

UNIMODALITY, CONVEXITY AND PROBABILITY INEQUALITIES

BY

S. W. DHARMADHIKARI and KUMAR JOGDEO

TECHNICAL REPORT NO. 147

SEPTEMBER 1979

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 78-07736

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



UNIMODALITY, CONVEXITY AND PROBABILITY INEQUALITIES

By

S. W. Dharmadhikari* and Kumar Jogdeo*

TECHNICAL REPORT NO. 147

SEPTEMBER 1979

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT

MCS 78-07736
Ingram Olkin, Project Director

*Prepared with the partial support of NSF Grant No. MCS-78-04014 at the University of Illinois, and NSF Grant No. MCS-79-04014.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

UNIMODALITY, CONVEXITY AND PROBABILITY INEQUALITIES

Abstract

Concepts of unimodality have different manifestations when extended from univariate to multivariate probability distributions. Convexity plays an important role and some related notions such as Schur convexity give rise to corresponding concepts of unimodality. Recent results on this subject are brought together and their interrelationship is studied in this expository article. A fundamental and intuitively obvious probability inequality associated with a symmetric unimodal density was first generalized by T. W. Anderson (1955). Since then various generalizations and other versions of this inequality have appeared in the literature. These inequalities are stated in this article and it is shown that unimodality provides an important tool in obtaining various monotonicity results in multivariate statistics.

Key Words and Phrases: Univariate unimodality, Strong unimodality, Multivariate unimodality, Peakedness, Star unimodality, Brün Minkowski inequality, Monotone unimodality, Central convex unimodality, Convolutions of unimodal distributions, Schur convexity, Axial modality, Association, Probabilities of rectangular regions, Monotonicity of power of tests in MONAVA.

UNIMODALITY, CONVEXITY AND PROBABILITY INEQUALITIES

0. Introduction.

The definition of unimodality is given in almost every elementary text book. However, the usual purpose of introducing the concept is just to point out that the distributions commonly found in statistical work are all unimodal. The use of unimodality as a tool to obtain probability inequalities as well as the study of its close association with the concept of convexity is of recent vintage and is somewhat confined to the articles appearing in research journals. The purpose of the present exposition is to develop the notion of univariate and multivariate unimodality, emphasizing the ideas of geometry. This approach is simple, intuitive and makes many results transparent.

In Chapter 1, the concept of univariate unimodality is examined from various approaches. These basic results prepare the ground for the multivariate generalizations which are discussed in Chapter 2. Chapter 3 is devoted to the notions of multivariate unimodality obtained by generalizing the concept of convexity and associated notions of symmetry and majorization. In the last chapter some typical applications to statistical problems are given emphasizing the importance of the various approaches.

1. Univariate Unimodality.

In this chapter we are concerned mainly with real random variables.

1.1 The standard definition. A definition applicable to all distributions on R is as follows.

Definition 1.1. A real random variable Z has a unimodal distribution about a mode (vertex) v if its distribution function F is convex on $(-\infty, v)$ and concave on (v, ∞) .

The following are simple consequences of the above definition.

(i) If F is UM (the abbreviation for "unimodal" or "unimodality") then the left and right derivatives exist everywhere except possibly at v .

(ii) If F is differentiable for every $x \neq v$, and if f denotes its derivative, then unimodality of F is equivalent to f being nondecreasing on $(-\infty, v)$ and nonincreasing on (v, ∞) .

In many textbooks this is adopted as the definition of UM and the property of UM is associated with f .

(iii) If F is UM then the modes of F form an interval. The extreme example where every point in the support of F is a mode, is that of a uniform distribution.

(iv) Another extreme case is that of a degenerate distribution which is unimodal. However, this is the only discrete distribution which is unimodal according to Definition 1.1. In Chapter 3 a different definition will be given to treat discrete distributions.

(v) The class of UM distributions is closed under weak limits. This follows from the closure properties of convex functions. This fact allows us to assume "smoothness" conditions such as existence of densities, etc. without loss of generality. Once a property is established for this smooth case, the results can be extended to the general case by taking limits.

(vi) A "mixture" of UM distributions with a common mode is UM with the same mode.

Definition 1.2. A nonnegative function g defined on R will be called unimodal if there exists v such that g is nondecreasing on $(-\infty, v)$ and nonincreasing on (v, ∞) .

In the following, it is seen that, on the real line, the concept of unimodality has several characterizing properties.

1.2 Characterizing Properties.

Theorem 1.1: Let F be the distribution function of a random variable Z with density f (with respect to Lebesgue measure). Then the following statements are equivalent.

(a) F is unimodal with mode 0. (a_1) Density f is nondecreasing for $x < 0$ and nonincreasing for $x > 0$. (a_2) For every $c > 0$, the set $\{x: f(x) \geq c\}$ is a bounded interval.

(b) F belongs to the closed convex hull of the class of uniform distributions on intervals with one endpoint at 0.

(c) There exists a pair of random variables (X, U) , where U is uniform on $[0, 1]$, and is independent of X , such that the distribution function of the product UX is F .

(d) Let V_a be the distribution function corresponding to the uniform distribution over $[-a/2, a/2]$. Then for every $a > 0$, the convolution $V_a * F$ is unimodal.

(e) For every $\delta > 0$, $P\{-\delta + x \leq Z \leq \delta + x\}$ as a function of x , is UM.

(f) For every bounded, nonnegative Borel measurable function g , $t E g(tZ)$ is nondecreasing in t for $t > 0$.

Proof: We state the basic ideas involved in the proof. As remarked earlier we may assume that F has density f . Let v_a denote the density of V_a .

The equivalence of (a) and (c) can be seen by first noting that UX represents a mixture of uniform distributions. This becomes more vivid when X is finitely discrete, in which case the density of UX looks like left and/or right staircases leading to a "platform" directly above 0. (The case where $P\{X=0\} > 0$ corresponds to $P\{Z=0\} > 0$. Since the mode is at 0 we are not concerned about the density there.) Conversely, if f is UM it can be approximated by such a (possibly double) staircase density which has a representation of the density of UX for some discrete X .

Khintchine (1938) originated this representation (c) of UM distributions and it is useful for generalizing the concept of unimodality to the multivariate case. An elegant proof of Khintchine's result is due to L. A. Shepp and can be found in Olshen and Savage (1970).

Statements (b) and (c) are equivalent because the mixing distribution involved in (b) is just the distribution of X mentioned in (c).

That (b) implies (d) follows from the elementary fact that the convolution of two uniform distributions with 0 in their supports, is unimodal (the density of the convolution is either triangular or trapezoidal) with 0 as a mode. Thus if F is a mixture of uniforms with supports having one

endpoint at 0 then $F * V_a$ is a mixture of unimodal distributions each having 0 as a mode and hence UM. Conversely, if $V_a * F$ is UM for arbitrarily small a then, since V_a approaches the degenerate distribution as $a \rightarrow 0$, F has to be UM.

To check the equivalence of (d) and (e) note that if Y has uniform density v_a and Y is independent of Z , then the density of $Z + Y$ at x can be seen to be $P\{x-a/2 \leq Z \leq x+a/2\}/a$.

The implication (b) \Rightarrow (f) needs to be verified for uniform distributions only. This verification is straightforward. Conversely, suppose (f) holds. Let v_0 be nonzero and let

$$g(x) = \begin{cases} 1 & \text{if } x \in [v_0 - \epsilon, v_0 + \epsilon] \\ 0 & \text{otherwise} . \end{cases}$$

Then recalling that f is the density of Z , we see that

$$t E g(tZ) = t \int_{-\infty}^{\infty} g(tz) f(z) dz = \int_{v_0 - \epsilon}^{v_0 + \epsilon} f\left(\frac{x}{t}\right) dx .$$

Thus if $t_1 > t_2 > 0$, then (f) shows that

$$\int_{v_0 - \epsilon}^{v_0 + \epsilon} f\left(\frac{x}{t_1}\right) dx = t_1 E g(t_1 Z) \geq t_2 E g(t_2 Z) = \int_{v_0 - \epsilon}^{v_0 + \epsilon} f\left(\frac{x}{t_2}\right) dx .$$

This holds for every $\epsilon > 0$ and thus for every v_0 , we have

$$f\left(\frac{v_0}{t_1}\right) \geq f\left(\frac{v_0}{t_2}\right)$$

whenever $t_1 > t_2 > 0$. Hence f is UM. ||

1.3 Convolutions. Besides taking mixtures another important operation in probability and statistics is that of convolution. From the above it is clear that convolutions with uniform distributions preserves the UM property. K.L. Chung (see [18]) pointed out that the convolutions of two UM distributions may not be UM. With this negative result in the background there have been three approaches in studying convolutions.

a) Strong Unimodality. As mentioned above there are UM distributions such as the uniform which yield UM distribution when convoluted with every UM distribution. Ibragimov (1956) called a distribution G "strongly UM" if for every unimodal F , the convolution $G * F$ is UM. He then proceeded to give an elegant characterization for a strongly UM G possessing density g , namely, $\log g$ is concave on the interior of its support.

b) Symmetry. Suppose we restrict our attention to the subclass of UM distributions which are symmetric about a mode. Then this subclass is closed under convolutions. Let f_1, f_2 be two symmetric UM densities. Since unimodality is preserved under a shift in location, without loss of generality we may assume that both densities are symmetric about 0. Now Khintchine's representation for the symmetric case implies that f_1 and f_2 are mixtures of uniform distributions

over intervals which are symmetric about 0. Thus the convolution $f_1 * f_2$ can be viewed as a mixture of convolutions of uniforms symmetric about 0. Since the convolution of two symmetric (about 0) uniforms is either triangular or trapezoidal, symmetric about 0, $f_1 * f_2$ is unimodal symmetric about 0. This result was first stated by Wintner (1938).

It should be noted that the condition of symmetry for one of the densities above does not suffice to yield unimodality for the convolution. For example, in [9] it is shown that if u_a and v_a denote uniform distributions on $[0, a]$ and $[-a/2, a/2]$ respectively and if

$$f_1 = \frac{1}{2} [u_a + u_b] ,$$

$$f_2 = \frac{1}{2} [v_a + v_b] ,$$

where $b > 3a > 0$, then $f_1 * f_2$ is not UM.

Another result regarding symmetry and unimodality is that if Z_1, Z_2 are independent and having some UM distribution then $Z_1 - Z_2$ has a symmetric UM distribution. A proof which is somewhat complicated is provided by Hodges and Lehmann (1954). Khintchine's representation can be used to obtain a simpler proof.

c) Index of Modality. Although the convolution of two unimodal distributions is generally not unimodal, intuitively it is clear that the convolution would inherit some behavior of a unimodal distribution. For example if both modes are at 0 then for $x > 0$, the convolution density

$h(x)$, say, might rise in some region but may not rise at a fast rate. Perhaps one can specify the maximal rate at which $h(x)$ might increase. Olshen and Savage (1970) made this precise by calling h to be α -modal (about 0) if $h(x)|x|^{1-\alpha}$ is UM in x . The smallest value of such an α is of

importance. It turns out that the above description is equivalent to the following.

Definition. The distribution of a random variable Z is said to be α -modal about a vertex v if for every bounded, nonnegative Borel measurable f , $t^\alpha E\{f(t(Z-v))\}$ is nondecreasing in t for $t > 0$.

Theorem: (Olshen-Savage). Convolution of an α -modal distribution with a β -modal distribution is $\alpha+\beta$ modal.

This shows that the convolution of two unimodal distributions is 2-modal. In view of Wintner's result (see 1.3b), it is natural to ask whether symmetry for one of the distributions can reduce the minimum index of modality. In [9], we answer this question as follows.

Theorem: (a) Let P_1, P_2 be distributions in R such that P_1 is symmetric unimodal about 0 and P_2 is α -modal about 0.

- (i) If $1/2 < \alpha \leq 1$ then $P_1 * P_2$ is $3/2$ -modal about 0.
- (ii) If $1 \leq \alpha \leq \beta$ then $P_1 * P_2$ is $(2+\alpha)/2$ modal about 0.
- (iii) If $\alpha \geq 2$ then $P_1 * P_2$ is α -modal about 0.

(b) Suppose both P_1 and P_2 above are symmetric then (by Wintner's result) the index $3/2$ in (i) above can be replaced by 1. However no improvement is possible in (ii) or (iii).

(c) Let P_1, P_2 be α and β modal respectively, where $\alpha > 1$ and $\beta > 1$. Then even if both are symmetric the index of $P_1 * P_2$ can be as high as $\alpha + \beta$.

Remark: It is curious to note that the symmetry helps only for unimodals. As (b) above shows $\alpha > 1$ may be very close to 1 while the index of $P_1 * P_2$ may remain as high as $3/2$. However, at $\alpha = 1$ it drops to 1.

1.4 Peakedness. For two unimodal distributions which are symmetric about a common vertex, it is meaningful to ask which distribution assigns more mass near the vertex (which will be assumed to be 0 for the following discussion). Z. W. Birnbaum (1948) introduced a concept of "peakedness" to answer this.

Let X, Y be real random variables. The distribution of X is said to be more peaked about 0 than Y if $|Y| \overset{s}{>} |X|$ (where $\overset{s}{>}$ denotes stochastically larger). Unimodality and symmetry play an important role in applying this concept to the convolutions.

Theorem (Birnbaum (1948)). Let X_1, Y_1 be independent and X_2, Y_2 be independent real random variables with distributions symmetric about 0 and those of X_2 and Y_1 being UM. Further suppose that $|X_i| \overset{s}{<} |Y_i|$, $i = 1, 2$. Then

$$|X_1 + X_2| \overset{s}{<} |Y_1 + Y_2| .$$

This theorem raises a question regarding the connection between UM and peakedness. By Khintchine's theorem, if X, Y have symmetric UM distributions then there exists a pair (Z_1, Z_2) of

nonnegative random variables and an independent random variable V having uniform distribution on $[-1, +1]$ such that X and Y have the same distribution as VZ_1 and VZ_2 . Further, distributions of Z_1 and Z_2 are uniquely determined by those of X and Y . Clearly, if $Z_2 \stackrel{s}{>} Z_1$ then $|Y| \stackrel{s}{>} |X|$. It is natural to ask whether the converse holds, i.e., suppose X is more peaked than Y . Then do the distributions of Z_1 and Z_2 , determined by X and Y , necessarily obey $Z_2 \stackrel{s}{>} Z_1$? If so, this would form a useful tool. Unfortunately the answer to this is in the negative.

Example: Let

$$P[Z_1 = 2] = \frac{2}{5} = 1 - P[Z_1 = 4] ,$$

$$P[Z_2 = 10] = \frac{2}{5} = 1 - P[Z_1 = 1] .$$

Thus Z_1, Z_2 are not stochastically ordered. However, by defining $X = VZ_1$ and $Y = VZ_2$, it is easy to see X is more peaked than Y .

2. Multivariate Analogues.

In Chapter 1 several equivalent ways of describing UM for distributions on the real line were considered. As we shall see, their multivariate analogues give rise to nonequivalent notions of UM, although convexity plays a crucial role throughout. To avoid unnecessary details, we will assume the existence of densities with respect to Lebesgue measure.

2.1 Definitions. One of the early definitions of multivariate UM is due to Anderson (1955). A straightforward analog of (a_2) in Theorem 1.1 would require a density f to satisfy: for every $c > 0$ the set $\{\underline{x}: f(\underline{x}) \geq c\}$ to be convex. However, without an additional symmetry condition not much can be done with this concept of UM. Thus Anderson (1955) adopted the following

Definition 2.1. A random n -vector X or its distribution with a density f is said to be convex UM about $\underline{0}$ if for every $c > 0$ the set $\{\underline{x}: f(\underline{x}) \geq c\}$ is a centrally symmetric convex set.

The above definition has a nice geometric interpretation. Its graph in R^{n+1} can be visualized as a "hyperhill" with convex contours. Further, the convexity condition implies that every planar section parallel to the $(n+1)$ st co-ordinate axis gives a graph of a univariate UM function or in other words, the restriction of f to every line $\sum \alpha_i x_i = \beta$ is UM.

Unfortunately, the above definition turns out to be rather restrictive. First of all, the class of convex UM distributions is not closed under mixtures.

Example 2.1. Let C, D be centrally symmetric convex sets such that $C \cup D$ is not convex. Then the uniform distributions on C and D are convex UM but any proper mixture of those two distributions is not convex UM.

Further, X_1 and X_2 could be independent identically distributed with a common unimodal symmetric distribution and still the joint distribution may not be convex UM.

Example 2.2. Consider

$$f(x) = \begin{cases} 3/8, & |x| \leq 1, \\ 1/8, & 1 < |x| \leq 2. \end{cases}$$

If g is the joint density of x_1, x_2 independent, each having density f , then the set

$$\{(x_1, x_2): g(x_1, x_2) \geq 3/64\}$$

is not convex.

Convolution of two convex UM distributions (in spite of symmetry) may not be convex UM. Certain marginals of convex UM may not be convex UM. An example illustrating these two facts is essentially due to Anderson and can be found in Das Gupta (1976).

Many of these properties can be salvaged by the following less restrictive definition.

Definition 2.2. A distribution in R^n is said to be central convex UM if it is in the closed convex hull of the set of all uniform distributions on centrally symmetric convex sets.

Clearly, convex UM is strictly stronger than central convex UM. These two classes of UM distributions will be seen to possess an important monotone property which is analogous to (e) of Theorem 1.1.

Definition 2.3. A random n -vector X is said to have a monotone UM distribution if for every $z \in R^n$ and for every centrally symmetric convex set C in R^n , $P[C + kz]$ is non-increasing in $k \in [0, \infty)$.

The above two definitions are based on the results of Sherman (1955) who gave a simplified proof of the basic theorem of Anderson (1955) establishing the monotone UM property for convex UM distributions. Sherman also extended the theorem to central convex UM distributions. Geometrically, the monotone UM property means that a centrally symmetric convex set receives a decreasing amount of probability as it moves away from the origin.

Recently, Kantor (1977) gave, for symmetric distributions in R^n , a definition of unimodality which can be looked upon as the Krein-Milman version of Definition 2.2. Kantor's definition is identical with Definition 2.2 for distributions on R and it seems likely that this result will hold in higher dimensions also.

A variant to the concept in Definition 2.2 would be to consider mixtures of uniform distributions on convex sets which cover the origin. This leads to mixtures of uniform distributions on star-shaped sets.

(Note that a set S is called star-shaped if there exists a point $\xi \in S$ such that $x \in S \Rightarrow$ the line segment joining x and ξ is contained in S . Intuitively, this means that one can "see" every point of the set S from ξ . The set of such points from which every point of S can be "seen" is called the kernel of S . It is well known that the kernel of a star-shaped set is convex. If symmetry condition is desired one may consider uniform distributions on centrally symmetric star-shaped sets. The convex hull of these will produce a class of "star" UM symmetric distributions. The following definition is due to Olshen and Savage (1970).

Definition 2.4. A random n -vector X is said to have an α -UM distribution about 0 if for every bounded nonnegative Borel measurable function

g on R^n , $t^\alpha E[g(tX)]$ is nondecreasing in $t \in [0, \infty)$. For $\alpha = n$, the distribution is said to be Star UM.

Olshen and Savage (1970) show that the density f of an α -modal distribution has a characteristic monotone property, namely, $t^{n-\alpha} f(tx)$ is nonincreasing in $t \in [0, \infty)$. Thus for a star UM distribution, the set $\{\underline{x}: f(\underline{x}) > c\}$ is star-shaped. Note that every planar section of the graph of the density passing through $\underline{0}$ is a graph of a symmetric univariate UM function. However, it need not represent a probability density function since the area under it may not be unity. It is easy to see that this property in fact suffices for the density to be star (centrally symmetric) UM.

Finally, one can "linearize" the requirements so that univariate properties can be exploited.

Definition 2.5. A random n -vector $\underline{X} = (X_1, \dots, X_n)$ is said to have linear UM distribution with vertex (or mode) at $\underline{0}$, if for every vector $\underline{a} \in R^n$, the linear combination $\sum a_i X_i$ has a univariate UM distribution about 0 .

This definition was given by us in [10] and also by Ghosh (1974). Although, there is no symmetry condition imposed one may either require central symmetry to start with or require that $\sum a_i X_i$ be symmetric about 0 for every \underline{a} . Surprisingly, as we shall see later, for the special case of uniform distributions, Definition 2.5 implies central symmetry.

Many properties involving convolutions, mixtures etc., of linear UM distributions can be easily verified from the corresponding univariate results. However, our main interest is to study how this definition relates to the other concepts of unimodality. This is discussed in the next subsection.

2.2 Inter-relationships. Of all the definitions given above, that of convex UM is the most restrictive. That it implies central convex UM, follows by a simple argument. The following theorem due to Fary and Radei (1949) and Sherman (1955) is the key for showing that central convex UM implies monotone UM.

Theorem 2.1. Let C and D be compact convex subsets of R^n and let V_n be the Lebesgue measure in R^n . Then $\psi(\tilde{x}) = V_n^{1/n}[C \cap (D+\tilde{x})]$ is concave on its support.

It should be noted that symmetry is not required in the above theorem. This is the basic moving set inequality.

Proof. According to Brün-Minkowski inequality, for two sets A, B in R^n and $0 < \lambda < 1$,

$$V_n^{1/n}(\lambda A + (1-\lambda)B) \geq \lambda V_n^{1/n}(A) + (1-\lambda)V_n^{1/n}(B).$$

To prove the theorem we have to show that if \tilde{x}_1 and \tilde{x}_2 are in the support of ψ then $\psi(\lambda\tilde{x}_1 + (1-\lambda)\tilde{x}_2) \geq \lambda\psi(\tilde{x}_1) + (1-\lambda)\psi(\tilde{x}_2)$. Let $A = C \cap (D+\tilde{x}_1)$, $B = C \cap (D+\tilde{x}_2)$. Then

$$C \cap (D + \lambda\tilde{x}_1 + (1-\lambda)\tilde{x}_2) \supset \lambda A + (1-\lambda)B.$$

To see this let $y \in \lambda A + (1-\lambda)B$. Then there exist \tilde{y}_1, \tilde{y}_2 such that $\tilde{y}_1 \in C \cap (D+\tilde{x}_1)$ and $\tilde{y}_2 \in C \cap (D+\tilde{x}_2)$ and $y = \lambda\tilde{y}_1 + (1-\lambda)\tilde{y}_2$. Convexity of C implies $y \in C$. Further $\tilde{y}_1 - \tilde{x}_1 \in D$ and $\tilde{y}_2 - \tilde{x}_2 \in D$ and due to convexity of D ,

$\lambda y_1 + (1-\lambda)y_2 - (\lambda x_1 + (1-\lambda)x_2) \in D$ or, $y \in D + \lambda x_1 + (1-\lambda)x_2$. Hence y belongs to the left side. Consequently, by the Brun-Minkowski inequality

$$\begin{aligned}\psi(\lambda x_1 + (1-\lambda)x_2) &\geq V_n^{1/n}(\lambda A + (1-\lambda)B) \\ &\geq \lambda V_n^{1/n}(A) + (1-\lambda)V_n^{1/n}(B) \\ &= \lambda \psi(x_1) + (1-\lambda)\psi(x_2) .\end{aligned}$$

Now we are in a position to state a theorem which will summarize the inter-relations between various concepts of unimodality.

Theorem 2.2.

- (a) Convex UM \Rightarrow Central convex UM
 - \Rightarrow Monotone UM
 - \Rightarrow Star UM,

while implication in the other direction does not hold, in general, for any of the four concepts.

- (b) Monotone UM \Rightarrow Linear UM.
- (c) Between star UM and linear UM no implication relation exists.
- (d) For uniform distributions in R^2

linear UM \Rightarrow convex UM.

Proof. We provide either the basic ideas of the proofs for the above assertions or give appropriate references.

(a) The first implication follows from the usual approximation technique. Let c_i be an increasing sequence of positive numbers and define $A_i = \{x: f(x) > c_i\}$. Then A_i are centrally symmetric convex sets which are nested and being measurable, f can be approximated by a mixture of uniform distributions on A_i . As the sequence $\{c_i\}$ becomes dense these approximations will converge to f . That the reverse implication does not hold was already shown by Example 2.1.

To prove that central convex UM implies monotone UM one may proceed as follows.

From Theorem 2.1 it follows that if C and D are centrally symmetric then $\psi(x)$ is concave and centrally symmetric and hence $\{\psi(x)\}^n$ is a convex UM function. This shows that the probability density which is uniform on a centrally symmetric convex set is monotone UM. The same conclusion follows if a probability density is a mixture of uniform distributions on centrally symmetric convex sets or, equivalently if it is central convex UM.

The reverse implication was conjectured by Sherman (1955). If true this would have been the multivariate analog of Khintchine's representation of a symmetric UM distribution on the line. The following example was proposed in [10] to show that monotone UM does not imply central convex UM. Let ABC be an equilateral triangle in R^2 with O at its centroid and $A'B'C'$ be its reflection through the origin. The bivariate distribution which we want to consider on this Star of David set is the $(1/2, 1/2)$

mixture of uniform distributions on ABC and $A'B'C'$. Since these are three pairs of vertices, each pair being symmetrically located and having the same density, a mixture of uniform distributions on centrally symmetric convex sets will assign density at the origin 3 times that at a vertex. However, the distribution prescribed above has density at 0 only twice that at a vertex. Thus the above mixture is not central convex UM. D. Wells (1978) showed that this distribution is indeed monotone UM settling the Sherman conjecture in the negative.

To verify the last implication in (a) suppose P is a monotone UM distribution in R^n and let Q_δ denote the uniform distribution on the ball C_δ in R^n with center at 0 . Since P is monotone UM it follows that the density $f_\delta(\underline{x})$ of the convolution $P * Q_\delta$, which is a constant multiple of $P(C_\delta + \underline{x})$, has the property that $f_\delta(k\underline{x})$ is nonincreasing in $k \in [0, \infty)$ for every \underline{x} . Thus f_δ is star UM. Letting $\delta \rightarrow 0$ we see that P is star UM.

That the reverse implication does not hold can be seen from the following example. Consider the triangle in R^2 with vertices at 0 , $(1, -1)$ and $(-1, -1)$. The uniform distribution defined on the union of this triangle with its reflection through the origin produces a star UM distribution which is clearly not monotone UM.

(b) The assertion will follow from the proof of Theorem 2.3.

In (c) it will be shown that linear UM does not imply even the weaker condition of star unimodality so that, a fortiori, the reverse implication in (b) does not hold.

(c) The example given at the end of the proof of (a) gives a star UM distribution which has a bimodal marginal on the y -axis and thus is not linear UM. The following example shows that a linear UM distribution

in R^2 may have circular symmetry and the density may have a crater at the origin! Let

$$f(x,y) = Ke^{-(x^2+y^2)/2} [e^{a(x^2+y^2)/2} - b], \quad x \in R, y \in R.$$

By choosing $0 < a < 1$ and $b < 1$ the constant K can be chosen to make f a density function. Let

$$h(t) = e^{-t} [e^{at} - b].$$

Then $h'(0) = (a-1)+b$. Therefore

$$b > (1-a) \Rightarrow h \text{ is increasing at } 0$$

$$\Rightarrow f(tx,tx) \text{ is increasing in}$$

$$t \in [0, \delta) \text{ for some } \delta > 0$$

$$\Rightarrow f \text{ is not star UM.}$$

Now the marginal density f_1 is given by

$$f_1(x) = K \sqrt{2\pi} \left\{ \frac{e^{-(1-a)x^2/2}}{(1-a)^{1/2}} - be^{-x^2/2} \right\}.$$

Therefore

$$f_1'(x) = K \sqrt{2\pi} \quad xe^{-x^2/2} \left\{ b - \sqrt{1-a} e^{ax^2/2} \right\}.$$

If $b < \sqrt{1-a}$ then

$$f_1'(x) \begin{matrix} > \\ \geq \\ < \end{matrix} 0 \text{ according as } x \begin{matrix} \leq \\ \geq \\ < \end{matrix} 0.$$

Hence for $b < \sqrt{1-a}$ the density f_1 is UM and symmetric. Thus for $1-a < b < \sqrt{1-a}$, the density f is linear UM but not star UM.

(d) This assertion was proved by us in [8] for polygonal sets and by G. Converse (1977) for more general sets. It is worthwhile to mention here that the requirement of a mode in Definition 2.5 is crucial for this result. Let P denote the uniform distribution on a compact set A in R^2 and suppose that A is the closure of its interior. This last condition avoids sets with isolated points and protruding line segments. If, under P , every linear function of the co-ordinates has a UM distribution, then the set A need not even be star-shaped. However, if there exists a modal point (v_1, v_2) in A , that is, if every $a_1 X_1 + a_2 X_2$ is UM about $a_1 v_1 + a_2 v_2$ then the set A is forced to be not only convex but also centrally symmetric. Intuitively, we feel that a similar result should hold in R^n with $n > 2$.

Remarks: (i) A monotone UM distribution has to be centrally symmetric. The univariate case of this assertion follows rather easily from Theorem 1.1 (e). The multivariate case reduces to the univariate case because of assertion (b) above.

(ii) In order to verify monotone unimodality, it suffices to have the monotone property of Definition 2.3 for every symmetric compact convex body. Further it can be shown that the set of all monotone UM distributions is closed under weak limits. These assertions are proved in [10].

(iii) The example in (c) above shows that the concept of linear UM is geometrically unnatural. If the density f in R^n is spherically symmetric then convex UM, monotone UM and star UM concepts are all equivalent. Linear UM, however, cannot be included in this list.

2.3 Marginal Distributions. It is important for applications to have a condition on a multivariate distribution inherited by its marginals. In the following discussion the marginals are considered to be projections on lower dimensional spaces, which are not necessarily of dimension one.

Theorem 2.3. (a) Marginals of convex UM are not convex UM in general. However, a uniform distribution on a centrally symmetric convex set has all its marginals convex UM.

- (b) Marginals of monotone UM are monotone UM.
- (c) Marginals of central convex UM are central convex UM.
- (d) Marginals of linear UM are (trivially) linear UM.
- (e) Marginals of star UM need not be star UM.

Proof. (a) The following example is due to Anderson and explicitly stated by Sherman (1955). Let f be a density in R^4 defined on the subset of $S = \{(x_1, x_2, y_1, y_2): |x_1| \leq 1, |x_2| \leq 1, |y_1| \leq 2, |y_2| \leq 5\}$. On S , let $f(x_1, x_2, y_1, y_2) = kg(x_1 + y_1, x_2 + y_2)$ where

$$g(t_1, t_2) = \begin{cases} 3, & \text{if } |t_1| \leq 1, |t_2| \leq 1 \\ 2, & \text{if } |t_1| \leq 1, 1 < |t_2| \leq 5 \\ 0, & \text{otherwise.} \end{cases}$$

Thus the sets where $f(x_1, x_2, y_1, y_2) > c$ are nested and centrally symmetric convex sets. Now fix $y_1 = 0.5$ and $y_2 = 4$ so that $g = 2$ for $-1 \leq x_1 \leq 0.5$, $-1 \leq x_2 \leq 1$ and 0 otherwise. Then the two dimensional (y_1, y_2) marginal density at this pair is given by

$$f_2(.5, 4) = k \int_{-1}^1 \int_{-1}^{0.5} g(x_1 + 0.5, x_2 + 4) dx_2 dx_1 = 6k.$$

Similarly, $f_2(1, 0) = 6k$. However at $(.75, 2)$ which is the midpoint of the line segment joining $(.5, 4)$ and $(.1, 0)$,

$$f_2(.75, 2) = 5k$$

and this particular section of f_2 is not unimodal. Thus a marginal of 2 or more dimensions derived from a convex UM may not be convex.UM. Assertion (b) however implies that one dimensional marginals of convex UM are necessarily UM.

To prove the assertion regarding uniform distributions write $R^n = \Omega_1 \times \Omega_2$ where $\Omega_1 = R^m$. Let P denote the uniform distribution on a compact convex body C in R^m and let P_1 be the marginal of P on Ω_1 . Let $D_1 \subset \Omega_1$ and $D_2 \subset \Omega_2$ be convex compact bodies. Let $D = D_1 \times D_2$ and V_r be r -dimensional Lebesgue measure. By Theorem 2.1,

$$\{P(D+z)\}^{1/n} = \frac{V_n^{1/n}[C \cap (D+z)]}{V_n^{1/n}[C]}, \quad z \in R^n,$$

is concave on its support, namely $C+D$. Letting $D_2 \rightarrow \Omega_2$, $P_1^{1/n}(D_1+x)$ is concave on its support, namely C_1+D_1 , where C_1 is the projection of C on Ω_1 . The density f_1 of P_1 is given by

$$f_1(x) = \liminf_{D_1 \rightarrow \{0\}} [P_1(D_1+x)/V_m(D_1)], \quad x \in \Omega_1.$$

Since the \liminf of concave functions is concave, $f_1^{1/n}$ is concave on its support C_1 . Thus f_1 is convex UM if C is centrally symmetric.

(b) Let $X = (X_1, \dots, X_n)$ have a monotone UM distribution denoted by P and let $Y = (X_1, \dots, X_m)$, $m < n$. Let $C \subset R^m$ be symmetric, convex and let $y \in R^m$ be non-zero. Write $D = C \times R^{n-m}$ for the cylinder set in R^n and write $x^* = (y, 0) \in R^n$. If Q denotes the distribution of Y then $Q(C+ky) = P[D+kx^*]$ which is nonincreasing in $k \in [0, \infty)$ due to monotone unimodality of P and the assertion follows.

(c) Marginal of a mixture is a mixture of marginals and the result follows from the assertion in (a) about uniform distributions and the fact that a convex UM distribution is also central convex UM.

(e) The example given in the proof of Theorem 2.2(a) to show that star UM does not imply monotone UM, has a bimodal marginal.

2.4 Convolutions. In statistics as well as in probability theory, convolution is one of the most common operations performed. Hence it is important to study whether a unimodality property is preserved under convolution. The basic approach would be to inspect the univariate case and examine the convolution under analogous conditions. Two conditions which were discussed in Chapter 1 were: a) symmetry, b, the shape of the density, such as log concavity. In most of the multivariate distributions the symmetry condition is that of central symmetry which arises naturally. In Chapter 3 we will discuss other notions of symmetry which will give rise to still another set of definitions of multivariate unimodality. Approach (b) does not have the same success as in the univariate case. More precisely, in order to remove the symmetry condition both the densities in the convolution have to satisfy the condition on the shape. In particular if the two densities are log concave then their convolution is log concave. This result was first proved by Prékopa (1973).

Another approach would be to study the index of modality of the convolution of two multivariate distributions as done by Olshen and Savage (1970).

In the following theorem we summarize the results about convolutions.

Theorem 2.4. Let \underline{X} and \underline{Y} be independent n -vector random variables with probability densities f and g . Then

- (a) f, g convex UM $\not\Rightarrow f * g$ convex UM.
- (b) f, g central convex UM $\Rightarrow f * g$ central convex UM.
- (c) f monotone UM, g central convex UM $\Rightarrow f * g$ monotone UM.

(d) $f, g \text{ star UM} \not\Rightarrow f * g \text{ star UM}.$

(e) $f, g \text{ logconcave (strongly unimodal)} \Rightarrow f * g \text{ logconcave}.$

Proof. (a) The density given in the example for the proof of Theorem 2.3(a) can be convoluted with a uniform distribution yielding a non-convex UM distribution. Simpler examples are easy to construct.

(b) Since the convolution of mixtures is a mixture of convolutions it suffices to show that convolution of two uniform distributions on centrally symmetric convex sets is central convex UM. However, this is an immediate consequence of Theorem 2.1. This result was first stated by Sherman (1955).

(c) By the same reasoning as in (b) it suffices to show the result when g is the density of a uniform distribution on a centrally symmetric convex set C say. Let P_1 be the probability measure corresponding to f , P_2 corresponding to g and P_3 corresponding to the uniform distribution on a centrally symmetric convex set D . We want to show that for every \underline{x} , $P_1 * P_2(D+k\underline{x})$ is nonincreasing in $k \in [0, \infty)$. However, D being centrally symmetric,

$$P_1 * P_2(D+k\underline{x}) = \int_D f * g(-\underline{y}+k\underline{x}) d\underline{y} = k \int f * g(-\underline{y}+k\underline{x}) dP_3(\underline{y}).$$

The right side is a constant multiple of the density of $P_1 * P_2 * P_3$ evaluated at $k\underline{x}$. However, $P_2 * P_3$ is central convex UM and thus a mixture of uniform distributions on centrally symmetric convex sets. Thus the required monotonicity result will follow if we show that the density of $P_1 * (\text{uniform on centrally symmetric convex set})$ at $k\underline{x}$ is nonincreasing in $k \in [0, \infty)$. However, P_1 is monotone UM and this follows from the definition of monotone unimodality. This simpler proof was shown to use by Professors M. Perlman and S. DasGupta.

(d) Note that for the univariate case a star UM symmetric distribution is the same as a symmetric unimodal distribution and thus by Wintner's theorem the convolution of such distributions are star UM. However, the following example shows that even with symmetry the convolution of two bivariate star UM distributions may not be star modal. Let X, Y be independent random variables with a common density $p(x) = |x|$, $-1 < x < 1$. According to the definition of α -modality, X and Y have 2-modal distributions. Therefore $(X, 0)$ and $(0, Y)$ are independent pairs each having star UM distributions (although the densities do not exist). Their sum has the density,

$$f(x, y) = |xy|, \quad |x| < 1, \quad |y| < 1,$$

which is 4-modal and not 2-modal.

(e) This result has been proved by several authors by using different techniques. Apparently the simplest so far seems to be the proof given by Brascamp and Lieb (1975). Their approach is to reduce the problem to the two dimensional case. Although the proof is simple it is too long to be reproduced here.

Remark. It may be conjectured that the convolution of two monotone UM distributions is monotone UM. Unfortunately, there is no convenient criterion to check monotone unimodality so that verification of this conjecture seems to be difficult.

3. Some Other Notions of Unimodality.

In this chapter we mainly ~~discuss~~ some generalizations of convexity and build the concept of unimodality utilizing these notions. Recall that in Chapter 2, centrally symmetric convex sets were the building blocks for defining central convex UM densities. The basic result associated with such a density is the moving set inequality where the set is chosen as one of the building blocks. Thus three ingredients are involved: the concepts of convexity, of symmetry, and a partial ordering among the sets to identify a "movement" of a set from the "center" of the density. The partial ordering is usually termed as majorization. The nature of symmetry is closely tied with the choice of convexity. In Chapter 2, our basic assumption is that of central symmetry. The only definition of multivariate unimodality where the symmetry does not play a major role is that of star UM, due to Olshen and Savage (1970). Instead of using uniform distributions on centrally symmetric convex sets, the building blocks are star shaped sets, with the origin being in the kernel of the set. The sacrifice of the symmetry and convexity conditions results in a weaker moving set inequality. From Definition 2.4 (of star UM), it can be seen that the "movement" of a set is really a contraction (or elongation) along the rays and potentially seems to be less useful.

In the following we consider two notions of symmetry. The first is based on permutation and leads to Schur convexity. The usual notion of a convex set requires every linear section to be an interval. For Schur convexity this is required only for those sections which are

perpendicular to the line L passing through the origin and the point $(1,1,\dots,1)$. Besides this property, if the set is permutation symmetric then it is called Schur convex.

A host of inequalities can be obtained by using Schur convex sets as building blocks. This is the subject of the forthcoming book by Marshall and Olkin (1979). We **show how** the basic results can be interpreted in terms of unimodality and moving set inequality in Subsection 3.1 and cite some typical applications of these in Chapter 4.

The second notion of symmetry considered is symmetry about the co-ordinate axes or equivalently sign-invariance. The corresponding convexity notion is similar to Schur convexity when rotated by 45° angle. We give more details and unimodality interpretations in Subsection 3.2. In Chapter 4 this will be used to derive some results related to positive dependence.

The above two symmetry notions and the corresponding unimodality concepts are particular cases of those based on general reflection groups. This approach has roots in a generalization of the moving set inequalities of Chapter 2 by Mudholkar (1966) and was developed by Eaton and Perlman (1977) and Conlon, Leon, Proschan and Sethuraman (1977). However, we do not consider this abstraction. One may go to the other extreme and consider spherical symmetry. Although unimodality based on uniform distributions on spheres has applications (see for example Wolfe (1975)), most of the notions in Chapter 2 coincide and many consequences follow fairly easily. We do not consider this case either.

3.1 Schur convexity and modality. We follow the same notation as in Marshall and Olkin (1974). First we treat the bivariate case.

Recall that L is the line passing through $(0,0)$ and $(1,1)$.

Let a and b be two vectors in R^2 . The vector a is said to be majorized by b and written $a \prec b$, if upon reordering the two components such that $a_1 \geq a_2$ and $b_1 \geq b_2$, $a_1 \leq b_1$ and $a_1 + a_2 = b_1 + b_2$. In other words (i) a and b lie on a straight line perpendicular to the line L and (ii) a is closer to the line L than b is.

A measurable set B is Schur convex if every section of B perpendicular to the line L is convex and is symmetric about L .

A measurable real function ϕ is called Schur concave if $a \prec b \Rightarrow \phi(a) \geq \phi(b)$ or equivalently if every section of ϕ perpendicular to the line L is unimodal and symmetric about the line.

By Khintchine's representation (Theorem 1.1b) it follows that the class of Schur concave density functions is the convex hull of uniform distributions on Schur convex sets and distributions degenerating on the line L . A density function which is Schur concave is called Schur modal.

Extension of these concepts to R^n is somewhat straightforward.

Let the co-ordinates of a and b in R^n be reordered so that

$a_1 \geq a_2 \geq \dots \geq a_n$, $b_1 \geq b_2 \geq \dots \geq b_n$. Then $a \prec b$ if

$$(i) \quad \sum_{i=1}^k a_i \leq \sum_{i=1}^k b_i, \quad \text{for } k = 1, \dots, n-1,$$

$$(ii) \quad \sum_{i=1}^n a_i = \sum_{i=1}^n b_i.$$

A measurable set B in R^n is Schur convex if $a \in B \Rightarrow$ the convex hull of the set of points obtained by reordering co-ordinates of a , is a subset of B .

A measurable function φ is Schur concave if $a \prec b \Rightarrow \varphi(a) \geq \varphi(b)$.

One of the happy coincidences of concepts of Schur convexity in \mathbb{R}^n is that these can be expressed completely in terms of bivariate concepts. For example, if $a \prec b$ then one can find intermediate points c_1, \dots, c_m such that $a \prec c_1 \prec c_2 \prec \dots \prec c_m \prec b$, where c_i and c_{i+1} are identical except for only 2 components. This can be used to show that for φ to be Schur concave, all its bivariate sections given by

$$\varphi_2(x_i, x_j) = \varphi(x_1, \dots, x_i, \dots, x_j, \dots, x_n)$$

are Schur concave. Here all the arguments other than x_i, x_j are kept fixed.

Note that the standard criterion for a differentiable function to be Schur concave is that for every \underline{x} ,

$$\left(\frac{\partial \varphi(\underline{x})}{\partial x_i} - \frac{\partial \varphi(\underline{x})}{\partial x_j} \right) (x_i - x_j) \leq 0,$$

for every i, j . It is clear that Schur convexity (or concavity) is a bivariate concept. As seen from the earlier discussion, the bivariate Schur modal function can be viewed as a function obtained by splicing symmetric unimodal functions along L . The main result, stating the related moving set inequality is also a "spliced version" of the univariate result of Theorem 1.1 (e).

Theorem 3.1. (Marshall and Olkin 1974).

Let $\underline{X} = (X_1, \dots, X_n)$ have a joint density that is Schur modal and let A be a Schur convex set. Then

$$h(\underline{z}) = P[\underline{X} \in A + \underline{z}]$$

is Schur concave in \underline{z} .

As pointed out earlier, it suffices to verify this result for the bivariate case, so that one can assume that $n = 2$. Let \underline{z}^* and \underline{z}^{**} be two vectors in R^2 such that $\underline{z}^* \prec \underline{z}^{**}$. Then the two sets $A + \underline{z}^*$ and $A + \underline{z}^{**}$ can be viewed as two positions of a set which is being moved away in the direction perpendicular to L . Since the density of \underline{X} is Schur modal, the sections of $A + \underline{z}^*$, perpendicular to L receive larger areas from the corresponding density sections than those for $A + \underline{z}^{**}$. Hence it follows that

$$P[\underline{X} \in A + \underline{z}^*] \geq P[\underline{X} \in A + \underline{z}^{**}] ,$$

establishing Schur concavity of h .

Similarly, by using the Wintner property it can be shown that the convolution of two Schur concave densities is Schur concave.

3.2 Sign Invariance and Axial Modality. A measurable set D in R^n is said to be sign invariant if

$$(\underline{x}_1, \dots, \underline{x}_n) \in D \Rightarrow (\epsilon_1 \underline{x}_1, \dots, \epsilon_n \underline{x}_n) \in D ,$$

where each ϵ_i is $+1$ or -1 , $i = 1, \dots, n$.

A measurable set D is said to be axially convex if D is sign invariant and $\underline{x} \in D \Rightarrow \underline{y} \in D$, whenever $|y_i| \leq |x_i|$, $i = 1, \dots, n$. Thus the concept of majorization relevant to axial convexity is simply the partial ordering

by absolute values. A non-negative (measurable) function ϕ on R^n is called axially concave (or axially modal) if y is sign invariant and $|y| \leq |x| \Rightarrow \phi(y) \geq \phi(x)$. Such a function can also be described as "decreasing in absolute value". Again, axial unimodality is a "spliced version" of univariate unimodality (with symmetry) where splicing is done along the axes. It is easy to verify the following theorem.

Theorem 3.2. (Jogdeo, 1977). (a) Let D be an axially convex set, $z \in R^n$, and X have an axially modal density. Then the function

$$g(z) = P[X \in D + z]$$

is axially concave.

(b) Convolutions of axially modal functions are axially modal.

3.3 Some Remarks. (i) Note that in Theorems 3.1 and 3.2, the argument in functions g and h causes movement of the sections of the set by the same magnitude. Thus the sets move without any distortion. This is not crucial. If one vector makes the sections move more than the other, with varying amounts, the probability inequality still holds.

(ii) The abstract version for a reflection group can be visualized from Schur and axial modality. However, the splicing of unimodal sections can be visualized in a variety of ways. It could be done along some curves provided the "convexity" is imposed along the same contours and if majorization follows the same path. Of course, the natural question would be how useful these notions are. At least for the concepts introduced in Sections 3.1 and 3.2 we will provide applications in Chapter 4.

3.4 Unimodality Based on a Notion of Peakedness. For a centrally symmetric distribution the concept of unimodality can be formulated by requiring the following property. Let A, B be a pair of centrally symmetric convex sets. If, for every such pair,

$$P[A \cap B] \geq P[A]P[B] ,$$

then the distribution is said to be "unimodal with a peakedness property" or P-modal. It can be checked that uniform distributions on centrally symmetric convex sets need not be P-modal! (These were unimodal according to every definition in Chapter 2.) Consider the uniform distribution on the unit square. The square obtained by joining the midpoints of the sides carries mass $1/2$ and can be viewed as the intersection of the following two sets. The first one is the union of the small (inner) square and the triangles at north-east and south-west corners and the second obtained by joining the small square to the triangles at the other corners. These hexagonal sets carry mass of $3/4$ each and their product is more than the mass $1/2$ carried by the intersection. The P-modality is utterly trivial for the univariate distributions and was shown for the bivariate normal recently by L. Pitt (1977). The problem for the multivariate normal distribution is still open. Since centrally symmetric convex sets can be viewed as intersections of centrally symmetric strips, the problem is equivalent to the following. Let \tilde{X} have a multinormal distribution with mean vector 0 , and let (I_1, I_2) be an arbitrary partition of the set of integers $\{1, 2, \dots, n\}$. The result

$$P[|X_i| \leq c_i, i=1, \dots, n] \geq P[|X_i| \leq c_i, i \in I_1] P[|X_j| \leq c_j, j \in I_2],$$

would imply that \tilde{X} is P-modal.

We consider similar inequalities in the next chapter in another context.

3.5. Remarks on the Discrete Case. The only discrete distribution which is UM according to Definition 1.1 is the degenerate distribution. Therefore an alternative definition is used in the discrete case. We consider only those distributions which are concentrated on the set of integers.

Definition 3.5.1. A distribution $\{p_n, -\infty < n < \infty\}$ is called unimodal about M if

$$p_n \geq p_{n-1} \quad \text{for } n \leq M$$

and

$$p_n \leq p_{n-1} \quad \text{for } n \geq M.$$

Definition 3.5.2. A distribution is called strongly unimodal if its convolution with every UM distribution is UM.

These definitions were given by Keilson and Gerber (1971), who proved that a distribution $\{p_n\}$ is strongly UM if, and only if,

$$p_n^2 \geq p_{n-1} p_{n+1} \quad \text{for all } n.$$

Thus Ibragimov's characterization of strong UM has a perfect discrete analog. However, the discrete case has its own special features which are not present in the general case.

Let \mathcal{D} denote the set of all distributions on the set of integers which are unimodal about 0. Then \mathcal{D} is a convex set whose extreme points are the uniform distributions on $\{-j, -j+1, \dots, k\}$, where j and k are nonnegative integers. This is the first point of departure from the general case. In the general case, one only gets the unilateral uniform distributions as the extreme points because the uniform distribution on $(-a, b)$ is a mixture of the uniform distributions on $(-a, 0)$ and $(0, b)$. Such a result cannot hold in the discrete case because the mass at 0 gets inflated. So, let u_{jk} denote the uniform distribution on the set $\{-j, -j+1, \dots, k\}$. The present authors (1976) proved the following theorem.

Theorem 3.3. A distribution $p = \{p_n, -\infty < n < \infty\}$ is UM about 0 if, and only if, it has a representation

$$\tilde{p} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_{jk} u_{jk},$$

where c_{jk} are nonnegative constants such that

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_{jk} = 1.$$

The above representation need not be unique. For example,

$$\frac{1}{2}[u_{k0} + u_{0k}] = \frac{2k+1}{2(k+1)} u_{kk} + \frac{1}{2(k+1)} u_{00}.$$

This provides the second point of departure from the general case, where there is always a unique representation of a unimodal distribution as a "mixture" of the uniform distributions.

For some additional results on discrete unimodality see the papers by Keilson and Gerber (1971) and the present authors (1976).

Much work needs to be done on the discrete case. For instance, can one apply the concept of α -unimodality to discrete distributions? What would be an appropriate definition of unimodality for distributions on the lattice points in the plane or even higher dimensional spaces?

4. Applications.

Most of the applications where convex unimodality is used are related to the multinormal density. The basic moving set inequality yields monotonicity properties for certain test procedures as well as points out how probabilities of certain regions change when correlations are changed.

Schur convexity arises in several applications. We will cite three.

(As stated earlier, the forthcoming book by Marshall and Olkin discusses several of these in detail.) Axial modality will be used to establish positive dependence called association for certain linear models. In order to facilitate the discussion of these concepts of positive dependence, we give a brief summary of the interrelations. Let (X,Y) be a pair of random variables and consider

$$(A) \quad \text{Cov}(X,Y) \geq 0$$

(B) Positive Quadrant Dependence: For every pair of nondecreasing (nonincreasing) functions f_1, g_1 ,

$$\text{Cov}[f_1(X), g_1(Y)] \geq 0$$

Equivalently, for every pair (x,y)

$$P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y] .$$

(see Lehmann (1966)).

(C) Association: (see Esary, Proschan and Walkup, 1967).

For every pair of nondecreasing (nonincreasing) functions f_2, g_2 on $R^2 \rightarrow R$,

$$\text{Cov}[f_2(X,Y), g_2(X,Y)] \geq 0 .$$

(D) Positive Regression Dependence: For every nondecreasing function f_1 , $E[f_1(X)|Y = y]$ is nondecreasing in y , for almost all y , or the above property holds when (X, Y, y) is replaced by (Y, X, x)

It is easy to show that $(D) == (C) == (B) == (A)$ (see, for example Esary, Proschan and Walkup (1967)). Property (B) and its multivariate analog have applications in confidence interval estimation (see, for example, Šidák (1968)). It is seen below that the concepts of unimodality and corresponding moving set inequalities play an important role in establishing the positive dependence properties. Multivariate association property for R^k can be defined by requiring covariance of every pair of nondecreasing (nonincreasing) functions on $R^k \rightarrow R$ to be nonnegative.

4.1 Probabilities of rectangular regions and their dependence on correlation under multinormal density. Let (X_1, \dots, X_k) be normally distributed with mean vector 0. Intuitively it is clear that $|X_i|$ are positively dependent in some sense. How can one define this precisely? Does this dependence increase with correlation coefficients? In the case of mutual independence the distribution function $P\{|X_i| \leq x_i, i = 1, \dots, k\}$ would be the product of individual distribution functions, so that at least one should have

$$(4.1) \quad P\{|X_i| \leq x_i, i = 1, \dots, k\} \geq \prod_{i=1}^k P\{|X_i| \leq x_i\}.$$

Let us examine the bivariate case. It is easy to see that conditional

probability, $P[|X_1| \leq x_1 | X_2 = x_2]$, as a function of x_2 depends only on $|x_2|$ (due to symmetry) and is decreasing in $|x_2|$ (due to unimodality) as long as the correlation between X_1, X_2 is nonzero. Thus $|X_2|$ is positively regression dependent on $|X_1|$. For studying the dependence of $P[|X_i| \leq x_i, i=1, \dots, k]$ on correlation coefficients, the following result of Slepian (1962) is useful. Let $\varphi(x_1, \dots, x_k)$ be the multinormal density and ρ_{ij} be the covariance between i^{th} and j^{th} component. Then for every pair $(i, j), i \neq j$,

$$(4.2) \quad \frac{\partial \varphi(x_1, \dots, x_k)}{\partial \rho_{ij}} = \frac{\partial^2}{\partial x_i \partial x_j} \varphi(x_1, \dots, x_k) .$$

From this it follows that if all the parameters are kept fixed except the correlation coefficient ρ_{ij} the probability $P[|X_i| \leq x_i, i = 1, \dots, k]$ is non-decreasing in ρ_{ij} . Now if (X_1, X_2) have a bivariate normal distribution with mean 0, variances one and correlation coefficient ρ then by (4.2)

$$\begin{aligned} \frac{d}{d\rho} P[|X_1| \leq c_1, |X_2| \leq c_2] &= \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \frac{\partial^2}{\partial x_1 \partial x_2} \varphi(x_1, x_2) dx_1 dx_2 \\ &= 2\{\varphi(c_1, c_2) - \varphi(-c_1, c_2)\} \underset{\leq}{\overset{\geq}{\approx}} 0 \text{ according as } \rho \underset{\leq}{\overset{\geq}{\approx}} 0 . \end{aligned}$$

Thus $P[|X_1| \leq c_1, |X_2| \leq c_2]$ increases when ρ is "increased along its direction". The following multivariate analog of this was proved by Šidák (1968) and its simpler proof based on the directional derivative was provided by one of us (1970).

Let \tilde{X} be multinormal with mean vector $\tilde{0}$ and unit variances (as will be seen from the statement of the results, this assumption can be made without loss of generality). Let $\text{cov}(X_1, X_j) = \lambda \rho_{1j}$, $j > 1$. Then for every $\tilde{c} = (c_1, \dots, c_k)$ where c_i are positive, $P[|X|_i \leq c_i, i=1, \dots, k]$ is nondecreasing in λ for $\lambda > 0$.

To see how the moving set inequality plays a crucial role we point out the key steps in the proof of the assertion. Note that if the density f is convex unimodal in R^n then under differentiability conditions, the moving set inequality (see Theorem 2.2) can be paraphrased as:

$$(4.3) \quad \int_E \sum_i a_i \frac{\partial f(\tilde{x} + \tilde{a})}{\partial x_i} \Pi \, dx_i \leq 0 \quad ,$$

for all $\tilde{a} \in R^n$ and for all centrally symmetric convex sets $E \subset R^n$.

Sidak's above-mentioned result is equivalent to

$$(4.4) \quad \sum_{j=2}^k \rho_{1j} \frac{\partial}{\partial \rho_{1j}} P[|X_i| \leq c_i, i=1, \dots, k] \geq 0 \quad .$$

Suppose B denotes the box $[-c_2, c_2] \times \dots \times [-c_k, c_k]$. Then the left side of (4.4) can be seen to be equal to

$$\begin{aligned} & 2 \sum_{j=2}^k \rho_{1j} \int_0^{c_1} \int_B \frac{\partial^2}{\partial x_1 \partial x_j} \varphi(x) \Pi \, dx_i \\ &= 2 \sum_{j=2}^k \rho_{1j} \int_B \frac{\partial}{\partial x_j} \varphi(c_1, x_2, \dots, x_k) \prod_{i=2}^k dx_i \quad . \end{aligned}$$

The integral in the last expression can be written in terms of the con-

ditional density g (say) of X_2, \dots, X_k given $X_1 = c_1$. Writing $y = (x_2, \dots, x_k)$ and $\rho = (\rho_{12}, \dots, \rho_{1k})$, we see that

$$(4.5) \quad \text{the left side of (4.4)} = \text{const.} \int_B \sum_{j=2}^k \rho_{1j} \frac{\partial}{\partial x_j} g(y - c_1 \rho) \prod_{i=2}^k dx_i$$

It follows from (4.3) that the right side of (4.5) is nonnegative.

An immediate consequence of this monotone property is that for every $c_i > 0$, $i = 1, \dots, k$

$$(4.6) \quad \begin{aligned} P[|X_i| \leq c_i; i=1, \dots, k] &\geq P[|X_1| \leq c_1] P[|X_j| \leq c_j; j=2, \dots, k] \\ &\geq \prod_{i=1}^k P[|X_i| \leq c_i]. \end{aligned}$$

Šidák (1968) points out via an example that the probability of a rectangular region could decrease if only one of the correlation coefficients is increased. Secondly, for $k \geq 3$, $|X_i| \leq c_i$ cannot be replaced by $|X_i| \geq c_i$ in (4.6). Thus in general the distribution of $|\tilde{X}| = (|X_1|, \dots, |X_k|)$ is not associated if \tilde{X} has multinormal distribution with mean vector 0 and arbitrary covariance matrix. We will study some conditions under which association can be established for $|\tilde{X}|$.

4.2 Applications in MANOVA. Let $Y: n \times p$ be a random matrix whose rows are independent multinormal vectors with a common covariance matrix

$\Sigma: p \times p$. Let $S: p \times p$ be independent of Y having Wishart distribution $W(\Sigma, p, m)$. We assume $p \leq \min(m, n)$. Let $EY = \mu$ and suppose tests are to be considered for testing $H_0: \mu = 0$ vs. $H_1: \mu \neq 0$.

By the usual invariance considerations it can be shown that an equivariant test procedure is based on the p largest characteristic roots of $YS^{-1}Y'$ whose joint distribution depends on the p largest characteristic roots of $\mu\Sigma^{-1}\mu'$. Let A be an invariant acceptance region in the space of (Y, S) for the above testing problem. Note that under the invariance condition A has to be centrally symmetric. Das Gupta, Anderson and Mudholkar (1964) proved that if A is convex in each row of Y (when other rows and S are held fixed) then the power function based on this acceptance region increases with the p largest characteristic roots (which are nonnegative) of $\mu'\Sigma^{-1}\mu$. It was shown that the above conditions of invariance and convexity are met by acceptance regions corresponding to Roy's maximum root test, Lawley-Hotelling trace test, likelihood ratio test and Pillai's trace test. Eaton and Perlman (1974) further showed that if A is convex (which is true for the first two procedures listed above) then the power function is a Schur convex function of the square roots of the p largest characteristic roots of $\mu'\Sigma^{-1}\mu$. This provides a useful comparison among the alternatives. For both of these results the moving set inequality plays a crucial role in the proofs.

4.3. Association of absolute values. Let $\tilde{X}, \tilde{Z}, \tilde{U}$ be k -vector random variables. Suppose the k components of \tilde{Z} are independent each having a symmetric unimodal distribution. Clearly \tilde{Z} has an axial modal

distribution (see 3.2). Further, suppose that \underline{Z} and \underline{U} are independent, where the components of \underline{U} are not necessarily independent. If

$$(4.7) \quad \underline{X} = \underline{Z} + \underline{U} ,$$

one can describe this linear model as "contaminated independence" where \underline{U} is the contaminant. We show that if the distribution $|\underline{U}|$ is associated then so is that of $|\underline{X}|$. In view of the fact that $(B) \Rightarrow (A)$ (see the introduction of this section) it suffices to show that the covariance of the indicators of the sets in R^k which are nonincreasing in absolute values have nonnegative covariance under the distribution of \underline{X} . Let I_1 and I_2 be indicators of such sets. Now

$$(4.8) \quad \begin{aligned} \text{Cov}[I_1(\underline{X}), I_2(\underline{X})] &= \text{Cov}\{E[I_1(\underline{X})|\underline{U}], E[I_2(\underline{X})|\underline{U}]\} \\ &\quad + E\{\text{Cov}[I_1(\underline{X}), I_2(\underline{X})|\underline{U}]\} . \end{aligned}$$

Now I_1, I_2 are "axially modal" sets and Theorem 2 shows that $E[I_1(\underline{X})|\underline{U}]$ and $E[I_2(\underline{X})|\underline{U}]$ are non-increasing in $|\underline{U}|$. Since $|\underline{U}|$ is associated, we see that the first term on the right side of (4.8) is nonnegative. Given \underline{U} , the vector \underline{X} has independent components and hence $|\underline{X}|$ is associated. Therefore $\text{Cov}[I_1(\underline{X}), I_2(\underline{X})|\underline{U}]$ is non-negative. It follows that the second term on the right side of (4.8) is also non-negative. Thus $|\underline{X}|$ is associated whenever $|\underline{U}|$ is associated.

As an application suppose that \tilde{X} has a multinormal distribution with $E(X_i) = c\lambda_i$, $\text{Var}(X_i) \geq \lambda_i^2$ and $\text{Cov}(X_i, X_j) = \lambda_i \lambda_j$, $i \neq j$. Then it is easy to show that \tilde{X} satisfies the model (4.7) with $\tilde{U} = (\lambda_1 V, \dots, \lambda_k V)$ where V is a real random variable. Since $|\tilde{U}|$ is trivially associated, it follows that $|\tilde{X}|$ is associated. Note that the contaminating variable \tilde{U} need not have a zero mean. These positive dependence properties have applications to χ^2 and multivariate t distributions; see [22] for details.

4.4 Schur convexity and majorization in the parameter space. In many applications the observations may be independent but the underlying parameter may take different values. In such cases, it is important to study the effect of this departure from homogeneity on statistics which are permutation invariant. If X_1, \dots, X_n are independent Poisson random variables with parameters $\lambda_1, \dots, \lambda_n$, respectively, then the distribution of $\sum X_i$ is the same when λ_i take different values, provided $\sum \lambda_i$ is kept fixed. However the situation is not the same when X_1, \dots, X_n are Bernoulli random variables with parameters p_1, \dots, p_n . The distribution of $Y = \sum X_i$ is clearly invariant under permutations of p_1, \dots, p_n . Hoeffding (1956) showed that a "tail probability"

$$(4.9) \quad h(p_1, \dots, p_n) = P[Y > y | p_1, \dots, p_n], \quad y \geq [\lambda + 2], \quad \lambda = \sum p_i,$$

is maximized when p_i 's are all equal to λ/n . (Here $[t]$ represents the largest integer less than or equal to t .) This result seems

natural because, for fixed $\lambda = \sum p_i$, $\text{Var}(Y) = \lambda - \sum p_i^2$ becomes maximum when p_i 's are all equal. Further, this expression for the variance can be easily seen to be Schur concave. This leads to the conjecture that $h(p_1, \dots, p_n)$ defined by (4.8) would be Schur concave. Indeed, Gleser (1975) proved this result. The study of the distribution of Y is important and can be applied to derive monotone convergence of the binomial probabilities. (See for example Jogdeo and Samuels (1968)).

There are several other distributions which arise as multivariate analogues and have vector valued parameters. Marshall and Olkin (1974) cite several "non-central" distribution functions (chi-square, t , F) which are Schur concave in their parameters. Since all these distributions are used in applications, such properties are of great value while finding probabilities of various regions under such distributions.

BIBLIOGRAPHY

- [1] Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170-176.
- [2] Birnbaum, Z.W. (1948). On random variables with comparable peakedness. Ann. Math. Statist., 19, 76-81.
- [3] Brascamp, H.J. and Lieb, E.H. (1975). Some inequalities for Gaussian measures, in "Functional Integration and Its Applications", (A.M. Arthurs, Ed.) Clarendon Press, Oxford.
- [4] Conlon, J.C., Leon, R., Proschan, F. and Sethuraman, J. (1977). G-ordered functions with applications in statistics. Florida State Univ., Dept. of Statistics, Technical Report 73.
- [5] Converse, G. (1977). Sets with a mode. Israel J. Math., 27, 124-130.
- [6] Das Gupta, S. (1976). A generalization of Anderson's theorem on unimodal functions. Proc. Amer. Math. Soc., 60, 85-91.
- [7] Das Gupta, S. Anderson, T.W. and Mudholkar, G.S. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. Ann. Math. Statist., 35, 200-205.
- [8] Dharmadhikari, S.W. and Jogdeo, K. (1973). A characterization of convexity and central symmetry for planar polygonal sets. Israel J. Math., 15, 356-366.
- [9] Dharmadhikari, S.W. and Jogdeo, K. (1974). Convolutions of α -modal distributions. Z. Wahrscheinlichkeitsth., 30, 203-208.
- [10] Dharmadhikari, S.W. and Jogdeo, K. (1976). Multivariate Unimodality. Ann. Statist., 4, 607-613.
- [11] Dharmadhikari, S.W. and Jogdeo, K. (1976). On characterizations of unimodality of discrete distributions. Ann. Inst. Statist. Math., 28, 9018.
- [12] Eaton, M.L. and Perlman, M.D. (1974). A monotonicity property of the power functions of some invariant tests for Manova. Ann. Statist., 2, 1022-1028.
- [13] Eaton, M.L. and Perlman, M.D. (1977). Reflection groups, generalized Schur functions and the geometry of majorization. Ann. Prob., 6, 829-860.
- [14] Esary, J., Proschan, F. and Walkup, W. (1967). Associated random variables with applications. Ann. Math. Statist., 38, 1466-1474.

- [15] Fáy, I. and Rédei, L. (1950). Der zentral symmetrische Kern und die zentral symmetrische Hülle von konvexen Körpern. Math. Ann., 122, 205-220.
- [16] Ghosh, P. (1974). On generalized unimodality. Comm. Statist., 3, 567-580.
- [17] Gleser, L.J. (1975). On the distribution of the number of successes in independent trials. Ann. Prob., 3, 182-188.
- [18] Gnedenko, B.V. and Kolmogorov, A.N., (1968). Limit Distributions for Sums of Independent Random Variables, Addison-Wesley, Cambridge, Massachusetts.
- [19] Hodges, J.L. and Lehmann, E.L. (1954). Matching in paired comparisons. Ann. Math. Statist., 25, 787-791.
- [20] Ibragimov, I.A. (1956). On the composition of unimodal distributions. Theor. Prob. Appl., 1, 255-266.
- [21] Jogdeo, K. (1970). A simple proof of an inequality for multivariate normal probabilities of rectangles. Ann. Math. Statist., 41, 1357-1359.
- [22] Jogdeo, K. (1977). Association and probability inequalities. Ann. Statist., 3, 495-504.
- [23] Jogdeo, K. and Samuels, S.M. (1968). Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. Ann. Math. Statist., 39, 1191-1195.
- [24] Kantor, M. (1977). Unimodality and dominance for symmetric random vectors. Trans. Amer. Math. Soc., 229, 65-85.
- [25] Keilson, J. and Gerber, H. (1971). Some results for discrete unimodality. J. Amer. Statist. Assoc., 66, 386-389.
- [26] Khintchine, A.Y. (1938). On unimodal distributions. Izv. Nauchno-Isled. Inst. Math. Mech. Tomsk. Gos. Univ., 2, 1-7.
- [27] Lehmann, E.L. (1966). Some concepts of dependence. Ann. Math. Statist., 37, 1137-1153.
- [28] Marshall, A.W., and Olkin, I. (1974). Majorization in multivariate distributions. Ann. Statist., 2, 1189-1200.
- [29] Marshall, A.W., and Olkin, I. (1979). Inequalities: The Theory of Majorization with Applications to Combinatorics, Probability, Statistics, and Matrix Theory. Aca. Press, New York (to appear).
- [30] Mudholkar, G.S. (1966). The integral of an invariant unimodal function over an invariant convex set--an inequality and applications. Proc. Amer. Math. Soc., 17, 1327-1333.

- [31] Olshen, R.A. and Savage, L.J. (1970). Generalized unimodality. J. Appl. Prob., 7, 21-34.
- [32] Pitt, L.D. (1977). A Gaussian correlation inequality for symmetric convex sets. Ann. Prob., 6, 470-474.
- [33] Prékopa, A. (1973). On logarithmic concave measures and functions. Acta. Sci. Math., 34, 335-343.
- [34] Sherman, S. (1955). A theorem on convex sets with applications. Ann. Math. Statist., 26, 763-766.
- [35] Sidak, Z. (1968). On multivariate normal probabilities of rectangles: their dependence on correlations. Ann. Math. Statist., 39, 1425-1434.
- [36] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. Bell System Tech. J., 41, 463-501.
- [37] Wells, D.R. (1978). A result in multivariate unimodality. Ann. Statist., 6, 926-931.
- [38] Wintner, A. (1938). Asymptotic Distributions and Infinite Convolutions. Edwards Brothers, Ann Arbor, Michigan.
- [39] Wolfe, S.J. (1975). On the unimodality of spherically symmetric stable distribution functions. J. Multiv. Anal. 5, 236-242.