Proposition 2.5</u>: For $x,y \in R^n$, the following are equivalent:

(i) $x \leq y$

(ii) $\tau(x) \leq \tau(y)$

(iii) $m[\tau(u),\tau(x)] \leq m[\tau(u),\tau(y)]$ for $u \in R^n$.

<u>Proof</u>: After noting that $\tau(v)$ is in the orbit of v for each $v \in R^n$, the equivalence of (i), (ii), and (iii) follows easily from Propositions 2.2, 2.3 and 2.4. □

The above result shows that if m can be calculated explicitly on F, then the majorization ordering can be characterized. To this end, we have

<u>Proposition 2.6</u>: For $u,v \in F$,

$$\sup_{g \in P_n} u'gv = u'v \tag{2.6}$$

<u>Proof</u>: This result (due to Hardy, Littlewood and Polya (1934)) is proved by induction. For $n = 2$, let u have coordinates $\alpha_1 \geq \alpha_2$ and v have

coordinates $\beta_1 \geq \beta_2$. Since P_2 has only two elements, the assertion (2.6) is

$$\alpha_1\beta_2 + \alpha_2\beta_1 \leq \alpha_1\beta_1 + \alpha_2\beta_2$$

or equivalently that

$$\alpha_1(\beta_1-\beta_2) + \alpha_2(\beta_2-\beta_1) = (\alpha_1-\alpha_2)(\beta_1-\beta_2) \geq 0.$$

But this latter inequality is true since $\alpha_1-\alpha_2 \geq 0$ and $\beta_1-\beta_2 \geq 0$.

Assume the result is true for $k = 2,\ldots, n$ and consider the case of $n+1$. For $u,v \in F \subseteq R^{n+1}$ with coordinates $\alpha_1 \geq\ldots\geq \alpha_{n+1}$ and $\beta_1 \geq\ldots\geq \beta_{n+1}$, and for $g \in P_{n+1}$, we have

$$u'gv = \sum_1^{n+1} \alpha_i\tilde{\beta}_i \tag{2.7}$$

where $\tilde{\beta}_1,\ldots, \tilde{\beta}_{n+1}$ is some permutation of $\beta_1,\ldots, \beta_{n+1}$. Let j be the smallest index such that $\tilde{\beta}_j = \beta_1$. If $j = 1$, then

$$u'gv = \alpha_1\beta_1 + \sum_2^{n+1} \alpha_i\tilde{\beta}_i$$

and the induction hypothesis gives

$$\sum_2^{n+1} \alpha_i\tilde{\beta}_i \leq \sum_2^{n+1} \alpha_i\beta_i$$

which yields (2.6). If $j > 1$, the case of $n = 2$ shows that

$$\alpha_1\tilde{\beta}_1 + \alpha_j\tilde{\beta}_j \leq \alpha_1\tilde{\beta}_j + \alpha_j\tilde{\beta}_1 = \alpha_1\beta_1 + \alpha_j\tilde{\beta}_1.$$

Hence (2.7) is bounded above by

$$\alpha_1\beta_1 + \alpha_j\tilde{\beta}_1 + \sum_{\substack{i=2\\i\neq j}}^{n+1} \alpha_i\tilde{\beta}_i.$$

Applying the induction hypothesis yields

$$\alpha_j\tilde{\beta}_1 + \sum_{\substack{i=2\\i\neq j}}^{n+1} \alpha_i\tilde{\beta}_i \leq \sum_{i=2}^{n+1} \alpha_i\beta_i$$

which implies (2.6). □

Proposition 2.7: For $x,y \in F$, the following are equivalent:

(i) $x \leq y$

(ii) $u'(y-x) \geq 0$ for all $u \in F$.

Proof: Suppose (i) holds. Then for all $u \in F$, Proposition 2.6 shows that $m[u,x] = u'x$ and $m[u,y] = u'y$. Thus, by Proposition 2.4, $u'x \leq u'y$ for all $u \in F$ so (ii) holds. Conversely, if (ii) holds, then

$$m[u,x] = u'x \leq u'y = m[u,y] \quad \text{for all } u \in F.$$

Since m is bi-invariant,

$$m[gu,x] = m[u,x] \leq m[u,y] = m[gu,y]$$

for all $g \in P_n$ and $u \in F$. But, as g ranges over P_n and u ranges over F, gu ranges over R^n. Hence

$$m[v,x] \leq m[v,y] \quad \text{for all } v \in R^n$$

which implies $x \leq y$. □

An alternative way to state Proposition 2.7 is in terms of the dual cone to F which is denoted by F^*. Recall (see the Appendix) that

$$F^* = \{w|u'w \geq 0 \quad \text{for all } u \in F\}.$$

Proposition 2.8: For $x,y \in F$, the following are equivalent

(i) $x \leq y$

(ii) $y - x \in F^*$.

Proof: Just use Proposition 2.7. □

To translate Proposition 2.8 into a useable analytic criterion for $x \leq y$ when x and y are in F, recall Example A.2 from the Appendix. In the notation of that example, t_i is the vector whose ith coordinate is 1, whose (i+1)th coordinate is -1, and the remaining coordinates are 0, $i = 1,\ldots, n-1$. Thus

$$F = \{x|t_i'x \geq 0, \ i = 1,\ldots, n-1\}$$

and $\{t_1,\ldots, t_{n-1}\} = T$ is a frame for F^*. Also, e_i is the vector in R^n whose first i coordinates are 1 and the remaining coordinates are 0, $i = 1,\ldots, n$. As shown in Example A.2, the set $\{e_1,\ldots, e_n,-e_n\} = E$ is a frame for F.

Proposition 2.9: Define the set B by

$$B = \{x|e_i'x \geq 0, \ i = 1,\ldots, n-1, \ e_n'x = 0\}.$$

Then $B = F^*$.

Proof: First observe that $e_i't_j = \delta_{ij}$ for $i = 1,\ldots, n$ and $j = 1,\ldots, n-1$. If $x \in F^*$, then $x = \Sigma_1^{n-1}b_it_i$ with $b_i \geq 0$ since T is a frame for F^*. Thus, $e_j'x = b_j \geq 0$ for $j = 1,\ldots, n-1$ and $e_n'x = 0$ so $x \in B$. Conversely suppose $x \in B$. Since $\{t_1,\ldots, t_{n-1},e_n\}$ is a basis for R^n and $e_n'x = 0$, x can be written as $x = \Sigma_1^{n-1}a_it_i$, for some real numbers $a_1,\ldots, a_{n-1}$. But $0 \leq e_j'x = a_j$ so the a_j are non-negative. Hence $x \in F^*$. Thus $B = F^*$. □

Proposition 2.10: Let x,y be elements of F with coordinates $\alpha_1 \geq\ldots\geq \alpha_n$

and $\beta_1 \geq \ldots \geq \beta_n$. The following are equivalent:

(i) $x \leq y$

(ii) $\Sigma_1^k \alpha_i \leq \Sigma_1^k \beta_i$, $k = 1, \ldots, n-1$ and $\Sigma_1^n \alpha_i = \Sigma_1^n \beta_i$.

Proof: For $x,y \in F$, it has been shown that $x \leq y$ iff $y-x \in F^*$. From Proposition 2.9, $y-x \in F^*$ iff

$$e_i'(y-x) \geq 0, \ i = 1, \ldots, n-1 \text{ and } e_n'(y-x) = 0.$$

These n conditions are exactly those given in (ii). □

The equivalence of Definition 2.1 and the classical definition of majorization is now an easy consequence of Proposition 2.10. Consider vectors x and y in R^n with coordinates $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. Let $a_{(1)} \geq \ldots \geq a_{(n)}$ denote the coordinates of $\tau(x) \in F$ and $b_{(1)} \geq \ldots \geq b_{(n)}$ denote the coordinates of $\tau(y) \in F$.

Theorem 2.11: For $x,y \in R^n$, the following are equivalent:

(i) $x \leq y$

(ii) $\Sigma_1^k a_{(i)} \leq \Sigma_1^k b_{(i)}$, $k = 1, \ldots, n-1$ and $\Sigma_1^n a_{(i)} = \Sigma_1^n b_{(i)}$.

Proof. Since $x \leq y$ iff $\tau(x) \leq \tau(y)$, the equivalence follows immediately from Proposition 2.10. □

Condition (ii) of Theorem 2.11 provides an easily verifiable condition to determine whether or not $x \leq y$. This condition is most useful in examples.

Section 2: The Path Lemma; Decreasing Functions

Many inequalities can be established by proving a certain function defined on R^n is decreasing (equivalently, Schur concave) in the majorization ordering. Here is the formal definition.

Definition 2.2: A real valued function f defined on R^n is decreasing in the majorization ordering if $x \leq y$ implies that $f(x) \geq f(y)$.

In what follows, we will simply say f is decreasing when the context is clear. Other writers call decreasing functions Schur concave functions because of Schur's basic contributions to the theory and applications of majorization (see Marshall and Olkin (1979) for a discussion). A function is increasing (Schur convex) if -f is decreasing.

Here are a couple of elementary facts.

Proposition 2.12: If f is decreasing, then f is P_n-invariant--that is, $f(x) = f(gx)$ for all $x \in R^n$ and $g \in P_n$. If f is P_n-invariant and concave, then f is decreasing.

Proof: The first assertion follows by noting that $x \le gx \le x$ for all $x \in R^n$ and $g \in P_n$. Hence if f is decreasing,

$$f(x) \ge f(gx) \ge f(x)$$

so f is P_n-invariant. For the second assertion, consider $x \le y$ so x has the representation

$$x = \sum_g \alpha_g gy$$

where the non-negative weights α_g add up to 1; that is, x is a convex combination of the gy, $g \in P_n$. The concavity of f implies

$$f(x) \ge \sum_g \alpha_g f(gy).$$

Since f is P_n-invariant, $f(gy) = f(y)$ for $g \in P_n$ and since the α_g add up to 1, we have $f(x) \ge f(y)$. □

A primary goal of this section is to provide some useful necessary and sufficient conditions that a function f be decreasing. Under a minor continuity assumption, the results of Marshall, Walkup and Wets (1967) discussed in the Appendix (see Theorem A.6) are directly applicable. To see this, first observe that decreasing functions f must be P_n-invariant and hence f is completely determined by its values on the convex cone F defined above. But on F (this is the convex set B in Theorem A.6), the majorization ordering is the same as the partial cone ordering defined by the convex cone F^* (Proposition 2.8). In other words, for $x,y \in F$, $x \leq y$ iff $y - x \in F^*$. Since F has a non-empty interior, Theorem A.6 yields

Proposition 2.13: Suppose f_1 defined on R^n is P_n-invariant and let f be the restriction of f_1 to F. Assume f_1 is continuous at the boundary of F. The following are equivalent

(i) f_1 is decreasing

(ii) For each frame vector t_i, $i = 1,\ldots, n-1$ of F^* and for each $x \in F$, the function $\beta \to f(x+\beta t_i)$ is decreasing for $\beta \geq 0$ as long as $x+\beta t_i \in F$.

Proof: If (i) holds, then f is decreasing on F and Theorem A.6 gives (ii). Conversely, given (ii), Theorem A.6 implies that f is decreasing on F. Since f_1 is P_n-invariant, this implies f_1 is decreasing on R^n. □

Proposition 2.14: (Ostrowski (1952)). Suppose f_1 defined on R^n is P_n-invariant and let f be the restriction of f_1 to F. Assume that f_1 has a differential on R^n. The following are equivalent

(i) f_1 is decreasing

(ii) For each x in the interior of F, say F^0,

$$\frac{\partial f}{\partial x_i}(x) \leq \frac{\partial f}{\partial x_{i+1}}(x), \quad i = 1,\ldots, n-1.$$

Proof: First assume (i). For each $x \in F^0$, and for each i, $f(x+\beta t_i)$ is decreasing in β for β in some non-degenerate interval $[0,\epsilon)$, and $\beta \to f(x+\beta t_i)$ is defined and differentiable in some interval (δ,ϵ) with $\delta < 0 < \epsilon$. Thus

$$\frac{d}{d\beta} f(x+\beta t_i)\Big|_{\beta=0} = \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_{i+1}}(x) \leq 0.$$

Hence (ii) holds. Conversely, if (ii) holds, we have that $\beta \to f(x+\beta t_i)$ is decreasing in β for $x \in F^0$ as long as $x+\beta t_i$ is in F. Thus, if $x \leq y$ with

$x,y \in F^0$, $f(x) \geq f(y)$. Since f_1 has a differential, f_1 is continuous so f is continuous on F. Thus, for any $x,y \in F$ with $x \leq y$, consider $u \in F^0$ so $x+\lambda u$ and $y+\lambda u$ are both in F^0 for $\lambda > 0$. Since $x+\lambda u \leq y+\lambda u$, the above argument gives

$$f(x+\lambda u) \geq f(y+\lambda u) \quad \text{for } \lambda > 0.$$

Letting $\lambda \to 0$ yields $f(x) \geq f(y)$ so f is decreasing on F. The P_n-invariance of f_1 implies that f_1 is decreasing. □

Slightly sharper results than those above concerning the decreasing functions can be obtained via a more detailed analysis of the majorization ordering. To motivate this analysis, recall the "path argument" used in the proof of Theorem A.6. In the present context, this "path argument" shows that if $x \leq y$ are both in F^0 (the interior of F), then there exists vectors $z_0, z_1, \ldots, z_m$ <u>all in F^0</u> with $z_0 = x$, $z_m = y$ and $z_{i+1} = z_i + \gamma_i u_i$ where $\gamma_i > 0$ and $u_i \in T = \{t_1, \ldots, t_{n-1}\}$--a frame for F^*. Obviously $z_i \leq z_{i+1}$ since $z_{i+1} - z_i \in F^*$. In fact, if we define

$$z_{i+1}(\beta) = z_i + \beta u_i; \quad 0 \leq \beta \leq \gamma_i$$

for $i = 0, \ldots, m-1$, then for $\beta_1 \leq \beta_2$, $z_{i+1}(\beta_1) \leq z_{i+1}(\beta_2)$. Thus, $z_{i+1}(\beta)$ defines a line segment connecting z_i and z_{i+1}, and as β increases, $z_{i+1}(\beta)$ increases in the majorization ordering. Denoting this "path" from x to y by $z_0 \to z_1 \to \ldots \to z_m$, we have a "monotone" path from x to y which lies in F. An important point is that the path remains in F. It is F where the ordering has been characterized. It was this path construction which provided Theorem A.6.

The above path construction fails when x and y are in the boundary of F with $x \leq y$. For example, take $n = 4$ and consider

$$y = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \quad x = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1 \end{pmatrix}.$$

Then $x \leq y$ by Theorem 2.1 and $x,y \in F$. But, it is easy to see that $x+\gamma t_i$,

$i = 1,2,3$ is not in F for any $\gamma > 0$. Thus, there can be no "path" in F from x to y if we are only allowed to start the path with vectors of the form $x+\gamma t_i$, $\gamma > 0$ where t_i is a frame vector for F^*. The construction below provides a path from x to y, which is in F, by enlarging the set of frame vectors. The technical details follow.

For integers i and j, with $1 \le i < j \le n$, let t_{ij} be the vector whose ith coordinate is 1, whose jth coordinate is -1, and whose remaining coordinates are zero. Observe that the n×n matrix

$$R_{ij} = I_n - t_{ij}t'_{ij} \tag{2.8}$$

is the element of P_n with permutes the ith and jth coordinate of $x \in R^n$. Further, let Δ be the set of all the t_{ij}, $1 \le i < j \le n$. It is easy to check that each $u \in \Delta$ is an element of F^*.

<u>Proposition 2.15</u>: Consider $x,y \in F$ with $x \ne y$ and $x \le y$. Then there exists a $u \in \Delta$ and a $\gamma > 0$ such that

(i) $x \le x+\gamma u \le y$

(ii) $x+\gamma u \in F$

(iii) The number of coordinates of $y - x$ which are zero is at least one less than the number of coordinates of $y-(x+\gamma u)$ which are zero.

<u>Proof</u>: Let the coordinates of x and y be $a_1 \ge \ldots \ge a_n$ and $b_1 \ge \ldots \ge b_n$ so

$$\begin{aligned} &\textstyle\sum_1^\alpha a_i \le \sum_1^\alpha b_i, \quad \alpha = 1,\ldots, n-1 \\ &\textstyle\sum_1^n a_i = \sum_1^n b_i \end{aligned} \tag{2.9}$$

Let j be the smallest index such that $a_j < b_j$ and let k be the largest index such that $a_k > b_k$. Such indices exist because $x \ne y$ and because of (2.9). From the definition of j and because of (2.9), $a_i = b_i$ for $i = 1,\ldots, j-1$. Also, we claim that

$$a_i = b_i \quad \text{for } i = k+1,\ldots, n \tag{2.10}$$

To verify this claim, observe that the definition of k implies $a_i \le b_i$ for $i = k+1, \dots, n$. But $\Sigma_1^k a_i \le \Sigma_1^k b_i$ so if $a_i > b_i$ for some $i = k+1, \dots, n$, we obtain a contradiction to $\Sigma_1^n a_i = \Sigma_1^n b_i$. Hence (2.10) holds. Now, define γ by

$$\gamma = \min\{b_j - a_j, a_k - b_k\} > 0$$

and pick $u = t_{jk} \in \Delta$. With these choices for γ and u, and the above remarks, it is now routine to verify (i), (ii) and (iii). □

<u>Proposition 2.16</u>: Consider $x, y \in F$ with $x \neq y$ and $x \le y$. Then there exists vectors $z_0, z_1, \dots, z_m$ in F such that

(i) $x = z_0 \le z_1 \le \dots \le z_m = y$

(ii) $z_{i+1} = z_i + \gamma_i u_i$ with $\gamma_i > 0$, $u_i \in \Delta$ for $i = 0, \dots, m-1$

(iii) $m \le n-1$.

<u>Proof</u>: First apply Proposition 2.15 to x and y to yield (with $z_0 = x$)

$$z_1 = z_0 + \gamma_0 u_0$$

satisfying (i), (ii) and (iii) of Proposition 2.15. If $z_1 = y$, we are done. If not, apply Proposition 2.15 to the pair z_1, y to yield

$$z_2 = z_1 + \gamma_1 u_1$$

satisfying (i), (ii) and (iii) of Proposition 2.15. Continuing this procedure until we have $z_m = y$ yields the claimed sequence of z's. Because of (iii) of Proposition 2.15, the procedure takes at most n steps. But, in fact, the procedure takes at most n-1 steps because $x \le y$ implies that y - x cannot have n - 1 zero coordinates; the sum of the coordinates of x and y are the same. □

The proof of Proposition 2.15 is a minor modification of the construction of so called T-transforms used by Muirhead (1903) and Hardy, Littlewood and Polya (1934). For a more complete discussion of T-transforms, see Marshall and Olkin (1979, p. 21).

It is now possible to give an alternative characterization of the decreasing functions. Recall that the vector $t_{12} \in \Delta$ has first coordinate 1, second coordinate -1, and the remaining coordinates are zero.

Proposition 2.17: Let f be a P_n-invariant function defined on R^n. The following are equivalent:

(i) f is decreasing

(ii) for each vector v which is perpendicular to t_{12} (that is, $v't_{12} = 0$), the function

$$\beta \to f(v+\beta t_{12}) \tag{2.11}$$

is decreasing on $[0,\infty)$.

Proof: First assume f is decreasing and consider v perpendicular to t_{12} (which we write as $v \perp t_{12}$ in what follows). For $0 \le \beta_1 \le \beta_2$, it must be shown that

$$f(v+\beta_1 t_{12}) \ge f(v+\beta_2 t_{12}). \tag{2.12}$$

But the matrix R_{12} defined in (2.8) is an element of P_n and satisfies

$$R_{12}v = v,\ R_{12}t_{12} = -t_{12}$$

Since $0 \le \beta_1 \le \beta_2$, we see that $v+\beta_1 t_{12}$ is a convex combination of $v+\beta_2 t_{12}$ and $R_{12}(v+\beta_2 t_{12}) = v-\beta_2 t_{12}$. Thus, $v+\beta_1 t_{12}$ is in $C(v+\beta_2 t_{12})$ and since f is decreasing, (2.12) holds.

Conversely, assume that (ii) holds and consider x and y with $x \le y$. It must be verified that

$$f(x) \ge f(y). \tag{2.13}$$

Since f is P_n-invariant, we can assume x and y are in F. Now, let $z_0 = x$,

$z_1, \ldots, z_m = y$ be the vectors given in Proposition 2.16. Since $z_i \le z_{i+1}$, it suffices to show that

$$f(z_i) \ge f(z_{i+1}) \tag{2.14}$$

for $i = 0, \ldots, m-1$. But from (ii) of Proposition 2.16, $z_{i+1} = z_i + \gamma_i u$ where $u \in \Delta$ and $\gamma_i > 0$ so it suffices to show that

$$\Psi(\gamma) = f(z_i + \gamma u) \tag{2.15}$$

is decreasing in γ for $\gamma \in [0,\infty)$. Since $u \in \Delta$, $u = t_{jk}$ for some integers j,k with $1 \le j < k \le n$. Write z_i as

$$z_i = v_i + \delta t_{jk} \tag{2.16}$$

where $v_i \perp t_{jk}$. Since $z_i \in F$, $0 \le t'_{jk} z_i = 2\delta$ so $\delta \ge 0$. Therefore, Ψ given in (2.15) is

$$\Psi(\gamma) = f(v_i + (\gamma+\delta)t_{jk}) \tag{2.17}$$

where $v_i \perp t_{jk}$. Now, let g be an element of $\mathcal{P}_n$ such that $gt_{jk} = t_{12}$--such a g clearly exists. Since f is invariant

$$\Psi(\gamma) = f(g(v_i + (\gamma+\delta)t_{jk})) = f(gv_i + (\gamma+\delta)t_{12}).$$

Because v_i is perpendicular to t_{jk} and g is an orthogonal transformation, gv_i is orthogonal to $gt_{jk} = t_{12}$. By (ii) with $v = gv_i$,

$$\beta \to f(gv_i + \beta t_{12})$$

is decreasing for $\beta \in [0,\infty)$. Since $\delta \ge 0$, this implies Ψ is decreasing on $[0,\infty)$, and the proof is complete. □

If f is $\mathcal{P}_n$-invariant, then for $v \perp t_{12}$ the function

$$\Psi(\beta) = f(v+\beta t_{12}), \quad \beta \in R^1 \tag{2.18}$$

is always symmetric on R^1. This follows from

$$f(v+\beta t_{12}) = f(R_{12}(v+\beta t_{12})) = f(v-\beta t_{12})$$

as in the proof of Proposition 2.17. Thus, another way to state this proposition is that f is decreasing iff Ψ given in (2.18) is unimodal (about 0)--the symmetry of Ψ is automatic when f is invariant. Since $v \perp t_{12}$, the first two coordinates of v are the same, say α. Let $\alpha_3, \ldots, \alpha_n$ be the remaining coordinates of v. Then (2.18) is just

$$\Psi(\beta) = f(\alpha+\beta, \alpha-\beta, \alpha_3, \ldots, \alpha_n) \tag{2.19}$$

and (ii) is that Ψ is decreasing on $[0,\infty)$. Of course, when f has a differential, we have

<u>Proposition 2.18</u>: Suppose f is P_n-invariant and f has a differential. The following are equivalent

(i) f is decreasing

(ii) $\frac{\partial f}{\partial x_1}(x) \le \frac{\partial f}{\partial x_2}(x)$ for all vectors x with $x_1 \ge x_2$ where x has coordinates $x_1, \ldots, x_n$.

<u>Proof</u>: Obvious from Proposition 2.17. □

We close this section with a few comments about doubly stochastic matrices which are discussed in the appendix. An early result in the development of majorization is

<u>Proposition 2.19</u> (Hardy, Littlewood and Polya (1929)). The following are equivalent:

(i) $x \le y$

(ii) $x = Qy$ for some doubly stochastic matrix Q.

<u>Proof</u>: That (i) implies (ii) is obvious from Definition 2.1 because a

convex combination of permutation matrices is doubly stochastic. The converse is a direct consequence of Birkhoff's Theorem in the Appendix. However a direct proof can also be given using the characterization of $\leq$ given in Theorem 2.11. For this direct proof, see the proof of Theorem A.4 in Marshall and Olkin (1979, p. 20). □

Section 3: The Convolution Theorem:

To motivate the main result of this section, suppose that f is a density function on R^n which is decreasing in the sense of Definition 2.2. Examples of such densities (which are most easily shown to be decreasing by using Proposition 2.17 or Proposition 2.18) are:

(i) $f(x) = h(\|x\|^2)$ where h is decreasing on $[0,\infty)$

(ii) $f(x) = \prod_1^n k(x_i)$ where k is a density on R^1 which is log concave

(iii) f corresponds to a normal distribution with mean 0 and covariance Σ which satisfies $\sigma_{ii} = \sigma_{11}$ for $i = 1,\ldots, n$ and $\sigma_{ij} = \sigma_{12}$ for $i \neq j$, $i,j = 1,\ldots, n$.

Consider a rectangular subset of R^n given by

$$A = \{x \mid c \leq x_i \leq d, \quad i = 1,\ldots, n\}$$

so A is permutation invariant and convex. Then the function

$$\Psi(\theta) = \int_A f(x-\theta)dx \tag{2.20}$$

is the probability of A as a function of the translation parameter $\theta \in R^n$. For example, A might be the acceptance region of a test designed to test $\theta = 0$ in which case $1 - \Psi(\theta)$ is the power function of the test. The problem is to say something about the behavior of the function Ψ as θ varies. Since A is $\mathcal{P}_n$-invariant and convex, it follows easily that the indicator function of A, say I_A, is a decreasing function on R^n. Thus,

$$\Psi(\theta) = \int I_A(x)f(x-\theta)dx = \int I_A(x)f(-(\theta-x))dx$$

which we recognize as the convolution of the two decreasing functions I_A and f_1 defined by $f_1(x) = f(-x)$. That f_1 is again a density and is decreasing is easily checked. Hence it seems natural to ask if Ψ, being the convolution of two decreasing functions, is itself a decreasing function (compare to the conclusion of Proposition 1.3). That this is in fact true was established in Marshall and Olkin (1974). This result immediately yields useful inequalities for the function Ψ. For example, over the set of θ's which satisfy $\Sigma\theta_i = 1$, Ψ is maximized when $\theta = \theta_0$ all of whose coordinates are n^{-1}. This is a consequence of $\theta_0 \le \theta$ for all θ's whose coordinates add up to 1 and the fact that Ψ is decreasing.

Here is the formal statement of one version of the convolution theorem.

Theorem 2.20: (Marshall and Olkin (1974)). Suppose f_1 and f_2 are Lebesgue measurable functions defined on R^n which are both decreasing in the sense of Definition 2.2. Suppose that $f_2 \ge 0$ and is integrable and suppose that f_1 is bounded. Then the convolution

$$h(y) = \int_{R^n} f_1(y-x)f_2(x)dx \tag{2.21}$$

is also decreasing.

Proof: The idea of the proof is to use Proposition 2.17 and Wintner's Theorem (Proposition 1.3). For simplicity of notation, set $t = t_{12}/\sqrt{2}$ where t_{12} is the vector in Proposition 2.17.

First observe that h is P_n-invariant since because f_1 and f_2 are and each $g \in P_n$ preserves Lebesgue measure. More specifically,

$$\begin{aligned} h(gy) &= \int f_1(gy-x)f_2(x)dx = \int f_1(y-g^{-1}x)f_2(x)dx \\ &= \int f_1(y-g^{-1}x)f_2(g^{-1}x)dx = \int f_1(y-x)f_2(x)dx = h(y). \end{aligned}$$

The second equality follows from the invariance of f_1, the third from the invariance of f_2, and the fourth because Lebesgue measure is invariant under the change of variable $x \to g^{-1}x$. Thus, to show h is decreasing it must be verified that for each vector $v \perp t$, the function

$$\beta \to h(v+\beta t) \tag{2.22}$$

is decreasing on $[0,\infty)$.

Let M be the linear subspace of R^n which is perpendicular to t, so R^n can be written as the orgthogonal direct sum

$$R^n = M + \text{span}(t). \tag{2.23}$$

Thus, each $x \in R^n$ can be written uniquely in the form

$$x = w + \gamma t, \quad w \in M, \ \gamma \in R^1$$

where $w \perp t$ since $w \in M$. Since the transformation from x to (w,t) is an orthogonal transformation, we have

$$h(v+\beta t) = \int_M \int_{-\infty}^{\infty} f_1(v+\beta t-w-\gamma t) f_2(w+\gamma t) d\gamma dw \tag{2.24}$$

where "dw" means Lebesgue measure on M. Since f_1 and f_2 are decreasing, Proposition 2.17 shows that for each $z \in M$,

$$\alpha \to f_i(z + \alpha t) \tag{2.25}$$

is a symmetric unimodal function on R^1 for $i = 1,2$.

Now, for v fixed, consider the function

$$K(w,\beta) = \int_{-\infty}^{\infty} f_1(v-w+(\beta-\gamma)t) f_2(w+\gamma t) d\gamma.$$

By assumption, the integral in (2.24) is well defined, so except for w in a Lebesgue null set $N \subseteq M$, $K(w,\beta)$ is a finite number. For $w \notin N$, (2.25) shows that $K(w,\cdot)$ is in fact the convolution on R^1 of two symmetric unimodal functions since $v-w \in M$ and $w \in M$. Thus, by Proposition 1.3, for $w \notin N$, $K(w,\cdot)$ is a symmetric unimodal function on R^1. Hence

$$h(v+\beta t) = \int_M K(w,\beta)dw = \int_{M\cap N^c} K(w,\beta)dw.$$

For $w \in M\cap N^c$, $K(w,\beta)$ is a decreasing function for $\beta \in [0,\infty)$ and hence so is the integral over $M\cap N^c$. Thus, the function given in (2.22) is decreasing on $[0,\infty)$ and the proof is complete. □

It should be emphasized that the key to the above proof is Wintner's Theorem together with the characterization of the decreasing functions given in Proposition 2.17. This technique is used again in a more general setting in Chapter 6.

Before extending Theorem 2.20 to functions f_1 which are unbounded, it is useful to discuss the class F of all $f:R^n \to R^1$ which are decreasing. A subset $B \subseteq R^n$ is called <u>monotone</u> if the indicator function I_B is in F. Hence a monotone set is necessarily invariant under permutations. Also, if B is P_n-invariant and convex, it follows easily that $I_B \in F$.

<u>Proposition 2.21</u>:

(i) A set B is monotone iff for each $x \in B$, $C(x) \subseteq B$.

(ii) A function f is decreasing iff for each $k \in R^1$, the set $B_k = \{x|f(x) \geq k\}$ is monotone.

<u>Proof</u>: The first assertion follows immediately from the definition of monotone. For the second assertion, first assume B_k is monotone for each k. To show f is decreasing, consider $x \leq y$ and pick $k = f(y)$. By (i), $x \in C(y) \subseteq B_k$ as $y \in B_k$ when $k = f(y)$. Thus,

$$f(x) \geq k = f(y)$$

so f is decreasing. Conversely, when f is decreasing, consider B_k and let $u \in B_k$. Then $f(u) \geq k$ so each $x \in C(u)$ satisfies $f(x) \geq f(u) \geq k$. Hence $C(u) \subseteq B_k$ and by (i), B_k is monotone. □

Now, observe that F is a convex cone of functions which is closed under minimum and maximum. That is, if k_1, k_2 are in F, then $h_1(x) =$

$\max\{k_1(x), k_2(x)\}$ and $h_2(x) = \min\{k_1(x), k_2(x)\}$ are also in F. Hence if $f_1 \in F$, then for $m = 1,2,\ldots,$

$$f_1^{(m)}(x) = \max\{-m, \min\{m, f_1(x)\}\}$$

is bounded in absolute value by m, is in F, and

$$f_1(x) = \lim_{m \to \infty} f_1^{(m)}(x).$$

To extend Theorem 2.20, consider a non-negative integrable function f_2 and assume that f_1 satisfies

$$\int |f_1(x)| f_2(y-x)dx < +\infty \tag{2.26}$$

for each $y \in R^n$. With $f_1^{(m)}$ as defined above, Theorem 2.20 shows that

$$h_m(y) = \int f_1^{(m)}(x) f_2(y-x)dx$$

is in F. Assumption (2.26) and the Dominated Convergence Theorem yield

$$\lim_{m \to \infty} h_m(y) = \int f_1(x) f_2(y-x)dx = h(y).$$

Since F is closed under the taking of pointwise limits, it follows that $h \in F$. Thus, Theorem 2.20 holds when (2.26) holds.

Section 4: Majorization on Subsets of R^n

In this section, we briefly discuss the validity of some of the previous results when majorization is restricted to a non-empty subset $X \subseteq R^n$. It is assumed that X is a P_n-invariant set--that is, $x \in X$ implies $gx \in X$ for all $g \in P_n$. A particularly important example of such an X is the set of all vectors in R^n whose coordinates are integers.

The first observation is that all the results and discussion in Section 1 are valid for any set X because these results concern only the meaning of

"$x \leq y$" which does not involve X. In other words, the discussion in Section 1 concerns two vectors x and y and the meaning of "$x \leq y$". However, there should be a minor modification in Proposition 2.3 because the appropriate domain of definition for m is $R^n x X$ so (ii) of Proposition 2.3 needs to be interpreted appropriately. With this one provision, Propositions 2.1 through and including Proposition 2.11 remain true as stated.

Now, again let X be a P_n-invariant subset of R^n and let f be a real valued function defined on X. As in Definition 2.2, f is decreasing (on X) if for $x,y \in X$, $x \leq y$ implies $f(x) \geq f(y)$. However, much care must be taken with regard to the validity of Propositions 2.12 through 2.18. In the next few paragraphs, some of the issues regarding these propositions will be discussed.

Since X is P_n invariant, the validity of the first assertion of Proposition 2.12 is clear. For the second assertion, X must be a set where concavity makes sense. For example, if X is a convex set, then the second claim is certainly true.

In the context of Proposition 2.13, the appropriate domain of definition of f is $X \cap F$ which is non-empty since X is non-empty and P_n invariant. However, $X \cap F$ must be convex and have a non-empty interior in order that Theorem A.6 apply. When $X \cap F$ does satisfy these assumptions, the argument used to prove Proposition 2.13 is valid since Theorem A.6 applies.

The function f in Proposition 2.14 is assumed to have a differential, and for this to make sense on X, the most natural assumption is that X is an open set. However, Proposition 2.13 is used in the proof of Proposition 2.14 and to use this, $X \cap F$ must be convex and have a non-empty interior. It may be possible to assume less than this, but things need to be checked very carefully.

The key idea in Propositions 2.15, 2.16 and 2.17 is the construction of a path from x to y which stays in F (when $x \leq y$). For the case at hand, the path must be constructed so that the nodes of the path (the vectors z_i in Proposition 2.16) stay in $X \cap F$. This depends crucially on the structure of X and needs to be checked in particular cases. But, there is one special case of great interest where these results are valid--namely when X is the set of all vectors in R^n whose coordinates are all integers (this X

is often called the <u>integer lattice</u> in R^n). The checking of the arguments in these Propositions for this particular X is routine and is left to the reader.

A discussion of Proposition 2.18 for a general set X is omitted. Proposition 2.19 is valid as it only involves a statement about $x \le y$.

We close this discussion with some remarks about the Convolution Theorem when X is the integer lattice in R^n. In this case X is an abelian group and counting measure on X is an invariant measure. That is, if μ denotes the measure which assigns measure 1 to each point of the countable set X, then for any subset $B \subseteq X$,

$$\mu(B+x) = \mu(B), \quad x \in X.$$

Let f_1 be a bounded function on X and let $f_2 \ge 0$ be integrable on the measure space (X,μ). The appropriate definition of the convolution is

$$h(y) = \int_X f_1(y-x)f_2(x)\mu(dx). \tag{2.27}$$

Suppose further that f_1 and f_2 are decreasing. The question is whether or not h need be decreasing. The answer is yes and a proof consists of "mimicking" that given for Theorem 2.20. The idea is to use the appropriate modification of Proposition 2.17 for the integer lattice and show that the function

$$\beta \to h(v+\beta t_{12}) \tag{2.28}$$

is decreasing on $\{0,1,2,\dots\}$ for each vector $v \in X$ which is perpendicular to t_{12}. The details are much the same as in the proof of Theorem 2.20, with one important exception. Namely, a discrete version of Wintner's Theorem is needed. Thus, suppose k_1 and k_2 are two symmetric real valued functions defined on $Z = \{0,\pm1,\pm2,\dots\}$ which are decreasing on $\{0,1,2,\dots\}$. Then the convolution

$$k_3(y) = \sum_x k_1(y-x)k_2(x)$$

is also symmetric and decreasing on {0,1,2,...}. The proof of this is much the same as the proof of Wintner's Theorem adapted to the case at hand. We note that an alternative proof of the Convolution Theorem for the integer lattice X is given in Chapter 3 as an application of the so called Composition Theorem.

Chapter 3: Majorization: Applications and Extensions

Many applications of majorization in statistics and probability provide at least partial solutions to the following rather general problem. Consider a parametric family of probability densities $f_1(x|\theta)$ on a space $\mathcal{X}$ (densities with respect to a fixed σ-finite measure μ) with $\theta \in \Theta$ and let h_1 be a real valued function. The problem is to "describe" the behavior of the function

$$\Psi_1(\theta) = \int h_1(x) f_1(x|\theta)\mu(dx) \tag{3.1}$$

Naturally, "solutions" to this problem require further assumptions and special structures on $\mathcal{X}$ and Θ. The convolution theorem (Theorem 2.20) is an example of a solution when $\mathcal{X} = \Theta = R^n$, $f_1(x|\theta) = f_0(x-\theta)$ is a translation family and both h_1 and f_0 are decreasing functions. In this case, Ψ_1 is a decreasing function which provides some information concerning the behavior of Ψ_1.

A related problem, which is connected to the discussion of the Behrens-Fisher problem given in Chapter 1, concerns linear combinations of real valued random variables. For example, suppose X is a random vector in R^n with a distribution function F. Given a real valued function h_2 defined on R^1, consider

$$\Psi_2(\theta) = \mathcal{E}h_2(\textstyle\sum_1^n \theta_i X_i) \tag{3.2}$$

where X has coordinates $X_1, \ldots, X_n$ and $\theta \in R^n$ has coordinates $\theta_1, \ldots, \theta_n$. Again, the problem is to describe the behavior of Ψ_2 under various assumptions on F when θ ranges over some set of interest. When $X_1, \ldots, X_n$ are i.i.d. random variables, and when h_2 is concave, the argument given to prove Proposition 1.6 shows that Ψ_2 is concave. Of course, this implies that Ψ_2 is decreasing in the sense of majorization.

Much of this chapter is devoted to results which use majorization and related notions to provide at least partial solutions to problems resembling those above. In Section 3.1, the Behrens-Fisher problem, introduced in Chapter 1, is discussed together with some related material.

An important generalization of the decreasing functions is introduced in Section 3.2 where the definition of the so called decreasing reflection (DR) functions is motivated. Some basic results concerning DR functions are given in Section 3.3 and a variety of applications appear in Section 3.4.

Many of the results and techniques presented here and in the previous chapter have extensions. Some of these extensions are treated, although rather incompletely, in Chapter 6. Thus, if some definitions and results appear to be phrased or formulated in what appears to be an unusual manner, the reader should keep in mind that the style of the presentation here has been selected with these extensions in mind.

Section 3.1. The Behrens-Fisher Problem and Related Topics

Recall that the problem introduced in Section 1.3 involved the study of a function

$$\Psi(w) = P\{Z \le c \sum_{1}^{m+n} w_i U_i\} \tag{3.3}$$

where $Z, U_1, \ldots, U_{m+n}$ are i.i.d. chi-squared random variables with one degree of freedom, c is a fixed positive constant and the weights w_i, i = 1,..., m+n are defined in Proposition 1.5. The problem is to give upper and lower bounds on Ψ. An immediate consequence of Proposition 1.6 is

Proposition 3.1: The function Ψ in (3.3) is decreasing (in the sense of Definition 2.2) on the set of w's which satisfy $0 \le w_i$ for i = 1,..., m+n and $w \ne 0$.

Proof: The proof of Proposition 1.6 shows that Ψ is P_n-invariant and concave on the given set of w's. The result follows from Proposition 2.12. □

Sharp bounds for (3.3) can now be had easily.

Proposition 3.2: Assume (without loss of generality) than $m \le n$. For any set of weights $w_1, \ldots, w_{m+n}$ as defined in Proposition 1.5, the following double inequality is valid:

$$P\{Z \le cm^{-1} \sum_1^m U_i\} \le \Psi(w) \le P\{Z \le c(m+n)^{-1} \sum_1^{m+n} U_i\}. \tag{3.4}$$

Proof: Consider the weight vectors $w^{(1)}$ and $w^{(2)}$ defined by: $w^{(1)}$ has its first m coordinates equal to m^{-1} and its remaining coordinates zero; $w^{(2)}$ has all its (m+n) coordinates equal to $(m+n)^{-1}$. Since Ψ is decreasing, (3.4) follows once it is verified that

$$w^{(2)} \le w \le w^{(1)} \tag{3.5}$$

for all weight vectors w as given in Proposition 1.5. However, the verification of (3.5) is routine if one uses Proposition 2.10. Thus (3.4) holds. □

The bounds given in (3.4) are obviously tight in the sense that there are weight vectors w under consideration which achieve both the upper and lower bounds. In this sense, Proposition 3.2 provides a complete solution to the problem posed in Chapter 1. The two inequalities in (3.4) are originally due to Hsu (1938).

A cursory examination of the proof of Proposition 1.6 immediately yields further results on decreasing functions. Recall that a random vector $X \in R^n$ has an exchangeable distribution (or X is exchangeable) if X and gX have the same distribution for all $g \in P_n$. Here is a result due to Marshall and Proschan (1965).

Proposition 3.3: Suppose $X \in R^n$, with coordinates $X_1, \ldots, X_n$, is exchangeable and let $H: R^n \to R^1$ be P_n-invariant and concave. For each vector $w \in R^n$ with coordinates $w_1, \ldots, w_n$ let

$$\Psi(w) = \mathcal{E}H(w_1X_1, \ldots, w_nX_n). \tag{3.5}$$

Then Ψ is P_n-invariant and concave, and hence decreasing.

<u>Proof</u>: For each w, let D(w) denote the n×n diagonal matrix with diagonal elements $w_1, \ldots, w_n$. For each $g \in \mathcal{P}_n$, it is easily verified that

$$D(gw) = gD(w)g'$$

and

$$D(aw+bv) = aD(w) + bD(v) \tag{3.6}$$

for $a,b, \in R^1$; $w,v \in R^n$. Now, (3.5) can be written

$$\Psi(w) = \mathcal{E}H(D(w)X) \tag{3.7}$$

Using the symmetry of H and the exchangeability of X, we have

$$\begin{aligned}\Psi(gw) &= \mathcal{E}H(D(gw)X) = \mathcal{E}H(gD(w)g'X) = \\ &= \mathcal{E}H(D(w)g'X) = \mathcal{E}H(D(w)X) = \Psi(w)\end{aligned}$$

so Ψ is invariant. The concavity of Ψ follows from (3.6) and the concavity of H--i.e., for $w,v \in R^n$ and $\alpha \in (0,1)$,

$$\begin{aligned}\Psi(\alpha w + (1-\alpha)v) &= \mathcal{E}H(D(\alpha w + (1-\alpha)v)X) \\ &= \mathcal{E}H(\alpha D(w)X + (1-\alpha)D(v)X) \\ &\geq \alpha\mathcal{E}H(D(w)X) + (1-\alpha)\mathcal{E}H(D(v)X) = \alpha\Psi(w) + (1-\alpha)\Psi(v).\end{aligned}$$

Thus Ψ is concave so by Proposition 2.10, Ψ is decreasing. □

In Proposition 3.3, Ψ need not be defined for all $w \in R^n$, but only in some symmetric convex subset of R^n--just so long as the expressions involved are well defined. The above argument remains valid. In fact, $X_1, \ldots, X_n$ need not even be real valued random variables, but can be random vectors--see Eaton and Olshen (1972) for an application. A particularly interesting function H is

$$H(x) = h\left(\sum_{1}^{n} x_i\right)$$

where x has coordinates $x_1, \ldots, x_n$ and h is a concave function on R^1. That this H satisfies the assumptions of Proposition 3.3 is easily checked.

Both the convolution theorem and Proposition 3.3 provide sufficient conditions that certain expectations of functions of a random vector be decreasing functions of a vector of parameters. In the convolution theorem, the parameters are translation parameters while in Proposition 3.3 the vector w is not a "traditional parameter", but in the context of Proposition 3.2, w is a vector of scale parameters. Before beginning the discussion of more general conditions under which the function defined in (3.1) is decreasing, we pause here to mention the problem of "coordinate systems." To wit, even though a P_n-invariant function $\Psi(\theta)$ may not be decreasing in $\theta \in R^n$, it is sometimes possible to introduce a new coordinate system say $\theta = h(\eta)$ with $\eta \in R^n$, so that the function

$$g(\eta) = \Psi(h(\eta))$$

is decreasing. An example will suffice.

<u>Example 3.1</u>: Consider independent random variables Y_1 and Y_2 with densities on $[0,\infty)$

$$f(y|\theta_i) = \frac{1}{\theta_i} \exp[-y/\theta_i], \quad y \geq 0$$

where $\theta_i > 0$, $i = 2$. Consider

$$\Psi(\theta) = P\{Y_1 + Y_2 \leq 1\}$$

for $\theta_i > 0$, $i = 1,2$. P. Diaconis has shown (unpublished work of 1976) that for $\theta_1 + \theta_2 \geq 1$, Ψ is decreasing, but for $\theta_1 + \theta_2 \leq 2/3$, Ψ is increasing. The method of proof is the verification of the derivative conditions in Proposition 2.18 and is not given here. However, if we set $X_i = \log Y_i$, i = 1,2, $\eta_i = \log \theta_i$, i = 1,2, and

$$g(\eta) = \Psi\left(e^{\eta_1}, e^{\eta_2}\right)$$

a bit of calculation shows that:

(i) The density of X_i is $h(x-\eta_i)$, $i = 1,2$ where

$$h(v) = e^v(\exp[-e^v]), \quad v \in R^1$$

(ii) The function g is given by

$$g(\eta) = \iint_{R^2} I_B(x_1,x_2)h(x_1-\eta_1)h(x_2-\eta_2)dx_1dx_2$$

where

$$B = \{u \in R^2 | e^{u_1} + e^{u_2} \le 1\}.$$

But, since B is a convex symmetric set, I_B is a decreasing function on R^2. Also, since h is a log concave function, it follows that $x \to h(x_1)h(x_2)$ is also decreasing on R^2. Since g is the convolution of two decreasing functions, g is decreasing. Obviously, this argument can be extended to more variables and other cases. The point of the above example is that sometimes a change of variable can yield a decreasing function from one which was not decreasing. The change of variable in this example was fairly obvious, but this is certainly not the case in general.

Section 3.2: Decreasing Reflection Functions; Motivation

In this section, we discuss an estimation problem which is intended to motivate the definition of a Decreasing Reflection Function given in the next section. In addition to having applications in ranking problems (see Eaton (1967)), these DR function play a role in showing functions of the form (3.1) are decreasing when neither Theorem 2.20 nor Proposition 3.3 are applicable. These applications are discussed in detail in the following two sections.

Consider independent random variables X_i, $i = 1,\ldots, n$ with $L(X_i) =$

$N(\theta_i, 1)$ $i = 1, \ldots n$--that is, the distribution of X_i is normal with mean θ_i and variance 1, $i = 1, \ldots, n$. Let X and θ be n-vectors whose coordinates are respectively $X_1, \ldots, X_n$ and $\theta_1, \ldots, \theta_n$. Then, the density of X given θ is

$$f(x|\theta) = \frac{1}{(\sqrt{2\pi})^n} \exp[-\frac{1}{2} \sum_1^n (x_i - \theta_i)^2] \tag{3.8}$$

where x an θ are in R^n. The vector $\theta \in R^n$ is assumed unknown and an observation $x \in R^n$ is available. The problem is to "estimate the order of the coordinates of θ." In other words, after seeing x, we are supposed to announce our guess concerning which coordinate of θ is largest, second largest,..., smallest. A more precise mathematical description of this problem is useful. In the notation of Chapter 2, let $F \subseteq R^n$ be the convex cone of vectors u whose coordinates satisfy $u_1 \geq \ldots \geq u_n$. Each parameter vector θ can be written

$$\theta = k\eta; \quad \eta \in F \quad \text{and} \quad k \in P_n. \tag{3.9}$$

Hence, η is the vector of ordered coordinates of θ and $k^{-1} = k'$ is a permutation matrix which puts the coordinates of θ in order--that is, $k'\theta = \eta$. The possible non-uniqueness of k caused by the equality of some of the coordinates of θ will not be an issue.

In terms of the parameterization $k\eta$ given in (3.9), the problem is to estimate k (or equivalently, k′) on the basis of the data x. Because of the symmetry in the density f in (3.8), it is intuitively clear that the answer to this estimation problem should be given by $k(x) \in P_n$ where

$$x = k(x)z; \quad z \in F, \quad k(x) \in P_n. \tag{3.10}$$

Namely, the estimated order of the coordinates of θ is just the observed order of the coordinates of x. In fact a bit more is true as is shown in Proposition 3.5, but before turning to this, it is useful to isolate two properties of f in (3.8).

Proposition 3.4: The density of f in (3.8) satisfies:

(i) $f(x|\theta) = f(gx|g\theta)$ for $x,\theta \in R^n$ and $g \in P_n$.

(ii) For integers j and k, $1 \le j \ne k \le n$, if x and θ satisfy $x_j \ge x_k$ and $\theta_j \ge \theta_k$, then

$$f(x|\theta) \ge f(x_{(j,k)}|\theta) \tag{3.11}$$

where $x_{(j,k)}$ is the vector x with its jth and kth coordinates interchanged.

Proof: That (i) holds is clear from the expression (3.8) for f. For assertion (ii), a bit of algebra shows that

$$\log f(x|\theta) - \log f(x_{(j,k)}|\theta) = (x_j - x_k)(\theta_j - \theta_k) \ge 0$$

so (3.11) holds. □

Proposition 3.5: For each $x \in R^n$ and $\eta \in F$,

$$f(x|k\eta) \le f(x|k(x)\eta), \quad k \in P_n \tag{3.12}$$

where k(x) is defined by (3.10). Thus, for each $\eta \in F$, the maximum likelihood estimator of k in the parameterization $\theta = k\eta$ is k(x).

Proof: Since $k^{-1}(x)x = z \in F$, (i) of Proposition (3.4) shows that (3.12) is equivalent to

$$f(k^{-1}x|\eta) \le f(k^{-1}(x)x|\eta) = f(z|\eta) \tag{3.13}$$

for each $k \in P_n$. Setting $y = k^{-1}x$, we see that y is a permutation of x so the ordered coordinates of y are just $z_1 \ge z_2 \ge \ldots \ge z_n$ which are the coordinates of z. Thus, it must be shown that

$$f(y|\eta) \le f(z|\eta). \tag{3.14}$$

Let $y_1, \ldots, y_n$ be the coordinates of y. If $y = z$, obviously (3.14) holds so assume $y \neq z$. Let $r(y,z)$ be the number of non-zero coordinates of $y-z$. Since $y \neq z$, $1 \leq r(y,z) \leq n$.

There are two cases.

Case 1: $y_1 < z_1$. In this case, let j be the largest index such that $y_j = z_1$ so $y_j > y_1$. Let $y^{(1)}$ be the vector with coordinates $y_1^{(1)} = y_j$, $y_j^{(1)} = y_1$, $y_i^{(1)} = y_i$ for $i \neq 1$, $i \neq j$. Applying (ii) of Proposition 3.4 with "$x = y^{(1)}$, $\eta = \theta$, $j = 1$ and $k = j$", we see that

$$\left.\begin{aligned} & f(y|\eta) \leq f(y^{(1)}|\eta) \\ \text{and} \quad & r(y^{(1)},z) \leq n - 1 \end{aligned}\right\} \tag{3.15}$$

Case 2: $y_1 = z_1$. In this case, set $y^{(1)} = y$ so again (3.15) holds.

Now, construct a vector $y^{(2)}$ by applying the Case 1 and Case 2 analysis to the second coordinates of $y^{(1)}$ and z. This yields

$$\left.\begin{aligned} & f(y^{(1)}|\eta) \leq f(y^{(2)}|\eta) \\ \text{and} \quad & r(y^{(2)},z) \leq n - 2 \end{aligned}\right\} \tag{3.16}$$

Applying this procedure n-1 times yields $y^{(1)}, \ldots, y^{(n-1)}$ such that

$$f(y|\eta) \leq f(y^{(1)}|\eta) \leq \ldots \leq f(y^{(n-1)}|\eta).$$

But $y^{(n-1)} = z$ since $r(y^{(n-1)},z) = 0$ and the proof is complete. □

The key to the above proof is the construction of the sequence $y^{(1)}, y^{(2)}, \ldots, y^{(n-1)}$. In fact, the proof shows that the conclusion of Proposition 3.5 remains valid for any function satisfying (i) and (ii) of Proposition 3.4. It is these two conditions which define a DR function. This is discussed carefully in the next section.

Section 3.3: Decreasing Reflection Functions: Basics

In order to give a proper definition and justify the term Decreasing Reflection (DR) function, a little notation is needed. In R^n, for any vector $u \neq 0$, let

$$R_u = I_n - 2\,\frac{uu'}{u'u} \tag{3.17}$$

where I_n is the n×n identity. The symmetry of R_u is clear, and the identity $R_u^2 = I_n$ is easily verified. Thus, R_u is a symmetric orthogonal matrix. Since

$$R_u u = -u, \text{ and } R_u v = v \tag{3.18}$$

for any $v \perp u$ (i.e., v is perpendicular to u), R_u has the geometric interpretation as the reflection across the hyperplane perpendicular to u--that is, across

$$H_u = \{v \mid u'v = 0\}.$$

In other words, R_u is the identity on H_u and is "minus the identity" on the one dimensional subspace span $\{u\}$. The term <u>reflection</u> is used for matrices of the form (3.17). Note that $R_{cu} = R_u$ for any real number $c \neq 0$. Also

$$R_{\Gamma u} = \Gamma R_u \Gamma'$$

for any n×n orthogonal matrix Γ.

The group $\mathcal{P}_n$ contains some reflections of particular interest. Let Δ_1 be the set of all vectors $u \in R^n$ which have (n-2) coordinates zero, one coordinate equal to one, and one coordinate equal to minus one. Notice that for each $u \in \Delta_1$, either $u \in \Delta$ or $-u \in \Delta$ where Δ is defined just after (2.8). Also, each R_{ij} in (2.8) is a R_u for some $u \in \Delta$ and conversely. The notation t_{12} is reserved for the particular vector in Δ given by

$$t_{12} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Now, let X and Y be symmetric subsets of R^n--that is, $gX = X$ for all $g \in P_n$ and the same for Y.

Definition 3.1: A real valued function f defined on $X \times Y$ is a Decreasing Reflection function (a DR function) if

(i) $f(x,y) = f(gx,gy)$ for $x \in X$, $y \in Y$, $g \in P_n$

(ii) for each $u \in \Delta$, $u'xu'y \geq 0$ implies that

$$f(x,y) \geq f(x,R_u y). \tag{3.20}$$

Before giving examples, the following result shows that (ii) need only be checked for $u = t_{12}$.

Proposition 3.6: Let f be a real valued function of $X \times Y$ which satisfies (i) of Definition 3.1. The following are equivalent:

(i) f is a DR function

(ii) the condition $t_{12}'xt_{12}'y \geq 0$ implies that

$$f(x,y) \geq f(x,R_{t_{12}} y). \tag{3.21}$$

Proof: Clearly (i) implies (ii) since $t_{12} \in \Delta$. Conversely assume (ii) holds and let $u \in \Delta$. Hence there exists a $g \in P_n$ such that $gu = t_{12}$. For $x \in X$ and $y \in Y$ such that $u'xu'y \geq 0$, it must be verified that (3.20) holds. For such an x and y,

$$\begin{aligned} u'xu'u &= u'g'gxu'g'gy = (gu)'gx(gu)'gy \\ &= t_{12}'(gx)t_{12}'(gy) \geq 0. \end{aligned}$$

Applying (3.21) with x replaced by gx and y replaced by gy, we have

$$f(gx,gy) \geq f(gx,R_{t_{12}} gy). \tag{3.22}$$

Using the assumed invariance of f, this yields

$$f(x,y) = f(gx,gy) \geq f(gx,R_{t_{12}}gy) = f(x,g'R_{t_{12}}gy).$$

But, from (3.19) with $g = \Gamma'$, it follows that

$$g'R_{t_{12}}g = R_{g't_{12}} = R_u$$

so (3.20) holds. □

The verification that a given function f defined on $X \times Y$ is a DR function is ordinarily most easily accomplished by checking that f is invariant (i.e., (i) of Definition (3.1) holds) and then verifying (ii) of Proposition 3.6. If x and y have coordinates $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, the condition $t_{12}'xt_{12}'y \geq 0$ simply means

$$(x_1-x_2)(y_1-y_2) \geq 0. \tag{3.23}$$

In other words, if $x_1 > x_2$ then $y_1 > y_2$ and if $x_1 < x_2$ then $y_1 < y_2$. When (3.23) holds, then (3.21) has to be verified. In the two classes of examples below (from Eaton (1967)), the verification is easy and is left to the reader.

Example 3.2: For symmetric subsets X and Y of R^n, consider f of the form

$$f(x,y) = \Psi_1(x)\Psi_2(y) \prod_{i=1}^{n} g(x_i,y_i) \tag{3.24}$$

where each Ψ_i is P_n-invariant and non-negative on R^n, $i = 1,2$ and g is a non-negative function of two real variables. The invariance of f is clear. If g has a MLR, then (ii) of Proposition 3.6 is easily checked. Conversely when Ψ_1, Ψ_2 and g are strictly positive, then (ii) of Proposition 3.6 holds iff g has a MLR. Thus Definition 3.1 can be thought of as one possible attempt at generalizing MLR. Examples of probability densities which can be written in the form (3.24) with x being the variable of the density and y being a vector of parameters include the multinomial density, the density of n independent Poisson random variables with different parameters, the density of n independent Gamma random variable with either different shape parameters (but the same scale parameter) or different scale parameters with the same shape parameter. □

Example 3.3: In this example, take $X = Y = R^n$ and consider an f of the form

$$f(x,y) = \Psi((x-y)'A(x-y))$$

where Ψ is a non-increasing function defined on R^1 and the n×n matrix A has the so-called intraclass correlation form:

$$A = \alpha I_n + \beta ee'$$

where e is the vector of ones in R^n and the scalars α and β satisfy $\alpha \geq 0$, $\alpha+n\beta \geq 0$. Because of the structure of A, f is invariant. Also, a calculation similar to that used in the proof of Proposition 3.4 shows that (3.21) holds. □

The analog of Proposition 3.5 for DR functions is

Proposition 3.7: Suppose f is a DR function defined on $X \times Y$. Fix $y \in X \cap F$. For $k \in P_n$ and $x \in X$,

$$\sup_{k \in P_n} f(kx,y) = f(k_0x,y)$$

where k_0 is any element of P_n such that $k_0x \in F$.

Proof: The proof is the same as the proof of Proposition 3.5. □

The connection between the DR functions and the decreasing functions (in the sense of Definition 2.2) follows. This result is due to Hollander et al. (1977).

Proposition 3.8: Suppose $X \subseteq R^n$ is a symmetric subset of R^n such that x,y

$\in \mathcal{X}$, implies that $x - y \in \mathcal{X}$. Let f_0 be a real valued function defined on $\mathcal{X}$. The following are equivalent:

(i) f_0 is decreasing

(ii) The function f defined on $\mathcal{X}\times\mathcal{X}$ by $f(x,y) = f_0(x-y)$ is a DR function.

Proof: First suppose f_0 is decreasing. Thus f_0 is $\mathcal{P}_n$ invariant so for each $g \in \mathcal{P}_n$,

$$f(gx,gy) = f_0(gx-gy) = f_0(g(x-y)) = f_0(x-y) = f(x,y).$$

Hence f is invariant. To show f is a DR function, (3.21) needs to be verified. For notational convenience, let $t = t_{12}$ and consider $x,y \in \mathcal{X}$ with $t'xt'y \geq 0$. Then, write $x = u + \alpha t$ with $u \perp t$ and $y = v + \beta t$ with $v \perp t$. Thus

$$f(x,R_t y) = f_0(x-R_t y) = f_0(u - v + (\alpha+\beta)t).$$

But, the condition $t'xt'y = \alpha\beta \geq 0$ implies that $|\alpha-\beta| \leq |\alpha+\beta|$ so the vector $x - y = u - v + (\alpha-\beta)t$ is in the line segment connecting $u - v - (\alpha+\beta)t = R_t(u - v + (\alpha+\beta)t)$ and $u - v + (\alpha+\beta)t$. Since f_0 is decreasing, this implies that

$$\begin{aligned} f_0(x-y) &= f_0(u - v + (\alpha-\beta)t) \\ &\geq f_0(u - v + (\alpha+\beta)t) = f_0(x-R_t y) \end{aligned}$$

so (3.21) holds.

Conversely, assume f is a DR function. Since f is invariant, f_0 is invariant. To show f_0 is decreasing we use (ii) of Proposition 2.17. Thus, for $0 \leq \beta_1 \leq \beta_2$, and $v \perp t$, it must be shown that

$$f_0(v + \beta_1 t) \geq f_0(v + \beta_2 t). \tag{3.25}$$

Set $x = v + \frac{1}{2}(\beta_1+\beta_2)t$ and $y = \frac{1}{2}(\beta_2-\beta_1)t$. Then $t'xt'y \geq 0$ so (3.21) holds. Hence

$$f_0(v + \beta_1 t) = f(x,y) \geq f(x,R_t y) = f_0(v + \beta_2 t),$$

so (3.25) holds. □

Of course there is a version of Proposition 3.8 which relates increasing functions to DR functions--namely, assume X is closed under addition. Then f_0 is increasing iff $f_0(x+y)$ is a DR function on $X \times X$. The proof of this assertion is essentially the same as that for Proposition 3.8.

We close this section with a few remarks about DR functions. Exploitation of (i) and (ii) of Definition 3.1 occurred in the work of Savage (1957), although the isolation of (i) and (ii) into a definition appeared in Eaton (1967) in work on ranking problems. In that paper, DR functions were said to have "Property M." Later Hollander et al. (1977) in their study of majorization and probability inequalities used "decreasing in transposition" for these functions and other functions defined on P_n. Such functions are called "arrangement increasing" in Marshall and Olkin (1979) (see 6.F).

The plethora of terminology concerning functions satisfying Definition 3.1 seems to be due to various authors' interpretations of the condition and applications of the results concerning such functions. There is a likelihood interpretation which I find rather appealing. First, write the function in question as $f(x,\theta)$ with $x \in X \subseteq R^n$ and $\theta \in \Theta \subseteq R^n$. Think of x as the variable in the density f and θ as the value of the parameter in Θ. For a given $u \in \Delta$, the hyperplane

$$H_u = \{z | u'z = 0\}$$

divides R^n into two parts:

$$H_u^+ = \{z | u'z \geq 0\}$$

and

$$H_u^- = \{z \mid u'z \le 0\}$$

with H_u being the intersection. The reflection R_u maps H_u^+ onto H_u^- and H_u^- onto H_u^+. The condition $u'xu'\theta \ge 0$ simply means x and θ are either both in H_u^+ or both in H_u^---that is, they are both on the same side of H_u. Hence, when $u'xu'\theta \ge 0$, then $u'xu'(R_u\theta) \le 0$ so x and $R_u\theta$ are on opposite sides of H_u. Now, the condition

$$f(x,\theta) \ge f(x,R_u\theta)$$

means that θ's on the same side of H_u as x are always more likely than the reflected θ. In other words, reflection decreases the likelihood when x and θ are on the same side of H_u. Hence the term decreasing reflection function.

Decreasing reflection functions are defined for other groups than P_n in Chapter 6, but are only discussed rather briefly since the main applications currently known are for the group P_n. However, the same interpretation given above, which depends only on the geometry of reflections, continues to hold.

Section 3.4: The Composition Theorem and First Applications

In this section, we first establish a result due to Hollander et al. (1977) concerning the composition of two DR functions. The applications here include an alternative proof of the Convolution Theorem and a decision theoretic treatment of the ranking problem discussed in Section 3.2.

Here is the Composition Theorem.

Theorem 3.9 (Hollander et al. (1977)). Let X, Y, and Z be symmetric subsets of R^n. Suppose f_1 is a DR function on $X \times Y$ and f_2 is a DR function on $Y \times Z$. Let μ be a P_n-invariant σ-finite measure defined on Y such that

$$f_3(x,z) = \int_Y f_1(x,y) f_2(y,z) \mu(dy) \tag{3.26}$$

is well defined for each $x \in X$ and $z \in Z$. Then f_3 is a DR function.

<u>Proof</u>: To show f_3 is P_n-invariant, we use the invariance of f_1, f_2 and the assumed invariance of μ to compute as follows. For $g \in P_n$,

$$f_3(gx,gz) = \int_Y f_1(gx,y)f_2(y,gz)\mu(dy) = \int_Y f_1(x,g^{-1}y)f_2(g^{-1}y,z)\mu(dy)$$
$$= \int_Y f_1(x,y)f_2(y,z)\mu(dy) = f_3(x,z),$$

so f_3 is invariant.

To complete the proof, condition (ii) of Proposition 3.6 needs to be verified. To this end, let $u = t_{12}$ and consider $x \in X$ and $z \in Z$ satisfying $u'xu'z \geq 0$. Thus, it must be shown that

$$\delta = f_3(x,z) - f_3(x,R_u z) \geq 0.$$

Let

$$Y^+ = \{y | y \in Y,\ u'y > 0\}$$

$$Y^- = \{y | y \in Y,\ u'y < 0\}$$

and

$$Y^0 = \{y | y \in Y,\ u'y = 0\},$$

so $Y = Y^+ \cup Y^- \cup Y^0$. Using the relation $R_u u = -u$ and the invariance of f_1 and f_2, an easy calculation shows that

$$\delta = \int_{Y^+} [f_1(x,y) - f_1(x,R_u y)][f_2(y,z) - f_2(y,R_u z)]\mu(dy).$$

This equality is a consequence of the relation $R_u Y^- = Y^+$ and the fact that the above integral over the set Y^0 is zero. Now, because $u'xu'z \geq 0$ and $y \in Y^+$, the integrand in the above expression for δ is non-negative. Hence $\delta \geq 0$ and the proof is complete. □

The Composition Theorem together with Proposition 3.8 provides an easy alternative proof of the Convolution Theorem (Theorem 2.20). To see this, consider h_1 and h_2 which are decreasing on R^n so that the convolution

$$h_3(x) = \int_{R^n} h_1(x-y)h_2(y)dy \tag{3.27}$$

is well defined. To show h_3 is decreasing, it must be verified that

$$f_3(x,z) = h_3(x-z)$$

is a DR function on $R^n \times R^n$. But

$$\begin{aligned} h_3(x-z) &= \int_{R^n} h_1(x-z-y)h_2(y)dy \\ &= \int_{R^n} h_1(x-y)h_2(y-z)dy \end{aligned} \tag{3.27}$$

where the second equality follows from the simple change of variable $y \to y+z$ and the translation invariance of Lebesgue measure. But both $f_1(x,y) = h_1(x-y)$ and $f_2(y,z) = h_2(y-z)$ are DR functions by Proposition 3.8. Thus, by the Composition Theorem with $X = Y = Z = R^n$, f_3 is a DR function so h_3 is decreasing. □

It is very natural to ask to what extent the Convolution Theorem is valid for spaces which are subsets of R^n. Since the natural setting for the Convolution Theorem is an additive group, we assume $X \subseteq R^n$ is a group under addition and assume that $\mu(dx)$ is a translation invariant measure on X. Both X and μ are assumed to be P_n-invariant. Suppose h_1 and h_2 are decreasing on X such that the convolution

$$h_3(x) = \int_X h_1(x-y)h_2(y)\mu(dy)$$

is defined. As long as Proposition 3.8 is valid for the space X, the argument in the previous paragraph holds without change. Thus, h_3 is again decreasing subject to the above proviso. For example, if X is the set of vectors in R^n whose coordinates are integers and μ is counting measure on X, the above argument holds. (See Section 2.4 for a discussion of majorization on other spaces than R^n).

The final example of this section concerns a decision theoretic extension of the "ranking problem" discussed in Section 3.2. The reader who is completely unfamiliar with the language of statistical decision theory may skip this example since it is not used in the sequel. However, for those with even a modest familiarity with decision theory, the arguments below are quite complete and the example is a nice application of the Composition Theorem. In essence, the following example provides a proof of a main result in Eaton (1967).

Example 3.4: Let X and Θ be symmetric subsets of R^n and suppose $f(x,\theta)$ is a density of X with respect to a P_n-invariant measure μ. It is assumed that f is a DR function. The statistical problem is to "rank the coordinates of θ" on the basis of an observation vector X with density $f(x,\theta)$. Now, a ranking of the coordinates of θ simply consists of some permutation of the vector a_0: whose coordinates are $n, n-1, \ldots, 1$. For example, if $n = 4$, then

$$a_0 = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} .$$

Let us agree that a given permutation of $a_0 \in R^4$ is our ranking; for example, the vector

$$\begin{pmatrix} 3 \\ 1 \\ 2 \\ 4 \end{pmatrix}$$

would assert that θ_4 is the largest (it receives the largest rank), θ_1 is the second largest, θ_3 is the third largest and θ_2 is the smallest. With this convention, an action space for the decision problem consists of the permutations of a_0.

Thus, for the general case, the action space for the decision problem is $A = \{ga_0 | g \in P_n\}$ where a_0 is the fixed vector with coordinates n,n-1,..., 1. Thus, each element $a \in A$ is just some permutation of n,n-1,..., 1--say $a_1, \ldots, a_n$; the interpretation is the action a gives the asserted ranks of the coordinates of θ. Thus, a_i is the asserted rank of θ_i, $i = 1, \ldots, n$ where large rank corresponds to large values.

Now, consider a loss function L defined on $A \times \Theta$ to R^1. It is assumed that -L (minus L) is a DR function. This means two things. First that L is invariant (i.e. $L(a,\theta) = L(ga, g\theta)$, $g \in P_n$) which seems very reasonable because of the symmetry of the problem. Second, using characterization (ii) of Proposition 3.6, this assumption means that if $a_1 > a_2$ and $\theta_1 > \theta_2$ (these are the first two coordinates of a and θ), then

$$-L(a,\theta) \geq -L(a, R_{t_{12}} \theta)$$

or equivalently, the loss for action

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{at} \quad \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}$$

is no larger than the loss for action

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{at} \quad \begin{pmatrix} \theta_2 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$$

In other words, with $a_3, \ldots, a_n$ and $\theta_3, \ldots \theta_n$ fixed, when $\theta_1 > \theta_2$, saying that the rank of θ_1 is a_1 and the rank of θ_2 is a_2 (when $a_1 > a_2$) gets no more loss than saying the rank of θ_2 is a_1 and the rank of θ_1 is a_2. This assumption seems to be a minimal requirement given our problem and the interpretation of $a \in A$.

Now, for simplicity we assume that each $x \in X$ has distinct coordinates.

This assumption is made so that some annoying technical details will not obfuscate the essence of the argument below. With this assumption, each x $\in X$ can be written uniquely as

$$x = k(x)z$$

where $z \in F$ and $k(x) \in P_n$ (recall that F is the convex cone of vectors whose coordinates are ordered largest to smallest). Recall that a decision rule, say δ, is a measurable function defined on X and taking values in A. Properties of a decision rule δ are measured in terms of the <u>risk function</u>

$$R(\delta,\theta) = \int L(\delta(x),\theta)f(x,\theta)\mu(dx) \tag{3.28}$$

which is just the expected loss from using δ at the parameter value θ. Also, given any distribution H on Θ, the average risk of δ is

$$R_H(\delta) = \int_\Theta R(\delta,\theta)H(d\theta).$$

A decision rule δ_H is a <u>Bayes rule</u> for H if

$$R_H(\delta_H) \le R_H(\delta) \quad \text{for all } \delta.$$

Now, we turn to the main result for the above ranking problem. Consider the decision rule δ_0 defined as follows. For $x \in X$, write $x = k(x)z$ with $z \in F$ and $k(x) \in P_n$. δ_0 is defined by

$$\delta_0(x) = k(x)a_0. \tag{3.29}$$

<u>Proposition 3.10</u>: For any distribution H on Θ with is P_n-invariant, δ_0 is a Bayes rule for H.

<u>Proof</u>: For any decision rule δ, it must be verified that

$$R_H(\delta_0) \le R_H(\delta). \tag{3.30}$$

To this end, let

$$Q(a,x) = \int L(a,\theta)f(x,\theta)H(d\theta).$$

Since -L and f are DR functions and H is P_n-invariant, the Composition Theorem implies that -Q is a DR function on $A \times X$. Write $a = ga_0$ for $g \in P_n$ and $x = k(x)z$, and note that both z and a_0 are in F. Since -Q is invariant, an application of Proposition 3.7 yields

$$-Q(ga_0,k(x)z) \le -Q(k(x)a_0,k(x)z)$$

for all $x \in X$ and $g \in P_n$. Hence

$$Q(a,x) \ge Q(k(x)a_0,x) \tag{3.31}$$

for $a \in A$ and $x \in X$. Since $\delta_0(x) = k(x)a_0$, (3.31) implies that for any decision rule δ,

$$Q(\delta(x),x) \ge Q(\delta_0(x),x). \tag{3.32}$$

Integrating both sides of (3.32) with respect to $\mu(dx)$ yields

$$\iint L(\delta(x),\theta)f(x,\theta)H(d\theta)\mu(dx) \ge \iint L(\delta_0(x),\theta)f(x,\theta)H(d\theta)\mu(dx)$$

which is just (3.30). Thus δ_0 is Bayes for any P_n-invariant H. □

The above result was proved in Eaton (1967) by showing essentially that Q is a DR function and then applying the argument given in Proposition 3.5. Of course, due to the work in Hollander et al. (1977), it is now clear the Composition Theorem together with the argument in Proposition 3.5 is what underlies this result. Proposition 3.10 provides a relatively easy proof of the fact that the decision rule δ_0 is both minimax and admissible.

<u>Proposition 3.11</u>: The decision rule δ_0 is minimax and admissible for any

loss function L such that -L is a DR function.

Proof: To show δ_0 is minimax, it must be verified (by the definition of minimax) that

$$\sup_\theta R(\delta_0,\theta) = \inf_\delta \sup_\theta R(\delta,\theta). \tag{3.33}$$

For any $\theta \in \Theta$, let H_θ be the distribution with mass $(n!)^{-1}$ at each $g\theta$, $g \in P_n$. Since the sup of an average is no greater than an average of sup's, we have

$$\sup_\theta R(\delta,\theta) \geq \sup_\theta \int R(\delta,\eta)H_\theta(d\eta). \tag{3.34}$$

Since for each θ, H_θ is a P_n-invariant distribution on Θ, Proposition (3.10) implies that

$$\int R(\delta,\eta)H_\theta(d\eta) \geq \int R(\delta_0,\eta)H_\theta(d\eta) = R(\delta_0,\theta)$$

where the last inequality is a consequence of the easily established identity

$$R(\delta_0,g\theta) = R(\delta_0,\theta) \tag{3.34}$$

for all $\theta \in \Theta$ and $g \in P_n$. Hence the right hand side of (3.33) is bounded below by $\sup_\theta R(\delta_0,\theta)$. But, trivially, the right hand side of (3.33) is bounded above by this expression. Hence (3.33) holds so δ_0 is minimax.

To show δ_0 is admissible, we argue by contradiction. Thus, assume δ_0 is not admissible so there exists a decision rule δ_1 such that

$$\left.\begin{array}{ll} R(\delta_1,\theta) \leq R(\delta_0,\theta) & \text{for all } \theta \in \Theta \\ R(\delta_1,\theta_0) < R(\delta_0,\theta_0) & \text{for some } \theta_0 \in \Theta. \end{array}\right\} \tag{3.35}$$

Using the notation above, let H_{θ_0} be the P_n-invariant distribution which puts mass $(n!)^{-1}$ at each $g\theta_0$, $g \in P_n$. Because of (3.35) and the fact that θ_0 gets positive mass from H_{θ_0}, integration of (3.35) yields

$$R_{H_{\theta_0}}(\delta_1) < R_{H_{\theta_0}}(\delta_0)$$

which contradicts the fact that δ_0 is Bayes for H_{θ_0}. □

Further refinements of these results as well as a number of examples and extensions can be found in Eaton (1967). Other applications to ranking problems can also be found in Hollander et al. (1977).

Section 3.5: Further Examples and Applications

In this section further examples and techniques are presented related to the problem of finding conditions under which the function in (3.1) is decreasing. Here is an example in which the technique is to verify the differential conditions of Proposition 2.18.

Example 3.5: Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables--that is,

$$X_i = \begin{cases} 1 & \text{with probability } p_i \\ 0 & \text{with probability } 1-p_i \end{cases}$$

for $i = 1, \ldots, n$. Fix $\Sigma_1^n p_i = n\lambda$. Gleser (1975) has shown that if x satisfies $0 \le x \le n\lambda-2$, then (Gleser (1975))

$$\Psi(p) = P\{\Sigma_1^n X_i \le x\} \tag{3.36}$$

is a decreasing function of the vector p whose coordinates $p_1, \ldots, p_n$ satisfy $\Sigma p_i = n\lambda$. As mentioned above, the technique is to verify the conditions of Proposition 2.17 via differentiation (i.e. as in Proposition

2.18). It is useful to interpret the above result. The mean of $\Sigma_1^n X_i$ is $n\lambda$ which is fixed. Since $x \le n\lambda-2$, (3.36) is a left tail probability. Intuitively, tail areas should decrease as variance decreases. But the variance of $\Sigma_1^n X_i$ is $\Sigma p_i(1-p_i)$ which is decreasing in the majorization ordering with $\Sigma p_i = n\lambda$ fixed. Hence the above confirms our intuition. However, it is shown by Gleser (1975) that this result is false for x's satisfying $n\lambda-2 < x \le n\lambda$. Thus, the result is somewhat delicate. For the details of the proof and other results, see Gleser (1975). □

The next example contains a result due to Y. Rinott (1973).

Example 3.6: Suppose $N \in R^k$ has a multinomial distribution M(k,p,n)--that is, N has integer coordinates $N_1, \ldots, N_k$ which satisfy $0 \le N_i$ and $\Sigma_1^k N_i = n$ where n is a positive integer, and $p \in R^k$ is a vector of probabilities, $p_1, \ldots, p_k$, which satisfy $\Sigma_1^k p_i = 1$. For future reference, we will write the probability function of N rather carefully. First, let $\mathcal{X}$ be the set of all vectors in R^k which have integer coordinates, and let μ be counting measure on $\mathcal{X}$. Let A be the set of vectors in $\mathcal{X}$, say x, whose coordinates satisfy $0 \le x_i$ and $\Sigma_1^k x_i = n$. Then the density of N, with respect to μ, given the parameter vector p, is

$$f(x|p) = \frac{n!}{x_1! \ldots x_k!} \prod_1^k p_i^{x_i} I_A(x) \tag{3.37}$$

where I_A is the indicator function of A. Of course, $f(\cdot|p)$ vanishes off the set A. Rinott proved that if h defined on A to R^1 is decreasing in the majorization sense, then

$$\Psi(p) = \mathcal{E}_p h(N) = \int_{\mathcal{X}} h(x) f(x|p) \mu(dx) \tag{3.38}$$

is also decreasing. Thus, Ψ is maximized when each p_i is 1/k and Ψ is minimized at the corners of the probability simplex. Rinott (1973) proved this result by verifying the differential condition in Proposition 2.18 and then he noted that a similar result follows for the Poisson distribution by averaging over n. An alternative method of proof was given by Nevius et al. (1977) which first establishes the Poisson result and then arrives at the multinomial via a conditioning argument. We will take this second route in proving Rinott's (1973) Theorem below. □

To introduce the methods developed in Proschan and Sethuraman (1977), Nevius et al. (1977) and Hollander et al. (1977), we first treat the case of Poisson variables and then take a careful look at the argument to see what makes it work. Let $X_1, \ldots, X_k$ be independent Poisson random variables with parameters $\theta_1, \ldots, \theta_k$. Thus X_i has a density on the integers given by

$$p(u|\theta_i) = \frac{e^{-\theta_i}\theta_i^u}{u!} I_{[0,\infty)}(u), \quad u = 0, \pm 1, \pm 2, \ldots \tag{3.39}$$

with respect to counting measure. As usual, $I_{[0,\infty)}$ is the indicator function of $[0,\infty)$. The sample space for the random vector X with coordinates $X_1, \ldots, X_k$ is, as usual, the set $\mathcal{X} \subseteq R^k$ which consists of all vectors which have integer coordinates. Let μ denote counting measure on $\mathcal{X}$ so the density of X is

$$f(x|\theta) = \prod_{i=1}^{k} \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!} I_{[0,\infty)}(x_i) \tag{3.40}$$

where $x \in \mathcal{X}$ has coordinates $x_1, \ldots, x_n$, and $\theta \in (0,\infty)^k = \Theta$ has coordinates $\theta_1, \ldots, \theta_k$. Now let h: $\mathcal{X} \to R^1$ be an increasing function (these are a bit more convenient than the decreasing functions in this example, but multiplication by a minus one changes from increasing to decreasing and vice versa) and consider

$$\Psi(\theta) = E_\theta h(X) = \int h(x) f(x|\theta)\mu(dx). \tag{3.41}$$

We now proceed to show that Ψ is increasing. To the end, recall that Ψ is increasing iff $\Psi(\theta+\eta)$ is a DR function (see the remark after Proposition 3.8). Hence we consider

$$\Psi(\theta+\eta) = \int_X h(x)f(x|\theta+\eta)\mu(dx). \tag{3.42}$$

But, the Poisson distribution is a <u>convolution family</u>--that is, the density f has the property that for $\theta,\eta \in \Theta$,

$$f(x|\theta+\eta) = \int_X f(x-y|\theta)f(y|\eta)\mu(dy) \tag{3.43}$$

which is most easily verified with characteristic functions. Substituting (3.43) into (3.42) and interchanging orders of integration yields

$$\begin{aligned}\Psi(\theta+\eta) &= \int\Big[\int h(x)f(x-y|\theta)\mu(dx)\Big]f(y|\eta)\mu(dy)\\ &= \int\Big[\int h(y+x)f(x|\theta)\mu(dx)\Big]f(y|\eta)\mu(dy).\end{aligned} \tag{3.44}$$

The second equality follows from a change of variable <u>and</u> the translation invariance of the measure μ on X. But, because p in (3.39) has a MLR, f in (3.40) is a DR function. Since h is increasing, h(y+x) is a DR function. Hence the Composition Theorem implies that

$$H(y,\theta) = \int_X h(y+x)f(x|\theta)\mu(dx)$$

is a DR function. Thus

$$\Psi(\theta+\eta) = \int H(y,\theta)f(y|\eta)\mu(dy)$$

and a second application of the Composition Theorem shows that Ψ is a DR function. Thus Ψ is increasing which yields

<u>Proposition 3.12</u>: If X has the density (3.40) on X and h: $X \to R^1$ is increasing (decreasing), then Ψ given by (3.41) is increasing (decreasing).

Now, the essentials in the above are that

$$\left.\begin{array}{ll}\text{(i)} & \mathcal{X} \text{ is a group under addition and the dominating} \\ & \text{measure for the density is translation invariant} \\ & \text{and permutation invariant} \\ \text{(ii)} & \Theta \text{ is closed under addition} \\ \text{(iii)} & \text{the density } f(x|\theta) \text{ is a DR function and satisfies} \\ & \text{the \underline{convolution property} (3.43).}\end{array}\right\} \qquad (3.45)$$

Here is another example where the same argument as above is valid.

Example 3.7. (Nevius et al. (1977)). Take $\mathcal{X} = R^k$, $\Theta = (0,\infty)^k$ and set

$$f(x|\theta) = \prod_{i=1}^{k} p(x_i|\theta_i) \qquad (3.46)$$

where

$$p(u|\eta) = \frac{u^{\eta-1}\lambda^{-\eta}\exp[-u/\lambda]}{\Gamma(\eta)} I_{(0,\infty)}(u).$$

Here $\lambda > 0$ is a fixed constant, $\Gamma(\cdot)$ denotes the gamma function, $\eta \in (0,\infty)$, and $u \in R^1$. The dominating measure μ on $\mathcal{X}$ is Lebesgue measure so f in (3.46) is the density of k independent gamma random variables with shape parameters $\theta_1,\ldots, \theta_k$ and a common scale parameter. That (i), (ii), and (iii) in (3.45) hold is easily verified so the argument given in the Poisson case is valid. Thus, if h is increasing (decreasing) on $\mathcal{X}$ to R^1, then

$$\Psi(\theta) = E_\theta h(X)$$

is increasing (decreasing) on Θ. □

Example 3.8: In this example, a proof of Rinott's (1973) result concerning the multinomial distribution (see Example 3.6) is given. Throughout this example, the notation established in Example 3.6 is used. Thus N has a M(k,p,n) distribution and N takes values in the set $A \subseteq \mathcal{X}$. The density of N on $\mathcal{X}$ is given by (3.37). If h is a decreasing function defined on A, the assertion is that

$$\Psi(p) = E_p h(N) \tag{3.47}$$

is decreasing. Since A is a finite set, h is bounded below on A by some constant--say M. Define h* on X by

$$h^*(x) = \begin{cases} h(x) & \text{if } x \in A \\ M & \text{if } x \notin A. \end{cases}$$

It is easy to verify that h* is decreasing on X. Let X have coordinates $X_1, \ldots, X_k$ which are independent, and X_i is Poisson with parameter p_i--the ith coordinate of p. Proposition 3.12 shows that

$$\Psi^*(p) = E_p h^*(X) \tag{3.48}$$

is decreasing. But, the conditional distribution of X given $\Sigma_1^k X_i = r$ is M(k,p,r), and the marginal distribution of $\Sigma_1^k X_i$ is Poisson with parameter 1. Thus

$$\Psi^*(p) = E_p E_p (h^*(X) | \Sigma_1^k X_i) = \sum_{r=0}^{\infty} E(h^*(X) | \sum_1^k X_i = r) q_r$$

where $q_r = P(\Sigma_1^k X_i = r)$. From the definition of h*, we have

$$\Psi^*(p) = q_n E(h^*(X) | \sum_1^k X_i = n) + M(1-q_n) = q_n \Psi(p) + M(1-q_n).$$

Since $q_n > 0$, it follows the Ψ is decreasing. □

For further examples and applications of the type above, the reader should consult Gleser (1975), Rinott (1973), Proschan and Sethuraman (1977), Hollander et al. (1977), Nevius et al. (1977), and Marshall and Olkin (1979, Chapters 3,11,12). Some nice applications to matching problems are discussed in Marshall and Olkin (1979, p. 304-305).

The final example in this section concerns the accuracy of confidence intervals for a mean based on the t-statistic when the observations are not normal, but only satisfy a weak symmetry condition. The relevant references for this example are Efron (1969) and Eaton (1970, 1974).

Example 3.9: Consider random variables $X_1,\ldots, X_n$ and assume a "linear model" type structure:

$$X_i = \mu + \epsilon_i, \quad i = 1,\ldots, n$$

where μ is an unknown parameter and $\epsilon_1,\ldots, \epsilon_n$ are random variables. To describe the assumption on the joint distribution of $\epsilon_1,\ldots, \epsilon_n$, let $\mathcal{D}_n$ denote the group of all n×n diagonal matrices whose diagonal elements are either 1 or -1. Thus, $\mathcal{D}_n$ has 2^n elements. It is assumed that the random vector ϵ, with coordinates $\epsilon_1,\ldots, \epsilon_n$ satisfies

$$\mathcal{L}(\epsilon) = \mathcal{L}(D\epsilon) \qquad \text{for } D \in \mathcal{D}_n. \tag{3.49}$$

In other words, the distribution of ϵ is $\mathcal{D}_n$-invariant--such distributions were said to have orthant symmetry by Efron (1969). If ϵ has a mean vector and satisfies (3.49), then the mean vector of ϵ must be zero.

Given the model above, one possible way to construct a confidence interval for μ is to use the t-statistic (as if $X_1,\ldots, X_n$ were i.i.d. $N(\eta,\sigma^2)$). In other words, let c_{n-1} be the $(1-\alpha)/2$ upper percentage point of a t_{n-1} distribution and use the interval

$$\bar{X} \pm \frac{c_{n-1}}{\sqrt{n(n-1)}} \left(\Sigma_1^n (X_i-\bar{X})^2\right)^{1/2}. \tag{3.50}$$

Of course, to evaluate the statistical properties of this procedure, we must try to calculate the probability that this random interval covers the parameter μ. This probability is

$$\delta = P\{|\bar{X}-\mu| \le \frac{c_{n-1}}{\sqrt{n(n-1)}} \left(\Sigma_1^n (X_i-\bar{X})^2\right)^{1/2}\} \tag{3.51}$$

Because of our model assumption, this can be written

$$\delta = P\{(\Sigma_1^n \epsilon_i)^2 \le \frac{n}{n-1} c_{n-1}^2 (\Sigma_1^n \epsilon_i^2 - \frac{1}{n}(\Sigma_1^n \epsilon_i)^2)\}$$

which, after some manipulation, is

$$\delta = P\{ \, |\Sigma_1^n \epsilon_i| / (\Sigma_1^n \epsilon_i^2)^{1/2} \le d_n\} \tag{3.52}$$

where

$$d_n = \left[1 + \frac{c_{n-1}^2 - 1}{n}\right]^{-1/2} c_{n-1}. \tag{3.53}$$

Now, from an inferential point of view, it would be useful to have lower bounds on (3.52) so that the constructed interval would have a guaranteed coverage probability. Equivalently, we will try to develop some upper bounds on

$$\beta = P\{|\Sigma_1^n \epsilon_i| / (\Sigma \epsilon_i^2)^{1/2} \ge d_n\}. \tag{3.54}$$

The assumption (3.49) on the distribution of ϵ implies that ϵ has the same distribution as the random vector

$$Z = \begin{pmatrix} U_1 \theta_1 \\ U_2 \theta_2 \\ \vdots \\ U_n \theta_n \end{pmatrix}$$

where $U_1, \ldots, U_n$ are i.i.d. random variables taking the values ± 1 each with probability 1/2. The distribution of the θ_i's is specified by

$$L\left(\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}\right) = L\left(\begin{pmatrix} |\epsilon_1| \\ \vdots \\ |\epsilon_n| \end{pmatrix}\right)$$

and the vector of U's is independent of the vector of θ's. A bit of reflection should convince the reader that these assertions are plausible--for a proof, see Efron (1969). Since ϵ and Z have the same distribution, (3.55) can be substituted into (3.54) yielding

$$\beta = P\{|\Sigma_1^n U_i \theta_i| / (\Sigma_1^n \theta_i^2)^{1/2} \geq d_n\} \tag{3.56}$$

Now, condition on $\theta_1, \ldots, \theta_n$, set $\xi_i = \theta_i/(\Sigma_1^n \theta_i^2)^{1/2}$ and let

$$\Psi(\xi) = P\{|\Sigma_1^n \xi_i U_i| \geq d_n\} \tag{3.57}$$

where $\xi_i \geq 0$, $\Sigma_1^n \xi_i^2 = 1$. Thus, to obtain upper bounds on β, it is sufficient to obtain upper bounds on $\Psi(\xi)$ which are valid for all ξ with $\Sigma_1^n \xi_i^2 = 1$, and $\xi_i \geq 0$.

Now, Efron (1969) argues as follows. Regard $\xi_1, \ldots, \xi_n$ as fixed constants satisfying $\xi_i \geq 0$ and $\Sigma_1^n \xi_i^2 = 1$. Look at the even moments (the odd moments are zero)

$$\mu_r(\xi) = E(\Sigma_1^n \xi_i U_i)^{2r} \tag{3.58}$$

for $r = 1, 2, \ldots$. Efron (1969) proves that $\mu_r(\xi)$ is bounded above by EV^{2r} where V is N(0,1). This suggests that $|V|$ is stochastically larger (at least approximately) than $|\Sigma_1^n \xi_i U_i|$ which in turn suggests that (3.57) is bounded above (at least approximately) by $P\{|V| \geq d_n\}$. But $P\{|V| \geq d_n\}$ is close to α because

(i) the t_{n-1} distribution is close to the N(0,1) for moderate and large n

(ii) d_n is close to c_{n-1} for moderate and large n, and for values of c_{n-1} which occur in practice.

Now it, in fact, is not true that $|V|$ is stochastically larger than $n^{-1/2}\Sigma_1^n U_i$. However, Efron's result shows that

$$E(\Sigma_1^n \xi_i U_i)^{2r} \leq EV^{2r} \tag{3.59}$$

for r = 1,2,... and all $\xi_i \geq 0$ satisfying $\Sigma_1^n \xi_i^2 = 1$. One way to try to sharpen this result is to extend the inequality (3.59) to functions other than $x \to x^{2r}$. That is, for what functions f is it true that

$$Ef(\Sigma_1^n \xi_i U_i) \leq Ef(V) \tag{3.60}$$

for all n and $\xi_1, \ldots, \xi_n$ satisfying $\xi_i \geq 0$, $\Sigma_1^n \xi_i^2 = 1$? Using majorization, this equation was partially answered in Eaton (1970,1974). The details follow.

<u>Proposition 3.13</u>: Let f be defined on R^1 to R^1. Suppose that f is symmetric and has a first derivative f'. Suppose further that

(i) for each $c \geq 0$ and for $t > 0$, the function

$$t \to t^{-1}[f'(c+t) - f'(c-t)]$$

is non decreasing on $(0,\infty)$.

For $\delta \in R^n$ with coordinates $\delta_1, \ldots, \delta_n$ which satisfy $\delta_i \geq 0$, $\Sigma_1^n \delta_i = 1$, let

$$g(\delta) = Ef(\Sigma_1^n \sqrt{\delta_i}\, U_i). \tag{3.61}$$

Then g is decreasing.

<u>Proof</u>: The proof is quite standard--the conditions of Proposition 2.18 will be verified. That g is P_n-invariant is clear since $U_1, \ldots, U_n$ are i.i.d. To verify (ii) of Proposition 2.18, set $W = \Sigma_3^n \sqrt{\delta_i} U_i$. By assumption, f is symmetric. Since W has a symmetric distribution and is independent of (U_1, U_2), for fixed $(\delta_3, \ldots, \delta_n)$, $g(\delta)$ in (3.61) can be written as an average (over c) of functions

$$g_c(\delta) = E_{U_1 U_2} f(\sqrt{\delta_1} U_1 + \sqrt{\delta_2} U_2 + c) \tag{3.62}$$

for $c \geq 0$. Since an average of decreasing functions is decreasing it suffices to show that g_c in (3.62) satisfies (ii) of Proposition 2.18 for each $c \geq 0$. Since f is symmetric $f'(-x) = -f'(x)$. Using this if we set $t_1 = \delta_1^{1/2} + \delta_2^{1/2}$ and $t_2 = \delta_1^{1/2} - \delta_2^{1/2}$, a computation yields

$$\Delta = \frac{\partial g_c}{\partial \delta_1} - \frac{\partial g_c}{\partial \delta_2}$$

$$= \frac{1}{2} Ef'(\sqrt{\delta_1}U_1 + \sqrt{\delta_2}U_2 + c)[\delta_1^{-1/2}U_1 - \delta_2^{-1/2}U_2]$$

$$= \frac{1}{8\sqrt{\delta_1\delta_2}t_1t_2}\{\frac{f'(c+t_2)-f'(c-t_2)}{t_2} - [\frac{f'(c+t_1)-f'(c-t_1)}{t_1}]\}.$$

Since $0 < t_2 \leq t_1$, assumption (i) shows that $\Delta \leq 0$. Thus (ii) of Proposition 2.18 holds so g is decreasing. □

Condition (i) of f in Proposition 3.13 is not so easy to check. A sufficient condition for (i) to hold is that f have three derivations and f''' be non-decreasing on $(0,\infty)$ (for a proof, see Eaton (1974)). In particular, the functions $x \to x^{2r}$, $r = 1,2,\ldots$ all satisfy (i).

Here are some immediate consequences of Proposition 3.13.

Proposition 3.14. Let f satisfy the assumptions of Proposition 3.13.

(i) $Ef(n^{-1/2}\Sigma_1^n U_i)$ is non-decreasing in n.

(ii) If there exists an $\epsilon > 0$ such that $E|f(n^{-1/2}\Sigma_1^n U_i)|^{1+\epsilon}$ is bounded in n, then for $L(Z) = N(0,1)$,

$$Ef(\textstyle\sum_1^n \sqrt{\delta_i}U_i) \leq Ef(Z) \tag{3.63}$$

for all n and all $\delta_1,\ldots, \delta_n$ satisfying $\delta_i \geq 0$ and $\Sigma_1^n \delta_i = 1$.

Proof: (i) follows from Proposition 3.13 by noting that the vector in R^n with coordinates $(n-1)^{-1/2},\ldots, (n-1)^{-1/2}, 0$ majorizes the vector with coordinates $n^{-1/2},\ldots, n^{-1/2}$.

For (ii), the Central Limit Theorem shows $\Sigma_1^n n^{-1/2}U_i$ converges in distribution to Z. Since f is continuous, $f(\Sigma_1^n n^{-1/2}U_i)$ converges in distribution to f(Z). The uniform boundedness of the $(1+\epsilon)$ absolute moment implies that

$$\lim_{n\to\infty} Ef(\Sigma_1^n n^{-1/2} U_i) = Ef(Z).$$

Since the sequence of expectations is non-decreasing in n (by part (i)), we have

$$Ef(\Sigma_1^n n^{-1/2} U_i) \le Ef(Z).$$

The conclusion (3.63) follows since $g(\delta)$ in (3.61) is maximized for $\delta \in R^n$ with all coordinates equal to $n^{-1/2}$. □

The assumptions in (ii) are easily shown to hold for $f(x) = x^{2r}$, $r = 1,2,\ldots$ so (3.63) give Efron's (1969) result in this case. The validity of (3.63) for other random variables than $U_1,\ldots, U_n$ is discussed in Eaton (1974). One application to probability inequalities for (3.57) (and hence for (3.51)) follows. Let $d > 0$ be a fixed number. Let f be a non-negative non-decreasing function on R^1 which satisfies $f(d) = 1$, and satisfies the assumptions of Proposition 3.14. A pointwise argument together with (3.63) gives

$$P\{|\Sigma_1^n \xi_i U_i| \ge d\} \le Ef(\Sigma_1^n \xi_i U_i) \le Ef(Z) \tag{3.64}$$

for $\xi_1,\ldots, \xi_n$ satisfying $\Sigma_1^n \xi_1^2 = 1$. The problem is now to choose f in a clever way to make $Ef(Z)$ as small as possible. This problem has not been solved completely, but there is some evidence to suggest that f's of the form

$$f_u(x) = \begin{cases} 0 & \text{if } |x| \le u \\ \dfrac{(|x|-u)^3}{(d-u)^3} & \text{if } |x| \ge u \end{cases}$$

for $0 \le u < d$ yield reasonably tight bounds for (3.64) (See Eaton (1974) for the argument which suggests such f's). Since f_u''' is non-decreasing and since $f_u(d) = 1$, (3.64) yields

$$P\{|\Sigma_1^n \xi_i U_i| \geq d\} \leq \inf_{0 \leq u < d} Ef_u(Z). \tag{3.65}$$

This infimum has not been computed explicitly, but some upper bounds are known which produce relatively good upper bounds when $d \geq 2$. See Eaton (1970,1974) for further details.

Chapter 4: Log Concavity and Related Topics

A main result of this chapter, due to Prekopa (1973), asserts that if $f(x,y)$ defined on $R^m \times R^n$ is log concave, then the "marginal function"

$$h(x) = \int_{R^n} f(x,y)dy \tag{4.1}$$

is log concave on R^m. This fact has a number of important consequences and applications. For example, results in Anderson (1955), Sherman (1955), Mudholkar (1966), and Davidovic et al. (1969) all follow easily from the above assertion. These results along with some applications are discussed below.

Section 1: Log concave functions

Let f be a non-negative real valued function defined on R^n.

Definition 4.1: The function f is log concave if for all $x,y \in R^n$ and $\alpha \in (0,1)$,

$$f(\alpha x + (1-\alpha)y) \geq f^{\alpha}(x)f^{1-\alpha}(y). \tag{4.2}$$

In some situations, a non-negative function h defined on a convex subset $D \subseteq R^n$ satisfies (4.2). In this case, observe that

$$f(x) = \begin{cases} h(x) & \text{if } x \in D \\ 0 & \text{if } x \notin D \end{cases}$$

is defined on all of R^n and also satisfies (4.2). For this reason, the domain of definition of log concave functions is always taken to be R^n.

Here are some elementary facts which are useful.

Proposition 4.1: If f is log concave on R^n, then for each $c \geq 0$, the sets $\{x|f(x) \geq c\}$ and $\{x|f(x) > c\}$ are convex.

Proof: Elementary. □

<u>Proposition 4.2</u>: Suppose B is a subset of R^n. Then I_B, the indicator function of B, is log concave iff B is a convex set.

<u>Proof</u>: Elementary. □

<u>Proposition 4.3</u>: Let f be non-negative and defined on R^n. Then f is log concave iff for each $x,y \in R^n$, the function

$$\Psi(t) = f(x + ty) \tag{4.3}$$

is log concave on R^1.

<u>Proof</u>: If f is log concave, then for $t_1, t_2, \in R^1$ and $\alpha \in (0,1)$,

$$\begin{aligned}\Psi(\alpha t_1 + (1-\alpha)t_2) &= f(x + (\alpha t_1 + (1-\alpha)t_2)y) \\ &= f(\alpha(x+t_1y) + (1-\alpha)(x+t_2y)) \geq f^{\alpha}(x+t_1y)f^{1-\alpha}(x+t_2y) \\ &= \Psi^{\alpha}(t_1)\Psi^{1-\alpha}(t_2).\end{aligned}$$

Conversely, if Ψ is log concave,

$$\begin{aligned}f(\alpha x + (1-\alpha)y) &= f(y + \alpha(x-y)) = \Psi(\alpha) \\ &\geq \Psi^{\alpha}(1)\Psi^{1-\alpha}(0) = f^{\alpha}(x)f^{1-\alpha}(y). \quad \square\end{aligned}$$

<u>Proposition 4.4</u>: Suppose f is log concave on R^n and A is an $n \times m$ matrix. Then the function h defined on R^m by $h(u) = f(Au)$ is log concave.

<u>Proof</u>: Elementary. □

<u>Proposition 4.5</u>: Suppose f is defined on $R^m \times R^n$ and is log concave. Fix u in R^m and define h on R^n by $h(v) = f(u,v)$. Then h is log concave.

<u>Proof</u>: Elementary. □

<u>Proposition 4.6</u>: Suppose f defined on R^m is log concave and define h on $R^m \times R^n$ by $h(u,v) = f(u)$. Then h is log concave. Also, if f_1 and f_2 are log

concave on R^m, then the product f_1f_2 is also log concave.

Proof: Elementary. □

Section 2: Prekopa's Theorem

Here is a special case of Prekopa's Theorem (1973) which in fact implies the general result alluded to in the introduction to this chapter.

Theorem 4.7. Suppose f defined on R^2 is log concave and let h be defined by

$$h(x) = \int_{-\infty}^{\infty} f(x,y)dy. \tag{4.4}$$

If $h(x) < +\infty$ for all $x \in R^1$, then h is log concave.

Because of its independent interest, the proof of Theorem 4.7 is given in the final section of this chapter. The proof is not that of Prekopa, but is modelled after an argument due to Brascamp and Lieb (1974). The main result of this section, which is an easy consequence of Theorem 4.7, follows.

Theorem 4.8 Prekopa (1973)). Let f be defined on $R^m \times R^n$ and suppose that f is log concave. Then, the function h defined on R^m by

$$h(u) = \int_{R^n} f(u,v)dv, \tag{4.5}$$

assumed to be finite for $u \in R^m$, is log concave.

Proof: Because Lebesgue measure, dv on R^n, is a product measure, an easy induction argument shows that it suffices to establish the claim for n = 1. For n = 1 and fixed $u_1, u_2 \in R^m$, define Ψ on R^2 by

$$\Psi(t,v) = f(u_1+tu_2,v). \tag{4.6}$$

That Ψ is log concave on R^2 is an easy consequence of the log concavity of f. However,

$$h(u_1+tu_2) = \int_{-\infty}^{\infty} f(u_1+tu_2,v)dv = \int_{-\infty}^{\infty} \Psi(t,v)dv.$$

Theorem 4.7 applied to Ψ shows that for u_1 and u_2 fixed, $t \to h(u_1+tu_2)$ is log concave on R^1. By Proposition 4.3, h is log concave on R_m. □

Proposition 4.9 (Davidovic et al. (1969)). Suppose f_1 and f_2 are log concave functions on R^n. Then the convolution

$$h(y) = \int_{R^n} f_1(y-x)f_2(x)dx$$

is log concave, if $h(y) < +\infty$ for $y \in R^n$.

Proof: The log concavity of f_1 and f_2 implies that

$$f(x,y) = f_1(y-x)f_2(x)$$

is log concave on R^{2n}. By Theorem 4.8,

$$h(y) = \int_{R^n} f(x,y)dx$$

is log concave on R^n. □

An immediate corollary to the above Proposition is a useful observation due to Sherman (1955).

Proposition 4.10 (Sherman (1955)). Let C_1 and C_2 be convex sets in R^n and let μ denote Lebesgue measure on R^n. If

$$g(x) = \mu(C_1 \cap (C_2+x))$$

is finite for $x \in R^n$, then g is log concave, so

$$\{x \mid g(x) \geq c\}$$

is a convex set for each $c \in R^1$.

<u>Proof</u>: Let f_1 be the indicator function of C_1 and let f_2 be the indicator function of the convex set $-C_2$. Then

$$g(x) = \int f_1(u) f_2(x-u) \mu(du)$$

is the convolution of two log concave functions. Hence g is log concave so by Proposition 4.1,

$$\{x \mid g(x) \geq c\}$$

is a convex set. □

We close this section with a brief discussion about random vectors which have log concave densities. Consider random vectors $U \in R^m$ and $V \in R^n$ and assume $(U,V) \in R^{m+n}$ has a joint density $f(u,v)$ which is log concave. It follows immediately from Theorem 4.8 that the marginal density of U, say

$$h(u) = \int f(u,v) dv \tag{4.7}$$

is log concave. Hence the set

$$D = \{u \mid h(u) > 0\}$$

is a convex set by Proposition 4.1.

For fixed $u \in R^m$, define $f_2(v \mid u)$ on R^n by

$$f_2(v|u) = \begin{cases} \frac{f(u,v)}{h(u)}, & u \in D \\ g(v), & u \notin D \end{cases}$$

where g is some fixed log concave density on R^n (for example, take g to be the density of a $N(0,I_n)$ distribution). It is well known that $f_2(\cdot|u)$ serves as a version of the conditional density of V given U = u. Moreover, for each fixed u, $f_2(\cdot|u)$ is clearly log concave on R^n. This shows that when the joint density is log concave, then one can select a version of the conditional density which is also log concave.

Section 3: Multivariate Unimodality and Anderson's Theorem

On the real line, it is rather obvious how a symmetric unimodal (about 0) function should be defined--namely, the definition given in Chapter 2. However, the choice of a definition of unimodality in higher dimensions is not so clear--even if attention is restricted to the symmetric case. In spite of its restrictiveness, the following rather strong definition used by Anderson (1955) has proved to be useful. (For a discussion of other notions of unimodality in R^n, see Dharmadhikari and Jogdeo (1976) and Das Gupta (1980).)

Definition 4.2: Let f be a real valued function defined on R^n. If f is symmetric ($f(x) = f(-x)$ for $x \in R^n$) and if for each real number r, the set

$$\{x|f(x) \geq r\} \subseteq R^n$$

is convex, f is called A-unimodal.

In most applications, A-unimodal functions are non-negative, but this is not required in the definition. If f is non-negative, symmetric and log concave, then by Proposition 4.1 f is A-unimodal. In particular, indicator functions of convex symmetric sets are A-unimodal. However, sums of A-unimodal functions need not be A-unimodal. This can be seen by taking f_i to be the indicator of a convex symmetric set for i = 1,2 and picking the sets so $f_1 + f_2$ is not A-unimodal.

One property that A-unimodal functions do possess is that they are decreasing on rays emanating from $0 \in R^n$. Before proving this, we state a formal definition.

Definition 4.3: A real valued function f defined on R^n is ray-decreasing if for each $x \in R^n$, the function

$$h(\beta) = f(\beta x), \quad \beta \in R^1 \tag{4.7}$$

is non-increasing on $[0,\infty)$.

Observe that the family of all ray-decreasing functions is a convex cone--that is, if f_1, f_2 are ray-decreasing and c_1, c_2 are non-negative constants, then $c_1 f_1 + c_2 f_2$ is also ray-decreasing.

Proposition 4.11: If f is A-unimodal, then for $x \in R^n$ fixed, $h(\beta) = f(\beta x)$, $\beta \in R^1$ is a symmetric unimodal function on R^1. Hence, f is ray-decreasing.

Proof: That h is symmetric is obvious since f is symmetric. For $0 \le \beta_1 \le \beta_2$, it must be shown that

$$h(\beta_1) \ge h(\beta_2). \tag{4.8}$$

Let $r = f(\beta_2 x)$. Since f is A-unimodal, the set

$$C = \{y | f(y) \ge r\}$$

is convex and symmetric. But $\beta_2 x \in C$ by construction so $-\beta_2 x \in C$. Since $0 \le \beta_1 \le \beta_2$, the point $\beta_1 x$ is in the line segment connecting $\beta_2 x$ and $-\beta_2 x$ so $\beta_1 x \in C$. Hence $h(\beta_1) = f(\beta_1 x) \ge r = h(\beta_2)$. Thus, f is ray-decreasing by definition. □

Theorem 4.12 (Anderson (1955)). Suppose f_1 and f_2 are non-negative A-unimodal functions. Then the convolution

$$f(y) = \int_{R^n} f_1(y-x)f_2(x)dx \tag{4.9}$$

is symmetric and ray-decreasing. In particular, $f(0) \geq f(y)$ for all y.

Proof: Since f_1 and f_2 are non-negative, f is well defined even though f(y) may be infinite for some y's in R^n. Let I_m be the indicator function of $\{x | x \in R^n, \|x\| \leq m\}$. It is easy to show that

$$f_{i,m}(x) = f_i(x)I_m(x), \quad i = 1,2$$

are both A-unimodal, and vanish off the compact set $\{x | x \in R^n, \|x\| \leq m\}$. The monotone convergence theorem shows that

$$f_m(y) = \int_{R^n} f_{1,m}(y-x)f_{2,m}(x)dx$$

converges pointwise to f(y) given in (4.9). Since the pointwise limit of symmetric and ray-decreasing functions is again symmetric and ray-decreasing, it suffices to prove the theorem for functions f_1 and f_2 which vanish off a compact set.

We now proceed with the proof under the assumption that f_1 and f_2 vanish off a compact set. That f is symmetric is easily established using the symmetry of f_1 and f_2. For i = 1,2, define K_i on $R^n \times (0,\infty)$ by

$$K_i(x,a) = \begin{cases} 1 & \text{if } f_i(x) \geq a \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, for each $a \in (0,\infty)$, $K_i(\cdot,a)$ is zero off a compact set. Also, for a fixed, $K_i(\cdot,a)$ is the indicator function of a bounded convex symmetric set since f_i is A-unimodal. Hence $K_i(\cdot,a)$ is log concave and symmetric, i = 1,2. Therefore, for fixed a_1 and a_2 in $(0,\infty)$,

$$H(y,a_1,a_2) = \int_{R^n} K_1(y-x,a_1)K_2(x,a_2)dx$$

is finite, is log concave (Proposition 4.9) and is obviously symmetric. Hence $H(\cdot, a_1, a_2)$ is symmetric and ray-decreasing. Because the ray-decreasing functions form a convex cone,

$$q(y) = \int_0^\infty \int_0^\infty H(y, a_1, a_2)\, da_1 da_2 \tag{4.10}$$

is also ray decreasing. But, the definition of K_i yields the relation

$$f_i(x) = \int_0^\infty K_i(x,a)\, da, \quad i = 1,2. \tag{4.11}$$

Using (4.11) in (4.10), Fubini's Theorem for non-negative functions implies

$$\begin{aligned} q(y) &= \int_0^\infty \int_0^\infty \int_{R^n} K_1(y-x, a_1) K_2(x, a_2)\, dx da_1 da_2 \\ &= \int_{R^n} \int_0^\infty \int_0^\infty K_1(y-x, a_1) K_2(x, a_2)\, da_1 da_2 dx \\ &= \int_{R^n} f_1(y-x) f_2(x)\, dx = f(y). \end{aligned}$$

Hence, f is ray-decreasing. □

<u>Example 4.1</u>: Consider an "elliptical" distribution on R^n--that is, a distribution with a density f of the form

$$f(x) = |\Sigma|^{-1/2} k(x'\Sigma^{-1}x) \tag{4.12}$$

where Σ: n×n is positive definitive. Assume that k is a non-increasing function defined on $[0,\infty)$. Since k is non-increasing, f is A-unimodal. The translation family generated by f has a density

$$p(x|\theta) = f(x-\theta), \quad x \in R^n,\ \theta \in R^n.$$

Let C be a convex symmetric subset of R^n and suppose the random vector X has density $p(\cdot|\theta)$. Then, Theorem 4.12 tells us that the probability that X is in C, as a function of the translation parameter θ, decreases on rays. To see this, the quantity of interest is

$$g(\theta) = P_\theta\{X \in C\} = \int_C f(x-\theta)dx = \int I_C(x)f(x-\theta)dx,$$

where I_C is the indicator function of C. The symmetry of f implies that $f(x-\theta) = f(\theta-x)$. Since I_C and f are A-unimodal, Theorem 4.12 shows that g is ray-decreasing. □

There are a couple of ways to extend Theorem 4.12 in fairly obvious but rather useful directions.

Proposition 4.13: Suppose $f_1 \geq 0$ is A-unimodal and $\int f_1(x)dx < +\infty$. Also suppose f_2 is a bounded function and is A-unimodal. Then

$$f(y) = \int f_1(y-x)f_2(x)dx$$

is symmetric and ray decreasing.

Proof: Because of our assumption f is well defined and finite for all $y \in R^n$. Let M be a bound for f_2 so $|f_2(x)| \leq M$ for all $x \in R^n$. Hence $f_2(x) + M \geq 0$ for all x, and $f_2 + M$ is A-unimodal.

By Proposition 4.12,

$$\begin{aligned} k(y) &= \int f_1(y-x)(f_2(x)+M)dx \\ &= \int f_1(y-x)f_2(x)dx + M(\int f_1(x)dx) = f(y) + c \end{aligned}$$

is ray-decreasing and symmetric. Here, c is just a constant. Thus f is ray-decreasing and symmetric. □

The extension of Proposition 4.13 to unbounded f_2's can sometimes be accomplished with truncation and limiting arguments--provided of course that enough conditions are assumed to justify taking limits. These extensions are left to the reader.

The next extension comes from Sherman (1955). It is based on the observation that the symmetric ray-decreasing functions form a convex cone. In particular, suppose f_1, f_2 and f_3 are non-negative A-unimodal functions. Then

$$k_i(y) = \int f_3(y-x)f_i(x)dx, \quad i = 1,2$$

are symmetric and ray-decreasing. Hence

$$k_1(y) + k_2(y) = \int f_3(y-x)[f_1(x) + f_2(x)]dx$$

is also symmetric and ray-decreasing even though $f_1 + f_2$ need not be A-unimodal. In order to describe this situation more formally, let C be the convex cone of all non-negative symmetric Borel measurable functions, h, on R^n which have the form

$$h(x) = \Sigma_1^r a_i f_i(x) \tag{4.13}$$

where $a_i \geq 0$, f_i is a non-negative A-unimodal function, and r is some positive integer.

Proposition 4.14: If h_1 and h_2 are in C, then

$$k(y) = \int_{R^n} h_1(y-x)h_2(x)dx$$

is symmetric and ray decreasing.

Proof: Since h_1 and h_2 have the form (4.13), k is a linear combination, with non-negative coefficients, of symmetric ray decreasing functions. Hence k is a symmetric ray decreasing function. □

Proposition 4.15: Suppose h_1 and h_2 can be expressed as

$$h_i(x) = \lim_{m\to\infty} h_{i,m}(x), \quad i = 1,2$$

where $h_{i,m} \in C$ for $i = 1,2$ and $m = 1,2,\ldots$ and $h_{i,m}(x) \le h_{i,m+1}(x)$ for all $x \in R^m$, $i = 1,2$ and $m = 1,2,\ldots$. Then

$$k(y) = \int_{R^n} h_1(y-x)h_2(x)dx$$

is symmetric and ray decreasing.

Proof: Define k_m by

$$k_m(y) = \int h_{1,m}(y-x)h_{2,m}(x)dx.$$

Then for each m, k_m is symmetric and ray decreasing. Because of our assumptions, for each y, $h_{1,m}(y-x)h_{2,m}(x)$ increases as $m \to \infty$ to the limit $h_1(y-x)h_2(x)$. Since all the functions involved are non-negative, the Monotone Convergence Theorem shows that k(y) is the pointwise limit of $k_m(y)$. It follows immediately that k is symmetric and ray decreasing. □

The point of the above discussion is that the convolution of functions in the convex cone C yields symmetric ray decreasing functions. Also, certain limiting arguments can be used to extend the validity of this convolution result--as long as enough assumptions are made to justify the limiting operations. Further, C is a convex cone so that positive linear combinations of elements of C also are functions whose convolutions are symmetric and ray-decreasing. Sherman (1955) uses a combination of uniform convergence and convergence in mean to study the convolutions in question.

Section 4: Mudholkar's Theorem:

The main result of this section, due to Mudholkar (1966), is perhaps best motivated by reinterpreting Theorem 4.12. Let G_0 be the two element group $\{I_n, -I_n\}$ thought of as a group acting on vectors in R^n. That is, $g \in G_0$ maps x into gx. First notice that a function f defined on R^n is symmetric iff $f(x) = f(gx)$ for $g \in G_0$. In other words, f is symmetric iff f is G_0-invariant. To reinterpret what ray-decreasing means for G-invariant functions, the argument used in the proof of Proposition 4.11 is relevant. To say that a symmetric function f is ray-decreasing is to say that for each $y \in R$ and each $\beta \in [-1,1]$, the inequality $f(\beta y) \geq f(y)$ holds. However, as β varies over $[-1,1]$, the vectors βy vary over the line segment connecting -y and y. That is, βy varies over the convex set generated by $\{-y,y\}$. Of course the set $\{-y,y\}$ is the orbit of y under the action of the group G_0; by definition, the G_0-orbit of y is $\{gy|g \in G_0\}$. Thus, a symmetric function f is ray-decreasing iff $f(x) \geq f(y)$ for all x in the convex set generated by the orbit of y. The parallel considerations in Chapter 2 on majorization are now fairly clear--namely, the convex set generated by the orbit of a point was used to define the majorization ordering (when the group is P_n) and the functions with the property that $f(x) \geq f(y)$ (for all x in the convex hull of the orbit of y) were called decreasing. These observations suggest that there may be a version of Theorem 4.12 for more general groups than $G_0 = \{\pm I_n\}$. That this is the case was discovered by Mudholkar (1966). We now proceed with the formal development.

Consider a group G which is a subgroup of the group of n×n orthogonal matrices. Thus each element of G defines a linear transformation on R^n. Given $y \in R^n$, let C(y) denote the convex set generated by $\{gy|g \in G\}$. We write $x \leq y$ to mean $x \in C(y)$, just as in the majorization case discussed at length in Chapter 2. It is easy to show $x \leq y$ and $y \leq z$ implies that $x \leq z$, and $x \leq y$ iff $C(x) \subseteq C(y)$. As usual, f defined on R^n is <u>decreasing</u> if $x \leq y$ implies that $f(x) \geq f(y)$. A function f defined on R^n is <u>G-invariant</u> if $f(x) = f(gx)$ for all $x \in R^n$ and $g \in G$. Any f which is decreasing is G-invariant because $x \leq gx \leq x$, $x \in R^n$ and $g \in G$.

<u>Proposition 4.16</u>: Suppose f is G-invariant and log concave. The f is decreasing.

<u>Proof</u>: For $x \in C(y)$, it must be shown that $f(x) \geq f(y)$. Set $\gamma = f(y)$ and consider

$$B = \{u | f(u) \geq \gamma\}.$$

It suffices to show $C(y) \subseteq B$. However, B is convex since f is log concave. Also $u \in B$ implies $gu \in B$ for all g since f is G-invariant. But $y \in B$ by definition so $\{gy | g \in G\} \subseteq B$. The convexity of B implies $C(y) \subseteq B$. □

<u>Definition 4.4</u>: A function $f: R^n \to R^1$ is <u>convex-unimodal</u> if for each $\gamma \in R^1$, $\{x | f(x) \geq \gamma\}$ is a convex set.

We now proceed to the statement and proof of the G-analogue of Theorem 4.12. An important observation in the previous case was that the class of symmetric ray-decreasing functions (the decreasing functions when $G = \{I_n, -I_n\}$) forms a convex cone. That same observation is important here--namely, the class of decreasing functions forms a convex cone. The dependence of the word "decreasing" on the group G is suppressed since G is fixed throughout the discussion.

<u>Theorem 4.17</u>: If f_1 and f_2 are non-negative, G-invariant, and convex-unimodal, then the convolution

$$h(y) = \int_{R^n} f_1(y-x)f_2(x)dx$$

is decreasing.

<u>Proof</u>: The proof is very similar to the proof of Theorem 4.12. The G-invariance of h follows from the invariance of f_1 and f_2 and the following calculation:

$$\begin{aligned} h(gy) &= \int_{R^n} f_1(gy-x)f_2(x)dx = \int f_1(g(y-g^{-1}x))f_2(x)dx \\ &= \int f_1(y-x)f_2(gx)dx = \int f_1(y-x)f_2(x)dx. \end{aligned}$$

The third equality follows from a change of variable and the fact that each g preserves Lebesgue measure. As in the proof of Theorem 4.12, it suffices to prove the Theorem for f_1 and f_2 which vanish off some compact set. This is assumed is what follows. For $a \geq 0$ and $i = 1,2$, define K_i on $R^n \times [0,\infty)$ by

$$K_i(x,a) = \begin{cases} 1 & \text{if } f_i(x) \geq a \\ 0 & \text{otherwise.} \end{cases}$$

As in the proof of Theorem 4.12,

$$h(y) = \int_0^\infty \int_0^\infty \int_{R^n} K_1(y-x,a_1)K_2(x,a_2)\,dx\,da_1\,da_2.$$

But, for a fixed, $K_i(\cdot,a)$ is the indicator function of a bounded convex G-invariant set. Thus, $K_i(\cdot,a)$ is a log concave G-invariant function. Hence,

$$K(y;a_1,a_2) = \int_{R^n} K_1(y-x,a_1)K_2(x,a_2)\,dx$$

is a log concave function (Proposition 4.9) and the G-invariance of $K(\cdot,a_1,a_2)$ is proved the same way the invariance of h is proved. Thus, for each a_1,a_2, $K(\cdot;a_1,a_2)$ is decreasing by Proposition 4.16. Since

$$h(y) = \int_0^\infty \int_0^\infty K(y;a_1,a_2)\,da_1\,da_2,$$

it follows immediately that h is decreasing. □

Now, all of the discussion concerning the extension of Theorem 4.12 (in Propositions 4.13, 4.14, 4.15) is valid for Theorem 4.17 because the class of decreasing functions is a convex cone. One simply replaces "symmetric and ray-decreasing" by "decreasing" and "A-unimodal" gets replaced by "G-invariant and convex-unimodal." The details of this are left to the reader.

The orginal proofs of Theorems 4.12 and 4.17 used the Brunn-Minkowski inequality. For a discussion of this and other aspects of unimodality, see Das Gupta (1980).

Section 5: Applications to MANOVA Problems:

Two applications to problems in multivariate analysis of variance (MANOVA) are given in this section. The first concerns a concentration property of Gauss-Markov estimators in certain linear models which include the multivariate linear model as a special case. In the second application, monotonicity properties of the power functions of some classical tests in MANOVA problems are discussed. The first application is an extension of results in Berk and Hwang (1984) while the material in the second example comes from Das Gupta et al. (1964) and Eaton and Perlman (1974).

The notion of concentration was studied on the real line by Birnbaum (1948) and was later extended to R^n by Sherman (1955). Because our applications involve random matrices, it is more convenient to formulate Sherman's definition for a finite dimensional inner product space $(V,(\cdot,\cdot))$.

Definition 4.5: Let P_1 and P_2 be two probability measures defined on the Borel sets of $(V,(\cdot,\cdot))$. The probability P_1 is more concentrated about 0 than P_2 if $P_1(C) \geq P_2(C)$ for all convex symmetric sets $C \subseteq V$. The notation $P_1 >_0 P_2$ is used to mean P_1 is more concentrated about 0 than P_2.

The following example of a concentration inequality is due to Anderson (1955).

Proposition 4.18: On R^n, suppose P_i is the probability measure corresponding to a multivariate normal distribution with mean 0 and covariance Σ_i, $i = 1,2$. If $\Sigma_2 - \Sigma_1$ is positive semi-definite, then $P_1 >_0 P_2$.

Proof: Let X_i be a random vector which has a $N_n(0,\Sigma_i)$ distribution. Since $\Delta = \Sigma_2 - \Sigma_1$ is positive semi-definite, there is a random vector Y, independent of X_1 and X_2, which is $N(0,\Delta)$ and $X_1 + Y$ has the same distribution as X_2. For any convex symmetric set C,

$$P_2(C) = Pr(X_2 \in C) = Pr(X_1 + Y \in C)$$

$$= \mathcal{E}\Pr\{X_1 + Y \in C \mid Y = y\} = \mathcal{E}\Pr\{X_1 \in C\text{-}y \mid Y = y\}.$$

Since X_1 and Y are independent and since the distribution of X_1 is log concave, Anderson's Theorem implies that

$$\Pr\{X_1 \in C\text{-}y \mid Y = y\} \le \Pr\{X_1 \in C\}$$

because $\Pr\{X_1 \in C\text{-}y\}$ is a symmetric ray-decreasing function of y. Thus

$$P_2(C) \le \Pr\{X_1 \in C\} = P_1(C)$$

for all convex symmetric sets so $P_1 >_0 P_2$. □

The proof of Proposition 4.18 is quite special because it uses the fact that the sum of two independent normal random vectors is again normal. Extensions of Proposition 4.18 to the so-called spherical distributions can be found in Das Gupta et al. (1971). This paper contains a useful result on the behavior of certain probabilities when the covariance matrix changes in special ways. In addition, there are a number of interesting examples included in this paper.

Here is the promised linear model example.

Example 4.2: It is assumed that the reader is somewhat familiar with the inner-product space version of the Gauss-Markov Theorem for linear models. However, for the sake of completeness, the theory is briefly outlined here following the treatment in Eaton (1983). Consider a random vector Y taking values in a finite dimensional inner product space $(V,(\cdot,\cdot))$. The mean vector of Y, say μ, is assumed to lie in a known linear subspace M of V. The covariance of Y, say $\Sigma = \mathrm{Cov}(Y)$, is assumed to be positive definite and to lie in some known set γ of positive definite covariances. Thus, the pair (M,γ) specifies the first and second moment structure of the "linear model."

To state the Gauss-Markov Theorem, some notation and assumptions are needed. First assume, without essential loss of generality, that $I \in \gamma$ where I denotes the identity linear transformation on V. The linear model specified by (M,γ) is <u>regular</u> if $\Sigma(M) \subseteq M$ for all $\Sigma \in \gamma$. Linear unbiased estimators for $\mu \in M$ are estimators of the form AY where A is a linear transformation from V to V which satisfies

(i) $A(V) \subseteq M$

(ii) $Ax = x$ for all $x \in M$.

Let $\mathcal{A}$ be all those linear transformations satisfying (i) and (ii). Further let A_0 denote the orthogonal projection onto M so $A_0 \in \mathcal{A}$. One version of the Gauss-Markov Theorem is

<u>Theorem 4.19</u>: If the linear model is regular, then

$$\mathrm{Cov}(A_0Y) = A_0\Sigma A_0' \leq A\Sigma A' = \mathrm{Cov}(AY) \tag{4.14}$$

for all $A \in \mathcal{A}$ and $\Sigma \in \gamma$. The inequality sign in (4.14) means that $A\Sigma A' - A_0\Sigma A_0'$ is positive semi-definite.

Examples of linear models which are regular include the so-called univariate linear model discussed in Scheffe (1959) and the classical multivariate linear model treated in Anderson (1958). A proof of Theorem 4.19 in the notation above can be found in Eaton (1983). Of course, $\hat{\mu} = A_0Y$ is called the Gauss-Markov estimator of μ for regular linear models.

Under the conditions in Theorem 4.19, inequality (4.14) suggests that $A_0Y-\mu$ is closer to zero than $AY-\mu$ because covariance measures dispersion about the mean and AY is an unbiased estimator of μ for all $A \in \mathcal{A}$. Thus, it is natural to ask whether or not the distribution of $A_0Y-\mu$ is more concentrated about $0 \in V$ than is the distribution of $AY-\mu$. It is this question to which we now turn.

For each $A \in \mathcal{A}$, notice that $AY-\mu = A(Y-\mu)$ because A satisfies condition (ii). Thus, questions concerning $AY-\mu$ only involve AZ where $Z = Y-\mu$ has mean 0 and $\mathrm{Cov}(Z) = \Sigma \in \gamma$ when $\mathrm{Cov}(Y) = \Sigma$. Since $AZ \in M$ for $A \in \mathcal{A}$, concentration of the distribution of AZ concerns the inner product space $(M,(\cdot,\cdot))$ rather than V. Formally, our question is this:

$$\left.\begin{array}{l}\text{Under what conditions on } \gamma \text{ and the distribution} \\ \text{of Z is it true that} \\ \qquad P\{A_0 Z \in C\} \geq P\{AZ \in C\} \\ \text{for all } \Sigma \in \gamma,\ A \in \mathcal{A}, \text{ and all convex} \\ \text{symmetric subsets } C \subseteq M?\end{array}\right\} \tag{4.15}$$

Proposition 4.20: If the linear model is regular and if for each $\mu \in M$ and $\Sigma \in \gamma$, Y has a normal distribution with mean μ and covariance Σ, then

$$P\{A_0Z \in C\} \geq P\{AZ \in C\} \tag{4.16}$$

for each $\Sigma \in \gamma$, $A \in \mathcal{A}$ and each convex symmetric subset of M.

Proof: Fix $\Sigma \in \gamma$ and $A \in \mathcal{A}$. Since Y is $N(\mu,\Sigma)$, $Z = Y-\mu$ is $N(0,\Sigma)$. Theorem 4.19 implies that

$$\mathrm{Cov}(A_0Z) \leq \mathrm{Cov}(AZ) \tag{4.17}$$

since the linear model is regular. Thus, on replacing R^n by M in Proposition 4.18, we see (4.16) holds for every convex symmetric subset of M. □

Now, the problem is how to weaken the distributional assumptions in Proposition 4.20 but still retain the inequality (4.16). To motivate a possible argument, consider AZ and notice that

$$AZ = AA_0Z + AQ_0Z = A_0Z + AQ_0Z \tag{4.18}$$

where $Q_0 = I - A_0$ is the orthogonal projection onto $M^{\perp}$--the orthogonal complement of M. The relation $AA_0 = A_0$ follows from $Ax = x$ for all $x \in M$. Thus, if C is a convex symmetric subset of M, we have

$$\begin{aligned} P\{AZ \in C\} &= P\{A_0Z + AQ_0Z \in C\} \\ &= \mathcal{E}P\{A_0Z + AQ_0Z \in C \mid Q_0Z = w\} \end{aligned}$$

$$= \mathcal{E}P\{A_0Z \in C\text{-}Aw \mid Q_0Z = w\}. \tag{4.19}$$

Thus, if the conditional distribution of A_0Z given $Q_0Z = w$ satisfies

$$P\{A_0Z \in C\text{-}u \mid Q_0Z = w\} \le P\{A_0Z \in C \mid Q_0Z = w\} \tag{4.20}$$

for each $u \in M$, it would follow that

$$P\{AZ \in C\} \le P\{A_0Z \in C\}. \tag{4.21}$$

In particular, if the conditional distribution of A_0Z given $Q_0Z = w$ has a density in M which is A-unimodal, then (4.20) follows by Theorem 4.12 with

(i) f_2 equal to the conditional density of A_0Z given $Q_0Z = w$
(ii) $y = u$ and f_1 equal to the indicator function of C.

The conditions in the next result are sufficient to make Theorem 4.12 applicable.

Proposition 4.21: Assume (M,γ) determines a regular linear model for Y. Suppose that for each $\Sigma \in \gamma$, the distribution of $Z = Y - \mu$ has a density f on V (with respect to Lebesgue measure) which satisfies

(i) f is log concave
(ii) $f(A_0x + Q_0x) = f(-A_0x + Q_0x)$

for $x \in V$ where $Q_0 = I - A_0$ and A_0 is the orthogonal projection onto M. Then, for each convex symmetric subset $C \subseteq M$,

$$P\{A_0Y\text{-}\mu \in C\} \ge P\{AY\text{-}\mu \in C\} \tag{4.22}$$

for all $A \in \mathcal{A}$.

Proof: For each $x \in V$, write $x = u + v$ where $u \in M$ and $v \in M^{\perp}$. Let

$$h(v) = \int_M f(u+v)du$$

where "du" denotes Lebesgue measure on M. For v fixed and for $u \in M$, define $f(u|v)$ by

$$f(u|v) = \begin{cases} \frac{f(u+v)}{h(v)} & \text{if } h(v) > 0 \\ g(u) & \text{if } h(v) = 0 \end{cases}$$

where g is the density of a normal distribution with mean 0 and covariance the identity on M. It was argued at the end of Section 4.2 that $f(\cdot|v)$ is log concave for each v. Because of assumption (ii), it follows that

$$f(-u|v) = f(u|v), \quad u \in M.$$

Thus, the conditional density $f(\cdot|v)$ of A_0Z given $Q_0Z = v$ is A-unimodal. Hence inequality (4.20) holds and thus (4.22) holds. □

Here is another case where (4.22) holds.

Proposition 4.22: Assume (M,γ) determines a regular linear model for Y. Suppose that for each $\Sigma \in \gamma$, the distribution of $Z = Y - \mu$ has a density f on V of the form

$$f(x) = |\Sigma|^{-1/2}h[(x,\Sigma^{-1}x)] \tag{4.23}$$

where h is a non-increasing function defined on $[0,\infty)$. Then for each convex symmetric subset $C \subseteq M$, (4.22) holds for all $A \in \mathcal{A}$.

Proof: Define K on $V\times(0,\infty)$ by

$$K(x,a) = \begin{cases} 1 & \text{if } f(x) \geq a \\ 0 & \text{if } f(x) < a. \end{cases}$$

Fix $a > 0$ and consider $K(\cdot,a)$. Because of the form of f in (4.23) and the assumption that h is non-increasing, $K(\cdot,a)$ has one of the following two forms for some constant $b \geq 0$:

$$\text{(i)}\quad K(x,a) = \begin{cases} 1 & \text{if } (x,\Sigma^{-1}x) \leq b \\ 0 & \text{otherwise.} \end{cases}$$

or

$$\text{(ii)}\quad K(x,a) = \begin{cases} 1 & \text{if } (x,\Sigma^{-1}x) < b \\ 0 & \text{otherwise.} \end{cases}$$

Since the argument is the same for both cases, we assume case (i) obtains. Hence $K(\cdot,a)$ is the indicator function of a bounded convex set so $K(\cdot,a)$ is log concave. We next verify that $K(\cdot,a)$ satisfies assumption (ii) of Proposition 4.21. Since the linear model is regular, $\Sigma(M) \subseteq M$ for each $\Sigma \in \gamma$, so $\Sigma^{-1}(M) \subseteq M$. The self adjointness of Σ^{-1} then implies $\Sigma^{-1}(M^{\perp}) \subseteq M^{\perp}$. These relations yield $\Sigma^{-1}A_0 = A_0\Sigma^{-1}$ and $\Sigma^{-1}Q_0 = Q_0\Sigma^{-1}$, so

$$A_0\Sigma^{-1}Q_0 = 0. \tag{4.24}$$

Therefore,

$$(x,\Sigma^{-1}x) = (A_0x+Q_0x,\Sigma^{-1}[A_0x+Q_0x]) = (A_0x,\Sigma^{-1}A_0x) + (Q_0x,\Sigma^{-1}Q_0x).$$

Because $K(\cdot,a)$ has the form (i), we see $K(\cdot,a)$ satisfies assumption (ii) of Proposition 4.21.

Now, let c be the Lebesgue measure of

$$\{x \mid (x,\Sigma^{-1}x) \leq b\}$$

so $c^{-1}K(\cdot,a)$ is a density which satisfies the assumption of Proposition 4.21. Thus, given a convex symmetric subset $C \subseteq M$, with

$$B_0 = \{x \mid A_0x \in C\}$$

and

$$B_1 = \{x | Ax \in C\},$$

Proposition 4.21 yields

$$\int_V I_{B_0}(x)K(x,a)dx \geq \int_V I_{B_1}(x)K(x,a)dx.$$

Integrating this inequality from 0 to ∞, and using the definition of K, we have

$$\int_V I_{B_0}(x)f(x)dx = \int_V\int_0^\infty I_{B_0}(x)K(x,a)dxda$$

$$\geq \int_V\int_0^\infty I_{B_1}(x)K(x,a)dxda = \int_V I_{B_1}(x)f(x)dx.$$

But the left side of this inequality is $P\{A_0Z \in C\}$ and the right side is $P\{AZ \in C\}$. Thus (4.22) holds for $A \in \mathcal{A}$. □

This ends the discussion of Example 4.2. □

Example 4.3: The final example of this chapter deals with the behavior of the power function of some tests in the classical MANOVA testing problem. The problem is considered in canonical form and it is assumed that the reader is somewhat familiar with the problem. A more complete description of the MANOVA problem and its reduction to canonical form can be found in Eaton (1983).

The data for the MANOVA problem in canonical form consists of a random matrix X: r×p and a symmetric positive definite random matrix S: p×p. It is assumed that X and S are independent and X has a normal distribution with mean matrix μ: r×p and a covariance $I_r \otimes \Sigma$. Thus, the rows of X are independent and each row has the covariance matrix Σ: p×p. The random matrix S is assumed to have a $W(\Sigma,p,n)$ distribution--that is, S has a

Wishart distribution with n degrees of freedom and expectation $n\Sigma$ where $n \geq p$. The problem is to test the null hypothesis H_0: $\mu=0$ versus the alternative H_1: $\mu\neq0$. For simplicity, it is assumed that $r \leq p$. The case of $r > p$ is similar.

This testing problem in invariant under the group of linear transformations $O_r \times Glp$ where O_r is the group of $r\times r$ orthogonal matrices and Glp is the group of $p\times p$ non-singular matrices. A group element (Γ, A) acts on a sample point by

$$(X,S) \rightarrow (\Gamma XA', A\Sigma A')$$

and on a parameter point by

$$(\mu,\Sigma) \rightarrow (\Gamma\mu A', A\Sigma A').$$

In this discussion, attention is restricted to non-randomized invariant tests. Such tests are functions of the eigenvalues of $XS^{-1}X'$--say $\lambda_1 \geq \ldots \geq \lambda_r$. The acceptance region of such a test is a subset of $\mathcal{L}_{r,p}\times\mathcal{S}_p$ where $\mathcal{L}_{r,p}$ is the vector space of all $r\times p$ real matrices and $\mathcal{S}_p$ is the vector space of all real $p\times p$ symmetric matrices. If C is the acceptance region of an invariant test, then $C \subseteq \mathcal{L}_{r,p}\times\mathcal{S}_p$ satisfies

$$(x,s) \in C \text{ implies } (\Gamma xA', AsA') \in C$$

for all $(\Gamma,A) \in O_r\times Glp$. The power function of an invariant test with acceptance region C is a function of the eigenvalues of $\mu\Sigma^{-1}\mu'$. Hence the power function of such a test can be written as

$$\pi(\theta,C) = 1 - P_{\mu,\Sigma}\{(X,S) \in C\} \qquad (4.25)$$

where $\theta \in R^r$ has coordinates $\theta_1 \geq \ldots \geq \theta_r > 0$ and $\theta_1^2, \ldots, \theta_r^2$ are the eigenvalues of $\mu\Sigma^{-1}\mu'$. The reason for using $\theta_1, \ldots, \theta_r$ rather than $\theta_1^2, \ldots, \theta_r^2$ as the argument of the power function will be clear in a moment.

Given a vector $\eta \in R^r$, define $\mu(\eta)$ to be

$$\mu(\eta) = \begin{bmatrix} \eta_1 & & 0 & \vdots & \\ & \eta_2 & & \vdots & 0 \\ 0 & & \ddots & \vdots & \\ & & \eta_r & \vdots & \end{bmatrix} \quad : \ r \times p$$

When $\Sigma = I_p$ and $\mu = \mu(\eta)$ in (4.25), the eigenvalues of $\mu\Sigma^{-1}\mu'$ are some permutation of $\eta_1^2, \ldots, \eta_r^2$. Therefore the power function $\pi(\theta, C)$ in (4.25) is determined by the function

$$\rho(\eta, C) = P_{\mu(\eta)}\{(X,S) \in C\} \tag{4.26}$$

where the probability in (4.26) is computed when $\mu = \mu(\eta)$ and $\Sigma = I_p$. More precisely, if $\theta_1^2 \geq \ldots \geq \theta_r^2 \geq 0$ are the eigenvalues of $\mu\Sigma^{-1}\mu'$, to compute $\pi(\theta, C)$, we just evaluate (4.26) for $\eta_i = \theta_i$, $i = 1, \ldots, r$ which yields

$$\pi(\theta, C) = 1 - \rho(\theta, C). \tag{4.27}$$

Now, we proceed with the analysis of $\rho(\cdot, C)$ defined in (4.26).

<u>Proposition 4.23</u>: (Das Gupta et al. (1964)). Let C be an invariant acceptance region of a test. Assume that C is convex in the ith row of $x \in \mathcal{L}_{p,r}$ when $s \in \mathcal{S}_p$ and the remaining rows of x are fixed, $i = 1, \ldots, r$. Then $\rho(\eta, C)$ is a symmetric unimodal function in each coordinate of η.

<u>Proof</u>: For notational simplicity, the proof is given for $i = 1$. Fix s, $x_2, \ldots, x_r$ and let

$$C_1 = \left\{x_1 \in R^p \,\middle|\, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}, s\right) \in C\right\}.$$

By assumption, C_1 is a convex subset of R^p. Let $\Gamma \in O_r$ be diagonal with (1,1) element equal to minus one and all other diagonals equal to plus one. Since C is invariant

$$\left[\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}, s\right] \in C$$

implies that

$$\left[\Gamma\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}, s\right] \in C$$

so

$$\left[\begin{pmatrix} -x_1 \\ \vdots \\ x_r \end{pmatrix}, s\right] \in C.$$

Hence $C_1 = -C_1$. Now, when $\mu = \mu(\eta)$ and $\Sigma = I_p$, the first row of X has a $N(\xi, I_p)$ distribution where

$$\xi = \begin{pmatrix} \eta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad . \tag{4.28}$$

Let g denote the density of a $N_p(0,I_p)$ distribution, Then, if X_1 is the first row of X,

$$EI_{C_1}(X_1) = \int_{R^p} I_{C_1}(x)g(x-\xi)dx \tag{4.29}$$

which is a symmetric ray-decreasing function of ξ since g is A-unimodal and C_1 is a symmetric convex set. Because of the special structure of ξ in (4.28), we conclude that (4.29) is a symmetric unimodal function of η_1. But, since the rows of X, say $X_1, \ldots, X_r$ and S are mutually independent, $\rho(\eta, C)$ in (4.26) can be computed by averaging (over $X_2, \ldots, X_r$, and S) functions of the form (4.29)--each of which is a symmetric unimodal function of η_1. Hence $\rho(\eta, C)$ is a symmetric unimodal function of η_1. □

When the assumption of Proposition 4.23 holds, we conclude that $\rho(\eta,C)$ is decreasing in each η_i when $\eta_i \geq 0$, $i = 1,\ldots, r$. Hence the power function $\pi(\theta,C)$ is increasing in each coordinate of θ. Das Gupta et al. (1964) show that the acceptance region of the likelihood ratio test satisfies the assumption of Proposition 4.23. The likelihood ratio test accepts iff

$$\prod_{1}^{r} (1 + \lambda_i) \leq k$$

where k is a fixed constant and $\lambda_1,\ldots, \lambda_r$ are the eigenvalues of $XS^{-1}X'$.

Stronger conclusions concerning $\rho(\eta,C)$ can be reached when it is assumed that $C \subseteq L_{p,r}\times S_p$ is a convex set. To describe these conclusions, some notation is needed. Let P_r denote the group of r×r permutation matrices and let D_r denote the group of r×r diagonal matrices with plus or minus ones on the diagonal. It is easy to show that

$$\left.\begin{aligned} &\mu(g\eta) = g\mu(\eta); \quad g \in D_r \\ \text{and} \quad &\mu(g\eta) = g\mu(\eta)\begin{pmatrix} g' & 0 \\ 0 & I \end{pmatrix}; \quad g \in P_r \quad . \end{aligned}\right\} \tag{4.30}$$

Here, I is the (p-r)×(p-r) identity matrix. Let $P_r \cdot D_r = \{g_1g_2 | g_1 \in P_r, g_2 \in D_r\}$. Then $P_r \cdot D_r$ is a group because $g_1g_2g_1' \in D_r$ whenever $g_1 \in P_r$ and $g_2 \in D_r$.

<u>Proposition 4.24</u> (Eaton and Perlman (1974)). If $C \subseteq L_{p,r}\times S_p$ is a convex set which is the acceptance region of an invariant test, then

(i) $\rho(h\eta,C) = \rho(\eta,C)$ for $\eta \in R^p$, $h \in P_r \cdot D_r$

(ii) $\rho(\cdot,C)$ is log concave.

<u>Proof</u>: Consider $\Sigma = I_p$ and a mean matrix $\mu \in L_{p,r}$ for X. The distributional assumptions made imply that the density of $X \in L_{p,r}$ is

$$p_1(x-\mu) = c_1 \exp[-\tfrac{1}{2}\operatorname{tr}(x-\mu)(x-\mu)']$$

where c_1 is a constant and the density of S on $\mathcal{S}_p$ is

$$p_2(s) = c_2(\det s)^{\frac{n-p-1}{2}} \exp[-\tfrac{1}{2}\operatorname{tr}(s)]I_1(s)$$

where I_1 is the indicator function of the convex set of positive definite $p\times p$ symmetric matrices, and c_2 is a constant. Note that $p_1(x)p_2(s)$ is a log concave function defined on the $rp + \frac{1}{2}p(p+1)$ dimensional vector space $\mathcal{L}_{p,r}\times\mathcal{S}_p$. Also, the indicator function of C is log concave on this space since C is convex by assumption. Therefore,

$$\Psi(\mu) = \iint I_C(x,s)p_1(x-\mu)p_2(s)dxds \tag{4.31}$$

is the convolution of two log concave functions evaluated at the point $(\mu,0) \in \mathcal{L}_{p,r}\times\mathcal{S}_p$. Hence $\Psi(\cdot)$ is log concave on $\mathcal{L}_{p,r}$ so

$$\rho(\eta,C) = \Psi(\mu(\eta)) \tag{4.32}$$

is log concave on R^r.

For $\Gamma_1 \in O_r$ and $\Gamma_2 \in O_p$, the invariance of C, p_1 and p_2 imply that Ψ given in (4.31) satisfies

$$\Psi(\mu) = \Psi(\Gamma_1\mu\Gamma_2), \quad \mu \in \mathcal{L}_{p,r}.$$

This together with (4.30) yields conclusion (i) of the proposition. □

Now, the group $\mathcal{P}_r\cdot\mathcal{D}_r$ induces a partial ordering on R^r as described in Section 4. Thus $u \le v$ iff u is in the convex hull of the $\mathcal{P}_r\cdot\mathcal{D}_r$ orbit of v. Under the assumptions of Proposition 4.24, Proposition 4.16 shows that $\rho(\cdot,C)$ is decreasing in this ordering on R^r. Therefore $\pi(\theta,C) = 1 - \rho(\theta,C)$ is increasing in this ordering. For two vectors θ and ξ which satisfy $\theta_1 \ge \ldots \ge \theta_r \ge 0$ and $\xi_1 \ge \ldots \ge \xi_r \ge 0$, the discussion in Example 6.2 (Chapter 6) shows that $\xi \le \theta$ in this ordering iff

$$\Sigma_1^j \xi_i \le \Sigma_1^j \theta_i, \quad j = 1,\ldots, r.$$

For any $\xi \le \theta$, we have

$$\pi(\theta,C) \ge \pi(\xi,C)$$

when C is convex and is the acceptance region of an invariant test.

An example of a test whose acceptance region is convex is provided by the Lawley-Hotelling trace test which accepts if

$$\Sigma_1^r \lambda_i \le k$$

where k is a fixed constant and $\lambda_1,\ldots, \lambda_r$ are the eigenvalues of $XS^{-1}X'$. This and other examples are provided in Eaton and Perlman (1974). □

Reliability theory is another area where log concavity has played a role. For example, see Savits (1985) for a definition of multivariate increasing failure rate and its relation to log concavity.

Section 6: Proof of Theorem 4.7

In the statement of Theorem 4.7, the non-negative function f defined on R^2 is assumed to be log concave. Then h, defined by

$$h(x) = \int_{-\infty}^{\infty} f(x,y)dy \tag{4.33}$$

is assumed to be finite for each $x \in R^1$. The claim is that h is log concave on R^1.

We first argue that it suffices to take f bounded with compact support. Let I_n be the indicator function of the set $\{(x,y) \in R^2 | x^2+y^2 \le n\}$. Thus $I_n(x,y)f(x,y)$ is log concave and has compact support. Now define f_n by

$$f_n(x,y) = \begin{cases} n & \text{if } I_n(x,y)f(x,y) \geq n \\ I_n(x,y)f(x,y) & \text{otherwise.} \end{cases}$$

Then f_n is log concave, is bounded and has compact support. Also, $f_n(x,y)$ increases monotonically to $f(x,y)$ as $n \to \infty$. By the Monotone Convergence Theorem

$$h_n(x) = \int_{-\infty}^{\infty} f_n(x,y)dy$$

converges pointwise to $h(x)$. Thus, if h_n is log concave, h is log concave.

Thus, we want to show h given in (4.33) is log concave when f is log concave, bounded and has compact support. The first step is the following.

Proposition 4.25: Let $C \subseteq R^2$ be a non-empty bounded convex set and define g on R^1 by

$$g(x) = \int I_C(x,y)dy \tag{4.34}$$

where I_C is the indicator function of C. On the set $D = \{x|g(x) > 0\}$, g is concave function.

Proof: If D is empty, there is nothing to prove so assume D is not empty. For x_1 and x_2 in D and $\alpha \in (0,1)$, it must be shown that

$$g(\alpha x_1 + (1-\alpha)x_2) \geq \alpha g(x_1) + (1-\alpha)g(x_2). \tag{4.35}$$

For each $x \in R^1$, let

$$C_x = \{y|(x,y) \in C\}.$$

Since C is convex, each C_x is a convex subset of R^1 (possibly empty).

However, C_{x_i}, $i = 1,2$ has positive Lebesgue measure because $g(x_i) = l(C_{x_i}) > 0$, $i = 1,2$ where l denotes Lebesgue measure. Thus, C_{x_i} has a non-empty interior which is an open interval--say (a_i,b_i) with $a_i < b_i$, $i = 1,2$. Thus $g(x_i) = b_i - a_i$, $i = 1,2$.

Now, we claim that

$$C_{\alpha x_1+(1-\alpha)x_2} \supseteq \alpha C_{x_1} + (1-\alpha)C_{x_2} \tag{4.36}$$

where the right hand side denotes the set of all points of the form $\alpha y_1 + (1-\alpha)y_2$ with $y_1 \in C_{x_1}$ and $y_2 \in C_{x_2}$. To verify the containment (4.36), observe that if $\alpha y_1 + (1-\alpha)y_2 \in \alpha C_{x_1} + (1-\alpha)C_x$ with $y_i \in C_{x_i}$, $i = 1,2$, then $(x_i,y_i) \in C$, $i = 1,2$ so

$$\alpha(x_1,y_1) + (1-\alpha)(x_2,y_2) = (\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \in C$$

by the convexity of C. Hence $\alpha y_1 + (1-\alpha)y_2$ is an element of $C_{\alpha x_1+(1-\alpha)x_2}$ by definition. Further, it is easy to show that the interior of $\alpha C_{x_1} + (1-\alpha)C_{x_2}$ is $(\alpha a_1+(1-\alpha)a_2, \alpha b_1+(1-\alpha)b_2)$. Thus,

$$\begin{aligned} g(\alpha x_1+(1-\alpha)x_2) &= l(C_{\alpha x_1+(1-\alpha)x_2}) \geq l(\alpha C_{x_1}+(1-\alpha)C_{x_2}) \\ &= \alpha b_1+(1-\alpha)b_2 - [\alpha a_1+(1-\alpha)a_2] = \alpha g(x_1)+(1-\alpha)g(x_2). \end{aligned}$$

Hence (4.35) holds. □

To complete the proof of Theorem 4.7, it must be shown that for $x_1,x_2 \in R^1$ and $\alpha \in (0,1)$, that

$$h(\alpha x_1+(1-\alpha)x_2) \geq h^{\alpha}(x_1)h^{1-\alpha}(x_2) \tag{4.37}$$

where h is given by (4.33) and f is log concave, bounded and has compact support. Obviously, we can assume $h(x_i) > 0$, $i = 1,2$ since otherwise (4.37) is trivial. Now, without loss of generality, assume that

$$\sup_y f(x_1,y) = \sup_y f(x_2,y) = c. \tag{4.38}$$

Remark. If the two suprema in (4.38) are not the same, replace f(x,y) in (4.33) by $\tilde{f}(x,y) = e^{bx}f(x,y)$ where b is a constant. Then $\tilde{f}$ is log concave, bounded and has compact support. Further h(x) becomes $e^{bx}h(x)$ so (4.37) remains the same. Clearly b can be chosen so (4.38) holds, since the two suprema are not 0 as $h(x_i) > 0$, $i = 1,2$. □

Note that $0 < c < +\infty$. For each $a > 0$, let

$$D(a) = \{(x,y) \mid f(x,y) \geq a\}$$

and note that

$$f(x,y) = \int_0^\infty I_{D(a)}(x,y)da$$

where $I_{D(a)}$ is the indicator function of the convex set D(a). Thus,

$$h(x) = \int_0^\infty\int_{-\infty}^\infty I_{D(a)}(x,y)dyda. \tag{4.39}$$

From the definition of c, it follows that

$$h(x_i) = \int_0^c\int_{-\infty}^\infty I_{D(a)}(x_i,y)dyda \tag{4.40}$$

for $i = 1,2$. Now fix $a \in (0,c)$ and let

$$g_a(x) = \int_{-\infty}^\infty I_{D(a)}(x,y)dy. \tag{4.41}$$

Since $a < c$, the log concavity of f implies that $\{y \mid f(x_i,y) \geq a\}$ is a convex set with a non-empty interior, $i = 1,2$. Hence

$$g_a(x) = l\{y \mid f(x,y) \geq a\}$$

satisfies $g_a(x_i) > 0$, $i = 1,2$. Thus, Proposition 4.25 implies that for each $a \in (0,c)$, g_a is concave on the interval $[x_1,x_2]$. Using this, we have

$$\begin{aligned} h(\alpha x_1+(1-\alpha)x_2) &= \int_0^c\int_{-\infty}^{\infty} I_{D(a)}(\alpha x_1+(1-\alpha)x_2,y)\,dy\,da \\ &\geq \int_0^c \{\alpha \int_{-\infty}^{\infty} I_{D(a)}(x_1,y)\,dy + (1-\alpha)\int_{-\infty}^{\infty} I_{D(a)}(x_2,y)\,dy\}\,da \\ &= \alpha h(x_1)+(1-\alpha)h(x_2) \geq h^{\alpha}(x_1)h^{1-\alpha}(x_2). \end{aligned}$$

The last inequality follows from the arithmetic mean-geometric mean inequality. Thus, (4.37) holds and the proof is complete. □

Chapter 5: The FKG Inequality and Association

The underlying problem with which both the FKG inequality and the notion of association deal concerns the question of trying to capture what one means by the vague statement that the coordinates of a random vector are "positively related." The word "positively" is to be interpreted here in the sense that we use positive in the expression "positive correlation" (as opposed to negative correlation). For example, consider a random vector X in R^2 with coordinates X_1 and X_2. In words, the idea that X_1 and X_2 are "positively related" should mean that if we are told X_1 is large, then the chance that X_2 is also large should be increased by this knowledge about X_1. Naturally, the same should hold with X_1 and X_2 interchanged so "positively related" is a reflexive notion. In terms of a conditional probability statement, the above intuitive idea is simply expressed as

$$P(X_1 \ge x_1 | X_2 \ge x_2) \ge P(X_1 \ge x_1) \tag{5.1}$$

which is to hold for all $x_1, x_2 \in R^1$. The condition

$$P(X_1 \ge x_1,\ X_2 \ge x_2) \ge P(X_1 \ge x_1)P(X_2 \ge x_2) \tag{5.2}$$

is equivalent to (5.1) and is symmetric in X_1 and X_2. This condition is discussed in Lehmann (1966) in his study of different notions of bivariate dependence.

A condition stronger than (5.1) is also suggested by an intuitive argument. Consider sets

$$B_1 = \{u | u \in R^2,\ \ u_i \ge x_i,\ \ i = 1,2\}$$
$$B_2 = \{u | u \in R^2,\ \ u_i \ge y_i,\ \ i = 1,2\}.$$

If X_1 and X_2 are "positively related," then how does the information "$X \in B_2$" affect the probability of B_1? Since "positively related" ought to mean that X_1 and X_2 tend to be relatively large together (and relatively small together), conditioning on the event "$X \in B_2$" should increase the probability of B_1. Again, in terms of conditional probability, this condition is just

$$P(B_1|B_2) \geq P(B_1)$$

which is equivalent to

$$P(B_1 \cap B_2) \geq P(B_1)P(B_2)$$

and this is equivalent to

$$EI_{B_1}(X)I_{B_2}(X) \geq EI_{B_1}(X)EI_{B_2}(X). \tag{5.3}$$

Here I_{B_i} denotes the indicator function of the set B_i, so (5.3) simply says that the covariance between I_{B_1} and I_{B_2} is non-negative for all sets B_1 and B_2 of the form indicated above. But, for any set B of the form

$$B = \{u | u \in R^2, \quad u_i \geq x_i, \quad i = 1,2\},$$

I_B is non-decreasing in each coordinate variable (with the other one held fixed). This suggests an even stronger condition than (5.3) as a candidate for the definition of "positively related." The coordinates X_1 and X_2 are <u>associated</u> (Esary, Proschan and Walkup (1967)) if

$$\text{cov}\{f_1(X), f_2(X)\} \geq 0 \tag{5.4}$$

for all functions f_i which are non-decreasing in each coordinate variable (with the other coordinate held fixed). This definition has an obvious extension to higher dimensions and is discussed in detail in this chapter. Sarkar (1969) and Fortuin, Ginibre and Kasteleyn (1971) gave sufficient conditions for (5.4) hold. The inequality (5.4) is often called the FKG inequality because of the work of Fortuin, Ginibre and Kasteleyn.

A second problem which is easy to motivate, but whose connection with the intuitive notion of "positively related" is not so clear, concerns the extension of stochastic ordering (on R^1) to higher dimensions. Recall from Proposition 1.1, that for two real valued random variables Z_1 and Z_2, Z_1 is stochastically smaller than Z_2 iff

$$Ef(Z_1) \le Ef(Z_2) \tag{5.5}$$

for all non-decreasing functions f defined on R^1 for which the expectations are defined. To extend things to R^n, introduce a partial ordering on R^n defined as follows:

$$x \le y \quad \text{iff } x_i \le y_i,\ i = 1, \ldots, n. \tag{5.6}$$

A real valued function f defined on R^n is <u>non-decreasing</u> if $x \le y$ implies $f(x) \le f(y)$. Clearly, f is non-decreasing iff f is non-decreasing in each coordinate variable (with the remaining variables held fixed). If X and Y are random vectors in R^n, then X is <u>stochastically smaller</u> than Y if

$$Ef(X) \le Ef(Y) \tag{5.7}$$

for all non-decreasing functions f defined on R^n for which the expectations exist. The problem is to give some useful sufficient conditions so that (5.7) holds. The conditions yielding (5.7) which are discussed in Preston (1974), Holley (1974), Kemperman (1977) and Edwards (1978) turn out to be very closely connected with conditions which yield inequality (5.4) on R^n. These conditions which are the principal topic in this chapter are also related to multivariate extensions of monotone likelihood ratio and are discussed at length below.

<u>Section 1: Association</u>

In this section, the basic properties of associated random variables are given following the original development in Esary, Proschan and Walkup (1967).

<u>Definition 5.1</u>: A real valued function f defined on R^n is <u>coordinatewise non-decreasing</u> if f is non-decreasing in each coordinate when the remaining coordinates are held fixed.

In terms of the partial ordering defined in (5.6), f is non-decreasing iff f is non-decreasing relative to the partial ordering. In what follows, we will use both of the terms non-decreasing and coordinatewise non-decreasing. Now, let X be a random vector in R^n with coordinates $X_1, X_2, \ldots, X_n$.

Definition 5.2: The random variables $X_1, \ldots, X_n$ are associated if

$$\mathrm{cov}\{f_1(X), f_2(X)\} \geq 0 \tag{5.8}$$

for all bounded coordinatewise non-decreasing functions f_1 and f_2. When $X_1, \ldots, X_n$ are associated, then we say that X is associated.

Remark 5.1: This definition is equivalent to the more usual definition of association which stipulates that (5.8) hold for all coordinatewise non-decreasing f_1 and f_2 for which the expectations exists. To see this, consider any coordinatewise non-decreasing f and set

$$f_M(x) = \begin{cases} M & \text{if } f(x) \geq M \\ f(x) & \text{if } -M < f(x) < M \\ -M & \text{if } f(x) \leq -M, \end{cases}$$

for $M > 0$. Clearly $|f_M(x)| \leq |f(x)|$ and

$$\lim_{M\to\infty} f_M(x) = f(x), \qquad x \in R^n.$$

Also, it is easily verified that f_M is coordinatewise non-decreasing. Hence if $|f(X)|$ has a finite expectation, then the dominated convergence theorem yields

$$\lim_{M\to\infty} Ef_M(X) = Ef(X).$$

Hence if (5.8) holds for all bounded coordinatewise non-decreasing f_1 and f_2, the dominated convergence theorem shows (5.8) holds for all coordinatewise non-decreasing f_1 and f_2 for which all the expectations are well defined. The boundedness condition in Definition 5.2 removes some annoying technical issues. □

One important consequence of association is the inequality

$$P\{X_i \geq a_i, i = 1,\ldots, n\} \geq P\{X_i \geq a_i, i = 1,\ldots, k\}P\{X_i \geq a_i, i = k+1,\ldots, n\} \quad (5.9)$$

which is valid for $k = 1,\ldots, n-1$, when $X_1,\ldots, X_n$ are associated. This follows from (5.8) by taking f_1 to be the indicator function of

$$B_1 = \{x | x \in R^n, \quad x_i \geq a_i, \quad i = 1,\ldots, k\}$$

and f_2 to be the indicator function of

$$B_2 = \{x | x \in R^n, \quad x_i \geq a_i, \quad i = k+1,\ldots, n\}.$$

For associated random variables, (5.9) and an induction argument establish

$$P\{X_i \geq a_i, \; i = 1,\ldots, n\} \geq \prod_{i=1}^{n} P\{X_i \geq a_i\}. \quad (5.10)$$

Here are some basic observations which allow the construction of associated random variables. First observe that for $n = 1$, X is always associated. In fact, this is a consequence of the following inequality due to Tchebyshev.

Proposition 5.1: Consider two functions h_1 and h_2 defined on R^1 which satisfy

$$(h_1(u)-h_1(v))(h_2(u)-h_2(v)) \geq 0 \quad (5.11)$$

for $u,v \in R^1$ (such functions are often called similarly ordered). For any random variable $Z \in R^1$ for which the expectations are defined,

$$\mathrm{cov}\{h_1(Z),h_2(Z)\} \geq 0. \tag{5.12}$$

Proof: Let W be an independent copy of Z. From (5.11)

$$E(h_1(Z)-h_1(W))(h_2(Z)-h_2(W)) \geq 0$$

which yields

$$2[Eh_1(Z)h_2(Z) - Eh_1(Z)Eh_2(Z)] \geq 0.$$

Hence (5.12) holds. □

Since (5.11) holds for h_1 and h_2 non-decreasing, $Z \in R^1$ is associated.

Proposition 5.2: Suppose $X \in R^m$ is associated and $Y \in R^n$ is associated. If X and Y are independent, then the random vector $(X,Y) \in R^{m+n}$ is associated. Further, if $U \in R^n$ has independent coordinates, then U is associated.

Proof: Let f_1 and f_2 be bounded non-decreasing functions on R^{m+n} and set

$$h_i(y) = E_X f_i(X,y), \qquad y \in R^n$$

for $i = 1,2$. Then h_i is bounded and is non-decreasing on R^n. Using the independence of X and Y and the assumption that X is associated, we have

$$\begin{aligned}\mathrm{cov}\{f_1(X,Y),f_2(X,Y)\} &= E_Y E_X f_1(X,Y)f_2(X,Y) - E_Y h_1(Y)E_Y h_2(Y)\\ &\geq E_Y[(E_X f_1(X,Y))(E_X f_2(X,Y))] - E_Y h_1(Y)E_Y h_2(Y) =\\ &= E_Y h_1(Y)h_2(Y) - E_Y h_1(Y)E_Y h_2(Y).\end{aligned}$$

However, the final term in the above expression is non-negative because Y is associated. Thus, the first assertion holds. The second assertion follows from the first via an easy induction argument. □

Proposition 5.3: If $X_1, \ldots, X_n$ are associated, then any subset of $X_1, \ldots,$

X_n is also associated.

Proof: This is clear. □

Proposition 5.4: Suppose $X \in R^m$ is associated and $h_1, \ldots, h_n$ are all non-decreasing functions defined on R^m. With $Y_i = h_i(X)$, $i = 1, \ldots, n$, the random vector $Y \in R^n$ with coordinates $Y_1, \ldots, Y_n$ is associated.

Proof: This follows from the following observation. If f defined on R^n is non-decreasing, then $\tilde{f}$ defined on R^m by

$$\tilde{f}(x) = f(h_1(x), \ldots, h_n(x)), \quad x \in R^m$$

is non-decreasing. Thus, for bounded non-decreasing f_1 and f_2 on R^n,

$$\tilde{f}_i(x) = f_i(h_1(x), \ldots, h_n(x)), \quad x \in R^m$$

is bounded and non-decreasing. Since X is associated, we have

$$\text{cov}\{f_1(Y), f_2(Y)\} = \text{cov}\{\tilde{f}_1(X), \tilde{f}_2(X)\} \geq 0$$

so Y is associated. □

In general, it is not an easy matter to decide whether or not a random vector $X \in R^n$ is associated. If the covariance matrix $\Sigma = \{\sigma_{ij}\}$ of X exists, then certainly each σ_{ij} must be non-negative since $\sigma_{ij} = \text{cov}\{X_i, X_j\}$. In the case that X has a multivariate normal distribution, it was only proved in the past few years (Pitt (1982)) that X is associated when $\sigma_{ij} \geq 0$. Pitt's proof is given later in this chapter.

Section 2: Extensions of MLR: Motivation

In this section we give a rather "soft argument" which yields a portion of a sufficient condition so that (5.7) holds. To describe this condition, first consider the usual lattice operations on R^n which are defined as follows. For vector x and y in R^n with coordinates $x_1, \ldots, x_n$ and $y_1, \ldots,$

y_n, $x \wedge y \in R^n$ has coordinates $\min(x_i, y_i)$, $i = 1, \ldots, n$ and $x \vee y \in R^n$ has coordinates $\max(x_i, y_i)$, $i = 1, \ldots, n$.

Suppose $X \in R^n$ has a density p_1 and $Y \in R^n$ has a density p_2, both densities with respect to Lebesgue measure. The problem is to find reasonable conditions under which

$$Ef(X) = \int f(x)p_1(x)dx \le \int f(x)p_2(x)dx = Ef(Y) \tag{5.13}$$

for all bounded non-decreasing functions f. The intuitive content of (5.13) is the random vector Y tends to be larger than X in the partial ordering on R^n given by (5.6).

Consider the following statistical problem. A data matrix $A{:}n\times 2$, having columns u and v, is given. The matrix A arose by taking one observation on X and one observation on Y. However, for each coordinate, the labels indicating which observation is on X and which is on Y, were lost. That is, in the ith row of A are observation u_i and v_i but we do not know which of these came from the X population and which from the Y population. The statistical problem is to unscramble the data. In other words, for each i say which coordinate of the ith row of A came from X and which came from Y.

Example 5.1: For two heart patients, suppose we have three measurements

$$\begin{array}{l} \text{systolic blood pressure} \\ \text{resting pulse rate} \\ \text{cholesteral count} \end{array} \quad \begin{pmatrix} 131 & 142 \\ 83 & 74 \\ 317 & 282 \end{pmatrix} = A$$

However, for a given variable, we do not know which measurement came from which patient. The problem is to assign measurements to patients. □

If we believe that Y tends to be bigger than X, then a plausible assignment is: In each row assign the larger observation to Y and the smaller observation to X. Assuming the two observations were independent, a likelihood justification of this method of assignment would run as follows. Let $\theta = (\theta_1, \ldots \theta_n)$ consist of n 2×2 matrices θ_i where each θ_i:

2×2 is either

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Let $\theta(A)$ have ith row $(u_i, v_i)\theta_i$ where (u_i, v_i) is the ith row of A.

If x is an observation on X, y is an observation on Y, and B:n×2 has first column x and second column y, then the likelihood of B is

$$h(B) = p_1(x)p_2(y). \tag{5.14}$$

Thus, if θ represents the true assignment of the data to the proper population, then

$$h(\theta(A)) = L(\theta) \tag{5.15}$$

is the likelihood function. Now, consider the condition

$$p_1(u)p_2(v) \le p_1(u \wedge v)p_2(u \vee v). \tag{5.16}$$

When (5.16) holds, then the likelihood function in (5.15) is maximized by any θ which results in $\theta(A)$ having first column $u \wedge v$ and second column $u \vee v$. Conversely, if L is maximized at such a θ, then (5.16) holds. Thus condition (5.16) is precisely the assumption that yields a maximum likelihood estimator of θ which corresponds to asserting that $\min\{u_i, v_i\}$ came from the ith coordinate of the X population, $i = 1, \ldots, n$.

This argument shows that (5.16) is a plausible candidate for trying to capture what one means (in R^n) by saying that "X tends to be smaller than Y." In the next section it is shown that (5.16) together with an assumption concerning the dominating measure for p_1 and p_2 (which is satisfied by Lebesgue measure) yields (5.13).

When $p_1 = p_2$ in (5.16), the condition becomes

$$p(u)p(v) \le p(u \wedge v)p(u \vee v). \tag{5.17}$$

In this case, the interpretation given above is no longer valid, but an alternative interpretation is possible. With A and θ as above, (5.17) means that the likelihood is maximized by a θ such that $\theta(A)$ has first column $u \wedge v$ and second column $u \vee v$. Thus, the likelihood is maximized by rearranging A in such a way all the smaller coordinates are together and all the larger coordinates are together. In other words, the coordinates of a random vector drawn from p tend to be "positively related" in the rather vague sense described in the introduction to this chapter. As is shown in the next section, (5.17) together with an assumption on a dominating measure implies (5.4).

Finally, we relate (5.17) to monotone likelihood ratio as discussed in Remark 1.2. Here is the classical definition of a TP_2 (totally positive of order 2) function.

Definition 5.3: Let r be a non-negative valued function defined on $X \times Y$ where X and Y are non-empty subsets of R^1. If for all $x_1 \le x_2$ in X and $y_1 \le y_2$ in Y

$$r(x_1,y_2)r(x_2,y_1) \le r(x_1,y_1)r(x_2,y_2) \tag{5.18}$$

then r is TP_2.

Of course, this is just the definition of MLR given in Chapter 1, but the interpretation of the second argument of r as a parameter has been removed.

Proposition 5.5: The function r is TP_2 iff r satisfies (5.17) for all u and v in $X \times Y$.

Proof: To show (5.17) implies (5.18), take $u = (x_1,y_2)$ and $v = (x_2,y_1)$, with $x_1 \le x_2$ and $y_1 \le y_2$. Then $u \wedge v = (x_1,y_1)$ and $u \vee v = (x_2,y_2)$ so (5.17) yields (5.18). For the converse, consider $u = (\alpha_1,\beta_1) \in X \times Y$ and $v = (\alpha_2,\beta_2) \in X \times Y$. Without loss of generality, assume $\alpha_1 \le \alpha_2$ (otherwise interchange u and v). There are two cases.

Case (i): $\beta_1 \leq \beta_2$. In this case $u \wedge v = u$ and $u \vee v = v$ so (5.17) holds trivially.

Case (ii): $\beta_2 < \beta_1$. In this case $u \wedge v = (\alpha_1, \beta_2)$ and $u \vee v = (\alpha_2, \beta_1)$. With $x_1 = \alpha_1$, $x_2 = \alpha_2$, $y_1 = \beta_2$ and $y_2 = \beta_1$, (5.18) yields (5.17). □

Thus, when $n = 2$, condition (5.17) is nothing but MLR or equivalently TP_2. For $n > 2$, functions which satisfy (5.17) are said to be multivariate totally positive of order 2 (MTP_2). These are discussed further below.

Section 3: The Basic Inequality

In this section we establish an inequality due to Ahlswede and Daykin (1979) which yields a sufficient condition for both (5.4) and (5.13). To formulate the inequality, let $\mathcal{X}^{(n)}$ be a product space

$$\mathcal{X}^{(n)} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_n$$

where each $\mathcal{X}_i$ is a Borel subset of R^1. Also, let $\mu = \mu_1 \times \mu_2 \times \ldots \times \mu_n$ be a product measure on $\mathcal{X}^{(n)}$ where μ_i is a σ-finite measure on $\mathcal{X}_i$, $i = 1, \ldots, n$.

Theorem 5.6 (Ahlswede and Daykin (1979)). Suppose p_i, $i = 1, \ldots, 4$ are non-negative functions defined on $\mathcal{X}^{(n)}$ which satisfy

$$p_1(x)p_2(y) \leq p_3(x \wedge y)p_4(x \vee y) \tag{5.19}$$

for $x, y \in \mathcal{X}^{(n)}$. Then the inequality

$$\int p_1 d\mu \int p_2 d\mu \leq \int p_3 d\mu \int p_4 d\mu \tag{5.20}$$

holds.

The induction proof of Theorem 5.6 which follows is due to Karlin and Rinott (1980). The next two lemmas constitute the essentials of the induction argument.

Lemma 5.7: In Theorem 5.6, assume that $n = 1$ and (5.19) holds. Then

(5.20) holds.

Proof: Suppressing "$d\mu$" under the integrals, the left hand side of (5.20) can be written

$$\iint_{\{x<y\}} p_1(x)p_2(y) + \iint_{\{x>y\}} p_1(x)p_2(y) + \iint_{\{x=y\}} p_1(x)p_2(y)$$
$$= \iint_{\{x<y\}} [p_1(x)p_2(y) + p_1(y)p_2(x)] + \iint_{\{x=y\}} p_1(x)p_2(x)$$

with a similar expression for the right hand side of (5.20). On the set $\{x=y\}$, (5.19) obviously yields $p_1(x)p_2(x) \le p_3(x)p_4(x)$ so

$$\iint_{\{x=y\}} p_1(x)p_2(x) \le \iint_{\{x=y\}} p_3(x)p_4(x)$$

Thus, it suffices to show that

$$\iint_{\{x<y\}} [p_1(x)p_2(y)+p_1(y)p_2(x)] \le \iint_{\{x<y\}} [p_3(x)p_4(y)+p_3(y)p_4(x)] \tag{5.21}$$

which is accomplished with the following pointwise argument. Set $a = p_1(x)p_2(y)$, $b = p_1(y)p_2(x)$, $c = p_3(x)p_4(y)$ and $d = p_3(y)p_4(x)$. Since $x < y$ in (5.21), (5.19) yields $a \le c$ and $b \le c$. But, (5.19) also gives $ab \le cd$. However,

$$c+d - (a+b) = (1/c)[(c-a)(c-b) + (cd-ab)]$$

which is non-negative. Hence (5.21) holds by a pointwise comparison of the two integrands. □

The crucial step in the induction is given in the next result.

Lemma 5.8: Suppose p_i, $i = 1, \ldots, 4$ defined on $X^{(n)}$ satisfy (5.19). For $x \in X^{(n)}$, write $x = (u,s)$ with $u \in X^{(n-1)}$ and $s \in X_n$. Define g_i on $X^{(n-1)}$ by

$$g_i(u) = \int_{X_n} p_i(u,s)\mu_n(ds) \tag{5.22}$$

for i = 1,..., 4. Then

$$g_1(u)g_2(v) \le g_3(u\wedge v)g_4(u\vee v) \tag{5.23}$$

for $u,v \in X^{(n-1)}$.

<u>Proof</u>: For notational convenience, the range of integration (namely X_n) is suppressed in the integrals below and we write ds for $\mu_n(ds)$. With this notation, for $u,v \in X^{(n-1)}$,

$$\begin{aligned} g_1(u)g_2(v) &= \iint p_1(u,s)p_2(v,t)dsdt = \iint_{\{s<t\}} p_1(u,s)p_2(v,t)dsdt \\ &+ \iint_{\{s>t\}} p_1(u,s)p_2(v,t)dsdt + \iint_{\{s=t\}} p_1(u,s)p_2(v,t)dsdt \\ &= \iint_{\{s<t\}} [p_1(u,s)p_2(v,t)+p_1(u,t)p_2(v,s)]dsdt + \iint_{\{s=t\}} p_1(u,s)p_2(v,t)dsdt \end{aligned} \tag{5.24}$$

with a similar expression holding for $g_3(u\wedge v)g_4(u\vee v)$. With x = (u,s) and y = (v,s), (5.19) implies that

$$p_1(u,s)p_2(v,s) \le p_3(u\wedge v,s)p_4(u\vee v,s). \tag{5.25}$$

Integration of (5.25) with respect to s yields

$$\iint_{\{s=t\}} p_1(u,s)p_2(v,t)dsdt \le \iint_{\{s=t\}} p_3(u\wedge v,s)p_4(u\vee v,t)dsdt.$$

Thus, to establish (5.23) is suffices to show that

$$\iint_{\{s<t\}} [p_1(u,s)p_2(v,t) + p_1(u,t)p_2(v,s)]dsdt$$
$$\leq \iint_{\{s<t\}} [p_3(u\wedge v,s)p_4(u\vee v,t) + p_3(u\wedge v,t)p_4(u\vee v,s)]dsdt. \tag{5.26}$$

This inequality is established using a pointwise argument as in the proof of Lemma 5.7. First, let $a = p_1(u,s)p_2(v,t)$, $b = p_1(u,t)p_2(v,s)$, $c = p_3(u\wedge v,s)p_4(u\vee v,t)$ and $d = p_3(u\wedge v,t)p_4(u\vee v,s)$. Since $s < t$ on the range of integration, condition (5.19) yields

$$a \leq c \qquad \text{and} \qquad b \leq c$$

by first taking $x = (u,s)$, $y = (v,t)$ and then taking $x = (u,t)$, $y = (v,s)$. However, the inequality

$$ab \leq cd$$

also follows easily from (5.19). Hence $c+d - (a+b) = c^{-1}[(c-a)(c-b) + (cd-ab)]$ which is non-negative. Thus (5.26) holds by a pointwise argument. □

Proof of Theorem 5.6: By Lemma 5.7, the Theorem holds for $n = 1$. Assume the result holds for $k = 1,\ldots, n-1$ and consider the assertion for $k = n$. Since the measure μ is a product measure, inequality (5.20) can be written

$$\int g_1 d\tilde{\mu} \int g_2 d\tilde{\mu} \leq \int g_3 d\tilde{\mu} \int g_4 d\tilde{\mu} \tag{5.27}$$

where the integrals are over $\mathcal{X}^{(n-1)}$,

$$\tilde{\mu} = \mu_1 \times\ldots\times \mu_{n-1},$$

and g_i is defined in (5.22). But (5.27) holds by Lemma 5.8 and the induction hypothesis, since the g_i, $i = 1,\ldots, 4$ satisfy (5.23). □

Theorem 5.9: (Preston (1974), Holley (1974), Kemperman (1977), Edwards (1978)). Suppose P_1 and P_2 are probability measures on $\mathcal{X}^{(n)}$. Further, assume there exists a product measure μ on $X^{(n)}$ such that P_i has a density ψ_i with respect to μ and

$$\psi_1(x)\psi_2(y) \le \psi_1(x \wedge y)\psi_2(x \vee y) \tag{5.28}$$

for all $x,y \in X^{(n)}$. Then

$$\int f(x)\psi_1(x)\mu dx \le \int f(x)\psi_2(x)\mu(dx) \tag{5.29}$$

for all coordinatewise non-decreasing functions f for which the two integrals exist.

Proof: First assume f is non-negative and set $p_1 = f\psi_1$, $p_2 = \psi_2$, $p_3 = \psi_1$ and $p_4 = \psi_2 f$. Then condition (5.19) is easily verified since f is non-negative and non-decreasing. Inequality (5.20) yields (5.29) since ψ_1 and ψ_2 are densities.

When f is bounded below by c, then f(x) + c is non-negative and non-decreasing. The first case yields

$$\int (f(x)+c)\psi_1(x)\mu(dx) \le \int (f(x)+c)\psi_2(x)\mu(dx)$$

which in turn gives (5.29) since ψ_1 and ψ_2 are densities. The general case is treated by a standard truncation and limiting argument. □

Theorem 5.10: (Sarkar (1969), Fortuin, Ginibre and Kasteleyn (1971)). Assume the random vector $X \in X^{(n)}$ has a density p with respect to a product measure μ. If p satisfies

$$p(x)p(y) \le p(x \wedge y)p(x \vee y) \tag{5.30}$$

for all $x,y \in X$, then X is associated. That is, $\mathrm{cov}\{f_1(X),f_2(X)\} \ge 0$ for all f_1 and f_2 which are bounded and non-decreasing.

Proof: Because $\mathrm{cov}\{\cdot,\cdot\}$ is invariant under translations of its arguments, we can assume that f_1 and f_2 are strictly positive without loss of generality. Hence

$$c = \int f_1 p du > 0.$$

With $p_2 = c^{-1} f_1 p$, $p_1 = p$ and $f = f_2$, Theorem (5.9) yields

$$\int f_2(x)p(x)\mu(dx) \quad \le \quad c^{-1}\int f_2(x)f_1(x)p(x)\mu(dx)$$

which is just the assertion that $\mathrm{cov}\{f_1(X), f_2(X)\} \ge 0$. □

Section 4: Multivariate Total Positivity

In general it can be rather difficult and tedious to check (5.30) for a density p. This section contains some useful criteria and examples which can facilitate the verification of (5.30) or (5.28).

Again consider a product space $\mathcal{X}^{(n)} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ where each $\mathcal{X}_i$ is a Borel subset of R^1.

Definition 5.4: A non-negative real valued function f defined on $\mathcal{X}^{(n)}$ is multivariate totally positive of order 2 (MTP_2) if

$$f(x)f(y) \le f(x \wedge y)f(x \vee y)$$

for all $x,y \in \mathcal{X}^{(n)}$.

Definition 5.5: A non-negative real valued function f defined on $\mathcal{X}^{(n)}$ is totally positive of order 2 in pairs (TP_2 in pairs) if for each pair of variables (with the remaining n-2 variables held fixed), f is TP_2 (according to Definition 5.3).

If f defined on $\mathcal{X}^{(n)}$ can be written in the form

$$f(x) = \prod_1^n g_i(x_i) \tag{5.31}$$

where each g_i is a non-negative function defined on $\mathcal{X}_i$, $i = 1, \ldots, n$, then clearly f is MTP_2. Hence, a density (with respect to a product measure) of independent random variables is MTP_2.

The argument used in Proposition 5.5 shows that if f is MTP_2, then f is TP_2 is pairs. That the converse is true under certain conditions is essentially due to Lorentz (1953).

Proposition 5.11: Suppose that f is TP_2 in pairs. Also assume that if $f(x)f(y) > 0$, then for each vector $z \in \mathcal{X}^{(n)}$ satisfying $x \wedge y \le z \le x \vee y$, $f(z) > 0$. Then f is MTP_2.

Proof: The argument used here is from Kemperman (1977). For $x,y \in \mathcal{X}^{(n)}$, it must be shown that

$$f(x)f(y) \le f(x \wedge y)f(x \vee y). \tag{5.32}$$

If $f(x)f(y) = 0$ then (5.32) holds so assume $f(x)f(y) > 0$. Let $u_i = \min\{x_i,y_i\}$ and $v_i = \max\{x_i,y_i\}$, $i = 1,\ldots, n$ where $x_1,\ldots, x_n$ and $y_1,\ldots, y_n$ are the coordinates of x and y. Writing x and y as row vectors, we can assume without loss of generality that

$$x = (v_1,\ldots,v_r, u_{r+1},\ldots, u_n)$$
$$y = (u_1,\ldots,u_r, v_{r+1},\ldots, v_n)$$

where $1 \le r \le n-1$. With $s = n-r$ and for $0 \le i \le r$, $0 \le j \le s$, let

$$x_{i,j} = (v_1,\ldots,v_i, u_{i+1},\ldots,u_r, v_{r+1},\ldots,v_{r+j}, u_{r+j+1},\ldots,u_n)$$

so $x_{i,j} \in \mathcal{X}^{(n)}$. Observe that $x_{0,0} = x \wedge y$, $x_{r,0} = x$, $x_{0,s} = y$, $x_{r,s} = x \vee y$, and $x \wedge y \le x_{i,j} \le x \vee y$.

Now $x_{i+1,j}$ and $x_{i,j+1}$ differ in at most two coordinates. Since f is TP_2 in pairs, this yields

$$f(x_{i+1,j})f(x_{i,j+1}) \le f(x_{i,j})f(x_{i+1,j+1}) \tag{5.33}$$

as a direct computation verifies. However the identity

$$\frac{f(x)f(y)}{f(x\wedge y)f(x\vee y)} = \prod_{i=0}^{r-1}\prod_{j=0}^{s-1} \frac{f(x_{i+1,j})f(x_{i,j+1})}{f(x_{i,j})f(x_{i+1,j+1})} \tag{5.34}$$

is easily established. But, (5.33) shows each term in the product is bounded above by 1 so (5.34) is bounded above by 1. Thus (5.32) holds. □

In general, TP_2 in pairs does not imply MTP_2. A counter example is given in Kemperman (1977). A rather complete discussion of this issue can be found in Perlman and Olkin (1980).

Here are some further observations which can be of help in verifying (5.30).

Proposition 5.12: If $f_1, \ldots, f_k$ are MTP_2 on $X^{(n)}$, then $f = \Pi f_i$ is MTP_2 on $X^{(n)}$.

Proof: Elementary. □

Proposition 5.13: For $1 < r < n$, suppose f defined on $X^{(r)}$ is MTP_2. Extend the definition of f to $X^{(n)}$ by

$$\tilde{f}(y,z) = f(y) \tag{5.35}$$

where $y \in X^{(r)}$ and $z \in X_{r+1} \times \ldots \times X_n$. Then $\tilde{f}$ is MTP_2 on $X^{(n)}$.

Proof: Elementary. □

Proposition 5.14: If f is MTP_2 on $X^{(n)}$ and $\mu = \mu_1 \times \ldots \times \mu_n$ is a product measure on X^n, for $1 < r < n$, define g on $X^{(r)}$ by

$$g(u) = \int \ldots \int f(u, x_{r+1}, \ldots, x_n)\mu_{r+1}(dx_{r+1}) \ldots \mu_n(dx_n).$$

Then g is MTP_2 on $X^{(r)}$.

Proof: An easy induction argument together with Lemma 5.8 yields the assertion. □

The following composition result should be compared to Theorem 3.9. In the TP_2 case, this proposition is sometimes called Karlin's Lemma (Karlin (1956)). For some applications of this in multivariate analysis, see Eaton (1983).

<u>Proposition 5.15</u>: On three product spaces $\mathcal{X}^{(p)}$, $\mathcal{Y}^{(q)}$, and $\mathcal{Z}^{(r)}$, suppose f(x,y) is MTP_2 on $\mathcal{X}^{(p)}\times\mathcal{Y}^{(q)}$ and suppose g(y,z) is MTP_2 on $\mathcal{X}^{(q)}\times\mathcal{Z}^{(r)}$. If ν is a σ-finite product measure on $\mathcal{Y}^{(q)}$, then

$$h(x,z) = \int f(x,y)g(y,z)\nu(dy) \tag{5.36}$$

is MTP_2 on $\mathcal{X}^{(p)}\times\mathcal{Z}^{(r)}$.

<u>Proof</u>: Extend the definition of f and g to $\mathcal{X}^{(p)}\times\mathcal{Y}^{(q)}\times\mathcal{Z}^{(r)}$ by

$$\begin{aligned} \tilde{f}(x,y,z) &= f(x,y) \\ \tilde{g}(x,y,z) &= g(y,z). \end{aligned} \tag{5.37}$$

Then $\tilde{f}$ and $\tilde{g}$ are MTP_2 on $\mathcal{X}^{(p)}\times\mathcal{Y}^{(q)}\times\mathcal{Z}^{(r)}$ by Proposition 5.13. Thus $\tilde{f}\tilde{g}$ is MTP_2 by Proposition 5.12. Integrating $\tilde{f}\tilde{g}$ over $\mathcal{Y}^{(q)}$ with respect to ν yields a MTP_2 function on $\mathcal{X}^{(p)}\times\mathcal{Z}^{(r)}$ by Proposition 5.14. This function is just h in (5.36). □

In the case of TP_2, there are a couple of useful criteria which together with Proposition 5.11 can be used to check for MTP_2.

<u>Proposition 5.16</u>: Suppose $\mathcal{X}^{(2)} = \mathcal{X}_1\times\mathcal{X}_2$ where $\mathcal{X}_i$ is an open subset of R^1, i = 1,2. If f is strictly positive on $\mathcal{X}^{(2)}$ and if f has a mixed partial derivative, then f is TP_2 iff

$$\frac{\partial^2}{\partial x_1 \partial x_2} \log f(x_1,x_2) \geq 0. \tag{5.38}$$

Proof: This well known criterion is given in Problem 6 in Lehmann (1959), p. 111. □

Proposition 5.17: Suppose a non-negative function h is defined on the difference set $\mathcal{X}_1 - \mathcal{X}_2$. Then f defined on $\mathcal{X}_1 \times \mathcal{X}_2$ by

$$f(x_1, x_2) = h(x_1 - x_2)$$

is TP_2 iff h is log concave.

Proof: The proof is left to the reader (see Lorentz (1953)). □

Example 5.1: Let h_i be a log concave function defined on R^1, $i = 1, \ldots, n$. Define f on $R^n \times R^n$ by

$$f(x,y) = \prod_{i=1}^{n} h_i(x_i - y_i). \tag{5.39}$$

To see that f is MTP_2 on $R^n \times R^n$, first use Proposition 5.17 to conclude that $h_i(x_i - y_i)$ is TP_2 on $R^1 \times R^1$ and hence its extension via (5.35) is MTP_2 on $R^n \times R^n$. By Proposition 5.12, f is MTP_2 as claimed. Note that f is actually a function of x-y. □

Example 5.2: For $x \in R^1$, define

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \le 0. \end{cases}$$

Since h is the indicator function of a convex set, h is log concave so h(x-y) is TP_2 on R^2. Using the argument given in Example 5.1, it follows that

$$f(u) = \prod_{1}^{n-1} h(u_i - u_{i+1}), \quad u \in R^n \tag{5.40}$$

which is the indicator function of $\{u | u \in R^n, u_1 > u_2 > \ldots > u_n\}$, is MTP_2.

If $X_1, \ldots, X_n$ are i.i.d. random variables with a density q on R^1 (with respect to Lebesgue measure), then the order statistic of $X_1, \ldots, X_n$, say $U_1 \geq U_2 \geq \ldots \geq U_n$ has a density on R^n (with respect to Lebesgue measure)

$$g(u) = \prod_1^n q(u_i) f(u)$$

It follows from Proposition 5.12 that g is MTP_2 since f given in (5.40) is MTP_2. □

Example 5.3: This example is from Dykstra and Hewitt (1978). On R^1, define h by

$$h(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

It is easily verified that h is log concave. Thus h(x-y) is TP_2 on R^2. Therefore, for $u \in R^n$ and for indices i and j, $i \neq j$, the function

$$u \rightarrow h(u_i - u_j)$$

is MTP_2 by Proposition 5.13. Therefore, by Proposition 5.12,

$$J(u) = \prod\prod_{i<j} h(u_i - u_j) \tag{5.41}$$

is MTP_2 on R^n.

Now, let S have a $W(I_p, p, n)$ distribution with $n \geq p$. It is known (see Anderson (1958)) that the density function of the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p > 0$ of S is

$$f(\lambda) = c \prod_{i=1}^p \lambda_i^\alpha \exp[-\tfrac{1}{2} \textstyle\sum_1^p \lambda_i] J(\lambda) I(\lambda) \tag{5.42}$$

where $\alpha = \frac{1}{2}(n-p-1)$, c is a constant, J is given by (5.41) and $I(\cdot)$ is the indicator function of $\{\lambda | \lambda \in R^p, \lambda_i > 0, i = 1, \ldots, p\}$. The function f in

(5.42) is a density with respect to Lebesgue measure on R^p. That f is MTP_2 follows from the fact that J is MTP_2 and easy applications of Proposition 5.13 and 5.12. □

Example 5.4: (Sarkar (1969)). Suppose $X \in R^p$ is $N_p(0,\Sigma)$ where Σ is non-singular. Thus, the density of X is

$$f(x) = (\sqrt{2\pi})^{-p}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}x'\Sigma^{-1}x]. \tag{5.43}$$

Since f is strictly positive, f is MTP_2 iff f is TP_2 in pairs. But, an application of Proposition 5.12 shows that f is TP_2 iff for each $i \neq j$,

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) = a_{ij} \geq 0$$

where a_{ij} is the (i,j) element of $-\frac{1}{2}\Sigma^{-1}$. Thus, f is MTP_2 iff the off diagonal elements of Σ^{-1} are non-positive. In particular, when $\Sigma = I_p$, f is MTP_2. Thus, it is natural to ask if there are other densities of the form

$$p(x) = h(\|x\|^2) \tag{5.44}$$

which are also MTP_2. For convenience, assume that $h > 0$ and h has two derivatives. Then p is MTP_2 iff p is TP_2 in pairs. Again, Proposition 5.12 shows p is TP_2 in pairs iff

$$\frac{\partial^2}{\partial x_i \partial x_j} \log h(\|x\|^2) = 4x_i x_j[h''(\|x\|^2)h(\|x\|^2) - (h'(\|x\|^2))^2]h^{-2}(\|x\|^2)$$

is non-negative for all $x \in R^p$. This is equivalent to the condition that

$$x_i x_j[h''(t)h(t) - (h'(t))^2] \geq 0 \tag{5.45}$$

for all $t > 0$ and all $x \in R^p$ with $\|x\|^2 = t$. But, if there is a $t > 0$ such that

$$h''(t)h(t) - (h'(t))^2 \neq 0, \tag{5.46}$$

then there is an x with $\|x\|^2 = t$ such that (5.45) is strictly negative. Hence p is MTP_2 iff

$$\frac{d^2}{dt^2} \log h(t) = 0 \tag{5.47}$$

which implies that $h(t) = c_1 e^{c_2 t}$ with $c_1 > 0$. Thus, the only smooth densities of the form (5.44) which are MTP_2 correspond to some $N(0,\sigma^2 I_p)$ distribution with $\sigma^2 > 0$. See Sampson (1983) for some related results. □

Section 5: Monotone Regression and Association

In this section, relations between monotone regression for a random vector, introduced by Lehmann (1955, 1966) and the previously described notions of MTP_2 and association are discussed. Basically, the results of this section show that MTP_2 implies monotone regression which in turn implies association. For some additional information concerning these and related ideas, the reader can consult Tong (1980, Chapter 5) and Karlin and Rinott (1980).

The following result in Esary et al. (1967) provides a sufficient condition for random variables $X_1, \ldots, X_n$ to be associated. However, the condition is rather difficult to check in practice.

Proposition 5.18: Suppose the random variables $X_1, \ldots, X_n$ satisfy the following condition:

$$\left.\begin{array}{l} \text{For each } i = 1,2,\ldots, n-1 \text{ and for each bounded} \\ \text{non-decreasing function } f \text{ defined on } R^1, \\ E(f(X_{i+1}) | X_1 = x_1, \ldots, X_i = x_i) \\ \text{is non-decreasing in } (x_1, \ldots, x_i) \in R^i. \end{array}\right\} \tag{5.48}$$

Then $X_1, \ldots, X_n$ are associated.

Proof: Since X_1 is associated, it suffices to verify that if $X_1, \ldots, X_i$ are associated and if (5.48) holds, then $X_1, \ldots, X_{i+1}$ are associated. Thus, let f_1 and f_2 be bounded non-decreasing functions defined on R^{i+1}. It must be shown that

$$\begin{aligned}\delta &= Ef_1(X_1, \ldots, X_{i+1})f_2(X_1, \ldots, X_{i+1}) \\ &\geq Ef_1(X_1, \ldots, X_{i+1})Ef_2(X_1, \ldots, X_{i+1}).\end{aligned}$$

Conditioning on $X_1 = x_1, \ldots, X_i = x_i$ and using the fact all one-dimensional random variables are associated,

$$\left.\begin{aligned}&E(f_1(x_1,\ldots,x_i,X_{i+1})f_2(x_1,\ldots, x_i,X_{i+1})|X_1=x_1,\ldots,X_i=x_i) \\ &\quad\geq E(f_1(x_1,\ldots, x_i,X_{i+1})|X_1=x_1,\ldots,X_i=x_i) \\ &\qquad\cdot E(f_2(x_1,\ldots,x_i,X_{i+1})|X_1=x_1,\ldots,X_i=x_i).\end{aligned}\right\} \tag{5.48a}$$

But, using (5.48) and the assumptions on f_i,

$$H_j(u_1,\ldots,u_i,x_1,\ldots,x_i) = E(f_j(u_1,\ldots,u_i,X_{i+1})|X_1=x_1,\ldots,X_i=x_i)$$

is non-decreasing and bounded on R^{2i}. Thus

$$H_j(x_1,\ldots,x_i,x_1,\ldots,x_i) = E(f_j(x_1,\ldots,x_i,X_{i+1})|X_1=x_1,\ldots,X_i=x_i)$$

is bounded and non-decreasing on R^i, $j = 1,2$. Since $X_1, \ldots, X_i$ are associated, the expectation over $X_1, \ldots, X_i$ of the right hand side of (5.48a) is bounded below by

$$Ef_1(X_1, \ldots, X_{i+1})Ef_2(X_1, \ldots, X_{i+1}).$$

This completes the proof of the basic induction step so the result follows. □

The next assertion shows that MTP_2 implies a condition stronger than (5.48). Consider a random vector $X \in R^n$ with coordinates $X_1, \ldots, X_n$ which has a density f with respect to a product measure $\mu = \mu_1 \times \ldots \times \mu_n$ on R^n.

<u>Proposition 5.19</u> (Sarkar (1969)). If f is MTP_2, then for any bounded function h defined on R^{n-i}, $1 \le i \le n-1$, the conditional expectation

$$E(h(X_{i+1}, \ldots, X_n) | X_1 = x_1, \ldots, X_i = x_i) \qquad (5.49)$$

is non-decreasing in $x_1, \ldots, x_i$.

<u>Proof</u>: First, the marginal density of $X_1, \ldots, X_i$ is

$$f_0(x_1, \ldots, x_i) = \int f(x_1, \ldots, x_n) \mu_{i+1}(dx_{i+1}) \ldots \mu_n(dx_n).$$

Thus, for $x_1 \le y_1, \ldots, x_i \le y_i$, a version of the conditional density of $X_{i+1}, \ldots, X_n$ given $X_1 = x_1, \ldots, X_i = x_i$ is

$$f_1(x_{i+1}, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{f_0(x_1, \ldots, x_i)}$$

and

$$f_2(x_{i+1}, \ldots, x_n) = \frac{f(y_1, \ldots, y_i, x_{i+1}, \ldots, x_n)}{f_0(y_1, \ldots, y_i)}$$

is a version of the conditional density of $X_{i+1}, \ldots, X_n$ given $X_1 = y_1, \ldots, X_i = y_i$. The verification that (5.49) is non-decreasing entails showing that

$$\int h(x_{i+1}, \ldots, x_n) f_1(x_{i+1}, \ldots, x_n) \mu_{i+1}(dx_{i+1}) \ldots \mu_n(dx_n)$$
$$\le \int h(x_{i+1}, \ldots, x_n) f_2(x_{i+1}, \ldots, x_n) \mu_{i+1}(dx_{i+1}) \ldots \mu_n(dx_n).$$

However, that f_1 and f_2 satisfy (5.28) is easily verified using the assumption that f is MTP_2. Theorem 5.9 yields the desired inequality. □

The implications in Propositions 5.18 and 5.19 are both strict. Examples of this can be found in Chapter 5 of Tong (1980) for the case of n = 2.

The argument used in the proof of Proposition 5.19 yields a method for proving certain expectation are non-decreasing functions of parameters. Suppose $X \in \mathcal{X}^{(m)}$ has a density which depends on a parameter $\theta \in \Theta^{(k)}$. Assume the density $f(x|\theta)$ is a density with respect to a product measure $\mu = \mu_1 \times \ldots \times \mu_m$ on $\mathcal{X}^{(m)}$. Here, both $\mathcal{X}^{(m)} \subseteq R^m$ and $\Theta^{(k)} \subseteq R^k$ are assumed to be product sets.

Proposition 5.20: Assume f regarded as a function on $\mathcal{X}^{(m)} \times \Theta^{(k)}$ is MTP_2. If h is non-decreasing on $\mathcal{X}^{(m)}$, then

$$\psi(\theta) = E_\theta h(X) \tag{5.50}$$

is non-decreasing on $\Theta^{(k)}$.

Proof: Consider θ and η in $\Theta^{(k)}$ with $\theta_i \le \eta_i$, $i = 1, \ldots, k$. Let

$$f_1(x) = f(x|\theta)$$

and

$$f_2(x) = f(x|\eta).$$

Because f is MTP_2 on $\mathcal{X}^{(m)} \times \Theta^{(k)}$, it is easy to verify that f_1 and f_2 satisfy (5.28). Theorem 5.9 yields the inequality $\psi(\theta) \le \psi(\eta)$. □

Example 5.5: In this example we sketch a result due to Perlman and Olkin (1980) concerning the unbiasedness of some invariant multivariate tests in the MANOVA problem. It is assumed that the reader is familiar with Example 4.3 as the notation and certain results given there are used here. Recall that the data for the canonical form of the MANOVA problem consists of X: $r \times p$ which is $N(\mu, I_r \otimes \Sigma)$ and S: $p \times p$ which is independent of X and has a $W(\Sigma, p, n)$ distribution. It is assumed that $r \le p$ as the case $r > p$ is similar. The problem is to test the null hypothesis H_0: $\mu = 0$ versus the alternative H_1: $\mu \neq 0$. As in Example 4.3, only invariant non-randomized tests are considered. Such tests are functions of the eigenvalues of $XS^{-1}X'$--say $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$; and the power function of such an invariant test is a function of the eigenvalues of $\mu\Sigma^{-1}\mu'$--say $\xi_1 \geq \ldots \geq \xi_r \geq 0$.

To describe the result of Perlman and Olkin, let $Y \subseteq R^p$ be the set

$$Y = \{y | y \in R^r,\ y_i \geq 0,\ i = 1, \ldots, r\}$$

so Y is a product space of dimension r with each element of the product being $[0,\infty)$. The random vector

$$\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{bmatrix} \in Y$$

whose coordinates are the eigenvalues of $XS^{-1}X'$ has a distribution on Y which depends on the vector

$$\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_r \end{bmatrix} \in Y$$

whose coordinates are the eigenvalues of $\mu\Sigma^{-1}\mu'$. The acceptance region of an invariant test of H_0 versus H_1 is a subset of

$$Y^0 = Y \cap \{y | y \in R^r,\ y_1 \geq \ldots \geq y_r\}.$$

An acceptance region $A \subseteq Y^0$ is monotone if

$$\left.\begin{array}{l} v \in A,\ u \in Y^0,\ u \leq v \\ \text{implies that } u \in A. \end{array}\right\} \qquad (5.51)$$

The corresponding test function

$$g(u) = 1 - I_A(u),\ u \in Y^0 \qquad (5.52)$$

is non-decreasing (in the coordinatewise ordering) on Y^0, when A is monotone.

Proposition 5.21 (Perlman and Olkin (1980)). If A is a monotone acceptance region, then the test determined by A, say g, is unbiased. That is,

$$E_0 g(\lambda) \le E_\xi g(\lambda) \tag{5.53}$$

where E_ξ denotes expectation computed with respect to the distribution of λ when the parameter value is ξ. More generally, if g: $Y^0 \to R^1$ is coordinatewise non-decreasing, then

$$E_0 g(\lambda) \le E_\xi g(\lambda) \tag{5.54}$$

assuming the expectations exist.

Remark: Conditions for strict inequality are given in Perlman and Olkin (1980), but those are not given here. □

Before describing the proof, we first outline the argument. In all that follows, it is assumed that S is $W(I_p, p, n)$ and X is $N(D_\xi, I_r \otimes I_p)$ where

$$D_\xi = \begin{pmatrix} \xi_1^{1/2} & & 0 & \vdots & \\ & \ddots & & \vdots & 0 \\ 0 & & \xi_r^{1/2} & \vdots & \end{pmatrix} : \; r \times p.$$

Since the concern here is with invariant tests, there is no loss of generality with this assumption. Let $\beta_1 \ge \ldots \ge \beta_r$ be the eigenvalues of XX'. The distribution of

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_r \end{pmatrix} \in Y$$

depends on ξ. It is first argued that

$$E_0 g_1(\beta) \le E_\xi g_1(\beta) \tag{5.55}$$

for functions g_1 which are non-decreasing on Y^0. Next, it is shown that the eigenvalues of $XS^{-1}X'$ have the same distribution as the eigenvalues of

$$D_\beta S^{-1} D_\beta = V$$

where

$$D_\beta = \left[\begin{array}{ccc:c} \beta_1^{1/2} & & 0 & \\ & \ddots & & 0 \\ 0 & & \beta_r^{1/2} & \end{array}\right] : \ r \times p$$

But, the vector α of eigenvalues of V is coordinatewise non-decreasing in the vector β. Thus, when $g(\alpha)$ is non-decreasing in α,

$$g_1(\beta) = g(\alpha(\beta))$$

is non-decreasing in β. This is then used to show that (5.55) implies (5.54) which proves Proposition 5.21.

Now, we turn to some technical details. First, let $\tau: Y \to Y^0$ denote the function which maps y into the vector of ordered coordinates of y. If $g_1: Y^0 \to R^1$ is non-decreasing on Y^0, then its extension to Y defined by

$$\bar{g}_1(y) = g_1(\tau(y))$$

is also non-decreasing because τ is coordinatewise non-decreasing. Since the random vector β is in Y^0, (5.55) follows if we can show

$$E\bar{g}_1(\beta) \le E_\xi \bar{g}_1(\beta). \tag{5.56}$$

This is established by appealing to Theorem 5.9 with ψ_1 taken to be the density of β when $\xi = 0$ and ψ_2 the density of β when the parameter is ξ. When $\xi = 0$, the density of β is given by $f_0(\beta)$ of (5.42) on the product space Y (with n replaced by p and p replaced by r). For an arbitrary ξ, the density of β is given by

$$f(\beta|\xi) = f_0(\beta)F(\beta|\xi)\exp[-\frac{1}{2}\sum_1^r \xi_i]$$

where

$$F(\beta|\xi) = \int_{O_r}\int_{O_p} \exp[\mathrm{tr}\Gamma D_\beta \Delta D'_\xi]\nu_1(d\Gamma)\nu_2(d\Delta).$$

Here ν_1 and ν_2 are the unique invariant (Haar) probability measures on the compact groups O_r and O_p. This representation of $f(\beta|\xi)$ is given in James (1961, 1964). For a general discussion of this type of representation, the reader can consult Eaton (1983, Chapter 7, Section 5) or Muirhead (1982, Chapter 3). Now, to apply Theorem 5.9 to establish (5.56), it must be verified that for $u,v \in Y$,

$$\begin{aligned} &f_0(u)f_0(v)F(v|\xi)\exp[-\frac{1}{2}\sum_1^r \xi_i] \le \\ &\quad f_0(u\wedge v)f_0(u\vee v)F(u\vee v|\xi)\exp[-\frac{1}{2}\sum_1^r \xi_i]. \end{aligned} \tag{5.57}$$

Since f_0 is MTP_2 (Example 5.3), (5.57) will follow if we can show that $F(\cdot|\xi)$ is coordinatewise non-decreasing on Y. To see this, define H on R^r by

$$H(x) = \int_{O_r}\int_{O_p} \exp[\mathrm{tr}\Gamma C_x \Delta D'_\xi]\nu_1(d\Gamma)\nu_2(d\Delta),$$

where

$$C_x = \begin{bmatrix} x_1 & & & & \vdots & \\ & x_2 & & 0 & \vdots & \\ & & \ddots & & \vdots & 0 \\ & 0 & & & x_r & \end{bmatrix} : \quad r \times p.$$

Because the exponential is convex, it follows easily that H is a convex function on R^r. Also, if g is any r×r orthogonal matrix with plus or minus ones on the diagonal, the relation $C_{gx} = gC_x$ and the invariance of the measure ν_1 show that $H(x) = H(gx)$. The convexity of H and this invariance imply that H is coordinatewise non-decreasing on Y since H is a convex even function of each argument. But, for $\beta \in Y$,

$$F(\beta|\xi) = H(\sqrt{\beta_1}, \ldots, \sqrt{\beta_r})$$

so F is coordinatewise non-decreasing. Thus Theorem 5.9 applies to yield inequality (5.56).

Now the second part of the argument proceeds as follows. Write X in its singular value decomposition

$$X = \psi_1 D_\beta \psi_2$$

where $\psi_1 \in O_r$, $\psi_2 \in O_p$ and D_β as defined above. The vector $\lambda \in Y^0$ of eigenvalues of

$$XS^{-1}X' = \psi_1 D_\beta \psi_2 S^{-1} \psi_2' D_\beta' \psi_1'$$

is the same as the vector of eigenvalues of

$$D_\beta (\psi_2 S \psi_2')^{-1} D_\beta'. \tag{5.58}$$

Now, X and S are independent and S is $W(I_p, p, n)$. Thus $\psi_2 S \psi_2'$ has the same distribution as S. Since (β, ψ_2) is independent of S, λ has the same distribution as the vector of eigenvalues, say γ, of

$$D_\beta S^{-1} D_\beta'. \tag{5.59}$$

Therefore, for any coordinatewise non-decreasing function g,

$$E_\xi g(\lambda) = E_\xi g(\gamma)$$

for any $\xi \in Y^0$.

Now, fix S, and consider the vector γ as a function of the vector β (via equation (5.59)). The claim is that each coordinate of $\gamma(\beta)$ is coordinatewise non-decreasing in β. To see this, consider β and $\delta \in Y^0$ with $\beta \leq \delta$. Then

$$D'_\delta D_\delta - D'_\beta D_\beta$$

is non-negative definite. But, $\gamma(\beta)$ for $D_\beta S^{-1} D'_\beta$ is the same as the vector of non-zero roots of

$$S^{-1/2} D'_\beta D_\beta S^{-1/2} \tag{5.60}$$

which is no larger than (in the sense of positive definiteness)

$$S^{-1/2} D'_\delta D_\delta S^{-1/2}. \tag{5.61}$$

This implies that the vector of eigenvalues of the matrix (5.61) is coordinatewise no smaller than the vector of eigenvalues of (5.60). Since the non-zero eigenvalues of (5.61) are $\gamma(\delta)$, it follows that $\gamma(\beta)$ is coordinatewise no larger than $\gamma(\delta)$.

To complete the proof, (5.54) is now verified. The important observation is that the vector of eigenvalues λ of $XS^{-1}X'$ has the same distribution as the vector of eigenvalues γ of $D_\beta S^{-1} D_\beta$ and for S fixed, $\gamma(\beta)$ is coordinatewise non-decreasing in the vector β of eigenvalues of XX'. Thus,

$$\begin{aligned} E_\xi g(\lambda) &= E_\xi g(\gamma) = E[E_\xi(g(\gamma(\beta))|S)] \\ &\geq E[E_0(g(\gamma(\beta))|S)] = E_0 g(\gamma). \end{aligned}$$

The inequality above follows from (5.55) applied with S fixed (β and S are independent) and $g_1(\beta) = g(\gamma(\beta))$. The proof is now complete.

An interesting open question is whether or not

$$\xi \to E_\xi g(\lambda), \quad \xi \in Y^0$$

is coordinatewise non-decreasing when g is non-decreasing. The argument of Perlman and Olkin (1980) seems not to be applicable to this question, but Perlman and Olkin have answered the question in the affirmative when ξ has only one non-zero coordinate. □

Section 6: Association and the Normal Distribution

Consider a random vector $X \in R^p$ with a multivariate normal distribution, say $N_p(\mu,\Sigma)$. If X is associated, then necessarily each element of $\Sigma = \{\sigma_{ij}\}$ is non-negative because

$$\sigma_{ij} = EX_iX_j - EX_iEX_j.$$

Recently, Pitt (1982) established the converse to this observation. This section is devoted to a discussion of Pitt's result and some related issues.

To begin the technical discussion, first observe that without loss of generality μ can be taken to be zero in discussions of association. This follows because a function f defined on R^p is non-decreasing in each coordinate iff f_μ defined by $f_\mu(x) = f(x-\mu)$ is non-decreasing in each coordinate. Our first task to show that if X is $N(0,\Sigma)$ and if each $\sigma_{ij} \geq 0$, then X is associated (Pitt (1982)). The details given below are slightly different than those in Pitt (1982), but the idea of the proof is from Pitt.

For a $N(0,\Sigma)$ random vector X in R^p, let Y also be $N(0,\Sigma)$ with Y independent of X. For $0 \leq \lambda \leq 1$, consider the random vector

$$Z_\lambda = \lambda X + \sqrt{1-\lambda^2}\, Y. \qquad (5.62)$$

Given bounded continuous non-decreasing functions f_1 and f_2 on R^p, set

$$H(\lambda) = Ef_1(Z_\lambda)f_2(X). \tag{5.63}$$

Then,

$$H(0) = Ef_1(Y)Ef_2(X) = Ef_1(X)Ef_2(X)$$

and

$$H(1) = Ef_1(X)f_2(X).$$

Thus, the inequality $H(1) \geq H(0)$ is equivalent to

$$Ef_1(X)f_2(X) \geq Ef_1(X)Ef_2(X), \tag{5.64}$$

which is equivalent to the assertion that X is associated.

Remark: That association is equivalent to the validity of (5.64) for all bounded continuous functions is due to Esary et al. (1967). This characterization of association is often very useful and is used here. □

Since f_1 and f_2 are bounded and continuous, H is continuous on [0,1]. The method of proof is to show that when $\sigma_{ij} \geq 0$ for all i,j and when f_1 and f_2 are suitably smooth, then H has a non-negative derivative on (0,1). This implies $H(0) \leq H(1)$ for smooth f_1 and f_2. The verification of (5.64) for bounded continuous functions involves an approximation argument. We now turn to the technical details.

The following lemma provides a crucial step in the main result.

Lemma 5.22: Let X be $N_p(0,\Sigma)$ and let $h: R^p \to R^1$ be bounded with and bounded and continuous partial derivatives. With Y and Z_λ as in (5.62) and for $1 \leq i \leq p$, set

$$H(\lambda) = Eh(Z_\lambda)X_i. \tag{5.65}$$

Then, for $0 < \lambda < 1$,

$$H'(\lambda) = \sum_{j=1}^{p} Eh_j(Z_\lambda)\sigma_{ij} \tag{5.66}$$

where

$$h_j(x) = (\frac{\partial}{\partial x_j}h)(x). \tag{5.67}$$

In particular, if h is non-decreasing and if $\sigma_{ij} \geq 0$ for all i,j, then $H'(\lambda) \geq 0$ and H is non-decreasing on [0,1].

<u>Proof</u>: Because of the assumption on h, the Dominated Convergence Theorem shows that for H given in (5.65)

$$H'(\lambda) = E[\frac{d}{d\lambda}h(Z_\lambda)]X_i.$$

With ∇h denoting the gradient of h, the chain rule yields

$$H'(\lambda) = E(X - \frac{\lambda}{\sqrt{1-\lambda^2}} Y)'[\nabla h(Z_\lambda)]X_i$$

so

$$\sqrt{1-\lambda^2}\; H'(\lambda) = E(\sqrt{1-\lambda^2}\; X - \lambda Y)'[\nabla h(Z_\lambda)]X_i. \tag{5.68}$$

With $S = Z_\lambda = \lambda X + \sqrt{1-\lambda^2}Y$ and $T = \sqrt{1-\lambda^2}X - \lambda Y$, the normality and independence of X and Y imply that S and T are independent $N_p(0,\Sigma)$. Also, $X = \lambda S + \sqrt{1-\lambda^2}T$ so

$$X_i = \lambda S_i + \sqrt{1-\lambda^2}\; T_i. \tag{5.69}$$

Using this in (5.68) yields

$$\sqrt{1-\lambda^2}\, H'(\lambda) = ET'[\nabla h(S)](\lambda S_i + \sqrt{1-\lambda^2}\, T_i).$$

Since $ET = 0$, the independence of S and T imply

$$ET'[\nabla h(S)](\lambda S_i) = 0$$

so

$$\begin{aligned}\sqrt{1-\lambda^2}\, H'(\lambda) &= \sqrt{1-\lambda^2}\, ET'[\nabla h(S)]T_i \\ &= \sqrt{1-\lambda^2} \sum_{j=1}^{p} E[h_j(S)T_jT_i] \\ &= \sqrt{1-\lambda^2} \sum_{j=1}^{p} Eh_j(S)\sigma_{ij}.\end{aligned}$$

Since $S = Z_\lambda$, (5.66) holds. The second assertion follows immediately. □

Proposition 5.23 (Pitt (1982)). Suppose X is $N(0,\Sigma)$ with $\sigma_{ij} \geq 0$ for all i and j. Then X is associated.

Proof: The first step in the proof is to verify (5.64) for f_1 and f_2 which are bounded and have bounded continuous partial derivatives. For such functions, again consider

$$H(\lambda) = Ef_1(Z_\lambda)f_2(X).$$

As argued above, it suffices to show that $H'(\lambda) \geq 0$ for $\lambda \in (0,1)$. Because of our assumptions on f_1 and f_2,

$$\begin{aligned}H'(\lambda) &= E\,\frac{d}{d\lambda} f_1(Z_\lambda)f_2(X) \\ &= E\Big(X - \frac{\lambda}{\sqrt{1-\lambda^2}}\, Y\Big)'[\nabla f_1(Z_\lambda)]f_2(X).\end{aligned}$$

As in the proof of Lemma 5.22, set $S = Z_\lambda$ and $T = \sqrt{1-\lambda^2}X - \lambda Y$ so $X = \lambda S + \sqrt{1-\lambda^2}T$. Then S and T are independent $N(0,\Sigma)$ and

$$\sqrt{1-\lambda^2}\ H'(\lambda) = \mathcal{E}T'[\nabla f_1(S)]f_2(\lambda S + \sqrt{1-\lambda^2}\ T). \tag{5.70}$$

Now fix S and for $1 \le i \le p$, set

$$g_i(S) = \mathcal{E}_T f_2(\lambda S + \sqrt{1-\lambda^2}\ T)T_i$$

where $\mathcal{E}_T$ means expectation over the distribution of T. Since S and T are independent and since $\Sigma = \{\sigma_{ij}\}$ satisfies $\sigma_{ij} \ge 0$ for all i,j, Lemma 5.22 implies that

$$g_i(S) \ge 0 \qquad \text{for } i = 1,\ldots, p. \tag{5.71}$$

However, (5.70) can be written

$$\sqrt{1-\lambda^2}\ H'(\lambda) = \mathcal{E}\sum_{i=1}^{p} g_i(S)f_{1,i}(S) \tag{5.72}$$

where

$$f_{1,i}(x) = (\frac{\partial}{\partial x_i}f_1)(x).$$

Since f_1 is non-decreasing, $f_{1,i}(S) \ge 0$ so (5.71) shows that (5.72) is non-negative for $\lambda \in (0,1)$. Hence (5.64) holds for f_1 and f_2 which are bounded and have bounded continuous partial derivatives.

To complete the proof, consider h_1 and h_2 which are bounded and continuous. For $\epsilon > 0$, let $\psi(\cdot|\epsilon)$ denote the density function of a $N_p(0,\epsilon^2 I_p)$ distribution. It is a routine exercise to show that

$$h_{i,\epsilon}(x) = \int_{R^p} h_i(y)\psi(x-y|\epsilon)dy$$

is bounded with bounded continuous partial derivatives and is non-decreasing in each argument. Further,

$$\lim_{\epsilon \to 0} h_{i,\epsilon}(x) = h_i(x), \qquad i = 1,2$$

for all $x \in R^n$ and

$$\sup_{0<\epsilon\le 1} \sup_x |h_{i,1}(x) - h_i(x)|, \qquad i = 1,2$$

is bounded. The first portion of the proof shows

$$\text{cov}\{h_{1,\epsilon}(X), h_{2,\epsilon}(X)\} > 0$$

and the above facts, together with the Dominated Convergence Theorem, show that

$$\lim_{\epsilon \to 0} \text{cov}\{h_{1,\epsilon}(X), h_{2,\epsilon}(X)\} = \text{cov}\{h_1(X), h_2(X)\} \ge 0.$$

Thus X is associated. □

Let X be a random vector in R^p. In applications, it is sometimes the case that one desires an inequality of the type

$$P\{X_i \le a_i,\ i = 1,\dots,\ p\} \ge \prod_1^p P\{X_i \le a_i\} \tag{5.73}$$

or of the type

$$P\{X_i \ge a_i,\ i = 1,\dots,\ p\} \ge \prod_1^p P\{X_i \ge a_i\}. \tag{5.74}$$

Both of these inequalities are valid when X is associated, but association is strictly stronger than these inequalities. For a simple counter example, see Tong (1980, Chapter 5). When X is $N(0,\Sigma)$ with each $\sigma_{ij} \ge 0$, then X is associated so (5.73) and (5.74) hold. An alternative proof of this is provided by Slepian's inequality (Slepian (1962)) which shows that certain probabilities increase as correlations increase (when X is $N(0,\Sigma)$). Slepian's result was generalized in Das Gupta et al. (1971) using a geometric argument due to Chartres (1963). Here is a formal statement of one version of this result.

Proposition 5.24: Let X be a random vector in R^p which has a density of the form

$$p(x|\Sigma) = |\Sigma|^{-1/2} r(x'\Sigma^{-1}x), \quad x \in R^p$$

where Σ is p×p and positive definite. Given real numbers $a_1, \ldots, a_p$, let

$$\psi(\Sigma) = P\{X_i \le a_i,\ i = 1, \ldots, p\}. \tag{5.75}$$

Then, for any (i,j) with $i \ne j$, $\psi(\Sigma)$ is non-decreasing in σ_{ij} when the other elements of Σ are held fixed.

Proof: See Section 5 in Das Gupta et al. (1971). □

When X is $N(0,\Sigma)$ with each $\sigma_{ij} \ge 0$, Proposition 5.24 shows (5.75) is bounded below by setting each $\sigma_{ij} = 0$ for all $i \ne j$ and inequality (5.73) follows from the independence properties of the normal.

In certain confidence set problems, lower bounds on probabilities of the form

$$P\{|X_i| \le a_i,\ i = 1, \ldots, p\} \tag{5.76}$$

are often desired. When X is $N(0,\Sigma)$, Sidak (1967) showed that

$$P\{|X_i| \le a_i,\ i = 1, \ldots, p\} \ge \prod_{i=1}^{p} P\{|X_i| \le a_i\}, \tag{5.77}$$

no matter what the covariance Σ is. Sidak's argument was the following.

When X is $N_p(0,\Sigma)$, partition Σ as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where Σ_{11} is $(p-1)\times(p-1)$ so $\Sigma_{22} \in (0,\infty)$. For $0 \le \lambda \le 1$, set

$$\Sigma_\lambda = \begin{bmatrix} \Sigma_{11} & \lambda\Sigma_{12} \\ \lambda\Sigma_{21} & \Sigma_{22} \end{bmatrix} . \qquad (5.78)$$

When Σ is positive definite (which we assume here), Σ_λ is positive definite. When X is $N(0,\Sigma_\lambda)$, Sidak (1967) showed that the probability in (5.76) is non-decreasing in $\lambda \in [0,1]$. Applying this result with $\lambda = 0$ and $\lambda = 1$ yields

$$P\{|X_i| \le a_i,\ i = 1,\ldots,\ p\} \ge P\{|X_i| \le a_i,\ i = 1,\ldots,\ p\text{-}1\}P\{|X_p| \le a_p\}$$

when X is $N(0,\Sigma)$. An easy induction argument now yields (5.77).

Sidak's proof used a conditioning argument along with Anderson's (1955) Theorem (Theorem 4.12 here). Das Gupta et al. (1971) extended Sidak's result to the so called elliptical distributions, again using Anderson's Theorem.

<u>Proposition 5.25</u>. Consider a density of the form

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

where Σ is p×p and positive definite. With Σ_λ as defined in (5.78), let X have $p(\cdot|\Sigma_\lambda)$ as its density. Also, let $\dot{X}$ denote the vector of the first p-1 coordinates of X. For a symmetric convex subset $C \subseteq R^{p-1}$, set

$$g(\lambda) = P\{\dot{X} \in C,\ |X_p| \le a_p\}.$$

Then g is non-decreasing on [0,1].

Proof: See Das Gupta et al. (1971), Section 2. An alternative proof which shows that the assumption of a density is unnecessary is given in Eaton (1982b), Section 5. □

An alternative method for attempting to establish (5.77) is to try to show the random variables $Y_i = |X_i|$, $i \geq 1, \ldots, p$ are associated. When X is $N(0,\Sigma)$, Karlin and Rinott (1981) showed that the density of $Y_1, \ldots, Y_p$ is MTP_2 iff there is a p×p diagonal matrix D with plus or minus ones on the diagonal such that $D\Sigma^{-1}D$ has non-positive off diagonal elements. In this case Y is associated (since MTP_2 implies association) so (5.77) holds. However, the Sidak result shows that MTP_2 is very much stronger than (5.77) since (5.77) holds for all covariances when X is normal. But, this discussion raises an interesting question: what are some useful conditions on the joint distribution of $X_1, \ldots, X_p$ which imply that $Y_1, \ldots, Y_p$ are associated (and hence that (5.77) hold)? Some results related to this question can be found in Jogdeo (1977), in Karlin and Rinott (1981), and in Bølviken (1982). An interesting alternative proof of Pitt's result can be found in Joag-dev, Perlman and Pitt (1983).

Chapter 6: Group Induced Orderings

The questions to be addressed in this chapter are all related to the one basic question: To what extent can the majorization results given in Chapters 2 and 3 be generalized to orderings induced on vector spaces by compact groups? The interest in this question arises partly from some interesting examples occurring in statistical and probabilistic problems which are described below. It will be clear from the discussion below that there are, at this point in time, many more questions than answers.

For subgroups of P_n, Rado (1952) considered some aspects of the above questions. As discussed in Chapter 4, Mudholkar (1966) considered group induced orderings on R^n and provided a generalization of Anderson's (1955) Theorem. For groups generated by reflections (see Section 3.3), Eaton and Perlman (1977) showed that analogues of many results valid for majorization continue to hold. Some of these results are discussed below, although the approach taken here is a bit different than in Eaton and Perlman (1977). In particular, the structure theory of reflection groups (see Benson and Grove (1971)) is not used. Other extensions of majorization results to the reflection group case can be found in Conlon et al. (1977).

Because of the geometric treatment of majorization given in Chapter 2, it is hoped that the development here appears to be quite natural. In general, this development follows Eaton (1982, 1984), but some modifications and extensions of that material are given below. For some related material, see Alberti and Uhlmann (1981) and Giovagnoli and Wynn (1985).

Section 1: The Ordering

Let $(V,(\cdot,\cdot))$ be a finite dimensional inner product space and let G be any closed subgroups of $O(V)$--the orthogonal group of the inner product space $(V,(\cdot,\cdot))$. The topology on $O(V)$ is the usual topology of the orthogonal group. The assumption that G is closed is mainly an assumption of convenience at this point and is satisfied in all the interesting examples that I know. Given $x \in V$, $C(x)$ denotes the convex hull of the G-orbit of x--that is, $C(x)$ is the convex hull of $\{gx|g \in G\}$. The dependence of $C(x)$ on G is usually suppressed as G remains fixed throughout most of the discussion. That $C(x)$ is compact follows easily from the assumption that G is closed.

Here is a natural extension of the majorization ordering.

Definition 6.1: For x and y in V, write $x \le y$ to mean that $x \in C(y)$. The relation $\le$ is called the G-induced ordering on $(V,(\cdot,\cdot))$.

Of course, when $V = R^n$, $(\cdot,\cdot)$ is the usual inner product on R^n and $G = P_n$, then the G-induced ordering is just majorization. As with majorization, $\le$ is actually a "pre-ordering" (see Marshall and Olkin (1979), p. 13, for a discussion), but for simplicity, we will just call $\le$ an "ordering." In addition, the dependence of $\le$ on G is suppressed notationally.

The analogues of Proposition 2.1 and 2.2 are:

Proposition 6.1: The following are equivalent:

(i) $x \le y$

(ii) $C(x) \subseteq C(y)$

(iii) $g_1 x \le g_2 y$ for some $g_1, g_2 \in G$.

Proof: The proof is essentially the same as the proof of Proposition 2.1. □

Proposition 6.2: The relation $\le$ is transitive--that is, $x \le y$ and $y \le z$ implies $x \le z$. If $x \le y$ and $y \le x$, then x is in the orbit of y and conversely.

Proof: The same as the proof of Proposition 2.2. □

As in the permutation group case, we observe that $x \in C(y)$ iff

$$(u,x) \le \sup_{g \in G}(u,gy) \qquad \text{for all } u \in V, \tag{6.1}$$

where $(\cdot,\cdot)$ denotes the inner product on V. This is a direct application of Proposition A.3 (with $B = C(y)$ and A equal to the orbit of y). Thus the function

$$m[u,y] = \sup_{g \in G}(y,gy) \tag{6.2}$$

plays the same role in the general case as in the case of $G = P_n$. The proofs of the following two propositions are the same as their counterparts (Propositions 2.3 and 2.4) in Chapter 2.

Proposition 6.3: For all $u,y \in V$, the function m in (6.2) satisfies

(i) $m[u,y] = m[y,u]$

(ii) $m[g_1u,g_2y] = m[u,y]$ for all $g_1,g_2 \in G$

(iii) $m[u,\cdot]$ is convex for each u.

Proposition 6.4: The following are equivalent:

(i) $x \le y$

(ii) $m[u,x] \le m[u,y]$ for all $u \in V$.

At this point, the general development diverges sharply from that in Chapter 2. In particular, there appears to be no natural choice for the convex cone F as in Chapter 2; and even in those cases where there is a natural choice for F, the important conclusion of Proposition 2.6 fails to hold in general. Specific examples to support these claims are given in the next section.

In order to proceed with a development parallel to that in Chapter 2, we now make the following assumptions:

(A.I)

(i) There is a closed convex cone $F \subseteq V$ such that for each x, there is a $g \in G$ with $gx \in F$.

(ii) For each $u,x \in F$, the function m in (6.2) is given by $m[u,x] = (u,x)$.

Because of (A.I), the ordering $\le$ on $V \times V$ is completely determined by the ordering restricted to $F \times F$. In other words, given x and $y \in V$, to decide whether or not $x \le y$, simply move x to $g_1x \in F$ and move y to $g_2y \in F$. By Proposition 6.1, $x \le y$ iff $g_1x \le g_2y$. However, for elements in F, we have

Proposition 6.5: Assume (A.I) holds and consider $x,y \in F$. The following are equivalent:

(i) $x \le y$

(ii) $(u,y-x) \ge 0$ for all $u \in F$.

Proof: Because (A.I) holds, the proof is the same as the proof of Proposition 2.7. □

Before turning to examples, there is one further technical issue with which we must deal. In order to apply Theorem A.6 to identify the decreasing functions, it is necessary that F have a non-empty interior. However, in some interesting examples, F does not have a non-empty interior as a subset of V. Let M = span{F} so M is a linear subspace of V. Then, F always has a nonempty interior as a subset of the vector space $(M,(\cdot,\cdot))$ (see the Rockafellar (1970)). Also, let F_M^* denote the dual cone of F when F is regarded as a convex cone in $(M,(\cdot,\cdot))$. In other words,

$$F_M^* = \{w | w \in M, \ (w,x) \ge 0 \quad \text{for all } x \in F\}.$$

A direct consequence of Proposition 6.5 is

Proposition 6.6: Assume (A.I) holds, and consider $x,y \in F$. The following are equivalent:

(i) $x \le y$

(ii) $y-x \in F_M^*$.

Proof: This is simply a restatement of Proposition 6.6. □

Thus, when (A.I) holds, the ordering restricted to F is a cone ordering defined by the dual cone $F_M^* \subseteq M$. Hence, if T is a frame for F and $x,y \in F$, then $x \le y$ iff $(t,x) \le (t,y)$ for all frame vectors $t \in T$. This is exactly the argument used in Proposition 2.10. Of course, a frame T^* for F_M^* will arise naturally in the discussion of the decreasing functions in Section 3.

Section 2. Examples

This section consists basically of a series of examples intended to illustrate the usefulness and restrictiveness of the assumption (A.I) in the previous section. In the examples where (A.I) holds, the cone F and a frame T for F are computed explicitly (as well as F^* and a frame T^* for F^*) because these are needed to obtain an analytic description of the ordering. In addition F^* and T^* are needed to describe the decreasing functions in these examples. Naturally, the reader should keep the majorization case in mind for comparative purposes since in this case, the objects of interest have been computed explicitly in Chapter 2.

Example 6.1: Let V be R^n and take G to be the group of coordinate sign changes, $\mathcal{D}_n$, on R^n. Thus $\mathcal{D}_n$ consists of all n×n diagonal matrices with each diagonal element equal to 1 or -1. For this example, take

$$F = \{x \mid x \in R^n,\ x_i \geq 0,\ i = 1,\ldots,\ n\}$$

so a frame for F is $T = \{\epsilon_1,\ldots,\ \epsilon_n\}$ where $\epsilon_1,\ldots,\ \epsilon_n$ is the standard orthonormal basis in R^n. Since $F^* = F$, T is also a frame for F^*. That (A.I)(i) holds is clear. To check (A.I)(ii), consider $x,y \in F$ and let $g \in \mathcal{D}_n$ have diagonal elements $d_1,\ldots,\ d_n$. Then,

$$\begin{aligned}(x,gy) &= \Sigma_1^n d_i x_i y_i \leq |\Sigma_1^n d_i x_i y_i| \\ &\leq \Sigma_1^n |d_i| x_i y_i = \Sigma_1^n x_i y_i = (x,y)\end{aligned}$$

so (A.I)(ii) holds. Thus (A.I) holds and and easy application of Proposition 6.6 shows that $x \leq y$ iff

$$|x_i| \leq |y_i|, \quad i = 1,\ldots,\ n. \ \square$$

Example 6.2: Again take V to be R^n and let G be the group generated by $\mathcal{P}_n$ and $\mathcal{D}_n$. This group is denoted by $\mathcal{P}_n \cdot \mathcal{D}_n$ as every element in G can be written in the form PD with $P \in \mathcal{P}_n$ and $D \in \mathcal{D}_n$. To see this, note that for each $P \in \mathcal{P}_n$ and $D \in \mathcal{D}_n$, $P^{-1}DP \in \mathcal{D}_n$ so that the set of elements of the form PD is a group. In this example, take

$$F = \{x | x_1 \geq x_2 \geq \ldots \geq x_n \geq 0\}.$$

A frame for F is given in Example A.3 and is $T = \{e_1, \ldots, e_n\}$ where e_i is the vector in R^n whose first i coordinates are one and the remaining coordinates are zero. Also in examples A.3, a frame $T^* = \{t_1, \ldots, t_n\}$ for F^* is given where t_i has its ith coordinate equal to one, its (i+1)th coordinate equal to minus one, and all remaining coordinates zero; i = 1,..., n-1. The vector t_n has its nth coordinate equal to one and all remaining coordinates are equal to zero. That assumption (A.I)(i) holds is clear. To verify (A.I)(ii), consider $x,y \in F$ and $g = PD \in G$. Then

$$(x,gy) = x'PDy = (P'x)'Dy \leq (P'x)'y = x'Py \leq x'y.$$

The first inequality follows from Example 6.1 and the second inequality follows from Proposition 2.6. Thus, (A.I)(ii) holds.

To express what the order means in this case, define a function τ on R^n to F as follows: $\tau(x)$ is the vector whose coordinates are denoted by $|x|_{(1)} \geq \ldots \geq |x|_n \geq 0$ obtained by ordering the numbers $|x_1|, \ldots, |x_n|$ where x has coordinates $x_1, \ldots, x_n$. Obviously, τ is the identity on F and given any $x \in R^n$, there is a $g \in \mathcal{P}_n \cdot \mathcal{D}_n$ such that $gx = \tau(x)$. Hence $x \leq y$ iff $\tau(x) \leq \tau(y)$. Since $\tau(x)$ and $\tau(y)$ are in F, we know $\tau(x) \leq \tau(y)$ iff

$$e_k'\tau(x) \leq e_k'\tau(y), \quad k = 1, \ldots, n$$

iff

$$\Sigma_1^k |x|_{(i)} \leq \Sigma_1^k |y|_{(i)}, \quad k = 1, \ldots, n.$$

Thus, it is easy to check whether $x \leq y$.

The ordering in this example should <u>not</u> be confused with the submajorization ordering discussed in Marshall and Olkin (1979, p. 10). □

<u>Example 6.3</u>: This example is somewhat more complicated that the first two. The vector space is $\mathcal{L}_{p,n}$--the space of a real n×p matrices. For

convenience, it is assumed that $n \geq p$. The inner product on $L_{p,n}$ is taken to be

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

where tr denotes trace and y' is the transpose of $y \in L_{p,n}$. The group in this example is $O_n \times O_p$ whose elements are written as pairs (g,h) with $g \in O_n$ and $h \in O_p$. As usual, $O_n(O_p)$ is the usual group of $n\times n$ $(p\times p)$ orthogonal matrices. The pair (g,h) defines an orthogonal transformation on $L_{p,n}$ by

$$(g,h)(x) = gxh'$$

where the right hand side means matrix multiplication. That this is an orthogonal transformation on $(L_{p,n},(\cdot,\cdot))$ is a consequence of

$$\begin{aligned}((g,h)x,(g,h)y) &= (gxh',gyh') = \text{tr } gxh'(gyh')' \\ &= \text{tr } gxh'hy'g' = \text{tr } gxy'g' = \text{tr } g'gxy' = \text{tr } xy' = (x,y).\end{aligned}$$

Now, the choice of F for this example is motivated by the Singular Value Decomposition Theorem which asserts that for any $x \in L_{p,n}$, there exists $(g,h) \in O_n \times O_p$ such that

$$gxh' = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \\ \cdots & \cdots & \cdots \\ & 0 & \end{pmatrix} \tag{6.3}$$

where $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$. The numbers $\lambda_1 \geq \ldots \geq \lambda_p$ are called the <u>singular values</u> of x and $\lambda_1^2, \ldots, \lambda_p^2$ are easily shown to be the eigenvalues of $x'x$. Define F to be all the matrices in $L_{p,n}$ which have the form (6.3). Thus, an $x \in F$ with elements x_{ij}, $i = 1,\ldots,n$, $j = 1,\ldots, n$ satisfies

$$\begin{cases} x_{ij} = 0 & \text{for } i \neq j \\ x_{11} \geq \ldots \geq x_{pp} \geq 0 & \end{cases}$$

and conversely, any such x with these properties is in F. The Singular Value Decomposition Theorem shows that (A.I)(i) holds with this choice for F.

The verification of (A.I)(ii) is much more delicate and depends on the following result (von Neumann (1937), Fan (1951)).

<u>Theorem 6.7</u>: Let x and y be elements in $\mathcal{L}_{p,n}$ with singular values $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ and $\mu_1 \geq \ldots \geq \mu_p \geq 0$ respectively. Then

$$\sup_{g\in\mathcal{O}_n, h\in\mathcal{O}_p} \operatorname{tr} gxh'y' = \Sigma_1^p \lambda_i \mu_i . \tag{6.4}$$

<u>Proof</u>: First, we treat the case n = p so x and y are square matrices. Using the Singular Value Decomposition Theorem, write

$$g_1 x h_1' = D_\lambda, \quad g_2 y h_2' = D_\mu \tag{6.5}$$

where D_λ and D_μ are n×n diagonal matrices with diagonal elements $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$. Thus, when n = p, the left hand side of (6.4) is

$$\sup_{g\in\mathcal{O}_n, h\in\mathcal{O}_n} \operatorname{tr} g g_1' D_\lambda h_1 h' h_2' D_\mu g_2 = \sup_{g\in\mathcal{O}_n, h\in\mathcal{O}_n} \operatorname{tr} g_2 g g_1' D_\lambda h_1 h' h_2' D\mu$$

$$= \sup_{g\in\mathcal{O}_n, h\in\mathcal{O}_n} \operatorname{tr} g D_\lambda h' D_\mu .$$

Thus, it must be shown that for $g \in \mathcal{O}_n$ and $h \in \mathcal{O}_n$,

$$\operatorname{tr} g D_\lambda h' D_\mu \leq \Sigma_1^n \lambda_i \mu_i \tag{6.6}$$

since equality in (6.6) is achieved by setting $g = h' = I_n$. Now, for $u,v \in \mathcal{L}_{n,n}$, consider the function

$$\psi(u,v) = \operatorname{tr} u D_\lambda v' D_\mu . \tag{6.7}$$

Obviously, $\psi(\cdot,v)$ is a linear function for each v and $\psi(u,\cdot)$ is a linear function for each fixed u--that is, ψ is a bilinear function. Routine calculations show that $\psi(u,v) = \psi(v,u)$ (so ψ is symmetric) and $\psi(u,u) \geq 0$ (so ψ is non-negative definite). Hence, the Cauchy-Schwarz inequality applied to ψ yields

$$\psi(u,v) \leq ((\psi(u,u))^{1/2}(\psi(v,v))^{1/2}. \tag{6.8}$$

Applying (6.8) to the left side of (6.6) gives the inequality

$$\text{tr } gD_\lambda hD_\mu \leq (\text{tr } gD_\lambda g'D_\mu)^{1/2}(\text{tr } hD_\lambda h'D_\mu)^{1/2}. \tag{6.9}$$

Thus, to establish (6.6), it suffices to establish the inequality

$$\text{tr } gD_\lambda g'D_\mu \leq \Sigma_1^n \lambda_i \mu_i \tag{6.10}$$

for $g \in O_n$. Using the fact that D_λ and D_μ are diagonal, writing out the left side of (6.10) yields

$$\text{tr } gD_\lambda g'D_\mu = \sum_i \sum_j \lambda_i s_{ij} \mu_j \tag{6.11}$$

where $s_{ij} = g_{ij}^2$ and g_{ij} is the (i,j) element of the matrix $g \in O_n$. Since $g \in O_n$,

$$\Sigma_i s_{ij} = \Sigma_j s_{ij} = 1$$

so the n×n matrix S with elements s_{ij} is doubly stochastic. Thus, the right side of (6.11) can be written

$$\lambda' S\mu \tag{6.12}$$

where $\lambda(\mu)$ has coordinates $\lambda_1, \ldots, \lambda_n$ $(\mu_1, \ldots, \mu_n)$. Since S is doubly stochastic, Birkhoff's Theorem (see the Appendix) implies that S is a convex combination of permutation matrices--say

$$S = \sum_k \alpha_k k$$

where the sum runs over P_n, $k \in P_n$, and the real numbers α_k satisfy $0 \le \alpha_k$ and $\Sigma\alpha_k = 1$. Thus (6.12) is equal to

$$\sum_k \alpha_k \lambda' k\mu. \tag{6.13}$$

However, by Proposition 2.6, $\lambda' k\mu \le \lambda'\mu$ for each $k \in P_n$. Since the α_k sum to 1, (6.13) is bounded above by

$$\lambda'\mu = \Sigma_1^n \lambda_i \mu_i. \tag{6.14}$$

Thus (6.11) is bounded above by $\Sigma_1^n \lambda_i \mu_i$ so (6.10) holds. Hence, the proof is complete for the case $p = n$.

Now, consider the case $p < n$. Given x and $y \in L_{p,n}$, construct $\tilde{x}$ and $\tilde{y}$ in $L_{n,n}$ as follows:

$$\tilde{x} = (x0); \quad \tilde{y} = (y0)$$

where "0" denotes a block of $n\times(n-p)$ zeroes. If $\lambda_1 \ge \ldots \ge \lambda_p \ge 0$ and $\mu_1 \ge \ldots \ge \mu_p \ge 0$ are the singular values of x and y respectively, then the singular values of x and y are just $\lambda_1, \ldots, \lambda_p, 0, \ldots, 0$ and $\mu_1, \ldots, \mu_p, 0, \ldots 0$ where there are $(n-p)$ zeroes following λ_p and μ_p. This follows by noting that the n eigenvalues of

$$\tilde{x}'\tilde{x} = \begin{pmatrix} x'x & 0 \\ 0 & 0 \end{pmatrix}$$

are just $\lambda_1^2, \ldots, \lambda_p^2, 0, \ldots, 0$ with $n-p$ zeroes following λ_p^2; and a similar statement concerning $\tilde{y}$. From the result for $p = n$ we have

$$\sup_{g\in O_n, h\in O_n} \operatorname{tr} g\tilde{x}h'\tilde{y}' = \Sigma_1^p \lambda_i \mu_i \tag{6.15}$$

since the last $n-p$ singular values of $\tilde{x}$ and $\tilde{y}$ are zero. Now, write

$$h = \begin{pmatrix} h_{11} & 0 \\ 0 & I_{n-p} \end{pmatrix}$$

where h_{11} is an arbitrary element of O_p. Since, for such an h,

$$\tilde{x}h'\tilde{y}' = xh_{11}'y',$$

it follows from (6.14) that

$$\sup_{g \in O_n, h_{11} \in O_p} \operatorname{tr} gxh_{11}'y' \leq \Sigma_1^p \lambda_i \mu_i. \tag{6.16}$$

However, the Singular Value Decomposition Theorem implies that there is equality in (6.16) with the appropriate choice of $g \in O_n$ and $h_{11} \in O_p$. □

To continue with our discussion of the example at hand, we now proceed to verify (A.I)(ii). The function m is

$$m[u,y] = \sup_{g \in O_n, h \in O_p} \operatorname{tr} guh'y' = \Sigma_1^p \lambda_i \mu_i \tag{6.17}$$

where $\lambda_1, \ldots, \lambda_p$ ($\mu_1, \ldots, \mu_p$) are the singular values of u (y, respectively). However, when u and y are in F, then u has the form

$$u = \begin{pmatrix} u_{11} & & 0 \\ & \ddots & \\ 0 & & u_{pp} \\ \cdots & \cdots & \cdots \\ & 0 & \end{pmatrix}$$

where $u_{11} \geq \ldots \geq u_{pp} \geq 0$. For such a u, it is clear that the ordered singular values of u are just $u_{11}, \ldots, u_{pp}$. Similar statements apply to y $\in$ F. This, together with (6.17), shows that when u and y are in F, then

$$m[u,y] = \operatorname{tr} uy' = (u,y)$$

so (A.I)(ii) holds.

For this example, observe that span(F) is just the subspace M of $L_{p,n}$ consisting of all n×p matrices of the form

$$x = \begin{pmatrix} x_{11} & & 0 \\ & \ddots & \\ 0 & & x_{pp} \\ \cdots & \cdots & \cdots \\ & 0 & \end{pmatrix}$$

where $x_{11}, \ldots, x_{pp}$ are arbitrary real numbers. Obviously, M has dimension p. A frame for F is easily constructed from Example 6.2 since the frame here is the "same" as the frame in Example 6.2. Let $T = \{e_1, \ldots, e_n\}$ where e_i in $L_{p,n}$ has its first i-diagonal elements equal to 1 and all other elements equal to zero. Obviously, T is a frame for F. The dual cone of F <u>in M</u>, say F_M^* has a frame $T^* = \{t_1, \ldots, t_p\}$ where, for $i = 1, \ldots, p-1$, t_i in $L_{p,n}$ has its (i,i) element equal to one, its (i+1,i+1) element equal to minus one and all other elements zero. The matrix t_p has its (p,p) element equal to one and all other elements are zero. Just as in Example 6.2, T^* is easily shown to be a frame for F_M^*. From the point of view of partial orderings, this example has appeared in Alberti and Uhlmann (1981) and Eaton (1982, 1984).

In order to interpret the ordering $x \le y$, let $\lambda_1 \ge \ldots \ge \lambda_p$ and $\mu_1 \ge \ldots \ge \mu_p$ be the singular values of x and y, respectively. Then representatives of x and y in F are

$$x_\lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \\ \cdots & \cdots & \cdots \\ & 0 & \end{pmatrix} \quad ; \quad y_\mu = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_p \\ \cdots & \cdots & \cdots \\ & 0 & \end{pmatrix} .$$

Since $x \le y$ iff $x_\lambda \le y_\mu$, it follows immediately from the structure of the frame for F that $x \le y$ iff

$$\Sigma_1^k \lambda_i \le \Sigma_1^k \mu_i ; \quad k = 1, \ldots, p.$$

Therefore, the ordering on $L_{p,n}$ can be described by saying that $x \leq y$ iff the singular values of x are "less than" the singular values of y in the sense of Example 6.2. □

Example 6.4: For this example, take V to the vector space S_p of all p×p real symmetric matrices, and use the inner product

$$(x,y) = \operatorname{tr} xy' \tag{6.18}$$

which equals tr xy since $y = y'$ in S_p. The group is O_p and each $g \in O_p$ defines an orthogonal transformation on $(S_p,(\cdot,\cdot))$ given by

$$x \to gxg'; \qquad x \in S_p,\ g \in O_p.$$

The Spectral Theorem for symmetric matrices asserts that for each $x \in S_p$, there is a $g \in O_p$ such that

$$gxg' = D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} \tag{6.19}$$

where $\lambda_1 \geq \ldots \geq \lambda_p$ are the eigenvalues of x. This suggests a frame for this example. Let F be all diagonal matrices $x \in S_p$ of the form (6.19) with diagonal elements $x_{11},\ldots, x_{pp}$ which satisfy $x_{11} \geq \ldots \geq x_{pp}$. Obviously F is a convex cone and (A.I)(i) holds for this choice of F. As in the last example, the verification of (A.I)(ii) requires a bit more work.

Proposition 6.8: For x and y in F,

$$\sup_{g \in O_n} \operatorname{tr} gxg'y' = \operatorname{tr} xy = (x,y) \tag{6.20}$$

and (A.I)(ii) holds.

Proof: With $g = I_p$ (the p×p identity matrix), $\operatorname{tr} gxg'y' = (x,y)$ so the left hand side of (6.20) is at least as large as the right hand side. Hence, to verify (6.20), it must be shown that for each $g \in \mathcal{O}_n$ and $x,y \in F$

$$\operatorname{tr} gxg'y \le \operatorname{tr} xy. \tag{6.21}$$

Consider x and y in F. Then for a real number c, $x+cI_p$ and $y+cI_p$ are both in F (and conversely). Replacing x and y by $x+cI_p$ and $y+cI_p$ in (6.21) yields

$$\begin{aligned}&\operatorname{tr} gxg'y + c\operatorname{tr} y + c\operatorname{tr} x + pc^2\\ &\quad\le \operatorname{tr} xy + c\operatorname{tr} y + c\operatorname{tr} x + pc^2\end{aligned} \tag{6.22}$$

which is obviously equivalent to (6.21) for all real c. Now, pick c large enough so that $\tilde{x} = x+cI$ and $\tilde{y} = y+cI$ have positive diagonal elements. That (6.21) holds for $\tilde{x}$ and $\tilde{y}$ follows immediately from Theorem 6.7 since the diagonal elements of $\tilde{x}$ (and $\tilde{y}$) are the singular values of $\tilde{x}$ (and $\tilde{y}$). Thus (6.21) holds for x and y and the proof is complete. □

Now, we proceed with a description of a frame for F and F^*. Since F is essentially the same convex cone as was used in Chapter 2 for majorization, the material there carries over to this case with essentially no change. Let $t_1,\ldots, t_{n-1}$ be elements of $\mathcal{S}_p$ where t_i has its (i,i) diagonal element equal to one, its (i+1,i+1) diagonal equal to minus one and the remaining elements of t_i are zero, $i = 1,\ldots, n-1$. Let M be the linear subspace of diagonal matrices in $\mathcal{S}_p$ so M = span(F). As in Chapter 2, $T^* = \{t_1,\ldots, t_{n-1}\}$ is a frame for F_M^*--the dual cone of F in M. Let $e_1,\ldots, e_n$ in M where e_i has its first i diagonal elements equal to one and all other elements equal to zero. Then $T = \{e_1,\ldots,e_n,-e_n\}$ is a frame for F.

To describe the ordering on $\mathcal{S}_p$, consider x and y in $\mathcal{S}_p$ with $x \le y$. Let D_λ (and D_μ) denote the diagonal matrices whose ith diagonal element is the ith largest eigenvalue of x (and y), $i = 1,\ldots, p$. Then $x \le y$ iff $D_\lambda \le D_\mu$. But since D_λ and D_μ are in F, $D_\lambda \le D_\mu$ iff

$$\begin{aligned}&\operatorname{tr} e_iD_\lambda \le \operatorname{tr} e_iD_\mu, \quad i = 1,\ldots, n-1\\ &\operatorname{tr} e_nD_\lambda = \operatorname{tr} e_nD_\mu.\end{aligned} \tag{6.23}$$

But (6.23) holds iff

$$\Sigma_1^k \lambda_i \le \Sigma_1^k \mu_i, \quad k = 1,\ldots, p$$
$$\Sigma_1^p \lambda_i = \Sigma_1^p \mu_i. \tag{6.24}$$

However, (6.24) is exactly the condition that the vector of eigenvalues of y majorize the vector of eigenvalues of x. Thus, the ordering on S_p induced by O_p has a direct relation to the majorization ordering. This relation was given in Karlin and Rinott (1981) using an argument which is rather different than the one given here. The argument above is from Eaton (1984). An alternative ordering on S_p is discussed in Alberti and Uhlmann (1981). □

Now, we turn to two examples where (A.I)(ii) does not hold.

Example 6.5: On the plane R^2, let G be the group consisting of four elements $\{I, g, g^2, g^3\}$ where g is rotation by 90° in the counter clockwise direction. Thus, the matrix of g (in the standard coordinate system for R^2) is

$$g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

An obvious choice for F in this example is

$$F = \{x \mid x_1 \ge 0,\ x_2 \ge 0\}$$

but any other 90°-wedge would do as well. Clearly (A.I)(i) holds. However (A.I)(ii) is false. For example, take

$$u = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

which are both in F. Since gu = v,

$$\sup_{h \in F} (hu, v) = (gu, v) = (v, v) = 1$$

while $(u,v) = 0$. Geometrically, $m[u,v]$ can be described as follows. Fix $v \in F$ and let u^* be a vector in the set $\{u, gu, g^2u, g^3u\}$ whose angle with v is no more than 45°. Then, it is easy to see $m[u,v] = (u^*,v)$. Of course, this example can easily be generalized to the group generated by the rotation through $2\pi/k$ where k is an integer which is at least 3. □

<u>Example 6.6</u>: For this example, again consider R^n and let G be the group

$$G = \{g | g = \pm k,\ k \in P_n\}.$$

Thus, an element of G is either a permutation matrix or minus a permutation matrix. For this example, take F to be

$$F = \{x | x \in R^n,\ x_1 \geq \ldots \geq x_n,\ \Sigma_1^n x_i \geq 0\}.$$

(A.I)(i) is easily shown to hold, but (A.I)(ii) fails. In fact, $m[u,v]$ can be calculated explicitly for this example. For $u,v \in F$, first observe that

$$\max_{k \in P_n} (ku)'v = u'v$$

which follows from Proposition 2.6. Since $v \in F$, $v_1 \geq \ldots \geq v_n$ so $-v_n \geq \ldots \geq -v_1$. Let v^* be

$$v^* = \begin{pmatrix} -v_n \\ -v_{n-1} \\ \vdots \\ -v_1 \end{pmatrix} \in F.$$

Then

$$\max_{k \in P_n} (-ku)'v = \max_{k \in P_n} (ku)'(-v) = \max_{k \in P_n} (ku)'v^* = u'v^*$$

where again the last equality is a consequence of Proposition 2.6. Hence, for $u,v \in F$

$$\max_{g\in G} (gu)'v = \max\{u'v, u'v^*\}.$$

When $n \geq 3$, it is not hard to show there are $u, v \in F$ such that $u'v < u'v^*$ so (A.I)(ii) can not hold for this example. □

Section 3: The Decreasing Functions

We now turn to the problem of describing the decreasing functions in the present context when (A.I) holds. Thus, on the inner product space $(V, (\cdot,\cdot))$, G is a closed subgroup of $O(V)$ and G induces and ordering as in Definition 6.1. A real valued function f defined on V is decreasing if $x \leq y$ implies $f(x) \geq f(y)$. Such a decreasing function must satisfy

$$f(x) = f(gx), \qquad x \in V,\ g \in G$$

because $x \leq gx \leq x$ for $x \in V$ and $g \in G$. That is, decreasing functions must be G-invariant.

Now, we assume (A.I) holds. Thus, there is a closed convex cone F so (A.I)(i) and (A.I)(ii) hold. With $M = \text{span}(F)$ and F_M^* denoting the dual cone of F in M, our previous results show that for $x, y \in F$,

$$x \leq y \quad \text{iff } y-x \in F_M^*.$$

Hence the ordering restricted to F is a cone ordering induced by the cone F_M^*. Since F has a non-empty interior in M, Theorem A.6 provides necessary and sufficient conditions for the characterization of the decreasing function. Here is the formal statement.

Proposition 6.9: Let f be a G-invariant function real valued function defined on V and let f_1 be the restriction of f to F. Let T^* be a frame for F_M^*, and assume f_1 is continuous at the boundary of F. The following are equivalent:

(i) f is decreasing

(ii) for each $t \in T^*$ and $x \in F$, the function $h(\lambda) = f_1(x+\lambda t)$ is decreasing in $\lambda \geq 0$ as long as $x+\lambda t$ is in F.

Proof: Since f is G-invariant, f is decreasing on V iff f_1 is decreasing on F. The equivalence of (i) and (ii) now follow directly from Theorem A.6. □

Of course, when f in Proposition 6.9 has a differential, then (ii) of Proposition 6.9 is easily expressed as a condition in terms of the differential. The writing out of these conditions in Examples 6.1 through 6.4 is a straightforward and tedious task which is left to the reader.

When assumption (A.I) does not hold, then the ordering ≤ restricted to F is, in general, not a cone ordering so Theorem A.6 is not available. In particular, characterizations of the decreasing functions are not known for Examples 6.5 and 6.6. However, a necessary condition for a function to be decreasing is given in Eaton (1975) (see Eaton and Perlman (1977), Proposition 2.2), but whether this condition is sufficient is not known.

Section 4: The Convolution Theorem

In this and the next section, we investigate conditions under which the Convolution Theorem (Theorem 2.20) and the material on DR kernels discussed in Chapter 3 can be extended to the general case. Basically, the results here show that the Convolution Theorem and the theory of DR kernels can be extended to the so called reflection groups. However, the development here is a bit different from the material in Eaton and Perlman (1977) and Conlon et al. (1977) because of the geometric approach taken in Chapters 2,3 and in this chapter,

To begin the technical development, again let $(V,(\cdot,\cdot))$ be a finite dimension inner product space. In this situation, a reflection is defined as follows. For $u \in V$, let $u \otimes u$ denote the linear transformation on V to V whose value at x is

$$(u \otimes u)(x) = (u,x)u.$$

Thus, for $u \neq 0$, $u \otimes u$ is a rank one self-adjoint linear transformation whose range is span{u} and whose null space is the orthogonal complement of span{u}. For $u \neq 0$, let

$$R_u = I - 2\frac{u \otimes u}{(u,u)} . \tag{6.25}$$

Now, it is easy to show $R_u u = -u$, $R_u v = 0$ if $v \perp u$, and $R_u^2 = I$. For these reasons, R_u is called a reflection--more precisely, R_u is the reflection across the hyperplane $\{v|v \perp u\}$.

As in the previous sections, the compact group $G \subseteq O(V)$ induces the ordering $\leq$ between elements of V. Throughout this section, assumption (A.I) is to hold. Thus the ordering $\leq$ when restricted to F is a cone ordering induced by the dual cone F_M^* where $M = \text{span}(F)$. In all of the examples to which the results below apply, $M = V$ but this need not be assumed in what follows. Let T^* be a frame for F_M^*. The key assumption in this section (aside from (A.I)) is

(A.II) For each $t \in T^*$, the reflection R_t is an element of G.

The three basic examples where (A.I) and (A.II) hold are:

(i) $V = R^n$, $G = P_n$

(ii) $V = R^n$, $G = D_n$

(iii) $V = R^n$, $G = P_n \cdot D_n$.

The verification of this claim is easy because F_M^* has been given for these three cases--the first in Chapter 2 and the other two in Examples 6.1 and 6.2.

Here are the implications of (A.II) for the characterization of the decreasing functions.

Proposition 6.10: Assume (A.I) and (A.II) hold and that f is a decreasing function. Then for each $t \in T^*$ and for each $v \perp t$, the function $h(\beta) = f(v+\beta t)$ is a symmetric unimodal function.

Proof: To verify the symmetry, we must show $h(\beta) = h(-\beta)$. But since $R_t \in G$ and $v \perp t$, the G-invariance of f implies

$$f(v+\beta t) = f(R_t(v+\beta t)) = f(v-\beta t)$$

so symmetry holds. To show h is decreasing on $[0,\infty)$, consider $0 \le \beta_1 \le \beta_2$. Thus, $v+\beta_1 t$ is in the line segment connecting $v+\beta_2 t$ and $v-\beta_2 t = R_t(v+\beta_2 t)$. Thus, because R_t is an element of G, $v+\beta_1 t \le v+\beta_2 t$ so

$$h(\beta_1) = f(v+\beta_1 t) \ge f(v+\beta_2 t) = h(\beta_2). \quad \square$$

<u>Proposition 6.11</u>: Assume (A.I) and (A.II) hold and let f be a G-invariant real valued function defined on V. Let f_1 be the restriction of f to F and assume f_1 is continuous at the boundary of F. If for each $t \in T^*$ and each $v \perp t$, $h(\beta) = f(v+\beta t)$ is decreasing on $[0,\infty)$, then f is decreasing.

<u>Proof</u>: Condition (ii) of Proposition 6.9 needs to be verified. Thus, consider $x \in F$ and $t \in T^*$. Write $x = v+\delta t$ where $v \perp t$ and $\delta = (t,x)/(t,t)$. Because $x \in F$ and $t \in F_M^*$, $\delta \ge 0$. Thus, for $\lambda \ge 0$, $x+\lambda t = v + (\lambda+\delta)t$ so

$$f(x+\lambda t) = f(v + (\lambda+\delta)t). \tag{6.26}$$

But, the assumption that $f(v+\beta t)$ is decreasing on $[0,\infty)$ implies the function of λ in (6.26) is decreasing on $[0,\infty)$ since $\delta \ge 0$. By Proposition 6.9, f is decreasing. $\square$

<u>Remark 6.1</u>: When $G = \mathcal{P}_n$, we have seen that Proposition 6.11 is true without the assumption that f_1 is continuous at the boundary of F. In the case that the group G is a reflection group (that is, G is equal to the smallest closed group generated by some set of reflections of the form (6.25)), Proposition 6.11 is also true without the assumption that f_1 is continuous at the boundary of F. This is proved in Eaton and Perlman (1977) where extensive use is made of the theory of reflection groups (see Benson and Grove (1971)). Whether the continuity assumption can be dispensed with in the present generality is not known. The groups $\mathcal{D}_n$ and $\mathcal{P}_n \cdot \mathcal{D}_n$ are reflection groups so the results in Eaton and Perlman (1977) apply to these. $\square$

In the present context, here is the Convolution Theorem.

Theorem 6.12: Assume (A.I) and (A.II) hold and let f_1 and f_2 be non-negative decreasing functions. Then the convolution

$$f(y) = \int_V f_1(y-x)f_2(x)dx \tag{6.27}$$

is decreasing.

Proof: Let h_n be the indicator function of $\{x|(x,x) \le n\}$, $n = 1,2,\ldots$ so h_n is a decreasing function and has compact support. Define $f_{i,n}$ by

$$f_{i,n}(x) = \begin{cases} n & \text{if } f_i(x)h_n(x) > n \\ f_i(x)h_n(x) & \text{if } f_1(x)h_n(x) \le n \end{cases}$$

for $i = 1,2$ and $n = 1,2,\ldots$. It is easy to check that $f_{i,n}$ is a decreasing function and $f_{i,n}$ converges monotonically to f_i. Thus, by the Monotone Convergence Theorem,

$$f_n(y) = \int_V f_{1,n}(y-x)f_{2,n}(x)dx \tag{6.28}$$

converges pointwise to f(y). Since the pointwise limit of decreasing functions is again a decreasing function, if suffices to show that f_n is a decreasing function.

Now, f_n defined in (6.28) is the convolution of two bounded functions with compact support. Hence f_n is continuous on V (see Kec (1982) for a proof of this well known result). Thus, it suffices to verify that f_n is G-invariant and for each $t \in T^*$ and $v \perp t$,

$$h(\lambda) = f_n(v+\lambda t) \tag{6.29}$$

is decreasing on $[0,\infty)$. The G-invariance of f_n follows from the G-invariance of $f_{1,n}$, $f_{2,n}$ and the fact that $G \subseteq O(V)$ so each $g \in G$ preserves the Lebesgue measure dx on V. The relevant calculation is

$$\begin{aligned} f_n(gy) &= \int f_{1,n}(gy-x)f_{2,n}(x)dx = \int f_{1,n}(y-g^{-1}x)f_{2,n}(x)dx \\ &= \int f_{1,n}(y-x)f_{2,n}(gx)dx = \int f_{1,n}(y-x)f_{2,n}(x)dx = f_n(y). \end{aligned}$$

The verification that $h(\lambda)$ in (6.29) is decreasing on $[0,\infty)$ proceeds as in the proof of Theorem 2.20. For $t \in T$, write V as the direct sum N + span$\{t_0\}$ where N is the orthogonal complement of span$\{t_0\}$, and $t_0 = t/(t,t)^{1/2}$. In (6.28), write $x = w+\gamma t_0$ where $w \in N$. Then

$$\begin{aligned} f_n(v+\lambda t) &= \int_N\int_{-\infty}^{\infty} f_{1,n}(v+\lambda t-w-\gamma t_0)f_{2,n}(w+\gamma t_0)d\gamma dw \\ &= \int_N\int_{-\infty}^{\infty} f_{1,n}(v-w+(\|t\|\lambda-\gamma)t_0)f_{2,n}(w+\gamma t_0)d\gamma dw. \end{aligned} \tag{6.30}$$

Since $f_{1,n}$ and $f_{2,n}$ are decreasing, Proposition 6.10 implies that

$$\gamma \to f_{2,n}(w+\gamma t_0)$$

is symmetric unimodal and

$$\beta \to f_{1,n}(v-w+\beta t_0)$$

is symmetric unimodal. Thus, the inside integral in (6.30) is the convolution of two symmetric unimodal functions of R^1 evaluated at $\|t\|\lambda$. By Wintner's Theorem, the inside integral is decreasing in $\|t\|\lambda$ for $\|t\|\lambda \geq 0$. Hence the inside integral is decreasing in $\lambda \geq 0$ since $\|t\| > 0$. Thus $f_n(v+\lambda t)$ is decreasing in $\lambda \geq 0$. □

Remark 6.2: The validity of the convolution Theorem when (A.I) holds but (A.II) does not (as in Examples 6.3 and 6.4) is an important open question. The Convolution Theorem is false for Example 6.5 (see Eaton (1984) for a counter example). This shows that some type of assumptions are needed.

However, the group introduced in example 6.6 is important in applications, and it would be useful to know whether or not the Convolution Theorem is valid for this case. □

Section 5: Reflections and DR Functions

In this section we establish general analogues of Proposition 3.8 and Theorem 3.9 for so called DR functions. The emphasis here is on reflections rather than the groups generated by reflections, but a brief discussion of reflection groups is appropriate.

In a finite dimensional vector space $(V,(\cdot,\cdot))$, let Δ be an arbitrary subset of V such that $0 \notin \Delta$. Using the notation of the previous section, each $t \in \Delta$ defines a reflection

$$R_t = I - 2\frac{t \otimes t}{(t,t)} . \tag{6.31}$$

Obviously $R_t = R_t' = R_t^{-1}$ for each reflection. The set of reflections generated by Δ is

$$\mathcal{R} = \{R_t | t \in \Delta\}.$$

Remark 6.3: A group $G \subseteq \mathcal{O}(V)$ is called a reflection group if there is some set $\mathcal{R}$ of reflections which generate (algebraically) G. Examples of reflection groups include $\mathcal{D}_n$ of Example 6.1, the permutation group $\mathcal{P}_n$, and the group $\mathcal{P}_n \cdot \mathcal{D}_n$ of Example 6.2. The reader can easily construct sets of reflections which generate these three groups. The structure of reflection groups is completely understood. All of the finite reflection groups are listed in Benson and Grove (1971). The infinite case is taken care of by the discussion in Eaton and Perlman (1977) and the result in Eaton and Perlman (1977) which asserts that every infinite irreducible reflection group is dense in $\mathcal{O}(V)$. These results are not used in what follows. □

Given the set Δ and the set of reflections $\mathcal{R}$ generated by Δ, here is what appears to be the appropriate definition of DR functions defined on $V \times V$.

Definition 6.2: If a real valued function f defined on V×V satisfies

(i) $f(R_t x, R_t y) = f(x,y)$ for $x,y \in V$ and $R_t \in \mathcal{R}$

(ii) for each $t \in \Delta$, $(t,x)(t,y) \geq 0$ implies that

$$f(x,y) \geq f(x, R_t y) \tag{6.32}$$

then, f is a decreasing reflection function (DR function).

The analogue of Proposition 3.8 in the present context follows.

Proposition 6.13: Let f_0 be a real valued function defined on V and define f on V×V by $f(x,y) = f_0(x-y)$. The following are equivalent:

(i) f is a DR function

(ii) for each $t \in \Delta$ and each $v \perp t$, the function $\beta \to f_0(v+\beta t)$ is symmetric and unimodal for $\beta \in R^1$.

Proof: Assume f is a DR function. With $x = v$ and $y = -\beta t$, (i) of Definition 6.2 implies

$$f_0(v+\beta t) = f(v,-\beta t) = f(R_t v, R_t(-\beta t)) = f(v,\beta t) = f_0(v-\beta t)$$

so symmetry in β holds. To establish unimodality, consider $0 \leq \beta_1 \leq \beta_2$, $t \in \Delta$ and $v \perp t$. With $y = \frac{1}{2}(\beta_2-\beta_1)t$ and $x = v+\frac{1}{2}(\beta_1+\beta_2)t$, $(t,x)(t,y) \geq 0$ so

$$f_0(v+\beta_1 t) = f(x,y) \geq f(x,R_t y) = f_0(v+\beta_2 t).$$

Hence $\beta \to f(v+\beta t)$ is decreasing on $[0,\infty)$ so (ii) holds.

Now, assume (ii) holds. Given $x,y \in V$ and $t \in \Delta$, write $x = v_1+\beta_1 t$ and $y = v_2+\beta_2 t$ with $v_i \perp t$ for $i = 1,2$. Then

$$\begin{aligned} f(x,y) &= f_0(x-y) = f_0(v_1-v_2+(\beta_1-\beta_2)t) = f_0(v_1-v_2-(\beta_1-\beta_2)t) \\ &= f_0(v_1-\beta_1 t-(v_2-\beta_2 t)) = f_0(R_t x-R_t y) = f(R_t x, R_t y). \end{aligned}$$

Thus (i) of Definition 6.2 holds. For $t \in \Delta$ and $x = v_1+\beta_1 t$, $y = v_2+\beta_2 t$, the condition $(t,x)(t,y) \geq 0$ is equivalent to the condition $\beta_1\beta_2 \geq 0$. This implies that $|\beta_1-\beta_2| \le |\beta_1+\beta_2|$ so that the vector $x-y = v_1-v_2+(\beta_1-\beta_2)t$ is a convex combination of the two vectors $v_1-v_2-(\beta_1+\beta_2)t$ and $v_1-v_2+(\beta_1+\beta_2)t$. Therefore, the symmetric unimodality of f_0 implies

$$\begin{aligned} f(x,y) &= f_0(x-y) = f_0(v_1-v_2+(\beta_1-\beta_2)t) \\ &\ge f_0(v_1-v_2+(\beta_1+\beta_2)t) = f_0(x-R_t y) = f(x,R_t y). \end{aligned}$$

Hence f is a DR function. □

Here is the version of Theorem 3.9 appropriate for the present context.

Proposition 6.14: Let f_1 and f_2 be non-negative DR functions defined on V×V. Suppose μ is a σ-finite measure on the Borel subsets of V such that μ is invariant under each of the transformations $R_t, t \in \Delta$. Assume that

$$f(x,z) = \int_V f_1(x,y) f_2(y,z) \mu(dy) \tag{6.33}$$

is finite for each $x,z \in V$. Then f is also a DR function.

Proof: The proof is essentially the same as that of Theorem 3.9 so just a sketch is given. That $f(R_t x, R_t z) = f(x,z)$ is easily established using the invariance of μ and the assumption that f_1 and f_2 are DR functions. To verify (ii) of Definition 6.2, consider $t \in \Delta$ and $x \in V$ satisfying $(t,x)(t,z) \ge 0$. It must be shown that

$$\delta = f(x,z) - f(x,R_t z) = \int_V f_1(x,y)[f_2(y,z) - f_2(y,R_t z)]\mu(dy) \tag{6.34}$$

is non-negative. Write the region V as $V = V_+ \cup V_0 \cup V_-$ where

$$\begin{aligned} V_+ &= \{y | (t,y) > 0\} \\ V_0 &= \{y | (t,y) = 0\} \\ V_- &= \{y | (t,y) < 0\}. \end{aligned}$$

The integral in (6.34) over the region V_0 is zero since $f_2(y,z) = f_2(y,R_t z)$

on V_0. The integral in (6.34) over V_- can be transformed into an integral over V_+ by making the change of variable $y \to R_t y$. Using the invariance assumptions on μ, f_1 and f_2 yields

$$\delta = \int_{V_+} [f_1(x,y) - f_1(x,R_t y)][f_2(y,z) - f_2(y,R_t z)]\mu(dy). \qquad (6.35)$$

Now, the condition $(t,x)(t,z) \geq 0$ together with the DR assumptions on f_1 and f_2 imply that the integrand in (6.35) is non-negative on the region of integration V_+. Hence $\delta \geq 0$. □

Remark 6.4: Versions of definition 6.2 and Propositions 6.13 and 6.14 are available when the domain of definition of the function involved is not V, but some subset of V which is invariant under each $R_t, t \in \Delta$. For example, suppose X and Y are invariant (under each R_t) subsets of V and let f be a real valued function defined on $X \times Y$. Then Definition 6.2 is the appropriate definition of a DR function for $x \in X$ and $y \in Y$. Similar modifications (but the same proofs) are easily made in Propositions 6.13 and 6.14. These are left to the reader. □

A version of the Convolution Theorem can also be formulated in the present context.

Proposition 6.15: Suppose f_1 and f_2 are non-negative functions defined on V which satisfy

(i) for $t \in \Delta$ and $v \perp t$, the function $\beta \to f_i(v+\beta t)$ is symmetric and unimodal, $i = 1,2$.

Then, the convolution

$$f(y) = \int_V f_1(y-x) f_2(x) dx \qquad (6.36)$$

also satisfies (i) above. Here, dx is Lebesgue measure on V.

Remark 6.5: The notion of decreasing functions (when a group $G \subseteq \mathcal{O}(V)$

induces an ordering) has been replaced by (i) in the present setting. We know (i) and "decreasing" are equivalent in the case that $V = R^n$ and $G = P_n$--see Proposition 2.17.

<u>Proof</u>: For $t \in \Delta$ and $v \perp t$,

$$f(v+\beta t) = \int_V f_1(v+\beta t-x)f_2(x)dx.$$

Without loss of generality, assume that $(t,t) = 1$ and let M be the subspace perpendicular to t. Write $x = w+\gamma t$ where $w \in M$ and $\gamma \in R^1$. This "orthogonal change of variable" yields

$$f(v+\beta t) = \int_M\int_{-\infty}^{\infty} f_1(v-w+(\beta-\gamma)t)f_2(w+\gamma t)d\gamma dw.$$

For v and w fixed, we recognize the inside integral as the convolution (on R^1) of two symmetric unimodal functions. Thus, by Wintner's Theorem, for each w,

$$\beta \to \int_{-\infty}^{\infty} f(v-w+(\beta-\gamma)t)f_2(w+\gamma t)d\gamma$$

is a symmetric unimodal function of β. Hence the same is true for $f(v+\beta t)$. □

We end this section with a few comments concerning the validity of Proposition 3.7 in the present reflection context. Again let $\Delta \subseteq V$ be a subset of V such that $0 \notin \Delta$. Let G be the group generated (algebraically) by $R = \{R_t | t \in \Delta\}$ so G is a reflection group. Suppose that f is a DR function according to Definition 6.2. Observe that

$$f(x,y) = f(gx,gy) \tag{6.37}$$

because each g is just a product of reflections in R and (6.37) holds for elements of R by definition. For fixed x and y, the problem is to describe where the function

$$\psi(g) = f(gx,y), \qquad g \in G$$

achieves its supremum (assuming it does achieve it supremum). In Proposition 3.7 when $G = P_n$, a solution to this problem is given in terms of the convex cone F. Thus, given G, one way to attack this problem is to try to construct a convex cone F_G so that Proposition 3.7 is true with P_n replaced by G and F replaced by F_G. When G is a finite reflection group, the existence of such F_G is a non-trivial fact--a proof of which the reader can find in Benson and Grove. A further discussion of topics related to the validity of Proposition 3.7 for reflection groups can be found in Eaton and Perlman (1977) and Conlon et al. (1977).

Appendix

In this appendix, some basic results in convex set theory are reviewed. In addition a few non-standard results, which have direct applications to the main body of the lectures, are covered. It is assumed that the reader is familiar with much of the material in Rockafellar (1970) which is used as a reference for proofs of standard results.

Because of the material in Chapter 6, the setting for the discussion here is in a finite dimensional inner product space, say $(V,(\cdot,\cdot))$ where $(\cdot,\cdot)$ denotes an inner product on the real vector space V assumed finite dimensional.

Definition A.1: A subset $B \subseteq V$ is convex if for all $x,y \in B$ and $\alpha \in [0,1]$, the vector $\alpha x + (1-\alpha)y \in B$.

Since $\{\alpha x+(1-\alpha)y;\alpha \in [0,1]\}$ is just the closed line segment connecting x and y, convex sets are those which contain all the closed line segments connecting points in the set. For a vector $\xi \in V$ with $\|\xi\| = 1$ and a real number c, the hyperplane

$$H_{\xi,c} = \{x|(\xi,x) = c\} \tag{A.1}$$

is convex. Also, the two closed half-spaces

$$\left.\begin{aligned} H^{+}_{\xi,c} &= \{x|(\xi,x) \geq c\} \\ H^{-}_{\xi,c} &= \{x|(\xi,x) \leq c\} \end{aligned}\right\} \tag{A.2}$$

are convex whose interiors do not intersect each other.

For any finite collection of vectors $x_1,\ldots, x_k$ in V, a sum of the form

$$\alpha_1 x_1 +\ldots+ \alpha_k x_k$$

with each α_i non-negative and $\Sigma\alpha_i = 1$, is called a convex combination of $x_1,\ldots, x_k$. It is easy to show that a set B is convex iff B contains all convex combinations of its elements. Given a subset $A \subseteq V$, the convex hull of A, denoted by S(A), is the set of all convex combinations of elements of

A. Obviously S(A) is a convex set.

Proposition A.1: For $A \subseteq V$, S(A) is equal to the intersection of all the convex sets which contain A.

Proof: Let B be the aforementioned intersection so B is convex. Since S(A) is convex and contains A, $B \subseteq S(A)$. On the other hand, $A \subseteq B$ and B is convex so B contains all convex combinations of elements of A--that is, $S(A) \subseteq B$. □

Proposition A.2: For $A \subseteq V$, let B be the intersection of all closed convex sets which contain A. Then B is the closure of S(A).

Proof: Since B is convex and contains A, $S(A) \subseteq B$. But B is closed so B contains the closure of S(A). On the other hand, the closure of S(A) is convex and contains A so $B \subseteq S(A)$. □

In general it is difficult to decide whether a given point is in a convex set. The following criterion is sometimes useful. Let A be a non-empty subset of V and let B be the closure of S(A).

Proposition A.3: The following are equivalent

(i) $x \in B$

(ii) $(u,x) \leq \sup_{z \in A} (u,z)$ for all $u \in V$.

Proof: If $x \in B$, then x is the limit of points in S(A). But if $y \in S(A)$, then $y = \Sigma\alpha_i z_i$ with $0 \leq \alpha_i \leq 1$, $\Sigma\alpha_i = 1$ and $z_i \in A$. Hence

$$(u,y) = \Sigma\alpha_i(u,z_i) \leq \max_i (u,z_i) \leq \sup_{z \in A} (u,z).$$

Since (ii) holds for all $y \in S(A)$, (ii) holds for points in the closure of S(A) by continuity of the inner product. Thus (i) implies (ii).

For (ii) implies (i), we will show that not (i) implies not (ii). Thus, consider $x \notin B$. Since both B and $\{x\}$ are convex and closed, and $\{x\}$ is bounded, there is a hyperplane which strictly separates B and $\{x\}$

(Rockafellar (1970), Corollary 11.4.2). That is, there is a $u \in V$ with $\|u\| = 1$ and a real number c such that $u'x > c$ and

$$u'y < c \quad \text{for all } y \in B.$$

Since $A \subseteq B$, this implies that (ii) cannot hold, so (ii) implies (i). □

A fundamental representation theorem for compact convex sets asserts that every compact convex set is equal to the convex hull of its set of extreme points (Rockafellar (1970), Corollary (18.5.1). An example of this representation theorem of consequence here is Birkhoff's Theorem (Birkhoff (1946)). To describe this result, let V be the n^2-dimensional vector space of $n \times n$ real matrices. An element $Q \in V$ is doubly stochastic if each element of Q is non-negative and, both row sums and column sums are equal to one. Let $B_n \subseteq V$ be the set of all doubly stochastic matrices. It is easy to see that B_n is compact. Thus, a knowledge of the extreme points of B_n would give representations of elements of B_n.

Recall that an $n \times n$ matrix P is a permutation matrix if in each row and each column of P, exactly one element is equal to one and the remaining elements are zero. If x is an n-dimensional coordinate vector and P is a permutation matrix, then Px is just some permutation of the coordinates of x. Hence P is an orthogonal linear transformation so $P^{-1} = P'$ which is also a permutation matrix. Let $\mathcal{P}_n$ denote the set of $n \times n$ permutation matrices. An easy combinatorial argument shows $\mathcal{P}_n$ has n! elements. Noting that $\mathcal{P}_n$ is closed under matrix multiplication, it follows immediately that $\mathcal{P}_n$ is a group--commonly called the group of permutation matrices. Obviously, $\mathcal{P}_n$ is contained in the set B_n of doubly stochastic matrices.

Theorem (Birkhoff (1946)). The group $\mathcal{P}_n$ is exactly the set of extreme points of B_n. Hence every doubly stochastic matrix is a convex combination of permutation matrices.

A proof of the above theorem is not given here. For a more thorough discussion of this result and references to a number of proofs, see Marshall and Olkin (1979) (p. 34).

Convex cones are the next topic of discussion. A subset $F \subseteq V$ is a <u>convex cone</u> if F is convex and if for each $\lambda \geq 0$ and $x \in F$, $\lambda x \in F$. This definition differs slightly from that in Rockafellar (1970) (p. 13), but is more suitable for our purposes.

Given a convex cone F, the <u>dual cone</u> to F is defined by

$$F^* = \{x | (x,u) \geq 0 \quad \text{for all } u \in F\}.$$

It is easily verified that F^* is a closed convex cone. When F is a non-empty closed convex cone, then $(F^*)^* = F$ (Rockafellar (1970), Theorem 14.1). In certain cases, F^* can be determined explicitly when F is not too complicated. Some examples are given below.

If F is a non-empty convex cone, a subset $T \subseteq F$ is a <u>positive spanning set</u> for F if every element of F is a finite linear combination of elements of T with the coefficients of the linear combination non-negative. Obviously F is a positive spanning set for F, but we are interested in minimal spanning sets which are called <u>frames</u>. In other words, T is a <u>frame</u> for F if T is a positive spanning set for F and no proper subset of T is.

In many cases, convex cones are defined as the intersection of a finite number of closed homogeneous half-spaces. That is, a finite number of non-zero vectors $u_1, \ldots, u_r$ are given and a closed convex cone F is defined by

$$F = \{x | (u_i, x) \geq 0, \quad i = 1, \ldots, r\}.$$

Let F_0 be the closed convex cone consisting of all vectors of the form $\Sigma_1^r a_i u_i$ with $a_i \geq 0$, $i = 1, \ldots, r$.

<u>Proposition A.4</u>: In the above notation, $F_0^* = F$ and consequently $F^* = F_0$. Further $\{u_1, \ldots, u_r\}$ is a positive spanning set for F^*. If $u_1, \ldots, u_r$ are linearly independent, then $\{u_1, \ldots, u_r\}$ is a frame for F^*.

<u>Proof</u>: To show $F_0^* = F$, first consider $x \in F$ so $(x, u_i) \geq 0$ $i = 1, \ldots, r$. Hence $(x,y) \geq 0$ for all $y = \Sigma_1^r a_i u_i$ with $a_i \geq 0$. Thus $x \in F_0^*$. But if $x \in F_0^*$, then $(x, u_i) \geq 0$ $i = 1, \ldots r$, so $x \in F$ and $F_0^* = F$. Thus, $F_0 = (F_0^*)^*$

$= F^*$. The second assertion follows by definition of $F_0(= F^*)$. The third assertion follows by noting that each $y \in F^*$ has a representation $y = \Sigma_1^r a_i u_i$ and this representation is unique by the linear independence. Thus, no subset of $\{u_1, \ldots, u_r\}$ can positively span F^* since each $u_i \in F^*$, $i = 1, \ldots, r$. □

Proposition A.5: Let $v_1, \ldots, v_n$ be a basis for the vector space $(V, (\cdot,\cdot))$ and let F be the convex cone generated by $v_1, \ldots, v_n$--that is, $y \in F$ iff $y = \Sigma a_i v_i$ with $a_i \geq 0$, $i = 1, \ldots, n$. Also, let $u_1, \ldots, u_n$ be the dual basis to $v_1, \ldots, v_n$--that is, $(u_i, v_j) = \delta_{ij}$. Then $\{u_1, \ldots, u_n\}$ is a frame for F^*.

Proof: Since $u_1, \ldots, u_n$ is a basis, each $x \in F^*$ can be written $x = \Sigma b_i u_i$. But, $0 \leq (x, v_j) = \Sigma b_i (u_i, v_j) = b_j$ so each $b_j \geq 0$. Conversely, any $x = \Sigma b_i u_i$ with each $b_j \geq 0$ is obviously in F^*. Thus $\{u_1, \ldots, u_n\}$ is a positive spanning set and hence a frame by linear independence. □

Here are three standard examples of closed convex cones, dual cones and frames.

Example A.1: With $V = R^n$, let $\epsilon_1, \ldots, \epsilon_n$ be the standard orthonormal basis for R^n. Set

$$F = \{x | \epsilon_i' x \geq 0, \quad i = 1, \ldots, n\}$$

so F consists of those vectors whose coordinates are all non-negative. A direct application of Propositions A.4 and A.5 shows that $F = F^*$ and $\{\epsilon_1, \ldots, \epsilon_n\}$ is a frame for both F^* and F.

Example A.2: Again take $V = R^n$ and define vectors $t_1, \ldots, t_{n-1}$ by: t_i has its ith coordinate equal to one, its (i+1)th coordinate equal the minus one, and all other coordinates are zero. Consider the closed convex cone given by

$$F = \{x | t_i' x \geq 0, \quad i = 1, \ldots, n-1\}.$$

Then, F consists of those vectors x whose coordinates $\alpha_1, \ldots, \alpha_n$ satisfy $\alpha_1 \geq \ldots \geq \alpha_n$. Since $t_1, \ldots, t_{n-1}$ are linearly independent, it follows from Proposition A.4 that $T^* = \{t_1, \ldots, t_{n-1}\}$ is a frame for F^*. To construct a frame for F, consider vectors $e_1, \ldots, e_n$ defined by: e_i has its first i coordinates equal to one and the remaining coordinates equal to zero. The claim is that $T = \{e_1, \ldots e_n, -e_n\}$ is a frame for F. That T positively spans F is easily checked. To see that T is a frame for F, first observe that $e_1, \ldots, e_n$ is a basis for R^n so no e_i, $1 \leq i \leq n-1$ can be deleted from T. But since e_n and $-e_n$ are both in F, neither of these can be deleted from T if T is to positively span F. □

<u>Example A.3</u>: For this example, let $t_1, \ldots, t_{n-1}$ be as in Example A.2 and let t_n be the vector whose nth coordinate is one and the rest of the coordinates are zero. Then $T^* = \{t_1, \ldots, t_n\}$ is a basis for R^n and by Proposition A.5 is a frame for the dual cone to

$$F = \{x \mid t_i'x \geq 0, \quad i = 1, \ldots, n\}.$$

Obviously, F consists of those vectors x with coordinates $\alpha_1, \ldots, \alpha_n$ which satisfy $\alpha_1 \geq \ldots \geq \alpha_n \geq 0$. To construct a frame for F, observe that $e_1, \ldots, e_n$ in Example 2.A is the dual basis to $t_1, \ldots, t_n$. Since $T^* = \{t_1, \ldots, t_n\}$ is a frame for F^*, Proposition A.5 shows that $T = \{e_1, \ldots, e_n\}$ is a frame for $(F^*)^* = F$. □

The remainder of this appendix is devoted to discussion of a result due to Marshall, Walkup and Wets (1967). Throughout the following discussion, $(V, (\cdot,\cdot))$ is a finite dimensional inner product space, B is a non-empty convex subset of V, and $G_0 \subseteq V$ is a non-empty convex cone. Using G_0, a relation on B is defined as follows:

$$x \leq y \quad \text{iff } y - x \in G_0. \tag{A.3}$$

Because G_0 is a convex cone, if $x \leq y$ and $y \leq z$, then $(y-x) + (z-y) = z-x \in G_0$ so $x \leq z$. That is, $\leq$ is a transitive. Henceforth, the relation $\leq$ is called a partial cone ordering. However, it is possible that $x \le y$ and $y \le x$, but $x \ne y$ (this is possible if G_0 contains a non-trivial subspace).

Given the partial cone ordering $\le$ on B, a real valued function $f:B \to R^1$ is decreasing if $x \le y$ implies $f(x) \ge f(y)$. The problem with which the Marshall, Walkup and Wets result deals is the characterization of decreasing functions in terms of the geometry of G_0. To motivate things, first observe that if $x \in B$ and $t \in G_0$, then the convexity of B implies that the set

$$\Lambda = \{\lambda | \lambda \ge 0,\ x+\lambda t \in B\} \tag{A.4}$$

is a subinterval of $[0,\infty)$ which contains 0. Also, if $0 \le \lambda_1 \le \lambda_2$ and λ_1, λ_2 are both in Λ, then $x+\lambda_1 t \le x+\lambda_2 t$ since $x+\lambda_2 t - (x+\lambda_1 t) = (\lambda_2-\lambda_1)t$ is an element of G_0. Hence, if f is decreasing it is necessary that

$$h(\lambda) = f(x+\lambda t), \quad \lambda \in \Lambda \tag{A.5}$$

be a decreasing function of λ. In particular, if $T \subseteq G_0$ is any positive spanning set, then for each $t \in T$, $h(\lambda)$ given in (A.5) must be decreasing. More particularly, when T is a frame for G_0, the above must hold. Under some regularity conditions, the converse of this observation holds--namely, if for each $x \in B$ and $t \in T$, the function of λ in (A.5) is decreasing for $\lambda \in \Lambda$, then f is decreasing. The utility of this result is that the frame T may have very few vectors in it, so checking that $h(\lambda)$ is decreasing can actually be carried out. In our statement of the result, T is not assumed to be a frame, but only a positive spanning set. In practice, one always tries to take T to be a frame since applying the result is easier for minimal spanning sets.

Here is the formal statement of the result.

Theorem A.6 (Marshall, Walkup and Wets (1967)). Assume the convex set $B \subseteq V$ has a non-empty interior and assume that $f:B \to R^1$ is continuous at $(\partial B)\cap B$ where ∂B is the boundary of B. Let T be a positive spanning set for the convex cone G_0. The following are equivalent

(i) f is decreasing

(ii) for each $t \in T$ and $x \in B$, the function $h(\lambda) = f(x+\lambda t)$ is decreasing in $\lambda \geq 0$ as long as $x+\lambda t \in B$.

Proof: That (i) implies (ii) is clear from the argument preceding the statement of the theorem. To show (ii) implies (i), consider $x \leq y$, $x,y \in B$. It must be verified that

$$f(x) \geq f(y). \tag{A.6}$$

First consider the case when x and y are in the interior of B (which is non-empty by assumption). Since $x \leq y$, the vector $y - x$ is in G_0 and hence is some finite linear combination of elements of T, say

$$y - x = \Sigma_1^r a_i t_i = u \tag{A.7}$$

where $a_i > 0$ and $t_i \in T$, $i = 1,\ldots, r$. Because x and y are in the interior of B, the convexity of B implies that the line segment

$$L = \{v | v = \alpha y + (1-\alpha)x, \quad 0 \leq \alpha \leq 1\}$$

is contained in the interior of B. Thus for some $\epsilon > 0$, the tubular neighborhood

$$N = \{w | \text{for some } v \in L, \ \|w-v\| < \epsilon\}$$

is contained in the interior of B. Now, select an integer k so large that

$$\frac{\|\Sigma_1^s a_i t_i\|}{k} < \epsilon \qquad \text{for } s = 1,\ldots, r. \tag{A.8}$$

Observe that the sequence of points

$$x_m = x + \frac{mu}{k}, \qquad m = 0,1,\ldots, k$$

are all on the line L and $x_m \leq x_{m+1}$ since $u \in G_0$. Thus, to verify (A.6), it suffices to verify

$$f(x_m) \geq f(x_{m+1}). \tag{A.9}$$

But, the sequence of points $z_0 = x_m$,

$$z_s = x_m + \frac{\Sigma_1^s a_i t_i}{k}, \qquad s = 1, \ldots, r$$

are all in N by (A.8). Further, $z_j \leq z_{j+1}$, $j = 0, \ldots, r-1$ since $a_j t_j / k$ is in G_0. Thus, it suffices to show that

$$f(z_j) \geq f(z_{j+1}). \tag{A.10}$$

However, the vector

$$z(\gamma) = z_j + \gamma \frac{a_{j+1} t_{j+1}}{k}, \qquad 0 \leq \gamma \leq 1$$

satisfies

$$\begin{cases} z(0) = z_j \\ z(1) = z_{j+1} \end{cases}$$

and $z(\gamma) \in N$ since N is convex. Thus, $z(\lambda)$ is in B for all λ in [0,1] and by assumption (ii) with $t = t_{j+1}$ and $\lambda = (\gamma a_{j+1})/k$, (A.10) follows and hence (A.6) holds for x and y in the interior of B with $x \leq y$.

Now suppose both x and y are in $\partial B \cap B$ with $x \leq y$. Select z in the interior of B and for $0 \leq \alpha \leq 1$, let

$$x_\alpha = (1-\alpha)x + \alpha z$$
$$y_\alpha = (1-\alpha)y + \alpha z.$$

Then, for $0 < \alpha \leq 1$, x_α and y_α are in the interior of B and $x_\alpha \leq y_\alpha$. By the argument above $f(x_\alpha) \geq f(y_\alpha)$ for all $\alpha \in (0,1]$. Letting $\alpha \to 0$ and using the continuity of f at $\partial B \cap B$, we see that $f(x) \geq f(y)$. When exactly one of x or y is in $\partial B \cap B$, a similar argument shows that $f(x) \geq f(y)$. □

Corollary A.7: Let f be as in Theorem A.6 and assume further that f has a differential, say df, for each x in the interior of B. Then f is decreasing iff $(df(x),t) \leq 0$ for each $t \in T$.

Proof: Since

$$\frac{d}{d\lambda} f(x+\lambda t)\Big|_{\lambda=0} = (df(x),t),$$

the result follows immediately. □

In most of our applications of Theorem A.6, the convex set B will be a closed convex cone F with a non-empty interior and G_0 will be the dual cone F^* of F. The three examples discussed earlier are important in applications and are discussed in detail in the relevant chapters.

References

R. Ahlswede and D.E. Daykin (1979). An inequality for weights of two families of sets, their unions and intersections. Z. Wahrsch. verw. Gebiete 93, 183-185.

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

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

T.W. Anderson (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

C.T. Benson and L.C. Grove (1971). Finite Reflection Groups. Bogden and Quigley, Tarrytown on Hudson, New York.

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

G. Birkhoff (1946). Tres observaciones sobre el algebra lineal. Univ. Nac. Tucuman Rev. Ser. A. 5, 147-151.

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

E. Bølviken (1982). Probability inequalities for the multivariate normal with non-negative partial correlations. Scan. J. Statist. 9, 49-58.

H.J. Brascamp and E.H. Lieb (1974). A logrithmic concavity theorem with some applications. Unpublished manuscript.

B.A. Chartres (1963). A geometrical proof of a theorem due to Slepian. SIAM Review, 5, 335-341.

J.C. Conlon, R. Leon, F. Proschan, and J. Sethuraman (1977). G-ordered functions, with applications in statistics, I, Theory. Report M432, Department of Statistics, Florida State University, Tallahassee, Florida.

S. Das Gupta, T.W. Anderson and G. Mudholkar (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. Ann. Math. Statist. 35, 200-205.

S. Das Gupta, M.L. Eaton, I. Olkin, M. Perlman, L.J. Savage and M. Sobel (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.

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

Ju. S. Davidovic, B.I. Korenbljum, and B.I. Hacet (1969). A property of logarithmically concave functions. Soviet Math. Doklady **10**, No. 2, 477-480 (English Translation).

S. Dharmadhikari and K. Joag-dev (1987). Unimodality, Convexity and Applications. To be published by Academic Press, New York.

S. Dharmadhikari and K. Jogdeo (1976). Multivariate unimodality. Ann. Statist. **4**, 607-613.

R.L. Dykstra and J.E. Hewett (1978). Positive dependence of the roots of a Wishart matrix. Ann. Statist. **6**, 235-238.

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

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

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

M.L. Eaton (1975). Orderings induced on R^n by compact groups with applications to probability inequalities--Preliminary Report. University of Minnesota Technical Report #251.

M.L. Eaton (1982). On group induced orderings, monotone functions and convolution theorems. University of Minnesota Technical Report, School of Statistics, University of Minnesota.

M.L. Eaton (1982b). A review of selection topics in multivariate probability inequalities. Ann. Statist. **10**, 11-43.

M.L. Eaton (1983). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

M.L. Eaton (1984). On group induced orderings, monotone functions, and convolution theorems. In Inequalities in Statistics and Probability, edited by Y.L. Tong. Institute of Mathematical Statistics Lecture Notes--Monograph Series.

M.L. Eaton and R.A. Olshen (1972). Random quotients and the Behrens-Fisher problem. Ann. Math. Statist. **43**, 1852-1860.

M.L. Eaton and M. Perlman (1974). A monotonicity property of the power function of some invariant tests for MANOVA. Ann. Statist. 2, 1022-1028.

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

M.L. Eaton and M. Perlman (1977). Generating $O_{(n)}$ with relections. Pac. J. Math. **73**, No. 1, 73-80.

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

B. Efron (1969). Student's t-test under symmetry conditions. J. Amer. Statist. Ass'n. **64**, 1278-1302.

J.D. Esary, F. Proschan, and D.W. Walkup (1967). Association of random variables with applications. Ann. Math. Statist. **38**, 1466-1474.

K. Fan (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci. **37**, 760-766.

C.M. Fortuin, J. Ginibre, and P.N. Kasteleyn (1971). Correlation inequalities on some partially ordered sets. Comm. Math. Phys. 22, 89-103.

A. Giovagnoli and H.P. Wynn (1985). G-majorization with applications to matrix orderings. Linear Algebra and Its Applications. **67**, 111-135.

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

B.V. Gnedenko, A.N. Kolmogorov (1954). Limit Distributions of Sums of Independent Random Variables. (Translated by K.L. Chung). Addison-Wesley, Cambridge, Mass.

J. Hajek (1962). Inequalities for the generalized Student's distributions. Sel. Transl. Math. Statist. Prob. 2, 63-74.

G.H. Hardy, J.E. Littlewood and G. Polya (1929). Some simple inequalities satisfied by convex functions. Messenger Math. **58**, 145-152.

G.H. Hardy, J.E. Littlewood, G. Polya (1934,1952). Inequalities, 1st and 2nd ed., Cambridge University Press, London.

M. Hollander, F. Proschan, and J. Sethuraman (1977). Functions decreasing in transposition and their applications in ranking problems. Ann. Statist. 5, 722-733.

R. Holley (1974). Remarks on the FKG inequalities. Comm. Math. Phys. 36, 227-231.

P.L. Hsu (1938). Statistical Research Memoirs. Department of Statistics, University College, London.

A.T. James (1961). The distribution of noncentral means with known covariance. Ann. Math. Statist. 32, 874-882.

A.T. James (1964). Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. 35, 475-501.

K. Joag-dev, M. Perlman and L. Pitt (1983). Association of normal random variables and Slepian's inequality. Ann. Probab. 11, 451-455.

K. Jogdeo (1977). Association and probability inequalities. Ann. Statist. 3, 495-504.

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

S. Karlin (1968). Total Positivity, Vol. I. Stanford University Press, Stanford, California.

S. Karlin and Y. Rinott (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. J. Multivariate Anal. 10, 467-498.

S. Karlin and J. Rinott (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. Ann. Statist. 9, 1035-1049.

W. Kecs (1982). The Convolution Product. Editura Academiei and D. Reidel Publishing Company (Holland).

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

W.H. Lawton (1965). Some inequalities for central and non-central distributions. Ann. Math. Statist. 36, 1521-1525.

W.H. Lawton (1968). Concentration of random quotients. Ann. Math. Statist. 39, 466-480.

E.L. Lehmann (1955). Ordered families of distributions. Ann. Math. Statist. **26**, 399-419.

E.L. Lehmann (1966). Some concepts of dependence. Ann. Math. Statist. 37, 1137-1152.

E.L. Lehmann (1959). Testing Statistical Hypotheses, John Wiley, New York.

G.G. Lorentz (1953). An inequality for rearrangement. Amer. Math. Monthly, **60**, 176-179.

A.W. Marshall and I. Olkin (1974). Majorization in multivariate distributions. Ann. Statist. **2**, 1189-1200.

A.W. Marshall and I. Olkin (1979). Inequalities: Theory of majorization and its applications. Academic Press, New York.

A.W. Marshall and F. Proschan (1965). An inequality for convex functions involving majorization. J. Math. Anal. Appl. **12**, 87-90.

A.W. Marshall, D.W. Walkup, and R.J.-B. Wets (1967). Order-preserving functions: applications to majorization and order statistics. Pac. J. Math. 23, 569-584.

M.R. Mickey and M.B. Brown (1966). Bounds on the distribution function of the Behrens-Fisher statistic. Ann. Math. Statist. **37**, 639-642.

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

R. Muirhead (1982). Aspects of Multivariate Statistical Theory. Wiley, New York.

R.F. Muirhead (1903). Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. Proc. Edinburgh Math. Soc. 21, 144-157.

J. von Neumann (1937). Some matrix-inequalities and metrization of matrix space. Tomsk. Univ. Rev. 1, 286-300.

E. Nevius, F. Proschan, and J. Sethuraman (1977). Schur functions in statistics. II. Stochastic majorization. Ann. Statist. **5**, 263-273.

A.M. Ostrowski (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. J. Math. Pures Appl. **9**, 253-292.

M. Perlman and I. Olkin (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. Ann. Statist. **8**, 1326-1341.

L. Pitt (1982). Positively correlated normal variables are associated. Ann. Probab. 2, 496-499.

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

C.J. Preston (1974). A generalization of the FKG inequalities. Comm. Math. Phys. 36, 233-241.

F. Proschan and J. Sethuraman (1977). Schur functions in statistics I: The preservation theorem. Ann. Statist. 2, 256-262.

R. Rado (1952). An inequality. J. London Math. Soc. 71, 1-6.

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

T. Rockafellar (1970). Convex Analysis. Princeton University Press, Princeton, New Jersey.

A. Sampson (1983). Positive dependence properties of elliptically symmetric distributions. J. Multivariate Anal. 13, 325-331.

T.K. Sarkar (1969). Some lower bounds of reliability. Technical Report 124. Department of Operations Research and Statistics, Stanford University.

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

T. Savits (1985). A multivariate IFR class. J. Appl. Prob. 22, 197-204.

H. Scheffe (1959). Analysis of Variance. Wiley, New York.

I.J. Schoenberg (1951). On Polya frequency functions, I. The totally positive functions and their Laplace transforms. J. Analyze Math. 1, 331-374.

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

Z. Sidak (1967). Rectangular confidence regions for the means of multivariate normal distributions. J. Amer. Statist. Assoc. 62, 626-633.

D. Slepian (1962). The one-sided barrier problem for Gaussian noise. Bell System Tech. J., 41, 463-501.

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

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

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.