Measures of location and asymmetry in the plane

David King Blough

Iowa State University
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame. If copyrighted materials were deleted you will find a target note listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. if necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International
300 N. ZEEB RD., ANN ARBOR, MI 481
Elough, David King

MEASURES OF LOCATION AND ASYMMETRY IN THE PLANE

Iowa State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Ph.D. 1982
Measures of location and asymmetry in the plane

by

David King Blough

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:
Signature was redacted for privacy.
In Charge of Major Work
Signature was redacted for privacy.
For the Major Department
Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa

1982
TABLE OF CONTENTS

I. INTRODUCTION 1
   A. Review of Univariate Case 1
   B. Statement of Problem and Summary of Results 13

II. CONSTRUCTION VIA THE STANDARD AND MARGINAL ORDERINGS 16
   A. Definitions 16
   B. Construction of the Location Region 23
      1. Properties 52
      2. Estimation 54
      3. Examples 66

III. CONSTRUCTION VIA THE CONDITIONAL ORDERING 72
   A. Definitions 72
   B. Construction of the Location Set 74
      1. Estimation 86
      2. Examples 93
   C. Loss Function Approach 96

IV. APPLICATIONS 122
   A. Measures of Asymmetry 123
   B. Ordering of Distributions 136

V. CONCLUSION 151
   A. Multivariate Generalization 151
   B. Summary 153
VI. REFERENCES 155
VII. ACKNOWLEDGMENTS 157
VIII. APPENDIX A 158
IX. APPENDIX B 165
A. Review of the Univariate Loss Function Approach 165
I. INTRODUCTION

A. Review of Univariate Case

There are a number of ways to characterize and describe univariate probability distributions. Among these are specification of the cumulative distribution function, which in the case of absolute continuity is equivalent to specifying the probability density function. Alternatively, one can specify the characteristic function of the distribution. While these functions serve to describe a distribution in its entirety, one is often interested only in a few summary values which describe some important characteristics of a distribution. These parameters fall into such categories as measures of location (i.e., where the distribution is located on the line) and measures of dispersion (i.e., what is the "spread" of the distribution). Much of classical probability and statistics deals with properties of these parameters as well as their estimation.

In as much as this paper will deal exclusively with measures of location, it is appropriate here to review some of the important results which have been obtained regarding such measures for the univariate case. Bickel and Lehmann (1975) give the following definition of a location parameter:
Definition 1.1: (measure of location/location parameter)

Let $\mathcal{F}$ be a suitably large class of univariate probability distribution functions. Define the functional $\mu(\cdot): \mathcal{F} \to \mathbb{R}$

Let $F \in \mathcal{F}$ and suppose $X$ is a random variable with distribution function $F$. Then $\mu(F)$ (or we will often use the notation $\mu(X)$) is a measure of location (or a location parameter) for $F$ (or $X$) if $\mu(F)$ satisfies the following three "axioms of location":

(1) Suppose $F \in \mathcal{F}$, $G \in \mathcal{F}$ and $F(x) \geq G(x), \forall x \in \mathbb{R}$. Then $\mu(F) \leq \mu(G)$

(2) $\mu(aX+b) = a\mu(X)+b, \forall a>0, b \in \mathbb{R}$

(3) $\mu(-X) = -\mu(X)$

These are generally accepted conditions any location parameter should satisfy. The first deals with the concept of stochastic order; a random variable $X$ (with distribution function $F$) is said to be stochastically smaller than a random variable $Y$ (with distribution function $G$) if $X$ tends to take on smaller values than $Y$. More precisely, this means $P(X \leq x) \geq P(Y \leq x), \forall x \in \mathbb{R},$ or equivalently $F(x) \geq G(x), \forall x \in \mathbb{R}$. In this case, it is natural to expect any measure of location for $F$ to be less than or equal to a corresponding measure of location for $G$.

The last two axioms describe the behavior of $\mu(\cdot)$ under
certain types of linear transformations.

Axiom (u2) can be decomposed into

(u2.1) $\mu(ax) = a\mu(x), \forall a > 0$

and

(u2.2) $\mu(x+b) = \mu(x) + b, \forall b \in \mathbb{R}$

Thus, (u2.1) states that $\mu(\cdot)$ is appropriately transformed under a change of scale in the sense that changing the scale of the distribution of $X$ and then determining $\mu(ax)$ yields the same result as when first finding $\mu(x)$ and then transforming to the new scale $a\mu(x)$. Similarly, (u2.2) states that $\mu(\cdot)$ is appropriately transformed under a translation.

Axiom (u3) states that $\mu(\cdot)$ is appropriately transformed under a reflection of the line about the origin. In addition to proving some simple results implied by these axioms, Bickel and Lehmann (1975) also show that these axioms are mutually independent; that is, any one cannot be deduced from the others.

Examples of location parameters in the univariate case are

(i) the mean: $\mu_1(F) = \int xdF(x)$

(ii) the median: $\mu_2(F) = F^{-1}\frac{1}{2}$

where $F^{-1}(y) = \sup\{x : F(x) \leq y\}$
(iii) the average of the first and third quartiles:
\[ \mu_3 = \frac{1}{2}[F^{-1}(\frac{1}{4}) + F^{-1}(\frac{3}{4})] \]

In fact, in the same paper, three broad classes of location parameters are defined and their asymptotic relative efficiencies are discussed. These classes are:

(i) \{\mu(F): \mu(F) = \frac{1}{2}[F^{-1}(u) + F^{-1}(1-u)], 0 < u \leq \frac{1}{2} \}

or, more generally,
\{\mu(F): \mu(F) = \int_0^1 F^{-1}(t)dK(t) \}

where K is any distribution on (0,1) which is symmetric about the point \( \frac{1}{2} \).

(ii) Take \( \mu(F) \) equal to that value of \( \theta \) which maximizes the quantity
\[ \int_0^\infty [L[F(x+\theta) - F(-x+\theta)] - L[F(x) - F(-x)]]dx \]

where L is an increasing, bounded, convex function on [0,1] with L(0) = 0.

(iii) Take \( \mu(F) \) equal to the value of \( \theta \) which minimizes the quantity
\[ \int \rho(x-\theta)dF(x) \]

where \( \rho \) is a positive, even, convex, twice differentiable function.

It is important to note that if \( X \) is symmetric about a point \( \theta \), then all of these classes reduce to the single point \( \mu(X) = \theta \). To be more specific, we have the following
Definition 1.2: (symmetric random variable/symmetric distribution function)

A random variable $X$ is symmetric about a point $\theta \in \mathbb{R}$ if

$X - \theta \sim -X + \theta,$

or equivalently, if $X - 2\theta \sim -X$. In terms of the distribution function, we say $F$ is symmetric about $\theta$ if

$F(x + \theta) = 1 - F(-x + \theta), \forall x \in \mathbb{R}.$

In case $X$ is not symmetric, many points may serve as location parameters. Doksum (1975) characterizes this set of location parameters and establishes some important results concerning it. He uses three different methods to construct a set of points, any one of which would serve as a "reasonable" measure of where a given distribution is located on the line. His first approach is as follows:

given a random variable $X$ with continuous, increasing distribution function $F$, define the support of $F$ by $S(F) = \{x : 0 < F(x) < 1\}$. $F$ will be approximated from above and below by two symmetric distribution functions. The points of symmetry $\bar{\theta}$ and $\hat{\theta}$ of these two distributions will then be used to define the location interval $[\hat{\theta}, \bar{\theta}]$. To this end, let $\overline{F}$ denote the distribution function of $-X$, i.e.,

$\overline{F}(x) = 1 - F(-x), \forall x \text{ s.t. } -x \in S(F).$
Let $H$ be the function which bisects all horizontal lines connecting $F$ and $\overline{F}$ (see Figure 1.1).

Then $H$ is a continuous, increasing distribution function which is symmetric about zero in the sense that if $Y$ has distribution function $H$, then so does $-Y$. $H$ can implicitly be defined as the distribution function whose inverse is given by

$$H^{-1}(u) = \frac{1}{2}[F^{-1}(u) + \overline{F}^{-1}(u)], \ 0 < u < 1.$$  

It should be noted that if $F$ is symmetric about 0, then $F$, $\overline{F}$ and $H$ all have the same shape and

$$H(x-\theta) = F(x), \ \forall x \in \mathbb{R}.$$  

When $F$ is not symmetric, Doksum approximates it from above and below as follows: define the lower point of symmetry of $F$ as

$$\theta_F^- = \sup\{\theta : H(x-\theta) \geq F(x), \ \forall x\}$$

Define the upper point of symmetry of $F$ as

$$\theta_F^+ = \inf\{\theta : H(x-\theta) \leq F(x), \ \forall x\}$$

If there is no $\theta$ s.t. $H(x-\theta) \geq F(x), \ \forall x$, take $\theta_F^+ = -\infty$. Similarly, let $\theta_F^- = +\infty$ if there is no $\theta$ s.t. $H(x-\theta) \leq F(x), \ \forall x$.

Define $[\theta_F^-, \theta_F^+]$ to be the location interval for $F$.

Doksum (1975) proves the following useful result:
Figure 1.1. Construction of a distribution function $H$, symmetric about zero.
Proposition 1.1: Let $\theta^*_F(x) = x - H^{-1}(F(x)), x \in S(F)$.

Then

$$\theta^-_F = \inf\{\theta^*_F(x) : x \in S(F)\}, \quad \theta^+_F = \sup\{\theta^*_F(x) : x \in S(F)\}$$

Proof: $H(x - \theta^*_F(x)) = H(H^{-1}(F(x))) = F(x)$.

$\inf \theta^*_F(x)$ is the largest $\theta$ s.t. $H(x - \theta) \geq F(x), \forall x$ and $\sup \theta^*_F(x)$ is the smallest $\theta$ s.t. $H(x - \theta) \leq F(x), \forall x$.

The result follows.

Note that $\theta^*_F(F^{-1}(u)) = F^{-1}(u) - H^{-1}(u)$

$$= F^{-1}(u) - \frac{1}{2}[F^{-1}(u) + F^{-1}(u)]$$

$$= \frac{1}{2}[F^{-1}(u) - F^{-1}(u)].$$

Recall now that $F(x) = 1 - F(-x), \forall x \text{ s.t. } -x \in S(F)$.

So if $u = F(x)$

$\Rightarrow 1 - u = F(-x)$

$\Rightarrow x = F^{-1}(u) = -F^{-1}(1-u)$

Hence, we can define $m^-_F(u) = \theta^*_F(F^{-1}(u)) = \frac{1}{2}[F^{-1}(u) + F^{-1}(1-u)]$.

Note that $m^-_F(u)$ is symmetric about $u = \frac{1}{2}$. Thus, the above proposition can be rephrased as

$$\theta^-_F = \inf\{m^-_F(u) : 0 < u \leq \frac{1}{2}\} \text{ and } \theta^+_F = \sup\{m^-_F(u) : 0 < u \leq \frac{1}{2}\}.$$ 

It should be noted that when $F$ cannot be assumed to be continuous and increasing, Doksum uses the following
modification of the above construction:

Let $F^{-1}(u) = \inf \{x: F(x) \geq u, x \in S(F)\}$

and

$F^{-1}(u) = \sup \{x: F(x) \leq u, x \in S(F)\}, \ 0 < u < 1.$

Let $H$ be the distribution function whose inverse is given by

$H^{-1}(u) = \frac{1}{2}[F^{-1}(u) + \overline{F}^{-1}(u)], \quad \text{when } u \in (0,1)$

is a point of continuity of either $F^{-1}$ or $\overline{F}^{-1}$. Thus, $H$ is defined at all points except where $F$ and $\overline{F}$ are both constant. At these points, $H$ is defined by right continuity. $H$ is symmetric about zero and we take

$\delta_F = \inf \{x - H^{-1}(F(x))\} \quad x \in S(F)$

and

$\overline{\delta}_F = \sup \{x - H^{-1}(F(x))\}. \quad x \in S(F)$

The question arises why not use some other symmetric distribution (for example, a $N(0,1)$ distribution) to approximate $F$? Doksum shows that using $H$, as constructed above, yields a certain optimality property that other distributions do not possess. It is natural to want the location interval to be as small as possible. That is, if $G$ is symmetric, and if $G$ serves as the approximating distribution for $F$, it is natural to minimize the horizontal distance between $F$ and $G$. Thus, $G$ should minimize

$\Delta(F,G) \overset{\text{def}}{=} \sup_u |F^{-1}(u) - G^{-1}(u)|$. Doksum proves the following result.
Theorem 1.1: In the class of distribution functions $G$ symmetric about $m_F = \text{median of } F$, the distribution $H(x-m_F)$ minimizes

$$
\Delta(F,G) = \sup_u |F^{-1}(u)-G^{-1}(u)|.
$$

Proof: Without loss of generality, assume $m_F = 0$, then $G^{-1}(u) = -G^{-1}(1-u)$. $F^{-1}(u) = -F^{-1}(1-u)$, so that

$$
\Delta(F,G) = \sup_u |F^{-1}(u)-G^{-1}(u)|
$$

$$
= \sup_u |F^{-1}(u)-G^{-1}(u)|
$$

Thus, $\Delta(F,G) = \frac{1}{2}[\Delta(F,G) + \Delta(F,G)]$

$$
\geq \frac{1}{2} \sup_u (|F^{-1}(u)-G^{-1}(u)| + |F^{-1}(u)-G^{-1}(u)|)
$$

$$
\geq \frac{1}{2} \sup_u |F^{-1}(u) - F^{-1}(u) - G^{-1}(u) + G^{-1}(u)|
$$

$$
= \Delta(F,H)
$$

which is the desired result.

The second approach Doksum uses is to consider the set of location parameters for a given distribution $F$; i.e., the points which satisfy the axioms given by Bickel and Lehmann. More specifically, Doksum defines a location parameter to be the real-valued location function $\theta$ (defined on the class $\mathcal{F}$ of distribution functions) satisfying location
axioms (u1)-(u3) for all F with finite support (i.e., with \( S(F) \subset [a,b], -\infty < a \leq b < \infty \)). Let \( \mathcal{D} \) denote the set of all real-valued location parameters \( \theta \). Then, for a given distribution function F, he defines

\[
L_F = \{ \theta_F: \theta \in \mathcal{D}, \theta_F \text{ exists} \}
\]

to be the location set for F.

Doksum's third approach involves the use of the function of symmetry. In the case of a symmetric distribution, there exists a unique constant \( \theta \) s.t. \( X + \theta \). This property can be generalized to any distribution function F (not necessarily symmetric) by letting \( \theta \) be a function for each F. Note that

\[
X - X + 2\theta
\]

\[
F(x) = \overline{F}(x-2\theta), \forall x \leftrightarrow F(x+2\theta) = \overline{F}(x), \forall x
\]

In general, define \( \theta_F(x) \) such that

\[
\overline{F}(x) = F(x+2\theta_F(x)), \text{ or more precisely,}
\]

\[
2\theta_F(x) = \sup\{ \theta: F(x) \leq \overline{F}(x-2\theta) \}, \quad x \in S(F)
\]  \hfill (1.1)

So, \( 2\theta_F(x) \) is the horizontal distance between F and \( \overline{F} \) at x. We can solve Equation (1.1) for \( \theta_F(x) \) to obtain

\[
\theta_F(x) = \frac{1}{2}[x - \overline{F}^{-1}(F(x))], \quad x \in S(F)
\]

When F is continuous, \( F(X) \sim U(0,1) \) so that \( X - 2\theta_F(X) = \)
$F^{-1}(F(X))$ has distribution function $F$. Thus, $X - 2\theta_F(X) \sim -X$.

Doksum defines $\theta_F(x)$ as a symmetry function in the sense that $X$, when shifted to the left by the amount $2\theta_F(X)$ has the same distribution as $-X$. The range of $\theta_F(\cdot)$ can be used as a location set for $F$: $R(\theta_F) = \{\theta_F(x): x \in S(F)\}$. Doksum establishes the following results concerning the symmetry function $\theta_F(x)$:

**Theorem 1.2:** Suppose $F$ is continuous. If $A$ is any function such that $X - 2A(X) \sim -X$ and $x - 2A(x)$ is nondecreasing for almost all $x(F)$, then $A(x) = \theta_F(x)$ for almost all $x(F)$.

**Theorem 1.3:** (i) For each $x$, $\theta_F(x)$ satisfies axiom (ul),
(ii) If $a > 0$, then for all $b \in \mathbb{R}$,
$$\theta_{ax+b}(ax+b) = a\theta_x(x) + b$$
(iii) If $F$ is increasing on $S(F)$, then,
$$\theta_{-X}(F^{-1}(F(x))) = -\theta_x(x)$$
(iv) Let $m_F$ denote the median of $F$. If $F$ is increasing then $\theta_F(m_F) = m_F$. If $E(X) < \infty$, then $E(\theta_F(X)) = E(X)$.

**Theorem 1.4:** If $X - \theta \sim -X + \theta$, then $P(\theta_F(x) = \theta) = 1$. If $\theta_F(x) = \theta$, $\forall x \in S(F)$, then $X - \theta \sim -X + \theta$. If $F$ is strictly increasing on $S(F)$, then $X - \theta \sim -X + \theta$ if and only if $\theta_F(x) = \theta$ for $x \in S(F)$.
Probably the most significant result of Doksum's work is the equivalence of the three methods of constructing a location set. He proves the following:

**Theorem 1.5:**

1. If $F$ is increasing and continuous on its support, then the closure of the location set $L_F$ equals the location interval $[\bar{\theta}_F, \bar{\theta}_F]$.

2. If $F$ is increasing and continuous on its support, then the closure of the range of $\theta_F(\cdot)$ equals the location interval $[\bar{\theta}_F, \bar{\theta}_F]$.

**B. Statement of Problem and Summary of Results**

It is the purpose of this paper to extend Doksum's results to multivariate distribution functions. The bivariate case will be emphasized due to the ease of geometric interpretation. Multivariate generalizations will be discussed briefly in the concluding chapter.

Two location regions will be obtained using two stochastic orderings. The standard ordering is used in Chapter II to construct a location region in the plane. By three approaches completely analogous to Doksum's methods, a well-defined closed convex location set is obtained, any point of which would be an appropriate measure of where the given distribution is located in the plane. It will then be shown that this region depends on the distribution function
only through its marginal distributions so that by using a weaker type of stochastic ordering based solely on the marginals yields exactly the same location region. Having thus defined the location set, its estimation is considered. An algorithm, based on a random sample from a bivariate distribution, is developed and implemented by way of a computer program given in the appendix. This program is applied to computer-generated samples from some common bivariate distributions and the resulting plots of the estimated location region are presented. The fact that this region is a consistent estimator of the true location region is also established.

Chapter III similarly uses Doksum's three approaches to construct a location region, this time using a more restrictive type of stochastic ordering defined through the conditional distributions of the given bivariate distribution function. This type of ordering yields, in general, a larger location region than that obtained by using the standard ordering. Also, results on estimation, similar to those obtained in Chapter II are presented. Finally, the exposition of a class of location parameters is undertaken. This class is obtained by minimizing the value of a double integral, similar in spirit to Bayesian estimation of a location parameter.
Chapter IV discusses some applications of the location region. In addition to the primary goal of obtaining measures of location for asymmetric bivariate distributions, it is apparent that the size of the location set can be used to characterize the degree of asymmetry of the distribution. Using some results from convex analysis, an intuitively appealing measure of asymmetry is obtained. This chapter concludes with a discussion of the use of location parameters to order distributions. This is done in general as well as for the special case of ordering components of a random vector.

The final chapter presents a multivariate generalization of the results of Chapters II-IV and concludes the paper with a summary.
II. CONSTRUCTION VIA THE STANDARD AND MARGINAL ORDERINGS

A. Definitions

It is necessary to begin with some definitions and notation. First, some important groups of transformations on $\mathbb{R}^2$ will be considered.

**Definition 2.1:** (transformations on the sample space $\mathbb{R}^2$)

(a) Let $\mathcal{E}$ denote the collection of all rigid motion transformations; i.e., $\mathcal{E}$ is the Euclidean group (a subgroup of the group of all nonsingular affine transformations) on $\mathbb{R}^2$. Thus,

$$\mathcal{E} = \{(H,k): H \text{ is an orthogonal transformation of } \mathbb{R}^2 \text{ into } \mathbb{R}^2, \text{ and } k \in \mathbb{R}^2 \text{ is constant}\}.$$  

Any element of $\mathcal{E}$ defines the mapping

$$x \mapsto Hx + k, \forall x \in \mathbb{R}^2.$$  

Geometrically, $\mathcal{E}$ is the collection of all rotations and reflections of the plane, followed by a translation.

(b) Let $\mathcal{P}$ denote the collection of all symmetric positive-definite linear transformations. Geometrically, $\mathcal{P}$ is the collection of all transformations which stretch or shrink the plane without inverting the axes.
Note: \( \mathcal{O} \) is not a group; while it is closed under inverses, it is not closed under products.

**Notation:** Orthogonal transformations of \( \mathbb{R}^2 \) will be denoted by \( H \). Any orthogonal transformation of the plane is either a rotation of the plane about the origin or a reflection of the plane about a line through the origin.

Rotations of the plane through a (counterclockwise) angle \( \alpha \) will be denoted by \( R_{\alpha} \). We have

\[
R_{\alpha} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad 0 < \alpha < 2\pi.
\]

Reflections of the plane about a line through the origin making an angle \( \theta = \frac{\alpha}{2} \) with the \( x_1 \)-axis will be denoted by \( T_{\theta} \). We have

\[
T = \begin{pmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{pmatrix}, \quad 0 < \alpha < 2\pi.
\]

Symmetric positive-definite transformations of the plane will be denoted by \( S \).

**Results from linear algebra:**

1. For any reflection \( T_{\alpha} \), we have

\[
T_{\alpha} = R_{\alpha} T_0, \quad \text{where } T_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
\]

and

\[
R_{\alpha} T_0 = T_0 R_{-\alpha}.
\]
(2) For any symmetric positive-definite transformation $S$, we have

$$S = H'DH,$$

where $H$ is orthogonal and $D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$ for some $d_1 > 0$, $d_2 > 0$. The collection of all positive definite diagonal matrices $\mathcal{D}^*$ is a group.

(3) Any nonsingular linear transformation $A$ can be expressed as $A = SH$ where $S$ is symmetric positive-definite and $H$ is orthogonal.

(4) An affine transformation $A$ is a transformation of the form $Ax = Tx + c$; i.e., $A$ is a linear transformation followed by a translation. The set of all nonsingular affine transformations (i.e., where $T$ is nonsingular) constitutes a group, the affine group $A_2(\mathbb{R})$.

(5) Any affine transformation carries the center of mass of a set to the center of mass of the transformed set.

(6) Any affine transformation carries parallel sets to parallel sets.

**Ordering points in the plane:** We will define

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Y$$

if $x_1 \leq y_1$ and $x_2 \leq y_2$. Hence, "$\leq$" on $\mathbb{R}^2$ is a partial ordering of $\mathbb{R}^2$, but it is not a total ordering.
**Definition 2.2:** (symmetry of distribution functions)

If the random vector $X$ has bivariate distribution function $F$, then $X$ and $F$ are said to be symmetric about the point

$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$

if

$$R_{\pi}(X-\theta) \sim X-\theta$$

This is equivalent to

$$\begin{pmatrix} X_1 - \theta_1 \\ X_2 - \theta_2 \end{pmatrix} \sim \begin{pmatrix} -X_1 + \theta_1 \\ -X_2 + \theta_2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} X_1 - 2\theta_1 \\ X_2 - 2\theta_2 \end{pmatrix} \sim \begin{pmatrix} -X_1 \\ -X_2 \end{pmatrix}.$$  

If $X$ has distribution function $F$, then $X$ and $F$ are said to be symmetric about the line $L: X_2 = ax_1 + b$ if

$$T_\beta (X - \begin{pmatrix} 0 \\ b \end{pmatrix}) \sim X - \begin{pmatrix} 0 \\ b \end{pmatrix},$$

where $\beta = 2 \arctan a$.

It is appropriate now to define a stochastic ordering for bivariate distributions. In the univariate case, the stochastic ordering defined in the preceding review of the univariate approach is most often used. In the bivariate case, there are a number of ways to order distributions. For the purposes of this paper, a number of relevant orderings will be considered. To this end, let
\( X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \) be a random vector with continuous, strictly increasing distribution function \( F \) and let \( Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \) be a random vector with continuous, strictly increasing distribution function \( G \). Denote the respective marginal distribution functions by

\[
F_{X_1} (x_1) = F(x_1, +\infty),
\]

\[
F_{X_2} (x_2) = F(+\infty, x_2),
\]

\[
G_{Y_1} (y_1) = G(y_1, +\infty),
\]

\[
G_{Y_2} (y_2) = G(+\infty, y_2),
\]

Denote the respective conditional distribution functions by

\[
F_{X_1|X_2} (x_1|x_2), \quad F_{X_2|X_1} (x_2|x_1),
\]

\[
G_{Y_1|Y_2} (y_1|y_2), \quad G_{Y_2|Y_1} (y_2|y_1).
\]

**Definition 2.3(a):** (standard stochastic ordering)

We say \( X \) is stochastically smaller than \( Y \) if

\[
F(x_1, x_2) \geq G(x_1, x_2) \quad \forall x_1, x_2
\]

This will be termed the standard stochastic ordering.
(b) (marginal stochastic ordering)

We say $X$ is marginally smaller than $Y$ in the $x_1$-direction iff $F_{X_1} (x_1) \geq G_{Y_1} (x_1), \forall x_1$. We say $X$ is marginally smaller than $Y$ in the $x_2$-direction iff $F_{X_2} (x_2) \geq G_{Y_2} (x_2), \forall x_2$. We say $X$ is marginally smaller than $Y$ iff $X$ is marginally smaller than $Y$ in both the $x_1$-direction and the $x_2$-direction.

(c) (conditional stochastic ordering)

We say $X$ is conditionally smaller than $Y$ in the $x_1$-direction iff $F_{X_1 | X_2} (x_1 | x_2) > G_{Y_1 | Y_2} (x_1 | x_2)$ and $F_{X_2} (x_2) \geq G_{Y_2} (x_2), \forall x_1, x_2$. We say $X$ is conditionally smaller than $Y$ in the $x_2$-direction iff $F_{X_2 | X_1} (x_2 | x_1) > G_{Y_2 | Y_1} (x_2 | x_1)$ and $F_{X_1} (x_1) \geq G_{Y_1} (x_1), \forall x_1, x_2$. We say $X$ is conditionally smaller than $Y$ iff $X$ is conditionally smaller than $Y$ in both the $x_1$-direction and the $x_2$-direction.

(d) (strong conditional stochastic ordering)

We say $X$ is strongly conditionally smaller than $Y$ in the $x_1$-direction iff $F_{X_1 | X_2} (x_1 | x_2) > G_{Y_1 | Y_2} (x_1 | x_2)$, $\forall x_1$, $\forall x_2 \leq y_2$, and $F_{X_2} (x_2) > G_{Y_2} (x_2), \forall x_2$. We say $X$ is strongly conditionally smaller than $Y$ in the $x_2$-direction iff $F_{X_2 | X_1} (x_2 | x_1) > G_{Y_2 | Y_1} (x_2 | y_1)$, $\forall x_2$, $\forall x_1 \leq x_1$, and $F_{X_1} (x_1) > G_{Y_1} (x_1), \forall x_1$. We say $X$ is strongly conditionally smaller than $Y$ iff $X$ is strongly conditionally smaller than $Y$ in both the $x_1$-direction and the $x_2$-direction.
Remarks: (1) If $X_1$ and $X_2$ are independent and if $Y_1$ and $Y_2$ are independent, then any type of conditional ordering implies $X$ is marginally less than $Y$.

(2) If $X_1$ and $X_2$ are exchangeable, and if $Y_1$ and $Y_2$ are exchangeable, then marginally (conditionally) smaller in the $x_1$-direction is equivalent to marginally (conditionally) smaller in the $x_2$-direction in the sense that one implies the other. Also in this case, one-directional ordering implies both directions.

(3) The standard stochastic ordering implies the marginal ordering.

(4) We will say $X$ is marginally (conditionally, strongly conditionally) larger than $Y$ in a given direction iff $Y$ is marginally (conditionally, strongly conditionally) smaller than $X$ in that direction.

(5) If $F_{X_1|x_2}(x_1|x_2)$ is a decreasing function of $x_2$, then the conditional and strong conditional orderings are equivalent. For in this case, if $x_2 < y_2$,

$$F_{X_1|x_2}(x_1|x_2) \geq F_{X_1|x_2}(x_1|y_2) \geq G_{Y_1|Y_2}(x_1|y_2)$$

implying the strong conditional ordering. Similar arguments apply if $F_{X_2|x_1}(x_2|x_1)$ or $G_{Y_2|Y_1}(y_1|x_1)$ are decreasing in $x_1$, or if $G_{Y_1|Y_2}(y_1|y_2)$ is decreasing in $y_2$.

It should be noted that the standard and strongly conditional orderings given in (a) and (d) above are
found extensively in the literature. The marginal and conditional orderings given in (b) and (c) are special cases of (a) and (d) respectively, and have been found to be appropriate for the purpose of this paper.

B. Construction of the Location Region

Method I: (approximating $F$ above and below by symmetric distributions)

We now wish to construct a symmetric distribution function $H$, and use it to approximate $F$ from above and below. While it is possible to construct such a symmetric bivariate distribution function by methods somewhat analogous to those in the univariate case, this is not necessary. Rather than extend the univariate results to the bivariate case, it is possible to reduce the bivariate case to the univariate case. This will be done by way of the marginal distribution functions. It will be shown that the boundaries of the location region depend on the bivariate distribution function $F$ only through its marginals.

On the other hand, distribution functions are based on "infinite" rectangles with sides parallel to the coordinate axes. It is desirable that the location region be independent of the orientation of the axes, and so we must view
F from all possible directions; that is, through all possible rotations $R_\alpha$, $\alpha \in (0, 2\pi]$. The location region will then be obtained by averaging in a certain sense the regions obtained in each direction.

Let $S(F) = \{(x_1, x_2): 0 < F(x_1, x_2) < 1\}$ be the support of $F$. Assume $F$ is continuous and strictly increasing in each argument. Let $R_\alpha$ denote the linear transformation representing a rotation of the plane about the origin through a (counter-clockwise) angle $\alpha (0 < \alpha < 2\pi)$.

Let $F_\alpha$ denote the distribution function of $R_\alpha(X)$ and let $F_{\alpha X_1}$ and $F_{\alpha X_2}$ denote the respective marginal distributions of $X_1$ and $X_2$ where the vector $(X_1, X_2)$ has bivariate distribution functions $F_\alpha$. Let $F_{\alpha X_1}$ and $F_{\alpha X_2}$ denote the c.d.f.'s of $-X_1$ and $-X_2$, respectively. Then for each $\alpha \in (0, 2\pi]$.

\[
F_{\alpha X_1}(x_1) = 1 - F_{\alpha X_1}(-x_1), \quad -x_1 \in S(F_{\alpha X_1})
\]

\[
F_{\alpha X_2}(x_2) = 1 - F_{\alpha X_2}(-x_2), \quad -x_2 \in S(F_{\alpha X_2})
\]

Let $H_{\alpha X_1}$ be the function which bisects all horizontal lines connecting $F_{\alpha X_1}$ and $F_{\alpha X_1}$, and let $H_{\alpha X_2}$ be the function which bisects all lines connecting $F_{\alpha X_2}$ and $F_{\alpha X_2}$. Then $H_{\alpha X_1}$ and $H_{\alpha X_2}$ are univariate distribution functions which are
symmetric about zero. \( H_{\alpha X_1} \) and \( H_{\alpha X_2} \) can also be defined to be the distribution functions whose inverses are given, respectively, by

\[
H^{-1}_{\alpha X_1}(u) = \frac{1}{2} [F^{-1}_{\alpha X_1}(u) + F^{-1}_{\alpha X_1}(u)] \\
H^{-1}_{\alpha X_2}(u) = \frac{1}{2} [F^{-1}_{\alpha X_2}(u) + F^{-1}_{\alpha X_2}(u)], \quad 0 < u < 1.
\]

Define the bivariate c.d.f. \( H_{\alpha} \) by

\[
H_{\alpha}(x_1, x_2) = H_{\alpha X_1}(x_1)H_{\alpha Y_2}(x_2)
\]

(Actually, any symmetric bivariate distribution function with marginals \( H_{\alpha X_1} \) and \( H_{\alpha X_2} \) would be sufficient here.)

Now \( H_{\alpha} \) is symmetric about the origin in the sense that if \( (X_1, X_2) \) has distribution function \( H_{\alpha} \), then so does \( (-X_1, X_2) \).

If \( F_{\alpha} \) is symmetric about the point \( (\theta_1, \theta_2) = \theta \), then \( F_{\alpha X_1} \), \( F_{\alpha X_2} \), and \( H_{\alpha X_1} \) all have the same shape and \( H_{\alpha X_1}(x_1 - \theta_1) = F_{\alpha X_1}(x_1), \forall x_1 \). Similarly, \( F_{\alpha X_2} \), \( F_{\alpha Y_2} \), and \( H_{\alpha X_2} \) all have the same shape and \( H_{\alpha X_2}(x_2 - \theta_2) = F_{\alpha X_2}(x_2), \forall x_2 \). If \( F_{\alpha} \) is not symmetric, it can be approximated by using the standard stochastic ordering as follows:

Let

\[
L_{\alpha} = \{ \theta : H_{\alpha}(x-\theta) > F_{\alpha}(x), \forall x \} \\
U_{\alpha} = \{ \theta : H_{\alpha}(x-\theta) < F_{\alpha}(x), \forall x \}
\]
Let \( \theta^*_F = \sup_L \theta_F, \theta^{**}_F = \inf_U \theta_F \)
(using the component-wise ordering given before).

Let \( A_\alpha = \) the rectangular region defined by

\[
\{ x: \theta^*_F \alpha \leq x \leq \theta^{**}_F \alpha \}
\]

(see Figure 2.1a).

The location region for \( F \) is then given by

\[
A = \bigcap_{\alpha \in [0,2\pi]} R_\alpha(A_\alpha),
\]

It was noted that in using the distribution function \( H \) to approximate \( F_\alpha \) from above and below the only important constraint on \( H_\alpha \) was that it have the appropriate marginals, \( H_\alpha x_1 \) and \( H_\alpha x_2 \). This leads one to wonder whether the marginal ordering could have been used in the construction of the sets \( U_\alpha \) and \( L_\alpha \). The following proposition shows that indeed the rectangle \( A_\alpha \) is the same under either the standard stochastic ordering or the marginal stochastic ordering.

**Proposition 2.1:** Let \( F \) be a bivariate distribution which is continuous and strictly increasing in each argument.

Let \( M_\alpha = \{ \theta = (\theta_1, \theta_2): H_\alpha x_1 (x_1 - \theta_1) \geq F_\alpha x_1 (x_1), \forall x_1 \} \)
Figure 2.1a. Diagram of a typical location region $A_\alpha$ for the distribution function $F_\alpha$, for $\alpha \in (0,2\pi]$.
Let

\[ \mathcal{M}_\alpha^{(1)} = \{ \theta_1 : \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \in \mathcal{M}_\alpha \}, \quad \mathcal{L}_\alpha^{(1)} = \{ \theta_1 : \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \in \mathcal{L}_\alpha \} \]

Then

\[ \mathcal{M}_\alpha^{(1)} = \mathcal{L}_\alpha^{(1)}. \]

**Proof:**

Since \( H_\alpha(x_1, x_2) > F_\alpha(x_1, x_2) \) for all \( x_1, x_2 \) implies

\[ H_{\alpha X_1}(x_1) > F_{\alpha X_1}(x_1), \quad \forall x_1, \]

it is clear that

\[ \mathcal{L}_\alpha \subseteq \mathcal{M}_\alpha \]

\[ \Rightarrow \quad \mathcal{L}_\alpha^{(1)} \subseteq \mathcal{M}_\alpha^{(1)}. \]

We now wish to show \( \mathcal{M}_\alpha^{(1)} \subseteq \mathcal{L}_\alpha^{(1)}. \)

Assume not. Assume \( \theta_1 \in \mathcal{M}_\alpha^{(1)} \) but \( \theta_1 \notin \mathcal{L}_\alpha^{(1)}. \)

Now if \( \theta_1 \in \mathcal{M}_\alpha^{(1)} \Rightarrow \theta^* = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \in \mathcal{M}_\alpha, \quad \forall \theta_2 \in \mathbb{R}. \) And, if \( \theta_1 \notin \mathcal{L}_\alpha^{(1)} \Rightarrow \theta^* = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \notin \mathcal{L}_\alpha, \quad \forall \theta_2 \in \mathbb{R}. \)

Also,

\[ \theta^* \in \mathcal{M}_\alpha \Rightarrow H_{\alpha X_1}(x_1 - \theta_1) > F_{\alpha X_1}(x_1), \quad \forall x_1 \]

and

\[ \theta^* \in \mathcal{L}_\alpha \Rightarrow \exists (x_1, x_2) \text{ s.t.} \]
If the support of $F_\alpha$ is bounded, then

$$H_\alpha x_1(x_1 - \theta_1) = \lim_{\theta_2 \to -\infty} H_\alpha(x_1 - \theta_1, x_2 - \theta_2) < F_\alpha(x_1, x_2) \leq F_\alpha x_1(x_1)$$

If the support of $F_\alpha$ is unbounded, then

$$H_\alpha x_1(x_1 - \theta_1) = \lim_{\theta_2 \to -\infty} H_\alpha(x_1 - \theta_1, x_2 - \theta_2) \leq F_\alpha(x_1, x_2)$$

(by strict monotonicity of $F$)

In either case, we have a contradiction.

$M_\alpha^{(1)} = L_\alpha^{(1)}$.

**Note:** Let $M^{(2)} = \{ \theta_2 : \theta = (\theta_1, \theta_2) \in M_\alpha \}$,

$$L^{(2)} = \{ \theta_2 : \theta = (\theta_1, \theta_2) \in L_\alpha \},$$

and

$$N_\alpha = \{ \theta : H_\alpha x_2(x_2 - \theta_2) \leq F_\alpha x_2(x_2), \forall x_2 \},$$

$$N_\alpha^{(1)} = \{ \theta_1 : \theta = (\theta_1, \theta_2) \in N_\alpha \},$$

$$N^{(2)} = \{ \theta_2 : \theta = (\theta_1, \theta_2) \in N_\alpha \},$$

$$U^{(1)} = \{ \theta_1 : \theta = (\theta_1, \theta_2) \in U_\alpha \},$$

$$U^{(2)} = \{ \theta_2 : \theta = (\theta_1, \theta_2) \in U_\alpha \}.$$
Arguments similar to those in the lemma yield

\[ M_\alpha^{(2)} = L_\alpha^{(2)} \]
\[ N_\alpha^{(1)} = U_\alpha^{(1)} \]
\[ N_\alpha^{(2)} = U_\alpha^{(2)}. \]

This result implies a certain redundancy in the construction of each rectangular region \( A_\alpha \). Since we consider all possible rotations of the distribution \( F \), the rotation corresponding to \( \alpha_0 = \frac{3\pi}{2} \) yields \( F_{\alpha_0 x_1} = F_{x_2} \). Thus, in the construction of the approximating distribution function \( H_\alpha \), it is sufficient that \( H_\alpha \) have first marginal distribution equal to the first marginal distribution of \( F_\alpha \) and second marginal distribution \( H_{\alpha x_2} \). That is, take

\[ H_\alpha(x_1, x_2) = F_{\alpha x_1}(x_1)H_{\alpha x_2}(x_2), \forall x_1, x_2 \]

If \( F_\alpha \) is symmetric about the line \( x_2 = \theta_2 \), then \( H_{\alpha x_2} \) and \( F_{\alpha x_2} \) will have the same shape, and \( H_{\alpha x_2}(x_2 - \theta_2) = F_{\alpha x_2}(x_2) \) \( \forall x_2 \). When \( F_\alpha \) is not symmetric, we can use \( H_\alpha \) to approximate \( F_\alpha \) marginally from above and below in the \( x_2 \)-direction as follows: Take

\[ L_\alpha = \{ \theta_2 : H_{\alpha x_2}(x_2 - \theta_2) \geq F_{\alpha x_2}(x_2), \forall x_2 \}, \]
\[ U_\alpha = \{ \theta_2 : H_{\alpha x_2}(x_2 - \theta_2) \leq F_{\alpha x_2}(x_2), \forall x_2 \}. \]

Let \( \theta^{* * \alpha}_{2} = \sup_{\theta_2} L_\alpha, \theta^{** \alpha}_{2} = \inf_{\theta_2} U_\alpha \).
Define the "infinite strip" $A^*_\alpha$ by

$$A^*_\alpha = \{(x_1, x_2) : \theta^* \leq x_2 \leq \theta^{**}, \ -\infty < x_1 < \infty\}$$

(see Figure 2.1b).

The location region then can be obtained as

$$A^* = \bigcap_{\alpha \in (0, 2\pi]} R_\alpha(B^*_\alpha).$$

Clearly, we could have defined $H_{\alpha}(x_1, x_2) = H_{\alpha x_1}(x_1)F_{\alpha x_2}(x_2)$ and approximated $F_{\alpha}$ marginally in the $x_2$-direction to obtain the same result.

Thus, using either the standard stochastic ordering or a marginal ordering in either the $x_1$- or $x_2$-direction yields the same location region. We will use the two interchangeably throughout this chapter.

It is worth noting at this point deficiencies in certain other types of stochastic orderings. A commonly used ordering for bivariate distribution functions is the Schur ordering. This ordering serves to induce a semi-order on the class of distribution functions (it is not antisymmetric), and as such does not provide an ordering strong enough to uniquely define a location region. To be more specific, attempting to approximate a distribution function from above and below (as was done in Method I), using the Schur ordering yields a location region symmetric about the equi-angular line. Thus, we can determine the location
Figure 2.1b. Diagram of a typical location region $A^*_\alpha$ for the distribution function $F_\alpha$, $\alpha \in (0, 2\pi]$
of a given distribution only up to a reflection about the line $y=x$. This weakness of the Schur ordering is also found more generally in orderings defined by way of finite reflection groups, and hence it is found in the multivariate generalization of these orderings as well.

Proposition 2.1 says, in effect, that we can reduce the bivariate problem to a univariate problem by considering the marginal distribution functions in each direction. Nonetheless, problems similar to those mentioned above are again encountered in the univariate case if the ordering used is not antisymmetric. For example, a univariate ordering introduced by van Zwet (1964) termed "c-precedes" also fails to determine a location interval. In fact, this ordering is independent of location and scale, and as such, it is inappropriate to use for the purposes of this paper.

**Method II:** (axioms of location)

We now extend the univariate axioms of location to the plane. $\theta_X = \hat{\theta}_F$ will be the candidate for measure of location for $F$. Desired properties of $\theta_F$:

(Bl) $\theta_X \leq \theta_Y$, whenever $X$ is stochastically smaller than $Y$ in terms of the standard ordering. (In view of Proposition 2.1, we can equivalently require (Bl') $\bar{\theta}_X < \bar{\theta}_Y$ whenever $X$ is marginally less than $Y$..)
(B2) \( \theta_{X+a} = \theta_X + a, \forall a \in \mathbb{R}^2 \)

(B3) \( \theta_{HX} = H\theta_X; H \) an orthogonal transformation of \( \mathbb{R}^2 \)

(B4) \( \theta_{SX} = S\theta_X; S \) a symmetric positive definite transformation of \( \mathbb{R}^2 \).

Remarks:

By (B2) and (B3), we have

1. If \( F \) is symmetric about the line \( x_2 = x_1 + b \), then
   \[ \theta_X' = (\theta_1, \theta_2) \] is on the line of symmetry; i.e.,
   \[ \theta_2 = a^2 \theta_1 + b. \]
   For by (B2) and (B3) if \( T_a \) denotes
   the reflection of the plane about the line
   \[ x_2 = ax_1, \]
   then
   \[ \theta_{X-\{0\}} = \theta_{X-\{b\}} = \theta_{Ta(X-\{0\})} = T_a^\theta_{X-\{0\}} = T_a^\theta_{X-\{b\}} \]
   Hence, \( \theta_{X-\{b\}} \) is invariant w.r.t. \( T_a \)
   \[ \Rightarrow \left( \begin{array}{c} \theta_1 \\ \theta_2 - b \end{array} \right) \]
   is on the line \( x_2 = ax_1 \)
   \[ \Rightarrow \theta_2 = a^2 \theta_1 + b. \]

2. If \( F \) is symmetric about the point \( \theta \), then
   \[ \theta_X = \theta. \]
   This follows from the fact that if
   \[ R_{\pi}(X-\theta) \sim X-\theta, \]
   then
   \[ \theta_{X-\theta} = \theta_{X-\theta} \]
   \[ = \theta_{R_{\pi}(X-\theta)} \]
\[ = R_\pi(\theta_X - \theta) \]
\[ = R_\pi(\theta_X - \theta) \]

\[ \Rightarrow \theta_X - \theta \] is invariant under the rotation \( R_\pi \). Hence,
\[ \theta_X - \theta = 0 \text{ or } \theta_X = \theta. \]

Remark: In view of result (2) from linear algebra and axiom (B3), we could change axiom (B4) to

\[ (B4') \theta_{DX} = D\theta_X, \text{ where } D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}, \quad d_1 > 0, \quad d_2 > 0. \]

This is possible since we then have

\[ \theta S_X = \theta_{H'DH_X} \]
\[ = H'\theta_{DH_X} \text{ (by axiom (B3))} \]
\[ = H'D\theta_{HX} \text{ (by axiom (B4'))} \]
\[ = H'DH\theta_X \text{ (by axiom (B3))} \]
\[ = S\theta_X, \]

implying axiom (B4).

Let \( \theta \) denote a functional defined on the class of distribution functions defined on \( \mathbb{R}^2 \).
Definition 2.4: (bivariate location parameter) \( \theta \).

will be called a location parameter iff it satisfies (B1)-(B4) for all \( F \) with finite support (i.e., with
\[ S(F) \subseteq [a,b] \times [c,d] \) for some \(-\infty < a, b, c, d < \infty \).

\( \mathcal{D} \) will denote the collection of all location
parameters \( \theta \).

For a given \( F \), the location set \( L_F \) is defined to be
the collection of values of the location parameters that
exist at \( F \). Thus,
\[ L_F = \{ \theta_F : \theta \in \mathcal{D}, \theta_F \text{ exists} \}. \]

There are many ways to define an order relation
between random vectors or their corresponding distribution
functions. Indeed, four such possible orderings appropriate
for use with measures of location will be employed in this paper. However, there appears to be no intrinsic reason why
one would choose one of these orderings over the other.
This might lead one to hope that location parameters could
be defined without reference to any such stochastic
ordering. Hence, a relevant question is whether the location
axioms can be reduced to only (B2)-(B4) and subsequently,
if location parameters can be defined only in terms of these
three axioms.

Suppose \( \{ u_i(X) \} \) is a collection of functionals defined
on \( \mathcal{S} \) into \( \mathbb{R}^2 \) which satisfy
(i) \( \mu_i(AX) = A\mu_i(X) \), \( \forall \) symmetric, positive definite \( A \).

(ii) \( \mu_i(X+b) = \mu_i(X) + b \), \( \forall b \in \mathbb{R}^2 \)

(iii) \( \mu_i(-X) = -\mu_i(X) \),

for all \( i \). Then \( \gamma(X) = \sum_i \alpha_i \mu_i(X) \) where \( \sum_i \alpha_i = 1 \) also satisfies (i)-(iii).

**Verification:**

(i) \[ \gamma(AX) = \sum \alpha_i \mu_i(AX) \]
\[ = \sum \alpha_i A\mu_i(X) \]
\[ = A\sum \alpha_i \mu_i(X) \]
\[ = A\gamma(X) \]

(ii) \[ \gamma(X+b) = \sum \alpha_i \mu_i(X+b) \]
\[ = \sum \alpha_i [\mu_i(X) + b] \]
\[ = \sum \alpha_i \mu_i(X) + b\sum \alpha_i \]
\[ = \sum \alpha_i \mu_i(X) + b \cdot \]
\[ = \gamma(X) + b \]

(iii) \[ \gamma(-X) = \sum \alpha_i \mu_i(-X) \]
\[ = \sum \alpha_i [-\mu_i(X)] \]
\[ = -\sum \alpha_i \mu_i(X) \]
\[ = -\gamma(X) \].
This result clearly extends to noncountable collections of location functionals.

Now, given any three affinely independent functionals \( \theta_1, \theta_2, \theta_3 \), satisfying (i)-(iii), the above result establishes the fact that any affine combination of \( \theta_1, \theta_2, \theta_3 \) also satisfies (i)-(iii). Hence, if the location region of a given distribution function contains at least three distinct, noncollinear points, then any point in the plane would satisfy (i)-(iii). Thus, the order relationship cannot be dispensed with.

As in the univariate case, we can establish the following result.

**Theorem 2.1:** If \( F \) is continuous and strictly increasing in each argument on its support, then the closure of the location set \( L_F \) is contained in the location region \( A \).

**Proof:** Consider the case where \( S(F) \subset [a, b] \times [c, d] \), \(-\infty < a, b, c, d < \infty \). Then, \( \theta_F^* \) and \( \theta_F^{**} \) are finite for every \( \alpha \in (0, 2\pi] \). For any given \( \alpha \), we have

\[
H_\alpha(x - \theta) \text{ is symmetric about the point } \theta \text{ (by remark (2) above and (B2)), and}
\]

\[
H_\alpha(x - \theta) \geq F_\alpha(x), \forall x \text{ and } \forall \theta \in L_\alpha
\]

\[
H_\alpha(x - \theta) \leq F_\alpha(x), \forall x \text{ and } \forall \theta \in U_\alpha
\]
Let \( \theta_{F_a} \) be any candidate for a measure of location for \( F_a \); let \( \theta_{F} \) be any corresponding measure of location for \( F \).

Then \( \frac{\theta_{F}}{\alpha} < \theta \quad \forall \theta \in U \)

and

\[
\frac{\theta_{F}}{\alpha} > \theta \quad \forall \theta \in L_a \quad \text{(by B1)}
\]

\( \Leftrightarrow \theta_{F_a} \) is a lower bound for \( U_a \) and \( \theta_{F_a} \) is an upper bd for \( L_a \).

\[
\Rightarrow \frac{\theta_{F_a}}{\alpha} < \frac{\theta_{F}}{\alpha} < \frac{\theta_{F}}{\alpha}^{**}
\]

\[
\Rightarrow \frac{\theta_{F}}{\alpha} \in A_a
\]

i.e.,

\[
\frac{\theta_{F_a}}{\alpha} X = R_{a} \frac{\theta_{F}}{\alpha} X = R_{a} \frac{\theta_{F}}{\alpha} A_a
\]

(by \( \text{B3} \))

\[
\Rightarrow \frac{\theta_{F_a}}{\alpha} \in R_{a} (A_a)
\]

Since this is true for every \( \alpha \),

\[
\Rightarrow \frac{\theta_{F_a}}{\alpha} \in \bigcap_{\alpha \in (0,2\pi)} R_{a} (A_a) = A
\]

\[
\Rightarrow L_F \subseteq A.
\]

The case where \( \theta_{F_a}^{**} \) and \( \theta_{F_a}^{**} \) are infinite similarly proved.

**Method III**: (the function of symmetry)

As defined previously, a distribution \( F \) is symmetric about a point \( \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \) iff
As in the univariate case, we can extend this idea component-wise to nonsymmetric distributions by allowing θ to be a function for each F. Hence, consider the function

\[
\theta_F(x) = \begin{pmatrix}
\frac{1}{2} [x_1 - F^{-1}_{\alpha x_1}(F_{\alpha x_1}(x_1))] \\
\frac{1}{2} [x_2 - F^{-1}_{\alpha x_2}(F_{\alpha x_2}(x_2))]
\end{pmatrix} = \begin{pmatrix}
\theta_{1_F}(x_1) \\
\theta_{2_F}(x_2)
\end{pmatrix},
\]

where the notation is as defined before, and \( \alpha \in (0,2\pi] \) is fixed. When \( F \) is continuous, \( F_{\alpha x_1}(X_1) \) and \( F_{\alpha x_2}(X_2) \) (where \( X_1 \) and \( X_2 \) have c.d.f. \( F_\alpha \)) are uniformly distributed on \((0,1)\) and

\[
X_1 - 2\theta_{1_F}(X_1) = F_{\alpha x_1}^{-1}(F_{\alpha x_1}(X_1)) \text{ has distribution } F_{\alpha x_1},
\]

\[
X_2 - 2\theta_{2_F}(X_2) = F_{\alpha x_2}^{-1}(F_{\alpha x_2}(X_2)) \text{ has distribution } F_{\alpha x_2}.
\]

Thus, the marginal distributions of

\[
\left( \begin{array}{c}
X_1 \\
X_2
\end{array} \right) - \begin{pmatrix}
2\theta_{1_F}(X_1) \\
2\theta_{2_F}(X_2)
\end{pmatrix},
\]

are the same as the respective marginal distributions of

\[
\left( \begin{array}{c}
-X_1 \\
-X_2
\end{array} \right)
\]

for each \( \alpha \in (0,2\pi] \) and \( F \) continuous.

Hence, \( \theta_F(x) \) is a marginal symmetry function in the sense that when the distribution \( \left( \begin{array}{c}
X_1 \\
X_2
\end{array} \right) \) is shifted by an amount \( 2\theta_F(X_2) \), then the marginal distributions are the same as the marginal distributions of \( \left( \begin{array}{c}
-X_1 \\
-X_2
\end{array} \right) \). We have immediately
the following facts regarding $\theta_F(x)$ (Doksum, 1975):

(1) For continuous $F_\alpha$, if $\Delta_1$ and $\Delta_2$ are any two functions s.t.

$x_1 - 2\Delta_1(x_1) \sim -x_1$ and

$x_2 - 2\Delta_2(x_2) \sim -x_2 \forall \alpha \in (0, 2\pi)$, and

$x_1 - 2\Delta_1(x_1)$ and $x_2 - 2\Delta_2(x_2)$ are nondecreasing functions for a.a. $x(F_\alpha)$, then $\Delta_1(x_1) = \theta_{1F_\alpha}(x_1)$ and $\Delta_2(x_2) = \theta_{2F_\alpha}(x_2)$ for a.a. $x(F)$.

(2) For each $\alpha \in (0, 2\pi)$, $\theta_{1F_\alpha}(x_1)$ and $\theta_{2F_\alpha}(x_2)$ both satisfy the univariate axioms of location:

(u1) $\theta_{F_\alpha} < \theta_{G}$ whenever $F(x) > G(x), \forall x \in \mathbb{R}$

(u2.1) $\theta_{aX} = a\theta_X, \forall a > 0$

(u2.2) $\theta_{X+b} = \theta_{X} + b \forall b \in \mathbb{R}$

(u3) $\theta_{-X} = -\theta_X$

Note that:

$$\theta_{1F_\alpha}(x_1) = \frac{1}{2}[x_1 - F_{aX_1}^{-1}(F_{aX_1}(x_1))]$$

$$\theta_{1F_\alpha}^{-1}(u)) = \frac{1}{2}[F_{aX_1}^{-1}(u) - F_{aX_1}^{-1}(u)]$$

$$= \frac{1}{2}[F_{aX_1}^{-1}(u) + F_{aX_1}^{-1}(1-u)]$$

$$= \frac{1}{2}[m_{F_{aX_1}}(u) + m_{F_{aX_1}}^{-1}(1-u)]$$

$$= m_{F_{aX_1}}^{-1}(u) \text{ (using Doksum's notation), } u \in (0, \frac{1}{2}).$$
Similarly, \( \theta_{2F_\alpha} (F_{aX_2}^{-1} (v)) = \frac{1}{2} [F_{aX_2}^{-1} (v) + F_{aX_2}^{-1} (1-v)] \)

\[ = m_{F_{aX_2}} (v), \quad v \in (0, \frac{1}{2}] \]

Define

\[ \theta_{\alpha X_1}^* = \inf \{ m_{F_{\alpha X_1}} (u) : 0 < u \leq \frac{1}{2} \}, \]

\[ \theta_{\alpha X_2}^* = \inf \{ m_{F_{\alpha X_2}} (v) : 0 < v \leq \frac{1}{2} \}, \]

\[ \theta_{\alpha X_2}^{**} = \sup \{ m_{F_{\alpha X_2}} (v) : 0 < v \leq \frac{1}{2} \}, \]

Then the closure of the range of \( \theta_{\alpha} (x), (x \in \mathbb{R}^2) \) equals \( B_\alpha \). (see Figure 2.2).

Note: \( B_\alpha \) is the Cartesian product of the univariate location intervals \( I_{aX_1} \) and \( I_{aX_2} \) for the marginal distributions \( F_{aX_1} \) and \( F_{aX_2} \), respectively. Hence, we can use the "marginal function of symmetry" to construct the following location region for \( F \): Let \( B = \bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha} (B_\alpha) \) be the (necessarily convex) location region for \( F \).
Figure 2.2. Use of the marginal functions of symmetry to construct the location rectangle in direction $\alpha$, $\alpha \in [0, 2\pi]$
Geometrically, we take a line which forms an angle $\alpha$ with the $x_1$-axis as shown above and move it "down" until there is probability $u$ "above" the line ($l_1$). Move a parallel line from below until there is probability $u$ "below" the line ($l_2$). The $m_{F_{\alpha X_1}}(u)$ is the line (dotted line in Figure 2.2), halfway between $l_1$ and $l_2$. We do the same thing for the angle $\alpha + \frac{3\pi}{2}$ to obtain $m_{F_{\alpha X_2}}(u)$. As $u$ varies between 0 and $\frac{1}{2}$, the intersection of all lines $m_{F_{\alpha X_1}}(u)$ with lines $m_{F_{\alpha X_2}}(u)$ yields the rectangle $B_{\alpha}$. In Figure 2.2, it has been rotated back to yield $R_{-\alpha}(B_{\alpha})$.

This is done for every $\alpha \in (0, 2\pi]$, and $B$ is then taken to be $\bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha}(B_{\alpha})$.

**Theorem 2.2:** Let $F$ be continuous and strictly increasing in each argument. Let $A$ be the location region for $F$ obtained via method I, $B$ the location region for $F$ obtained via method III. Then $A = B$.

**Proof:** Using the previous notation, we have for each $\alpha \in (0, 2\pi]$
\[ \frac{\theta^*}{\alpha} = \sup_{\theta} L = (\sup_{\alpha} L^{(1)}, \sup_{\alpha} L^{(2)}) \]
\[ = (\sup_{\alpha} M^{(1)}, \sup_{\alpha} M^{(2)}) \text{ (by Proposition 2.1)} \]
\[ = (\sup_{\alpha} \{\theta_1: H_{\alpha x_1}(x_1 - \theta_1) \leq F_{\alpha x_1}(x_1), \forall x_1\}, \]
\[ \sup_{\alpha} \{\theta_2: H_{\alpha x_2}(x_2 - \theta_2) \geq F_{\alpha x_2}(x_2), \forall x_2\}) \]
\[ = (\inf_{\alpha} \{m_{\alpha x_1}^{(1)}(u): 0 < u \leq \frac{1}{2}\}, \inf_{\alpha} \{m_{\alpha x_2}^{(2)}(v): 0 < v \leq \frac{1}{2}\}) \]
\[ \text{ (by Doksum's result for the univariate case)} \]
\[ = (\frac{\theta^*}{\alpha x_1}, \frac{\theta^*}{\alpha x_2}) \]

Similarly,
\[ \frac{\theta^{**}}{\alpha} = \inf_{\theta} U = (\inf_{\alpha} U^{(1)}, \inf_{\alpha} U^{(2)}) \]
\[ = (\inf_{\alpha} N^{(1)}, \inf_{\alpha} N^{(2)}) \text{ (by Proposition 2.1)} \]
\[ = (\inf_{\alpha} \{\theta_1: H_{\alpha x_1}(x_1 - \theta_1) \leq F_{\alpha x_1}(x_1), \forall x_1\}, \]
\[ \inf_{\alpha} \{\theta_2: H_{\alpha x_2}(x_2 - \theta_2) \leq F_{\alpha x_2}(x_2), \forall x_2\}) \]
\[ = (\sup_{\alpha} \{m_{\alpha x_1}^{(1)}(u): 0 < u \leq \frac{1}{2}\}, \sup_{\alpha} \{m_{\alpha x_2}^{(2)}(v): 0 < v \leq \frac{1}{2}\}) \]
\[ \text{ (by Doksum)} \]
\[ = (\frac{\theta^{**}}{\alpha x_1}, \frac{\theta^{**}}{\alpha x_2}) \]

Hence, \( A_{\alpha} \) obtained in method I equals \( B_{\alpha} \) obtained in method III for each \( \alpha \in (0, 2\pi) \).

\[ A = \bigcap_{\alpha \in (0, 2\pi)} R_-(A_{\alpha}) = \bigcap_{\alpha \in (0, 2\pi)} R_-(B_{\alpha}) = B. \]
In the univariate case, any point \( \theta \) in the location interval \( I_F \) of a distribution with c.d.f. \( F(x) \) can be characterized as

\[
\theta = \frac{1}{2}[F^{-1}(u) + F^{-1}(1-u)] = m_F(u) \quad \text{for some } u \in (0, \frac{1}{2}].
\]

That is, if \( \theta \in I_F \), then \( \exists u \in (0, \frac{1}{2}] \) s.t. \( \theta = m_F(u) \).

A natural generalization of this to the bivariate case would be to characterize any point \( \theta' \) in the location region \( K_F \) (say) of a bivariate distribution with c.d.f. \( F(x,y) \) as

\[
\theta' = (m_{F_{X_1}}(u), m_{F_{X_2}}(v)) \quad \text{for some } u, v \in (0, \frac{1}{2}].
\]

Here, two values \( u \) and \( v \) serve to characterize \( \theta' \). It follows immediately from the univariate axioms that \( \theta \) so defined satisfies the following bivariate axioms:

1. \( \frac{\partial \theta}{\partial x} + d = \frac{\partial \theta}{\partial x} + d, \quad \forall d \in \mathbb{R}^2 \)
2. \( A\theta = A\theta \) where \( A \) is one of the following transformations:

\[
\begin{pmatrix}
-1 & 0 \\
0 & 1 \\
\end{pmatrix}; \begin{pmatrix}
1 & 0 \\
0 & -1 \\
\end{pmatrix}; \begin{pmatrix}
-1 & 0 \\
0 & -1 \\
\end{pmatrix}; \begin{pmatrix}
d_1 & 0 \\
0 & d_2 \\
\end{pmatrix}, \quad d_1 > 0, \ d_2 > 0.
\]
3. \( F(x) \geq G(x), \quad \forall x \in \mathbb{R}^2 \) implies \( \theta_F \leq \theta_G \).

The set \( \mathcal{K}_F = \{ \theta: \theta' = (m_{F_{X_1}}(u), m_{F_{X_2}}(v)), u, v \in (0, \frac{1}{2}] \} \) is a rectangle, and in fact is the Cartesian product of the
location intervals for the marginal distributions $F_{X_1}$ and $F_{X_2}$ of $F$.

If $\theta$ so defined also satisfied the axiom for rotations (i.e., $\theta_{R_{\alpha}X} = R_{\alpha}\theta_X$), then $\theta$ would indeed be a bivariate location parameter. But it is apparent that in general, any $\theta \in K_F$ fails to satisfy the axiom for rotations, and hence $K_F$ needs to be restricted in some way to obtain a subset of $\theta$-values which not only satisfy (1)-(3), but are also invariant under rotations.

To this end, consider the set $B$ obtained via method III. Recall $B = \bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha}(B_{\alpha})$, where $B_{\alpha}$ is the Cartesian product of the location intervals of $F_{\alpha X_1}$ and $F_{\alpha X_2}$. Thus, $B$ represents a sort of "averaging" of $K_F$ over all possible rotations $R_{\alpha}$, and we also note that $B \subset K_F$. We now wish to prove that any $\theta \in B$ is a location parameter; i.e., it satisfies location axioms (B1)-(B4). As will be shown later, it suffices to show $\theta \in B$ satisfies (1)-(3) given in this section, and it satisfies the axiom for rotation.

Now $\theta \in B \Rightarrow \theta \in \bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha}(B_{\alpha})$

$\Rightarrow \theta \in R_{-\alpha}(B_{\alpha}); \ \forall \alpha \in (0, 2\pi]$

$\Rightarrow R_{\alpha}(\theta) \in B_{\alpha}; \ \forall \alpha \in (0, 2\pi]$

Hence, we can characterize $\theta \in B$ as follows:

$\theta \in B \Rightarrow \exists$ two functions $u(\cdot), v(\cdot)$ s.t.
Two properties of \( u(\cdot) \) are evident:

1. \( u(x + \frac{3\pi}{2}) = v(x) \), so that \( \theta \) is essentially characterized by the function \( u(\cdot) \) alone.
2. \( u(\cdot) \) is periodic with period \( 2\pi \).

Hence, in the univariate case, it is a specific element of the interval \( (0, \frac{1}{2}] \) which serves to characterize any element in \( I_F \), while in the bivariate case it is a function whose range is the interval \( (0, \frac{1}{2}] \) which serves to characterize any element of \( B \).

We now introduce some terminology:

**Definition 2.5:** (direction function)

For each \( \theta \in B \), the corresponding function \( u(\cdot) : (0, 2\pi] \rightarrow (0, \frac{1}{2}] \) will be called the direction function of \( \theta \) with respect to \( F \).

The axioms of location under discussion can be divided into two categories: (i) those axioms which deal with transformations which do not change the direction the coordinate axes. These are precisely axioms (1)-(3) mentioned at the beginning of this section, and (ii) that
axiom which deals with transformations which do change the direction of the coordinate axes. These are precisely the rotations. To be more explicit, we have the following table:

Table 2.1. Classification of bivariate axioms: standard ordering

<table>
<thead>
<tr>
<th>A plane involving no direction change</th>
<th>Rotation axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $F(x) &gt; G(x), \forall x \in \mathbb{R}^2$</td>
<td>(1) $R_{\alpha} \theta_x = \theta_{R_{\alpha} x}, \forall \alpha \in (0, 2\pi]$</td>
</tr>
<tr>
<td>$\theta_F \leq \theta_G$</td>
<td></td>
</tr>
<tr>
<td>(2) $\hat{\theta}_{x + d} = \hat{\theta}_x + \hat{d}, \forall d \in \mathbb{R}^2$</td>
<td></td>
</tr>
<tr>
<td>(3) $A \hat{\theta}<em>x = \hat{\theta}</em>{A x},$ where $A$ is one of the following transformations:</td>
<td></td>
</tr>
<tr>
<td>$\begin{pmatrix} -1 &amp; 0 \ 0 &amp; 1 \end{pmatrix}, \begin{pmatrix} 1 &amp; 0 \ 0 &amp; -1 \end{pmatrix}, \begin{pmatrix} -1 &amp; 0 \ 0 &amp; -1 \end{pmatrix},$</td>
<td></td>
</tr>
<tr>
<td>$\begin{pmatrix} d_{1} &amp; 0 \ 0 &amp; d_{2} \end{pmatrix}, d_{1} &gt; 0, d_{2} &gt; 0.$</td>
<td></td>
</tr>
</tbody>
</table>

In verifying that any point $\theta \in B$ satisfies these axioms, it is necessary to invoke the direction function $u(\cdot)$. So given $\theta \in B$, we then have at our disposal the entire corresponding direction function $u(\cdot)$. For each axiom, it is necessary to specify the appropriate $\theta$ of the transformed distribution and the check that the necessary relationship
holds between these two points. In the case of the axioms for no-direction change, $u(0)$ will yield the appropriate point. For the axiom for a direction change, that is a rotation through an angle $\alpha$, $u(\alpha)$ will yield the appropriate point.

More specifically, consider axiom (1) of the no-direction change axioms given above:

Assume $F(x) \geq G(x)$, $\forall x \in \mathbb{R}^2$ and let $\theta_F \in B_F$. Then $\exists$ a direction function $u(\cdot)$ for $\theta_F$. Consider now the point $\theta'_G = (m_{G_{X_1}}(u(0)), m_{G_{X_2}}(v(0)))$ (provided of course this point is in $B_G$; otherwise, $\theta_G$ is undefined and the axiom holds vacuously). It is immediately obvious from the univariate case that $\theta_F \leq \theta_G$ as desired.

The remaining no-direction change axioms follow similarly, using $u(0)$. We see, in fact, that using $u(0)$ is simply a marginal generalization of the univariate results.

Consider now the axiom for a rotation through an angle $\alpha$. Let $G$ denote the c.d.f. of $R_{\alpha}X$. So given $\theta_{\alpha}X \in B_F$, a direction function $u(\cdot)$ for $\theta_{\alpha}X$. Consider now the point $\theta'_G = \theta_{R_{\alpha}X} = (m_{G_{X_1}}(u(\alpha)), m_{G_{X_2}}(v(\alpha)))$. Then since

$$\theta_F = \theta_{R_{\alpha}X} = R_{-\beta} \begin{pmatrix} m_{F_{\beta X_1}}(u(\beta)) \\ m_{F_{\beta X_2}}(v(\beta)) \end{pmatrix}, \forall \beta,$$
in particular,
\[ \Theta_X = R_{-\alpha} \begin{pmatrix} m_{F}^{\alpha X_1} (u(\alpha)) \\ m_{F}^{\alpha X_2} (v(\alpha)) \end{pmatrix} \]
\[ = R_{-\alpha} \begin{pmatrix} m_{G}^{X_1} (u(\alpha)) \\ m_{G}^{X_2} (v(\alpha)) \end{pmatrix} \quad \text{(since } G = F_{\alpha} \text{)} \]
\[ = R_{-\alpha} (\Theta_{-\alpha} X). \]

\[ R_{\alpha} (\Theta_{-\alpha} X) = \Theta_{-\alpha} X', \text{ as desired.} \]

Thus, \( \Theta_X \) satisfies the axiom for rotations and we thus have \( \Theta_X \) satisfies all the bivariate axioms of location:
It satisfies (Bl) and (B2) because they are no-direction axioms. Now since any reflection \( T_{\alpha} \) can be expressed as
\[ T_{\alpha} = R_{\alpha} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \text{ } \Theta_X \]
satisfies the axiom for orthogonal transformations (B3). Finally, since any symmetric p.d. matrix \( S \) can be written as \( S = H'DH \) where \( D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \), \( d_1 > 0, d_2 > 0 \) and \( H \) is orthogonal, \( \Theta_X \) is seen to satisfy axiom (B4).

We have thus proved the following theorem, providing the reverse inclusion to that established in Theorem 2.1:
Theorem 2.3: The region $B$ defined via method II is contained in the closure of the location set $L_F$.

Hence, $L_F = A = B$ and all three methods yield essentially the same location region for the bivariate distribution $F$.

**Corollary:** $\theta_{AX} = A\theta_X$ for any nonsingular matrix $A$.

**Proof:** It is known that any nonsingular matrix $A$ can be expressed as $A = SH$ where $S$ is symmetric positive definite and $H$ is orthogonal. The result now follows from axioms (B3) and (B4).

**Remark:** For fixed $\alpha$, $a$ and $b$, $\theta_{F^a(s)} = \begin{pmatrix} F^{-1}(a) \\ F^{-1}(b) \end{pmatrix}$ is not a location parameter. Consider $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where $X_1$ and $X_2$ are two independent $\chi^2$-random variables with $m_1$ and $m_2$ degrees of freedom, respectively. Consider $\mu = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ where $m_1$ and $m_2$ are the respective medians of $X_1$ and $X_2$. Now, the transformation $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ represents a shear, and we have $A\theta_X \neq \theta_{AX}$. This follows upon noting that $A\theta_X = \begin{pmatrix} m_1 + m_2 \\ m_2 \end{pmatrix}$ while $\theta_{AX} = \begin{pmatrix} m_3 \\ m_2 \end{pmatrix}$ where $m_3$ is the median of a $\chi^2$-random variable with $n_1 + n_2$ degrees of freedom. The result is obtained since $m_1 + m_2 \neq m_3$. 
1. Properties

We will now show that, in some sense, for each $\alpha \in (0, 2\pi]$ the distribution $H_\alpha$ is the symmetric distribution whose shifted versions best approximate $F_\alpha$. The approximating distribution should be such that the resulting upper and lower points of symmetry should be as close as possible. Geometrically, this amounts to making the rectangle $A_\alpha$ have as small an area as possible. In view of Proposition 2.1, we can consider the marginal distributions $F_{\alpha X_1}$ and $F_{\alpha X_2}$ of $F_\alpha$ for each $\alpha \in (0, 2\pi]$. With $H_\alpha$ as defined above, it is evident that $H_{\alpha X_1}$ and $H_{\alpha X_2}$ are each univariate distributions symmetric about zero. Hence, if $G$ is any symmetric distribution of symmetric marginals $G_{X_1}^*$ and $G_{X_2}^*$ and $G$ approximates $F$, we seek to simultaneously minimize the horizontal distance between $F_{X_1}$ and $G_{X_1}^*$, and $F_{X_2}$ and $G_{X_2}^*$. That is, $G$ should minimize

$$\Delta_2(F, G) = \sup_{0 < u < 1} \left\{ |F_{X_1}^{-1}(u) - G_{X_1}^{-1}(u)| + |F_{X_2}^{-1}(u) - G_{X_2}^{-1}(u)| \right\}$$

$$= \sup_{0 < u < 1} |F_{X_1}^{-1}(u) - G_{X_1}^{-1}(u)|$$

$$+ \sup_{0 < u < 1} |F_{X_2}^{-1}(u) - G_{X_2}^{-1}(u)|$$

def
def

$$= \Delta_1(F_{X_1}, G_{X_1}) + \Delta_1(F_{X_2}, G_{X_2})$$

With regard to the univariate case, Doksum (1975) proves the following theorem (a restatement of Theorem 1.1):
Theorem 1.1: In the class of univariate distributions, G symmetric about $m_F = \text{median of } F$, the distribution $H(x-m_F)$ minimizes $\Delta_1(F,G)$.

For reasons of symmetry, it is desirable to let F and G in Theorem 1.1 have the same median m. Then F and G will be approximated by shifts of the same distribution G. In the bivariate case then, H is the symmetric distribution whose shifted versions best approximate F in the sense of the theorem for each $\alpha \in (0,2\pi]$.

Thus, method I yields for each $\alpha$, a rectangular region $A_\alpha$ with the smallest area as compared to any other such region obtained by approximating $F_\alpha$ above and below by symmetric distributions. Hence, in this sense the resulting region $A = \bigcap_{\alpha \in (0,2\pi]} R_\alpha(A)$ is optimally small, and we have the following bivariate analogue to Theorem 3.

Theorem 2.4: In the class of distribution functions $G_\alpha$ symmetric about the point $(m_{F_\alpha X_1}, m_{F_\alpha X_2})$ for each $\alpha \in (0,2\pi]$, where $m_{F_\alpha X_1} = \text{median of } F_\alpha X_1$, $m_{F_\alpha X_2} = \text{median of } F_\alpha X_2$, the distributions $H_{\alpha X_1}^\alpha (x_1-m_{F_\alpha X_1})$ and $H_{\alpha X_2}^\alpha (x_2-m_{F_\alpha X_2})$ minimize $\Delta_1(F_{\alpha X_1},G_{\alpha X_1})$ and $\Delta_1(F_{\alpha X_2},G_{\alpha X_2})$, respectively. Hence, $H_{\alpha}^\alpha (x_1-m_{F_\alpha X_1}, x_2-m_{F_\alpha X_2})$ minimizes $\Delta_2(F_\alpha,G_\alpha)$. Thus, the region $A = \bigcap_{\alpha \in (0,2\pi]} R_\alpha(A_\alpha)$ will have the smallest area by using $H_{\alpha}^\alpha$ for each $\alpha$. 
It is worth noting that each of the three methods of constructing a location region provides a corresponding justification of why the region be convex. The first method yields a region A which is convex since it is the intersection of rectangles; the second method yields a region $L_F$ which is convex owing to the following fact: if $\{\theta_\gamma(X)\}$ is any collection of location functionals, then any convex combination of these functionals is also a location functional, and is, therefore, in $L_F$; the third method yields a region B which is convex since it is the intersection of infinite "strips". Of course, the fact that A, B and $L_F$ are essentially equivalent means that any one of the above justifications is sufficient. It is interesting to note though the variety of reasons for convexity.

2. Estimation

Given a random sample of size n from a bivariate distribution with distribution function $F(x_1, x_2)$, it is desirable to construct from these n points an estimate of the location region $B_F$.

Consider first the univariate case: Suppose we have a random sample $x_1, x_2, \ldots, x_n$ from a population with distribution function $G(x)$. As given by Doksum (1975) the natural estimator of the function of symmetry is given by

$$\hat{\theta}_n(x) = \frac{1}{2}[x - \bar{G}_n(G_n(x))]$$
where \( G_n \) and \( \overline{G}_n \) are the empirical distribution functions of \( X_1, X_2, \ldots, X_n \) and \(-X_1, -X_2, \ldots, -X_n\), respectively. Now evaluating \( \hat{\theta}_n(x) \) at the order statistics gives

\[
\hat{\theta}_n(x(i)) = \frac{1}{2}[x(i) - \overline{G}_n[G_n(x(i)))]
\]

\[
= \frac{1}{2}[x(i) - \overline{G}_n(\frac{i}{n})]
\]

\[
= \frac{1}{2}[x(i) + x(n-i+1)], \quad i = 1, 2, \ldots, n,
\]

since \( \overline{G}_n^{-1}\left(\frac{i}{n}\right) = -x(n-i+1) \).

Hence, the natural estimates of the endpoints of the univariate location interval are given by

\[
\hat{\theta}_n = \min_{1 \leq i \leq n} \frac{1}{2}[x(i) + x(n-i+1)]
\]

and

\[
\hat{\theta}_n = \max_{1 \leq i \leq n} \frac{1}{2}[x(i) + x(n-i+1)].
\]

Now this can be applied to the bivariate case since by method III, the bivariate location region is constructed by considering the bivariate c.d.f. \( F(x_1, x_2) \) marginally in every direction.

So, suppose we have a random sample

\[
(x_1^{(1)}, x_1^{(2)}), (x_2^{(1)}, x_2^{(2)}), \ldots, (x_1^{(n)}, x_2^{(n)})
\]

from a population with c.d.f. \( F(x_1, x_2) \). Let \( R_\alpha \) denote the linear transformation which represents a rotation of the plane through an angle \( \alpha \):
\[ R_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \ \alpha \in (0, \frac{\pi}{2}). \]

Let \( \{a_1, a_2, \ldots, a_N\} \) be any finite collection of angles from the interval \((0, \frac{\pi}{2})\). Then \( R_{a_1}, \ldots, R_{a_N} \) will represent the directions from which the sample c.d.f. \( F_n(x_1, x_2) \) will be "viewed". Two considerations concerning the \( a_j \)'s are evident:

1. The \( a_i \)'s should be equally spaced over the interval \((0, \frac{\pi}{2})\) in order to "see" \( f \) from as wide a range of directions as possible. This will yield a maximum amount of information about \( F \).

2. \( N \) should be chosen large enough so that a relatively high degree of resolution of the sample location region can be achieved. In practice, some compromise between computer time costs and degree of resolution must be made (see Appendix A).

Let \( (x_i, x_j) = R_{a_j}(x_i) \); \( i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, N \). From the univariate case then, the natural estimates of \( \theta_F^*, \theta_F^{**}, \theta_F^*, \) and \( \theta_F^{**} \) are given by

\[
\hat{\theta}_F^* = \min_{1 \leq i \leq n} \frac{1}{2} [x(i) + x(n-i+1)]
\]

\[
\hat{\theta}_F^{**} = \max_{1 \leq i \leq n} \frac{1}{2} [x(i) + x(n-i+1)]
\]
Hence, by method III, the location region in the direction \( \alpha_j \) (a rectangle) is estimated by

\[
\hat{\beta}_{n\alpha_j} = [\hat{\beta}^*_n\alpha_j X_1, \hat{\beta}^{**}_n\alpha_j X_1] \times [\hat{\beta}^*_n\alpha_j X_2, \hat{\beta}^{**}_n\alpha_j X_2],
\]

yielding an estimated location region

\[
\hat{\beta}_F^n = \bigcap_{j=1}^N \mathbb{R}^{-\alpha_j} (\hat{\beta}_{n\alpha_j})
\]

It will now be shown that in some sense \( \hat{\beta}_F^n \) is a consistent estimator of \( B_F \). This will be done in terms of the pseudo-metric \( \rho \) given by

\[
\rho(A, B) = P(A \Delta B)
\]

where \( P \) is the probability measure on \( \mathbb{R}^2 \) induced by \( F, A \) and \( B \) are any two \( P \)-measurable sets and \( A \Delta B \) is the symmetric difference of \( A \) and \( B \). We will prove that \( \hat{\beta}_F^n \) converges almost surely to \( B_F \) in this pseudo-metric.

Consider first the following lemma:

**Lemma 2.1:** Let \( \{X_n\}, n = 1, 2, 3, \ldots \) be a sequence of \( k \)-variate random vectors; i.e., \( X'_n = (X_{1n}, X_{2n}, \ldots, X_{kn}), n = 1, 2, 3, \ldots \). Then \( \{X'_n\} \) converges to a \( k \)-variate random vector \( X' = (X_1, X_2, \ldots, X_k) \) a.s. iff
\[ X_{jn} \to X_j \text{ a.s. for } j = 1,2,\ldots,k. \]

**Proof:** Since \(|X_{jn} - X_j| < |X_n - X|\) for \(j = 1,2,\ldots,k\), it is clear that

\[ |X_n - X| \to 0 \text{ a.s. implies } |X_{jn} - X_j| \to 0 \text{ a.s. for } j = 1,2,\ldots,k. \]

Conversely, given any \(\varepsilon > 0\), \(\exists\) an integer \(M_j\) and a null set \(N_j\) s.t. for all \(n > M_j\) and for all \(\omega \notin N_j\)

\[ |X_{jn} - X_j| < \frac{\varepsilon}{\sqrt{k}}, \quad j = 1,2,\ldots,k. \]

Let \(N = \bigcup_j N_j\), \(M = \max \{M_1,\ldots,M_k\}\). Then, for all \(\omega \notin N\), \(n > M\) we have

\[
|X_n - X| = \sqrt{\sum_{j=1}^{k} |X_{jn} - X_j|^2} < \sqrt{\sum_{j=1}^{k} \frac{\varepsilon^2}{k}} = \varepsilon
\]

This establishes the desired result.

Now it is known from the univariate case that \(\hat{\theta}^*_F\), \(\hat{\theta}^{**}_F\), \(\hat{\theta}^*_\alpha\), \(\hat{\theta}^{**}_\alpha\) are strongly consistent estimators of \(\theta^*_F\), \(\theta^{**}_F\), \(\theta^*_\alpha\), \(\theta^{**}_\alpha\), respectively.

Denote the corners of the rectangular region \(\hat{\theta}^*_F\) as follows:
Denote the corresponding corners of the rectangular region $B_{\alpha_j}$ as follows:

\begin{align*}
\hat{b}_{\alpha_j1} &= \begin{pmatrix}
\hat{\theta}_F^{* \alpha_j x_1} \\
\hat{\theta}_F^{* \alpha_j x_2}
\end{pmatrix} \\
\hat{b}_{\alpha_j2} &= \begin{pmatrix}
\hat{\theta}_F^{* \alpha_j x_1} \\
\hat{\theta}_F^{* \alpha_j x_2}
\end{pmatrix} \\
\hat{b}_{\alpha_j3} &= \begin{pmatrix}
\hat{\theta}_F^{* \alpha_j x_1} \\
\hat{\theta}_F^{* \alpha_j x_2}
\end{pmatrix} \\
\hat{b}_{\alpha_j4} &= \begin{pmatrix}
\hat{\theta}_F^{* \alpha_j x_1} \\
\hat{\theta}_F^{* \alpha_j x_2}
\end{pmatrix}
\end{align*}
Then by the preceding lemma, we have

$$||\hat{\epsilon}^{n}_{a_{ij}} - b_{a_{ij}}|| \to 0 \text{ a.s. for } i = 1, 2, 3, 4; \forall j$$

$$||R_{-a_{j}}(\hat{\epsilon}^{n}_{a_{ij}}) - R_{-a_{j}}(b_{a_{ij}})|| =$$

$$||\hat{\epsilon}^{n}_{a_{ij}} - b_{a_{ij}}|| \to 0 \text{ a.s. for } i = 1, 2, 3, 4; \forall j.$$

Now we can interpret this result in terms of the rectangular regions $\hat{\alpha}_{n_{a_{ij}}}$ and $B_{a_{ij}}$, $(j = 1, 2, \ldots, N)$. We know there is a null set $O_{j}$ s.t. for $\omega \not\in O_{j}$ the corners of $\hat{\alpha}_{n_{a_{ij}}}$ converge pointwise to the corresponding corners of $B_{a_{ij}}$ as $n \to \infty$, for each $j$. Let $O = \bigcup_{j=1}^{N} O_{j}$. Then $O$ is a null set, and for $\omega \not\in O$, the corners of $\hat{\alpha}_{n_{a_{ij}}}$ converge pointwise to the corresponding corners of $B_{a_{ij}}$ as $n \to \infty$, for all $j$. Clearly, then for $\omega \not\in O$, $\hat{\alpha}_{n_{a_{ij}}}$ converges to $B_{a_{ij}}$ in the sense that

$$\limsup_{n} \hat{\alpha}_{n_{a_{ij}}} = \liminf_{n} \hat{\alpha}_{n_{a_{ij}}} = B_{a_{ij}}, \forall j$$

$$\Rightarrow \limsup_{n} R_{-a_{j}}(\hat{\alpha}_{n_{a_{ij}}}) = \liminf_{n} R_{-a_{j}}(\hat{\alpha}_{n_{a_{ij}}}) = R_{-a_{j}}(B_{a_{ij}}), \forall j$$

In this sense, we can say that

$$\lim_{n \to \infty} R_{-a_{j}}(\hat{\alpha}_{n_{a_{ij}}}) = R_{-a_{j}}(B_{a_{ij}}) \text{ a.s., } \forall j.$$
\[ \rho(R_{-\alpha_j}(\hat{S}_n\alpha_j), R_{-\alpha_j}(B_{\alpha_j})) = P(R_{-\alpha_j}(\hat{S}_n\alpha_j) \Delta R_{-\alpha_j}(B_{\alpha_j})) \rightarrow 0 \text{ a.s. as } n \rightarrow \infty, \forall j. \]

This fact, along with the following lemma will help establish the following result.

**Lemma 2.2**: Let \( F \) be the distribution function of the bivariate random vector \( X \). Assume \( F \) is strictly increasing in each argument and absolutely continuous with respect to Lebesgue measure on \( \mathbb{R}^2 \). Fix \( \alpha_0 \in (0, 2\pi] \) and let \( \{\alpha_j\}_{j=1} \) be any sequence of points in \( (0, 2\pi] \) s.t. \( \lim_{j \rightarrow \infty} \alpha_j = \alpha_0 \). Let \( R_{\alpha_j} \) denote the rotation of the plane through an angle \( \alpha_j \). Let \( F_{\alpha_j} \) denote the distribution function of \( R_{\alpha_j}(X) \) (necessarily absolutely continuous with respect to Lebesgue measure for \( j = 0, 1, 2, \ldots \)), and let \( f_{\alpha_j} \) denote the density of \( F_{\alpha_j} \), respectively, for \( j = 0, 1, 2, \ldots \). If

\[
\lim_{j \rightarrow \infty} f_{\alpha_j}(x_1, x_2) = f_{\alpha_0}(x_1, x_2) \text{ for a.e. } (x_1, x_2),
\]

then

\[
\lim_{j \rightarrow \infty} \sup_{0 < u < \frac{\pi}{2}} |\theta_{F_{\alpha_j}}(u) - \theta_{F_{\alpha_0}}(u)| = 0 \text{ and } \lim_{j \rightarrow \infty} \sup_{0 < u < \frac{\pi}{2}} |\theta_{F_{\alpha_j}}(u) - \theta_{F_{\alpha_0}}(u)| = 0.
\]

In this case then, the location rectangles \( A_{\alpha_j} \) for \( F_{\alpha_j} \) converge uniformly to \( A_{\alpha_0} \).
Proof: By Scheffé's Theorem on converge of densities,

\[ \lim_{j \to \infty} f_{\alpha_j}(x_1, x_2) = f_{\alpha_0}(x_1, x_2) \quad \text{for a.a. } (x_1, x_2) \]

implies

\[ \lim_{j \to \infty} \iint_{\mathbb{R}^2} |f_{\alpha_j}(x_1, x_2) - f_{\alpha_0}(x_1, x_2)| \, dx_1 \, dx_2 = 0 \]

This implies

\[ \lim_{j \to \infty} \iint_{B} |f_{\alpha_j}(x_1, x_2) - f_{\alpha_0}(x_1, x_2)| \, dx_1 \, dx_2 = 0 \]

uniformly for all Borel sets B. This yields

\[ \lim_{j \to \infty} F_{\alpha_j X_1}(x_1) = F_{\alpha_0 X_1}(x_1), \quad \forall x_1 \]

and

\[ \lim_{j \to \infty} F_{\alpha_j X_2}(x_2) = F_{\alpha_0 X_2}(x_2), \quad \forall x_2. \]

The convergence is uniform in both cases by Polya's Theorem.

We will now show \( F_{\alpha_j X_1}(x_1) \) uniformly \( \xrightarrow{j \to \infty} F_{\alpha_0 X_1}(x_1) \) implies

\[ F_{\alpha_j X_1}(u) \xrightarrow{\text{uniformly}} F_{\alpha_0 X_1}(u). \]

Let \( \varepsilon > 0 \) be given. Then, since \( F_{\alpha_0 X_1}(x+\varepsilon) > F_{\alpha_0 X_1}(x), \forall x, \)

\[ \lim_{j \to \infty} F_{\alpha_j X_1}(x+\varepsilon) > F_{\alpha_0 X_1}(x), \forall x. \]
Hence, \( J \) a positive integer. \( J_1 \) s.t. \( \forall j > J_1 \) and \( \forall x \)

\[
F_{\alpha_j X_1}(x + \epsilon) > F_{\alpha_0 X_1}(x)
\]

This implies \( \forall j > J_1 \) and \( \forall x \)

\[
x + \epsilon > F_{\alpha_j X_1}^{-1}(F_{\alpha_0 X_1}(x))
\]

If we let \( u = F_{\alpha_0 X_1}(x) \), then \( \forall j > J_1 \) and \( \forall u \)

\[
F_{\alpha_0 X_1}^{-1}(u) + \epsilon > F_{\alpha_j X_1}^{-1}(u)
\]

A similar argument yields the existence of a positive integer \( J_2 \) s.t. \( \forall j > J_2 \) and \( \forall u \)

\[
F_{\alpha_0 X_1}^{-1}(u) - \epsilon < F_{\alpha_j X_1}^{-1}(u)
\]

So, for \( j > \max\{J_1, J_2\} \) and \( \forall u \), we have

\[
|F_{\alpha_j X_1}^{-1}(u) - F_{\alpha_0 X_1}^{-1}(u)| < \epsilon
\]

Thus, \( F_{\alpha_j X_1}^{-1}(u) \) converges uniformly to \( F_{\alpha_0 X_1}^{-1}(u) \) on \( (0, \frac{1}{2}] \). A similar argument yields the uniform convergence of

\[
F_{\alpha_j X_2}^{-1}(u) \text{ to } F_{\alpha_0 X_2}^{-1}(u) \text{ on } (0, \frac{1}{2}].
\]

Hence,

\[
\theta_{F_{\alpha_j X_1}}(u) = \frac{1}{2}[F_{\alpha_j X_1}^{-1}(u) + F_{\alpha_j X_1}^{-1}(1-u)]
\]
converges uniformly to $\theta_{F_{a_0X_1}}(u) = \frac{1}{2}[F_{a_0X_1}^{-1}(u) + F_{a_0X_1}^{-1}(1-u)]$ on $(0, \frac{1}{2}]$.

Similarly, $\theta_{F_{a_jX_2}}(u)$ converges uniformly to $\theta_{F_{a_0X_2}}(u)$ on $(0, \frac{1}{2}]$.

Now any point in $A_{a_0}$ can be characterized as

$$\theta_{F_{a_0}}(u,v) = \begin{cases} \theta_{F_{a_0}}(u) \\ \theta_{F_{a_0}}(v) \end{cases}$$

for $u, v \in (0, \frac{1}{2}]$

and by the above results,

$\theta_{F_{a_j}}(u,v)$ converges uniformly to $\theta_{F_{a_0}}(u,v)$. Thus, the desired result is established.

**Theorem 2.5:** Under the assumptions given in the preceding lemma, the estimated location region $\hat{B}_n$ converges a.s. in the pseudo-metric $\rho$ to $B_F$.

**Proof:** $B_F$ can be constructed in a countable sequence of steps as follows:

1. construct $B_1 = B_{a_1}$ with $a_1 = 0$
2. construct $B_2 = B_{a_2}$ with $a_2 = \frac{\pi}{2}$

and compute $B_{a_1} \cap R_{-a_2}(B_{a_2})$. 
(3) construct $B_3 = B_{\alpha_3}$ with $\alpha_3 = \frac{\pi}{4}$

and $B_4 = B_{\alpha_4}$ with $\alpha_4 = \frac{3\pi}{4}$

and compute $\bigcap_{j=1}^{4} R_{-\alpha_j} (B_{\alpha_j})$

Continue in this way, constructing $2^{k-1}$ equally spaced rectangles at the $k^{th}$ stage. Clearly then, by Lemma 2.2

$$B_F = \bigcap_{j=1}^{\infty} R_{-\alpha_j} (B_{\alpha_j})$$

and

$$\bigcap_{j=1}^{2^{N-1}} R_{-\alpha_j} (B_{\alpha_j}) \uplus B_F \text{ as } N \to \infty.$$ 

$$\Rightarrow \lim_{N \to \infty} P\left( \bigcap_{j=1}^{2^{N-1}} R_{-\alpha_j} (B_{\alpha_j}) \right) = P(B_F).$$

Let $\varepsilon > 0$ be given. Then there exists a positive integer $N$ s.t.

$$\rho\left( \bigcap_{j=1}^{N} R_{-\alpha_j} (B_{\alpha_j}), B_F \right) = P\left( \bigcap_{j=1}^{N} R_{-\alpha_j} (B_{\alpha_j}) \right) - P(B_F) < \frac{\varepsilon}{2}.$$ 

Also, there exist positive integers $M_j$ s.t. $\forall n > M_j$

$$\rho(R_{-\alpha_j} (B_{\alpha_j}), R_{-\alpha_j} (\hat{B}_{\alpha_j})) < \frac{\varepsilon}{2N} \text{ a.s. for } j = 1, 2, \ldots, N.$$ 

Let $M = \max\{M_1, M_2, \ldots, M_N\}$. Then $\forall n > M$ we have
66

\[
\rho(\hat{\beta}_{F_n}, B_F) = \rho(\bigcap_{j=1}^{N} R_{-\alpha_j}(\hat{\beta}_{n\alpha_j}), B_F) \\
\leq \rho(\bigcap_{j=1}^{N} R_{-\alpha_j}(\hat{\beta}_{n\alpha_j}), \bigcap_{j=1}^{N} R_{-\alpha_j}(B_{\alpha_j})) + \\
\rho(\bigcap_{j=1}^{N} R_{-\alpha_j}(B_{\alpha_j}), B_F) \\
\leq P(\bigcup_{j=1}^{N} (R_{-\alpha_j}(\hat{\beta}_{n\alpha_j}) \Delta R_{-\alpha_j}(B_{\alpha_j}))) + \\
P((\bigcap_{j=1}^{N} R_{-\alpha_j}(B_{\alpha_j})) - B_F) \\
\leq \sum_{j=1}^{N} P(R_{-\alpha_j}(\hat{\beta}_{n\alpha_j}) \Delta R_{-\alpha_j}(B_{\alpha_j})) + \\
\left[ P((\bigcap_{j=1}^{N} R_{-\alpha_j}(B_{\alpha_j})) - P(B_F)) \right] \\
< N \cdot \frac{\varepsilon}{2N} + \frac{\varepsilon}{2} \text{ a.s.} \\
= \varepsilon.
\]

3. Examples

What follows are four examples of the above estimation procedure. In each case, eighteen rectangles are estimated at increments of 5°, evenly spaced from 0° to 90°. The program used in the construction of \( \hat{\beta}_{F_n} \) is given in Appendix A. For example, the first diagram is obtained by generating a random sample of 300 bivariate observations from a \( N_2(0^t, 10_t) \) population. Then, a subroutine is used to rotate the points through the appropriate angle \( \alpha_j \). The corners of
the rectangular region $\hat{B}_{n\alpha_j}$ are then computed using the above algorithm. Finally, these four points are rotated back through an angle $-\alpha_j$ using the rotation subroutine, and the corresponding rectangle $R_{-\alpha_j}(\hat{B}_{n\alpha_j})$ is then plotted. The estimated location region is the intersection of the so-computed 18 rectangles.
Figure 2.3. Estimated location region for $X \sim N_2(0, I_2)$; $n=300$
Figure 2.4. Estimated location region for $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where $X_1 \sim B(3, 5); X_2 \sim B(2, 6)$ $X_1$ and $X_2$ independent; $n=300$
Figure 2.5. Estimated region for $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $X_1$ and $X_2$
i.i.d. $B(3,5)$; $n=300$
Figure 2.6. Estimated location region for $X \sim N_2(0, \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix})$; $n=300$
III. CONSTRUCTION VIA THE CONDITIONAL ORDERING

A. Definitions

In the preceding chapter a location region was constructed using the standard stochastic ordering. This region was proved to be equivalent to the region which would be obtained using the marginal stochastic ordering. In this sense, the bivariate problem was seen to be reducible to a univariate problem, and thus the results obtained by Doksum (1975) could then be applied. In this same spirit a third stochastic ordering is possible, one which is based upon the univariate conditional distribution functions.

In this chapter, we will use the conditional ordering to construct a location region by again appealing to the three approaches used by Doksum. Once having obtained this region, its estimation will then be considered, and the asymptotic convergence of the sample region to the true region will be established. Finally, a different approach to obtaining location parameter will be considered. We will show that the value of \( \theta \) which minimizes the double integral

\[
\iint_{\mathbb{R}^2} W(||x-\theta||) \, dF(x)
\]

satisfies the location axioms (B2)-(B4). Here, \( W \) is essentially a positive, even, convex function. The author found that
none of the stochastic orderings given to this point 
(standard, marginal, or conditional) was sufficiently 
strong enough to establish the respective axioms (Bl), (Bl') 
or (Bl") . But the strong conditional ordering does enable 
one to establish a corresponding order axiom. Thus, a 
whole class of location parameters will be obtained, and in 
certain special cases these can be shown to provide examples 
of location parameters under the marginal and the conditional 
stochastic orderings.

We begin by establishing some relevant notation and then 
restating the definition of the conditional ordering.

Let F be a continuous, strictly increasing bivariate 
distribution function. If X is a random vector with 
distribution function F, let Fα denote the distribution 
function of z = Rα (X), αε(0,2π]. Thus, FZ1|Z2(x1|x2) is a 
conditional distribution function for the rotated distribu-
tion Fα. We will use the notation 

\[ F_{\alpha X_1|X_2}(x_1|x_2) = F_{Z_1|Z_2}(x_1|x_2) \]

for such conditional distributions. Similarly,

\[ F_{\alpha X_2|X_1}(x_2|x_1) = F_{Z_2|Z_1}(x_2|x_1). \]

As before, FαX1 and FαX2 denote the respective marginal 
distribution functions of the rotated distribution Fα; i.e.,
Definition 3.1: (conditional stochastic ordering)

Suppose $X$ is a random variable with c.d.f. $F$ and $Y$ is a random variable with c.d.f. $G$. We say $F$ is conditionally smaller than $G$ in the $x^-_1$-direction iff $F_{x_1|(x_1,x_2)} \leq G_{y_1|(y_1,y_2)}(x_1|x_2)$ and $F_{x_2} \geq G_{y_2}(x_2), \forall x_1, x_2 \in \mathbb{R}$.

We say $F$ is conditionally smaller than $G$ in the $x^-_2$-direction iff $F_{x_2|x_1} \geq G_{y_2|x_1}(x_2|x_1)$ and $F_{x_1} \geq G_{y_1}(x_1), \forall x_1, x_2 \in \mathbb{R}$.

We say $F$ is conditionally smaller than $G$ iff $F$ is strongly smaller than $G$ in both the $x^-_1$-direction and the $x^-_2$-direction.

Using this definition of stochastic ordering, we will now construct a location region in the plane for a given bivariate distribution function $F$.

B. Construction of the Location Set

Method 1: We will approximate $F$ from above and below with a distribution symmetric about a line. This will be done in all directions.

Let $\alpha \in (0, 2\pi]$ be fixed and for each $x_1 \in \mathbb{R}$, consider the conditional distribution function $F_{x_2|x_1}(x_2|x_1)$. If,
conditional on \(X_1=x_1\), \(X_2\) has distribution function \(F_{\alpha|X_2|X_1}\), let \(F_{\alpha|X_2|X_1}\) denote the distribution of \(-X_2\), conditional on \(X_1=x_1\). As before, let \(F_{\alpha|X_2|X_1}\) denote the function which bisects all horizontal lines connecting \(F_{\alpha|X_2|X_1}\) and \(\bar{F}_{\alpha|X_2|X_1}\). Thus, \(F_{\alpha|X_2|X_1}\) is the distribution function whose inverse is given by

\[
H^{-1}_{\alpha|X_2|X_1}(u)=\frac{1}{2}(F^{-1}_{\alpha|X_2|X_1}(u)+\bar{F}^{-1}_{\alpha|X_2|X_1}(u)), \quad u\in(0,1).
\]

Hence, for each value of \(X_1\), \(F_{\alpha|X_2|X_1}(x_2|x_1)\) is a univariate distribution function which is symmetric about zero. If \(F_{\alpha}\) is symmetric about the line \(x_2=\theta_2\), then \(F_{\alpha|X_2|X_1}(x_2|x_1)\) and \(\bar{F}_{\alpha|X_2|X_1}(x_2|x_1)\) will have the same shape, \(\forall x_1\) and \(\bar{F}_{\alpha|X_2|X_1}(x_2-x_1|x_1) = F_{\alpha|X_2|X_1}(x_2|x_1), \quad \forall x_1, x_2.

When \(F_{\alpha}\) is not symmetric, we can approximate it from above and below as follows:

Define the bivariate distribution function \(H_{\alpha}^*\) by

\[
H_{\alpha}^*(x_1,x_2) = \int_{-\infty}^{x_1} H_{\alpha|X_2|X_1}(x_2|x_1^*) dF_{\alpha|X_1^*}(x_1^*).
\]

If \(H_{\alpha}^*\) has density \(h_{\alpha}^*\), then

\[
h_{\alpha}^*(x_1,x_2) = h_{\alpha|X_2|X_1}(x_2|x_1) f_{\alpha|X_1}(x_1)
\]

\[
\Rightarrow h_{\alpha|X_1}(x_1) = f_{\alpha|X_1}(x_1)
\]

\[
\Rightarrow h_{\alpha|X_2|X_1}(x_2|x_1) = \frac{h_{\alpha}^*(x_1,x_2)}{h_{\alpha|X_1}(x_1)}
\]
for each value of \( x_1 \), \( H_{ax_2}^* (x_2 | x_1) \) is a univariate distribution function symmetric about zero. Also, 
\[
H_{ax_2}^* (x_2) = \int_{-\infty}^{\infty} H_{ax_2}^* (x_2 | x_1^*) dF_{ax_1} (x_1^*)
\]
\[
\Rightarrow 1 - H_{ax_2}^* (-x_2) = \int_{-\infty}^{\infty} [1 - H_{ax_2}^* (-x_2 | x_1^*)] dF_{ax_1} (x_1^*)
\]
\[
= \int_{-\infty}^{\infty} H_{ax_2}^* (x_2 | x_1^*) dF_{ax_1} (x_1^*)
\]
\[
= H_{ax_2}^* (x_2)
\]
\[
\Rightarrow H_{ax_2}^* \text{ is a univariate distribution function symmetric about zero.}
\]

It is also clear that \( H_{ax_1}^* (x_1) = F_{ax_1} (x_1), \forall x_1 \in \mathbb{R} \). Finally, \( H_{ax}^* \) is a bivariate distribution function symmetric about the line \( x_2 = 0 \).

To prove this, we must show
\[
H_{ax}^* (x_1, x_2) = H_{ax_1}^* (x_1) - H_{ax}^* (x_1, -x_2), \forall x_1, x_2 \in \mathbb{R}.
\]
But,
\[
H^*_\alpha(x_1) - H^*(x_1',-x_2) =
\]
\[
F_{\alpha x_1}(x_1) - \int_{-\infty}^{x_1} H_{\alpha x_2|x_1}(-x_2|x_1^*) dF_{\alpha x_1}(x_1^*) =
\]
\[
F_{\alpha x_1}(x_1) - \int_{-\infty}^{x_1} (1-H_{\alpha x_2|x_1}(x_2|x_1^*)) dF_{\alpha x_1}(x_1^*) =
\]
\[
F_{\alpha x_1}(x_1) - [F_{\alpha x_1}(x_1) - H^*_\alpha(x_1',x_2)] =
\]
\[
H^*_\alpha(x_1',x_2).
\]

Now let \(S_\alpha(x_1) = \{ \theta_2 : H^*_\alpha x_2|x_1(x_2,-\theta_2|x_1) \geq F_{\alpha x_2|x_1}(x_2|x_1), \forall x_2 \in \mathbb{R} \}\)

and define \(S_\alpha = \bigcap_{x_1} S_\alpha(x_1)\).

Let \(\theta^*_2 \in S_\alpha\). Then \(H^*_\alpha x_2|x_1(x_2,-\theta^*_2|x_1) \geq F_{\alpha x_2|x_1}(x_2|x_1), \forall x_1, x_2\)

and since \(H^*_\alpha(x_1) = F_{\alpha x_1}(x_1), \forall x_1\) we see that \(H^*_\alpha(x_1',x_2-\theta^*_2)\) is conditionally smaller than \(F_{\alpha}\) in the \(x_2\)-direction, and \(H^*_\alpha(x_1',x_2-\theta^*_2)\) is symmetric about the line \(x_2 = \theta^*_2\).

We can similarly define
\[
T_\alpha(x_1) = \{ \theta_2 : H^*_\alpha x_2|x_1(x_2,-\theta_2|x_1) \leq F_{\alpha x_2|x_1}(x_2|x_1), \forall x_2 \in \mathbb{R} \}\)

and let \(T_\alpha = \bigcap_{x_1} T_\alpha(x_1)\).
Then for $\theta^{**} \in T_\alpha$, $H^*(x_1, x_2 - \theta^{**})$ is conditionally larger than $F_\alpha$ the $x_2$-direction, and $H^*(x_1, x_2 - \theta^{**})$ is symmetric about the line $x_2 = \theta^{**}$.

Let $\underline{\theta}_2 = \inf T_\alpha$, $\overline{\theta}_2 = \sup S_\alpha$.

Then the lines $x_2 = \underline{\theta}_2$ and $x_2 = \overline{\theta}_2$ can be thought of as lower and upper lines of symmetry for $F_\alpha$ in the $x_2$-direction.

Let $A_\alpha = \{(x_1, x_2): \underline{\theta}_2 \leq x_2 \leq \overline{\theta}_2, -\infty < x_1 < \infty\}$

$= \mathbb{R} \times I_\alpha$ where $I_\alpha = [\underline{\theta}_2, \overline{\theta}_2]$.

The location region for $F$ is then

$A = \bigcap_{\alpha \in (0, 2\pi]} R_\alpha (A_\alpha)$.

**Method II:** (axioms of location)

What follows is a restatement of the bivariate axioms of location with the first axiom replaced by the conditional stochastic ordering defined above:

**(B1")**: $\frac{\theta_X}{Y} \leq \frac{\theta_Y}{Y}$ whenever $X$ is conditionally smaller than $Y$.

**(B2)**: $\frac{\theta_{X+a}}{Y} = \frac{\theta_X}{Y} + a$, $\forall a \in \mathbb{R}^2$

**(B3)**: $\frac{\theta_{HX}}{Y} = H \frac{\theta_X}{Y}$, $H$ an orthogonal transformation of $\mathbb{R}^2$. 
(B4): \( \theta^*_X = S \); \( S \) a symmetric positive definite
transformation of \( \mathbb{R}^2 \).

As before, let \( \mathcal{B} \) denote the collection of all location
functionals \( \theta \) (see Definition 2.4) and define the location
set \( L_F \) for the distribution function \( F \) by

\[
L_F = \{ \theta_F : \theta \in \mathcal{B}; \theta_F \text{ exists} \}.
\]

**Method III:** (the function of symmetry)

Let \( \theta_{F|X_2 \mid X_1} (u \mid x_1) \) be the function of symmetry for the
univariate distribution function \( F_{\alpha X_2 \mid X_1} \) where \( u \in (0, \frac{1}{2}] \); i.e.,

\[
\theta_{F|X_2 \mid X_1} (u \mid x_1) = \frac{1}{2}[F_{\alpha X_2 \mid X_1=1} (u) + F_{\alpha X_2 \mid X_1=0} (1-u)],
\]

\( u \in (0, \frac{1}{2}] \).

Let

\[
\theta^*_{F|X_2 \mid X_1} (x_1) = \inf_{0 < u \leq \frac{1}{2}} \theta_{F|X_2 \mid X_1} (u \mid x_1),
\]

\[
\theta^*_{F|X_2 \mid X_1} (x_1) = \sup_{0 < u \leq \frac{1}{2}} \theta_{F|X_2 \mid X_1} (u \mid x_1)
\]

Then

\[
I_{\alpha} (x_1) = [\theta^*_{F|X_2 \mid X_1} (x_1), \theta^*_{F|X_2 \mid X_1} (x_1)] \text{ is the location}
\]

interval for \( F_{\alpha X_2 \mid X_1} \)

Now, let \( B_{\alpha} = \mathbb{R} \times \left( \bigcup_{x_1} I_{\alpha} (x_1) \right) \), which is the infinite strip
obtained by taking the Cartesian product of \( \mathbb{R} \) and the union of the ranges of the symmetry functions \( \theta_{F_{\alpha X_2} | X_1} (u|x_1) \), the union being over all \( x_1 \) (see Figure 3.1). Hence, we can take as a location region for \( F \) the set \( B = \bigcap_{\alpha \in (0,2\pi]} R_{-\alpha}(B_{\alpha}). \)

We have the following results concerning these three methods of construction:

**Theorem 3.1:** If \( F \) is continuous and strictly increasing in each argument, then \( \overline{F} \subseteq \bar{A}. \)

**Proof:** The proof is completely analogous to the proof of Theorem 2.1.

**Theorem 3.2:** If \( F \) is continuous and strictly increasing in each argument, then \( A = B. \)

**Proof:** We know that for each \( \alpha \in (0,2\pi], \)

\[
\sup S_{\alpha}(x_1) = \sup \{ \theta_2: \theta_{F_{\alpha X_2} | X_1} (x_2 - \theta_2 | x_1) \geq F_{\alpha X_2} | X_1 (x_2 | x_1), \forall x_2 \}
\]

\[
= \inf_{0<u<\frac{1}{2}} \theta_{F_{\alpha X_2} | X_1} (u | x_1) \quad \text{(by Proposition 1.1)}
\]

\[
= \theta_{F_{\alpha X_2} | X_1} (x_1)
\]
Similarly, \[ \inf T(x_1) = \sup_{0<u<1} \theta_F^{ax_2|x_1} (u|x_1) \]
\[ = \overline{r}_{ax_2|x_1} (x_1) \]

Hence,
\[ A_\alpha = \mathbb{R}[\theta_{2\alpha}, \overline{\theta}_{2\alpha}] \]
\[ = \mathbb{R}[\sup S_\alpha, \inf T_\alpha] \]
\[ = \mathbb{R}[\sup \bigcap_{x_1} S_\alpha(x_1), \inf \bigcap_{x_1} T_\alpha(x_1)] \]
\[ = \mathbb{R}[\inf \bigcup_{x_1} S_\alpha^c(x_1), \sup \bigcup_{x_1} T_\alpha^c(x_1)] \]
\[ = \mathbb{R}[\{ \bigcup_{x_1} S_\alpha^c(x_1) \} \{ \bigcup_{x_1} T_\alpha^c(x_1) \}] \]
\[ = \mathbb{R}[\{ \bigcup_{x_1} [S_\alpha^c(x_1) \bigcap T_\alpha^c(x_1)] \} \]
\[ = \mathbb{R}[\{ \inf S_\alpha^c(x_1), \inf T_\alpha^c(x_1) \}] \]
\[ = \mathbb{R}[\{ \sup S_\alpha(x_1), \inf T_\alpha(x_1) \}] \]
\[ = \mathbb{R}[\{ \sup [\theta_F^{ax_2|x_1}, \overline{\theta}_F^{ax_2|x_1}] (x_1) \}] \]
\[ = \mathbb{R}[\{ \bigcup_{x_1} I_\alpha(x_1) \}] \]
\[ = B_\alpha \]

So, \[ A_\alpha = B_\alpha, \forall \alpha \]
\[ \Rightarrow A = B. \]
Figure 3.1. Construction of location strip $B_\alpha$. $I_\alpha(x_1^*)$ is a typical location interval for $F_{x_2|x_1}(x_1^*)$;

$B_\alpha = \mathbb{R} \times \bigcup_{x_1 \in S(F_{\alpha x_1})} I_\alpha(x_1)$
We would now like to establish the converse of Theorem 3.1. This would then imply that all three methods yield essentially the same location region.

To establish $A = B L^t$, it is necessary to be able to characterize any point of $B=A$. Now,

$$\theta \in B$$

$$\Rightarrow \theta \in R_{-\alpha}(B_{\alpha}), \forall \alpha$$

$$\Rightarrow \theta \in R_{-\alpha}(\{\mathbb{R} \times [\bigcup \mathcal{I}_{\alpha}(x_1)\})}, \forall \alpha$$

$$\Rightarrow R_{\alpha}(\theta) \in \bigcup_{x_1} [\mathbb{R} \times I_{\alpha}(x_1)], \forall \alpha$$

$$\Rightarrow \exists x_1(\alpha) \text{ s.t. } R_{\alpha}(\theta) \in R_{\alpha}I_{\alpha}(x_1(\alpha)), \forall \alpha$$

$$\Rightarrow \exists x_1(\alpha), u(\alpha) \text{ s.t.}$$

$$R_{\alpha}(\theta) = \begin{pmatrix} \theta_F^{\beta x_2} | x_1^{(\beta)} \\ \theta_F^{\alpha x_2} | x_1^{(\alpha)} \end{pmatrix}, \text{ where}$$

$$\beta = \alpha + \frac{3\pi}{2}$$

$$u(\cdot): (0,2\pi] \rightarrow (0,\frac{1}{2}]$$

and

$$x_1(\alpha): (0,2\pi] \rightarrow R_{-\alpha}(S(F_{\alpha}x_1))$$
But then

\[ R_a(\theta) = \begin{pmatrix} \theta F_{aX_1|X_2} (v(a) \mid x_2(a)) \\ \theta F_{aX_2|X_1} (u(a) \mid x_1(a)) \end{pmatrix}, \quad \forall a \]

where \( u(\cdot) \) is the direction function and \( x_1(\cdot) \) is a function which specifies from which conditional distribution \( \theta \) is computed for in a given direction, with

\[ v(a) = u(a + \frac{3\pi}{2}), \quad \forall a \]

and

\[ x_2(a) = x_1(a + \frac{3\pi}{2}), \quad \forall a. \]

As in the previous chapter, we can divide the location axioms as follows:

**Table 3.1. Classification of bivariate axioms: conditional ordering**

<table>
<thead>
<tr>
<th>Axioms involving no direction change</th>
<th>Rotation axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ( F ) conditionally less than ( G ) implies ( \theta_F &lt; \theta_G )</td>
<td>( R_a \theta F_X = \theta R_a X ), ( \forall a \in (0, 2\pi] )</td>
</tr>
<tr>
<td>(2) ( \theta_{X+a} = \theta_{X+a} ), ( \forall a \in \mathbb{R}^2 )</td>
<td>( R_a \theta X = \theta R_a X ), ( \forall a \in (0, 2\pi] )</td>
</tr>
</tbody>
</table>
| (3) \( \theta_{AX} = A\theta_{X} \), where \( A \) is of the form \[
\begin{pmatrix}
-1 & 0 \\
0 & 1 \\
d_1 & 0 \\
0 & d_2
\end{pmatrix}; \quad d_1 > 0, \quad \zeta_2 > 0
\] | \( R_a \theta X = \theta R_a X \), \( \forall a \in (0, 2\pi] \) |
Now,
\[ \theta \in B \Rightarrow \theta = \begin{pmatrix} \theta_{r|x_1|X_2}^{x_1(x_1)}(v(x_2|x_1(x_1)) \\ \theta_{r|x_1|X_2}^{x_1(x_1)}(u(x_1|x_1)) \end{pmatrix}, \forall \theta. \]

For the directional axiom, \( \theta \in B \Rightarrow \theta \in F_r \) by an argument completely analogous to the marginal ordering case. In fact, in the nondirectional case, axioms (1) and (2) above are also similarly proved. Consider case (3):

For \( D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}, \ d_1 > 0, d_2 > 0, \alpha = 0 \) and

\[
D_\theta = \begin{pmatrix} d_1 \theta_{x_1|x_2}(v(x_2|x_1(x_1)) \\ d_2 \theta_{x_2|x_1}(u(x_1|x_1)) \end{pmatrix} = \begin{pmatrix} \theta_{d_1|x_1|x_2}(v|x_2) \\ \theta_{d_2|x_2|x_1}(u|x_1) \end{pmatrix}
\]

Similarly,

for \( T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \)

\[ T_\theta = \begin{pmatrix} \theta_{x_1|x_2}(v|x_2) \\ -\theta_{x_2|x_1}(u|x_1) \end{pmatrix} = \begin{pmatrix} \theta_{x_1|-x_2}(v|-x_2) \\ \theta_{-x_2|x_1}(u|x_1) \end{pmatrix} \]
The other transformations are similarly proved. We thus have any point in $B = A$ satisfies the location axioms and so is in $\bar{L}_F$. Thus, we have

**Theorem 3.3:** $\bar{L}_F = A = B$.

**Remarks:**
1. Using the conditional form of stochastic ordering we have generated a location region which again is easily seen to be convex.
2. If, in a given direction $\alpha$, $F_\alpha(x_1, x_2) = F_\alpha x_1(x_1) F_\alpha x_2(x_2)$ then the rectangle obtained by method I (for example using $\alpha$ and $\alpha + \frac{3\pi}{2}$), will coincide with the rectangle obtained in method I of the previous chapter using the marginal ordering. In general though, Theorem 2.4 indicates that the rectangle obtained using the conditional ordering will be at least as large as the rectangle obtained by way of the marginal ordering.

1. **Estimation**

It is desirable to develop an estimate of the location region based on the conditional stochastic ordering. Whereas the marginal ordering leads to the use of the empirical marginal distribution functions (in a representative set of directions), it is apparent that we must
now consider the empirical conditional distribution functions
(again, in a representative set of directions). However, given
a random bivariate sample of \( n \) points in the plane, it is
evident that a "discretized" version of these conditional
distributions must be used in order to obtain meaningful
empirical distributions. That is, we will use the fact that

\[
F_X|Y(x|y) = \lim_{0<h \to 0} F_X|Y(x|y-h < Y < y+h)
\]

under suitable restrictions on the joint c.d.f. of \( X \) and \( Y \).

Let \( S \) represent the set of points obtained in a random
sample of size \( n \) from a bivariate population with c.d.f. \( F \).
Hence, \( S = \{(x_i, y_i): i = 1, 2, \ldots, n\} \). We will partition \( S \)
as follows:

Let \( y(n) = \max_{1 \leq i \leq n} y_i \), \( y(1) = \min_{1 \leq i \leq n} y_i \) and take

\[
h_k = \frac{y(n) - y(1)}{k}
\]

where \( k \) is some positive integer. Let

\[S_i = \{(x, y) \in S: y(1) + (i-1)h_k < y \leq y(1) + i \cdot h_k\}\]

for \( i = 1, 2, \ldots, k \).

Then \( S = \bigcup_{i=1}^{k} S_i \) and the \( S_i \) are disjoint. Then we can
obtain approximate empirical conditional distribution func­
tions \( F_{nX|Y} \) as follows:

Let \( n_i \) = number of points in \( S_i \), and let \( N_i \) = the set
of subscripts of points which are in \( S_i \); \( i = 1, \ldots, k \).

Then for \( y(1) + (i-1) \cdot h_k < y \leq y(1) + i \cdot h_k \)
Hence, an estimate of the location interval for $F_{X|Y}$ is given by $[\hat{\theta}_{i,k,n}, \hat{\theta}_{i,k,n}]$ where

$$\hat{\theta}_{i,k,n} = \min_{j \in N_i} \frac{1}{2} \left[ x(j) + x(n_i - j + 1) \right]$$

$$\hat{\theta}_{i,k,n} = \max_{j \in N_i} \frac{1}{2} \left[ x(j) + x(n_i - j + 1) \right]$$

So the estimate of the endpoints of the union of the conditional location intervals is given by

$$\hat{\theta}_n = \min_{1 \leq i \leq k} \hat{\theta}_{i,k,n}$$

$$\hat{\theta}_n = \max_{1 \leq i \leq k} \hat{\theta}_{i,k,n}$$

Let $\hat{B}_{n0} = \{(x,y) : \hat{\theta}_n \leq x \leq \hat{\theta}_n \}$.

Now, rotate the points through an angle, $\alpha_j$ and repeat the procedure. And obtain $\hat{B}_{\alpha_j n}$. Hence, $\hat{B}_n = \bigcap_{j=1}^{K} R_{-\alpha_j} (\hat{B}_{n\alpha_j})$ for some $K$.

We now wish to show that the above estimate of the location region is consistent (in the same sense as defined for the marginal ordering estimate).

We first need the following results, given as lemmas:
Lemma 3.1: Given $X_1, X_2, \ldots$, a sequence of i.i.d. random variables and a sequence of intervals $I_1, I_2, I_3, \ldots$, s.t.

$$P_X(I_j) > 0, \forall j \text{ and } \lim_{n \to \infty} P_X(I_n) = 0,$$

where $P_X$ is the probability measure induced by $X_1$.

Let $n_1 < n_2 < n_3 \ldots$ be a subsequence of the positive integer given recursively by $n_1 = 1$ and

$$n_{k+1} = n_k + \zeta(k) \quad \text{where}$$

$$\zeta(k) = \langle (P_X(I_k)[1-P_X(I_k)])^{-2} \rangle, \quad \text{where } \langle x \rangle \text{ denotes the smallest integer greater than or equal to } x.$$

Let $m_k = \zeta(k) P_X(I_k)$, $k = 1, 2, \ldots$

Define

$$Y_{n_k}, j = \begin{cases} 1 \text{ if } X_{n_k+j} \in I_k, \\ 0 \text{ otherwise} \end{cases}$$

and let $S_k = \sum_{j=n_k+1}^{n_{k+1}} Y_{n_k}, j$.

Then the distribution function of

$$\frac{S_k - \zeta(k) P_X(I_k)}{\sqrt{\zeta(k) P_X(I_k)[1-P_X(I_k)]}}$$

tends to $\Phi$ as $k \to \infty$.

Proof: $S_k \sim \text{Binomial} (\zeta(k), m_k)$. We have

$$E[Y_{n_k}, j, n_k] = P_X(I_k),$$

$$\text{var}(Y_{n_k}, j, n_k) = P_X(I_k)[1-P_X(I_k)], \text{ and}$$

$$E[Y_{n_k}, j, n_k - P_X(I_k)]^3 = P_X(I_k)[1-P_X(I_k)][P_X(I_k)]^2$$

$$+ [1-P_X(I_k)]^2.$$
for \( j = 1, 2, \ldots, \zeta(k) \).

Let \( B_k = \{ \zeta(k)P_X(I_k)\} \left[ 1 - P_X(I_k) \right] \left\{ \left[ P_X(I_k) \right]^2 + \left[ 1 - P_X(I_k) \right]^2 \right\}^{1/3} \)

and

\[ c_k = \{ \zeta(k)P_X(I_k)\} \left[ 1 - P_X(I_k) \right] \right]^{1/2} \]

Then, \( \lim_{k \to \infty} \frac{B_k}{c_k} = \lim_{k \to \infty} \frac{\left\{ \left[ P_X(I_k) \right]^2 + \left[ 1 - P_X(I_k) \right]^2 \right\}^{1/3}}{\left\{ \zeta(k)P_X(I_k)\} \left[ 1 - P_X(I_k) \right] \right\}^{1/6}} \)

\[ = \lim_{k \to \infty} \left\{ \left[ P_X(I_k)\right]^2 \right\}^{1/6} \left\{ \left[ 1 - P_X(I_k) \right]^2 \right\}^{1/3} \]

\[ = 0. \]

Hence, by Liapunov's CLT, the distribution of

\[ \frac{S_k - m_k}{\sqrt{m_k \left[ 1 - P_X(I_k) \right]}}, \]

tends to \( \phi \) as \( k \to \infty \).

**Lemma 3.2:** Given \( X_1, X_2, X_3, \ldots \) a sequence of i.i.d. random variables and a sequence of intervals \( I_1, I_2, I_3, \ldots \) s.t.

\[ P_X(I_j) > 0, \forall j \text{ and } \lim_{n \to \infty} P_X(I_n) = 0. \]

Let \( n_1 < n_2 < n_3 \ldots \) be a subsequence of positive integers given recursively by \( n_1 = 1 \) and
\[ n_{k+1} = n_k + \langle p_X(I_k) \rangle \left[ 1 - p_X(I_k) \right]^{-2} \]

\[ = n_k + \xi(k), \text{ say, where } \langle x \rangle \text{ denotes the smallest integer greater than or equal to } x. \]

Let \( A_k \) be the event that at least \( m_k = \xi(k)p_X(I_k) \) of \( X_1, X_2, \ldots, X_{n_{k+1}} \) fall into the interval \( I_k \).

Then \( p_X(A_k \ i.o.) = 1. \)

**Proof:** We wish to apply the converse part of the Borel-Cantelli lemma, but the \( A_k \)'s are not independent events.

So consider the events \( C_k \) where \( C_k \) is the event that at least \( m_k \) of \( X_{n_{k+1}}, X_{n_{k+2}}, \ldots, X_{n_{k+1}} \) fall into \( I_k \). The \( C_k \)'s are independent since they are disjoint.

Also, \( \{C_k \ i.o.\} \subset \{A_k \ i.o.\} \)

We need to show \( \sum_{k=1}^{\infty} p_X(C_k) = \infty. \)

Let
\[ Y_{jn_k} = \begin{cases} 1 & \text{if } X_{n_k+j} \in I_k \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \ldots, \xi(k) \]

and let \( S_k = \sum_{j=n_k+1}^{n_{k+1}} Y_{jn_k}. \)

Then
\[ S_k \sim \text{Binomial (} \xi(k), p_X(I_k) \text{)} \]

and by Liaponouv's CLT, we have
\[ \lim_{k \to \infty} p_X(C_k) = \lim_{k \to \infty} p_X(S_k \geq m_k) \]
\[ = P(Z > 0) \text{ (by Lemma 3.1)} \]
\[ = \frac{1}{2} \]
\[ \sum_{k=1}^{\infty} P_X(C_k) = \infty \]

\[ P_X(C_k \text{ i.o.}) = 1 \text{ by Borel-Cantelli} \]

\[ P_X(A_k \text{ i.o.}) = 1. \]

**Theorem 3.4:** Suppose \( X_1 \) and \( X_2 \) have a joint distribution function \( F \) which is absolutely continuous, strictly increasing in each argument and has bounded support. Let \( f_{\alpha}(x,y) \) denote the joint density of \( R_\alpha(X) \) and let \( f_{\alpha x_1}(x_1), f_{\alpha x_2}(x_2) \) denote the respective marginal densities. Assume that \( f_{\alpha x_2}(x_2) > 0, \forall x_2 \in S(F_{\alpha x_2}) \) and \( \forall \alpha \in (0,2\pi] \). For every \( \alpha_0 \in (0,2\pi] \) and any sequence \( \{\alpha_j\}_{j=1}^{\infty} \) in \( (0,2\pi] \) converging to \( \alpha_0 \) assume

\[ \lim_{j \to \infty} f_{\alpha_j}(x_1, x_2) = f_{\alpha_0}(x_1, x_2) \]

for a.a. \((x_1, x_2) \in S(F_{\alpha_0})\).

Then \( \hat{B}_n \) as previously defined by way of the conditional ordering converges a.s. in the pseudo-metric \( \rho \) to the location set \( B_\mu \).

**Proof:** It is necessary to show each infinite strip \( \hat{B}_{n\alpha_j} \) is a consistent estimator of \( B_{\alpha_j} \), where \( \alpha_j \) is one of a representative set of directions along which we view the sample.

Fix \( \alpha_j \) and let \( \{P_k\}_{k=1}^{\infty} \) be a sequence of partitions of \( S(F_{\alpha_j x_2}) \) s.t. \( P_{k+1} \) is a refinement of \( P_k \forall k = 1,2,3,\ldots, \)

and the norm of the partitions tends to zero:

\[ \lim_{k \to \infty} ||P_k|| = 0. \]
For any $y_0 \in S(F_{X_2}^\alpha)$, we know by Lemma 3.1 that a sequence of intervals $I_k(y_0)$, $k = 1, 2, \ldots$ s.t. $P_{X_1}^\alpha(x|y_0) \to 0$ as $k \to \infty$ and $I_k(y_0)$ is an interval defined by the partition $P_k$.

We can approximate $P_{X_1}^\alpha(x|y_0)$ by $P_{n_kX_1|X_2}(x|I_k(y_0))$ where $n_k$ is given as in the preceding lemma. Using this lemma along with an application of the strong law of large numbers yields $P_{X_1|X_2}(x|y_0)$, $\forall x, y_0$.

$\hat{\sigma}_{\alpha_j}$ and $\hat{\sigma}_{n\alpha_j}$ are consistent estimators of $\sigma_{\alpha_j}$ and $\sigma_{n\alpha_j}$, respectively.

Now by an argument completely analogous to that given in Theorem 2.5, we have $\hat{\sigma}_{F \alpha}$ is a consistent estimator of $\sigma_B$.

2. **Examples**

What follows is an example of the above estimation procedure. Again, eighteen equally spaced rectangles are estimated from a computer-generated sample of size 300. In each direction, discrete estimates of 25 conditional distributions are used. For comparison, a plot of the estimated location region obtained by using the marginal ordering of Chapter IV is also given. Note that under the conditional ordering, a much larger estimate is obtained.
Figure 3.2. Estimated location region for $X \sim N_2\left(\begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}\right)$ using the conditional ordering; $n=300$
Figure 3.3. Estimated location region for $X \sim N_2 \left( \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} \right)$ using the marginal ordering; $n=300$
C. Loss Function Approach

In this and the preceding chapters, three methods of obtaining a set of location parameters for a given distribution were presented by way of three stochastic orderings. The second or axiomatic approach in each case can be used to determine if a given functional is a location parameter; we simply verify that it satisfies the four axioms of location. In this way, it is apparent that the bivariate mean \( \mu \) is a location parameter. However, in general, it is much more difficult to give any other examples of bivariate location parameters.

In the hope of generating a class of bivariate location parameters, we appeal to a result from the univariate case. Consider a random variable \( X \) with distribution function \( F \). It can be shown (see Appendix B) that the value of \( \theta \) which minimizes the integral

\[
\int \left| x - \theta \right| \lambda dF(x), \quad \lambda > 1
\]

satisfies the univariate location axioms. Thus, this class of parameters (one for each value of \( \lambda \)) provides many examples of location parameters, including the mean (\( \lambda = 2 \)) and the median (\( \lambda = 1 \)).

A more general problem is to find the value of \( \theta \) which minimizes the integral
\[ p(x-e) dF(x) \]

where \( p \) is a positive, even, convex, twice differentiable function. We can extend this problem to the multivariate case by considering the integral

\[ \int_{\mathbb{R}^n} W(||x-\theta||) dF(x) \quad (3.2) \]

where \( X \) is a random vector with distribution function \( F \), \( || \cdot || \) denotes the Euclidean norm, and \( W \) is an even, convex, symmetric function \( W: \mathbb{R} \to \mathbb{R}^+ \setminus \{0\} \). In terms of a solution to this problem, DeGroot and Rao (1966) have proved the following result, stated here as a theorem:

**Theorem 3.5:** Suppose \( W: \mathbb{R} \to \mathbb{R} \) satisfies the following two growth conditions: for all \( x>0 \) and some \( 0<c<\infty \) and for each \( a>1 \) and some \( k_a>1 \),

1. \( W(2x) \leq cW(x) \)
2. \( W'(ax) \geq K_a W'(x) \) (where \( W'(x) = \frac{dW(x)}{dx} \))

Then, there exists a unique value of \( \theta \), say \( \theta^* \) s.t.

\[ \int_{\mathbb{R}^n} W(||x-\theta^*||) dF(x) \leq \int_{\mathbb{R}^n} W(||x-\theta||) dF(x) \]

for all \( \theta \in \mathbb{R}^n \). In fact, \( \theta^* \) minimizes Equation (3.2) iff it satisfies the equation.
Consider now the bivariate analogue of Equation (3.2):

$$\int_{\mathbb{R}^2} W(|x-\theta|) dF(x)$$

We will consider the special case where

$$W(x) = |x|^\lambda, \lambda \geq 1.$$ 

It will be noted at the end of this chapter that if $F$ is symmetric in its arguments, then in fact, $W$ must be of this form.

Hence, with the goal of providing examples of bivariate location parameters, we wish to show that for each $\lambda \geq 1$, the value of $\theta$ which minimizes the double integral

$$\int_{\mathbb{R}^2} |x-\theta|^\lambda dF(x)$$

(3.3)

satisfies the four bivariate axioms of location. By Theorem 3.5, the existence and uniqueness of such a $\theta$ are guaranteed. In fact, we know that the value of $\theta$ which minimizes Equation (3.3) must satisfy

$$\int_{\mathbb{R}^2} |x-\theta|^\lambda \cdot c' (x-\theta) dF(x) = 0,$$

where $c \in \mathbb{R}^2$ is an arbitrary fixed vector. This implies
\[
\int \int \left[ \frac{1}{\sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2}} \right] \lambda - 2 (x_1-\theta_1) dF(x_1, x_2) = \{x_1 > \theta_1\}
\]
\[
\int \int \left[ \frac{1}{\sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2}} \right] \lambda - 2 (\theta_1 - x_1) dF(x_1, x_2) = \{x_1 < \theta_1\}
\]
and
\[
\int \int \left[ \frac{1}{\sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2}} \right] \lambda - 2 (x_2-\theta_2) dF(x_1, x_2) = \{x_2 > \theta_2\}
\]
\[
\int \int \left[ \frac{1}{\sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2}} \right] \lambda - 2 (\theta_2 - x_2) dF(x_1, x_2) = \{x_2 < \theta_2\}
\]
We first establish axiom (B2):

Suppose \( X \) has joint distribution function \( F \). Let \( \bar{F} \) be the distribution function of \( Y = X + k \), for \( k \in \mathbb{R}^2 \) fixed. Then for arbitrary \( C \subseteq \mathbb{R}^2 \), we know
\[
\int \int \left| Y - \bar{\theta}_F \right|^{\lambda - 2} (Y - \bar{\theta}_F)' C \ d\bar{F}(Y) = 0
\]
Now, make a change of variable: let \( X = Y - k \), \( Y = W + k \). Then the Jacobian \( J = 1 \) and
\[
\int \int \left| W - (\bar{\theta}_F - k) \right|^{\lambda - 2} (W - (\bar{\theta}_F - k))' C \ d\bar{F}(W) = 0
\]
\Rightarrow \bar{\theta}_X = \bar{\theta}_F - k = \bar{\theta}_X + k
\Rightarrow \bar{\theta}_X + k = \bar{\theta}_X + k
Consider now axiom (B3):
Suppose $X$ has distribution function $F$. Consider
\[ \int \int_{\mathbb{R}^2} |x - \theta_F| \lambda^{-2} (x - \theta_F)' c \, dF(x) = 0 \text{ for arbitrary fixed } c \in \mathbb{R}^2. \]

Let $\overline{F}$ denote the distribution function of $Y = HX$ where $H$ is an orthogonal transformation and consider
\[ \int \int_{\mathbb{R}^2} |y - \theta_{\overline{F}}| \lambda^{-2} (y - \theta_{\overline{F}})' c \, d\overline{F}(y) = 0 \]

We make change of variable: let $w = H^{-1}y = H'y$. So,
\[ \int \int_{\mathbb{R}^2} |y - \theta_{\overline{F}}| \lambda^{-2} (y - \theta_{\overline{F}})' c \, d\overline{F}(y) = 0 \]
\[ \rightarrow \int \int_{\mathbb{R}^2} |Hw - \theta_F| \lambda^{-2} (Hw - \theta_F)' c \, dF(w) = 0 \]
\[ \rightarrow \int \int_{\mathbb{R}^2} |H(w - H'\theta_F)| \lambda^{-2} [H(w - H'\theta_F)]' c \, dF(w) = 0 \]
\[ \rightarrow \int \int_{\mathbb{R}^2} |H(w - H'\theta_F)| \lambda^{-2} [H(w - H'\theta_F)]' Hc \, dF(w) = 0 \]
(since $Hc$ is an arbitrary fixed element of $\mathbb{R}^2$.)
\[ \rightarrow \int \int_{\mathbb{R}^2} |(w - H'\theta_F)| \lambda^{-2} (w - H'\theta_F)' c \, dF(w) = 0 \]
\[ \rightarrow \theta_X = \theta_F = H'\theta_F = H'\theta_{HX} \]
\[ \rightarrow \theta_{HX} = H\theta_X \]
Consider axiom (B4):

Let \( X \) have joint distribution function \( F \).

Suppose \( F \) is the distribution function of \( Y = DX \) where \( D \) is of the form

\[
\begin{pmatrix}
  d_1 & 0 \\
  0 & d_2
\end{pmatrix}; \quad d_1 > 0, d_2 > 0.
\]

Then for \( c \in \mathbb{R}^2 \) arbitrary, we know

\[
\int \int |y - \theta_F| \lambda^2 (y - \theta_F)'c \, dF(y) = 0
\]

We make a change of variable: let \( w = D^{-1}y = \begin{pmatrix} 1/d_1 & 0 \\ 0 & 1/d_2 \end{pmatrix}y \)

\( y = Dw \),

\( J = d_1 \cdot d_2 \)

\[
\int \int |Dw - \theta_F| \lambda^2 (Dw - \theta_F)'c \, dF(w) = 0
\]

\[
\int \int |D(w - D^{-1}\theta_F)| \lambda^2 (w - D^{-1}\theta_F)'c \, dF(w) = 0
\]

\[
\int \int |D(w - D^{-1}\theta_F)| \lambda^2 (w - D^{-1}\theta_F)'c^* \, dF(w) = 0,
\]

where \( c^* = Dc \) is some arbitrary fixed element of \( \mathbb{R}^2 \)

If \( d_1 = d_2 \), then we have
\[
\int \int |w-D^{-1}a_F| \lambda^{-2} (w-D^{-1}a_F)' \sigma^* \, dF(w) = 0
\]
\[
\Rightarrow \theta_X = \theta_F = D^{-1}a_F = D^{-1}a_{DX}
\]
\[
\Rightarrow D\theta_X = \theta_{DX}
\]

Note: For the subgroup \{ \begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} : d \in \mathbb{R}^+ \} of the group \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} : a, b \in \mathbb{R}^+ \) we have the desired property. The need for taking \( d_1 = d_2 \) is a result of the fact that the Euclidean metric gives "equal weights" to the components of a vector \( \bar{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \).

We now wish to establish that the value of \( \theta \) minimizing Equation (3.2) satisfies one or more of the order axioms (Bl), (Bl'), and (Bl''). The author attempted to use the standard, marginal and conditional orderings to prove this and failing to do so, resorted to the use of the strong conditional ordering. With this ordering, it can be shown that for values of \( \theta_X \) and \( \theta_Y \) which minimize their respective integrals in Equation (3.2) we have: (Bl'') \( \theta_X < \theta_Y \) whenever \( X \) is strongly conditionally less than \( Y \). To establish this, some preliminaries are necessary. First, consider the left-hand side of Equation (3.4) as a function of \( \theta_1 \).

For each \( x_1, x_2, \) and \( \theta_2 \), let

\[
h(\theta_1) = \sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2} + 2\lambda^{-2}(x_1-\theta_1); \theta_1 < x_1
\]
Then \( h'(\theta_1) = - \frac{\sqrt{(x_1-\theta_1)^2 + (x_2-\theta_2)^2}}{\lambda-4(\lambda-1)(x_1-\theta_1)^2 + (x_2-\theta_2)^2} \)

\( \Rightarrow h'(\theta_1) < 0, \forall \theta_1 \) and \( \lambda \geq 1. \)

Hence, for all \( \lambda \geq 1, \) \( h(\theta_1) \) is decreasing for \( \theta_1 < x, \) for all \( x, x_2. \) So, for \( \theta_1^1 < \theta_1^2, \) we have

\[
\int_{x_1 > \theta_1^1} \left[ \sqrt{(x_1-\theta_1^1)^2 + (x_2-\theta_2^1)^2} \right] \lambda-2 (x_1-\theta_1^1) \, dF(x_1, x_2) > \int_{x_1 > \theta_1^2} \left[ \sqrt{(x_1-\theta_1^2)^2 + (x_2-\theta_2^2)^2} \right] \lambda-2 (x_1-\theta_1^2) \, dF(x_1, x_2)
\]

(since we are integrating a nonnegative function over a smaller set)

\[
\int_{x_1 > \theta_1^1} \left[ \sqrt{(x_1-\theta_1^1)^2 + (x_2-\theta_2^1)^2} \right] \lambda-2 (x_1-\theta_1^1) \, dF(x_1, x_2) > \int_{x_1 > \theta_1^2} \left[ \sqrt{(x_1-\theta_1^2)^2 + (x_2-\theta_2^2)^2} \right] \lambda-2 (x_1-\theta_1^2) \, dF(x_1, x_2)
\]

(since the integrand is a decreasing function of \( \theta_1 \))

Hence, for each value of \( \theta_2 \in \mathbb{R}, \) the LHS of Equation (3.4) is a decreasing function of \( \theta_1. \)

Similarly, for every \( \theta_1 \in \mathbb{R}, \) the LHS of Equation (3.5) is a decreasing function of \( \theta_2. \)

Consider now the RHS of Equation (3.4) as a function of \( \theta_1 \) for each \( \theta_2, \) and \( x_1, x_2: \) Let
\[
g(\theta_1) = \left[ (x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \right]^{\lambda-2} (\theta_1 - x_1); \ \theta_1 > x_1
\]

\[
g'(\theta_1) = \left[ (x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \right]^{\lambda-4} \left\{ (\lambda-1) (\theta_1 - x_1)^2 + (x_2 - \theta_2)^2 \right\}
\]

\[
g'(\theta_1) > 0, \ \forall \theta_1 \text{ and } \lambda \geq 1.
\]

Hence, for all \( \lambda \geq 1 \), \( g(\theta_1) \) is increasing for \( \theta_1 > x_1 \) for each \( \theta_2 \) and for all \( x_1, x_2 \).

So for \( \theta_1 < \theta_1'' \), we have

\[
\int_{x_1 < \theta_1''} \left[ (x_1 - \theta_1'')^2 + (x_2 - \theta_2)^2 \right]^{\lambda-2} (\theta_1'' - x_1) \, dF(x_1, x_2) \geq \int_{x_1 < \theta_1} \left[ (x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \right]^{\lambda-2} (\theta_1 - x_1) \, dF(x_1, x_2) \quad (\text{since we are integrating a nonnegative function over a smaller set})
\]

\[
\int_{x < \theta_1'} \left[ (x_1 - \theta_1')^2 + (x_2 - \theta_2)^2 \right]^{\lambda-2} (\theta_1' - x_1) \, dF(x_1, x_2) \quad (\text{since the integrand is an increasing function of } \theta_1')
\]

Hence, for every \( \theta_2 \in \mathbb{R} \), the RHS of Equation (3.4) is an increasing function of \( \theta_1 \).

Similarly, for every \( \theta_1 \in \mathbb{R} \), the RHS of Equation (3.5) is
an increasing function of $\theta_2$.

So, as in the univariate case, we have to show the following: if $X$ (with distribution function $F$) is strongly conditionally less than $Y$ (with distribution function $G$) then

\[
\begin{align*}
(1) & \quad \int \int |x-\theta_1| \cdot (x_1-\theta_1) dF(x) \leq \int \int |x-\theta_1| \cdot (x_1-\theta_1) dG(x) \\
& \quad \text{if } x_1 > \theta_1
\end{align*}
\]

and

\[
\begin{align*}
\int \int |x-\theta_1| \cdot (x_1-\theta_1) dF(x) & \geq \int \int |x-\theta_1| \cdot (x_1-\theta_1) dG(x) \\
& \quad \text{if } x_1 < \theta_1
\end{align*}
\]  

and

\[
\begin{align*}
(2) & \quad \int \int |x-\theta_2| \cdot (x_2-\theta_2) dF(x) \leq \int \int |x-\theta_2| \cdot (x_2-\theta_2) dG(x) \\
& \quad \text{if } x_2 > \theta_2
\end{align*}
\]

and

\[
\begin{align*}
\int \int |x-\theta_2| \cdot (x_2-\theta_2) dF(x) & \geq \int \int |x-\theta_2| \cdot (x_2-\theta_2) dG(x) \\
& \quad \text{if } x_2 < \theta_2
\end{align*}
\]  

in Figure 3.4 we have

\[
\begin{align*}
\xi_0 &= \left\{(\theta_1, \theta_2) : \int \int |x-\theta_1| \cdot (x_1-\theta_1) dF(x) \\
& \quad \text{if } x_1 > \theta_1 \right\} \\
& = \int \int |x-\theta_1| \cdot (x_1-\theta_1) dF(x) \\
& \quad \text{if } x_1 < \theta_1
\end{align*}
\]
Figure 3.4. Relationship among location parameters for $F$ and $G$ using the loss function approach.
$\lambda_3 = \{(\theta_1, \theta_2) : \int \int |x-\theta_1|^{\lambda-2}(x_1-\theta_1) dF(x) \}
\{x_1>\theta_1\}

= \int \int |x-\theta_1|^{\lambda-2}(\theta_1-x_1) dF(x) \}
\{x_1<\theta_1\}

\lambda_2 = \{(\theta_1, \theta_2) : \int \int |x-\theta_1|^{\lambda-2}(x_1-\theta_1) dG(x) \}
\{x_1>\theta_1\}

= \int \int |x-\theta_1|^{\lambda-2}(\theta_1-x_1) dG(x) \}
\{x_1<\theta_1\}

\lambda_4 = \{(\theta_1, \theta_2) : \int \int |x-\theta|^{\lambda-2}(x_2-\theta_2) dG(x) \}
\{x_2>\theta_2\}

= \int \int |x-\theta|^{\lambda-2}(\theta_2-x_2) dG(x) \}
\{x_2<\theta_2\}

Relations (3.6) and (3.7) imply $\lambda_3$ is "below" $\lambda_4$ and $\lambda_1$ is "to the left of" $\lambda_2$. This is turn implies $\theta_{\lambda} \leq \theta_{\lambda_2}$.

Now let $\mathcal{M}(E)$ denote the class of all probability measures on $E$, where $E$ is a complete, separable metric space. Suppose $E$ is endowed with a closed partial ordering, $\leq$. Let $\mathbf{B}$ be the Borel $\sigma$-field of $E$.

We will take $\mathcal{L}*(E)$ to be the class of all $\mathbf{B}$-measurable, increasing, real-valued functions on $E$, and we will take
\( \mathcal{J}(E) \) to be the family of \( \mathcal{B} \)-measurable subsets of \( E \) for which the indicator function is increasing.

**Definition 3.2:** (stochastic ordering)

Given the above notation, we say that \( P_1 \in \mathcal{M}(E) \) is stochastically smaller than \( P_2 \in \mathcal{M}(E) \) \( (P_1 < P_2) \) iff

\[
\int_{E} f dP_1 \leq \int_{E} f dP_2, \quad \forall f \in \mathcal{I}(E) \text{ for which the integrals are well-defined.}
\]

This definition is clearly equivalent to the following:

\[ P_1 < P_2 \iff P_1(A) < P_2(A), \quad \forall A \in \mathcal{J}(E). \]

**Note:** If \( E = \mathbb{R} \), and \( F(G) \) is the distribution function associated with the measure \( P_1(P_2) \), then \( P_1 < P_2 \) is equivalent to \( F(x) > G(x) \), \( \forall x \in \mathbb{R} \). This result does not extend to \( E = \mathbb{R}^n \) for \( n > 2 \).

**Definition 3.3:** (stochastic kernel)

Suppose \( E_1 \) and \( E_2 \) are complete, separable metric spaces with partial ordering \( \leq \) and Borel fields \( \mathcal{B}_1 \) and \( \mathcal{B}_2 \), respectively. A stochastic kernel in \( E_1 \times E_1 \) is a function \( k: E_1 \times E_2 \to [0,1] \) s.t.

1. \( k(\cdot, A) \) is measurable, \( \forall A \in \mathcal{B}_2 \), and
2. \( k(x, \cdot) \in \mathcal{M}(E_2) \), for each \( x \in E_1 \).
If $k$ is such a kernel and $P \in \mathcal{M}(E)$, then $P^*k \in \mathcal{M}(E \times E)$ determined by

$$(P^*k)(A \times A') = \int_{A'} k(x, A')P(dx).$$

$P^k$ is used to denote the second marginal distribution of $P^*k$.

**Definition 3.4:** (stochastically monotonic)

A stochastic kernel $k$ in $E \times E$ is said to be stochastically monotonic iff $k(x, \cdot) \leq k(y, \cdot)$, $\forall x < y$. A stochastic kernel in $E \times E$ is called upward iff $k(x, \cdot)$ is a measure with support in $\{y \in E : y \geq x\}$, $\forall x \in E$.

The following theorem, proved by Kamae, Krengel and O'Brien (1977) is stated below for future reference:

**Theorem 3.6:** For $P_1, P_2 \in \mathcal{M}(E)$, the following conditions are equivalent:

(i) $P_1 < P_2$

(ii) $\exists \lambda \in \mathcal{M}(E \times E)$ with support in $K = \{(x, y) \in E \times E : x < y\}$ with first marginal $P_1$ and second marginal $P_2$.

(iii) $\exists$ a real-valued random variable $Z$ and two measurable functions $f$ and $g : \mathbb{R} \to E$ with $f \leq g$ s.t. the distribution of $f(Z)$ is $P_1$ and the distribution of $g(Z)$ is $P_2$.

(iv) $\exists$ two $E$-valued random variables $X_1$ and $X_2$ s.t. $X_1 < X_2$ a.s. and the distribution of $X_1$ is $P_i$ $(i=1,2)$. 
(v) \exists \text{ an upward kernel } k \text{ on } E \times E \text{ s.t. } P_2 = P^k_1.

(vi) P_1(B) \leq P_2(B) \text{ for all closed } B \in \mathcal{B}(E).

Now let \( E^n = \prod_{i=1}^{n} E_i \), where each \( E_i \) is a partially ordered Polish space. Consider the product topology on \( E^n \) and define the component-wise partial ordering of points in \( E^n \) by

\[
x^n = (x_1, x_2, \ldots, x_n), \quad y^n = (y_1, \ldots, y_n)
\]

Then \( E^n \) is also a partially ordered Polish space. Kamae, Krengel and O'Brien (1977) also prove the following result:

**Theorem 3.7**: Let \( E_1, \ldots, E_n \) be partially ordered Polish spaces, \( P_1, Q_1 \in \mathcal{M}(E_1) \), \( P_1 \leq Q_1 \) and let \( p_i, q_i \ (i = 2, \ldots, n) \) be stochastic kernels on \( E^{i-1} \times E_i \) s.t.

\[
p_i(x^{i-1}, \cdot) \leq q_i(y^{i-1}, \cdot)
\]

whenever \( x^{i-1} \leq y^{i-1} \). Then,

\[
P_1 * P_2 * \cdots * P_n \leq Q_1 * Q_2 * \cdots * Q_n
\]

**Note**: Equation (3.9) implies \( P_k \leq Q_k \), where \( P_k \) and \( Q_k \) are the kth marginals of the measures in Equation (3.9). Also, if \( p_i \) is stochastically monotonic, then Equation (3.8) need only be assumed for \( x^{i-1} = y^{i-1} \) since then, if \( x^{i-1} \leq y^{i-1} \),

\[
p_i(x^{i-1}, \cdot) \leq p_i(y^{i-1}, \cdot) \leq q_i(y^{i-1}, \cdot)
\]

so that Equation (3.8) still holds. Similarly, we need only assume Equation (3.8)
(3.8) for \( x^{i-1} = y^{i-1} \) if \( q_1 \) is stochastically monotonic.

We can now apply this result to establish axiom (Bl"):

Let \( E_1 = E_2 = \mathbb{R} \). We use the component-wise ordering on \( \mathbb{R}^2 \) as given before. Assume \( X \) (with distribution function \( F \)) is strongly conditionally less than \( Y \) (with distribution function \( G \)). Then, in particular,

\[
(1) \quad F_{X_1}(x_1) \geq G_{Y_1}(x_1), \quad \forall x_1 \in \mathbb{R}
\]

and

\[
(2) \quad F_{X_2|X_1}(x_2|x_1) \geq G_{Y_2|Y_1}(x_2|x_1), \quad \forall x_2 \in \mathbb{R}
\]

s.t. \( x_1 \leq y_1 \)

Let \( f_{\theta_1, \theta_2}(x_1) = \int I_{\{x_2 > \theta_2\}}(x_2) \left[ \frac{(x_1 - \theta_1)^2}{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2} \right]^\lambda - 2 \right.
\]

\[\times (x_2 - \theta_2) dF_{X_2|X_1}(x_2|x_1)
\]

and

\[g_{\theta_1, \theta_2}(y_1) = \int I_{\{x_2 > \theta_2\}}(x_2) \left[ \frac{\sqrt{y_1 - \theta_1^2}}{\sqrt{(y_1 - \theta_1^2)^2 + (x_2 - \theta_2)^2}} \right]^\lambda - 2 \]

\[\times (x_2 - \theta_2) dG_{Y_2|Y_1}(x_2|y_1)
\]

Then for \( x_1 \leq y_1 \), \( f_{\theta_1, \theta_2}(x_1) \leq g_{\theta_1, \theta_2}(y_1) \). This follows by

\[I_{\{x_2 > \theta_2\}} \left[ \frac{(x_1 - \theta_1)^2}{(x_2 - \theta_2)^2} \right]^\lambda - 2 (x_2 - \theta_2), \text{ as a function of } x_2 \text{ is in } \mathfrak{A}^*(\mathbb{R}), \text{ and by (2) above.}
\]

Now by Theorem 3.3 (v), \( \Xi \) an upward kernel \( k \) on \( \mathbb{R}^2 \) s.t.
\[ Q_1 = P_1^k \] where \( P_1 \) is the measure corresponding to \( F_{x_1} \) and \( Q_1 \) is the measure corresponding to \( G_{y_1} \).

Hence,
\[
\int f_{\theta_1, \theta_2, \lambda}(x_1) dF_{x_1}(x_1) = \int \int f_{\theta_1, \theta_2, \lambda}(x_1) k(x_1, dy_1) P_1(dx_1)
\]
\[ Y_1 \geq x_1 \]
\[
\leq \int \int g_{\theta_1, \theta_2, \lambda}(y_1) k(x_1, dy_1) P_1(dx_1)
\]
\[ Y_1 \geq x_1 \]
\[
= \int \int g_{\theta_1, \theta_2, \lambda}(y_1) k(x_1, dy_1) P_1(dx_1)
\]
\[ = \int g_{\theta_1, \theta_2, \lambda}(y_1) Q_1(dy_1) \]

But this is precisely
\[
\int \int \sqrt{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2} \lambda^{-2} (x_2 - \theta_2) dF(x_1, x_2)
\]
\[ x_2 > \theta_2 \]
\[
\int \int \sqrt{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2} \lambda^{-2} (x_2 - \theta_2) dG(x_1, x_2)
\]
\[ x_2 > \theta_2 \]

A similar argument yields
\[
\int \int \sqrt{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2} \lambda^{-2} (\theta_2 - x_2) dF(x_1, x_2)
\]
\[ x_2 < \theta_2 \]
\[
\int \int \sqrt{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2} \lambda^{-2} (\theta_2 - x_2) dG(x_1, x_2)
\]
\[ x_2 < \theta_2 \]
Assume now

(1) \( F(x_2) \geq G(y_2), \forall x_2 \in \mathbb{R} \)

and

(2) \( F_{x_2|x_1}(x_2|x_1) \geq G_{y_2|x_2}(y_2|x_2), \forall x_1 \in \mathbb{R} \)

s.t. \( x_2 \leq y_2 \).

An argument similar to the one given above then yields Equation (3.6).

Hence, for \( \theta_F = \{\theta_{1F}, \theta_{2F}\} \) and \( \theta_G = \{\theta_{1G}, \theta_{2G}\} \), Equation (3.6) implies \( \theta_{1F} \leq \theta_{1G} \) and Equation (3.7) yields \( \theta_{2F} \leq \theta_{2G} \). Thus, we have established \( \theta_F \leq \theta_G \) under the strong conditional ordering of \( F \) and \( G \).

Remarks: (1) If we assume \( F_{x_2|x_1}(x_2|x_1) \) is stochastically monotonic, i.e., \( F_{x_2|x_1}(x_2|x_1) \geq F_{x_2|x_1}(x_2|y_1) \),

\( \forall x_1 \leq y_1 \), then the above ordering need be defined for \( x_1 = y_1 \) \( (x_2 = y_2) \). Hence, we are reduced to the conditional ordering and axiom (Bl") is then verified.

(2) In the case of independence, i.e.,

\( F_{x_1,x_2}(x_1,x_2) = F_{x_1}(x_1)F_{x_2}(x_2) \) and

\( G_{y_1,y_2}(y_1,y_2) = G_{y_1}(y_1)G_{y_2}(y_2) \), then we are reduced to the marginal ordering and axiom (Bl') is then verified.
It is not appropriate to use the strong conditional ordering to construct a location region. For having obtained $H_{\alpha X_2 | X_1}$ for a particular value of $X_1$, say $X_1 = x_1^*$, we would have to use this $H_{\alpha X_2 | X_1}$ to approximate not only $F_{\alpha X_2 | X_1}$ for $X_1 = x_1^*$ but also all distributions $F_{\alpha X_2 | X_1}$ for $x_1 < x_1^*$. This would, in general, yield location intervals for $F_{\alpha X_2 | X_1}$, $\forall x_1 < x_1^*$ which are strictly larger than those obtained by the conditional ordering. Thus, there would be points in these intervals which do not satisfy the axioms of location.

Hence, with the above modification of axiom (B4') and the stochastic ordering given in axiom (B1''), the value of $\theta = \theta_{\lambda}(\lambda)$ which minimizes the double integral

$$\iint ||X - \theta||^\lambda dF(x), \quad \lambda > 1$$

satisfies the bivariate axioms of location. This result is not surprising in view of the nature of the Euclidean norm. The norm "combines" information from the two components of a random vector $X$, and hence, it is not sufficient merely to view the vector marginally as is done in Chapter II. In the original development of the location region, no use is made of any correlation which may exist among the components of $X$. With the Euclidean norm though, it is
necessary to incorporate such information into the stochastic ordering axiom by making the axiom more restrictive. That is, information on the ordering of the marginal distributions is not enough; the appropriate ordering of the conditional distributions (which thus takes intercorrelation among the components into account) has thus been established to be the type of ordering found in the hypotheses of Theorem 3.8. If stochastic monotonicity is assumed, axioms \((B1'')\) and \((B1'''')\) are equivalent and so we have a class of location parameters in the location region obtained in this chapter; if independence is assumed, axioms \((B1')\) and \((B1'')\) are equivalent and so we have a class of location parameters in the location region obtained in Chapter II.

We now extend a result of Bickel and Lehmann (1975) to a special multivariate result.

Theorem 3.6: Consider the class \(\mathcal{F}\) of n-dimensional distribution functions which are symmetric in the sense that

\[ F \in \mathcal{F} \iff F(x_1, \ldots, x_n) = F(x_{\pi(1)}, \ldots, x_{\pi(n)}) \]

where \(\pi \in S_n\), the symmetric group of order n. Suppose that \(\mathcal{F}\) is convex, contains all point masses, is closed under changes of scale of the form \((\sigma x_1, \ldots, \sigma x_n)\), \(\sigma > 0\) and contains a c.d.f. \(F^0\) which is symmetric about zero.

Let \(W: \mathbb{R} \to \mathbb{R}\) be a positive, even, convex function which
is twice differentiable and satisfies the growth conditions of Theorem 3.1.

Let $\mu: \mathbb{R}^n \to \mathbb{R}^n$ be a location parameter; i.e., $\mu(F)$ is the value of $\theta$ which minimizes

$$\int_{\mathbb{R}^n} W(||x-\theta||)dF(x).$$

Let $\Sigma$ denote the $n \times n$ diagonal matrix given by

$$\Sigma = \text{diag}\{\sigma, \ldots, \sigma\}, \sigma > 0$$

For any given distribution $F \in \mathcal{F}$, denote by $F_\Sigma$ the distribution defined by $F_\Sigma(x) = F(\Sigma^{-1}x)$. Assume $\mu(F_\Sigma) = \Sigma \mu(F), \forall F \in \mathcal{F}$.

Finally, assume

$$\frac{3}{\partial t_i} \int_{\mathbb{R}^n} W'(||x-t||) ||x-t||^{-1} \frac{1}{||x-t||} (x-t) dF(x) |_{F=F_0} =$$

$$\int_{\mathbb{R}^n} \left\{ \frac{3}{\partial t_i} \right\} W'(||x-t||) ||x-t||^{-1} \frac{1}{||x-t||} (x-t) dF(x) |_{F=F_0}$$

$$- \int_{\mathbb{R}^n} \left\{ W''(||x-t||) ||x-t||^{1} \frac{1}{||x-t||} (x-t) + W'(||x-t||) ||x-t||^{-1} (x-t) \right\} dF^0(x)$$

for $i = 1, 2, \ldots, n$, and

$$\int_{\mathbb{R}^n} \left\{ W''(||x||) ||x||^{1} \frac{1}{||x||} (x) + W'(||x||) ||x||^{-1} (x) \right\} ||x||^{-3} \frac{1}{||x||} dF^0(x) < \infty$$

for all $t \in \mathbb{R}^n$ and all $i = 1, 2, \ldots, n$.

Then, $W(x) = c|x|^\lambda$ for some $\lambda > 1, c > 0$. 
Proof: The measure \( u(F) \) is the solution of

\[
\int_{\mathbb{R}^n} W'(|x-\theta|)|x-\theta|^{-1}c'(x-\theta)\,dF(x) = 0
\]

where \( c \in \mathbb{R}^n \) is an arbitrary fixed vector. In particular then, \( u(F) \) is the solution of

\[
\int_{\mathbb{R}^n} W'(|x-\theta|)|x-\theta|^{-1}1'(x-\theta)\,dF(x) = 0.
\]

Let \( f(\theta,F) = f(u(F),F) = \int_{\mathbb{R}^n} W'(|x-\theta|)|x-\theta|^{-1}1'(x-\theta)\,dF(x) \)

Now

\[
f(\theta,F) = 0
\]

\[
\begin{align*}
\Delta f = \left( \frac{\partial f}{\partial \theta} \right)' \, d\theta + \frac{\partial f}{\partial F} & = 0 \\
\left. \frac{df}{dF} \right|_{F=F^0} & = \left( \frac{\partial f}{\partial \theta} \right)' \bigg|_{\theta=0} \\
& + \frac{\partial f}{\partial F} \bigg|_{F=F^0} = 0
\end{align*}
\]

where

\[
\left( \frac{\partial f}{\partial \theta} \right)' = \left( \frac{\partial f}{\partial \theta_1}, \ldots, \frac{\partial f}{\partial \theta_n} \right)\bigg|_{\theta=\theta}
\]

But

\[
\left. \frac{\partial f}{\partial \theta_i} \right|_{\theta=\theta} = -\left( W''(|x|)|x|1'x + W'(||x||)|x|^{-3}x_i \right) dF^0(x)
\]

= K (say) \( Wi = 1, \ldots, n \) because \( F^0 \) is symmetric in its arguments by assumption.
So \( KL \frac{d\delta}{dF} = -\frac{\partial f}{\partial F}\bigg|_{F=F_0} \).

Now let \( \delta_x \in \mathfrak{F} \) be the distribution which puts unit mass at the point \( x \). Then

\[
\frac{1}{1'} \left[ \lim_{\varepsilon \to 0} \frac{u((1-\varepsilon)F_0^0 + \varepsilon \delta_x(F_0^0))}{\varepsilon} \right] = -K^{-1} \frac{\partial f}{\partial F}\bigg|_{F=F_0^0}
\]

Now,

\[
\frac{\partial f}{\partial F}\bigg|_{F=F_0^0} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{ \int_n W'(||x-0||) ||x-0||^{-1/2} (y-0)d(F_0^0(y)) + \varepsilon \delta_x \right\}
\]

\[
- \int_n W'(||y-0||) ||y-0||^{-1/2} (y-0)dF_0^0(y)
\]

\[
= -\int_n W'(||x||) ||x||^{-1/2} ydF_0^0(y) + W'(||x||) ||x||^{-1/2} x
\]

\[
= W'(\sqrt{n}||x||) \sqrt{n} \text{ sgn } x,
\]

since \( x = x \cdot 1 \) in order that \( x \in \mathfrak{F} \). (x \in \mathfrak{F})

Hence,

\[
\frac{1}{1'} \left[ \lim_{\varepsilon \to 0} \frac{u((1-\varepsilon)F_0^0 + \varepsilon \delta_x) - u(F_0^0)}{\varepsilon} \right] = -K^{-1} \sqrt{n} \text{ sgn } x W'(\sqrt{n}||x||)
\]
Similarly,
\[
\lim_{\varepsilon \to 0} \frac{\mu \left\{ \left[ (1-\varepsilon)F^0 + \varepsilon \delta_x \right] - \mu(F^0) \right\}}{\varepsilon} = -K_\sigma \sqrt{n} \sgn x W'(\sigma\sqrt{n}|x|)
\]

where
\[
K_\sigma = - \int \left\{ w^n(||x||) ||y||l'y \right\} \left\{ w'(||y||) ||y||l'y \right\}
\]

\[
\frac{1}{\sigma^2} \int \frac{\{ \sigma^2 w^n(\sigma||y||)||y||l'y \}}{\sigma} \left\{ w'(\sigma||y||)||y||l'y \right\} \frac{-3 \sigma y_i dF^0(y)}{\sigma}
\]

for all \( i = 1, 2, \ldots, n \).

But, since \( \mu(F^0) = E \mu(F) \), we have
\[
\frac{\sqrt{n} \sgn x W'(\sqrt{n}|x|)}{K} = \frac{\sqrt{n} \sgn x W'(\sigma\sqrt{n}|x|)}{K_\sigma}
\]

\[
\frac{\sigma W'(\sqrt{n}|x|)}{K} = \frac{W'(\sigma\sqrt{n}|x|)}{K_\sigma}
\]

Differentiating w.r.t. \( x (x>0) \) yields
\[
\frac{W''(\sqrt{n}\ x)}{K} = \frac{W''(\sigma\sqrt{n}\ x)}{K_\sigma}
\]

\[
\frac{W''(\sqrt{n}\ \sigma x)}{W''(\sqrt{n}\ x)} = \frac{W'(\sqrt{n}\ \sigma x)}{\sigma W'(\sqrt{n}\ x)}, \ x>0
\]

Now let \( x = \frac{1}{\sqrt{n}} \). Then
\[
\frac{W''(\sigma)}{W'(\sigma)} = \frac{1}{\sigma} \frac{W''(1)}{W'(1)}
\]

\[
\Rightarrow \log W'(\sigma) = \frac{W''(1)}{W'(1)} \log \sigma + c^*
\]

\[
W'(\sigma) = e^{c^*} |\sigma| \frac{W''(1)}{W'(1)}
\]

\[
W(\sigma) = \frac{W'(1)e^{c^*} |\sigma| \frac{W''(1)}{W'(1)} + 1}{W'(1) + W''(1)}
\]

= \sigma|\sigma|^\lambda, \ c > 0, \ \lambda \geq 1.

Remarks: 
(1) The requirement that $F(\mathbb{F})$ be symmetric in its arguments means that $F$ is the n-dimensional c.d.f. of a random vector, the components of which are exchangeable random variables.

(2) This requirement of exchangeability is a result of using the Euclidean norm. This norm is symmetric in its components $\Rightarrow F(\mathbb{F})$ should be too.

(3) $F(\mathbb{F})$, symmetric in its arguments $\Rightarrow$ the location region for $F$ will be a line segment (possibly infinite) lying on the line $x_1 = x_2 = \ldots = x_n$.

(4) Under certain regularity conditions on $F'$ we could probably get a result similar to that of Groeneveld and Meeden (1981).

(5) If we regard $F$ as the posterior distribution of some n-dimensional location parameter $\theta$, given
the data $x, \mu(F)$ can be thought of as the Bayes estimator of $\theta$. All the foregoing results then apply to this estimator.
IV. APPLICATIONS

Different stochastic orderings have been used to construct a location region in the plane for a given bivariate distribution function $F$. We can now use this region for two purposes: to characterize the degree of asymmetry in $F$ and to establish order relationships among distributions.

More specifically, in this chapter two applications of the location region $L_P$ will be given. Having obtained $L_P$ via either the marginal or the conditional ordering, we know $L_P$ is a closed convex set in the plane. It is intuitively clear that in either case, the larger $L_P$ is, the more $F$ deviates from symmetry. In the first section of this chapter, this idea will be made rigorous using results from convex analysis. In addition, for $L_P$ obtained by way of the marginal ordering, an asymptotic result concerning the size of $L_P$ will be presented. The second section will deal with the use of $L_P$ in both a point-wise as well as an overall manner to order distribution functions. It is possible to define a stochastic ordering of two distributions $F$ and $G$ by using subsets of location parameters; that is, subsets of $L_F$ and $L_G$. Finally, we can order the components of a random vector by using the location region as a whole.
A. Measures of Asymmetry

Having defined the location region for a bivariate distribution function $F$, it is possible to use this set to characterize the degree of asymmetry of $F$. Since the location region under either stochastic ordering is a closed convex set in the plane, many results from convex analysis are applicable. Kelly and Weiss (1979) have developed the necessary tools needed for our purpose.

The following propositions and definitions are appropriate for this discussion: (for the greatest generality, $\mathbb{R}^n$ will be used throughout).

**Notation**: For $a \in \mathbb{R}^n$, $a \neq 0$, $k \in \mathbb{R}$, a hyperplane in ortho­gonal to $a$ will be denoted by $\{x: a'x = k\}$. As $k$ varies through all real values, all hyperplanes orthogonal to the vector $a$ are obtained. ($\mathbb{R}^n$ will denote Euclidean $n$-space, i.e., $\mathbb{R}^n$ with $\| \cdot \|$, the Euclidean metric.) Capital letters such as $A$ will represent points in $\mathbb{R}^n$. If $A \in \mathbb{R}^n$, $a$ will denote the vector of $n$ components of $A$ with respect to some basis for $\mathbb{R}^n$.

For $x, y \in \mathbb{R}^n$, the distance between $x$ and $y$, denoted $d(x, y) = d(X, Y)$ will be defined by

$$d(x, y) = \|x - y\|.$$
Definition 4.1: (open and closed half-spaces)

If $a'x = k$, $a \neq 0$ is the representation of a hyperplane $\mathcal{H}$, then the graphs of $a' \cdot x > k$ and $a' \cdot x < k$ are the opposite sides or opposite open half-spaces of $\mathcal{H}$. The graphs of $a' \cdot x > k$ and $a' \cdot x < k$ are the opposite closed half-spaces of $\mathcal{H}$. The hyperplane $\mathcal{H}$ is the face of the four half-spaces.

Definition 4.2: (supporting hyperplane)

A hyperplane $\mathcal{H}$ in $\mathbb{R}^n$ is a supporting hyperplane of a set $\mathcal{A}$ if $\mathcal{H}$ intersects the closure of $\mathcal{A}$ and a closed side of $\mathcal{H}$ contains $\mathcal{A}$. Points in the set $\mathcal{H} \cap \overline{\mathcal{A}}$ are said to be contact points of $\mathcal{H}$ with $\overline{\mathcal{A}}$, and $\mathcal{H}$ is said to support $\mathcal{A}$ at each contact point.

Definition 4.3: (supporting half-space)

An open or closed half-space supports a set $\mathcal{A}$ if it contains $\mathcal{A}$ and its face supports $\mathcal{A}$.

Definition 4.4: (slabs)

The open slab between two parallel hyperplanes $\mathcal{H}_1$ and $\mathcal{H}_2$ is the intersection of the $\mathcal{H}_1$ side of $\mathcal{H}_2$ with the $\mathcal{H}_2$ side of $\mathcal{H}_1$. The closed slab between $\mathcal{H}_1$ and $\mathcal{H}_2$ is the intersection of the corresponding closed half-spaces. The hyperplanes $\mathcal{H}_1$ and $\mathcal{H}_2$ are the faces of both slabs, and the distance between these faces is the width of both slabs. A direction of a slab is a direction normal to both faces.
Definition 4.5: (supporting slabs)
An open or closed slab supports a nonempty set $\mathbf{A}$ if it contains $\mathbf{A}$ and both its faces support $\mathbf{A}$.

Definition 4.6: (width of a set)
A nonempty set $\mathbf{A}$ in $\mathbb{R}^n$ has width zero in direction $\mathbf{u}$ if $\mathbf{A}$ is contained in a hyperplane with direction $\mathbf{u}$.
The set has width $h$ in the direction $\mathbf{u}$ if there is a slab of width $h$ and direction $\mathbf{u}$ that supports $\mathbf{A}$. The width of $\mathbf{A}$ in the direction of $\mathbf{u}$ will be denoted by $W(\mathbf{u})$.

Two results which follow from these definitions are now given:

Theorem 4.1: A nonempty bounded set $\mathbf{A}$ in $\mathbb{R}^n$ has a width in every direction.

Corollary: If a set $\mathbf{A}$ is not contained in a hyperplane with direction $\mathbf{u}$ and $\mathbf{A}$ is bounded, then there are exactly two hyperplanes with direction $\mathbf{u}$ that support $\mathbf{A}$. The closed slab between these hyperplane supports $\mathbf{A}$, and its width is the nonzero width of $\mathbf{A}$ in the direction $\mathbf{u}$.

We now wish to develop an explicit expression for $W(\mathbf{u})$, the width function, as well as prove some important properties that $W(\mathbf{u})$ possesses.

Let $\mathbf{L}$ be a line through the origin. Then, the orthogonal projection $\mathbf{y} = f(\mathbf{x})$ of a set $\mathbf{R}$ into $\mathbf{L}$ has a simple explicit
representation (see Figure 4.1). Let \( u \) be a vector in the
direction of \( \mathcal{L} \) s.t. \(|u| = 1\). Thus, \( x \in \mathcal{L} \) iff

\[
x = \eta u, \quad \text{for some } \eta \in \mathbb{R}.
\]

If \( p \in \mathbb{R}, \) then the hyperplane through \( p \) and perpendicular
to \( \mathcal{L} \) has equation

\[
\mathcal{H}: (x-p) \cdot u = 0
\]

The intersection of \( \mathcal{H} \) with \( \mathcal{L} \) corresponds to the \( \eta \)-solution
of

\[
(\eta u - p) \cdot u = 0
\]

\( \Rightarrow \eta = u \cdot p. \)

\( \Rightarrow f(p) = (u \cdot p)u \) and so in general \( f \) is the mapping

\[
f(x) = (x \cdot u)u, \quad x \in \mathbb{R}. \tag{4.1}
\]

Now, suppose that \( \mathcal{H}_A \) and \( \mathcal{H}_B \) are two hyperplanes which
are perpendicular to \( \mathcal{L} \) at \( A \) and \( B \), with \( O \) between \( A \) and \( B \).
Let \( P \) be a contact point in \( \mathcal{H}_A \), \( Q \) a contact point in \( \mathcal{H}_B \).
Let \( u, v \) be opposite unit vectors in \( \mathcal{L} \) with directions \( OA, \)
\( OB \) respectively.

The width of \( \mathcal{K} \) in directions \( u, v \) is

\[
d(\mathcal{H}_A, \mathcal{H}_B) = d(A, B) = d(O, A) + d(O, B)
\]

From Equation (4.1), \( A = f(P) = (u \cdot p)u \)

\( \Rightarrow d(O, A) = ||(u \cdot p)u|| = u \cdot p \) since \( u \cdot p > 0. \)
Figure 4.1. Convex set $R$ in the plane with supporting lines $\mu_A$ and $\mu_B$. 
Similarly, $d(0,B) = |(v' \cdot q) v| = v' \cdot q$

$$d(N_A, N_B) = u' \cdot P + v' \cdot q \quad (4.2)$$

If $R$ is closed, then $P, Q \in R$ and $u' \cdot P$ and $v' \cdot q$ in Equation (4.2) can be described in a more intrinsic way. Let $H_0$ be the hyperplane perpendicular to $L$ at $0$. Hence, for each $X \in R$, $u' \cdot x = \pm d(0,y)$ where $Y = f(X)$ is the orthogonal projection of $R$ into $L$. Also, $u \cdot x$ is positive, zero, or negative according as $X$ is in the $A$ side of $H_0$, is in $H_0$, or is in the $B$ side of $H_0$. Thus,

$$u' \cdot P = \sup\{u' \cdot x : x \in R\}$$

Similarly, $v' \cdot q = \sup\{ v' \cdot x : x \in R\}$. This yields the following definition.

**Definition 4.7: (support function)**

If $R$ is a nonempty, compact set, the support function of $R$ is the function $H$ defined at each unit vector $u$ by

$$H(u) = \sup\{u' \cdot x : x \in R\}$$

We note that for fixed $u$, the function $u \cdot x$ is continuous at all $X$ and hence achieves a maximum on any compact set $R$. Thus, the support function $H$ of such a set has a real function value at each point $u$ on the unit sphere $S(0,1)$.

We can summarize this in the following theorem.
Theorem 4.2: If two parallel supporting hyperplanes to a compact set \( \mathcal{R} \) have unit normal direction \( u \), and are such that 0 is between them, then the width of \( \mathcal{R} \) in the direction of \( u \) is given by

\[
W(u) = H(u) + H(-u).
\]

The artificial condition about the position of the origin can be removed via the following theorems.

Theorem 4.3: If \( \mathcal{R} \) is a nonempty compact set and \( \mathcal{R}' \) is the \( a \)-translation of \( \mathcal{R} \), then the width functions \( W \) and \( W' \) of these sets are the same, and the support functions \( H \) and \( H' \) are related by \( H'(u) = H(u) + a'u \).

Theorem 4.4: If \( \mathcal{R} \) is a nonempty, compact set with support function \( H \) and width function \( W \) defined on \( S(0,1) \), then

\[
W(u) = H(u) + H(-u).
\]

Finally, two important properties of the width function \( W \) are given in the following theorem and its corollary.

Theorem 4.5: The width function for a nonempty, compact set is continuous on \( S(0,1) \).

Corollary: If \( \mathcal{A} \) is a nonempty compact set in \( \mathbb{R}^n (n>1) \), then its width function \( W \) assumes a minimal value in some direction \( u_1 \) and a maximal value in some direction \( u_2 \).
$W(u_1) < W(u_2)$, then $W$ assumes every value intermediate to $W(u_1)$ and $W(u_2)$. If $u>2$, $W$ assumes every intermediate value infinitely often and at least once in the plane of $u_1$, $u_2$.

We can use these results now to characterize the degree of asymmetry in a bivariate distribution $F$. Suppose $F$ has bounded support. Then the location set $L_F$ of $F$ is a nonempty, compact, convex set in the plane. Intuitively, the width of $L_F$ in a given direction $u$ is a measure of how far $F$ deviates from symmetry in that direction. If $W(\cdot)$ denotes the width function of $L_F$, then $W(u) = 0$, $\forall u \in S(0,1)$ implies $L_F$ is a single point $\emptyset$, and hence $F$ is symmetric about $\emptyset$. If $u^* \in S(0,1)$ s.t. $W(u^*) = 0$, then $F$ is symmetric about a line perpendicular to $u^*$. The converse of each of these two statements is true, too. Thus, the width function contains all the information necessary to describe the asymmetric nature of $F$.

We summarize these cases below:

1. $W(u) = 0 \ \forall u \in S(0,1)$ iff $F$ is symmetric about a point.

2. $W(u) = 0$ for exactly two values of $u \in S(0,1)$ iff $F$ is symmetric about a line perpendicular to $u$.

3. $W(u) > 0 \ \forall u \in S(0,1)$ iff $F$ is asymmetric in every direction and $L_F$ is a plane set.
So the number of zeros of $W$ serves first of all to categorize $F$: since $u \in \mathcal{S}(), l) u = \begin{cases} \cos a, \\ \sin a \end{cases}, a \in (0, 2\pi)$ let $W^*(a) = W(u)$. A plot of $W^*$ then provides information on how $F$ deviates from symmetry in each direction (see Figure 4.2).

We now consider an asymptotic result obtained by Doksum (1975). We assume $F$ is a univariate distribution function whose support is the interval $(a, b)$, $-\infty < a \leq b < \infty$. We extend the definition of $\hat{\theta}_n(x)$ (see Chapter I) from $[x(1), x(n)]$ to $[a, b]$ by

$$
\hat{\theta}_n(x) = \frac{1}{2}(x + x(n)), \quad X(n) \leq x < b
$$

$$
\hat{\theta}_n(x) = \frac{1}{2}(x + x(1)), \quad a < x \leq X(1).
$$

We also extend the definition of $\theta_F(x)$ from $S(F) = (a, b)$ to $[a, b]$ by $\theta_F(x) = \frac{1}{2}[X - F^{-1}(F(x))]$, $x \in [a, b]$. Then

$$
\sqrt{n} \left[ \hat{\theta}_n(x) - \theta_F(x) \right]
$$

is a member of the class $D$ of functions on $[a, b]$ which are right-continuous and have left-hand limits. On this space, the Skorohod topology will be used. Doksum proves the following theorem.

**Theorem 4.6:** Suppose $F(t)$ has a continuous, positive derivative $f$ on $S(F) = [a, b]$, $-\infty < a < b < \infty$. Let $W_0(t)$ denote a Brownian Bridge on $[0, 1]$, i.e., $W_0(t)$ is a Gaussian process with mean zero and covariance function $s(1-t)$, $0 \leq s \leq t \leq 1$. Then

$$
n[\hat{\theta}_n(x) - \theta_F(x)]
$$

converges weakly to the Gaussian process
Figure 4.2. Examples of width functions in the case of (a) symmetry about a point, (b) symmetry about a line in direction $\alpha^\ast$, (c) complete asymmetry
\[ \frac{W_0(F(x)) + W_0(1-F(x))}{2f(F^{-1}(1-F(x)))} \]

Now consider \( L_F \), the location region for the bivariate distribution function \( F \), obtained by way of the marginal ordering. Then, for each \( \alpha \in (0, 2\pi) \), the empirical width function \( \hat{W}_n^*(\alpha) \) is the width of the estimated location region \( \hat{b}_n \) in the direction \( \alpha \). Hence, for each \( \alpha \in (0, 2\pi) \), two points \( x_\alpha, y_\alpha \in S(F_{\alpha X_1}) \) s.t.

\[ \hat{W}_n^*(\alpha) = [\hat{\theta}_{an}(x_\alpha) - \hat{\theta}_{an}(y_\alpha)], \]

where \( \hat{\theta}_{an}(\cdot) \) is the empirical function of symmetry for \( F_{\alpha X_1} \).

By Theorem 4.6, for \( x_\alpha < y_\alpha \), and \( F_{\alpha X_1}' = f_{\alpha X_1}' \),

\[ \sqrt{n} \left( \begin{array}{c} \hat{\theta}_{an}(x_\alpha) - \theta_{F_{\alpha X_1}}(x_\alpha) \\ \hat{\theta}_{an}(y_\alpha) - \theta_{F_{\alpha X_1}}(y_\alpha) \end{array} \right) \xrightarrow{\text{asymp}} N_2 \left( 0, \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right), \]

where

\[ a_{11} = \frac{f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(x_\alpha)))}{2[f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(x_\alpha)))]^2}; \]

\[ a_{12} = a_{21} = \frac{f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(x_\alpha)))f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(y_\alpha)))}{f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(x_\alpha)))f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(y_\alpha)))}; \]

\[ a_{22} = \frac{f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(y_\alpha)))}{2[f_{\alpha X_1}(F_{\alpha X_1}^{-1}(1-F_{\alpha X_1}(y_\alpha)))]^2}. \]

We summarize these results in the following theorem:
Theorem 4.7: If $F$ is an absolutely continuous bivariate distribution function with compact support, and if $\hat{W}^*(\cdot)$ denotes the empirical width function for $F$, then for each $\alpha \in (0, 2\pi]$ $x_\alpha, y_\alpha \in S(F_{\alpha X_1})$ s.t. $\sqrt{n}[\hat{W}^*(\alpha) - (\theta_{F_{\alpha X_1}}(x_\alpha) - \theta_{F_{\alpha X_1}}(y_\alpha))]$ is asymptotically normal with mean zero and variance

$$\frac{F_{\alpha X_1}(x_\alpha)}{2f_{\alpha X_1}(F^{-1}_{\alpha X_1}(1-F_{\alpha X_1}(x_\alpha)))} + \frac{F_{\alpha X_1}(y_\alpha)}{2f_{\alpha X_1}(F^{-1}_{\alpha X_1}(1-F_{\alpha X_1}(y_\alpha)))} - \frac{F_{\alpha X_1}(x_\alpha)}{f_{\alpha X_1}(F^{-1}_{\alpha X_1}(1-F_{\alpha X_1}(x_\alpha)))f_{\alpha X_1}(F^{-1}_{\alpha X_1}(1-F_{\alpha X_1}(y_\alpha)))}$$

for $x_\alpha \leq y_\alpha$

In particular, when $F_{\alpha X_1}$ is symmetric about zero,

$$\sqrt{n} \hat{W}^*(\alpha) \xrightarrow{\text{asy.}} N(0, \frac{F_{\alpha X_1}(x_\alpha)}{2[f_{\alpha X_1}(x_\alpha)]^2} + \frac{F_{\alpha X_1}(y_\alpha)}{2[f_{\alpha X_1}(y_\alpha)]^2} - \frac{F_{\alpha X_1}(x_\alpha)}{f_{\alpha X_1}(x_\alpha)f_{\alpha X_1}(y_\alpha)}$$

for $x_\alpha \leq y_\alpha$

The width function contains all information on how $F$ deviates from symmetry. Thus, the graph of $W$ is useful in determining the nature of asymmetry of $F$; for example, the number of zeros of $W$ indicate which of the above categories $F$ falls into. On the other hand, if a single measure of asymmetry is required, various functions of $W$ can be used.

We can use $W$ (or $W^*$) to obtain some overall descriptive measures of asymmetry for $F$ as follows:
(1) the average value of $W^*$:
\[ \mathcal{R} = \frac{1}{2\pi} \int_{0}^{\pi} W^*(\alpha) d\alpha \]

(note: $W^*$ being continuous on a compact set implies $W^*$ is Riemann integrable)

(2) diameter of $S = \sup_{0<\alpha<2\pi} W^*(\alpha) = \max_{u \in S(0,1)} W(u) = \delta$

There is some justification for considering the diameter of $L_F$. Let $\mathcal{I}^*$ be the class of all distribution functions with compact support.

Let $C[0,2\pi]$ be the class of continuous functions on $[0,2\pi]$. We have a mapping $\mathcal{I}^* \rightarrow C[0,2\pi]$ where each distribution function $F$ with compact support in $\mathcal{I}^*$ is mapped into its width function $W^*_F(\cdot)$ in $C[0,2\pi]$.

Consider the following metric:
\[ d(W^*_F, W^*_G) = \max\{|W^*_F(x)-W^*_G(x)| : x \in [0,2\pi]\}. \]

Let $F^\theta$ be any distribution function symmetric about $\theta$, $F^\theta$ with compact support. Then $W^*_F^\theta(x) = 0$, $\forall x \in [0,2\pi]$. For arbitrary $F \in \mathcal{I}^*$, we can then measure how far $F$ is from symmetry by
\[ d(F^\theta, W^*_F) = \max\{|W^*_F(x)-W^*_F(x)| : x \in [0,2\pi]\} \]
\[ = \max\{W^*_F(x) : x \in [0,2\pi]\} \]
\[ = \text{diameter } L_F \]
\[ = \delta \]
B. Ordering of Distributions

Having constructed the location region (under either the marginal or the conditional stochastic ordering), it is intuitively appealing to define a class of stochastic orderings in terms of points in the location set. The following discussion deals with the location region obtained using the marginal ordering but it is easily modified for use with the location region obtained using the conditional ordering.

Let $\mathcal{F}$ be the class of all bivariate distribution functions which are continuous and increasing in each argument. The location region $\bar{L}_F$ for each $F \in \mathcal{F}$ is then a well-defined closed convex set in $\mathbb{R}^2$. Suppose $F, G \in \mathcal{F}$ are given. For each $\theta_F \in \bar{L}_F, \theta_G \in \bar{L}_G$, let $u_\theta_F(\cdot)$ and $u_\theta_G(\cdot)$ denote the respective direction functions. For each $\alpha \in (0, 2\pi]$, let

$$\Omega_{F,G}^\alpha = \{(\theta_F, \theta_G) \in \bar{L}_F \times \bar{L}_G : u_\theta_F(\alpha) = u_\theta_G(\alpha) \}$$

and

$$u_\theta_F(\alpha + \frac{3\pi}{2}) = u_\theta_G(\alpha + \frac{3\pi}{2})$$

Let $S_{F,G}^\alpha$ be any subset of $\Omega_{F,G}^\alpha$.

**Definition 4.8: (S$^\alpha$-ordering)**

We say $F \in \mathcal{F}$ is $S$-stochastically smaller than $G \in \mathcal{F}$ in direction $\alpha$ (denoted $F \prec_{S^\alpha} G$) iff
Thus, given \( L_F, L_G \) and nonempty \( \mathbb{S}_F, \mathbb{S}_G \), the above orderings are well-defined. On the other hand, it is evident that if we replace axiom (B1') by the above ordering, and then attempt to construct \( L_F \), we will in general obtain an empty set. For example, suppose

\[
\theta_F = \begin{pmatrix} m_{F_1}(\frac{1}{2}) \\ m_{F_2}(\frac{1}{2}) \end{pmatrix} = \begin{pmatrix} F_{X_1}(\frac{1}{2}) \\ F_{X_2}(\frac{1}{2}) \end{pmatrix} \epsilon L_F
\]

and

\[
\theta_G = \begin{pmatrix} m_{G_1}(\frac{1}{2}) \\ m_{G_2}(\frac{1}{2}) \end{pmatrix} = \begin{pmatrix} G_{X_1}(\frac{1}{2}) \\ G_{X_2}(\frac{1}{2}) \end{pmatrix} \epsilon L_G
\]

and \( \mathbb{S}_{F,G} = \{ (\theta_F, \theta_G) \} \). If we now use the ordering \( \mathbb{S}_0 \) to construct a location region for \( F \), say as we did in Chapter II it is clear that

\[
A_\alpha = \begin{pmatrix} m_{F_1}(\frac{1}{2}) \\ m_{F_2}(\frac{1}{2}) \end{pmatrix}, \forall \alpha \in (0, 2\pi]
\]

But, in general, \( R_{-\alpha}(A_\alpha) \neq A_0 \) so that \( L_F = \bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha}(A_\alpha) \) is empty.

The problem here is that we are trying to do too much. Having obtained \( L_F \) (using either the standard or marginal ordering) it is valid to parallel the univariate case and define the stochastic ordering \( \mathbb{S}_0 \). On the other hand, to then use this new ordering to replace axiom (B1') and then
construct the location region (as is possible in the univariate case) leads to the above contradiction. So, Definition 4.8, standing alone, does define a whole class of bivariate stochastic orderings. This example does provide insight into how one can actually construct a functional satisfying location axioms (B2)-(B4). It is not possible to show that either the standard or marginal ordering axiom is satisfied by this functional.

We can, however, use $S^\alpha$-ordering in the special case where $S^\alpha_{F,G}$ is a singleton to construct a location parameter as follows.

Consider now a weaker definition of stochastic ordering: Define $F < G$ iff $\theta_F \leq \theta_G$ where $(\theta_F, \theta_G) \in S^0_{F,G}$ are any measures of location for $F$ and $G$, respectively. Clearly, if $F$ is stochastically smaller than $G$ as first defined, then $F$ is stochastically smaller than $G$ as defined above. Hence, the set of location functionals satisfying this weaker definition is contained in $L_F$.

With this in mind, we can construct an element of this larger set by considering points of the form

$$\theta^*\{a\} = \frac{m_F^{-1}(a)}{m_F(a)}$$

If we let $CH(S)$ denote the convex hull generated by all points in a set $S$, consider $B(a, b) = CH_{R_\beta}(\theta^*\{a\})$ and let $\frac{1}{m_F(a)} = \theta^*\{b\} = \frac{1}{m_F(a)} = \frac{1}{m_F(b)} = \frac{1}{m_F(b)}$. 

138
Then \( \theta_F^* [a, b] \) can be seen to satisfy axioms (B2), (B3) and the modified axiom (B4') mentioned in the section describing the loss function approach. For \( \theta_F^* [a, b] \) to satisfy axiom (B4'), it is necessary and sufficient that for 
\[ D = \text{diag}\{d_1, d_2\}, \quad d_1 > 0, \quad d_2 > 0 \] 
we have
\[
R_\alpha DR_{\alpha} \theta_F^* [x] = \theta_F^* \alpha D x \quad \forall \alpha \in (0, 2\pi)
\]

But the following example shows that Equation (4.3) is not necessarily true, and what follows in a modification of the above procedure for the construction of a location parameter to satisfy axioms (B2)-(B4').

**Example 4.1:** To show \( R_\alpha DR_{\alpha} \theta_F^* [x] \neq \theta_F^* \alpha D x \) with \( R_\alpha \) a rotation through an angle \( \alpha \) and 
\[
D = \begin{pmatrix} d_1 & 0 \\ 0 & 1 \end{pmatrix}, \quad d_1 > 0.
\]
Consider
\[
f(x, y) = 3x^2; \quad 0 < x < 1, \quad 0 < y < 1. \quad \text{Then,}
\]
\[
F(x, y) = x^3 y; \quad 0 < x < 1, \quad 0 < y < 1
\]
\[
F_{x_1} (x) = x^3; \quad 0 < x < 1
\]
\[
F_{x_2} (y) = y; \quad 0 < y < 1
\]
Let
\[
D = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u \\ v \end{pmatrix} = D \begin{pmatrix} x \\ y \end{pmatrix}
\]
\[ f(u,v) = \frac{3}{10^6} u^2; \quad 0 < u < 100, \quad 0 < v < 1 \]

\[ F(u,v) = \frac{1}{10^6} u^3v; \quad 0 < u < 100, \quad 0 < v < 1 \]

\[ F_U(u) = \frac{1}{10^6} u^3; \quad 0 < u < 100 \]

\[ F_V(v) = v; \quad 0 < v < 1 \]

Let

\[ R_{\pi/4} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}; \quad R_{\pi/4}^{DR_{-\pi/4}} = \frac{1}{2} \begin{pmatrix} 101 & 99 \\ 99 & 101 \end{pmatrix}. \]

Let \( f \) be the density of \( R_{\pi/4} \{ x \} \):

\[ f(x,y) = \frac{3}{2}(x+y)^2 \quad \text{on } S. \]

(see Figure 4.3(a))

Let \( g \) be the density of \( R_{\pi/4}^{D} \{ x \} \):

\[ g(x,y) = \frac{3}{2 \cdot 10^6}(x+y)^2 \quad \text{on } S'. \]

(see Figure 4.3(b))

We have

\[ F_{X_1}(x) = \begin{cases} 
(x + \frac{\sqrt{2}}{2})^4, & -\frac{\sqrt{2}}{2} \leq x \leq 0 \\
\sqrt{2}x - x^4 + \frac{1}{4}, & 0 \leq x \leq \frac{\sqrt{2}}{2} 
\end{cases} \]
Figure 4.3. Support regions $S(a)$ and $S'(b)$ for Example 4.1
\[ F_{X_2}(y) = \begin{cases} 
y^4, & 0 \leq y \leq \frac{\sqrt{2}}{2} \\
\sqrt{2y} - (y - \frac{\sqrt{2}}{2})^4 - \frac{3}{4}, & \frac{\sqrt{2}}{2} \leq y \leq \sqrt{2}
\end{cases} \]

\[ G_{X_1}(x) = \begin{cases} 
\frac{1}{10^6}(x + \frac{\sqrt{2}}{2})^4, & -\frac{\sqrt{2}}{2} \leq x \leq 0 \\
\frac{1}{10^6}((x + \frac{\sqrt{2}}{2})^4 - x^4), & 0 \leq x \leq \frac{99\sqrt{2}}{2} \\
\sqrt{2}x - \frac{1}{10^6}x^4 - 74, & \frac{99\sqrt{2}}{2} \leq x \leq 50\sqrt{2}
\end{cases} \]

\[ G_{X_2}(y) = \begin{cases} 
\frac{1}{10^6}y^4, & 0 \leq y \leq \frac{\sqrt{2}}{2} \\
\frac{1}{10^6}[y^4 - (y - \frac{\sqrt{2}}{2})^4], & \frac{\sqrt{2}}{2} \leq y \leq 50\sqrt{2} \\
\sqrt{2}y - \frac{1}{10^6}(y - \frac{\sqrt{2}}{2})^4 - 75, & 50\sqrt{2} \leq y \leq \frac{101\sqrt{2}}{2}
\end{cases} \]

Consider now the medians \((a = b = \frac{1}{2})\):

\[
\text{med}_{F_{X_1}} = 0.1774782566 \\
\text{med}_{F_{X_2}} = 0.884585038 \\
\text{med}_{G_{X_1}} = 55.76880661 \\
\text{med}_{G_{X_2}} = 56.47591339
\]
Now,
\[ R_{\pi/2}D_{\pi/2}^{\Omega}R_{\pi/2} \left( \frac{1}{2} \right) = \frac{1}{2} \begin{bmatrix} 101 & 99 \\ 99 & 101 \end{bmatrix} \begin{bmatrix} 0.1774782566 \\ 0.8845850380 \end{bmatrix} \]
\[ = \begin{bmatrix} 52.74961134 \\ 53.45671812 \end{bmatrix} \]
\[ \neq \begin{bmatrix} 55.76880661 \\ 56.47591339 \end{bmatrix} \]
\[ = \frac{\theta}{\pi/2}D_{\Pi/2} \left( \frac{1}{2} \right) \]

We need the following preliminary result:

**Lemma 4.1:** Consider the transformation \( D = \begin{pmatrix} d_1 & 0 \\ 0 & 1 \end{pmatrix}, \quad d_1 > 0. \)

Let \( L: ax + by = c \) be any given line. Let \( D \left( \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \right) = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \). Then

\[ P(X \leq L) = P(U \leq D(L)). \]

**Proof:** Suppose \( a = P(aX_1 + bX_2 \leq c), \quad 0 \leq a \leq 1 \)

**Case (1):** \( b \neq 0 \)

Consider the transformation \( D = \begin{pmatrix} d_1 & 0 \\ 0 & 1 \end{pmatrix}, \quad d_1 > 0. \)

Now the line \( aU_1 + bU_2 = c \)

or equivalently, \( U_2 = \frac{a}{b}U_1 + \frac{c}{b} \)

is transformed by \( D \) into the line

\[ U_2 = \frac{a}{bd}U_1 + \frac{c}{b} \]

\[ \Rightarrow aU_1 + bdU_2 = cd \]
Hence, for \( \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} d_1 \\ 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} dX_1 \\ X_2 \end{pmatrix} \),

\[
P(aU_1 + bdU_2 \leq cd) = P(adX_1 + bdX_2 \leq cd)
\]

\[
= P(aX_1 + bX_2 \leq c)
\]

\[
= a.
\]

**Case (2):** \( b=0 \)

Suppose

\[
a = P(aX_1 \leq c), \ a \neq 0
\]

\[
= F_{X_1}(\frac{c}{a})
\]

The line \( aU_1 = c \) is transformed by \( D \) into the line \( aU_1 = cd \). Hence,

\[
P(aU_1 \leq cd) = P(adX_1 \leq cd)
\]

\[
= P(aX_1 \leq c)
\]

\[
= a.
\]

Suppose \( F \) is the c.d.f. of the random vector \( X \) and \( F_\alpha \) is the c.d.f. of \( R_\alpha (X) \). Consider the family of lines

\[
L_{F_\alpha} (u) = \{(x,y): y = m_{F_\alpha X_2} (u)\}, \ a \in (0,2\pi]
\]

for each \( u \in (0,\frac{1}{2}] \).

Here \( F_{\alpha X_2} \) is the marginal distribution with respect to \( X_2 \) of \( X = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = R_\alpha (X) \) and \( m_{F_{\alpha X_2}} (u) = \frac{1}{2}[F_{\alpha X_2}^{-1} (u) + F_{\alpha X_2}^{-1} (1-u)] \).
Let $B_F(u)$ denote the convex hull generated by all points of intersection of the lines \( \{ R_\alpha(L_F(u)) : \alpha \in (0, 2\pi) \} \) and let $\theta_F(u)$ denote the center of mass of $B_F(u)$:

\[
\theta_F(u) = K^{-1} \int \int x \, dA, \quad \text{where} \quad K = \int \int dA.
\]

**Theorem 4.8:** For each $u \in (0, 1/z)$, $\theta_F(u)$ is a location parameter.

**Proof:** We must show $\theta_F(u)$ satisfies the aforementioned axioms.

**Axiom (B2):** For $a = (a_1, a_2) \in \mathbb{R}^2$, let $G$ be the c.d.f. of $X + a$, $G_\alpha$ the c.d.f. of $R_\alpha(X + a) = R_\alpha(X) + R_\alpha(a) = R_\alpha(X) + (b_1\alpha, b_2\alpha)$, (say). By the univariate case, we know

\[
m_{F_\alpha X_2}(u) + b_2\alpha = m_{G_\alpha X_2}(u)
\]

\[\Rightarrow L_F(u) + R_\alpha(a) = L_{G_\alpha}(u); \forall \alpha, \forall u\]

\[\Rightarrow R_\alpha(L_F(u)) + a = R_\alpha(L_{G_\alpha}(u)), \forall \alpha\]

\[\Rightarrow B_F(u) + a = B_{G}(u)\]

\[\Rightarrow \theta_F(u) + a = \theta_G(u)\]

i.e., $\theta_{X+\alpha}(u) = \theta_{X}(u) + a$.
Axiom (B3): Consider first the transformation \( T_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \) (i.e., a reflection about the x-axis). Let \( G \) be the c.d.f. of \( T_0(X) \), \( G_\alpha \) the c.d.f. of \( R_\alpha(T_0(X)) = T_0(R_\alpha(X)) \). Now

\[
\{R_\alpha(L_{G_\alpha}(u)): \alpha \in (0,2\pi]\} = \{R_\alpha(L_{R_\alpha T_0 X}(u)): \alpha \in (0,2\pi]\}
\]

\[
= \{R_\alpha(L_{T_0 R_\alpha X}(u)): \alpha \in (0,2\pi]\}
\]

\[
= \{R_\alpha(T_0(L_{R_\alpha X}(u))): \alpha \in (0,2\pi]\}
\]

(by the invariance case)

\[
= \{T_0(R_\alpha(L_{R_\alpha X}(u))): \alpha \in (0,2\pi]\}
\]

\[
= T_0\{R_\alpha(L_{R_\alpha X}(u)): \alpha \in (0,2\pi]\}
\]

\[\rightarrow B_\alpha(u) = T_0(B_X(u))\]

\[\rightarrow \theta(u) = \theta(T_0X(u))\]

i.e., \( T_0\theta_X(u) = \theta_{T_0 X}(u) \). (4.3)

Next, consider the rotation \( R_\beta \). Let \( G \) be the c.d.f. of \( R_\beta(X) \), \( G_\alpha \) the c.d.f. of \( R_\alpha(R_\beta(X)) = R_{\alpha+\beta}(X) \). (Note: \( G_\alpha = F_{\alpha+\beta} \)). Then,

\[
R_\beta\{R_\alpha(L_{G_\alpha}(u)): \alpha \in (0,2\pi]\}
\]

\[
= \{R_-((\alpha+\beta))(L_{F_{\alpha+\beta}}(u)): \alpha \in (0,2\pi]\}
\]

\[
= \{R_-\alpha(L_{F_\alpha}(u)): \alpha \in (0,2\pi]\}
\]
\[
\begin{align*}
  &\Rightarrow R^{-\alpha}(B_G(u)) = B_F(u) \\
  &\Rightarrow B_G(u) = R_\beta(B_F(u)) \\
  &\Rightarrow \theta_G(u) = R_\beta(\theta_F(u)) \\
  \end{align*}
\]

i.e., \( R_\beta(\theta_X(u)) = \theta_{R_\beta X}(u) \). \tag{4.4}

Finally, since any reflection \( T_\alpha \) can be expressed as

\( T_\alpha = R_\alpha T_0 \), we have

\[
\begin{align*}
  \theta_{T_\alpha X}(u) &= \theta_{R_\alpha T_0 X}(u) \\
  &= R_\alpha(\theta_{T_0 X}(u)) \quad \text{(by (4.4))} \\
  &= R_\alpha T_0(\theta_X(u)) \quad \text{(by (4.3))} \\
  &= T_\alpha(\theta_X(u)), \text{ establishing axiom B(3).}
\end{align*}
\]

**Axiom (B4):** Consider first the transformation \( D = \begin{pmatrix} d_1 & 0 \\ 0 & 1 \end{pmatrix} \), \( d_1 > 0 \). Let \( G \) be the c.d.f. of \( D(X) \) and \( G_\alpha \) the c.d.f. of \( R_\alpha(D(X)) \).

For every \( \alpha \in (0,2\pi) \), \( \exists \beta \in (0,2\pi) \) s.t.

\[
D_1(R_\alpha(L_F(u))) = R_\beta(L_G(u)).
\]

In fact, \( \beta \) is s.t. \( d_1 \arctan \beta = \arctan \alpha \).

This follows immediately from Lemma 1 (and the fact that any affine transformation takes parallel lines into parallel lines). (So, if \( x \) is a point of intersection of two lines: \( x = R_{-\alpha_1}(L_{F_{\alpha_1}}(u)) \cap R_{-\alpha_2}(L_{F_{\alpha_2}}(u)) \) then \( D_1 x \) is also a point...
of intersection: \( D_1^X = R_{-\beta_1} (L_{\beta_1}^G (u)) \cap R_{-\beta_2} (L_{\beta_2}^G (u)) \) where
\[
d_1 \arctan \beta_i = \arctan \alpha_i \quad (i = 1, 2).
\]

\[\rightarrow D_1 (B_F (u)) = B_G (u)\]

\[\rightarrow D_1 (\theta_F (u)) = \theta_G (u)\]

i.e., \( D_1 \theta_X (u) = \theta_{D_1 X} (u) \).

A similar argument holds for transformations of the form

\[
D_2 = \begin{pmatrix} 1 & 0 \\ 0 & \alpha_2 \end{pmatrix}, \quad \alpha_2 > 0 \quad \text{and so for transformations}
\]

\[
D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}, \quad d_1 > 0, \quad d_2 > 0.
\]

Hence, given two distribution functions \( F \) and \( G \), we can construct \( \theta_F (u) \) and \( \theta_G (u) \) as above, and then define \( F \) smaller than \( G \) iff

\[
\theta_F (u) \leq \theta_G (u).
\]

This can be done for each value of \( u \in (0, \frac{1}{2}] \) to obtain an entire class of stochastic orderings. In fact, we can more generally let \( S \) be any subset of \((0, \frac{1}{2}]\) and define \( F \) smaller than \( G \) iff

\[
\theta_F (u) \leq \theta_G (u), \quad \forall u \in S.
\]

There are many ways to define an ordering of component random variables in a random vector. Snijders (1981) considers a number of definitions, and provides an important result establishing their equivalence. His primary concern
is the following testing problem:

Given a random vector $X = \begin{pmatrix} X \\ Y \end{pmatrix}$ with distribution function $F$, we wish to test

$$H: F(x,y) = F(y,x)$$

versus an alternative under which $Y$ is in some sense stochastically smaller than $X$. One approach is to rotate the distribution through a (counterclockwise) angle of $45^\circ$ and then test for symmetry about the $y$-axis versus the alternative that the rotated distribution is stochastically positive. Snyders considers a group of transformations $G$, where for $g \in G$, $g(x,y) = (g(x), g(y))$, and $g$ is an increasing injection $g: \mathbb{R} \to \mathbb{R}$. The transformed or rotated problem is not invariant with respect to this group. Hence, Snyders considers the original testing problem with an appropriate alternative which is invariant under $G$. To this end, he gives the following definition of bivariate asymmetry, indicating that in the random vector $\begin{pmatrix} X \\ Y \end{pmatrix}$ $Y$ is stochastically smaller than $X$.

**Definition 4.9:** (stochastic ordering for components of a random vector)

Given a random vector $X = \begin{pmatrix} X \\ Y \end{pmatrix}$, we say that $X$ is stochastically larger than $Y$ iff

$$P((X,Y) \in A) \geq P((Y,X) \in A), \forall A \in \mathcal{R}$$
where

\[ R = \{ \begin{align*} & (x, y) \in \mathbb{R}^2 : & \text{if } (x_1, y_1) \in A, & x_2 \geq x_1, & y_2 \leq y_1, \\
& & \text{then } (x_2, y_2) \in A \} \]

It is not unreasonable to use the location region to obtain a natural ordering for the components of \( X \).

Consider the location region \( L_F \) obtained by using the marginal ordering. Let

\[ S_M = \{ (\theta_{F_X}(u), \theta_{F_Y}(v)) : \text{if } L_F \text{ is the location region obtained by using the conditional ordering, let} \\
& \text{let } S_C = \{ (\theta_{F_X}(u|y), \theta_{F_Y}(v|x)) : \text{for } x \in S(F_X), y \in S(F_Y), \\
& \text{then } (x_2, y_2) \in A \} \}
\]

Suppose further that \( L_F \) lies entirely below the equi-angular line. Then, \( m_{X,Y} \geq m_{F_X}(v), \text{ for } x \in S(F_X), y \in S(F_Y), \]

Thus, in the spirit of Definition 4.8, we have \( Y \leq X \). In this way, the location region, taken as a whole can be used to define when one component of a random vector is \( S_M \)-stochastically larger than the other component.

Similarly, if \( L_C \) is the location region obtained by using the conditional ordering, let

\[ S_C = \{ (\theta_{F_X}(u|y), \theta_{F_Y}(v|x)) : \text{for } x \in S(F_X), y \in S(F_Y), \\
& \text{then } (x_2, y_2) \in A \} \}
\]

If \( L_F \) lies entirely below the equi-angular line, then it is appropriate to say \( Y \leq X \).
V. CONCLUSION

A. Multivariate Generalization

The construction of the location region in Chapters II and III was accomplished by reducing the bivariate problem to a univariate problem. In Chapter II, the marginal distributions were used to construct a location rectangle, obtained as the Cartesian product of the univariate marginal location intervals. All possible rotations of the random vector were then considered, a location rectangle obtained for each, and then by intersecting the inversely rotated rectangles, a location region was constructed. A similar process in Chapter III yielded a location region under a more restrictive type of stochastic ordering. Again, the bivariate problem was reduced to a univariate problem, this time by considering not only the marginal distributions, but also all conditional distributions as well. In each direction, a location rectangle was obtained as the Cartesian product of the union of the location intervals, where the union was taken over all values of the conditioning variable. By averaging again over all possible rotations, a location region was constructed.

These procedures can be applied to the general multivariate case in the same way. Thus, with the marginal ordering, hyperrectangles can be constructed by forming the
Cartesian product of the ranges of the marginal symmetry functions. By then considering all orthogonal transformations of the random vector, an analogous averaging can be done, this time averaging over all orthogonal transformations. Similarly, in the conditional ordering case, location hyperrectangles are obtained as the Cartesian product of the union of the ranges of the respective univariate conditional functions of symmetry. This is done for every orthogonal transformation and then averaged. In both cases, all three methods of approach yield essentially the same location region, the proofs completely analogous to those in the bivariate case. Also, the regions are again closed convex sets.

The results on estimation are also easily extended but the inability to portray the estimated location region graphically in $\mathbb{R}^n$ for $n>3$ makes the usefulness of such extensions questionable.

Since results on the loss function approach, measures of location and measures of asymmetry were originally presented in multivariate form, extensions of these results are immediate.

The applications dealing with stochastic orderings are also easily extended. It seems reasonable that a test of whether a multivariate distribution is symmetric about the equi-angular line versus one component is stochastically
larger than the remaining variables could have applications in linear models; specifically, the analysis of variance.

B. Summary

This paper has extended the univariate results obtained by Doksum regarding measures of location and asymmetry to the multivariate case. The method of approach reduced the multivariate problem to the univariate case by considering the respective marginal and conditional distributions. Two appropriate stochastic orderings were defined in each case, and a location region was then constructed. These methods were used in the construction of these regions, paralleling Doksum's approach in the univariate case. The resulting location region is a closed, convex, set. For the bivariate case, estimation of the location regions was discussed, and computer programs were presented which produced plots of the estimated regions in the plane. Examples were given and properties of these sample regions were also discussed.

These graphs of the estimated location region serve as a general and perhaps subjective guide to check if the assumption of symmetry is valid. They are not intended to be used as formal tests of the hypothesis of symmetry; indeed, the alternate hypothesis in such tests is so large, that the power of such tests would undoubtedly be very low. More specific alternatives are necessary, and such
tests can be found in the literature (see for example, Snijders (1981)).

Finally, two applications of the location region were presented. These include using points in the location region as measures of location for a given distribution, as well as using the region as a whole to characterize the degree of asymmetry in a distribution. The latter was done by considering the width function, a tool taken from convex analysis. Again, the graph of the empirical width function should serve as a guide to check the hypothesis of symmetry, much as a full normal plot aids in the checking of normality. In addition, it is possible to use the ideas developed to define stochastic orderings of both multivariate distribution functions and components of multivariate random vectors.
VI. REFERENCES


VII. ACKNOWLEDGMENTS

I wish to express my appreciation to Glen Meeden for his guidance and encouragement throughout my work on this dissertation.
VIII. APPENDIX 'A

Given below are two FORTRAN programs used to construct estimates for the location region $B_p$. The first program and its corresponding subprograms are used to obtain an estimated region under the marginal stochastic ordering. This program is given in section I. To obtain an estimated location region under the conditional stochastic ordering, only the main program from section I needs to be altered and one additional subprogram is also needed. These are given in section II.

I. Marginal ordering:

C MAIN PROGRAM
DOUBLE PRECISION DSEED
REAL X(300), Y(300), THETAX(300), THETAY(300), SX(300), SY(300)
REAL P1(2), P2(2), P3(2), P4(2), RP1(2), RP2(2), RP3(2), RP4(2)
REAL A(2), RA(2), SC(4), YC(4), DEGREE
INTEGER LOCSML, LOCBIG, SMALLX, SMALLY, BIGX, BIGY, N
C SET SAMPLE SIZE AT 300
N=300
C INITIALIZE THE SEED FOR THE PSEUDO-RANDOM NUMBER
C GENERATOR: IN THIS EXAMPLE, AN IMSL SUBROUTINE CALLED
C GGNML IS USED TO GENERATE STANDARD NORMAL RANDOM
C VARIABLES
DSEED = 11111. D0
C GENERATE X-COORDINATES
CALL GGNML (DSEED,N,X)
C GENERATE Y-COORDINATES
DSEED = 345678. D0
CALL GGNML (DSEED,N,Y)
C INITIALIZE FIRST DIRECTION
DEGREE = 0.0
C SET 9 INCH LENGTH OF AXES FOR FINAL PLOT
SL = 9.0
YL = 9.0
C    ROTATE THE SAMPLE THROUGH DEGREES
25 DO 30 K=1,N
    A(1) = X(K)
    A(2) = Y(K)
    CALL ROTATE (A,RA,DEGREE)
    SX(K) = RA(1)
    SY(K) = RA(2)
30 CONTINUE
C    SORT THE X AND Y COORDINATES
    CALL SORT(SX,N)
    CALL SORT(SY,N)
C    COMPUTE THE MARGINAL FUNCTIONS OF SYMMETRY EVALUATED
    AT THEIR RESPECTIVE ORDER STATISTICS
    DO 40 L=1,N
        THETAX(L) = (SX(L)+SX(N-L+1))/2.0
        THETAY(L) = (SY(L)+SY(N-L+1))/2.0
40 CONTINUE
C    COMPUTE THE COORDINATES OF THE FOUR CORNERS OF THE
C    LOCATION RECTANGLE AND ROTATE THEM BACK -DEGREES
    SMALLX = LOCML(THETAX,1,N)
    SMALLY = LOCML(THETAY,1,N)
    BIGX = LOCBIG(THETAX,1,N)
    BIGY = LOCBIG(THETAY,1,N)
    P1(1) = THETAX(SMALLX)
    P1(2) = THETAY(SMALLY)
    P2(1) = THETAX(SMALLX)
    P2(2) = THETAY(BIGY)
    P3(1) = THETAX(BIGX)
    P3(2) = THETAY(BIGY)
    P4(1) = THETAX(BIGX)
    P4(2) = THETAY(SMALLY)
    CALL ROTATE(P1,RP1,(-1.0)*DEGREE)
    CALL ROTATE(P2,RP2,(-1.0)*DEGREE)
    CALL ROTATE(P3,RP3,(-1.0)*DEGREE)
    CALL ROTATE(P4,RP4,(-1.0)*DEGREE)
    XC(1) = RP1(1)
    XC(2) = RP2(1)
    XC(3) = RP3(1)
    XC(4) = RP4(1)
    YC(1) = RP1(2)
    YC(2) = RP2(2)
    YC(3) = RP3(2)
    YC(4) = RP4(2)
C PRINT OUT THE COORDINATES OF B-HAT SUB DEGREE
WRITE(6,50) DEGREE, RP1(1), RP1(2), RP2(1), RP2(2),
C RP3(1), RP3(2), RP4(1), RP4(2)
50 FORMAT('DEGREE=', F5.1, 2X, 4('(' , F10.3, ',','F10.3,')',2X))
C PLOT THE RECTANGLE
CALL GRAPH(4,XC,YC,11,6,XL,YL,0.10,-0.40,0.10,-0.40,
C 'X-AXIS;',','Y-AXIS;',','EST LOC REGION;',',';')
XL=0.0
YL=0.0
C INCREMENT DEGREE
DEGREE = DEGREE + 5.0
IF(DEGREE .LT. 90) GO TO 25
STOP
END
SUBROUTINE ROTATE(POINT,RPOINT,DEGREE)
C THIS IS A SUBROUTINE TO ROTATE A POINT (INPUT) THROUGH
C AN ANGLE ANPHA
C THE ROTATED POINT IS RPOINT (OUTPUT).
REAL POINT(2), RPOINT(2), ALPHA, DEGREE, PI
DATA PI/3.1415926/
ALPHA = DEGREE*PI/180.0
RPOINT(1) = POINT(1)*COS(ALPHA) - POINT(2)*SIN(ALPHA)
RPOINT(2) = POINT(1)*SIN(ALPHA) + POINT(2)*COS(ALPHA)
RETURN
END
SUBROUTINE SORT(VECTOR,N)
C THIS IS A SIMPLE SORTING SUBROUTINE WHICH WILL ARRANGE
C THE N ELEMENTS OF THE ARRAY VECTOR INTO INCREASING ORDER
INTEGER N, TOP, SMLEST, LOCSML
REAL VECTOR(N)
TOP=1
10 SMLEST = LOCSML(VECTOR, TOP, N)
CALL SWITCH(VECTOR(TOP), VECTOR(SMLEST))
TOP=TOP+1
IF(TOP.LT.N) GO TO 10
RETURN
END
INTEGER FUNCTION LOCSML(A, FROM, TO)
C THIS FUNCTION LOCATES THE POSITION OF THE SMALLEST
C ELEMENT OF THOSE NUMBERS IN THE ARRAY A FROM A(FROM)
C TO A(TO).
INTEGER FROM, TO, I
REAL A(TO)
LOCML = FROM
I = FROM+1
IF(I.GT.TO) RETURN
IF(A(I) .LT. A(LOCML)) LOCSML+I
I=I+1
GO TO 10
END

INTEGER FUNCTION LOCBIG(A,FROM,TO)
C THIS FUNCTION LOCATES THE POSITION OF THE LARGEST ELEMENT
C OF THOSE NUMBERS IN THE ARRAY A FROM A(FROM) TO A(TO)
INTEGER FROM, TO, I
REAL A(TO)
LOCBIG = FROM
I = FROM+1
IF (I.GT.TO) RETURN
IF (A(I).GT. A(LOCBIG)) LOCBIG=I
I=I+1
GO TO 10
END

SUBROUTINE SWITCH (A,B)
C THIS IS A SUBROUTINE TO INTERCHANGE THE CONTENTS
C OF THE VARIABLES A AND B
REAL A, B, COPYA
COPYA = A
A=B
B = COPYA
RETURN
END
II. Conditional ordering:

C MAIN PROGRAM
DOUBLE PRECISION DSEED
REAL SX(300), SY(300), THTACY(300), THTACX(300)
REAL TCXMIN(25), TCXMAX(25), TCYMIN(25), TCYMAX(25)
REAL P1(2), P2(2), P3(2), P4(2)
REAL RP1(2), RP2(2), RP3(2), RP4(2)
REAL A(2), RA(2), XC(4), YC(4)
REAL DEGREE, SIGMA(3), RVEC(300,2), WKVEC(2), CX(300), CY(300)
REAL LENDPT, RENDPT, DELTA
INTEGER SMLCY, SMLCX, BIGCY, BIGCX
INTEGER LOCSML, LOCBIG, I
N=300
DSEED = 44444.D0
WKVEC(1) = 0.0
SIGMA(1) = 1.0
SIGMA(2) = 0.5
SIGMA(3) = 1.0
CALL GGNML (DSEED, N, 2, SIGMA, 300, RVEC, WKVEC, IER)
DEGREE = 0.0
XL = 9.0
YL = 9.0
25 DO 30 K=1,N
    A(1) = RVEC(K,1)
    A(2) = RVEC(K,2)
    CALL ROTATE (A,RA,DEGREE)
    SX(K) = RA(1)
    SY(K) = RA(2)
30 CONTINUE
CALL CSORT(SX,XY,N)
DELTA=(SX(N)-SX(1))/25.0
DO 35 J=1,25
    I=0
    LENDPT = SX(1) + (J-1)*DELTA
    RENDPT = SX(1) + J*DELTA
    DO 32 K=1,N
        IF (SX(K).GE.LENDPT.AND.SX(K).LE.
            RENDPT) GO TO 31
    GO TO 32
31    I = I+1
    CY(I) = SY(K)
32 CONTINUE
CALL SORT(CY,I)
DO 33 L=1,I
    THTACY(L) = (CY(L)+CY(I-L+1))/2.0
CONTINUE
SMLCY = LOCSML(THTACY,1,I)
BIGCY = LOCBIG(THTACY,1,I)
TCYMIN(J) = THTACY(SMLCY)
TCYMAX(J) = THTACY(BIGCY)
CONTINUE
SMALLY = LOCSML(TCYMIN,1,25)
BIGY = LOCBIG(TCYMAX,1,25)
CALL CSORT(SY,SX,N)
DELTA = (SY(N)-SY(1))/25.0
DO 45 J=1,25
I=0
LENDPT = SY(1) + (J-1)*DELTA
RENDEPT = SY(1) + J*DELTA
DO 42 K=1,N
IF (SY(K).GE.LENDPT.AND.SY(K).LE.RENDEPT)
    GO TO 41
GO TO 42
I=I+1
CX(I) = SX(K)
CONTINUE
CALL SORT (CX,I)
DO 43 L=1,I
THTACX(L) = (CX(L)+CX(I-L+1))/2.0
CONTINUE
SMLCX = LOCSML (THTACX,1,I)
BIGCX = LOCBIG(THTACX,1,I)
TCXMIN(J) = THTACX(SMLCX)
TCMAX(J) = THTACX(BIGCX)
CONTINUE
SMALLX = LOCSML(TCXMIN,1,25)
BIGX = LOCBIG(TCXMAX,1,25)
P1(1) = TCXMIN(SMALLX)
P1(2) = TCYMIN(SMALLY)
P2(1) = TCXMIN(SMALLX)
P2(2) = TCYMAX(BIGY)
P3(1) = TCXMAX(BIGX)
P3(2) = TCYMAX(BIGY)
P4(1) = TCXMAX(BIGX)
P4(2) = TCYMIN(SMALLY)
CALL ROTATE (P1,RP1,(-1.0)*DEGREE)
CALL ROTATE (P2,RP2,(-1.0)*DEGREE)
CALL ROTATE (P3,RP3,(-1.0)*DEGREE)
CALL ROTATE (P4,RP4,(-1.0)*DEGREE)
XC(1) = RP1(1)
SC(2) = RP2(1)
SC(3) = RP3(1)
SC(4) = RP4(1)
YC(1) = RP1(2)
YC(2) = RP2(2)
YC(3) = RP3(2)
YC(4) = RP4(2)
WRITE(6,50) DEGREE, RP1(1), RP1(2), RP2(1), RP2(2),
       RP3(1), RP3(2), RP4(1), RP4(2)
50 FORMAT('0DEGREE=', F5.1,2X,4('(',F10.3,')',F10.3,''),
       2X) )
CALL GRAPH(4,SX,YC,11,6,SL,YL,1.0,-4.0,1.0,-4.0,
       'X-AXIS;','Y-AXIS;','EST LOC REGION;',';')
DEGREE = DEGREE + 5.0
IF(DEGREE.LT.90) GO TO 25
STOP
END

SUBROUTINE CSORT(KEYS, OTHER, N)
INTEGER N, TOP, SMLEST, LOCSML
REAL KEYS (N), OTHER (N)
TOP = 1
10 SMLEST = LOCSML(KEYS,TOP,N)
   CALL SWITCH (KEYS(TOP), KEYS(SMLEST))
   CALL SWITCH (OTHER(TOP), OTHER(SMLEST))
   TOP = TOP+1
   IF(TOP.LT.N) GO TO 10
RETURN
END
Bickel and Lehmann (1975) consider using as a measure of location the value $\theta = \theta_F$ which minimizes the integral
\[ \int \rho(x-\theta) dF(x), \] (B.1)
where $\rho: \mathbb{R} \to \mathbb{R}$ is an even, convex, nonnegative and twice differentiable function.

They prove the following result:

**Theorem B.1:** Suppose that $\theta_F$ is defined as minimizing Equation (B.1) on a set $\mathcal{F}$ which is convex, contains all point masses, is closed under changes of scale, and contains a distribution $F^0$ symmetric about zero s.t.
\[ \int \rho''(x) dF^0(x+t) < \infty \text{ for all } t \text{ and} \]
\[ \frac{d}{dt} \int \rho'(x-t) dF^0(x) = -\int \rho''(x-t) dF^0(x) \]
For any given distribution $F$, denote by $F_\sigma$ the distribution defined by $F_\sigma(x) = F(\frac{x}{\sigma})$. Suppose that
\[ \theta_{F_\sigma} = \sigma \theta_F \text{ for all } \sigma > 0. \]

Then, $\rho'(x) = c|x|^\lambda \text{ sgn } x$ for some $\lambda > 0$, $c > 0$.

This result implies that we can restrict our attention to functions $\rho$ of the form $\rho(x) = |x|^\lambda$, $\lambda \geq 1$. We now prove that for such a $\rho$, $\theta_F$ is a location functional. Though the
proof is not new, it is included here because it provides insight into the multivariate case.

Consider the following integral:
\[ \mu_\lambda(\theta) = \int |x - \theta|^\lambda dF(x); \lambda \geq 1 \]

Let \( \theta_F(\lambda) \) be the value of \( \theta \) which minimizes \( \mu_\lambda(\theta) \) for each \( \lambda \geq 1 \); i.e.,
\[ \mu_\lambda(\theta_F(\lambda)) = \inf_\theta \int |x - \theta|^\lambda dF(x). \]

DeGroot and Rao (1963) show that \( \theta_F(\lambda) \) satisfies the following equation.
\[ \int_{x > \theta} (x - \theta)^{\lambda-1} dF(x) = \int_{x < \theta} (\theta - x)^{\lambda-1} dF(x) \]

**Proposition B.1:** (1) \( \overline{\mu}_F(\theta) = \int_{x > \theta} (x - \theta)^{\lambda-1} dF(x) \)

is a decreasing function of \( \theta \).

(2) \( \underline{\mu}_F(\theta) = \int_{x < \theta} (\theta - x)^{\lambda-1} dF(x) \)

is an increasing function of \( \theta \).

**Proof:** (1) Assume \( \theta_1 < \theta_2 \)

\[ \Rightarrow -\theta_1 > -\theta_2 \]

\[ \Rightarrow x - \theta_1 > x - \theta_2 \]

\[ \Rightarrow (x - \theta_1)^{\lambda-1} \geq (x - \theta_2)^{\lambda-1}, \text{ since } \lambda \geq 1. \]
\[
\Rightarrow \int_{x>\theta_1} (x-\theta_1)^{\lambda-1} dF(x) \geq \int_{x>\theta_2} (x-\theta_1)^{\lambda-1} dF(x)
\]
since we are integrating over a smaller set

\[
\geq \int_{x>\theta_2} (x-\theta_2)^{\lambda-1} dF(x)
\]
by Equation (B.2)

This proves (1). Result (2) follows similarly.

**Proposition B.2:** \( \theta_F(\lambda) \) satisfies the univariate axioms of location for each \( \lambda > 1 \).

**Proof:** (ul) Assume \( F(x) \geq G(x), \forall x \).

Consider the function \( h_\theta(x) = (x-\theta)^{\lambda-1}, x>\theta, \lambda > 1 \).

Now \( h_\theta \) is a nonnegative increasing function of \( x \) for each fixed \( \theta \).

For each positive integer \( n \), define

\[
A^\theta_{k,n} = \{x : h_\theta(x) > \frac{k-1}{2^n}\}, \quad k = 1, 2, \ldots, n \cdot 2^n,
\]

\[
B^\theta_n = \{x : h_\theta(x) > n\}.
\]

Define the sequence of simple functions \( \{s^\theta_n(x)\} \) where

\[
s^\theta_n(x) = \frac{1}{2^n} \sum_{i=1}^{n \cdot 2^n} I_{A^\theta_{i,n}}(x) + n I_{B^\theta_n}(x), \text{ where}
\]

\[
I_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}
\]
Since $h_\theta$ is increasing, $A_{k,n}^\theta = A_{k+1,n}^\theta$, $\forall k$ and

\[ A_{k,n}^\theta = \{a_{k,n}, \infty\} \] where $h_\theta(a_{k,n}) = \frac{k-1}{2^n}.$

On $B_{n}^\theta \cap [s_n^\theta(x) - h_\theta(x)] < \frac{1}{2^n}$ and since

\[ \lim_{n \to \infty} B_{n}^\theta = \emptyset, \] we have \[ \lim_{n \to \infty} s_n^\theta(x) = h_\theta(x); \] for all $x.$

Actually, $s_n^\theta \uparrow h_\theta$ pointwise. Hence,

\[ \int_{x > \theta} (x-\theta)^{\lambda-1}dF(x) = \lim_{n \to \infty} \int_{x > \theta} s_n^\theta(x)dF(x) \quad \text{(LMCT)} \]

\[ = \lim_{n \to \infty} \frac{1}{2^n} \sum_{k=1}^{n-2} \int_{A_{k,n}^\theta} dF(x) \]

\[ = \lim_{n \to \infty} \frac{1}{2^n} \sum_{k=1}^{n-2} [1-F(a_{k,n})] \]

\[ \leq \lim_{n \to \infty} \frac{1}{2^n} \sum_{k=1}^{n-2} [1-G(a_{k,n})] \quad \text{(by assumption)} \]

\[ = \lim_{n \to \infty} \frac{1}{2^n} \sum_{k=1}^{n-2} \int_{A_{k,n}^\theta} dG(x) \]

\[ = \lim_{n \to \infty} \int_{x > \theta} s_n^\theta(x)dG(x) \]

\[ = \int_{x > \theta} (x-\theta)^{\lambda-1}dG(x) \quad \text{(LMCT)} \]
A similar argument shows

$\int_{x<\theta} (\theta-x)^{\lambda-1}dF(x) \geq \int_{x<\theta} (\theta-x)^{\lambda-1}dG(x).$

If we let $\mu_P(\theta) = \int_{x>\theta} (x-\theta)^{\lambda-1}dF(x),$

$\mu_G(\theta) = \int_{x>\theta} (x-\theta)^{\lambda-1}dG(x)$

$\mu_P(\theta) = \int_{x<\theta} (\theta-x)^{\lambda-1}dF(x)$

$\mu_G(\theta) = \int_{x<\theta} (\theta-x)^{\lambda-1}dG(x)$

we have

$\mu_P(\theta_F(\lambda)) = \mu_P(\theta_F(\lambda))$

and

$\mu_G(\theta_G(\lambda)) = \mu_G(\theta_G(\lambda))$ (by DeGroot and Rao, 1963).

Using Proposition B.2, see Figure B.1.
Figure B.1. The functions $\bar{\mu}_F$, $\mu_F$, $\bar{\mu}_G$, and $\mu_G$
We have the following claim:

$$\theta_F(\lambda) \leq \theta_G(\lambda), \quad \forall \lambda > 1$$

Proof of claim: Fix $\lambda > 1$ and assume $\theta_F(\lambda) > \theta_G(\lambda)$.

$$\Rightarrow \mu_F(\theta_F(\lambda)) \leq \mu_F(\theta_G(\lambda)) \quad \text{and} \quad \bar{\mu}_F(\theta_F(\lambda)) \geq \bar{\mu}_F(\theta_G(\lambda))$$

But $\mu_F(\theta_F(\lambda)) = \bar{\mu}_F(\theta_F(\lambda))$ implies

$$\bar{\mu}_F(\theta_G(\lambda)) \leq \bar{\mu}_F(\theta_F(\lambda)) = \mu_F(\theta_F(\lambda)) \leq \mu_F(\theta_G(\lambda))$$

$$\Rightarrow \bar{\mu}_G(\theta_G(\lambda)) \leq \bar{\mu}_F(\theta_G(\lambda))$$

$$\leq \bar{\mu}_F(\theta_F(\lambda))$$

$$= \mu_F(\theta_F(\lambda))$$

$$\leq \mu_F(\theta_G(\lambda))$$

$$\leq \mu_G(\theta_F(\lambda))$$

$$\Rightarrow \mu_F(\theta_G(\lambda)) = \bar{\mu}_F(\theta_G(\lambda)). \quad (B.3)$$

But DeGroot and Rao (1963) show that since $W(x) = |x|^\lambda$ is strictly convex, $\theta_F(\lambda)$ is unique. Hence, the contradiction Equation (B.3) establishes the claim and thus axiom (ul) is proven.
(u2): Let $Y = aX + b$. Then $F_Y(x) = F_X\left(\frac{x-b}{a}\right)$, $a > 0$

Now

$$\int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}dF_Y(x) = 0.$$

Hence,

$$\int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}dF_Y(x) = 0$$

$$\Rightarrow \int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}dF_X\left(\frac{x-b}{a}\right) = 0$$

$$\Rightarrow \int |ax + b - \theta_{F_Y}(\lambda)|^{\lambda-1}dF_X(x) = 0$$

$$\Rightarrow \theta_{F_Y}(\lambda) - b = \frac{\theta_{F_Y}(\lambda) - b}{a}$$

$$\Rightarrow \theta_{F_Y}(\lambda) = \frac{\theta_{F_Y}(\lambda) - b}{a} \quad \text{or}$$

$$\theta_{F_Y}(\lambda) = \theta_{aX + b}(\lambda) = aX + b$$

$$\Rightarrow \theta_{F_X}(\lambda) = \theta_{aX + b}(\lambda) + b = a\theta_{F_X}(\lambda) + b.$$

(\text{13}): Let $Y = -X$. $F_Y(x) = 1 - F_X(-x)$

Now

$$\int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}dF_Y(x) = 0.$$  

Hence,

$$\int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}d[1-F_X(-x)] = 0$$

$$\Rightarrow \int |x-\theta_{F_Y}(\lambda)|^{\lambda-1}dF_X(x) = 0$$
\[ \Rightarrow \int |x - (-\theta_{\mathcal{F}_Y}(\lambda))|^{\lambda - 1} \mathcal{F}_X(x) = 0 \]

\[ \Rightarrow \theta_{\mathcal{F}_X}(\lambda) = \theta_{\mathcal{F}_Y}(\lambda) = -\theta_{-\mathcal{F}_X}(\lambda) = -\theta_{\mathcal{F}_Y}(\lambda). \]

Hence, \( \theta_{\mathcal{F}_X}(\lambda) \) satisfies (u1)-(u3) and thus is a univariate location parameter.