SIMPLICIAL, CIRCULAR AND RECTANGULAR DATA DEPTH

BY

ANDREW J. MORRIS

A Thesis Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

SIMPLICIAL, CIRCULAR AND RECTANGULAR DATA DEPTH

BY

ANDREW J. MORRIS

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

Abstract

In general, the notion of data depth enables us to describe how deep or central a given p-dimensional point x is with respect to a p-variate distribution F. The sample version of data depth is nonparametric in nature, and enables us to order multivariate data in a centre-outward ranking, giving us a notion of order statistics in multiple dimensions. Liu (1990) defined a particular depth function based on simplices that has come to be widely studied and applied.

In this thesis, Liu's simplicial depth function, along with its empirical estimate, are examined, with focus on some of their important properties, including continuity and consistency, respectively. Some new properties are asserted and examined, including a tractable form for the simplicial depth function, as well as an upper bound on the mean of a random version of simplicial depth.

Two new depth functions are defined which, in two dimensions, are based on circles and rectangles, as Liu's simplicial depth is based on triangles. Properties analogous to those for simplicial depth are asserted for our new circular and rectangular depth functions, and the three are compared to illustrate some benefits and downfalls of each.

One important application of simplicial depth is studied, namely, that of nonparametric multivariate quality control, based on the work of Liu (1995). Control charts based on simplicial depth are defined for a multivariate process, and an analogue is presented using our new rectangular depth function. Rectangular depth posesses some distinct advantages over simplicial depth in its application to these charts, specifically, the reduction in the size of the reference sample required in order to construct the charts.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis supervisor, Dr. Dean Slonowsky for his countless hours of work and discussion with me. His patience, dedication, ingenuity and friendship have made my work so much more enjoyable. He was always available for help and expressed a genuine interest in both me and my work. Needless to say, I look forward to working again with him on my doctoral thesis.

I would also like to thank the members of my examining committee, whose comments and questions enhanced both my understanding and interest in the subject: Dr. Xikui Wang of the Department of Statistics, and Dr. Kirill Kopotun of the Department of Mathematics.

I wish to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for my PGS-A scholarship, and also the Department of Statistics, whose financial support has greatly contributed to my studies.

I must also thank the various Department of Statistics members for their help and encouragement over the last six years of undergraduate and graduate studies: Dr. Smiley Cheng, who has offered me not only the opportunity to gain invaluable experience in teaching, but also the genuine interest and support of a friend; Dr. James Fu, under whom I worked on my first research project in Statistics, and who is always willing to offer his expertise on virtually any statistical problems I have; Dr. Ken Mount, my first Statistics Professor, who has never failed to push me to my full potential as a student; Dr. Brian Macpherson, who is always willing to lend a hand, whether it relates to my thesis, my teaching, or any other concerns I may have; and Dr. Liqun Wang, who is always around the corner and ready to help. I would also like to extend a big thank you to our Department secretary, Margaret Smith, whose smile and friendly help none of us could do without.

This dissertation is devoted to my parents, David and Jenni, and my best friend, André Lavergne. My parents have been a constant source of love and encouragement, and have been there for me in every way I could have asked. As the first member of my family with a Masters degree, I know they are extremely proud of me. I hope to continue to make them proud in my future studies, and all aspects of my life. André has helped me with the practical part of this thesis, drawing graphs, finding resources, editing my work, and is always eager to help in any way he can. In addition, he has pushed me to attain my potential, and is always there for me without condition.

I would also like to mention my niece, Amelia Kate, who reminds me every day that there are more important things in life than Statistics and this thesis.

Contents

Chapter 1

Introduction

1.1 General Motivation

The notion of order statistics for univariate random variables is both straightforward and simply defined. Let $X_1, X_2, ..., X_n$ be a random sample representing one-dimensional data from a distribution. Then the kth order statistic $X_{[k]}$ is defined to be the kth smallest of $X_1, X_2, ..., X_n$. In particular, $X_{[1]} = \min_{1 \leq i \leq n} X_i$ while $X_{[n]} = \max_{1 \leq i \leq n} X_i$.

There is, however, no simple definition of order statistics in the general p-dimensional case. For example, which data point is "greater", $(2, 7)$ or $(4, 3)$? More complicated still, how do we compare the points $(-2, 3, 1, 9)$ and $(6, 0, 7, -4)$? The problem becomes increasingly more difficult as the dimension p increases.

The problem of multivariate order statistics can be dealt with by the notion of data depth. This area of statistics was first proposed by Mahalanobis

(1936) and came to prominence some four decades later with the work of Tukey (1975).

The basic principle of data depth is that multi-dimensional observations are ordered based on how central or "deep" they are, with respect to a "depth function", whose shape is determined by (but is different from) the distribution (probability density function) from which the observations arise.

In the context of data depth, if $X_1, X_2, ..., X_n$ is a random sample from a multi-dimensional distribution, the kth order statistic $X_{[k]}$ is now defined to be that value of $X_i, i = 1, 2, ..., n$ with the kth **greatest** depth (Liu, 1990).

Data depth has become a valuable statistical tool, both in theory and application. One very desirable property of sample depth is that it is non-parametric in nature, and so it lends its use to multivariate problems in which the underlying multivariate distribution is either unknown, or does not belong to a specific class of distributions. In particular, we must often deal with the problem that multivariate normality is not a valid assumption in many applications, and the notion of data depth can help us in these circumstances. This is illustrated in the use of data depth in multivariate quality control.

Univariate quality control methods are readily available, frequently used, and often easy to understand by practitioners. Even when the distributional form of a variable is unknown, some relatively simple nonparametric control charts are available. However, quite often we may wish to simultaneously examine several quality characteristics of a process, as individual variable examination ignores the correlation structure of the various attributes. There are some multivariate quality control techniques currently in common use, but they rely on an often unrealistic assumption of multivariate normality. In contrast, the notion of data depth allows us to construct nonparametric multivariate control charts that will be useful regardless of the underlying probability distribution (see Liu (1995) and Stoumbos & Jones (2000)).

1.2 The Notion of Data Depth

Before giving specific examples of depth functions which have already appeared in the literature, we state a rough heuristic definition of data depth. See Appendix A for all relevant notation.

Definition 1.2.1 Let F be a probability distribution on $\mathbf{R}^p$, $p \geq 1$. A *data depth* $D(x)$ is a measure of how deep or central a given point $x \in \mathbf{R}^p$ is with respect to the underlying distribution F. In the case of *sample data depth* $D_n(x)$, we examine how deep or central a given point $x \in \mathbf{R}^p$ is with respect to the data cloud $X_1, X_2, ..., X_n$.

For the remainder of this section, $X_1, X_2, ..., X_n$ is an independently and identically distributed (i.i.d.) random sample from a cumulative distribution function (c.d.f.) F on $\mathbf{R}^p$, where each sample point X_i can be viewed as a $1 \times p$ row vector. We now examine several useful forms of data depth, stated in the chronological order of their discovery.

Definition 1.2.2 The *Mahalanobis depth* $D^m(x)$ [Mahalanobis, (1936)] at $x \in \mathbf{R}^p$, a row vector, with respect to F is defined as

$$D^m(x) = [1 + (x - \mu_F)\Sigma_F{}^{-1}(x - \mu_F)^T]^{-1},$$

a real number, where μ_F and Σ_F are the mean vector and dispersion matrix of F, respectively.

Note that $\mu_F = (\mu_1, \mu_2, ..., \mu_p)$ and $\Sigma_F = [\sigma_{ij}]$ (a $p \times p$ matrix), i.e. the ith entry in μ_F is $E(X_{0i})$ and the entry σ_{ij} in the ith row and the jth column of $\Sigma_F\,(i, j = 1, 2, ..., p, i \neq j)$ is the covariance of X_{0i} and X_{0j}, and the values $\sigma_{ii}\,(i = 1, 2, ..., p)$ on the diagonal are the variances of the X_{0i}, where $X = (X_{01}, X_{02}, ..., X_{0p})$ has c.d.f. F. The sample Mahalanobis depth $D_n^m(x)$ is calculated by replacing these terms with their respective sample estimates, $\bar{x}$ and S. Here, $\bar{x} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_p)$ and $S = [s_{ij}]$, i.e. the ith entry in $\bar{x}$ is $\bar{x}_i$ and the entry s_{ij} in the ith row and the jth column of $S\,(i, j = 1, 2, ..., p, i \neq j)$ is the sample covariance and the values $s_{ii}\,(i = 1, 2, ..., p)$ on the diagonal are the sample variances of the given data.

If F has a density with elliptic contours, then $D^m(x)$ is, intuitively, a good measure of how close a point x is to the mean or "centre" μ_F of the underlying distribution F. As must be the case, observations close to the centre will have a higher depth value. Indeed, although nonparametric in nature, Mahalanobis depth is clearly best suited to situations where the underlying density $f(x)$ has elliptic contours. The remaining notions of depth discussed here do not have that shortcoming.

Definition 1.2.3 In general, a *half-space* is defined to be $H = \{x \in \mathbf{R}^p | x$ lies on or below (above) a $(p-1)$-dimensional hyperplane$\}$. For example, when $p = 1$, $H = (-\infty, \alpha]$ or $[\alpha, \infty)$ for some $\alpha \in \mathbf{R}$. When $p = 2$, half-spaces are determined by a line; that is, $H = \{(x, y) \in \mathbf{R}^p | y \leq ax + b\}$ or $H = \{(x, y) \in \mathbf{R}^p | y \geq ax + b\}$ for some $a, b \in \mathbf{R}$. When $p = 3$, a half-space is determined by a plane, $z = ax + by + c$, where $a, b, c \in \mathbf{R}$, and so on.

Definition 1.2.4 The *half-space depth* or *Tukey depth* $D^h(x)$ [Hodges (1955), Tukey (1975)] at $x \in \mathbf{R}^p$ with respect to F is defined as

$$D^h(x) \;=\; \inf\{\text{F(H): H is a closed half-space} \subseteq \mathbf{R}^p \text{ and } x \in H\}.$$

The sample version of half-space depth $D_n^h(x)$ is calculated by replacing $F(H)$ with the proportion of all data points falling in the closed half-space H, where $x \in H$.

$D^h(x)$ gives us the lowest probability under F of any halfspace H containing x. Clearly, if F has a unimodal and symmetric distribution about θ, $\theta \in \mathbf{R}^p$, then $D^h(x)$ for values of $x \in \mathbf{R}^p$ near θ will have higher values of $D^h(x)$, while values of x in low-probability regions will have depths close to zero.

Definition 1.2.5 The *convex hull peeling depth* D_n^{cd} [Barnett (1976)] of a data point X_k with respect to the data cloud $X_1, X_2, ..., X_n$ is defined to be the "level" of convex layer to which X_k belongs.

To determine the level of the convex layer of X_k, we start by constructing the convex hull which encloses all of the sample points $X_1, X_2, ..., X_n$. (See Appendix A for the definition of a convex hull.) All data points X_i, $i = 1, 2, ..., n$ on the perimeter of this convex hull are designated to belong to the first convex layer and all such points are then removed. The process is repeated with all remaining data points and those on the perimeter of the convex hull of the remaining set of points constitute the second convex layer.

We keep repeating the process until no points remain. As such, any point X_k on the perimeter of the convex hull of all remaining points during the jth iteration of the process is said to belong to the jth convex layer, i.e. the jth level. For an illustration, see Figure 1.1. The higher the level of a point, the greater its depth. Note that only a sample version of depth is defined for the convex peeling technique. Note also that this method represents only one version of convex peeling. See Huber (1972) and Eddy (1982) for further discussion on other versions.

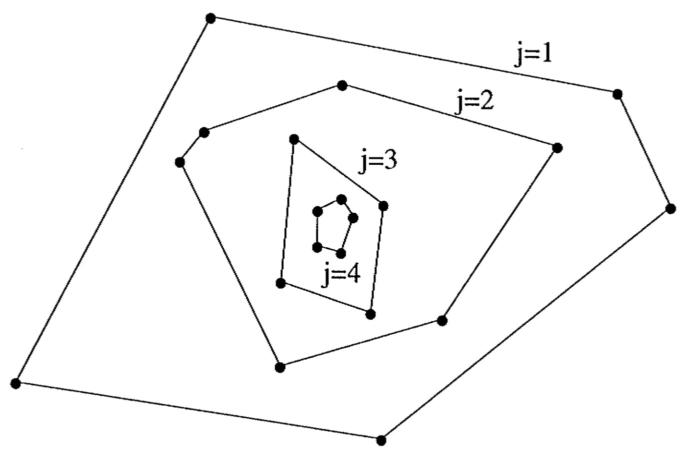

Figure 1.1: An illustration of the convex hull peeling technique. In this illustration, there are four convex layers (i.e. levels).

Definition 1.2.6 The *Oja depth* $D^o(x)$ [Oja (1983)] at x with respect to F is defined as

$$D^o(x) = [1 + E[volume(S(x, X_1, ..., X_p))]]^{-1},$$

where $S[x, X_1, ..., X_p]$ is the closed simplex with vertices x, and p random observations $X_1, ..., X_p$ from F (see Section 2.1 for the definition of a simplex). The sample Oja depth $D_n^o(x)$ at $x \in \mathbf{R}^p$ with respect to the data cloud $X_1, X_2, ...X_n$ is defined as

$$D^o{}_n(x) = \binom{n}{p}^{-1} [1 + \sum_* \{volume(S[x, X_{i_1}, ..., X_{i_p}])\}^{-1}],$$

where $*$ indicates all p-plets $(i_1, ..., i_p)$ such that $1 \leq i_1 < \cdots < i_p \leq n$. Intuitively, points close to "thicker" parts of the distribution (or data cloud, in the sample case) will form simplices with smaller volumes than observations close to the perimeter of the distribution or data cloud, and so will have a greater Oja depth.

Definition 1.2.7 Given $x_1, x_2, ..., x_p \in \mathbf{R}^p$, a *major side* is that half-space of $\mathbf{R}^p$ bounded by the hyperplane containing $\{x_1, x_2, ..., x_p\}$ which has probability ≥ 0.5 under F.

Definition 1.2.8 The *majority depth* $D^{mj}(x)$ [Singh (1991)] of $x \in \mathbf{R}^p$ with respect to F is defined as

$$D^{mj}(x) = P(x \text{ is in the major side determined by } (X_1, ..., X_p)),$$

where $X_1, X_2, ..., X_p$ are i.i.d. random variables with c.d.f. F.

The *sample majority depth* $D_n^{mj}(x)$ at $x \in \mathbf{R}^p$ with respect to the data cloud $X_1, X_2, ..., X_n$ is defined as

$$D_n^{mj}(x) = \binom{n}{p}^{-1} \sum_* \mathbb{I}(x \text{ is in the major side determined by } \{(X_{i_1}, ..., X_{i_p})\}),$$

where $*$ is defined as above and $\mathbb{I}$ is the indicator function. That is,

$$\mathbb{I}(A) = \begin{cases} 1, & \text{when } A \text{ occurs,} \\ 0, & \text{otherwise} \end{cases}$$

In the sample case, the major side is that side of the hyperplane containing the majority of the data points. In two dimensions, a hyperplane is generated by passing a (unique) line through two random points X_{i_1} and X_{i_2}. The major side is that side of the line containing more of the data points. In three dimensions, a hyperplane is generated by passing a (unique) plane through the three random points X_{i_1}, X_{i_2} and X_{i_3}. The major side is that side of the plane containing the majority of the data points. We generate one hyperplane for each combination of p data points. Obviously, the more major sides that contain a point x, the greater its majority depth.

Definition 1.2.9 The *likelihood depth* $D^l(x)$ [Fraiman and Meloche (1996)] of x with respect to F is defined simply as its probability density, and so $D^l(x) = f(x)$.

The sample version of $D^l(x)$ can be any consistent density estimate at x. (It should be noted that, in general, likelihood and depth are two distinct, albeit complimentary concepts).

1.3 Overview of Thesis

This thesis concentrates on a specific notion of depth not listed above, namely Liu's simplicial depth, a notion of data depth that has received much attention since its introduction in Liu (1990).

Liu based her depth function on random p-dimensional simplices. She originally proposed the idea as a means to obtain a notion of a multivariate median, which is not formally or uniquely defined for a general distribution F on $\mathbf{R}^p, p \geq 2$. Liu suggested using that value $x \in \mathbf{R}^p$ with the highest simplicial depth $D^s(x)$ as the multivariate median μ. The sample multivariate median $\hat{\mu}_n$ is then defined analogously as the sample point X_i attaining the highest sample simplicial depth $D_n^s(x)$. In Section 2.1, we formally define these notions. In Sections 2.2 and 2.3, several properties of the simplicial depth function $D^s(x)$ are given, including its continuity, monotonicity and maximum value. Some of the results will require an additional assumption of angular symmetry, which is defined in Section 2.3. The unbiasedness and consistency of $D_n^s(x)$ will also be asserted. All of these results have appeared in Liu (1990).

Section 2.4 deals with some new properties of Liu's simplicial depth not discussed in her 1990 paper or, to our knowledge, any subsequent papers. Included is an explicit form for the depth function $D^s(x)$ in the case of $p = 2$ dimensions, which has heretofore gone uncalculated. Also, an upper bound for $E(D^s(X))$ is given, where $D^s(X)$ is a random version of simplicial depth. The bound, given in Theorem 2.4.6 may be crude compared to Liu's upper bound for $D^s(x)$ itself, but removes the restrictive assumption of angular symmetry.

After having examined the various forms of data depth, with particular emphasis on Liu's simplicial depth, we are led to define and explore two new notions of depth. Like simplicial depth, which in two dimensions is based on the triangle, our two newly defined depth functions will be based on other simple geometic shapes; circles and rectangles. In Chapter 3, we define both the circular depth $D^c(x)$ and the rectangular depth $D^r(x)$. We examine properties similar to those explored for simplicial depth in Chapter 2. We define and study unbiased estimators $D^c_n(x)$ and $D^r_n(x)$ for the circular and rectangular depth functions, respectively. The advantages and disadvantages of using these forms of depth over simplical depth are discussed in Section 3.3. In any dimension p, both of our new depth functions are based on only two points in space, rather than the $p+1$ points required for simplicial depth. Whereas this implies a larger variance for the two new depth functions as compared to simplicial depth, our new notions of depth have many merits over simplicial depth. For example, they are easier to visualize, and the actual functional form of the depth functions are far simpler than that of simplicial depth. In addition, the empirical estimates $D^c_n(x)$ and $D^r_n(x)$ are quick to calculate in practice, in contrast to $D^s_n(x)$. (See Liu, Parelius & Singh (1999) for a discussion on the computational complexities in the evaluation of $D^s_n(x)$.)

Chapter 4 details one specific multiparameter control chart — that for individual observations — that is nonparametric in nature and based on Liu's simplicial depth. We extend the results to make a new kind of individuals multivariate control chart, based this time on our rectangular depth from Chapter 3. A discussion on the required sample size is provided, and the advantage of using rectangular depth for these charts is highlighted. Extending these charts to circular depth will be examined in future work.

This thesis also contains two appendices. Appendix A provides some necessary notation, definitions and results which form the mathematical (probabilistic) foundation for many of our calculations and results throughout the thesis. Appendix B contains some graphical displays which are very illustrative in the calculation of one of our results, namely that of the functional (tractable) form of Liu's simplicial depth $D^s(x)$ when $p = 2$.

Chapter 2

Liu's Simpicial Depth

Since its inception in Liu (1990), simplicial depth has become a widely studied and applied depth function. When F is bivariate, i.e. $p = 2$, it is based on the triangle, a simple geometric shape. This enables us to better visualize the associated notion of depth, and enhances our understanding of the concept substantially.

2.1 Simplicial Depth and Sample Simplicial Depth

Suppose we have an i.i.d. bivariate data set $X_1, X_2, ..., X_n$ with c.d.f. F. With any three data points $X_{i_1}, X_{i_2}, X_{i_3}$, we can form the closed triangle with vertices X_{i_1}, X_{i_2} and X_{i_3} (denoted $\Delta(X_{i_1}, X_{i_2}, X_{i_3})$). If we use every combination of three data points, $\binom{n}{3}$ triangles will be generated from our sample. To any point $x \in \mathbf{R}^2$, we can associate the proportion of the generated triangles which enclose the point x. Intuitively, this proportion should be relatively large if x is "deep" in the data cloud, and lower on its periphery.

The p-dimensional generalization of a triangle is a simplex, hence the name of Liu's depth function. A simplex in p dimensions is the convex hull formed by $p+1$ distinct points in $\mathbf{R}^p$. In one dimension, the "simplex" is simply a line segment. In two dimensions, it is a triangle. In three dimensions, it is a pyramid, and so on. We begin with the definition of the sample simplicial depth function in the bivariate case, since it is easier to visualize:

Definition 2.1.1 The *sample simplicial depth* $D_n^s(x)$ for a point $x \in \mathbf{R}^2$ is equal to the proportion of all triangles $\Delta(X_{i_1}, X_{i_2}, X_{i_3}), 1 \leq i_1 < i_2 < i_3 \leq n$ which contain x. That is,

$$D_n^s(x) = \binom{n}{3}^{-1} \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{I}(x \in \Delta(X_{i_1}, X_{i_2}, X_{i_3})). \tag{2.1}$$

We see that the depth given by Equation (2.1) is clearly the proportion of all generated triangles containing the fixed point x. $D_n^s(x)$ is the empirical (sample) version of the true simplicial depth $D^s(x)$:

Definition 2.1.2 The *simplicial depth* $D^s(x)$ for a point $x \in \mathbf{R}^2$ is equal to the probability that x is contained in the random triangle $\Delta(X_1, X_2, X_3)$. That is,

$$D^s(x) = P(x \in \Delta(X_1, X_2, X_3)), \tag{2.2}$$

where X_i, $i = 1, 2, 3$ are i.i.d. with c.d.f. F.

When the density function $f(x)$ is unimodal and symmetric about a point $\theta \in \mathbf{R}^2$, it is intuitively clear (and indeed can be proven) that $D^s(x)$ assumes higher values for x near θ, which coincides with the mode and mean of F. As such, θ can be viewed as the "centre" of the distribution. Conversely, when $f(x)$ is not symmetric, $D^s(x)$ can be used to define a centre (multivariate median) of F.

To motivate the definition of Liu's simplicial multivariate median, we consider the univariate version of simplicial depth:

$$D^s(x) = P(x \in \overline{X_1X_2}),$$

where $x \in \mathbf{R}$ and X_1 and X_2 are i.i.d. with c.d.f. F. $\overline{X_1X_2}$ represents the closed line segment connecting X_1 and X_2. That is, $\overline{X_1X_2} = [X_1, X_2] \cup [X_2, X_1]$, a disjoint union ($= [\min(X_1, X_2), \max(X_1, X_2)]$). When F is absolutely continuous,

$$\begin{aligned} D^s(x) &= P(X_1 \le x \le X_2) + P(X_2 \le x \le X_1) \\ &= P(X_1 \le x)P(x \le X_2) + P(X_2 \le x)P(x \le X_1) \\ &= F(x)(1 - F(x)) + F(x)(1 - F(x)) \\ &= 2F(x)(1 - F(x)). \end{aligned} \tag{2.3}$$

As such, we have that

$$\frac{d}{dF(x)} D^s(x) = 2 - 4F(x) = 0$$

when $F(x) = 0.5$, i.e. x is the median of F, and since the second derivative of $D^s(x)$ is negative, the median is that value which maximizes the simplicial depth. Hence, Liu (1990) proposes the following definition:

Definition 2.1.3 A *bivariate simplicial median* μ is any point $x \in \mathbf{R}^2$ which maximizes the simplicial depth. If there is a finite number of such points, we can uniquely define the simplicial median as the average of those values.

Similarly, the sample version of the bivariate median is then:

Definition 2.1.4 Given an i.i.d. random sample $X_1, X_2, ..., X_n$ with c.d.f. F on $\mathbf{R}^2$, the *sample bivariate simplicial median* $\hat{\mu}_n$ is that data point X_i which attains the highest sample simplicial depth. If there is more than one point X_i attaining this highest value, we define $\hat{\mu}_n$ as the average of those values.

All of these concepts can easily be extended to higher dimensions. For a distribution F on $\mathbf{R}^p$, the triangle in Definitions 2.1.1 and 2.1.2 is replaced by the simplex whose vertices are formed by $p+1$ independent observations from F. In general, given $p+1$ distinct points $x_1, x_2, ..., x_{p+1} \in \mathbf{R}^p$, we define the simplex

$$S(x_1, x_2, ..., x_{p+1}) = \left\{ x \in \mathbf{R}^p : x = \sum_{i=1}^{p+1} \alpha_i x_i, \ \sum_{i=1}^{p+1} \alpha_i = 1, \ \ \alpha_i \geq 0 \,\forall\, i \right\}.$$

The general definition of the sample simplicial depth in any dimension p is as follows:

Definition 2.1.5 The *sample simplicial depth* $D_n^s(x)$ for a point $x \in \mathbf{R}^p$ is equal to the proportion of all simplices $S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})$, $1 \leq i_1 < i_2 < \cdots < i_{p+1} \leq n$ which contain x. That is,

$$D_n^s(x) = \binom{n}{p+1}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{p+1} \leq n} \mathbb{I}(x \in S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})), \qquad (2.4)$$

where $X_1, X_2, ..., X_n$ is a random sample from F, and $p + 1 \leq n$.

Definition 2.1.6 The *simplicial depth* $D^s(x)$ for a point $x \in \mathbf{R}^p$ is equal to the probability that x is contained in the random simplex $S(X_1, X_2, ..., X_{p+1})$ generated by the $p+1$ i.i.d. observations $X_1, X_2, ..., X_{p+1}$:

$$D^s(x) = P(x \in S(X_1, X_2, ..., X_{p+1})). \tag{2.5}$$

The definitions of the p-dimensional simplicial median and sample simplicial median follow similarly:

Definition 2.1.7 The *multivariate simplicial median* μ is that value (or the average of those values) $x \in \mathbf{R}^p$ that maximizes the depth function given by Equation (2.5).

Definition 2.1.8 The *multivariate sample simplicial median* $\hat{\mu}_n$ is that data value X_i (or the averages of those data values) which maximizes the function given in Equation (2.4).

Given a point $x \in \mathbf{R}^p$ and $p+1$ random observations from F, our next task is to determine whether x is contained in the simplex generated by these $p+1$ points. We can check whether $x \in S(x_1, x_2, ..., x_{p+1})$ by solving the system of linear equations:

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_{p+1} x_{p+1}, \ \alpha_1 + \alpha_2 + \cdots + \alpha_{p+1} = 1, \tag{2.6}$$

under the constraint that $\alpha_i \geq 0 \ \forall \ i$.

Remark 2.1.9 For a nondegenerate simplex (which occurs almost surely when taking random observations from an absolutely continuous distribution, i.e. $P(X_1, X_2, ..., X_{p+1}$ are "co-hyperplanar") $= 0$), this system with $p+1$ unknowns $\alpha_1, \alpha_2, ..., \alpha_{p+1}$ has a unique solution, and x is in the interior of the simplex if and only if $\alpha_1, \alpha_2, ..., \alpha_{p+1}$ are all positive.

Example 2.1.10 From Figure 2.1, it is obvious that the bivariate point $x_0 = (5,4)$ is inside the triangle $\Delta(x_1, x_2, x_3)$, where $x_1 = (1,2), x_2 = (4,6), x_3 = (6,3)$. We can verify this fact by solving Equation (2.6):

$$\begin{aligned} \alpha_1(1,2) + \alpha_2(4,6) + \alpha_3(6,3) = (5,4), \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \Rightarrow \alpha_1 + 4\alpha_2 + 6\alpha_3 = 5 \\ 2\alpha_1 + 6\alpha_2 + 3\alpha_3 = 4 \\ \alpha_1 + \alpha_2 + \alpha_3 = 1 \end{aligned}$$

$$\Rightarrow \alpha_1 = \frac{1}{17}, \alpha_2 = \frac{6}{17}, \alpha_3 = \frac{10}{17},$$

and by Equation (2.6), we conclude that x_0 is in fact in the simplex.

If we look at the point $x_0' = (2,5)$ in Figure 2.1, it is clear that it is not contained in the triangle. We use Equation (2.6) to verify this:

$$\begin{aligned} \alpha_1(1,2) + \alpha_2(4,6) + \alpha_3(6,3) = (2,5), \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \Rightarrow \alpha_1 + 4\alpha_2 + 6\alpha_3 = 2 \\ 2\alpha_1 + 6\alpha_2 + 3\alpha_3 = 5 \\ \alpha_1 + \alpha_2 + \alpha_3 = 1 \end{aligned}$$

$$\Rightarrow \alpha_1 = \frac{8}{17}, \alpha_2 = \frac{14}{17}, \alpha_3 = -\frac{5}{17},$$

which does not satisfy the constraint that $\alpha_i \geq 0 \; \forall \; i$, and so we conclude that x_0' is not in the simplex.

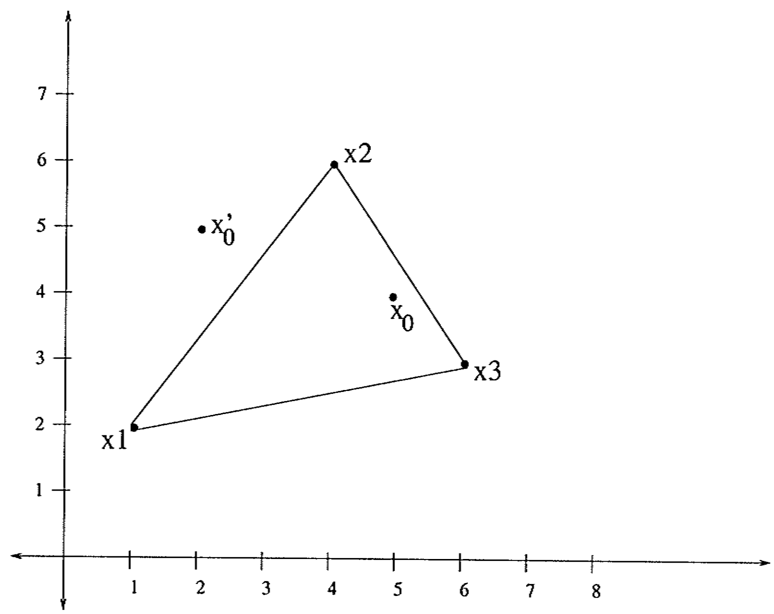

Figure 2.1: The triangle $\Delta(x_1, x_2, x_3)$ generated by the points $x_1 = (1, 2), x_2 = (4, 6), x_3 = (6, 3)$.

In practice, especially when working in higher dimensions, it is much easier to use (or create) a computer program to make these determinations, rather than row reducing large matrices.

2.2 Properties of Liu's Simplicial Depth

We will now look at some important results from Liu's (1990) paper which describe the behaviour of $D^s(x)$. Theorems 2.2.1 and 2.2.4 are stated but not proven in Liu (1990). Theorems 2.2.2 and 2.2.3 are stated and proven in Liu (1990), but the proofs are given here in more detail, and in the general case of p dimensions, whereas the proofs in Liu (1990) were only given for $p = 2$.

Theorem 2.2.1 *$D^s(x)$ is invariant under affine transformations. That is, if A is a non-singular $p \times p$ matrix and $b \in \mathbf{R}^p$, then*

$$D^s_{A,b}(Ax + b) = D^s(x),$$

where $D^s_{A,b}(y)$ is the probability that $y \in \mathbf{R}^p$ is contained inside the simplex with vertices $AX_{i_j} + b, j = 1, 2, ..., p + 1$. In our case, $y = Ax + b$, and we view all elements in $\mathbf{R}^p$ as $p \times 1$ column vectors.

Proof. It is enough to show that

$$x \in S(x_1, x_2, ..., x_{p+1}) \Leftrightarrow Ax + b \in S(Ax_1 + b, Ax_2 + b, ..., Ax_{p+1} + b)$$

Now,

$$\begin{aligned} x \in S(x_1, x_2, ..., x_{p+1}) \quad &\Leftrightarrow \quad x = \sum_{i=1}^{p+1} \alpha_i x_i \text{ where } \sum_{i=1}^{p+1} \alpha_i = 1,\ \alpha_i \geq 0 \,\forall\, i \\ &\Leftrightarrow \quad Ax + b = A\left(\sum_{i=1}^{p+1} \alpha_i x_i\right) + b \\ &\qquad = \sum_{i=1}^{p+1} \alpha_i (Ax_i) + b \end{aligned}$$

$$= \sum_{i=1}^{p+1} \alpha_i (Ax_i + b)$$
$$\Leftrightarrow \quad Ax + b \in S(Ax_1 + b, Ax_2 + b, ..., Ax_{p+1} + b)$$

This completes the proof. Note that the second to last step is possible since $b = \sum_{i=1}^{p+1} \alpha_i b$. $D_n^s(x)$ is also invariant under affine transformations by a similar argument. **QED**

In other words, instead of forming simplices using $X_{i_1}, X_{i_2}, ..., X_{ip+1}$ and finding the proportion of these simplices containing x, we can equivalently form simplices using $AX_{i_1}+b, AX_{i_2}+b, ..., AX_{i_{p+1}}+b$, and find the proportion of the latter simplices containing $Ax + b$. The proportion will be the same, as asserted by this property.

The following result asserts that $D^s(x)$ vanishes uniformly fast as $||x|| \to \infty$. The norm, $||x||$ is defined in Appendix A.

Theorem 2.2.2 *For any c.d.f. F on $\mathbf{R}^p$, $\sup_{||x|| \geq M} D^s(x) \to 0$ as $M \to \infty$.*

Proof. Let $X_1, X_2, ..., X_{p+1}$ be i.i.d. with c.d.f. F. Given any $x \in \mathbf{R}^p$, we observe that $\{x \in S(X_1, X_2, ..., X_{p+1})\} \subseteq \bigcup_{i=1}^{p+1}\{||X_i|| \geq ||x||\}$. This is because, if x is further from the origin than any of the $p+1$ random points, it clearly cannot be contained in the simplex generated by the $p+1$ points. Using the above inclusion and Lemma A.2.1, we get

$$\begin{aligned} D^s(x) &= P(x \in S(X_1, X_2, ..., X_{p+1})) \\ &\leq P\left(\bigcup_{i=1}^{p+1}\{||X_i|| \geq ||x||\}\right) \\ &\leq \sum_{i=1}^{p+1} P(||X_i|| \geq ||x||) \text{ (by subadditivity)} \end{aligned}$$

$$= \quad (p+1)P(||X_1|| \geq ||x||).$$

Note that the last step is possible by the identical distributions of the X_i. So clearly, since $P(||X_1|| \geq ||x||)$ is decreasing in $||x||$, it follows that

$$\sup_{||x|| \geq M} D^s(x) \leq \sup_{||x|| \leq M} (p+1)P(||X_1|| \geq ||x||) \leq (p+1)P(||X_1|| \geq M)$$

Since $P(||X_1|| \geq ||M||) \to 0$ as $M \to \infty$, we get the desired result, namely

$$\sup_{||x|| \geq M} D^s(x) \to 0 \,\text{as}\, M \to \infty.$$

This completes the proof. **QED**

The following result asserts the continuity of the simplicial depth function.

Theorem 2.2.3 *If F is an absolutely continuous distribution on $\mathbf{R}^p$, then $D^s(x)$ is continuous at every $x \in \mathbf{R}^p$.*

Proof. Let $X_1, X_2, ..., X_{p+1}$ be i.i.d. with c.d.f. F. We let $\{x_n\}$ be a sequence in $\mathbf{R}^p$ such that $x_n \to x$, and show that

$$|D^s(x) - D^s(x_n)| \leq (p+1)P(CH(X_1, X_2, ..., X_p) \cap \overline{xx_n}),$$

where $CH(X_1, X_2, ..., X_p)$ is the convex hull of p points in $\mathbf{R}^p$. Note that this represents a "face" of the simplex $S(X_1, X_2, ..., X_{p+1})$.

Note that, in the context of intersecting line or hyperplane segments, "$\cap$" refers to the two crossing one another at some point. This differs from our usual use of the intersection symbol "$\cap$", meaning the intersection of two sets or events.

A random simplex can contribute to the difference $D^s(x) - D^s(x_n)$ only if it contains one point but not the other. This however implies that $\overline{xx_n}$ passes through exactly one face of the simplex $S(X_1, X_2, ..., X_{p+1})$. See Figure 2.2 for an illustration for the case when $p = 2$.

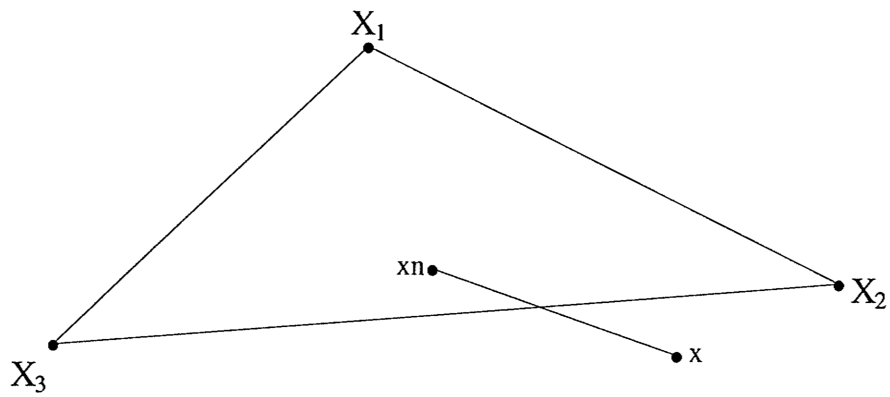

Figure 2.2: $x_n \in \Delta(X_1, X_2, X_3)$ and $x \notin \Delta(X_1, X_2, X_3)$, and so $\overline{xx_n} \cap \overline{X_2X_3}$. Therefore, this simplex contributes to the difference $D^s(x) - D^s(x_n)$.

For any two events A and B, $P(A \backslash B) = P(A) - P(A \cap B) \geq P(A) - P(B)$. Therefore, if we define

$$\begin{aligned} A &= [x \in S(X_1, X_2, ..., X_{p+1})] \\ B &= [x_n \in S(X_1, X_2, ..., X_{p+1})], \end{aligned}$$

we have, by Lemma A.2.1,

$$\begin{aligned} D^s(x) - D^s(x_n) &= P(A) - P(B) \\ &\leq P(x \in S(X_1, X_2, ..., X_{p+1}) \cap x_n \notin S(X_1, X_2, ..., X_{p+1})) \\ &\leq P\left(\bigcup_{i=1,2,...,p+1} (CH(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_{p+1}) \cap \overline{xx_n}) \right) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{p+1} P(CH(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_{p+1}) \cap \overline{xx_n}) \\
&= (p+1)P(CH(X_1, X_2, ..., X_p) \cap \overline{xx_n}) \\
&\leq (p+1)P(\overline{X_1X_2...X_p} \cap \overline{xx_n}),
\end{aligned}
$$

where $\overline{X_1X_2...X_p}$ is the p-dimensional hyperplane containing the points $X_1, X_2,...,$ X_p. This follows since, $CH(X_1, X_2, ..., X_p) \subseteq \overline{X_1X_2...X_p}$ and since $X_1, X_2, ..., X_p$ are identically distributed. It can similarly be shown (or simply understood by symmetry) that

$$D^s(x_n) - D^s(x) \leq (p+1)P(\overline{X_1X_2...X_p} \cap \overline{xx_n}).$$

As such, we have

$$|D^s(x) - D^s(x_n)| \leq (p+1)P(\overline{X_1X_2...X_p} \cap \overline{xx_n}).$$

We define the event

$$A_n = \{\overline{X_1X_2...X_p} \cap \overline{xx_n}\} \;\; \forall \;\; n.$$

Then

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \left[\bigcup_{k=n}^{\infty} A_k \right] = \{x \in \overline{X_1X_2...X_p}\}.$$

By Lemma A.2.2, we know that

$$\limsup_{n\to\infty} P(A_n) \leq P(\limsup_{n\to\infty} A_n) = P(x \in \overline{X_1X_2...X_p}) = 0, \qquad (2.7)$$

since F is absolutely continuous.

We can make the above assertion because of the following: We note that, by conditioning on $X_1, X_2, ..., X_{p-1}$ (Lemma A.2.4),

$$
\begin{aligned}
& P(x \in \overline{X_1 X_2 ... X_p}) \\
= & \iint_{(\mathbf{R}^2)^{p-1}} P(x \in \overline{X_1 X_2 ... X_p} | X_1 = x_1, X_2 = x_2, ..., X_{p-1} = x_{p-1})) \\
& f(x_1) f(x_2) \cdots f(x_{p-1}) dx_1 dx_2 \cdots dx_{p-1} \\
= & \iint_{(\mathbf{R}^2)^{p-1}} P(x \in x_1 x_2 ... X_p) f(x_1) f(x_2) \cdots f(x_{p-1}) dx_1 dx_2 \cdots dx_{p-1} \\
= & \iint_{(\mathbf{R}^2)^{p-1}} 0 f(x_1) f(x_2) \cdots f(x_{p-1}) dx_1 dx_2 \cdots dx_{p-1} = 0,
\end{aligned}
$$

by the absolute continuity of F. That is,

$$
\begin{aligned}
P(x \in \overline{x_1 x_2 ... X_p}) &= P(X_p \in \overline{x x_1 x_2 ... x_{p-1}}) \\
&= P(X_p \text{ lies on a } p\text{−dimensional hyperplane}) = 0
\end{aligned}
$$

Therefore, $|D^s(x) - D^s(x_n)| \to 0$ as $n \to \infty$. This completes the proof, using the sequential characterization of continuity. **QED**

Theorem 2.2.4 *$D_n^s(x)$ is an unbiased estimator for $D^s(x)$.*

Proof. Let $X_1, X_2, ..., X_p$ be i.i.d. with c.d.f. F. Since expectation is a linear operator, we have

$$
\begin{aligned}
E\left[D_n^s(x)\right] &= E\left[\binom{n}{p+1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{p+1} \le n} \mathbb{I}(x \in S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}}))\right] \\
&= \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{p+1} \le n} E\left[\mathbb{I}(x \in S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}}))\right] \\
&= \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{p+1} \le n} P(x \in S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})) \\
&= \binom{n}{p+1}^{-1} \binom{n}{p+1} P(x \in S(X_1, X_2, ..., X_{p+1}))
\end{aligned}
$$

$$= P(x \in S(X_1, X_2, ..., X_{p+1}))$$
$$= D^s(x).$$

This completes the proof. **QED**

It can be shown that

$$Var(D_n^s(x)) = E[(D_n^s(x) - D_n(x))]^2 \rightarrow 0 \tag{2.8}$$

as $n \rightarrow \infty$ (See Lemma 3.3.1.) A consequence of Equation (2.8) is that $D_n^s(x)$ is a weakly consistent estimator of $D^s(x)$ at each fixed $x \in \mathbf{R}^p$, i.e. $D_n^s(x) \rightarrow D^s(x)$ in probability for every fixed $x \in \mathbf{R}^p$. (See Appendix A for the definition of convergence in probability.) The following theorem, which appears in Liu (1990), asserts something much stronger, namely, the uniform strong consistency of $D_n^s(x)$. Its proof uses Gilvenko-Cantelli classes and other in-depth notions in probability theory, and is therefore omitted in this thesis.

Theorem 2.2.5 *If F is an absolutely continuous distribution on $\mathbf{R}^p$ with bounded density f, then D_n^s is uniformly consistent, i.e.*

$$\sup_{x \in \mathbf{R}^p} |D_n^s(x) - D^s(x)| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

(See Appendix A for the definition of almost sure convergence).

In fact, Liu (1990) gives conditions under which $\hat{\mu}_n$, the sample simplicial median, is a strongly consistent estimator of μ.

Theorem 2.2.6 *If F is absolutely continuous on $\mathbf{R}^p$ with bounded density $f(x)$ which does not vanish in a neighbourhood of μ and if $D^s(x)$ is uniquely maximized at μ, then $\hat{\mu}_n \to \mu$ almost surely as $n \to \infty$.*

2.3 Angularly Symmetric Distributions and Their Depth

In this section, we study the class of angularly symmetric distributions, a class for which it is natural to speak of a "median point". This will be reflected in the properties of the associated depth functions. We will first define a better known class of distributions.

Definition 2.3.1 A random variable $X \in \mathbf{R}^p$ has a *centrally symmetric distribution about $\theta \in \mathbf{R}^p$* if

$$X - \theta \stackrel{d}{=} \theta - X,$$

where "$\stackrel{d}{=}$" denotes equality in distribution.

Zuo and Serfling (2000) give the following lemma:

Lemma 2.3.2 *Definition 2.3.1 is equivalent to stating that*

$$P(X - \theta \in H) = P(X - \theta \in -H)$$

for each closed halfspace $H \subseteq \mathbf{R}^p$, where $-H = \{-x : x \in H\}$.

In other words, a random variable X which follows a centrally symmetric distribution about θ has the same probability of falling in any halfspace

$H \subseteq \mathbf{R}^p$ as it does in the halfspace $-H \subseteq \mathbf{R}^p$, which is the mirror image of H across the point of symmetry θ.

Definition 2.3.3 A random variable $X \in \mathbf{R}^p$ has an *angularly symmetric distribution about* $\theta \in \mathbf{R}^p$ if

$$\frac{X-\theta}{||X-\theta||} \stackrel{d}{=} \frac{\theta - X}{||X-\theta||}.$$

Lemma 2.3.4 *Definition 2.3.3 is equivalent to stating that*

$$P(X - \theta \in H^*) = P(X - \theta \in -H^*)$$

for each closed halfspace $H^* \subseteq \mathbf{R}^p$ *which passes through the origin.*

In other words, a random variable X which follows a distribution which is angularly symmetric about θ has the same probability of falling in any halfspace H^* determined by θ as it does in the halfspace which is the mirror image of H^* across the point of angular symmetry. Consequently, if F is absolutely continuous, $P(H^*) = \frac{1}{2}$ for every half-space determined by θ. In this sense, θ can be viewed as a multivariate median of F.

By Definitions 2.3.1 and 2.3.3, it follows that X has an angularly symmetric distribution about θ if and only if $(X-\theta)/(||X-\theta||)$ has a centrally symmetric distribution about the origin. Clearly, X is (angularly) symmetric about θ if and only if $X - \theta$ is (angularly) symmetric about the origin.

Central symmetry (or just symmetry as it is usually called) is stronger than angular symmetry, as is apparent from Definitions 2.3.1 and 2.3.3. That is, symmetry implies angular symmetry. All multivariate normal distributions are both symmetric and thus angularly symmetric about their mean vector μ. The difference between symmetry and angular symmetry is apparent in the following example:

Example 2.3.5 Consider a bivariate random variable X which has density

$$f(x,y) = \begin{cases} \frac{1}{2\pi} & \text{when } x \geq 0, y \geq 0, x^2 + y^2 \leq 4 \\ \frac{2}{\pi} & \text{if } x \leq 0, y \leq 0, x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

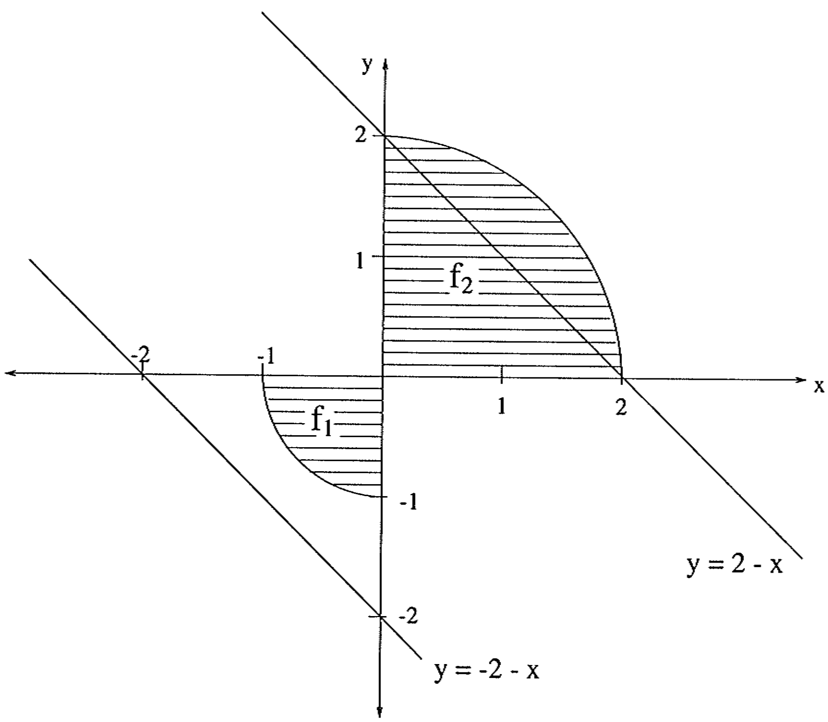

Figure 2.3: An illustration of a probability density function $f(x)$ that is angularly symmetric but not symmetric. The support of $f(x)$ is shaded. f is constant with $f_1(x) = \frac{2}{\pi}$ and $f_2(x) = \frac{1}{2\pi}$.

See Figure 2.3 for an illustration of this probability density function (p.d.f.) X obviously has an angularly symmetric about the origin. Any halfspace determined by $(0,0)$ will have equal probability on either side. However, the condition for central symmetry does not hold. If we consider the halfspace $H = \{(x,y): \ y \geq 2-x\}$ and its mirror image about the origin, $-H = \{(x,y): \ y \leq -2-x\}$, we see that

$$P(X \in H) > 0 \neq P(X \in -H) = 0.$$

Since the condition in Lemma 2.3.2 must hold for all halfspaces $H \subseteq \mathbf{R}^2$, it follows that the distribution of X is angularly symmetric about the origin, but not symmetric.

Theorem 2.3.6 *Every univariate p.d.f. $f(x)$ is angularly symmetric about its median M.*

Proof: There is only one hyperplane that passes through the origin, namely the "line" $x = 0$. This creates two closed halfspaces $H^* = [0, \infty]$ and $-H^* = [-\infty, 0]$. By definition of the median, $P(X - M \in H^*) = \frac{1}{2} = P(X - M \in -H^*)$, and the proof is complete. **QED**

The following two theorems require that the c.d.f. F be angularly symmetric. Their proofs can be found in Liu (1990).

The first result is that of monotonicity, a useful property for a depth function. It states that the depth decreases steadily as we move away from the centre of the distribution in any given direction.

Theorem 2.3.7 *If F is absolutely continuous and angularly symmetric about the origin, then $D^s(\alpha x)$ is a monotone nonincreasing function in $\alpha \geq 0 \ \forall \ x \in$ $\mathbf{R}^p$.*

Theorem 2.3.8 *If F is an absolutely continuous distribution on $\mathbf{R}^p$ and it is angularly symmetric about a point $\theta \in \mathbf{R}^p$, then $D^s(\theta) = 2^{-p}$.*

Note that Theorems 2.3.7 and 2.3.8 imply that $D^s(x)$ attains its maximum at its point of angular symmetry θ, where its depth is $D^s(\theta) = 2^{-p}$, and that for every point $x \in \mathbf{R}^p$, we have $D^s(x) \leq 2^{-p}$. Consequently, by Definition 2.1.7, θ is the multivariate simplicial median of X.

One application of the above property is in testing the centre of angular symmetry. As mentioned above, if F is angularly symmetric, then $D^s(x)$ is maximized at the centre of angular symmetry and takes there the value 2^{-p}. As such, if θ_0 is a hypothesized centre of angular symmetry for some density function $f(x)$, then a large value of $(2^{-p} - D^s_n(\theta_0))$ is an indication of the null hypothesis being false. For a more in-depth discussion of the inference procedure, see Gregory (1977).

2.4 New Properties of Liu's Simplicial Depth

Liu (1990) proposed the simplicial depth function and established some very important properties thereof. The unbiasedness and consistency of $D^s_n(x)$ were asserted and proven. It was shown that $D^s(x) \to 0$ uniformly as $||x||$ tends to infinity, and that $D^s(x)$ is continuous and affine invariant. For angularly symmetric distributions, it was also shown that $D^s(x)$ is a monotone nonincreasing function, which attains its maximum value of 2^{-p} at the point of angular symmetry.

We now examine some properties of the simplicial depth function $D^s(x)$ that, to our knowledge, have heretofore gone unmentioned or uncalculated in the literature. Unless otherwise noted, all results in this section are proven for the case of $p = 2$. This case is the simplest non-trivial setting for Liu's depth function, as well as the easiest to visualize. All of our theorems and proofs can be extended or generalized to the p-dimensional case by making the obvious modifications.

One difficulty encountered with Liu's simplicial depth function is that it is generally intractable. That is, there are no results published that enable us to actually calculate $D^s(x)$ for a given distribution F. This is due to the complexity of the mathematical description of a "random simplex". In Theorem 2.4.3, we will find an expression for this function, which will clearly illustrate its complexity. How do we go about calculating the probability that a fixed point x_0 will fall within a random triangle generated by i.i.d. random variables X_1, X_2, X_3?

We will do this with the help of the process of conditioning. We ask the question: Given the values $X_1 = x_1$ $(=(x_{11}, x_{12}))$ and $X_2 = x_2$ $(=(x_{21}, x_{22}))$, where must the value of X_3 fall in order for the triangle $\Delta(X_1, X_2, X_3)$ to contain a fixed point x_0 $(=(x_{01}, x_{02}))$? It turns out that this is not a simple problem to solve. The conditional probability

$$P(x_0 \in \Delta(X_1, X_2, X_3) | X_1 = x_1, X_2 = x_2) = P(x_0 \in \Delta(x_1, x_2, X_3))$$

is itself conditional on the ordering of the components of x_0, x_1 and x_2. The first components (which we normally speak of as the x-components) of the three points can be ordered in $3! = 6$ ways, as can the second (y-) components. As such, we have to consider 36 separate cases. For any given case, in order for x_0 to be in the triangle, X_3 must fall in the region (labelled A) bounded by L_1 and L_2, the line segments formed by passing through x_0 and x_1, and between x_0 and x_2, respectively. See Figure 2.4 for an illustration.

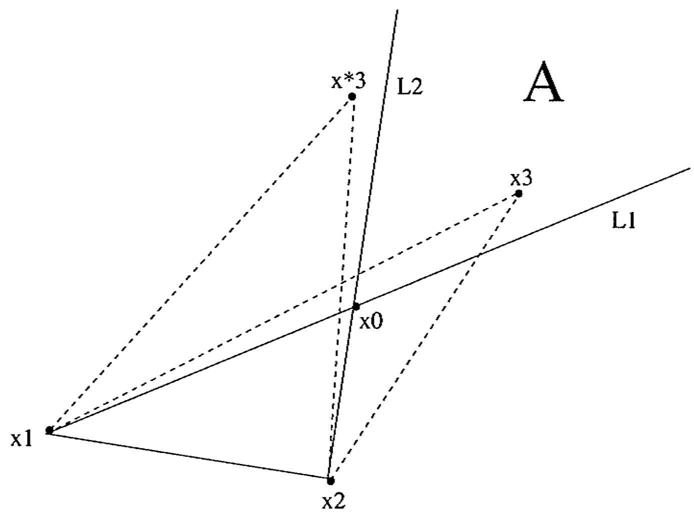

Figure 2.4: $x_3 \in A$, and so $x_0 \in \Delta(x_1, x_2, x_3)$, whereas $x_3^* \notin A$, and so $x_0 \notin \Delta(x_1, x_2, x_3)$.

We find equations for the lines, using our usual notion of x- and y-components:

For any point $(x, y) \in \mathbf{R}^2$ on the line L_1, we have that

$$\frac{y - x_{02}}{x - x_{01}} = \frac{x_{12} - x_{02}}{x_{11} - x_{01}}$$

$$\Rightarrow \quad y = \frac{(x_{12} - x_{02})(x - x_{01})}{x_{11} - x_{01}} + x_{02}$$

$$\Rightarrow \quad y = \left(\frac{(x_{12} - x_{02})x_{01}}{x_{01} - x_{11}} + x_{02} \right) + \left(\frac{x_{12} - x_{02}}{x_{11} - x_{01}} \right) x,$$

which is the equation for L_1. Equivalently, L_1 can be expressed as a function of y.

Similarly, we can find the equation of L_2:

$$\frac{y - x_{02}}{x - x_{01}} = \frac{x_{22} - x_{02}}{x_{21} - x_{01}}$$

$$\Rightarrow \quad y = \frac{(x_{22} - x_{02})(x - x_{01})}{x_{21} - x_{01}} + x_{02}$$

$$\Rightarrow \quad y = \left(\frac{(x_{22} - x_{02})x_{01}}{x_{01} - x_{21}} + x_{02} \right) + \left(\frac{x_{22} - x_{02}}{x_{21} - x_{01}} \right) x,$$

which is the equation for L_2. Equivalently, L_2 can be expressed as a function of y.

Example 2.4.1 We consider the case for $x_{01} < x_{11} < x_{21}$ and $x_{22} < x_{12} < x_{02}$. We can see from Figure B.6 that even this case must be divided into two further cases. The region in which X_3 must fall in order for $x_0 \in \Delta(x_1, x_2, X_3)$ depends on whether the angle between x_0 and x_1 is steeper than the angle between x_0 and x_2. Now, in the first case (pictured on top in Figure B.6),

$$P(x_0 \in \Delta(x_1, x_2, X_3)) = P(X_3 \in A) = \int_{-\infty}^{x_{01}} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31}.$$

Similarly, in the second case (pictured on the bottom),

$$P(x_0 \in \Delta(x_1, x_2, X_3)) = P(X_3 \in A) = \int_{-\infty}^{x_{01}} \int_{L_2}^{L_1} f(x_{31}, x_{32}) dx_{32} dx_{31}.$$

So in total, regardless of the aforementioned angles,

$$P(x_0 \in \Delta(x_1, x_2, X_3)) = \left| \int_{-\infty}^{x_{01}} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31} \right|.$$

This last result is derived from the calculus property

$$\int_c^d \int_{a(y)}^{b(y)} f(x, y) dx dy = - \int_c^d \int_{b(y)}^{a(y)} f(x, y) dx dy.$$

Fortunately, we only need to separate a case into two distinct cases when the second highest x- and y-components both belong to the same point (i.e. either $(x_0, x_1,$ or $x_2)$). The other cases generally go as follows:

Example 2.4.2 Consider now the case for $x_{21} < x_{01} < x_{11}$ and $x_{12} < x_{22} < x_{02}$. We see from Figure B.28 that the slopes of the two lines L_1 and L_2 do not matter in this case; we will get the same result either way:

$$\begin{aligned} P(x \in \Delta(x_1, x_2, X_3)) = P(X_3 \in A) &= \int_{x_{02}}^{\infty} \int_{L_1}^{x_{01}} f(x_{31}, x_{32}) dx_{31} dx_{32} \\ &+ \int_{x_{01}}^{\infty} \int_{L_2}^{\infty} f(x_{31}, x_{32}) dx_{32} dx_{31}. \end{aligned}$$

As mentioned, there are 36 separate cases we have to consider. For a complete illustration of each case, see Appendix B. We are now ready to give the general result:

Theorem 2.4.3 *Liu's simplicial depth function $D^s(x)$ can be expressed as:*

$$D^s(x) = \sum_{i=1}^{12} \left[\iint_{\mathbb{R}^2} \iint_{\mathbb{R}^2} \mathbb{I}(A_i') h_i(x_{11}, x_{12}; x_{21}, x_{22}) f(x_1) f(x_2) dx_1 dx_2 \right] \tag{2.9}$$

where A_i' is the set of all $(x_{11}, x_{12}; x_{21}, x_{22}) \in \mathbf{R}^4$ which satisify the condition in the event A_i in the proof that follows, where X is replaced by x therein, and where

$$\begin{aligned}
h_1(x_{11}, x_{12}; x_{21}, x_{22}) &= \left| \int_{-\infty}^{x_{01}} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31} \right| \\
h_2(x_{11}, x_{12}; x_{21}, x_{22}) &= \left| \int_{x_{01}}^{\infty} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31} \right| \\
h_3(x_{11}, x_{12}; x_{21}, x_{22}) &= \left| \int_{-\infty}^{x_{02}} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{31} dx_{32} \right| \\
h_4(x_{11}, x_{12}; x_{21}, x_{22}) &= \left| \int_{x_{02}}^{\infty} \int_{L_1}^{L_2} f(x_{31}, x_{32}) dx_{31} dx_{32} \right| \\
h_5(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{-\infty}^{x_{01}} \int_{L_2}^{\infty} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
&\quad + \int_{x_{02}}^{\infty} \int_{x_{01}}^{L_1} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
h_6(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{-\infty}^{x_{02}} \int_{L_2}^{x_{01}} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
&\quad + \int_{x_{01}}^{\infty} \int_{-\infty}^{L_1} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_7(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{-\infty}^{x_{02}} \int_{x_{01}}^{L_1} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
&\quad + \int_{-\infty}^{x_{01}} \int_{-\infty}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_8(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{x_{02}}^{\infty} \int_{L_2}^{x_{01}} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
&\quad + \int_{x_{01}}^{\infty} \int_{L_1}^{\infty} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_9(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{x_{02}}^{\infty} \int_{L_1}^{x_{01}} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
&\quad + \int_{x_{01}}^{\infty} \int_{L_2}^{\infty} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_{10}(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{-\infty}^{x_{02}} \int_{x_{01}}^{L_2} f(x_{31}, x_{32}) dx_{31} dx_{32}
\end{aligned}$$

$$
\begin{aligned}
& \quad + \int_{-\infty}^{x_{01}} \int_{-\infty}^{L_1} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_{11}(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{-\infty}^{x_{02}} \int_{L_1}^{x_{01}} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
& \quad + \int_{x_{01}}^{\infty} \int_{-\infty}^{L_2} f(x_{31}, x_{32}) dx_{32} dx_{31} \\
h_{12}(x_{11}, x_{12}; x_{21}, x_{22}) &= \int_{x_{02}}^{\infty} \int_{x_{01}}^{L_2} f(x_{31}, x_{32}) dx_{31} dx_{32} \\
& \quad + \int_{-\infty}^{x_{01}} \int_{L_1}^{\infty} f(x_{31}, x_{32}) dx_{32} dx_{31}.
\end{aligned}
$$

Proof. We define events

$$
\begin{aligned}
A_1 &= [x_{01} < X_{11}] \cap [x_{01} < X_{21}] \\
A_2 &= [X_{11} < x_{01}] \cap [X_{21} < x_{01}] \\
A_3 &= [x_{02} < X_{12} \cap x_{02} < X_{22}] \cap [(X_{21} < x_{01} < X_{11}) \cup (X_{11} < x_{01} < X_{21})] \\
A_4 &= [X_{12} < x_{02} \cap X_{22} < x_{02}] \cap [(X_{21} < x_{01} < X_{11}) \cup (X_{11} < x_{01} < X_{21})] \\
A_5 &= [X_{11} < x_{01} < X_{21}] \cap [X_{12} < x_{02} < X_{22}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| < \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_6 &= [X_{11} < x_{01} < X_{21}] \cap [X_{12} < x_{02} < X_{22}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| > \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_7 &= [X_{11} < x_{01} < X_{21}] \cap [X_{22} < x_{02} < X_{12}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| < \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_8 &= [X_{11} < x_{01} < X_{21}] \cap [X_{22} < x_{02} < X_{12}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| > \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_9 &= [X_{21} < x_{01} < X_{11}] \cap [X_{12} < x_{02} < X_{22}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| < \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_{10} &= [X_{21} < x_{01} < X_{11}] \cap [X_{12} < x_{02} < X_{22}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| > \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_{11} &= [X_{21} < x_{01} < X_{11}] \cap [X_{22} < x_{02} < X_{12}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| < \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right] \\
A_{12} &= [X_{21} < x_{01} < X_{11}] \cap [X_{22} < x_{02} < X_{12}] \cap \left[\left|\frac{X_{22} - x_{02}}{X_{21} - x_{01}}\right| > \left|\frac{X_{12} - x_{02}}{X_{11} - x_{01}}\right|\right].
\end{aligned}
$$

$A_1, A_2, \ldots, A_{12}$ are pairwise mutually exclusive and exhaustive. Therefore, by

Lemma A.2.3,

$$\begin{aligned}
D^s(x_0) &= P(x_0 \in \Delta(X_1, X_2, X_3)) \\
&= \sum_{i=1}^{12} P(x_0 \in \Delta(X_1, X_2, X_3 | A_i) P(A_i) \\
&= \sum_{i=1}^{12} P(x_0 \in \Delta(X_1, X_2, X_3 \cap A_i) \\
&= \sum_{i=1}^{12} \int_{(\mathbf{R}^2)^2} \mathbb{I}(A_i') h_i(x_{11}, x_{12}; x_{21}, x_{22}) f(x_1) f(x_2) dx_1 dx_2.
\end{aligned}$$

In particular by Lemma A.2.5,

$$\begin{aligned}
& P(x_0 \in \Delta(X_1, X_2, X_3) \cap A_i) \\
= \; & E[\mathbb{I}(x_0 \in \Delta(X_1, X_2, X_3))\mathbb{I}(A_i)] \\
= \; & E(E[\mathbb{I}(x_0 \in \Delta(X_1, X_2, X_3))\mathbb{I}(A_i) | (X_1, X_2)]) \\
= \; & \iint_{\mathbf{R}^2} \iint_{\mathbf{R}^2} E[\mathbb{I}(x_0 \in \Delta(x_1, x_2, X_3))\mathbb{I}(A_i')] f(x_1) f(x_2) dx_1 dx_2 \\
= \; & \iint_{\mathbf{R}^2} \iint_{\mathbf{R}^2} \mathbb{I}(A_i') P(x_0 \in \Delta(x_1, x_2, X_3)) f(x_1) f(x_2) dx_1 dx_2 \\
= \; & \iint_{\mathbf{R}^2} \iint_{\mathbf{R}^2} \mathbb{I}(A_i') h(x_{11}, x_{12}; x_{21}, x_{22}) f(x_1) f(x_2) dx_1 dx_2
\end{aligned}$$

Note that we also used the identical distributions of X_1 and X_2. The rest of the proof follows from tedious but straighforward calculations similar to those in Examples 2.4.1 and 2.4.2, based on the cases depicted in Appendix B. Note that we have grouped the 36 original cases into 12 by "gathering terms". The terms that were gathered into the respective cases will also be outlined in Appendix B. **QED**

It is clear from Theorem 2.4.3 that the functional form of $D^s(x)$ is very complex. Even $P(A_1), P(A_2), ..., P(A_{12})$ would be very difficult to calculate for most distributions. We notice that it would be even more difficult to calculate $D^s(x)$ for a distribution $f(x)$ with a finite support. Even if we were to consider the basic Uniform$[0,1]^2$ distribution on $\mathbf{R}^2$, we would have to break each of the 12 cases up even further to account for the possibility that L_1 and L_2 may intersect the edge of the support on any of the four lines $x = 0, x = 1, y = 0$, or $y = 1$.

At the very least, the preceding calculation offers a perspective of why we need the estimator $D_n^s(x)$ in order to estimate $D^s(x)$.

Liu (1990) proved for any absolutely continuous angularly symmetric distribution F on $\mathbf{R}^p$, that $D^s(x) \leq 2^{-p}\ \forall\ x \in \mathbf{R}^p$. The family of angularly symmetric distributions is large, but we might have to deal with distributions which do not have this property. We will now determine an upper bound for the **mean** of a random version of simplicial depth function in two dimensions, for **any** continuous bivariate distribution F. This will involve the concept of extreme points of a random sample, a topic studied extensively in Efron (1965).

Definition 2.4.4 Given a random sample $X_1, X_2, ..., X_n$, the discrete random variable $E^{(n)}$ denotes the number of extreme points in the convex hull determined by $X_1, X_2, ..., X_n$.

In a bivariate sample, we picture spikes sticking out from the plane at every point X_i from the random sample. We wrap a string around the periphery of the spikes, and the number of spikes touched by the string corresponds to the number of extreme points in the sample.

Given X, X_1, X_2, X_3, an i.i.d. random sample with c.d.f. F on $\mathbf{R}^p$, we define the random variable $D^s(X)$, where, on the event $\{X = x\}$, $x \in \mathbf{R}^p$, $D^s(X) = D^s(x)$. That is, $D^s(X)$ can be viewed as a random version of $D^s(x)$.

Theorem 2.4.5 *Let F be* **any** *absolutely continuous distribution on $\mathbf{R}^2$. Then $E(D^s(X)) \leq \frac{1}{4}$.*

Proof: Let X_1, X_2, X_3, X_4 be i.i.d. random variables with absolutely continuous c.d.f. F. Then, conditioning on X_4, by the "law of the unconscious statistician",

$$\begin{aligned} P(X_4 \in \Delta(X_1, X_2, X_3)) &= \iint_{\mathbf{R}^2} P(x_4 \in \Delta(X_1, X_2, X_3)) f(x_4) dx_4 \\ &= E(D(X_4)) = E(D(X)), \end{aligned}$$

and so

$$\begin{aligned} P(E^{(4)} = 3) &= P(X_4 \in \Delta(X_1, X_2, X_3)) \\ &\quad + P(X_3 \in \Delta(X_1, X_2, X_4)) \\ &\quad + P(X_2 \in \Delta(X_1, X_3, X_4)) \\ &\quad + P(X_1 \in \Delta(X_2, X_3, X_4)) \\ &= 4E(D^s(X)). \qquad (2.10) \end{aligned}$$

Equation (2.10) follows from the fact that, if there are only three extreme points in a sample of four points, then the convex hull, a triangle, must contain the fourth point. By mutual exclusivity,

$$P(E^{(4)} = 3) = 4E(D^s(X)) \Rightarrow P(E^{(4)} = 4) = 1 - 4E(D^s(X)).$$

This follows from the fact that there must be either three or four extreme points in a sample of size four from an absolutely continuous distribution F. By definition of expectation, it follows that

$$\begin{aligned} E(E^{(4)}) &= 3(4E(D^s(X))) + 4(1 - 4E(D^s(X))) \\ &= 4 - 4E(D^s(X)). \end{aligned}$$

Since

$$\begin{aligned} 3 &\leq E^{(4)} \leq 4 \\ \Rightarrow 3 &\leq E(E^{(4)}) \leq 4 \\ \Rightarrow 3 &\leq 4 - 4E(D^s(X)) \leq 4 \\ \Rightarrow -1 &\leq -4E(D^s(X)) \leq 0 \\ \Rightarrow 0 &\leq E(D^s(X)) \leq \frac{1}{4}, \end{aligned}$$

this completes the proof. **QED**

We now extend the result to the general case of p dimensions.

Theorem 2.4.6 *Let F be* any *absolutely continous distribution on $\mathbf{R}^p$. Then*

$$E(D^s(X)) \leq \frac{1}{p+2}.$$

Proof: Let $X_1, X_2, ..., X_{p+2}$ be i.i.d. random variables with absolutely continuous c.d.f. F. Then, conditioning on X_{p+2},

$$\begin{aligned} &P(X_{p+2} \in S(X_1, X_2, ..., X_{p+1})) \\ = &\int_{\mathbf{R}^p} P(x_{p+2} \in S(X_1, X_2, ..., X_{p+1})) f(x_{p+2}) dx_{p+2} \\ = &E(D^s(X_{p+2})) = E(D^s(X)), \end{aligned}$$

and so

$$
\begin{aligned}
& P(E^{(p+2)} = p+1) \\
= \ & P(X_{p+2} \in S(X_1, X_2, ..., X_{p+1})) + P(X_{p+1} \in S(X_1, X_2, ..., X_p, X_{p+2})) \\
& + \cdots + P(X_2 \in S(X_1, X_3, ..., X_{p+2})) + P(X_1 \in S(X_2, X_3, ..., X_{p+2})) \\
= \ & (p+2)E(D^s(X)). \qquad (2.11)
\end{aligned}
$$

Equation (2.11) follows from the fact that, if there are only $p+1$ extreme points in a sample of $p+2$ points, then the convex hull, a simplex, must contain the $(p+2)th$ point. By mutual exclusivity,

$$P(E^{(p+2)} = p{+}1) = (p{+}2)E(D^s(X)) \Rightarrow P(E^{p+2} = p{+}2) = 1{-}(p{+}2)E(D^s(X)).$$

This follows from the fact that there must be either $p+1$ or $p+2$ extreme points in a sample of size $p+2$ from an absolutely continuous distribution F. By definition of expectation, it follows that

$$
\begin{aligned}
E(E^{(p+2)}) &= (p+1)((p+2)E(D^s(X))) + (p+2)(1-(p+2)E(D^s(X))) \\
&= (p+2) - (p+2)E(D^s(X)).
\end{aligned}
$$

Since

$$
\begin{aligned}
p+1 &\leq E^{(p+2)} \leq p+2 \\
\Rightarrow p+1 &\leq E(E^{(p+2)}) \leq p+2 \\
\Rightarrow p+1 &\leq (p+2)-(p+2)E(D^s(X)) \leq p+2 \\
\Rightarrow -1 &\leq -(p+2)E(D^s(X)) \leq 0 \\
\Rightarrow 0 &\leq E(D^s(X)) \leq \frac{1}{p+2},
\end{aligned}
$$

our upper bound for any dimension p is $E(D^s(X)) \leq \frac{1}{p+2}$. **QED**

Recall the result of Liu from Section 3.3: If F is absolutely continuous and angularly symmetric, then $D^s(x) \leq 2^{-p} \; \forall \; x \in \mathbf{R}^p$. Hence, under these conditions, $E(D^s(X)) \leq 2^{-p}$ by conditioning on X. The precision of Theorem 2.4.6 relative to Liu's result for angularly symmetric distributions decreases with the dimension p, but our main goal was to attain an upper bound for $E(D^s(X))$ for **any** continuous distribution F, which is not necessarily angularly symmetric.

Chapter 3

New Types of Data Depth Based on Other Simple Geometric Shapes

In Chapter 2, we thoroughly examined the definition and many properties of Liu's simplicial depth function $D^s(x)$. In two dimensions, $D^s(x)$ is based on a triangle, a simple geometric shape. In this chapter, we examine two new depth functions in particular, the circular depth $D^c(x)$ and the rectangular depth $D^r(x)$. In two dimensions, these are based on equally simple geometric objects, namely, circles and rectangles, respectively. We derive results for these two depth functions similar to some of those asserted for Liu's simplicial depth function $D^s(x)$ in Chapter 3. We will also discuss the advantages and disadvantages of using circular or rectangular depth functions as opposed to simplicial depth.

3.1 Circular Depth

For any two distinct points x_1 and x_2 in the plane, we can generate a unique closed disc $\odot(x_1, x_2)$ containing both points, with the centre of the circle located at the centre of the two points. That is, the points x_1 and x_2 will be at opposite sides of the disc, and the distance between them will be the circle's diameter. See Figure 3.1 for an illustration. If we have an i.i.d. random sample $X_1, X_2, ..., X_n$ from a distribution F on $\mathbf{R}^2$, we can generate $\binom{n}{2}$ such discs, and for any point $x \in \mathbf{R}^2$, we can find the proportion of discs which contain x. Naturally, the higher the proportion, the greater the depth of the point in the data cloud.

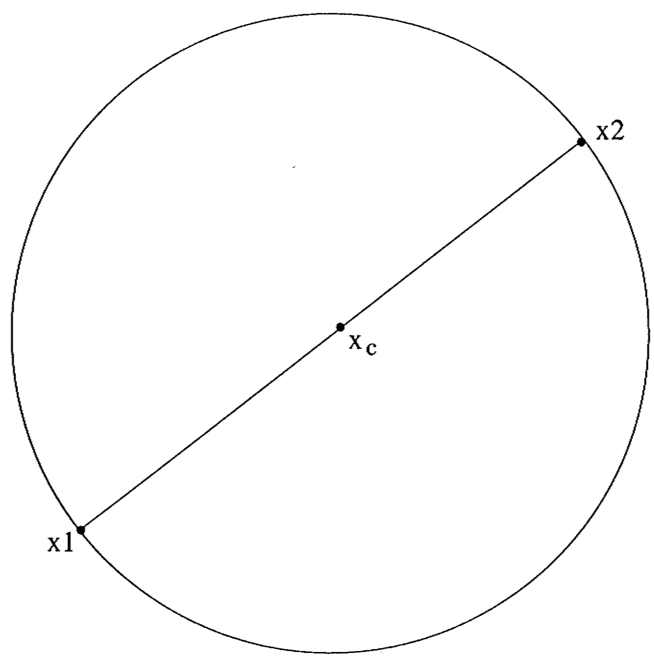

Figure 3.1: The circle $\odot(x_1, x_2)$ with centre $\frac{x_1+x_2}{2}$ and diameter $||x_1 - x_2||$.

Definition 3.1.1 Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample with c.d.f. F. Then the *sample circular depth* $D_n^c(x)$ for a point $x \in \mathbf{R}^2$ is equal to the proportion of all discs $\bigodot(X_{i_1}, X_{i_2})$, $1 \leq i_1 < i_2 \leq n$ which contain x, where $\bigodot(X_{i_1}, X_{i_2})$ is defined as above. That is,

$$D_n^c(x) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{I}(x \in \bigodot(X_{i_1}, X_{i_2})). \tag{3.1}$$

$D_n^c(x)$ is the empirical (sample) version of the true simplicial depth $D^c(x)$:

Definition 3.1.2 The *circular depth* $D^c(x)$ for a point $x \in \mathbf{R}^2$ is equal to the probability that x is contained in the random disc $\bigodot(X_1, X_2)$. That is,

$$D^c(x) = P(x \in \bigodot(X_1, X_2)), \tag{3.2}$$

where X_i, $i = 1, 2$ are i.i.d. with c.d.f. F.

Note that the univariate form of $D^c(x)$ reduces to the exact form of Equation (2.3) that we had for the simplicial depth $D^s(x)$.

All of these concepts can easily be extended to higher dimensions. For a distribution F on $\mathbf{R}^p$, the disc in Definitions 3.1.1 and 3.1.2 is replaced by the closed ball $C(X_1, X_2)$ formed by two independent observations X_1 and X_2 from F. This time, the p-dimensional ball will be that unique ball passing through both X_1 and X_2 and with diameter $||X_1 - X_2||$ and centre $\frac{X_1+X_2}{2}$. The general definition of the sample circular depth in any dimension p is as follows:

Definition 3.1.3 The *sample circular depth* $D_n^c(x)$ for a point $x \in \mathbf{R}^p$ is equal to the proportion of all closed balls $C(X_{i_1}, X_{i_2})$, $1 \leq i_1 < i_2 \leq n$ which contain x. That is,

$$D_n^c(x) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathbb{I}(x \in C(X_{i_1}, X_{i_2})), \tag{3.3}$$

where $X_1, X_2, ..., X_n$ is an i.i.d. random sample from F.

Similarly, the circular depth of a point $x \in \mathbf{R}^p$ is defined as:

Definition 3.1.4 The *circular depth* $D^c(x)$ for a point $x \in \mathbf{R}^p$ is equal to the probability that x is contained in the random closed ball $C(X_1, X_2)$ generated by the two i.i.d. observations X_1 and X_2:

$$D^c(x) = P(x \in C(X_1, X_2)). \tag{3.4}$$

We now look at some important properties of $D^c(x)$, as well as its empirical estimate $D_n^c(x)$. The first three resemble those for simplicial depth found in Section 2.2.

Theorem 3.1.5 *$D_n^c(x)$ is an unbiased estimator for $D^c(x)$, for any $x \in \mathbf{R}^p$.*

Proof. By the linearity of the expectation operator,

$$\begin{aligned}
E\left[D_n^c(x)\right] &= E\left[\binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathbb{I}(x \in C(X_{i_1}, X_{i_2}))\right] \\
&= \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} E\left[\mathbb{I}(x \in C(X_{i_1}, X_{i_2}))\right] \\
&= \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} P(x \in C(X_{i_1}, X_{i_2})) \\
&= \binom{n}{2}^{-1} \binom{n}{2} P(x \in C(X_1, X_2)) \\
&= P(x \in C(X_1, X_2)) \\
&= D^c(x).
\end{aligned}$$

This completes the proof. **QED**

Theorem 3.1.6 *For any c.d.f. F on $\mathbf{R}^p$, $\sup_{||x|| \geq M} D^c(x) \to 0$ as $M \to \infty$.*

Proof. Let X_1, X_2 be i.i.d. with c.d.f. F. Given any $x \in \mathbf{R}^p$, we observe that $\{x \in C(X_1, X_2)\} \subseteq \cup_{i=1}^{2} \{||X_i|| \geq ||x||\}$. This is because, if x is further from the origin than either of the two random points, it clearly cannot be contained in the ball generated by the two points. Using the above inclusion, we get

$$\begin{aligned} D^c(x) &= P(x \in C(X_1, X_2)) \\ &\leq P\left(\bigcup_{i=1}^{2} \{||X_i|| \geq ||x||\}\right) \\ &\leq \sum_{i=1}^{2} P(||X_i|| \geq ||x||) \text{ (by subadditivity)} \\ &= 2P(||X_1|| \geq ||x||). \end{aligned}$$

Note that the last step is possible by the identical distributions of X_1 and X_2. So clearly, since $P(||X_1|| \geq ||x||)$ is decreasing in $||x||$, it follows that

$$\sup_{||x|| \geq M} D^c(x) \leq \sup_{||x|| \geq M} 2P(||X_1|| \geq ||x||) \leq 2P(||X_1|| \geq M).$$

Since $P(||X_1|| \geq M) \to 0$ as $M \to \infty$, we get the desired result. This completes the proof. **QED**

Our next assertion is that of the continuity of $D^c(x)$.

Theorem 3.1.7 *If F is an absolutely continuous distribution on $\mathbf{R}^p$, then $D^c(x)$ is continuous on $\mathbf{R}^p$.*

Proof. We prove the theorem for $p = 2$. A more general proof for any dimension p would follow analogously.

Let X_1, X_2 be i.i.d. with c.d.f. F. To establish continuity at $x \in \mathbf{R}^2$, we let $\{x_n\}$ be a sequence in $\mathbf{R}^2$ such that $x_n \rightarrow x$, and we will show that

$$|D^c(x) - D^c(x_n)| \leq P(\partial(\bigodot(X_1, X_2)) \cap \overline{xx_n}) \rightarrow 0$$

as $n \rightarrow \infty$, where $\partial(\odot(X_1, X_2))$ denotes the boundary of $\odot(X_1, X_2)$.

Note again that, in the context of intersecting line (or arc) segments, "$\cap$" refers to the two crossing one another at some point. This differs from our usual use of the intersection symbol "$\cap$", meaning the intersection of two sets or events.

A random disc can contribute to the difference $D^c(x) - D^c(x_n)$ only if it contains one point but not the other. This however implies that there must be a point on the boundary of the disc $\odot(X_1, X_2)$ which intersects the line segment $\overline{xx_n}$. See Figure 3.2 for an illustration for the case when $p = 2$.

For any two events A and B, $P(A \backslash B) = P(A) - P(A \cap B) \geq P(A) - P(B)$. Therefore, if we define

$$\begin{aligned} A &= [x \in \bigodot(X_1, X_2)] \\ B &= [x_n \in \bigodot(X_1, X_2)], \end{aligned}$$

then we have that

$$\begin{aligned} D^c(x) - D^c(x_n) &= P(A) - P(B) \\ &\leq P((x \in \bigodot(X_1, X_2)) \cap (x_n \notin \bigodot(X_1, X_2))) \\ &\leq P(\partial(\bigodot(X_1, X_2)) \cap \overline{xx_n}). \end{aligned}$$

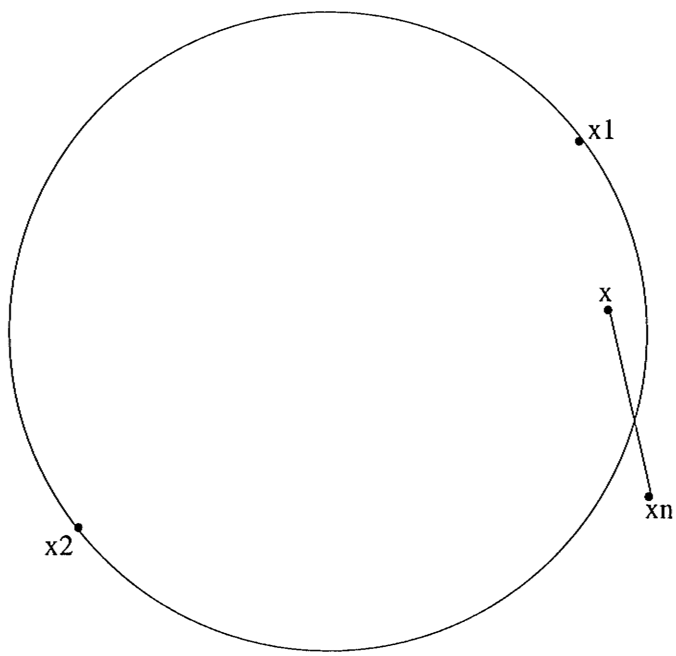

Figure 3.2: $x \in \bigodot(x_1, x_2)$ and $x_n \notin \bigodot(x_1, x_2)$, so $\overline{xx_n} \cap \partial(\bigodot(x_1, x_2))$. Therefore, this circle contributes to the difference $D^c(x) - D^c(x_n)$.

It can similarly be shown (or simply understood by symmetry) that

$$D^c(x_n) - D^c(x) \leq P(\partial(\bigodot(X_1, X_2)) \cap \overline{xx_n}).$$

As such, we have

$$|D^c(x) - D^c(x_n)| \leq P(\partial(\bigodot(X_1, X_2)) \cap \overline{xx_n}),$$

where X_1 and X_2 are i.i.d. with c.d.f. F.

We define the events

$$A_n = \{\partial(\bigodot(X_1, X_2)) \cap \overline{xx_n}\} \ \forall \ n.$$

Then

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} A_k\right] = \{x \in \partial(\bigodot(X_1, X_2))\}.$$

By Lemma A.2.2, we know that

$$\limsup_{n\to\infty} P(A_n) \leq P(\limsup_{n\to\infty} A_n) = P(x \in \partial(\bigodot(X_1, X_2))) = 0, \qquad (3.5)$$

since F is absolutely continuous. We can make the above assertion because of the following: We note that, by conditioning on X_1 (Lemma A.2.4),

$$\begin{aligned} & P[\partial(x \in \bigodot(X_1, X_2))] \\ = & \iint_{\mathbf{R}^2} P(x \in \partial(\bigodot(x_1, X_2)) f(x_{11}, x_{12}) dx_{11} dx_{12}. \end{aligned} \qquad (3.6)$$

We must now ask, where must X_2 lie (given fixed points x_0 and x_1) in order for $x = x_0$ to lie on the boundary of the disc $\bigodot(x_1, X_2)$?

We note that X_2 must be a point such that the distance from x_0 to the centre of the disc $\bigodot(x_1, x_2)$ is equal to the distance from x_1 to the centre.

This must be the case, as x_0 and x_1 are both on the boundary of the disc, and so must be equidistant to the center. In other words, we must have:

$$
\begin{aligned}
& \left(x_{01} - \frac{x_{11} + x_{21}}{2}\right)^2 + \left(x_{02} - \frac{x_{12} + x_{22}}{2}\right)^2 \\
= & \left(x_{11} - \frac{x_{11} + x_{21}}{2}\right)^2 + \left(x_{12} - \frac{x_{12} + x_{22}}{2}\right)^2
\end{aligned}
$$

which implies

$$
\begin{aligned}
& (2x_{01} - x_{11} - x_{21})^2 + (2x_{02} - x_{12} - x_{22})^2 \\
= & (2x_{11} - x_{11} - x_{21})^2 + (2x_{12} - x_{12} - x_{22})^2 \\
\Rightarrow & 4x_{01}^2 + x_{11}^2 + x_{21}^2 - 4x_{01}x_{11} - 4x_{01}x_{21} + 2x_{11}x_{21} \\
& +4x_{02}^2 + x_{12}^2 + x_{22}^2 - 4x_{02}x_{12} - 4x_{02}x_{22} + 2x_{12}x_{22} \\
= & x_{11}^2 + x_{21}^2 - 2x_{11}x_{21} + x_{12}^2 + x_{22}^2 - 2x_{12}x_{22} \\
\Rightarrow & x_{01}^2 - x_{01}x_{11} - x_{01}x_{21} + x_{02}^2 - x_{02}x_{12} - x_{02}x_{12} = 0 \\
\Rightarrow & (x_{11} - x_{01})x_{21} + (x_{12} - x_{02})x_{22} + x_{01}^2 + x_{02}^2 - x_{01}x_{11} - x_{02}x_{22} = 0 \\
\Rightarrow & x_{22} = \frac{x_{01}^2 + x_{02}^2 - x_{01}x_{11} - x_{02}x_{12} + (x_{11} - x_{01})x_{21}}{x_{02} - x_{12}} = L(x_{21}). \quad (3.7)
\end{aligned}
$$

(See the illustration in Figure 3.3.) In order for x_0 to be in the disc $\bigodot(x_1, X_2)$, we see that X_2 must lie along this straight line, a one-dimensional subspace of $\mathbf{R}^2$. But by the absolute continuity of F, the probability of this happening is zero.

From Equations (3.6) and (3.7), we thus have that

$$P[x \in \partial(\bigodot(X_1, X_2))]$$

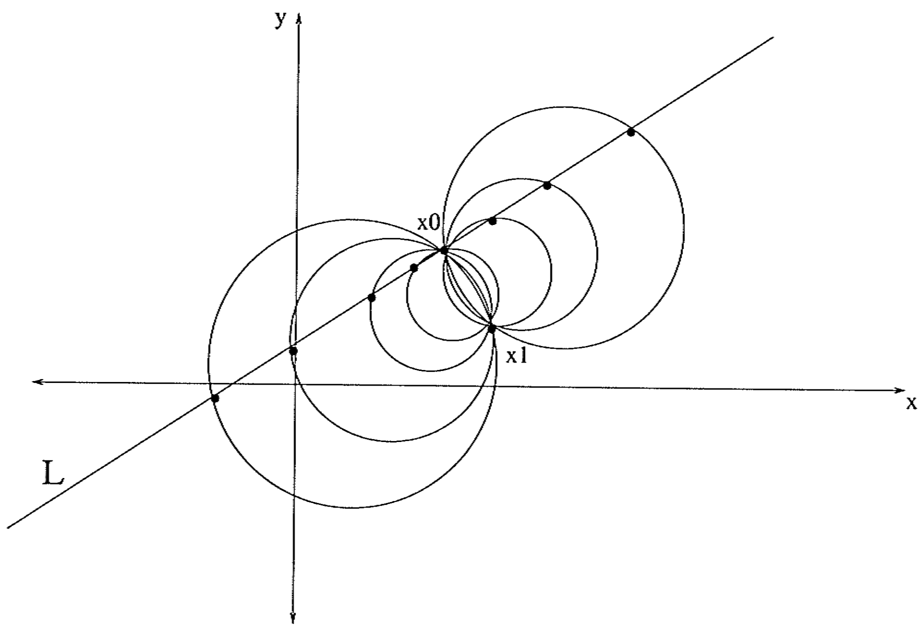

Figure 3.3: In order for $x_0 \in \partial(\bigodot(x_1, X_2))$, X_2 must lie on the line L.

$$
\begin{aligned}
&= \iint_{\mathbf{R}^2} P(X_2 \in L) f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&= \iint_{\mathbf{R}^2} 0 f(x_{11}, x_{12}) dx_{11} dx_{12} = 0. \qquad (3.8)
\end{aligned}
$$

Therefore, $|D^c(x) - D^c(x_n)| \to 0$ as $n \to \infty$. This completes the proof, using the sequential characterization of continuity. **QED**

We would now like to find a tractable form for $D^c(x)$ in two dimensions. Similar to our derivation of $D^s(x)$ in Chapter 2, we must identify where two points must lie in the plane in order for a fixed point x_0 to be in the disc $\odot(X_1, X_2)$. Again, we use the method of conditioning.

Theorem 3.1.8 *Given a fixed point* $x_0 = (x_{01}, x_{02})$ *in the plane,*

$$
\begin{aligned}
D^c(x_0) &= P\{x_0 \in \bigodot(X_1, X_2)\} \\
&= \iint_{\mathbf{R}^2} \mathbb{I}(B) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{L(x_{21})} f(x_{21}, x_{22}) dx_{22} dx_{21} \right] f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&\quad + \iint_{\mathbf{R}^2} \mathbb{I}(B^c) \left[\int_{-\infty}^{\infty} \int_{L(x_{21})}^{\infty} f(x_{21}, x_{22}) dx_{22} dx_{21} \right] f(x_{11}, x_{12}) dx_{11} dx_{12},
\end{aligned}
$$

where

$$
\begin{aligned}
B &= \{(x_{11}, x_{12}) \in \mathbf{R}^2 | x_{12} > x_{02}\} \\
B^c &= \{(x_{11}, x_{12}) \in \mathbf{R}^2 | x_{12} < x_{02}\} \qquad (3.9)
\end{aligned}
$$

$$
L(x_{21}) = \frac{x_{01}(x_{11} - x_{01}) + x_{02}(x_{12} - x_{02}) + (x_{01} - x_{11}) x_{21}}{x_{12} - x_{02}}.
$$

Proof. Toward a conditioning argument, we fix a value $x_1 = (x_{11}, x_{12}) \in \mathbf{R}^2$ for X_1 and we would like to determine where X_2 must lie in order for a fixed point x_0 to lie in the disc $\odot(x_1, X_2)$, where the disc is defined as above.

We note that we will have $x_0 \in \odot(x_1, x_2)$ if the distance from x_0 to the centre of the disc $\odot(x_1, x_2)$ is less than the radius of the circle. See Figure 3.4 for an illustration.

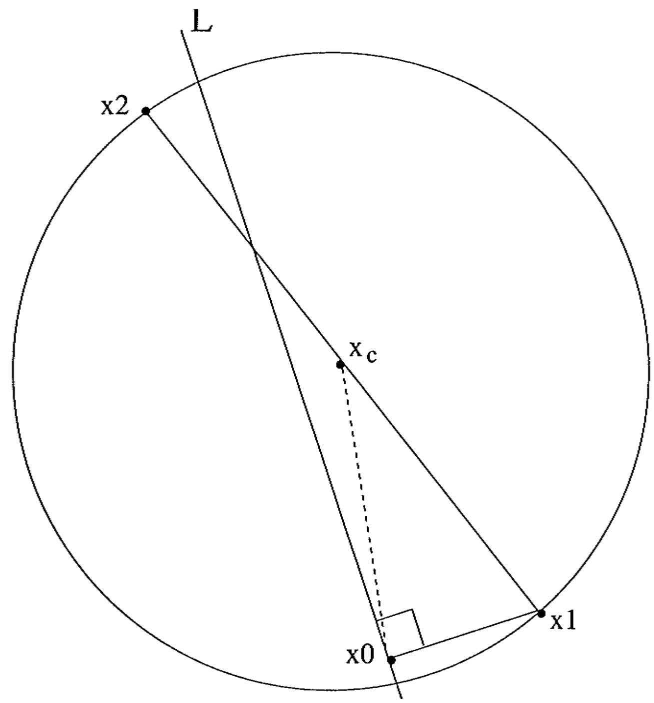

Figure 3.4: X_2 must lie on the opposite side of L from x_1 in order for $x_0 \in \odot(x_1, X_2)$.

That is, we must have that

$$\sqrt{\left(x_{01} - \frac{x_{11} + x_{21}}{2}\right)^2 + \left(x_{02} - \frac{x_{12} + x_{22}}{2}\right)^2} \leq \frac{1}{2}\sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2}$$

$$\Rightarrow (2x_{01} - x_{11} - x_{21})^2 + (2x_{02} - x_{12} - x_{22})^2 \leq (x_{11} - x_{21})^2 + (x_{12} - x_{22})^2$$

$$\Rightarrow \quad 4x_{01}^2 + x_{11}^2 + x_{21}^2 - 4x_{01}x_{11} - 4x_{01}x_{21} + 2x_{11}x_{21}$$

$$
\begin{aligned}
& +4x_{02}^2 + x_{12}^2 + x_{22}^2 - 4x_{02}x_{12} - 4x_{02}x_{22} + 2x_{12}x_{22} \\
\leq \; & x_{11}^2 + x_{21}^2 + x_{12}^2 + x_{22}^2 - 2x_{11}x_{21} - 2x_{12}x_{22} \\
\Rightarrow \; & x_{01}^2 - x_{01}x_{11} - x_{01}x_{21} + x_{11}x_{21} + x_{02}^2 - x_{02}x_{12} - x_{02}x_{22} + x_{12}x_{22} \leq 0 \\
\Rightarrow \; & (x_{11} - x_{01})x_{21} + x_{01}(x_{01} - x_{11}) + (x_{12} - x_{02})x_{22} + x_{02}(x_{02} - x_{12}) \leq 0 \\
\Rightarrow \; & (x_{12} - x_{02})x_{22} \leq x_{01}(x_{11} - x_{01}) + x_{02}(x_{12} - x_{02}) + (x_{01} - x_{11})x_{21}
\end{aligned}
$$

$$
\Rightarrow \begin{cases} x_{22} \leq L(x_{21}) & \text{if } x_{12} > x_{02}, \\ x_{22} > L(x_{21}) & \text{if } x_{12} < x_{02}, \end{cases}
$$

where

$$
L(x_{21}) = \frac{x_{01}(x_{11} - x_{01}) + x_{02}(x_{12} - x_{02}) + (x_{01} - x_{11})x_{21}}{x_{12} - x_{02}}. \tag{3.10}
$$

Assuming X_1 and X_2 are i.i.d. random variables from a distribution F, we condition with respect to the event $A = [X_{12} > x_{02}]$. Since A and A^c are exhaustive and mutually exclusive, then by Lemma A.2.3 we have

$$
D^c(x_0) \;=\; P([x_0 \in \odot(X_1, X_2)] \cap A) + P([x_0 \in \odot(X_1, X_2)] \cap A^c).
$$

Now, by Lemma A.2.5, for B defined in the statement of the theorem, we have

$$
\begin{aligned}
& P([x_0 \in \odot(X_1, X_2)] \cap A) \\
= \; & E[\mathbb{I}(x_0 \in \odot(X_1, X_2))\mathbb{I}(A)] \\
= \; & E[E(\mathbb{I}(x_0 \in \odot(X_1, X_2))\mathbb{I}(A)|X_1)] \\
= \; & \iint_{\mathbf{R}^2} E[\mathbb{I}(x_0 \in \odot(x_1, X_2))\mathbb{I}(B)] f(x_{11}, x_{12}) dx_{11} dx_{12}
\end{aligned}
$$

$$= \iint_{\mathbf{R}^2} \mathbb{I}(B) P(x_0 \in \bigodot(x_1, X_2)) f(x_{11}, x_{12}) dx_{11} dx_{12}$$
$$= \iint_{\mathbf{R}^2} \mathbb{I}(B) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{L(x_{21})} f(x_{21}, x_{22}) dx_{22} dx_{21} \right] f(x_{11}, x_{12}) dx_{11} dx_{12}.$$

Following the same steps, we can show that

$$\begin{aligned} & P([x_0 \in \bigodot(X_1, X_2)] \cap A^c) \\ = & \iint_{\mathbf{R}^2} \mathbb{I}(B^c) \left[\int_{-\infty}^{\infty} \int_{L(x_{21})}^{\infty} f(x_{21}, x_{22}) dx_{22} dx_{21} \right] f(x_{11}, x_{12}) dx_{11} dx_{12}, \end{aligned}$$

and the proof is complete. **QED**

We will now seek to find a functional (i.e. tractable) form of the variance of $D_n^c(x)$.

Theorem 3.1.9 *The variance of $D_n^c(x)$ is*

$$Var(D_n^c(x)) = \binom{n}{2}^{-1} \left[D^c(x_0) + 2(n-2) D_{(2)}^c(x_0) + (3-2n)(D^c(x_0))^2 \right],$$

where

$$\begin{aligned} D_{(2)}^c(x_0) = & \iint_{\mathbf{R}^2} \mathbb{I}(B) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{L(x_{21})} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 dx_{11} dx_{22} + \\ & \iint_{\mathbf{R}^2} \mathbb{I}(B^c) \left[\int_{-\infty}^{\infty} \int_{L(x_{21})}^{\infty} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 dx_{11} dx_{22}, \end{aligned} \tag{3.11}$$

and where B and B^c are as defined in Equation (3.9) and $L(x_{21})$ is as defined in Equation (3.10).

Proof. We know that

$$Var[D_n^c(x)] = E[D_n^c(x)]^2 - (E[D_n^c(x)])^2$$

$$= E\left[\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \mathbb{I}(x_0 \in \bigodot(X_i, X_j))\right]^2 + \left(E\left[\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \mathbb{I}(x_0 \in \bigodot(X_i, X_j))\right]\right)^2 . \quad (3.12)$$

Now we consider only the first term of the sum in Equation (3.12):

$$\begin{aligned}
& E\left[\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \mathbb{I}(x_0 \in \bigodot(X_i, X_j))\right]^2 \\
&= \binom{n}{2}^{-2} E\left[\sum_{1\le i<j\le n} \mathbb{I}(x_0 \in \bigodot(X_i, X_j))\right]^2 \\
&= \binom{n}{2}^{-2} \left\{\left(E\left[\sum_{*}[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))]\right]\right)\right. \\
&\quad + \left(E\left[\sum_{**}[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))]\right]\right) \\
&\quad \left. + \left(E\left[\sum_{***}[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))]\right]\right)\right\}, \quad (3.13)
\end{aligned}$$

where

$$\begin{aligned}
* &= \{1 \le i < j \le n, 1 \le k < l \le n, i = k, j = l\} \\
** &= \{1 \le i < j \le n, 1 \le k < l \le n, \\
&\qquad (\text{neither } (i = k, j = l) \text{ nor } (i \ne j \ne k \ne l))\} \\
*** &= \{1 \le i < j \le n, 1 \le k < l \le n, i \ne j \ne k \ne l\}.
\end{aligned}$$

In other words, * represents all pairs (i, j) and (k, l) such that $i = k$ and j =l. That is, the sum is over all identical pairs of variables. This is equivalent to squaring each term within the sum. In the second term, ** represents all pairs (i, j) and (k, l) with exactly one equal index. This will be the most complicated sum to determine. Finally, *** represents all pairs (i, j) and (k, l) with no indices in common, so the two terms are independent and the expectation of their product will the the product of their expectations.

Now, continuing with Equation (3.13), we have

$$\begin{aligned} = & \binom{n}{2}^{-2} \left\{ \left(\left[\sum_{*} E[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))] \right] \right) \right. \\ & + \left(\left[\sum_{**} E[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))] \right] \right) \\ & \left. + \left(\left[\sum_{***} E[\mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \times \mathbb{I}(x_0 \in \bigodot(X_k, X_l))] \right] \right) \right\}. \end{aligned} \tag{3.14}$$

Now we must determine how many terms are being added for each of the three sums. For *, it is quite obvious that there are precisely $\binom{n}{2}$ terms for which $i = k$ and $j = l$. Moreover, squaring the indicator function just gives us the indicator again, so we can remove the square. For ***, for every (i, j), for which there are $\binom{n}{2}$ combinations, in order for $i \neq j \neq k \neq l$ (i.e. all distinct), we must have k and l be two of the other $n - 2$ possible indices. As such, there are $\binom{n}{2}\binom{n-2}{2}$ ways this can happen. That leaves us with **, the most complex case. In all, there must be $\binom{n}{2}^2$ combinations of pairs of indices, and so there are $\binom{n}{2}^2 - \binom{n}{2}\left(1 + \binom{n-2}{2}\right) = \binom{n}{2}(2(n-2))$ terms in the second sum. Because each expectation within each sum is equal for any values of the indices, continuing from Equation (3.14), we have

$$\begin{aligned} = & \binom{n}{2}^{-2} \left\{ \binom{n}{2} E[\mathbb{I}(x_0 \in \bigodot(X_1, X_2))] \right. \\ & + \binom{n}{2}(2(n-2)) E[\mathbb{I}(x_0 \in \bigodot(X_1, X_2)) \times \mathbb{I}(x_0 \in \bigodot(X_1, X_3))] \end{aligned}$$

$$
\begin{aligned}
&+\binom{n}{2}\binom{n-2}{2}E[\mathbb{I}(x_0 \in \bigodot(X_1,X_2)) \times \mathbb{I}(x_0 \in \bigodot(X_3,X_4))]\Big\} \\
= &\binom{n}{2}^{-1}\Big\{P(x_0 \in \bigodot(X_1,X_2)) \\
&+2(n-2)E[\mathbb{I}(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3))] \\
&+\binom{n-2}{2}[P(x_0 \in \bigodot(X_1,X_2))]^2\Big\} \\
= &\binom{n}{2}^{-1}\Big\{D^c(x_0) + 2(n-2)E[\mathbb{I}(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3))] \\
&+\binom{n-2}{2}(D^c(x_0))^2\Big\}. \qquad (3.15)
\end{aligned}
$$

Now we are left with the middle term. We will use conditioning. Note that

$$x_0 \in [\bigodot(X_1,X_2) \cap \bigodot(X_1,X_3)]$$

implies X_2 and X_3 must be on one side of the line L, which is perpendicular to $\overline{x_0X_1}$ and passes through x_0. See Figure 3.5 for an illustration. L is defined in Equation (3.10) and is random, depending on the value of X_1.

Now we have

$$
\begin{aligned}
&E[\mathbb{I}((x_0 \in \bigodot(X_1,X_2)) \cap (x_0 \in \bigodot(X_1,X_3)))] \\
= \; &P(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3)). \qquad (3.16)
\end{aligned}
$$

Continuing from Equation (3.16) and examining Figure 3.5, if we define $A = [X_{12} > x_{02}]$, applying conditioning using Lemmas A.2.3, A.2.4 and A.2.5,

$$
\begin{aligned}
&P(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3)) \\
= \; &P(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3) \cap A) \\
&+P(x_0 \in \bigodot(X_1,X_2) \cap x_0 \in \bigodot(X_1,X_3) \cap A^c)
\end{aligned}
$$

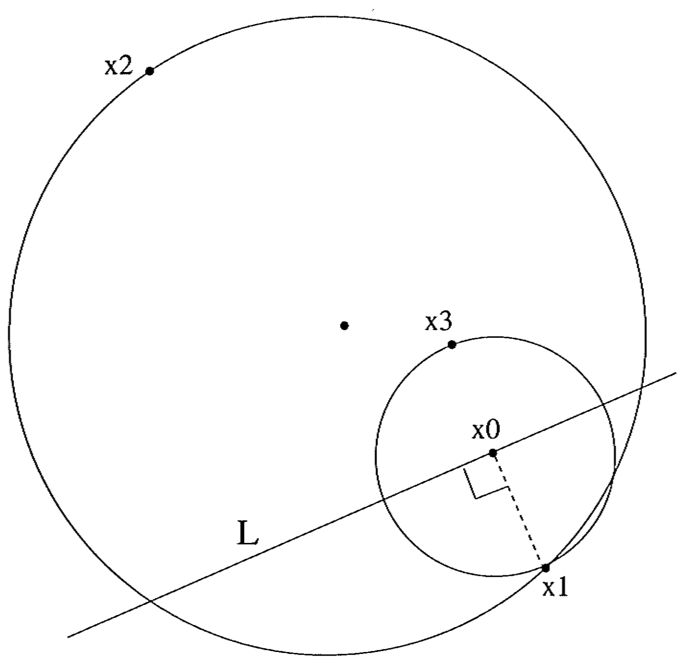

Figure 3.5: X_2 and X_3 must **both** lie on the opposite side of L from x_1 in order for $x_0 \in \odot(x_1, X_2)$ **and** $x_0 \in \odot(x_1, X_3)$.

$$
\begin{aligned}
&= \iint_{\mathbf{R}^2} \mathbb{I}(B) P(x_0 \in \bigodot(x_1, X_2) \cap x_0 \in \bigodot(x_1, X_3)) f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&\quad + \iint_{\mathbf{R}^2} \mathbb{I}(B^c) P(x_0 \in \bigodot(x_1, X_2) \cap x_0 \in \bigodot(x_1, X_3)) f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&= \iint_{\mathbf{R}^2} \mathbb{I}(B) (P(x_0 \in \bigodot(x_1, X_2)))^2 f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&\quad + \iint_{\mathbf{R}^2} \mathbb{I}(B^c) (P(x_0 \in \bigodot(x_1, X_2)))^2 f(x_{11}, x_{12}) dx_{11} dx_{12} \\
&= \iint_{\mathbf{R}^2} \mathbb{I}(B) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{L(x_{21})} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 f(x_{11}, x_{12}) dx_{11} dx_{22} \\
&\quad + \iint_{\mathbf{R}^2} \mathbb{I}(B^c) \left[\int_{-\infty}^{\infty} \int_{L(x_{21})}^{\infty} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 f(x_{11}, x_{12}) dx_{11} dx_{22}
\end{aligned}
\tag{3.17}
$$

We also used the independence and identical distributions of X_2 and X_3 in the above calculations. We denote the above probability as $D^c_{(2)}(x_0)$ to facilitate our notation.

In total, from Equation (3.15), we have

$$
\begin{aligned}
& E\left[\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbb{I}(x_0 \in \bigodot(X_i, X_j)) \right]^2 \\
&= \binom{n}{2}^{-1} \left[D^c(x_0) + 2(n-2) D^c_{(2)}(x_0) + \binom{n-2}{2} (D^c(x_0))^2 \right], \quad (3.18)
\end{aligned}
$$

and so

$$
\begin{aligned}
& Var(D^c_n(x_0)) \quad (3.19) \\
&= E(D^c_n(x_0))^2 - [E(D^c_n(x_0))]^2 \\
&= \binom{n}{2}^{-1} \left[D^c(x_0) + 2(n-2) D^c_{(2)}(x_0) + \left[\binom{n-2}{2} - \binom{n}{2} \right] (D^c(x_0))^2 \right] \\
&= \binom{n}{2}^{-1} \left[D^c(x_0) + 2(n-2) D^c_{(2)}(x_0) + (3-2n)(D^c(x_0))^2 \right]. \quad (3.20)
\end{aligned}
$$

This completes the proof. **QED**

The variance for $D_n^c(x)$ in any dimension p can simply be obtained by altering the variance in two dimensions by replacing the line L with the $(p-1)$-dimensional hyperplane H which is perpendicular to $\overline{x_0 X_1}$, and by using p-tuple integration rather than double integration. Otherwise, it has the same functional form as the case for $p = 2$.

We now find an upper bound for our circular depth function $D^c(x)$.

Theorem 3.1.10 *The circular depth $D^c(x)$ has value less than or equal to $\frac{1}{2}$ for every $x \in \mathbf{R}^p$. That is,*

$$D^c(x) \leq \frac{1}{2} \; \forall \; x \in \mathbf{R}^p.$$

Proof. We state the proof for $p = 2$. Recall that

$$D^c(x_0) \quad = \quad P(x_0 \in \bigodot(X_1, X_2)). \tag{3.21}$$

We can also write the circular depth function in two dimensions as

$$\begin{aligned} D^c(x_0) &= P(x_0 \in \bigodot(X_1, X_2)|X_{12} > x_{02})P(X_{12} > x_{02}) \\ &\quad + P(x_0 \in \bigodot(X_1, X_2)|X_{12} < x_{02})P(X_{12} < x_{02}) \\ &= P(X_{22} < L(X_{21})|X_{12} > x_{02})P(X_{12} > x_{02}) \\ &\quad + P(X_{22} > L(X_{21})|X_{12} < x_{02})P(X_{12} < x_{02}), \end{aligned} \tag{3.22}$$

where L is defined in Equation (3.10). Now, let

$$\begin{aligned} a = P(X_{22} < L(X_{21})|X_{12} > x_{02}) \text{ and} \\ b = P(X_{12} < x_{02}). \end{aligned} \tag{3.23}$$

Then from Equation (3.22), we have

$$D^c(x_0) \quad = \quad a(1-b) + b(1-a) = a + b - 2ab.$$

We maximize the depth by taking partial derivatives with respect to both a and b and setting them to zero:

$$\begin{aligned} \frac{\partial D^c}{\partial a} &= 1 - 2b = 0 \\ \frac{\partial D^c}{\partial b} &= 1 - 2a = 0 \\ &\Rightarrow \quad a = b = \frac{1}{2} \end{aligned} \tag{3.24}$$

and so by Equation (3.24), we have that

$$D^c(x_0) \leq \frac{1}{2} + \frac{1}{2} - 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{2}.$$

This completes the proof. To extend the proof to p dimensions, we simply replace L by H, the $(p-1)$-dimensional hyperplane which is perpendicular to $\overline{x_0 X_1}$. **QED**

Theorem 3.1.11 *If X_1, X_2, X_3 are i.i.d. random variables with c.d.f. F, then the probability that x is contained in both $\odot(X_1, X_2)$ and $\odot(X_1, X_3)$ is less than or equal to $\frac{1}{4}$ for every $x \in \mathbf{R}^p$. That is,*

$$D^c_{(2)}(x) \leq \frac{1}{4} \ \forall \ x \in \mathbf{R}^p.$$

Proof. Again we give the proof for $p = 2$ dimensions. Recall from Equation (3.11) that

$$\begin{aligned} D^c_{(2)}(x_0) \quad = \quad & \iint_{\mathbf{R}^2} \mathbb{I}(B) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{L(x_{21})} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 dx_{11} dx_{22} \\ & + \iint_{\mathbf{R}^2} \mathbb{I}(B^c) \left[\int_{-\infty}^{\infty} \int_{L(x_{21})}^{\infty} f(x_{21}, x_{22}) dx_{22} dx_{21} \right]^2 dx_{11} dx_{22}. \end{aligned}$$

However, similar to what we did in Equation (3.22), we can express $D^c_{(2)}(x_0)$ as

$$\begin{aligned} D^c_{(2)}(x_0) &= P(x_0 \in \bigodot(X_1, X_2) \cap x_0 \in \bigodot(X_1, X_3) | X_{12} > x_{02}) P(X_{12} > x_{02}) \\ &\quad + P(x_0 \in \bigodot(X_1, X_2) \cap x_0 \in \bigodot(X_1, X_3) | X_{12} < x_{02}) P(X_{12} < x_{02}) \\ &= P(X_{22} < L(X_{21}), X_{32} < L(X_{31}) | X_{12} > x_{02}) P(X_{12} > x_{02}) \\ &\quad + P(X_{22} > L(X_{21}), X_{32} > L(X_{31}) | X_{12} < x_{02}) P(X_{12} < x_{02}) \\ &= P(X_{22} < L(X_{21}) | X_{12} > x_{02}) P(X_{32} < L(X_{31}) | X_{12} > x_{02}) \\ &\quad \times P(X_{12} > x_{02}) \\ &\quad + P(X_{22} < L(X_{21}) | X_{12} > x_{02}) P(X_{32} < L(X_{31}) | X_{12} > x_{02}) \\ &\quad \times P(X_{12} < x_{02}) \\ &= (P(X_{22} < L(X_{21}) | X_{12} > x_{02}))^2 P(X_{12} > x_{02}) \\ &\quad + (P(X_{22} > L(X_{21}) | X_{12} < x_{02}))^2 P(X_{12} < x_{02}) \end{aligned} \tag{3.25}$$

The above calculation follows the same steps as in Equation (3.22), again using the identical distributions of X_2 and X_3 and now, the fact that $[X_{22} < L(X_{21})]$ and $[X_{32} < L(X_{31})]$ are conditionally independent given $[X_{12} > x_{02}]$.

We let

$$\begin{aligned} a &= P(X_{22} < L(X_{21}) | X_{12} > x_{02}) \\ b &= P(X_{12} < x_{02}), \end{aligned}$$

then Equation (3.25) becomes

$$D^c_{(2)}(x_0) = a^2(1-b) + (1-a)^2 b.$$

We maximize this function by taking partial derivatives:

$$\begin{aligned}
\frac{\partial D^c_{(2)}}{\partial b} &= -a^2 + (1-a)^2 = 1 - 2a = 0 \;\Rightarrow\; a = \frac{1}{2} \\
\frac{\partial D^c_{(2)}}{\partial a} &= 2a(1-b) - 2(1-a)b = 2a - 2b = 0 \;\Rightarrow\; b = a = \frac{1}{2} \\
&\Rightarrow\; b = \frac{1}{2} \qquad (3.26)
\end{aligned}$$

and so by Equation (3.25), we have that

$$D^c_{(2)}(x_0) \le \left(\frac{1}{2}\right)^2 + \frac{1}{2} - 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}.$$

This completes the proof. To extend the proof to p dimensions, we simply replace L by H, the $(p-1)$-dimensional hyperplane which is perpendicular to $\overline{x_0 X_1}$. **QED**

Lemma 3.1.12 *For any $p \ge 2$, the variance of the sample circular depth can be bounded above by $\frac{3}{(4n-6)n(n-1)} + \frac{n-2}{n(n-1)}$ for all $n \ge 2$. That is,*

$$Var(D^c_n(x_0)) \le \frac{3}{(4n-6)n(n-1)} + \frac{n-2}{n(n-1)} \;\forall\, n \ge 2.$$

Proof. It is sufficient to prove prove the result for $p = 2$. Recall from Equation (3.19) that

$$\begin{aligned}
Var(D^c_n(x_0)) &= E(D^c_n(x_0))^2 - [E(D^c_n(x_0))]^2 \\
&= \binom{n}{2}^{-1} \left[D^c(x_0) + 2(n-2)D^c_{(2)}(x_0) + (3-2n)(D^c(x_0))^2\right] \\
&= \frac{2}{n(n-1)} \left[D^c(x_0) + 2(n-2)D^c_{(2)}(x_0) + (3-2n)(D^c(x_0))^2\right] \\
&= \frac{2}{n(n-1)}D^c(x_0) + \frac{4n-8}{n(n-1)}D^c_{(2)}(x_0) + \frac{6-4n}{n(n-1)}D^c(x_0)^2. \qquad (3.27)
\end{aligned}$$

The second term in the sum is strictly positive for $n \geq 2$ (which we assume), and so

$$\frac{4n-8}{n(n-1)} D^c_{(2)}(x_0) \leq \frac{4n-8}{n(n-1)} \times \frac{1}{4} = \frac{n-2}{n(n-1)} \tag{3.28}$$

by Theorem 3.1.11. We now maximize the sum of the first and third terms in Equation (3.27). Maximizing the sum $\frac{2}{n(n-1)} D^c(x_0) + \frac{6-4n}{n(n-1)} D^c(x_0)^2$ is equivalent to maximizing the sum of the numerators, as $n(n-1) \geq 0 \ \forall \ n$. Letting $D^c(x_0) = z$, define

$$g(z) = 2z + (6-4n)z^2.$$

To maximize $g(z)$ with respect to z (for $n \geq 2$):

$$\begin{aligned} \frac{dg}{dz} &= 2 + (12-8n)z = 0 \\ &\Rightarrow \ z = \frac{1}{4n-6}. \end{aligned} \tag{3.29}$$

By Equations (3.28) and (3.29), substituting $\frac{1}{4n-6}$ for $D^c(x_0)$, we have that

$$\begin{aligned} Var(D^c_n(x_0)) &= \frac{2}{n(n-1)} D^c(x_0) + \frac{4n-8}{n(n-1)} D^c_{(2)}(x_0) + \frac{6-4n}{n(n-1)} D^c(x_0)^2 \\ &\leq \left(\frac{2}{n(n-1)}\right)\left(\frac{1}{4n-6} + \frac{n-2}{n(n-1)} + \frac{6-4n}{(4n-6)^2 n(n-1)}\right) \\ &= \frac{3}{(4n-6)n(n-1)} + \frac{n-2}{n(n-1)}. \end{aligned}$$

This completes the proof. **QED**

Note that Lemma 3.1.12 does not necessarily give the optimal upper bound on $Var(D^c_n(x_0))$, as both $D^c(x_0)$ and $D^c_{(2)}(x_0)$ are functions of a and

b, and $\frac{2}{n(n-1)}D^c(x_0) + \frac{6-4n}{n(n-1)}D^c(x_0)^2$ and $\frac{4n-8}{n(n-1)}D^c_{(2)}(x_0)$ will not necessarily achieve their respective maxima for the same pair (a, b). Nonetheless, it provides a very useful (i.e. explicit) upper bound for the variance.

Corollary 3.1.13 *$D_n^c(x)$ is a weakly consistent estimator of $D^c(x)$, i.e. $D_n^c(x)$ converges to $D^c(x)$ in probability for any $x \in \mathbf{R}^p$.*

The proof follows from Lemma 3.1.12 and Markov's inequality applied to $Y = |D_n^c(x) - D^c(x)|$ (see Lemma A.2.6).

3.2 Rectangular Depth

We now focus on another alternative to simplicial depth, based on an equally basic geometric shape. For any two points x_1 and x_2 in the plane, we can generate a unique rectangle $R(x_1, x_2)$ with sides parallel to the x- and y-axes and diagonal corners x_1 and x_2. See Figure 3.6 for an illustration. If we have an i.i.d. random sample $X_1, X_2, ..., X_n$ from a distribution F, we can generate $\binom{n}{2}$ such rectangles, and for any point $x \in \mathbf{R}^2$, we can find the proportion of such rectangles which contain x. Again, the higher this proportion, the greater the depth of x in the data cloud.

Definition 3.2.1 The *sample rectangular depth* $D_n^r(x)$ for a point $x \in \mathbf{R}^2$ is equal to the proportion of all rectangles $R(X_{i_1}, X_{i_2})$, $1 \leq i_1 < i_2 \leq n$ which contain x, where $R(X_{i_1}, X_{i_2})$ is defined as above. That is,

$$D_n^r(x) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{I}(x \in R(X_{i_1}, X_{i_2})). \tag{3.30}$$

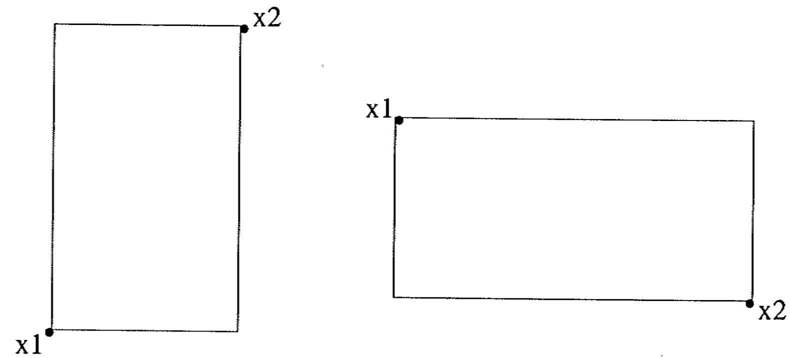

Figure 3.6: Two possible rectangles generated by points x_1 and x_2.

$D_n^r(x)$ is the empirical (sample) version of the true simplicial depth $D^r(x)$:

Definition 3.2.2 Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample from a c.d.f. F. The *rectangular depth* $D^r(x)$ for a point $x \in \mathbf{R}^2$ is equal to the probability that x is contained in the random rectangle $R(X_1, X_2)$. That is,

$$D^r(x) = P(x \in R(X_1, X_2)), \tag{3.31}$$

where X_i, $i = 1, 2$ are i.i.d. with c.d.f. F.

Note that the univariate form of $D^r(x)$ reduces to the exact form of Equation (2.3) that we had for both the simplicial depth $D^s(x)$ and circular depth $D^c(x)$.

For a distribution F on $\mathbf{R}^p$, the random rectangle in Definitions 3.2.1 and 3.2.2 is replaced by the random closed box $B(X_1, X_2)$ formed by two independent observations X_1 and X_2 from F. The p-dimensional box will be that unique box with diagonal corners X_1 and X_2. The general definition of the sample rectangular depth in any dimension p is as follows:

Definition 3.2.3 The *sample rectangular depth* $D_n^r(x)$ for a point $x \in \mathbf{R}^p$ is equal to the proportion of all boxes $B(X_{i_1}, X_{i_2})$, $1 \le i_1 < i_2 \le n$ which contain x. That is,

$$D_n^r(x) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathbb{I}(x \in B(X_{i_1}, X_{i_2})), \tag{3.32}$$

if $X_1, X_2, ..., X_n$ is an i.i.d. random sample from F.

Similarly, the rectangular depth of a point $x \in \mathbf{R}^p$ is defined as:

Definition 3.2.4 The *rectangular depth* $D^r(x)$ for a point $x \in \mathbf{R}^p$ is equal to the probability that x is contained in the random box $B(X_1, X_2)$ generated by the two i.i.d. observations X_1 and X_2:

$$D^r(x) = P(x \in B(X_1, X_2)) \tag{3.33}$$

As done in Section 3.1 for circular depth, we establish several key properties for rectangular depth resembling those for simplicial depth. We begin by asserting the unbiasedness of $D_n^r(x)$, $x \in \mathbf{R}^p$:

Theorem 3.2.5 *$D_n^r(x)$ is an unbiased estimator for $D^r(x)$.*

Proof. By the linearity of the expectation operator,

$$\begin{aligned} E\left[D_n^r(x)\right] &= E\left[\binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathbb{I}(x \in B(X_{i_1}, X_{i_2}))\right] \\ &= \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} E\left[\mathbb{I}(x \in B(X_{i_1}, X_{i_2}))\right] \\ &= \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} P(x \in B(X_{i_1}, X_{i_2})) \end{aligned}$$

$$
\begin{aligned}
&= \binom{n}{2}^{-1}\binom{n}{2}P(x \in B(X_1, X_2)) \\
&= P(x \in B(X_1, X_2)) \\
&= D^r(x).
\end{aligned}
$$

This completes the proof. **QED**

Theorem 3.2.6 *For any c.d.f. F on $\mathbf{R}^p$, $\sup_{||x||\geq M} D^c(x) \to 0$ as $M \to \infty$.*

Proof. Let X_1, X_2 be i.i.d. with c.d.f. F. Given any $x \in \mathbf{R}^p$, we observe that $\{x \in B(X_1, X_2)\} \subseteq \bigcup_{i=1}^{2}\{||X_i|| \geq ||x||\}$. This is because, if x is further from the origin than either of the two random points, it clearly cannot be contained in the box generated by the two points. Using the above inclusion, we get

$$
\begin{aligned}
D^r(x) &= P(x \in B(X_1, X_2)) \\
&\leq P\left(\bigcup_{i=1}^{2}\{||X_i|| \geq ||x||\}\right) \\
&\leq \sum_{i=1}^{2} P(||X_i|| \geq ||x||) \text{ (by subadditivity)} \\
&= 2P(||X_1|| \geq ||x||).
\end{aligned}
$$

Note that the last step is possible by the identical distributions of X_1 and X_2. So clearly, since $P(||X_1|| \geq ||x||)$ is decreasing in $||x||$, it follows that

$$
\sup_{||x||\geq M} D^r(x) \leq \sup_{||x||\geq M} 2P(||X_1|| \geq ||x||) \leq 2P(||X_1|| \geq M).
$$

Since $P(||X_1|| \geq M) \to 0$ as $M \to \infty$, we get the desired result. This completes the proof. **QED**

We now assert the continuity of $D^r(x)$:

Theorem 3.2.7 *If F is an absolutely continuous distribution on $\mathbf{R}^p$, then $D^r(x)$ is continuous on $\mathbf{R}^p$.*

Proof. We prove the theorem for $p = 2$. A similar proof for any dimension p would follow. To establish continuity at $x \in \mathbf{R}^2$, we take a sequence $\{x_n\}$ in $\mathbf{R}^2$ such that $x_n \to x$, and we will show that

$$|D^r(x) - D^r(x_n)| \to 0$$

as $n \to \infty$.

Let $X_1, X_2, ..., X_n$ be an i.i.d. sample from a c.d.f. F. The random rectangle $R(X_1, X_2)$ can contribute to the difference $D^r(x) - D^r(x_n)$ only if it contains one point but not the other. This however implies that there must be exactly one line segment of the four in our rectangle which intersects the line segment $\overline{xx_n}$. See Figure 3.7 for an illustration. For notational purposes, we define the lines

$$\begin{aligned} L_V(X_i) &= \text{vertical line with constant value } X_{i1} \\ L_H(X_i) &= \text{horizontal line with contant value } X_{i2}. \end{aligned}$$

For any two events A and B, $P(A \backslash B) = P(A) - P(A \cap B) \geq P(A) - P(B)$. Therefore, if we define the events

$$\begin{aligned} A &= [x \in R(X_1, X_2)] \\ B &= [x_n \in R(X_1, X_2)], \end{aligned}$$

we have

$$D^r(x) - D^r(x_n) = P(A) - P(B)$$

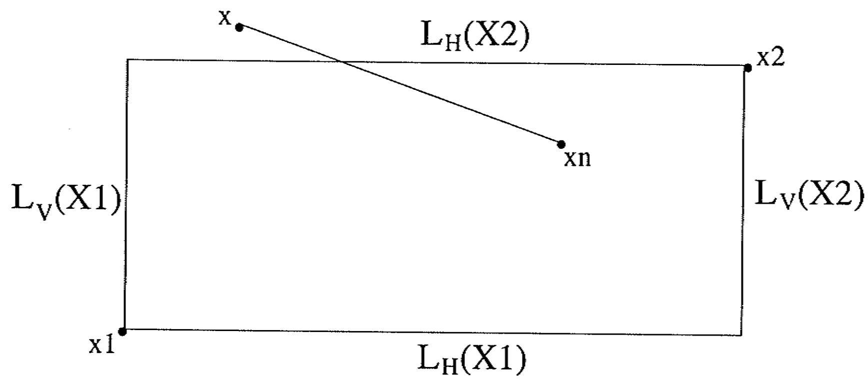

Figure 3.7: $x_n \in R(X_1, X_2)$ and $x \notin R(X_1, X_2)$, so $\overline{xx_n} \cap L_H(X_2)$. Therefore, this rectangle contributes to the difference $D^r(x) - D^r(x_n)$.

$$
\begin{aligned}
&\leq P[x \in R(X_1, X_2) \cap x_n \notin R(X_1, X_2)] \\
&\leq P([L_V(X_1) \cap \overline{xx_n}] \cup [L_H(X_1) \cap \overline{xx_n}] \cup \\
&\quad [L_V(X_2) \cap \overline{xx_n}] \cup [L_H(X_2) \cap \overline{xx_n}]) \\
&\leq P([L_V(X_1) \cap \overline{xx_n}]) + P([L_H(X_1) \cap \overline{xx_n}]) \\
&\quad + P([L_V(X_2) \cap \overline{xx_n}]) + P([L_H(X_2) \cap \overline{xx_n}]).
\end{aligned}
$$

Note again that, in the context of intersecting line segments, "$\cap$" refers to the two lines crossing one another at some point. This differs from our usual use of the intersection symbol "$\cap$", meaning the intersection of two sets or events.

It can similarly be shown (or simply understood by symmetry) that

$$
\begin{aligned}
D^r(x_n) - D^r(x) &\leq P(L_V(X_1) \cap \overline{xx_n}) + P(L_H(X_1) \cap \overline{xx_n}) \\
&\quad + P(L_V(X_2) \cap \overline{xx_n}) + P(L_H(X_2) \cap \overline{xx_n}).
\end{aligned}
$$

As such, we have

$$\begin{aligned}
|D^r(x) - D^r(x_n)| &\leq P(L_V(X_1) \cap \overline{xx_n}) + P(L_H(X_1) \cap \overline{xx_n}) \\
&\quad + P(L_V(X_2) \cap \overline{xx_n}) + P(L_H(X_2) \cap \overline{xx_n}) \\
&= 2P(L_V(X_1) \cap \overline{xx_n}) + 2P(L_H(X_1) \cap \overline{xx_n}), \quad (3.34)
\end{aligned}$$

since X_1 and X_2 are identically distributed.

We define the events

$$A_n = [L_V(X_1) \cap \overline{xx_n}] \ \forall\ n$$

$$B_n = [L_H(X_1) \cap \overline{xx_n}] \ \forall\ n.$$

Then, if $x_0 = (x_{01}, x_{02})$,

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \left[\bigcup_{k=n}^{\infty} A_k \right] = [x \in L_V(X_1)] = [X_{11} = x_{01}].$$

By Lemma A.2.2, and since we know that F is absolutely continuous (hence has absolutely continuous marginals),

$$\limsup_{n\to\infty} P(A_n) \leq P(\limsup_{n\to\infty} A_n) = P(X_{11} = x_{01}) = 0. \qquad (3.35)$$

By a similar argument, it can be shown that $\limsup_{n\to\infty} P(B_n) = 0$ Therefore, $|D^r(x) - D^r(x_n)| \to 0$ as $n \to \infty$. This completes the proof, using the sequential characterization of continuity. **QED**

We give an explicit form for the rectangular depth $D^r(x)$ in two dimensions:

Theorem 3.2.8 *Given a fixed point $x_0 = (x_{01}, x_{02})$ in the plane,*

$$\begin{aligned} D^r(x_0) &= P\{x_0 \in R(X_1, X_2)\} \\ &= \left[2\int_{-\infty}^{x_{02}}\int_{-\infty}^{x_{01}} f(x,y)dxdy\right]\left[\int_{x_{02}}^{\infty}\int_{x_{01}}^{\infty} f(x,y)dxdy\right] \\ &+ \left[2\int_{x_{02}}^{\infty}\int_{-\infty}^{x_{01}} f(x,y)dxdy\right]\left[\int_{-\infty}^{x_{02}}\int_{x_{01}}^{\infty} f(x,y)dxdy\right]. \end{aligned}$$

Proof: From Figure 3.8, we see that the only way for x_0 to be in the rectangle $R(X_1, X_2)$ is for X_2 to lie in the "quadrant" diagonal from the "quadrant" containing X_1. We define a quandrant as one of the four sections of the plane generated by passing lines through x_0 that are parallel to the x- and y-axes. In other words:

$$\begin{aligned} [x_0 \in R(X_1, X_2)] &= \{[X_{11} < x_{01} < X_{21}] \cap [X_{12} < x_{02} < X_{22}]\} \cup \\ & \quad \{[X_{11} < x_{01} < X_{21}] \cap [X_{22} < x_{02} < X_{12}]\} \cup \\ & \quad \{[X_{21} < x_{01} < X_{11}] \cap [X_{12} < x_{02} < X_{22}]\} \cup \\ & \quad \{[X_{21} < x_{01} < X_{11}] \cap [X_{22} < x_{02} < X_{12}]\} \\ &= A \cup B \cup C \cup D. \end{aligned}$$

Note that A, B, C and D are mutually exclusive and exhaustive.

Now, by the independence of X_1 and X_2,

$$\begin{aligned} P(A) &= P([X_{11} < x_{01}] \cap [X_{12} < x_{02}] \cap [x_{01} < X_{21}] \cap [x_{02} < X_{22}]) \\ &= P([X_{11} < x_{01}] \cap [X_{12} < x_{02}])P([X_{21} > x_{01}] \cap [X_{22} > x_{02}]) \\ &= \left[\int_{-\infty}^{x_{02}}\int_{-\infty}^{x_{01}} f(x_{11}, x_{12})dx_{11}dx_{12}\right]\left[\int_{x_{02}}^{\infty}\int_{x_{01}}^{\infty} f(x_{21}, x_{22})dx_{21}dx_{22}\right]. \end{aligned}$$

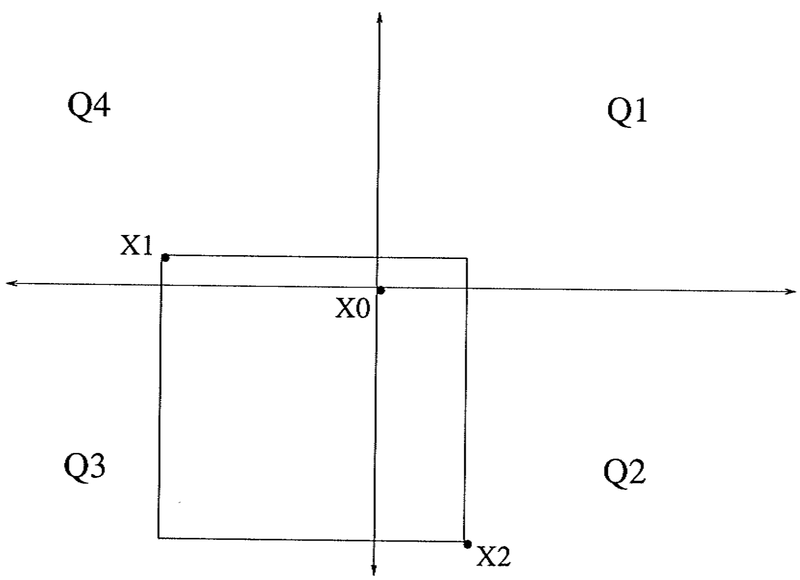

Figure 3.8: X_1 and X_2 must lie in opposite (diagonal) "quadrants" in order for $x_0 \in R(X_1, X_2)$.

Similarly, since X_1 and X_2 are identically distributed,

$$
\begin{aligned}
P(D) &= P([X_{21} < x_{01}] \cap [X_{22} < x_{02}] \cap [x_{01} < X_{11}] \cap [x_{02} < X_{12}]) \\
&= \left[\int_{-\infty}^{x_{02}} \int_{-\infty}^{x_{01}} f(x_{21}, x_{22}) dx_{21} dx_{22}\right] \left[\int_{x_{02}}^{\infty} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12}\right] \\
&= P(A).
\end{aligned}
$$

By a similar argument, it can be shown that

$$
P(B) = P(C) = \left[\int_{x_{02}}^{\infty} \int_{-\infty}^{x_{01}} f(x,y) dx dy\right] \left[\int_{-\infty}^{x_{02}} \int_{x_{01}}^{\infty} f(x,y) dx dy\right].
$$

This completes the proof. **QED**

We now give an explicit form for the variance of the sample rectangular depth function:

Theorem 3.2.9 *The variance of the sample rectangular depth function $D_n^r(x_0)$ is*

$$
Var(D_n^r(x_0)) = \binom{n}{2}^{-1} \left[D^r(x_0) + 2(n-2)D_{(2)}^r(x_0) + (3-2n)(D^r(x_0))^2\right],
$$

where

$$
D_{(2)}^r(x_0) = a^2 b + b^2 a + c^2 d + d^2 c,
$$

in which

$$
\begin{aligned}
a &= \int_{x_{02}}^{\infty} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12} \\
b &= \int_{-\infty}^{x_{02}} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12} \\
c &= \int_{-\infty}^{x_{02}} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12} \\
d &= \int_{x_{02}}^{\infty} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12}.
\end{aligned}
$$

Proof. Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample from a c.d.f. F. We know that

$$
\begin{aligned}
Var[D_n^r(x_0)] &= E[D_n^r(x_0)]^2 - (E[D_n^r(x_0)])^2 \\
&= E\left[\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \mathbb{I}(x_0 \in R(X_i, X_j))\right]^2 \\
&\quad + \left(E\left[\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \mathbb{I}(x_0 \in R(X_i, X_j))\right]\right)^2 .
\end{aligned}
\tag{3.36}
$$

The first term in Equation (3.36) yields the expansions in Equation (3.14), with R replacing $\odot$. Hence, by the argument following Equation (3.14), we have

$$
\begin{aligned}
& E[D_n^r(x_0)]^2 \qquad (3.37) \\
= & \binom{n}{2}^{-2} \left\{ \binom{n}{2} E[\mathbb{I}(x_0 \in R(X_1, X_2))] \right. \\
& + \binom{n}{2} (2(n-2)) E[\mathbb{I}(x_0 \in R(X_1, X_2)) \times \mathbb{I}(x_0 \in R(X_1, X_3))] \\
& \left. + \binom{n}{2} \binom{n-2}{2} E[\mathbb{I}(x_0 \in R(X_1, X_2)) \times \mathbb{I}(x_0 \in R(X_3, X_4))] \right\} \\
= & \binom{n}{2}^{-1} \{P(x_0 \in R(X_1, X_2)) \\
& + 2(n-2) E[\mathbb{I}(x_0 \in R(X_1, X_2) \cap x_0 \in R(X_1, X_3))] \\
& \left. + \binom{n-2}{2} [P(x_0 \in R(X_1, X_2))]^2 \right\} \\
= & \binom{n}{2}^{-1} \{D^r(x_0) + 2(n-2) E[\mathbb{I}(x_0 \in R(X_1, X_2) \cap x_0 \in R(X_1, X_3))] \\
& \left. + \binom{n-2}{2} (D^r(x_0))^2 \right\}
\end{aligned}
\tag{3.38}
$$

Now we are left with the middle term. We'll use conditioning. Note that

$$x_0 \in [R(X_1, X_2) \cap R(X_1, X_3)]$$

implies X_2 and X_3 must both be in the "quadrant" opposite X_1, where the quadrants, $Q1, Q2, Q3, Q4$ are formed by passing lines through x_0 parallel to the x- and y-axes. In particular,

$$\begin{aligned}
Q1 &= \{(x,y) \in \mathbf{R}^2 | x > x_{01}, y > x_{02}\} \\
Q2 &= \{(x,y) \in \mathbf{R}^2 | x > x_{01}, y < x_{02}\} \\
Q3 &= \{(x,y) \in \mathbf{R}^2 | x < x_{01}, y < x_{02}\} \\
Q4 &= \{(x,y) \in \mathbf{R}^2 | x < x_{01}, y > x_{02}\}.
\end{aligned}$$

Since $Q1$ and $Q3$ are opposite quadrants and $Q2$ and $Q4$ are also opposite, then

$$\begin{aligned}
& E[\mathbb{I}(x_0 \in R(X_1, X_2) \cap x_0 \in R(X_1, X_3))] \\
= \; & P(x_0 \in R(X_1, X_2) \cap x_0 \in R(X_1, X_3)) \\
= \; & P(X_2 \in Q1, X_3 \in Q1, X_1 \in Q3) \\
& + P(X_2 \in Q2, X_3 \in Q2, X_1 \in Q4) \\
& + P(X_2 \in Q3, X_3 \in Q3, X_1 \in Q1) \\
& + P(X_2 \in Q4, X_3 \in Q4, X_1 \in Q2) \\
= \; & P(X_2 \in Q1)P(X_3 \in Q1)P(X_1 \in Q3) \\
& + P(X_2 \in Q2)P(X_3 \in Q2)P(X_1 \in Q4) \\
& + P(X_2 \in Q3)P(X_3 \in Q3)P(X_1 \in Q1) \\
& + P(X_2 \in Q4)P(X_3 \in Q4)P(X_1 \in Q2)
\end{aligned}$$

$$
\begin{aligned}
&= [P(x_{01} < X_{11}, x_{02} < X_{12})]^2 P(x_{01} > X_{11}, x_{02} > X_{12}) \\
&\quad +[P(x_{01} < X_{11}, x_{02} > X_{12})]^2 P(x_{01} > X_{11}, x_{02} < X_{12}) \\
&\quad +[P(x_{01} > X_{11}, x_{02} > X_{12})]^2 P(x_{01} < X_{11}, x_{02} < X_{12}) \\
&\quad +[P(x_{01} > X_{11}, x_{02} < X_{12})]^2 P(x_{01} < X_{11}, x_{02} > X_{12}). \qquad (3.39)
\end{aligned}
$$

Note that the second last and last steps in Equation (3.39) are due, respectively, to the independence and identical distributions of X_1, X_2 and X_3. Finally, continuing with Equation (3.39), we have

$$
\begin{aligned}
&= \left[\int_{-\infty}^{x_{02}} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12}\right] \\
&\quad \times \left[\int_{x_{02}}^{\infty} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12}\right]^2 \\
&\quad + \left[\int_{x_{02}}^{\infty} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12}\right] \\
&\quad \times \left[\int_{-\infty}^{x_{02}} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12}\right]^2 \\
&\quad + \left[\int_{x_{02}}^{\infty} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12}\right] \\
&\quad \times \left[\int_{-\infty}^{x_{02}} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12}\right]^2 \\
&\quad + \left[\int_{-\infty}^{x_{02}} \int_{x_{01}}^{\infty} f(x_{11}, x_{12}) dx_{11} dx_{12}\right] \qquad (3.40) \\
&\quad \times \left[\int_{x_{02}}^{\infty} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}) dx_{11} dx_{12}\right]^2
\end{aligned}
$$

We denote $P(x_0 \in R(X_1, X_2) \cap x_0 \in R(X_1, X_3))$ as $D^c_{(2)}(x_0)$ to facilitate our notation. Repeating Equations (3.18) and (3.19) with r in the place of c, the proof is complete. **QED**

Analogous to that for circular depth, the variance for $D^r_n(x)$ in any dimension p can simply be obtained by altering the above variance calculations in two dimensions by replacing the quadrants $Qi, i = 1, 2, 3, 4$ by $p-$ dimensional boxes $Qi, i = 1, 2, ..., 2^p$, and by using p-tuple integration rather than double integration.

In the *Lagrange multiplier method*, we can minimize or maximize an equation

$$H(x_1, x_2, ..., x_n)$$

subject to m restrictions

$$\psi_i(x_1, x_2, ..., x_n) = 0, \;\; i = 1, 2, ..., m$$

by solving the equations

$$\frac{\partial H}{\partial x_i} + \lambda_1 \frac{\partial \psi_1}{\partial x_i} + \lambda_2 \frac{\partial \psi_2}{\partial x_i} + \cdots + \lambda_m \frac{\partial \psi_m}{\partial x_i} = 0, \;\; i = 1, 2, ..., n.$$

There are $n + m$ unknowns and $n + m$ equations, and so a unique solution, $x_1, x_2, ..., x_n$ may exist. We will use this method to find an upper bound on $D^r(x_0)$ and $D^r_{(2)}(x_0)$:

Theorem 3.2.10 *The rectangular depth $D^r(x)$ has value less than or equal to $\frac{1}{2}$ for every $x \in \mathbf{R}^p$. That is,*

$$D^r(x) \leq \frac{1}{2} \; \forall \; x \in \mathbf{R}^p.$$

Proof. We will begin with the proof the theorem for $p = 2$. Recall from the proof of Theorem 3.2.8 that

$$\begin{aligned} D^r(x_0) &= [P(X_2 \in Q1)][P(X_1 \in Q3)] + [P(X_2 \in Q2)][P(X_1 \in Q4)] \\ &\quad + [P(X_2 \in Q3)][P(X_1 \in Q1)][P(X_2 \in Q4)][P(X_1 \in Q2)] \end{aligned}$$

$$
\begin{aligned}
&= 2\{[P(X_1 \in Q1)][P(X_1 \in Q3)]\} \\
&\quad +2\{[P(X_1 \in Q2)][P(X_1 \in Q4)]\} \\
&= 2ab + 2cd,
\end{aligned} \tag{3.41}
$$

where

$$
\begin{aligned}
a &= P(X_1 \in Q1) \\
b &= P(X_1 \in Q3) \\
c &= P(X_1 \in Q2) \\
d &= P(X_1 \in Q4).
\end{aligned}
$$

Note that the last step in Equation (3.41) is possible by the identical distributions of X_1 and X_2. We now want to maximize $2ab + 2cd$, subject to the restrictions

$$
\begin{aligned}
\psi_1(a,b,c,d) &= a+b+c+d-1 = 0 \\
\psi_2(a,b,c,d) &= ab \geq 0, \text{ and} \\
\psi_3(a,b,c,d) &= cd \geq 0.
\end{aligned}
$$

We can equivalently maximize $ab + cd$. We let

$$
f = ab + cd + \lambda_1(a+b+c+d) + \lambda_2(ab) + \lambda_3(cd).
$$

Using the Lagrange multiplier method, we have

$$
\begin{aligned}
\frac{\partial f}{\partial a} &= b + \lambda_1 + b\lambda_2 = 0 \\
\frac{\partial f}{\partial b} &= a + \lambda_1 + a\lambda_2 = 0 \\
\frac{\partial f}{\partial c} &= d + \lambda_1 + d\lambda_3 = 0 \\
\frac{\partial f}{\partial d} &= c + \lambda_1 + c\lambda_3 = 0.
\end{aligned} \tag{3.42}
$$

From Equations (3.42), we have that

$$\begin{aligned}(1+\lambda_2)b &= (1+\lambda_2)a \quad \Rightarrow a=b \text{ and} \\ (1+\lambda_3)d &= (1+\lambda_3)c \quad \Rightarrow c=d. \end{aligned} \tag{3.43}$$

From $\psi_1(a,b,c,d)=0$, we have

$$a+b+c+d=1 \quad \Rightarrow 2a+2c=1 \quad \Rightarrow c=\frac{1}{2}-a,$$

and so

$$ab+cd=a^2+c^2=a^2+(\frac{1}{2}-a)^2=2a^2-a+\frac{1}{4}.$$

But since $a=b$ and $a+b+c+d=1$, it follows that $a \leq \frac{1}{2}$, and so

$$ab+cd=2a^2-a+\frac{1}{4}\leq 2\left(\frac{1}{2}\right)^2-\frac{1}{2}+\frac{1}{4}=\frac{1}{4}.$$

Therefore we have that

$$D^r(x_0)=2ab+2cd\leq\frac{1}{2}.$$

Note that conditions ψ_2 and ψ_3 are equivalent to saying that a,b,c and d are all greater than or equal to zero. Although the product of two negatives gives a positive, if a and b were both negative, then c and d would have to be as well. Conditions ψ_2 and ψ_3 would be satisfied, but the first condition, ψ_1, would be violated. This is therefore impossible, and this completes the proof for $p=2$. Note that $D^r(x_0)$ can in fact attain this maximum value when either

$$(a,b,c,d)=\left(\frac{1}{2},\frac{1}{2},0,0\right) \quad \text{or} \quad (a,b,c,d)=\left(0,0,\frac{1}{2},\frac{1}{2}\right).$$

In p dimensions, we would have

$$\begin{aligned} D^r(x_0) &= \sum_{i=1}^{2^p} P(X \in Qi)P(X \in -Q_i) \\ &= \sum_{i=2}^{2^p} a_i b_i, \end{aligned}$$

where $a_i = P(X \in Qi), b_i = P(X \in -Qi)$, and where $-Qi$ is the quadrant opposite of Qi. By a simple extension of the above argument, $D^r(x_0) \leq \frac{1}{2}$ and $D^r(x_0) = \frac{1}{2}$ when $a_{i_0} = b_{i_0} = \frac{1}{2}$ for some i_0, and $a_i = b_i = 0$ for all $i \neq i_0$. **QED**

Theorem 3.2.11 *If X_1, X_2, X_3 are i.i.d. with c.d.f. F, the probability that a fixed point x_0 is contained in both $B(X_1, X_2)$ and $B(X_1, X_3)$ is less than or equal to $\frac{1}{4}$. That is,*

$$D^r_{(2)}(x_0) \leq \frac{1}{4} \ \forall \ x_0 \in \mathbf{R}^p.$$

Proof. We will begin with the proof for $p = 2$. Recall from Theorem 3.2.9 that

$$\begin{aligned} D^r_{(2)}(x_0) &= [P(X_2 \in Q1)]^2[P(X_1 \in Q3)] + [P(X_2 \in Q2)]^2[P(X_1 \in Q4)] \\ &\quad +[P(X_2 \in Q3)]^2[P(X_1 \in Q1)][P(X_2 \in Q4)]^2[P(X_1 \in Q2)] \\ &= a^2b + b^2a + c^2d + d^2c, \end{aligned} \tag{3.44}$$

where a, b, c and d are defined as in Theorem 3.2.10. Note that the last step in Equation (3.44) is possible by the identical distributions of X_1 and X_2.

To maximize Equation (3.44), we proceed as follows:

$$a^2b + b^2a + c^2d + d^2c = ab(a+b) + cd(c+d) \leq \ ab + cd$$

since $a+b \leq 1$ and $c+d \leq 1$. But we found the maximum value for $ab+cd$ to be $\frac{1}{4}$ in the proof for Theorem 3.2.10. This completes the proof. We note that $D^r_{(2)}(x_0)$ can actually attain this value if

$$(a,b,c,d) = \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) \quad \text{or} \quad (a,b,c,d) = \left(0, 0, \frac{1}{2}, \frac{1}{2}\right).$$

The proof for any dimension p would follow analogously to that stated at the end of the proof for Theorem 3.2.10. **QED**

We now find an upper bound for the variance of $D^r_n(x)$.

Lemma 3.2.12 *For any dimension $p \geq 2$, the variance of the sample rectangular depth has an upper bound of $\frac{3}{(4n-6)n(n-1)} + \frac{n-2}{n(n-1)}$ for all $n \geq 2$. That is,*

$$Var(D^r_n(x_0)) \leq \frac{3}{(4n-6)n(n-1)} + \frac{n-2}{n(n-1)} \;\; \forall\, n \geq 2.$$

Proof. In light of Theorems 3.2.10, 3.2.9 and 3.2.11, the proof is identical to that of Lemma 3.1.12. **QED**

Another consequence of Theorem 3.2.9 is the following corollary:

Corollary 3.2.13 *$D^r_n(x)$ is a weakly consistent estimator of $D^r(x)$, i.e. $D^r_n(x)$ converges to $D^r(x)$ in probability for any $x \in \mathbf{R}^p$.*

The proof follows from Lemma 3.1.12 and Markov's inequality applied to $Y = |D^r_n(x) - D^r(x)|$ (see Lemma A.2.6).

3.3 A Comparison of Our Three Geometric Depth Functions

Now that we have examined the simplicial, circular and rectangular depth functions, we can compare them and discuss some of the potential advantages and disadvantages of their use in practice.

To date, Liu's simplicial depth has proven to be a very useful and valuable tool, both in theory and in practice. As discussed in Chapter 2, the sample simplicial depth function $D_n^s(x)$ has some very desirable properties, including unbiasedness and consistency. The simplicial depth function $D^s(x)$ was shown to be continuous and to decrease to zero as x gets further and further away from the "centre" of the distribution F. Furthermore, for angularly symmetric distributions, $D^s(x)$ was shown in Liu (1990) to be monotone non-increasing, and to have a maximum of 2^{-p}, attained at the point of angular symmetry.

We obtained many similar results for both the circular and rectangular depth functions. Both $D_n^c(x)$ and $D_n^r(x)$ were found to be unbiased and weakly consistent. (The property of strong consistency will be examined in my doctoral thesis.) The circular and rectangular depth functions $D^c(x)$ and $D^r(x)$ were both shown to be continuous. Conditions on F ensuring monotonicity will also be examined in my doctoral thesis. The maximum values for both $D^c(x)$ and $D^r(x)$ were found to be $\frac{1}{2}$, **regardless** of the probability distribution F.

A question of obvious interest is the use of circular and rectangular depth in defining (locating) the centre of a multivariate distribution. The omission of this topic in this thesis was deliberate.

While attending the DIMACS Workshop on Data Depth (held May 12-14, 2003 at Rutgers University, U.S.A.), my supervisor became acquainted with a group at Penn State University (Pittsburgh, U.S.A.) that were pursuing this very direction. This group, consisting of Professor Thomas P. Hettmansperger and graduate students Ryan T. Elmore and Fengjuan Xuan studied the analogues of $\hat{\mu}_n$ and μ (see Section 2.1) using circles and rectangles instead of simplices. Included was a study of the variance and consistency of $\hat{\mu}_n$. At the time of completion of this thesis, their work had not yet been published, or to our knowledge, been completely drafted. This explains our omission of their work in the bibliography.

It is important to note that, while they formulated the definition of circular and rectangular depth simultaneously, yet independent of us, they had not, to my supervisor's knowledge, studied any of the properties for them found in this chapter.

3.3.1 Advantages of Circular and Rectangular Depth

In Theorem 2.4.3, it was shown that the "tractable" form of $D^s(x)$ is extremely difficult, if not impossible, to calculate in practice. On the contrary, the tractable forms for $D^c(x)$ and $D^r(x)$ may be tedious, depending on the form of $f(x)$, but are nonetheless much simpler to calculate and the reasoning motivating the formulae for their calculations is much more apparent.

Calculating the sample simplicial depth, $D^s_n(x)$ can also be quite tedious, and represents a hurdle in the practical applications of Liu's simplicial depth. Among other related complexities, to determine whether a point lies in a given simplex $S(X_1, X_2, ..., X_p)$, we must solve a system of $p+1$ linear equations. In contrast, the methods for determining whether a point falls in a ball $C(X_1, X_2)$ or a box $B(X_1, X_2)$ for the circular and rectangular depths, respectively, are comparatively much simpler and straightforward, especially if p is large.

We have also calculated explicit tractable forms for the variances of $D_n^c(x)$ and $D_n^r(x)$. In principle we could have proceeded along the same lines to "calculate" the variance of Liu's sample simplicial depth function $D_n^s(x)$. Here, rather than separating the pairs of pairs of indices $\{i, j\}$ and $\{k, l\}$ into three cases, we would have had to separate the pairs of triplets of indices $\{i, j, k\}$ and $\{t, u, v\}$ into four cases which are more difficult to describe and count. One can only imagine how much more complex the calculation of the variance would be. For example, even in the plane, the expansion of $E[D_n^s(x_0)]^2$ would include terms of the form

$$P((x_0 \in \Delta(X_1, X_2, X_3)) \cap (x_0 \in \Delta(X_1, X_4, X_5))),$$

which are much more complicated (i.e. involve many more cases) than the already-complex $D^s(x_0)$ itself. (See Theorem 2.4.3).

Related to the variance, Liu (1990) states the following lemma, the proof of which can be found in Serfling (1980):

Lemma 3.3.1 *Let F be a distribution on $\mathbf{R}^p$ and $X_1, X_2, ..., X_n$ a random* *sample from F. Let*

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, X_{i_2}, ..., X_{i_m})$$

be a U-statistic with kernel h of degree m. If h is bounded by a constant c, then for any $r \geq 2$,

$$E(U_n - E(U_n))^r \leq \frac{K}{n^{\frac{r}{2}}},$$

where K depends only on c.

The simplicial depth, circular depth and rectangular depth functions are all U-statistics with respective kernels $\mathbb{I}(x \in \Delta(X_1, X_2, ..., X_{p_1}))$ (of degree $p+1$), $\mathbb{I}(x \in C(X_1, X_2))$ (of degree 2) and $\mathbb{I}(x \in B(X_1, X_2))$ (of degree 2). As such, if we let $r = 2$, Lemma 3.3.1 gives an upper bound of $\frac{K}{n}$ for the variances of all three sample depth functions, where K is some constant. Clearly, Lemmas 3.1.12 and 3.2.12 improve on this result, in that they yield specific, although possibly not optimal rates of convergence for $Var(D_n^r(x_0))$ and $Var(D_n^c(x_0))$.

3.3.2 Advantages of Simplicial Depth

Up to this point, we have focused on advantages of circular and rectangular depth over simplicial depth. One of the potential drawbacks of using the circular and rectangular depth functions is that we do not get as many balls or boxes generated for a given sample size as we would simplices in the case of simplicial depth. Recall that, given a sample of size n in $\mathbf{R}^p$, we can generate $\binom{n}{p+1}$ simplices, but only $\binom{n}{2}$ balls or boxes. If $p = 2$, this is a ratio of

$$\frac{n-2}{3} \quad (\geq 1 \; \forall \, n \geq 5).$$

For example, if we have a sample of size $n = 20$, we get six times as many triangles generated as we would circles or rectangles. The discrepancy becomes even greater as n and p increase. As such, for a given sample size, $D_n^s(x)$ will give a better estimate for $D^s(x)$ than $D_n^c(x)$ and $D_n^r(x)$ will for $D^c(x)$ and $D^r(x)$, respectively.

We also notice that a simplex is formed in a way that (at least for unimodal, continuous and monotone densities $f(x)$) will enclose a "more reasonable" region than will a ball or a box. This is due to the fact that all extreme points of the simplex have come directly from our sample, and so they likely are found in regions of substantial probability. In contrast, only two of the points on a ball or two corners of a box actually came from our sample of data values. As such, they may contain points in areas of very low probability. To further illustrate, if $f(x)$ has finite support, $C(X_1, X_2)$ and $R(X_1, X_2)$ will, with positive probability, contain points that are not even in the support of $f(x)$. A possible solution to this latter problem is to simply alter our definitions of $D_n^c(x)$ and $D_n^r(x)$ by multiplying each $\mathbb{I}(x \in C(X_1, X_2))$ (respectively, $\mathbb{I}(x \in R(X_1, X_2))$), by an indicator function $\mathbb{I}(x \in S)$ where S is the support of $f(x)$. In contrast, note that a random simplex, $S(X_1, X_2, ..., X_{p+1})$ will lie entirely in the support of f, provided the support is a convex subset of $\mathbf{R}^p$. For an illustration of the above discussion, see Figure 3.9.

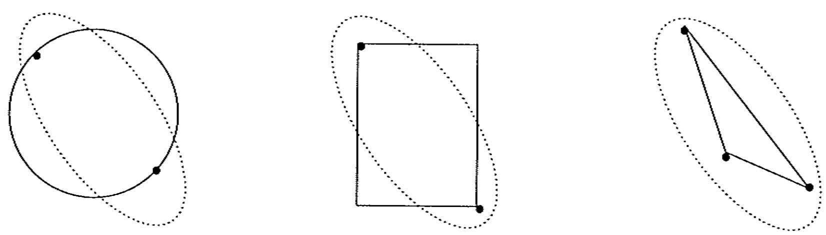

Figure 3.9: The circle $\odot(X_1, X_2)$, the rectangle $R(X_1, X_2)$ and the triangle $\Delta(X_1, X_2, X_3)$ superimposed on the support (dotted region) of f, respectively.

Chapter 4

The Use of Data Depth in Quality Control

Univariate quality control charts are commonly used in industry to monitor quality characteristics of items being produced. We now examine some problems associated with the use of these charts in a multivariate setting, and we examine the application of Liu's simplicial depth in this important area of statistics.

4.1 Statistical Process Control

Roughly stated,

Definition 4.1.1 *Statistical process control* (or *quality control* as it is often referred to) is the set of methods for monitoring and improving the quality characteristics of a process.

Quality control may be employed, for example, on the production line of an automobile manufacturer. In the context of statistical process control, *quality* does not refer to luxurious options such as leather seats and high-tech stereo systems. Rather, *quality* refers to conformance to requirements, or as quality expert Joseph M. Juran (1992) describes it, "fitness for use".

For a vehicle to be fit for use, we are not concerned with the clarity of sound of the stereo system, but rather the reliability and smooth operation of the automobile. We need all the parts to properly fit together so as to function properly and safely. After all, faulty brakes will be of far graver consequence than seats which may not be as soft as leather.

In *statistical process control* (SPC), we are concerned with controlling the variability of a process. If certain quality characteristics vary excessively, they may not conform to requirements. As such, variability is seen as the enemy of quality.

Moore (1995) gives the following definitions:

Definition 4.1.2 A variable (quality characteristic) that continues to be described by the same c.d.f. over time is said to be *in control.*

Definition 4.1.3 *Control charts* are graphical tools used to monitor the control of a process and alert us when the process has been disturbed.

Shewhart (1931) introduced the notion of control charts to monitor a single quality characteristic of a process. Shewhart $\bar{x}$ charts are used to monitor the process mean when a variable is assumed to follow a normal distribution with mean μ and standard deviation σ. The chart consists of a centre line drawn at μ and parallel lines drawn at the upper and lower control limits (UCL and LCL, respectively), drawn at $\mu \pm 3\frac{\sigma}{\sqrt{n}}$. Values of $\bar{x}$ are plotted on the chart for samples taken over time.

As long as the process is in control, points should be randomly scattered about the centre line, and, by properties of the normal distribution, a point will only fall outside the control limits 0.3% of the time. When we see such a point (known as an *out-of-control signal*), the process is stopped to determine if the mean or variance of the process has changed significantly, or if the signal was simply the result of an exceptional sample (which we expect to occur randomly in 3 of 1000 cases, even when the process is in control). In the former case, the problem must be investigated and action taken to correct the shift in parameters before resuming production. As such, it is hoped that any problem can be detected and remedied as soon as possible, so time and money are not wasted manufacturing items that are not fit for use.

Note that a point falling outside the control limits is not the only potential out-of-control signal. A trend of increasing or decreasing points over time is also an indication of a gradual shift in the process mean, and should be investigated.

Furthermore, the assumption of known mean and variance are not often valid. In such cases, Shewhart control charts based on sample means and ranges are available and frequently used.

Due to time or cost constraints, it is often not feasible to take samples of size greater than one, and so the Central Limit Theorem cannot be invoked. In such cases, individual control charts are available. Most any SPC textbook will provide a more in-depth discussion of Shewhart control charts. See

Montgomery (2001), for example.

4.2 The Use of Data Depth in Multivariate Quality Control

We now consider one very important application of data depth, namely, its use in the longstanding and difficult problem of multivariate quality control. The discussion in this section is based in large part on that found in Stoumbos and Jones (2000).

4.2.1 The Problem

The traditional approach to control charts has been parametric and univariate. The problem encountered with these methods is two-fold. Firstly, most methods include the assumption that a specific quality characteristic's measurements follow a normal distribution. We often suspect (or know) this not to be the case, and when only small sample sizes are feasible, as is the case with individuals control charts, we cannot rely on the invocation of the central limit theorem. Furthermore, most control charts examine only one quality characteristic, when in reality, most items produced must be "fit for use" with respect to several different correlated variables. The examination of all variables separately is of little use, as it ignores the correlation structure of the characteristics.

Hotelling (1947) introduced the first multivariate control chart, a Shewhart-type chart that was later extended by Jackson and Morris (1957) using principle component analysis. Such multivariate charts, however, involve the often-unrealistic assumption of multivariate normality. In addition, when the production run is short, we may not have sufficient data to estimate all the necessary parameters. Coleman (1997) asserted that he "would never believe the multivariate normal assumption for industrial data", and that he "cannot believe that there are tests for multivariate normality with sufficient power for practical sample sizes." His conclusion was simple: "Distribution-free multivariate SPC is what we need."

4.2.2 The Use of Liu's Simplicial Depth in Multivariate SPC

Liu (1995) proposed a new Shewhart-type control chart based on her notion of simplicial depth (Liu, 1990). The chart is nonparametric in nature, and can be used to detect a significant change in the centre or variation of a process. The central idea underlying these control charts is to "reduce each multivariate measurement to a univariate index — namely, its relative centre-outward ranking induced by a data depth measure" (Liu, 1995).

Individuals Control Charts Based on Liu's Simplicial Depth

Liu (1995) used the name *r chart* for her multivariate control chart based on simplicial depth. To avoid confusion with the commonly used *R chart* for sample ranges, Stoumbos and Jones (2000) refer to the chart as the *simplicial depth for individuals chart* (SDI chart), as we will do here.

Let $X_1, X_2, ..., X_n$ be a random sample. Assume that, when the process is in control, the X_i have absolutely continuous distribution F on $\mathbf{R}^p$. When the process is out of control, the X_i have absolutely continuous distribution G on $\mathbf{R}^p$. Let $X_1^*, X_2^*, ..., X_n^*$ be an i.i.d. reference sample from F, where $n \geq (p+1)$, which we use to construct an SDI chart. After the reference sample is selected, we take a further sample of independent p-variate observations $X_1, X_2, ...$ at specified sampling points $i = 1, 2, ...$ respectively. The following definition is taken from Stoumbos and Jones (2000):

Definition 4.2.1 The *control statistic* for the SDI chart is

$$r_n(X_i) = \frac{1}{n+1}\left[1 + \sum_{j=1}^{n} \mathbb{I}(D_n^s(X_j^*) \leq D_n^s(X_i))\right], \ \forall\ i = 1, 2, ..., \qquad (4.1)$$

where $D_n^s(X_j^*)$ and $D_n^s(X_i)$ are calculated with respect to the *expanded reference sample* (ERS) $X_1^*, X_2^*, ..., X_n^*, X_i$, for $i = 1, 2, ...$

The control statistic $r_n(X_i)$ represents the proportion of observations in the ERS that have simplicial depths at least as low as $D_n^s(X_i)$. Large values of $r_n(X_i)$ indicate that the point X_i is relatively deep within the data cloud generated by the ERS, and so there is no cause for concern. Exceptionally low values of $r_n(X_i)$ indicate that X_i is an outlying point and may be indicative of the process being out-of-control. As such, fixing a pre-assigned control limit g, at the ith sampling point, if $r_n(X_i) \leq g$, we stop sampling and declare the process out-of-control. If $r_n(X_i) > g$, the process is declared in control and the sampling is continued.

Liu's SDI chart plots i against $r_n(X_i)$. Because of the centre-outward notion of ranking associated with data depth, this chart only has, i.e. only requires, a lower control limit, g. The control limit g, as is the case for most any control chart, is chosen in a manner so as to attain a reasonable "average run length" when the process is in control:

Definition 4.2.2 The *average run length* (ARL) of a process is the expected number of observations to be taken before an out-of-control signal occurs.

When the process is in control (i.e. X_i's represent a random sample from c.d.f. F), we would like for the ARL to be high, as we do not wish to frequently stop the process to examine causes of the signal when in fact the process is in control. The inverse of the ARL computed under the in-control distribution F is known as the *false alarm rate* (FAR). Conversely, we would like the ARL to be low when the process is out-of-control (i.e. the data represent a random sample from c.d.f. G, which is different from F).

The FAR is simply the probability of an observation from F being represented as out-of-control. For the classic univariate Shewhart control charts with the control limits at $\mu \pm 3\frac{\sigma}{\sqrt{n}}$, this value is known to be 0.0027. Assuming observations are i.i.d. with c.d.f. F (when the process is in control), the waiting time Y for an out-of-control signal is a geometric random variable with parameter 0.0027. As such,

$$E(Y) = \text{ARL} = \frac{1}{\text{FAR}} = \frac{1}{0.0027} = 370.4.$$

In general, the in-control ARL of the SDI chart is the mean of a geometric distribution,

$$\text{ARL} = \frac{1}{\alpha},$$

where $\alpha = \text{FAR}$.

Liu and Singh (1993) showed that, when F is absolutely continuous,

$$r_n(X_i) \xrightarrow{d} U[0,1] \text{ as } n \to \infty, \tag{4.2}$$

where "$\xrightarrow{d}$" denotes convergence in distribution and $U[0,1]$ is the uniform distribution on $[0,1]$. In other words, if we take a sufficiently large reference sample, $r_n(X_i)$ will behave approximately like a $U[0,1]$ random variable. Equation 4.2 suggests that we use $\alpha = g$ as the lower control limit in our SDI chart. See Liu (1995) for illustrations of such charts.

The Effect of the Reference Sample Size in SDI Charts

Liu (1995) recommended using a reference sample of size $n = 500$ when $p = 2$, and more in higher dimensions. However, the recommendation was more heuristic than mathematically justified. As such, Stoumbos and Jones (2000) investigated the problem of determining the smallest required reference sample size for SDI charts, in order that $r_n(X_i)$ could even possibly attain a value lower than g. Stoumbos and Jones (2000) state and prove the following theorem:

Theorem 4.2.3 *Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample from a c.d.f. F on $\mathbf{R}^p$. The minimum sample simplicial depth any sample point can have is $\frac{p+1}{n}$.*

Proof. Clearly, a sample point $X_i \in \{X_1, X_2, ..., X_n\}$ will attain the minimum sample simplicial depth if and only if $X_i \notin S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})$ whenever $i \notin \{i_1, i_2, ..., i_{p+1}\}$. That is, X_i is contained only inside the closed simplices for which X_i is a vertex. X_i is a vertex of precisely $\binom{n-1}{p}$ simplices. In total, $\binom{n}{p+1}$ simplices are generated. As such,

$$\begin{aligned} D_n^s(X_i) &= \binom{n}{p+1}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq n} \mathbb{I}(X_i \in S(X_{i_1}, X_{i_2}, ..., X_{i_{p+1}})) \\ &= \binom{n-1}{p} \binom{n}{p+1}^{-1} \\ &= \left(\frac{(n-1)!}{p!(n-p-1)!}\right) \left(\frac{(p+1)!(n-p-1)!}{n!}\right) = \frac{p+1}{n}. \end{aligned}$$

The proof is complete. **QED**

Stoumbos and Jones (2000) show that a sample point assumes this minimum sample simplicial depth if and only if it is an extreme point in the convex hull determined by $X_1, X_2, ..., X_n$. They also assert and prove the fact that at least $p+1$ points share the minimum simplicial depth. As such, lower bounds for $r_n(X_i)$ are given as

$$r_n(X_i) \geq \frac{E_i^{(n+1)}}{n+1} \geq \frac{p+1}{n+1} \text{ for } i = 1, 2, ..., \tag{4.3}$$

where $E_i^{(n+1)}$ is the number of extreme points in the ERS $X_1^*, X_2^*, ..., X_n^*, X_i$. (See Definition 2.4.4).

Equation (4.3) implies that the smallest possible FAR that can be used in an SDI chart is $g = \alpha = \frac{p+1}{n+1}$. For any value of g lower than this, the control chart will never signal, thereby defeating our purpose. The largest attainable in-control ARL that can be attained is thus $\frac{1}{\alpha} = \frac{n+1}{p+1}$. When $n = 500$ and $p = 2$, this corresponds to an ARL of $\frac{501}{3} = 167$. Note however that this represents the best-case scenario, as $\frac{p+1}{n+1}$ is the absolute lower bound for $r_n(X_i)$. It is quite possible that there will be more than $p+1$ extreme sample points in the expanded reference sample, and so our control limit must be at least $\frac{E_i^{(n+1)}}{n+1}$.

Since $g = \frac{E_i^{(n+1)}}{n+1}$ is a random variable, the distribution of $E_i^{(n+1)}$ must be examined. A comprehensive examination of asymptotic distributions for $E_i^{(n+1)}$ for various distributions F is provided in Stoumbos and Jones (2000), based on earlier work, such as that in Efron (1965). They use the expectation of the extreme number of points in the lower bound of $r_n(X_i)$, rather than the number of extreme points $E_i^{(n+1)}$ itself. They obtain some remarkable conclusions, notably that Liu severely underestimated the number of points required in the reference sample to obtain a reasonable in-control ARL.

For example, as calculated in Stoumbos and Jones (2000), using $n = 500$ data points from a bivariate standard normal distribution gives an expected minimum FAR of 0.0215, corresponding to an in-control ARL of only 46.53. Using $n = 500$ data points from a bivariate uniform distribution on the unit circle gives an expected minimum FAR of 0.0530, corresponding to an in-control ARL of only 18.86. In quality control, false alarm rates of at most 0.0027 are usually used, corresponding to an in-control ARL of 370.4.

Also from Stoumbos and Jones (2000), to achieve an expected minimum FAR of 0.0027 for a bivariate standard normal distribution, a reference sample of size $n = 4,816$ is required. For a trivariate standard normal distribution, $n = 11,498$ observations are required. Worse still, $n = 44,347$ bivariate uniform observations on the unit disc and $n = 2,745,678$ trivariate uniform observations on the unit sphere are required to achieve an expected minimum FAR of 0.0027. The reference sample size requirements increase quickly as the number of dimensions increases. (See Stombous and Jones (2000)). Still, nonparametric multivariate control charts such as the SDI charts require far fewer reference sample observations than, say, Hotelling's (1947) T^2 chart, which requires the estimation of several parameters as well.

4.2.3 The Use of Rectangular Depth in Individuals Control Charts

Replacing simplicial depth with rectangular depth and simplices with boxes, we can construct an individuals control chart analogous to the SDI chart using rectangular depth. We will call it a *Rectangular Depth for Individuals (RDI) chart.*

Definition 4.2.4 The *control statistic* for the RDI chart is

$$r_n(X_i) = \frac{1}{n+1}\left[1 + \sum_{j=1}^{n} \mathbb{I}(D_n^r(X_j^*) \leq D_n^r(X_i))\right], \tag{4.4}$$

where $D_n^r(X_j^*)$ and $D_n^r(X_i)$ are calculated with respect to the expanded reference sample $X_1^*, X_2^*, ..., X_n^*, X_i$, for $i = 1, 2, ...$

The control statistic $r_n(X_i)$ represents the proportion of all observations in the ERS that have rectangular depths at least as low as $D_n^r(X_i)$. The chart is constructed and monitored in the same manner as the SDI chart.

In my doctoral thesis, I will investigate, among other things, conditions ensuring the limit law in Equation (4.2) when $r_n(X_i)$ is defined with respect to rectangular depth, as per Definition 4.4. Following the same argument as that for SDI charts, such a result would suggest that we take $g = \alpha =$ FAR as the lower limit in our RDI charts.

The Effect of the Reference Sample Size in RDI Charts

In light of the above discussions, we would like to determine the smallest required reference sample size for the RDI chart in order that $r_n(X_i)$ can attain its lower bound g (using 0.0027 as an example). But firstly, we state an analogue to Theorem 4.2.3.

Theorem 4.2.5 *Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample from a c.d.f. F on $\mathbf{R}^p$. The minimum sample rectangular depth any sample point can have is $\frac{2}{n}$, regardless of the dimension p.*

Proof: Clearly, a sample point $X_i \in \{X_1, X_2, ..., X_n\}$ will attain the minimum sample rectangular depth if and only if $X_i \notin B(X_{i_1}, X_{i_2})$ whenever $i \notin \{i_1, i_2\}$. That is, X_i is contained in only those boxes for which X_i is one of the generating corners. X_i is a generating corner of precisely $n-1$ boxes. In total, $\binom{n}{2}$ boxes are generated. As such,

$$\begin{aligned} D_n^r(X_i) &= \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathbb{I}(X_i \in B(X_{i_1}, X_{i_2})) \\ &= \frac{n-1}{\binom{n}{2}} \\ &= (n-1)\left(\frac{2!(n-2)!}{n!}\right) = \frac{2}{n}. \end{aligned}$$

The proof is complete. **QED**

It can easily be shown that, in any sample, the maximum number of sample points that can share the minimum sample rectangular depth is $2p$. For example, in the case of $p = 2$, a maximum of four points do not lie in any rectangles other than those generated by the point itself. These four points correspond to those with the minimum and maximum values of the $x-$ and $y-$ components. This argument obviously extends to higher dimensions. Each sample point X_i possessing either the minimum or maximum value of some component will have the minimum sample rectangular depth. As such, sharp lower bounds for $r_n(X_i)$ are given as

$$r_n(X_i) \geq \frac{2p}{n+1} \text{ for i = 1, 2,...,} \tag{4.5}$$

unlike Equation (4.3), which actually depends on the number of extreme points in the expanded reference sample, which in turn is a function of the distribution F of the random variable X.

Equation (4.5) implies that the smallest possible FAR that can be used in an RDI chart is $g = \alpha = \frac{2p}{n+1}$. When $n = 500$, this corresponds to an ARL of $\frac{501}{4} = 125.25$.

Theorem 4.2.6 *To achieve a minimum FAR of 0.0027 for an RDI chart for* **any** *bivariate distribution, we require 1481 observations.*

Proof. By Equation (4.5), the minumum FAR is $\frac{2p}{n+1}$, so we have

$$\begin{aligned} \frac{4}{n+1} &= 0.0027 \\ \Rightarrow n+1 = \frac{4}{0.0027} &= 1481.48 \\ \Rightarrow n &= 1480.48, \end{aligned}$$

which we round up to 1481. The proof is complete. **QED**

Note that this represents a distinct advantage over using an SDI chart. Firstly, the required sample size does not depend on the distribution of the X_i's, nor on any heavy calculations. Secondly, and more importantly, many fewer observations are required in the construction of an RDI chart. From Section 4.2, to construct an SDI chart when F is bivariate standard normal, a reference sample of 4,816 observations is required, whereas 44,347 observations are required in the case of a uniform distributionon on the unit disc. These represent, respectively, sample sizes over times and 29 times greater than those required for an RDI chart for the same random variables.

Theorem 4.2.7 *To achieve a minimum FAR of 0.0027 for an RDI chart for* **any** *trivariate distribution, we require 2222 observations.*

Proof. By Equation 4.5, the minumum FAR is $\frac{2p}{n+1}$, so we have

$$\begin{aligned} \frac{6}{n+1} &= 0.0027 \\ \Rightarrow n+1 = \frac{6}{0.0027} &= 2222.22 \\ \Rightarrow n &= 2221.22, \end{aligned}$$

which we round up to 2222. The proof is complete. **QED**

Note that this again represents a distinct advantage over using an SDI chart. From Section 4.2, to construct an SDI chart when F is trivariate standard normal, 11,498 observations are required in the reference sample, whereas 2,745,678 observations are required in the case of a uniform distribution on the unit sphere. These represent, respectively, sample sizes over 5 times and 1235 times greater than those required for an RDI chart for the same distributions, a remarkable improvement, especially in the latter case.

It is obvious from the preceding examples that the relative efficiency of the RDI chart in terms of required sample size with respect to the SDI chart continues to increase with the dimension p. Although not quantified in this thesis, these results should be tempered by the fact that $D_n^r(x)$ will converge more slowly to $D^r(x)$ than $D_n^s(x)$ will to $D^s(x)$, for the reasons discussed in Section 3.3.2. Therefore, the rate of convergence in the rectangular depth analogue of Equation (4.2) may be considerably slower. These issues will be examined in my doctoral thesis.

It should be noted that while $r_n(X_i)$ can, in principle be defined using circular depth, the determination of the maximum number of sample points X_i attaining the minimum sample circular depth $D_n^c(X_i)$ is rather geometrically and mathematically complicated. The notion of the potential use of circular depth in quality control will be explored in my doctoral thesis.

Appendix A

Statistical Notation and Basic Results in Probability

A.1 Basic Terminology and Notation

In this thesis, we will frequently work with points in p-dimensional Euclidean space, denoted $\mathbf{R}^p, p \geq 1$. We express points $x_0, x_1, x_2, \ldots$ in $\mathbf{R}^p$ by $x_0 = (x_{01}, x_{02}, \ldots, x_{0p}), x_1 = (x_{11}, x_{12}, \ldots, x_{1p}), x_2 = (x_{21}, x_{22}, \ldots, x_{2p}), \ldots$ When there is no ambiguity in doing so, we will sometimes use x and y to refer either to points in $\mathbf{R}^p$, or to numbers in $\mathbf{R}$.

The distance between x_1 and x_2 in $\mathbf{R}^p$ is defined as

$$||x_1 - x_2|| = \sqrt{\sum_{i=1}^{p}(x_{1i} - x_{2i})^2}.$$

Given a random variable X taking values in $\mathbf{R}^p$, the *cumulative distribu-*

tion function (c.d.f.), $F : \mathbf{R}^p \longrightarrow \mathbf{R}$ of x is defined as

$$F(x) = P(X \leq x), \quad x \in \mathbf{R}^p,$$

where the relation $X \leq x$ is defined co-ordinate-wise.

F (equivalently, X) is said to be *absolutely continuous* if F has a density function, $f : \mathbf{R}^p \longrightarrow [0, \infty)$. That is,

$$\int_{\mathbf{R}^p} f(x)dx = 1$$

and

$$F(B) = P(X \in B) = \int_B f(x)dx$$

for any Borel set, $B \subseteq \mathbf{R}^p$. In particular,

$$F(x_0) = \int_{-\infty}^{x_{0p}} \cdots \int_{-\infty}^{x_{02}} \int_{-\infty}^{x_{01}} f(x_{11}, x_{12}, ..., x_{1p})dx_{11}dx_{12} \cdots dx_{1p}.$$

If X has density f on $\mathbf{R}^p$ and $h : \mathbf{R}^p \longrightarrow \mathbf{R}$, then the *expectation* of $h(X)$ is

$$E[h(X)] = \int_{\mathbf{R}^p} h(x)f(x)d(x).$$

This is the so-called "law of the unconscious statistician".

Two continuous random variables X_1 and X_2 are said to be independent if and only if

$$P(X_1 \leq x_1, X_2 \leq x_2) = P(X_1 \leq x_1)P(X_2 \leq x_2) \;\; \forall \;\; x_1, x_2 \in \mathbf{R}^p$$

The two random variables are said to be i.i.d. (*independently and identically distributed*) if and only if they are independent and both have the same distribution F in $\mathbf{R}^p$.

A set $A \subseteq \mathbf{R}^p$ is said to be *convex* if, given any $x, y \in A$,

$$\alpha x + (1-\alpha)y \in A \;\; \forall\, \alpha \in [0,1].$$

Given $x_1, x_2, ..., x_n \in \mathbf{R}^p$, the *convex hull*, denoted $CH(x_1, x_2, ..., x_n)$ is the smallest closed, convex subset of $\mathbf{R}^p$ containing the points $x_1, x_2, ..., x_n$.

If X_n is a sequence of random variables, we say that X_n *converges in probability* to a random variable X if, for every fixed $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \rightarrow 0 \;\; \text{as}\, n \rightarrow \infty.$$

The sequence X_n is said to *converge almost surely (a.s.)* to X if

$$P\left(\lim_{n\rightarrow\infty} X_n = X\right) = 1.$$

A.2 Basic Results in Probability

The following lemmas provide some basic tools from probability, frequently used throughout the thesis. Their basic proofs can be found in most standard probability texts. See Billingsley (1986), for example.

Lemma A.2.1 *If P is a probability measure and $A_1, A_2, ..., A_n$ are any events, then*

(i) (monotonicity) $A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)$,

(ii) (subadditivity) $P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$.

Lemma A.2.2 *(Fatou's lemma)*

Given events $A_1, A_2, \ldots$, $\limsup_{n\to\infty} P(A_n) \leq P(\limsup_{n\to\infty} A_n)$ *, where*

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} A_k\right].$$

Lemma A.2.3 *(Conditioning 1)*

Given mutually exclusive, exhaustive events $B_1, B_2, \ldots, B_n$ *and an event* A*,*

$$\begin{aligned} P(A) &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i). \end{aligned}$$

Lemma A.2.4 *(Conditioning 2)*

Given an event A *and an absolutely continuous random variable* X *with p-dimensional density* $f(x), x \in \mathbf{R}^p$

$$P(A) = \int_{\mathbf{R}^p} P(A|X = x) f(x) dx.$$

Lemma A.2.5 *(Conditioning 3)*

Given a random variable X *and another random variable* Y*, the latter absolutely continuous with p-dimensional density* $f(y), y \in \mathbf{R}^p$*,*

$$\begin{aligned} E(X) &= E(E(X|Y)) \\ &= \int_{\mathbf{R}^p} E(X|Y = y) f(y) dy. \end{aligned}$$

Lemma A.2.6 *(Markov's Inequality)*

Given a random variable, $Y \geq 0$ *and a constant* $\alpha > 0$*,*

$$P[Y \geq \alpha] \leq \frac{1}{\alpha} E(Y).$$

Appendix B

Illustrations for Calculating Simplicial Depth in Two Dimensions

This appendix contains the 36 graphs necessary in formulating the proof for Theorem 2.4.3.

Recall that the probability of X_3 being a point such that x_0 lies in the triangle $\Delta(x_1, x_2, X_3)$ for fixed values of x_0, x_1 and x_2 is dependent on the ordering of the x- and y- components of x_0, x_1 and x_2.

The following graphs represent the illustrations for each combination of these orderings. Recall that probabilities of the 36 cases were grouped into 12 expressions h_i, $i = 1, 2, ..., 12$. The case (i.e. the function h_i) to which each diagram belongs will be given in the caption. In each graph, we denote L_1 that unique line which passes through both x_0 and x_1, whereas L_2 is that unique line passing through both x_0 and x_2. We denote as A the region in which X_3 must fall in order to have $x_0 \in \Delta(x_1, x_2, X_3)$.

Note that some cases are divided into two illustrations, as the region A depends on the slopes of the lines L_1 and L_2.

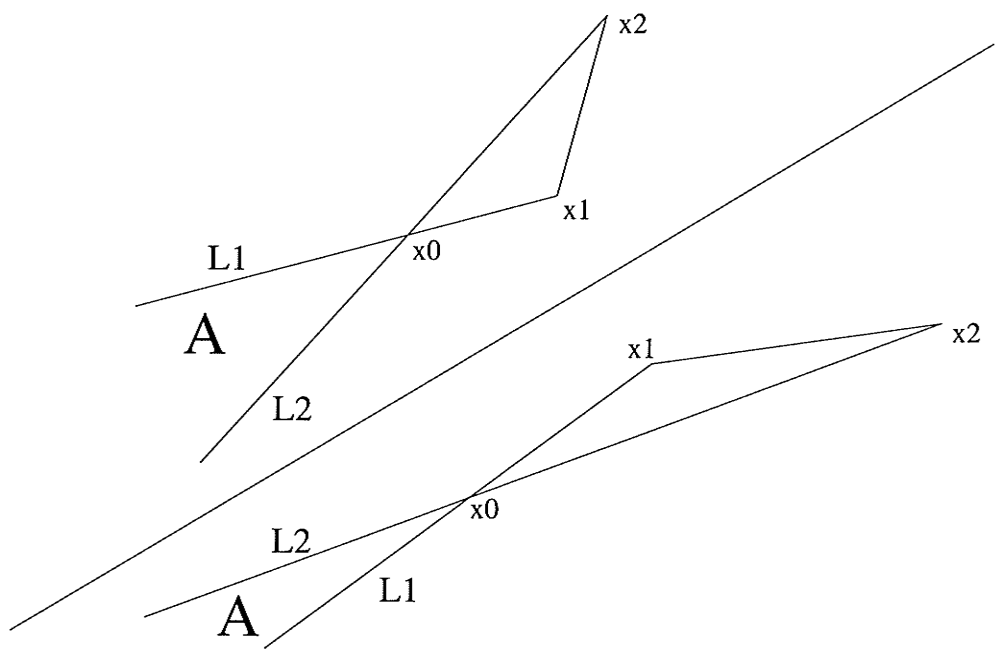

Figure B.1: The case where $x_{01} < x_{11} < x_{21}$ and $x_{02} < x_{12} < x_{22}$. Both sub-cases contribute to the function h_1 in Theorem 2.4.3.

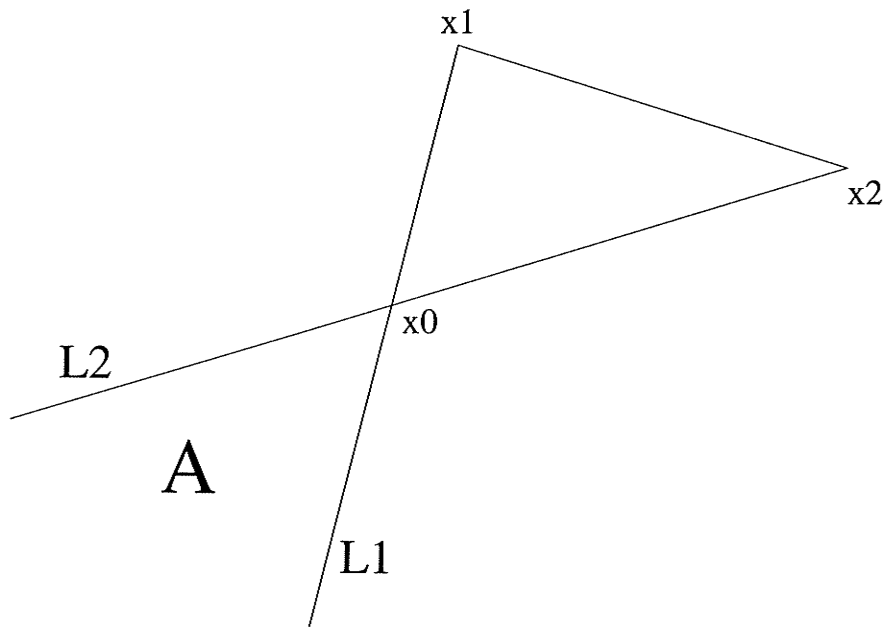

Figure B.2: The case where $x_{01} < x_{11} < x_{21}$ and $x_{02} < x_{22} < x_{12}$. This case contributes to the function h_1 in Theorem 2.4.3.

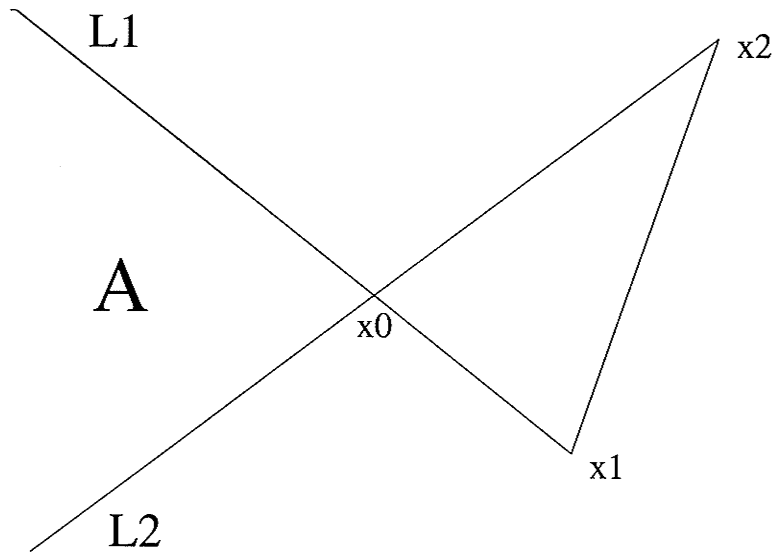

Figure B.3: The case where $x_{01} < x_{11} < x_{21}$ and $x_{12} < x_{02} < x_{22}$. This case contributes to the function h_1 in Theorem 2.4.3.

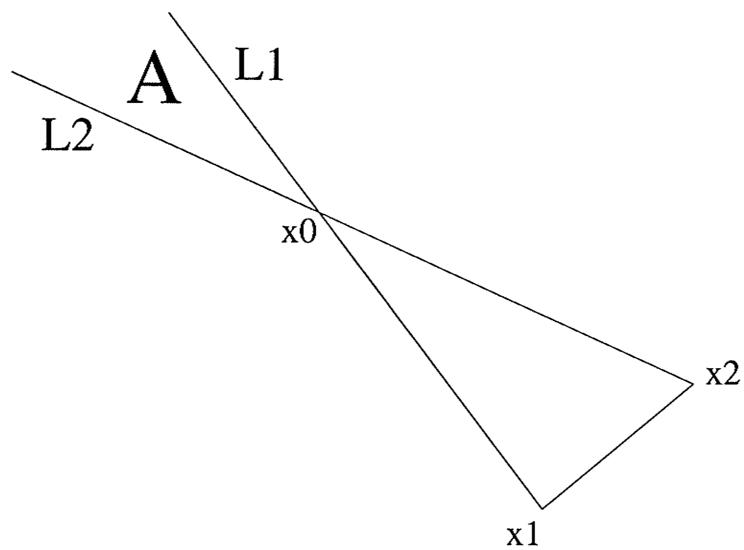

Figure B.4: The case where $x_{01} < x_{11} < x_{21}$ and $x_{12} < x_{22} < x_{02}$. This case contributes to the function h_1 in Theorem 2.4.3.

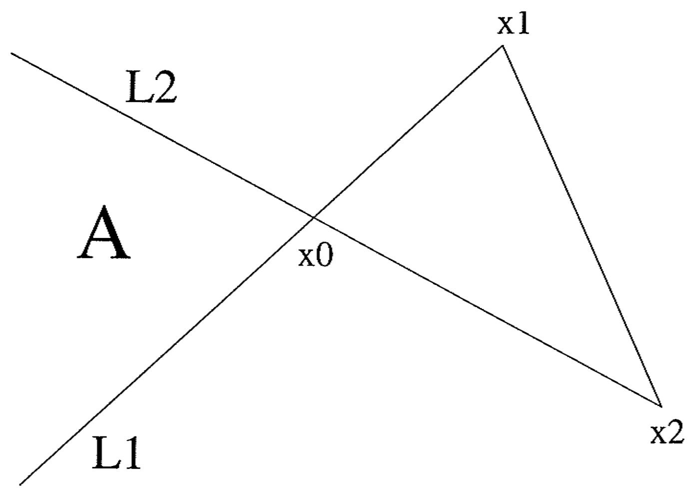

Figure B.5: The case where $x_{01} < x_{11} < x_{21}$ and $x_{22} < x_{02} < x_{12}$. This case contributes to the function h_1 in Theorem 2.4.3.

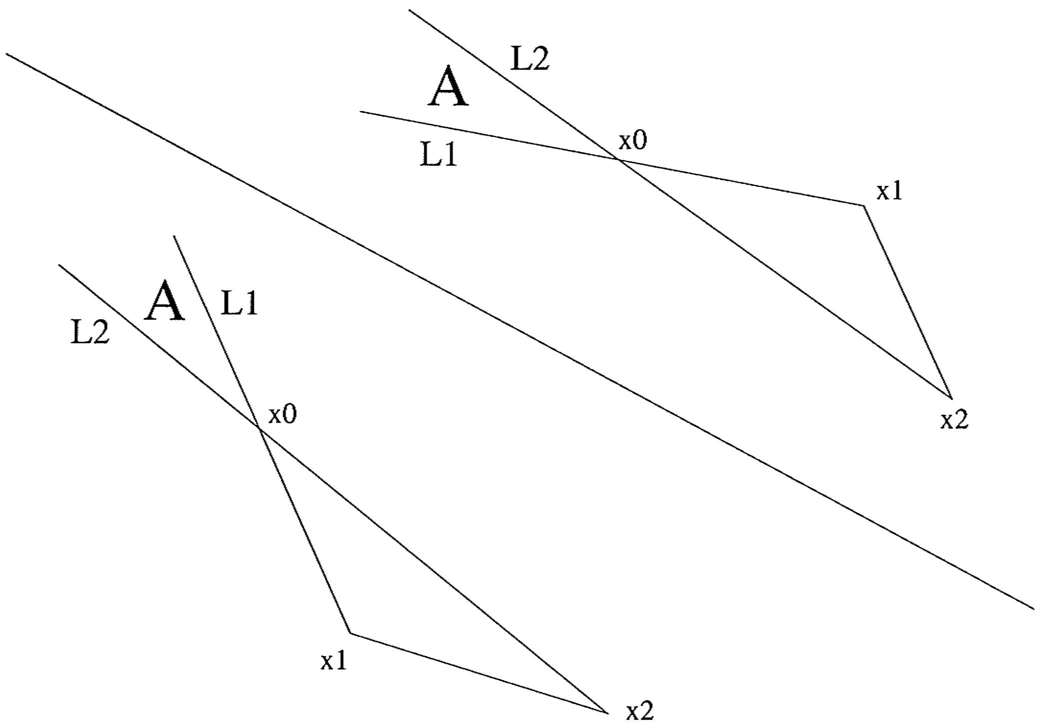

Figure B.6: The case where $x_{01} < x_{11} < x_{21}$ and $x_{22} < x_{12} < x_{02}$. Both sub-cases contribute to the function h_1 in Theorem 2.4.3.

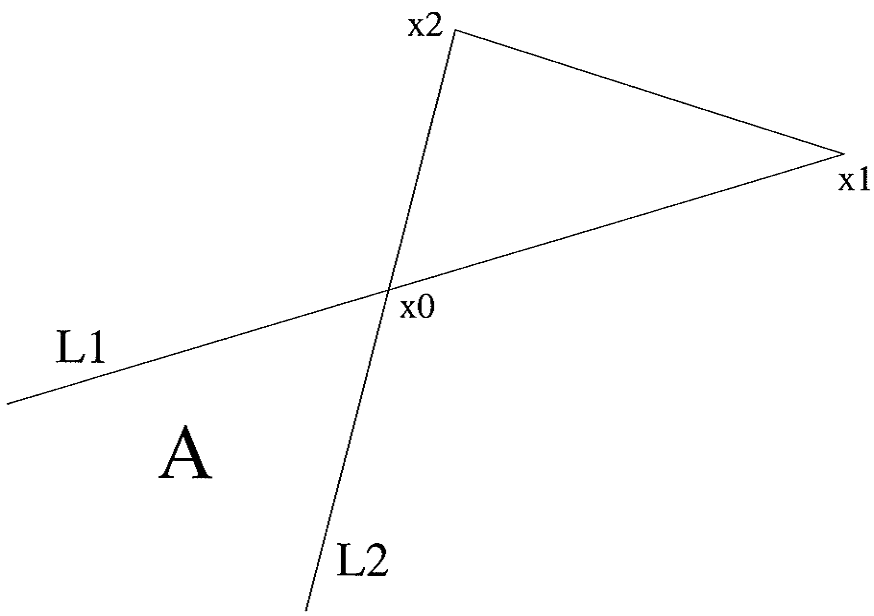

Figure B.7: The case where $x_{01} < x_{21} < x_{11}$ and $x_{02} < x_{12} < x_{22}$. This case contributes to the function h_1 in Theorem 2.4.3.

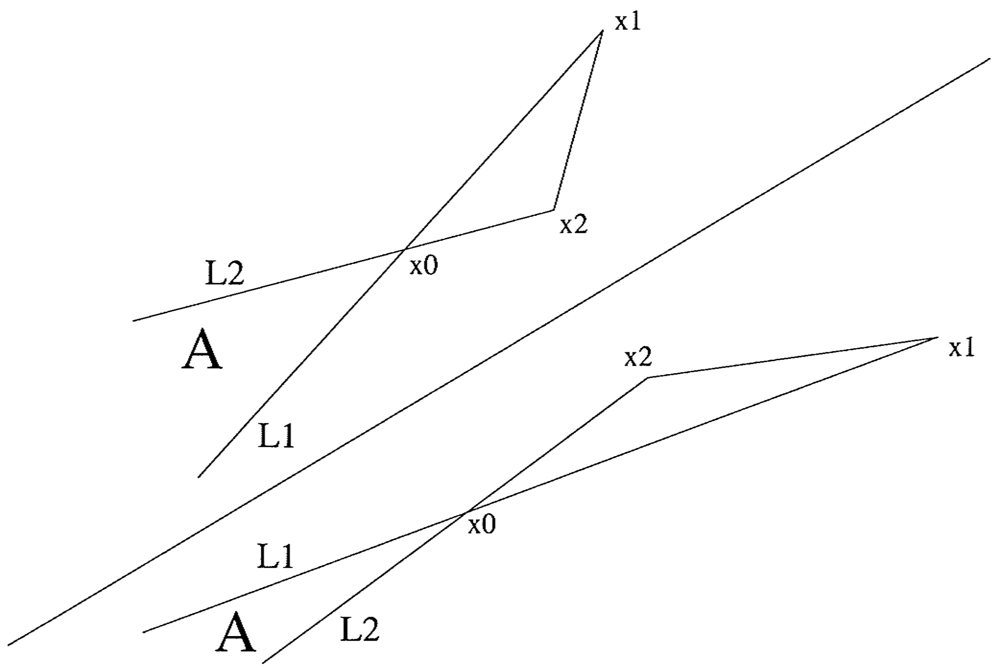

Figure B.8: The case where $x_{01} < x_{21} < x_{11}$ and $x_{02} < x_{22} < x_{12}$. Both sub-cases contribute to the function h_1 in Theorem 2.4.3.

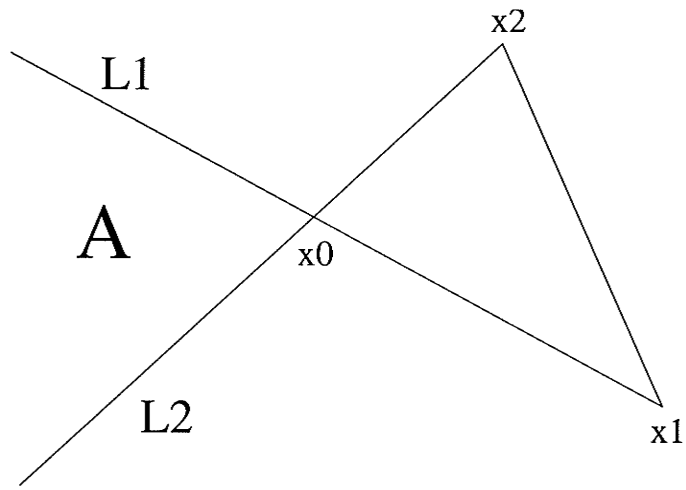

Figure B.9: The case where $x_{01} < x_{21} < x_{11}$ and $x_{12} < x_{02} < x_{22}$. This case contributes to the function h_1 in Theorem 2.4.3.

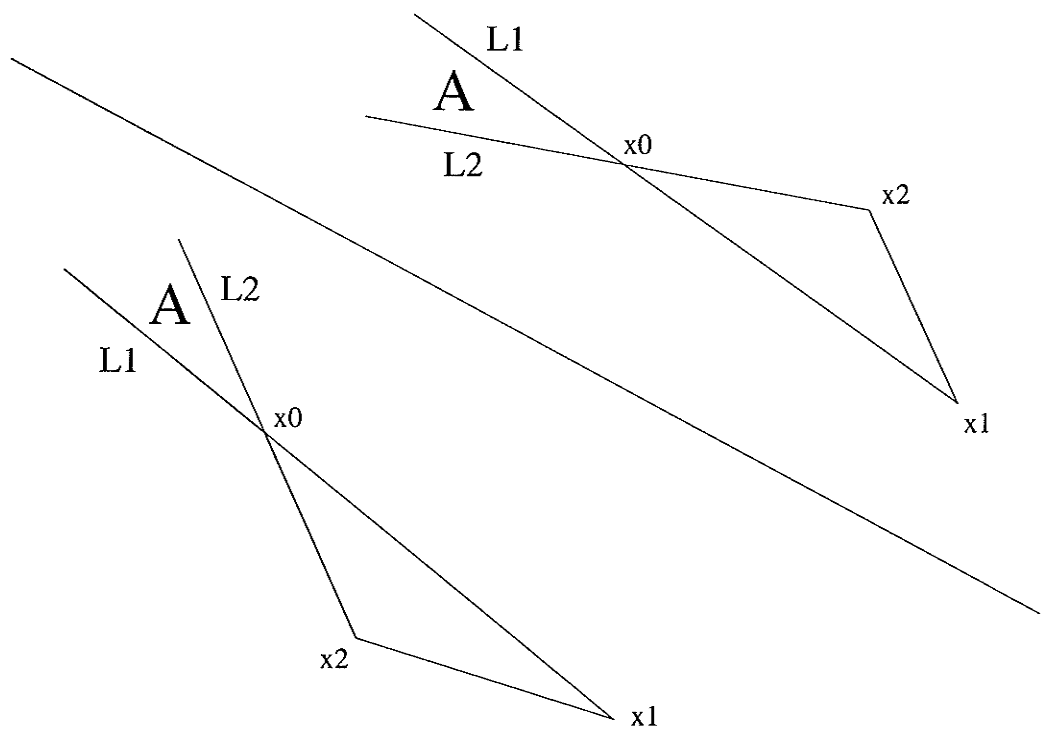

Figure B.10: The case where $x_{01} < x_{21} < x_{11}$ and $x_{12} < x_{22} < x_{02}$. Both sub-cases contribute to the function h_1 in Theorem 2.4.3.

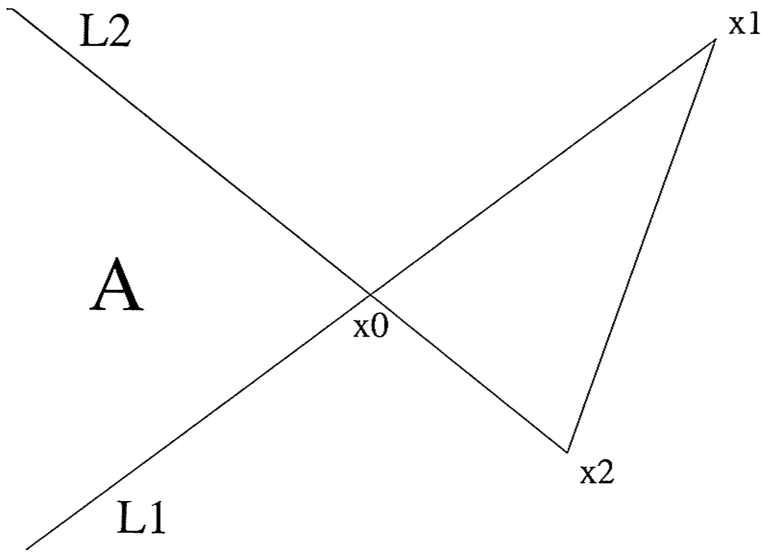

Figure B.11: The case where $x_{01} < x_{21} < x_{11}$ and $x_{22} < x_{02} < x_{12}$. This case contributes to the function h_1 in Theorem 2.4.3.

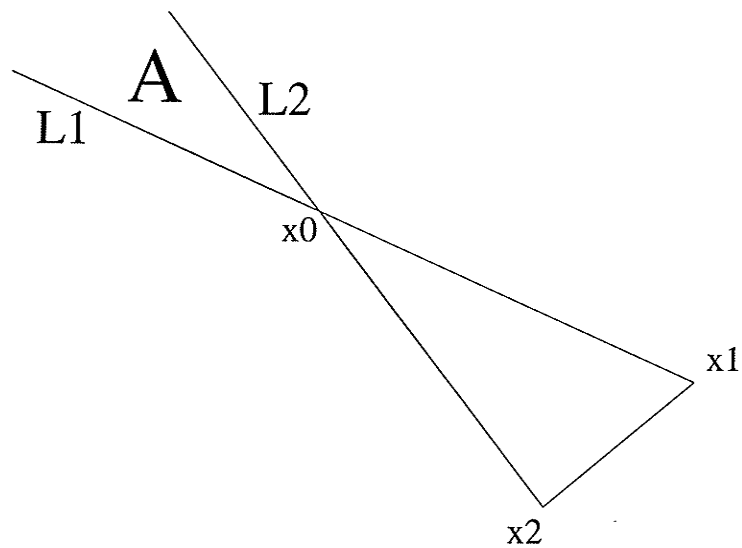

Figure B.12: The case where $x_{01} < x_{21} < x_{11}$ and $x_{22} < x_{12} < x_{02}$. This case contributes to the function h_1 in Theorem 2.4.3.

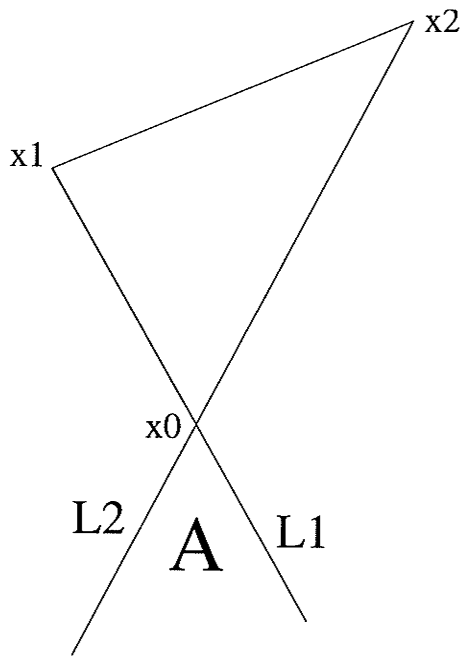

Figure B.13: The case where $x_{11} < x_{01} < x_{21}$ and $x_{02} < x_{12} < x_{22}$. This case contributes to the function h_3 in Theorem 2.4.3.

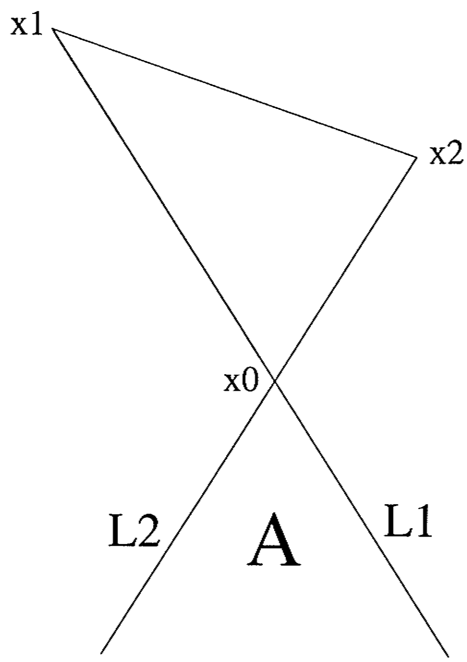

Figure B.14: The case where $x_{11} < x_{01} < x_{21}$ and $x_{02} < x_{22} < x_{12}$. This case contributes to the function h_3 in Theorem 2.4.3.

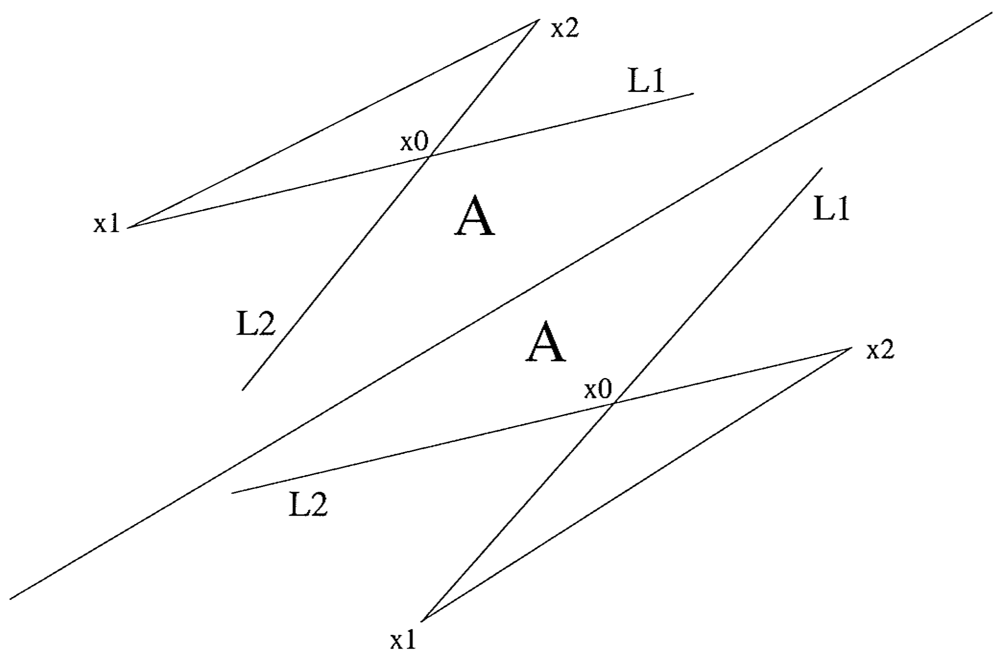

Figure B.15: The case where $x_{11} < x_{01} < x_{21}$ and $x_{12} < x_{02} < x_{22}$. These sub-cases contribute to the functions h_6 (top case) and h_5 (bottom case) in Theorem 2.4.3.

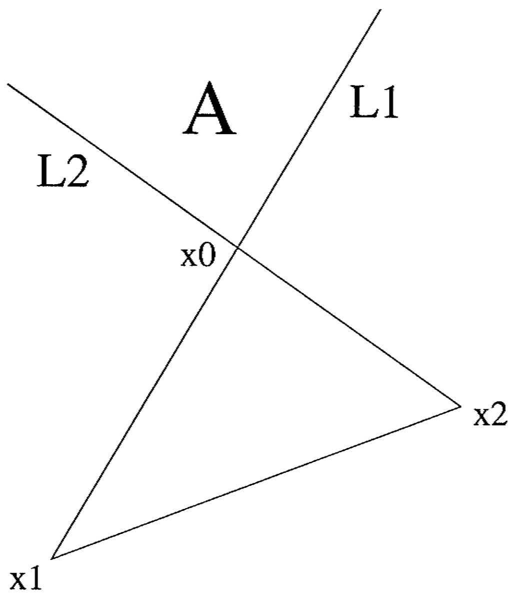

Figure B.16: The case where $x_{11} < x_{01} < x_{21}$ and $x_{12} < x_{22} < x_{02}$. This case contributes to the function h_4 in Theorem 2.4.3.

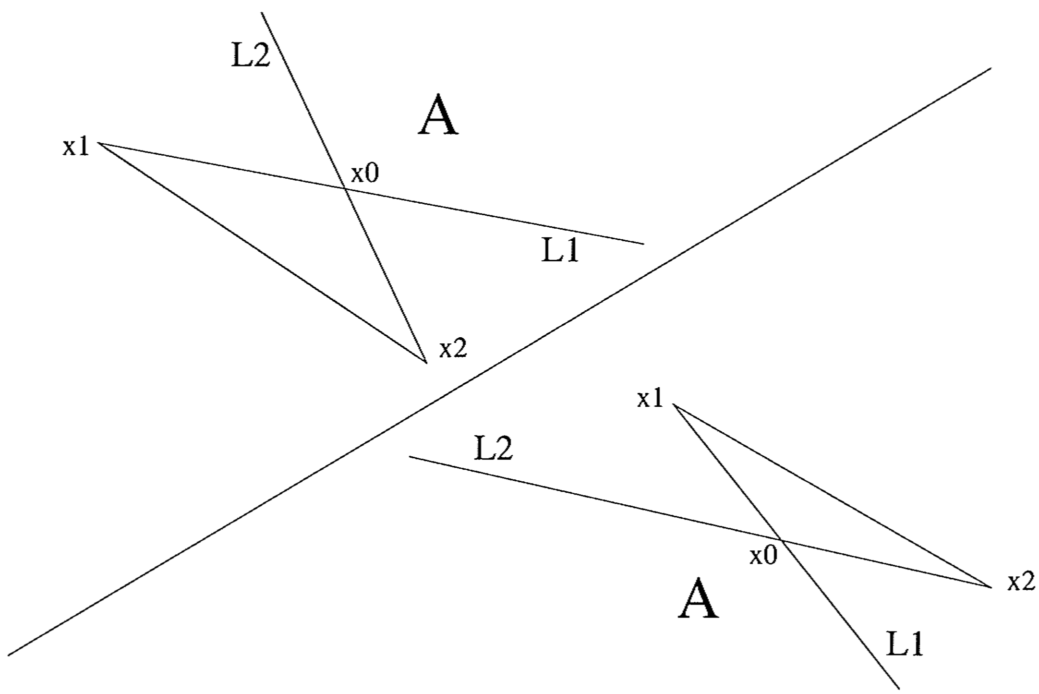

Figure B.17: The case where $x_{11} < x_{01} < x_{21}$ and $x_{22} < x_{02} < x_{12}$. These sub-cases contribute to the functions h_8 (top case) and h_7 (bottom case) in Theorem 2.4.3.

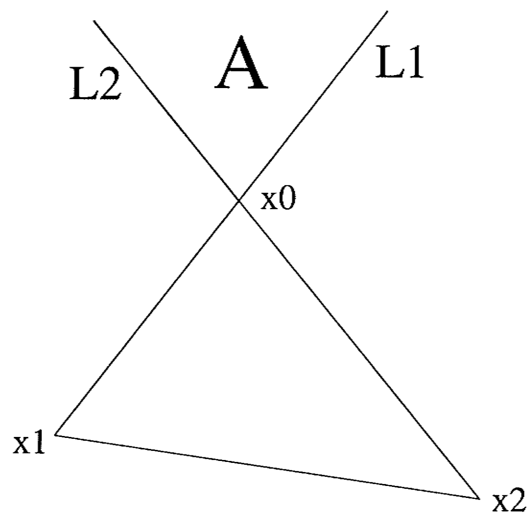

Figure B.18: The case where $x_{11} < x_{01} < x_{21}$ and $x_{22} < x_{12} < x_{02}$. This case contributes to the function h_4 in Theorem 2.4.3.

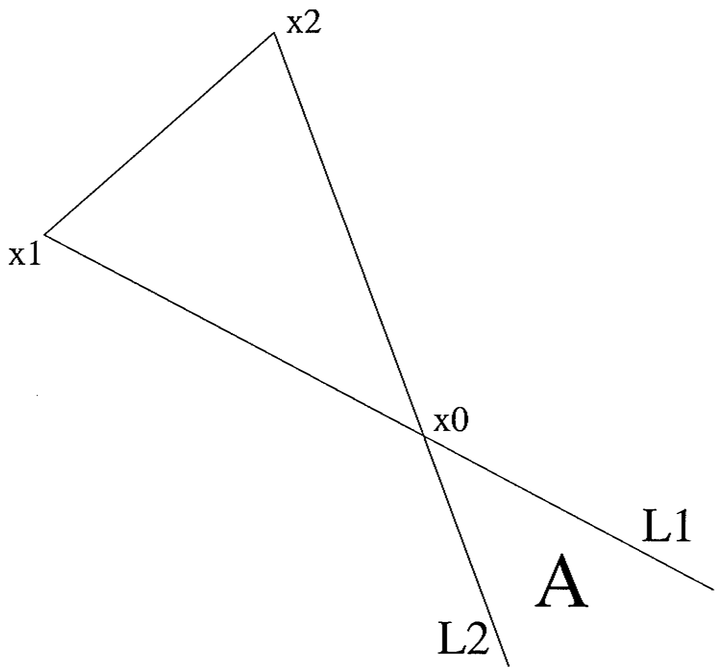

Figure B.19: The case where $x_{11} < x_{21} < x_{01}$ and $x_{02} < x_{12} < x_{22}$. This case contributes to the function h_2 in Theorem 2.4.3.

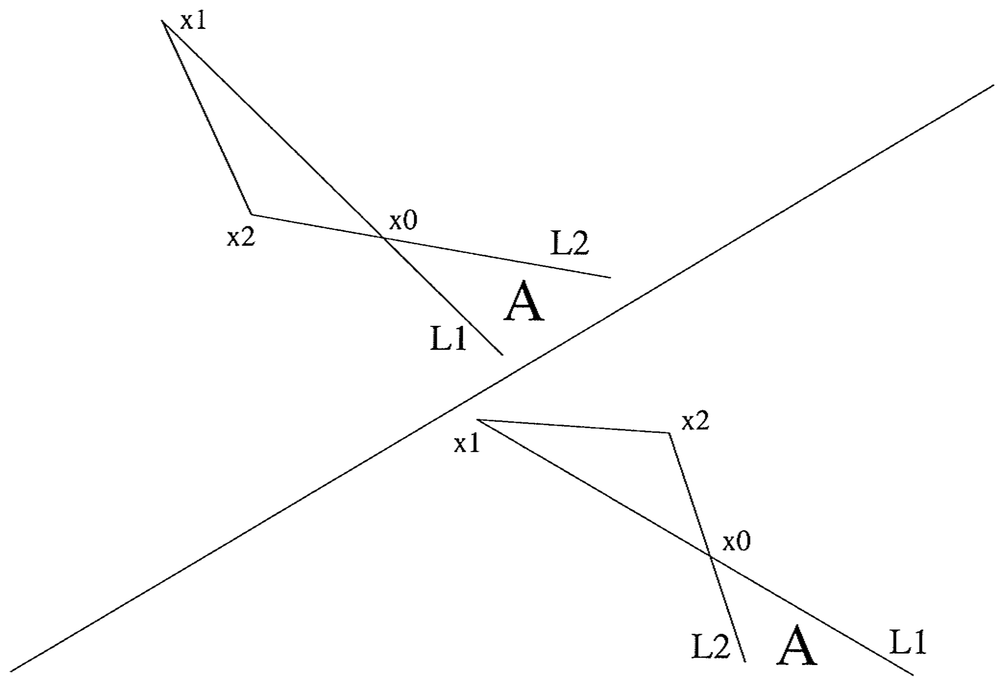

Figure B.20: The case where $x_{11} < x_{21} < x_{01}$ and $x_{02} < x_{22} < x_{12}$. Both sub-cases contribute to the function h_2 in Theorem 2.4.3.

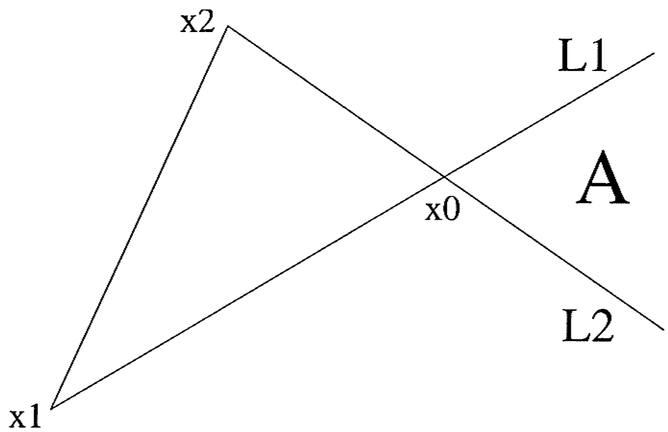

Figure B.21: The case where $x_{11} < x_{21} < x_{01}$ and $x_{12} < x_{02} < x_{22}$. This case contributes to the function h_2 in Theorem 2.4.3.

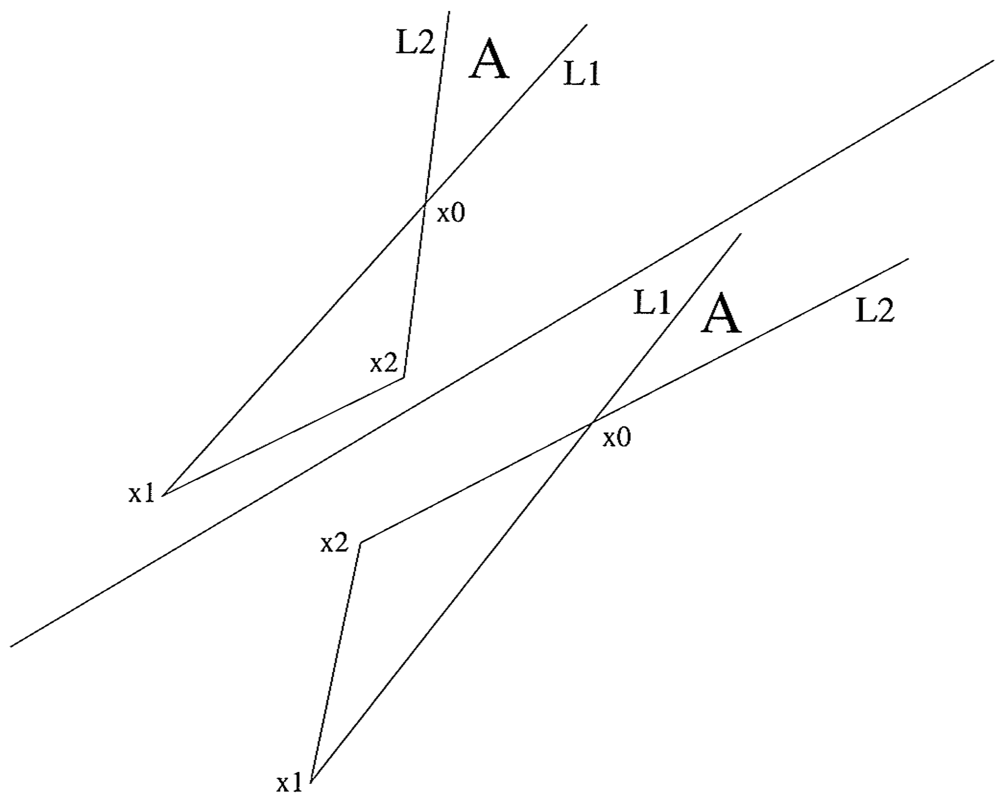

Figure B.22: The case where $x_{11} < x_{21} < x_{01}$ and $x_{12} < x_{22} < x_{02}$. Both sub-cases contribute to the function h_2 in Theorem 2.4.3.

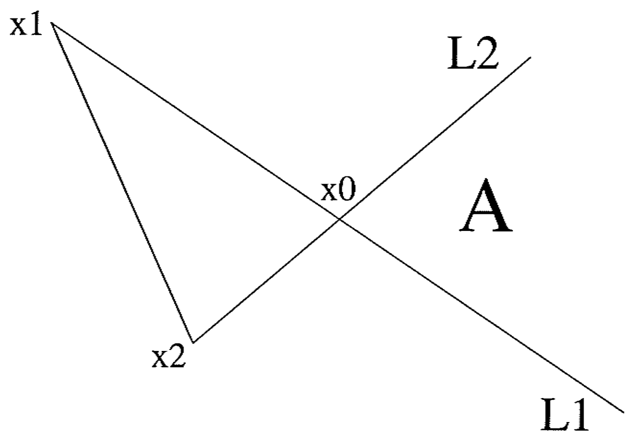

Figure B.23: The case where $x_{11} < x_{21} < x_{01}$ and $x_{22} < x_{02} < x_{12}$. This case contributes to the function h_2 in Theorem 2.4.3.

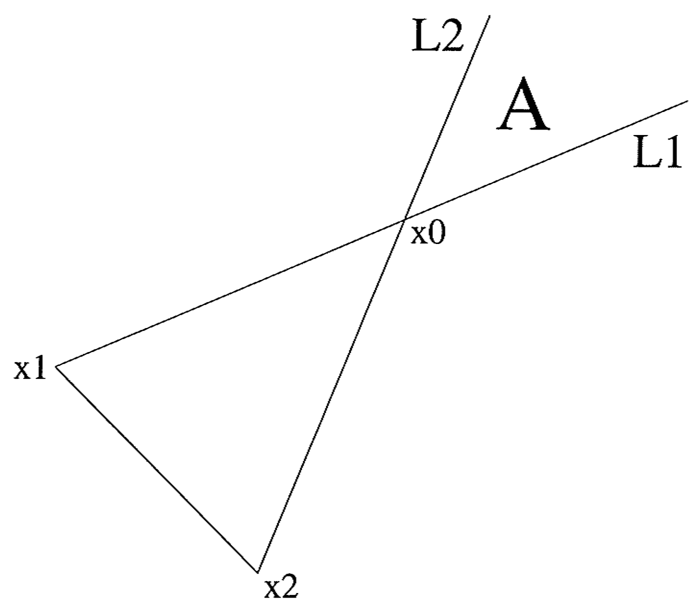

Figure B.24: The case where $x_{11} < x_{21} < x_{01}$ and $x_{22} < x_{12} < x_{02}$. This case contributes to the function h_2 in Theorem 2.4.3.

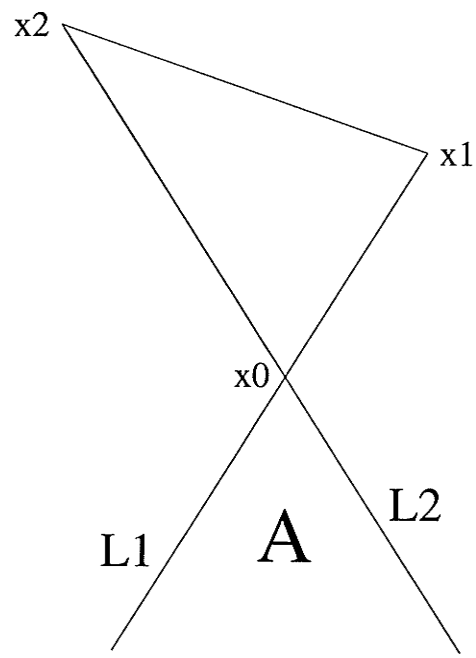

Figure B.25: The case where $x_{21} < x_{01} < x_{11}$ and $x_{02} < x_{12} < x_{22}$. This case contributes to the function h_3 in Theorem 2.4.3.

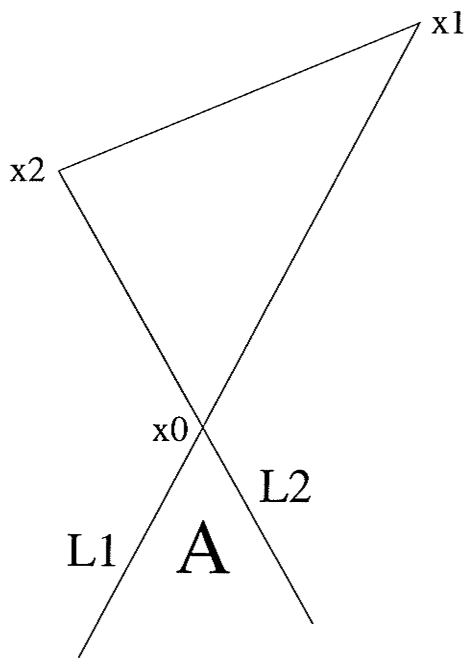

Figure B.26: The case where $x_{21} < x_{01} < x_{11}$ and $x_{02} < x_{22} < x_{12}$. This case contributes to the function h_3 in Theorem 2.4.3.

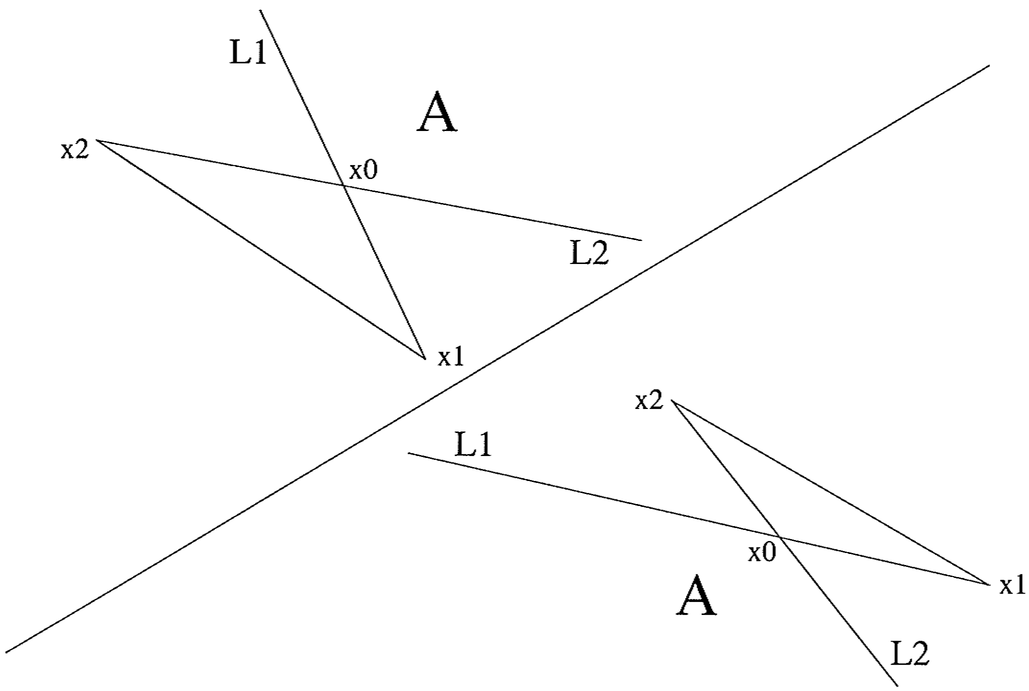

Figure B.27: The case where $x_{21} < x_{01} < x_{11}$ and $x_{12} < x_{02} < x_{22}$. These sub-case contribute to the functions h_9 (top case) and h_{10} (bottom case) in Theorem 2.4.3.

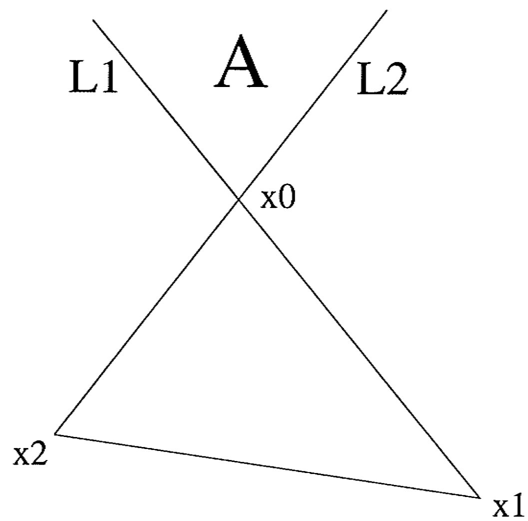

Figure B.28: The case where $x_{21} < x_{01} < x_{11}$ and $x_{12} < x_{22} < x_{02}$. This case contributes to the function h_4 in Theorem 2.4.3.

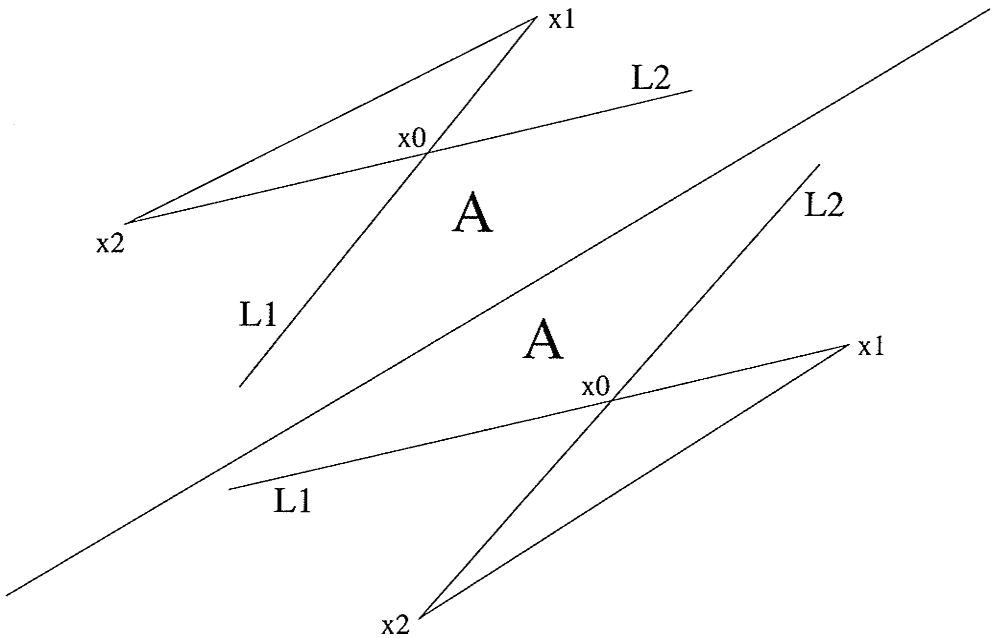

Figure B.29: The case where $x_{21} < x_{01} < x_{11}$ and $x_{22} < x_{02} < x_{12}$. These sub-cases contribute to the functions h_{11} (top case) and h_{12} (bottom case) in Theorem 2.4.3.

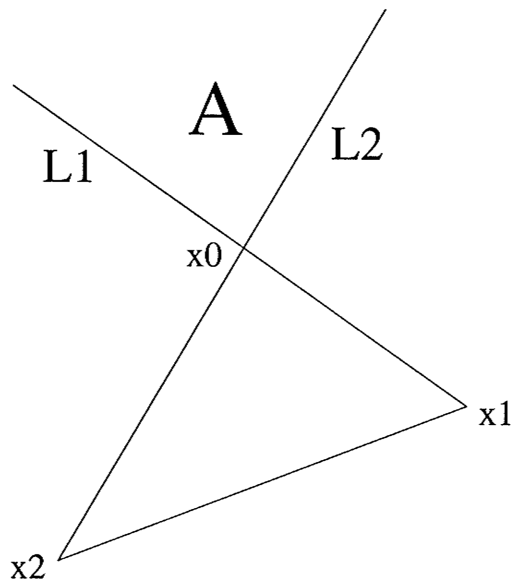

Figure B.30: The case where $x_{21} < x_{01} < x_{11}$ and $x_{22} < x_{12} < x_{02}$. This case contributes to the function h_4 in Theorem 2.4.3.

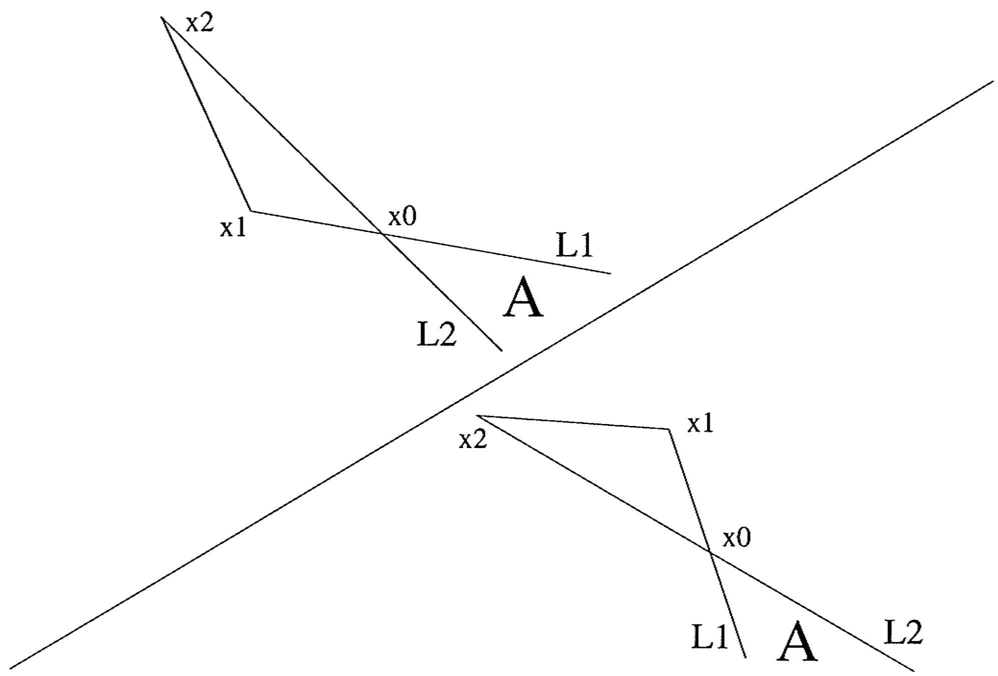

Figure B.31: The case where $x_{21} < x_{11} < x_{01}$ and $x_{02} < x_{12} < x_{22}$. Both sub-cases contribute to the function h_2 in Theorem 2.4.3.

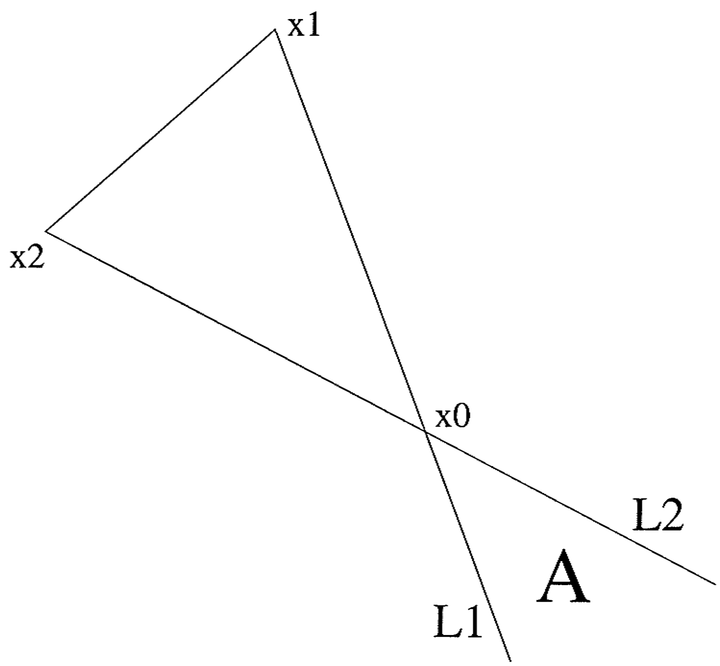

Figure B.32: The case where $x_{21} < x_{11} < x_{01}$ and $x_{02} < x_{22} < x_{12}$. This case contributes to the function h_2 in Theorem 2.4.3.

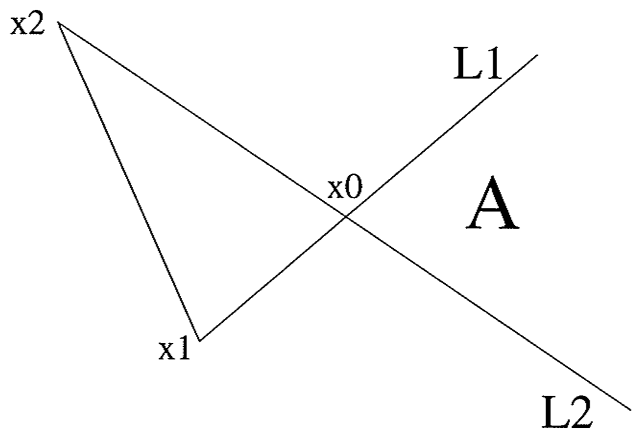

Figure B.33: The case where $x_{21} < x_{11} < x_{01}$ and $x_{12} < x_{02} < x_{22}$. This case contributes to the function h_2 in Theorem 2.4.3.

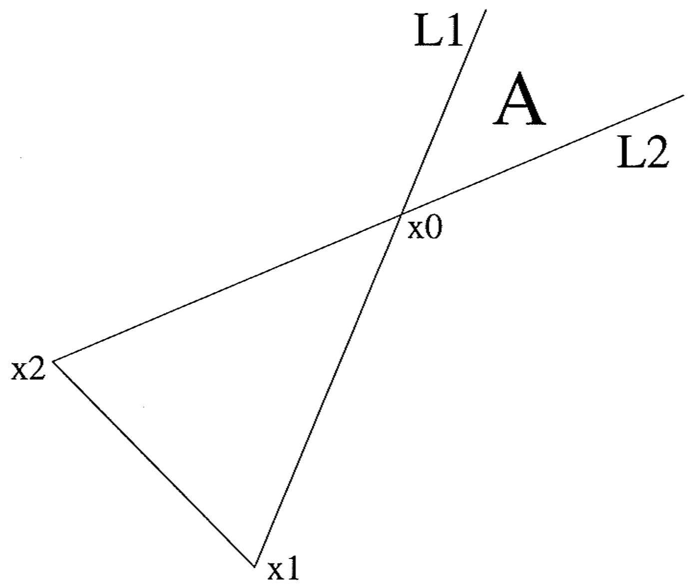

Figure B.34: The case where $x_{21} < x_{11} < x_{01}$ and $x_{12} < x_{22} < x_{02}$. This case contributes to the function h_2 in Theorem 2.4.3.

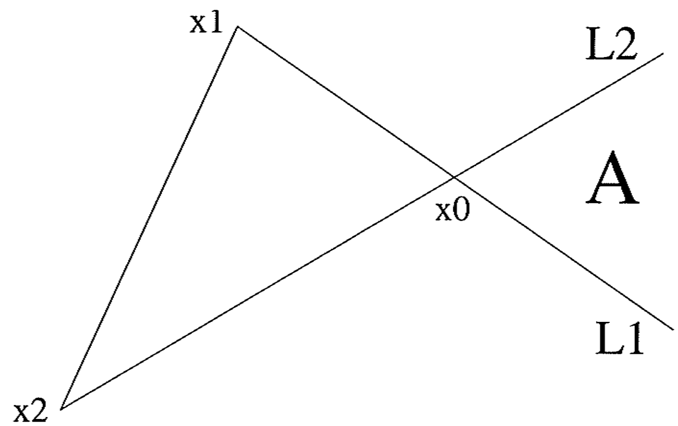

Figure B.35: The case where $x_{21} < x_{11} < x_{01}$ and $x_{22} < x_{02} < x_{12}$. This case contributes to the function h_2 in Theorem 2.4.3.

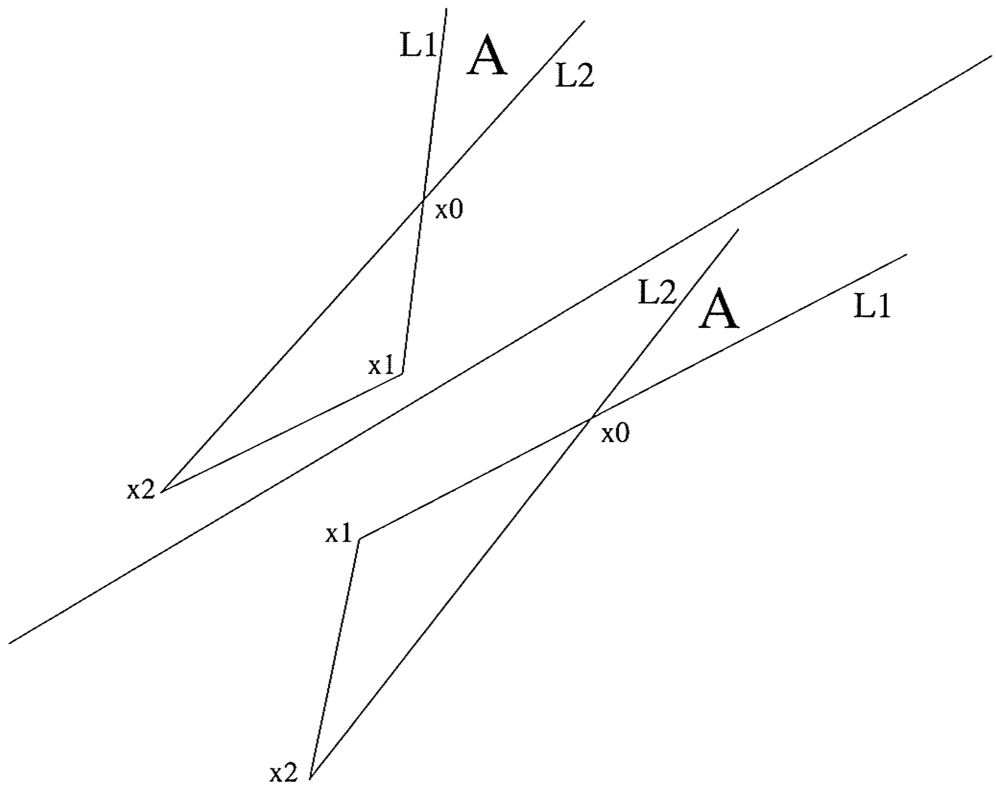

Figure B.36: The case where $x_{21} < x_{11} < x_{01}$ and $x_{22} < x_{12} < x_{02}$. Both sub-cases contribute to the function h_2 in Theorem 2.4.3.

Bibliography

[1] Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society* (Series A) **139**, 319-354.

[2] Billingsley, P. (1986). *Probability and Measure* (2nd edition), John Wiley and Sons.

[3] Coleman, D.E. (1997). "Individual Contributions" in: A discussion on statistically-based process monitoring and control, editors D.C. Montgomery and W.H. Woodall. *Journal of Quality Technology* **29**, 148-149.

[4] Eddy, W. (1982). Convex hull peeling. In *Compstat* (H. Caussinus et al., eds.), Physica 42-47.

[5] Efron, B. (1965). The convex hull of a random set of points. *Biometrika* **52**, 331-343.

[6] Fraiman, R. and Meloche, J. (1996). Multivariate L-estimation. *Test* **8**, 255-217.

[7] Gregory, G.G. (1977). Large sample theory for U-statistics and tests of fit. *Annals of Statistics* **5**, 110-123.

[8] Hodges, J.L. (1955). A bivariate sign test. *The Annals of Mathematical Statistics* **26**, 523-527.

[9] Hotelling, H. (1947). Multivariate Quality Control, in *Techniques of Statistical Analysis*, McGraw-Hill.

[10] Huber, P. (1972). Robust statistics: a review. *The Annals of Mathematical Statistics* **43**, 1041-1067.

[11] Jackson, J.E. and Morris, R.H. (1957). An application of MVQC to photographic processing. *Journal of the American Statistical Association* **52**, 186-199.

[12] Juran, J.M. (1992). *Juran on Quality by Design – The New Steps for Planning Quality into Goods and Services*, The Free Press.

[13] Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18**, 405-414.

[14] Liu, R.Y. (1995). Control Charts for Multivariate Processes. *Journal of the American Statistical Association* **90**, 1380-1387.

[15] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association* **88**, 252-260.

[16] Liu, R.Y., Parelius, J.M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Annals of Statistics* **27**, 783-858.

[17] Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proceedings of the National Academy of Science of India*, **12**, 49-55.

[18] Montgomery, D.C. (2001). *Introduction to Statistical Quality Control*, John Wiley and Sons.

[19] Moore, D. (1995). *The Basic Practice of Statistics*, W.H. Freeman and Company.

[20] Oja, H. (1983). Descriptive statistics for multivariate distribution. *Statistics and Probability Letters* **1**, 227-333.

[21] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons.

[22] Shewhart, W.A. (1931). *Economic Control of Quality of Manufactured Product*, Van Nostrand.

[23] Singh, K. (1991). Majority depth. Unpublished manuscript.

[24] Stoumbos, Z.G. and Jones, L.A. (2000). On the properties and design of individuals control charts based on simplicial depth. *Nonlinear Studies* **7**, 147-178.

[25] Tukey, J.W. (1979). Mathematics and picturing data. *Proceedings of the 1975 International Congress of Mathematics* **2**, 523-531.

[26] Zuo, Y. and Serfling, R. (2000). On the performance of some robust nonparametric location measures relative to a general notion of mul-

tivariate symmetry. *Journal of Statistical Planning and Inference* **84**, 55-79.