QUANTIFICATIONS OF RANDOM VARIABLES

XIAOSONG YAN

Quantifications of Random Variables

by

©Xiaosong Yan

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Statistics
Memorial University of Newfoundland

May 1996

St. John's Newfoundland Canada

0-612-36218-3

Abstract

Order restricted inference is an important field in statistical science. The utilization of ordering informations can increase the efficiency of statistical inference procedures in several senses, see Ayer, Brunk, Ewing, Reid, and Silverman (1955), Robertson and Wright (1974), Barlow and Ubhaya (1971), Lee (1981) and Kelly (1989).

In this thesis we review some basic theories about the least squares regressions, particularly the isotonic regressions. We give a simplified proof of an iterative procedure proposed by Dykstra (1983) for least squares problems.

We investigate the properties of the orderings of real-valued functions from several aspects. Some definitions are extended and their properties are generalized. We also show that the concept of closed convex cones and their duals is important in estimating procedures as well as in testing procedures. We demonstrate that some seemingly different problems have actually the same likelihood ratio test statistics and critical regions.

We observe that the orders of real-valued functions and the orders of random variables are closely related and statistical inference regarding these two orders behave similarly. A class of bivariate quantifications are defined based on these two orders. This bivariate notion has direct interpretation and appealing properties. More important, it characterizes a degree of positive dependence among random variables and therefore makes it possible to study the positive dependence of random variables by using the theories of the orders of real-valued functions and the orders of random variables.

We consider several estimation problems under order restrictions. We propose an algorithm that finds the nonparametric maximum likelihood estimates of a stochastically bounded survival function in finite steps, usually two or three steps. Simulation study shows that in general, utilizing the prior knowledge of a lower bound and an upper bound may reduce the point-wise MSE's and the amount of reduction in MSE's could be substantial for small and moderate sample sizes for a pair of sharp bounds. We obtain the estimates of a multinomial parameter under various order constraints for a general multinomial estimation procedure defined by Cressie and Read (1984).

We also consider the problem of simulating tail probabilities with a known stochastic bound. The proposed procedure may increase the efficiency of simulation significantly.

Acknowledgments

I am deeply indebted to my supervisor Dr. Chu-In Charles Lee. He has always been generous with his time and ideas. This work would not have been completed without his able guidance, constant encouragement, and financial support.

I am grateful to the School of Graduate Studies and the Department of Mathematics and Statistics, Memorial University of Newfoundland for supporting me with a Graduate Fellowship and a Teaching Assistantship.

I would like to thank the Department of Mathematics and Statistics, Memorial University of Newfoundland for providing me with a friendly atmosphere and necessary facilities to finish my program.

I would like to thank Drs. T. Robertson, F. T. Wright and V. Gadag who carefully read this thesis and made a number of valuable comments on improving its presentation.

Finally, I would like to express my sincere appreciation to my parents for their encouragement and care. I heartily thank my wife for her understanding and love.

Xiaosong Yan

St. John's

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Statistical inference under order restrictions is an important field in statistical science. Many types of problems are concerned with identifying meaningful structure in real world situations. Structure characterized by order restrictions arises in numerous settings and has many useful applications. For example, the failure rate of a component may increase as it ages; treatment responses may be stochastically dominated by a control. The books of Barlow, Bartholomew, Bremner and Brunk (1972), and Robertson, Wright and Dykstra (1988) are two classical monographs on this field and contain many important problems.

The utilization of ordering informations can increase the efficiency of statistical inference procedures in several senses. For example, 'isotonizing' estimates can reduce total square error (Ayer, Brunk, Ewing, Reid, and Silverman 1955) and maximum absolute error (Robertson and Wright 1974, and Barlow and Ubhaya 1971). Little was known about the pointwise properties of MLE's under order restrictions until Lee (1981) considered the problem of estimating linearly ordered normal means. He showed that in this case mean square error is reduced for every individual mean by using order restricted MLE's. An even stronger result for the same problem was obtained by Kelly (1989): the absolute error of each individual estimate under order restriction is stochastically smaller than that of the sample mean. Lee (1988) also observed that these pointwise properties do not hold, in general, for partial order restrictions.

A number of quantifications have appeared in literature to characterize the order of vectors in R^k. The well known isotonic regression arises from the maximum likelihood estimation of normal means under an isotonic restriction with respect to a quasi-order of the populations. Its usefulness is greatly enhanced by the fact that it solves a wide variety of restricted estimation problems in which the objective function may take many different forms other than the sum of squares. Its application include the maximum likelihood estimation of ordered normal variances, ordered binomial parameters (bioassay), ordered Poisson means, ordered multinomial parameters as well as a variety of problems from other areas, such as inventory theory and reliability analysis. In addition, the application of isotonic regression can be readily extended to some other important problems by the theory of Fenchel's duality. Other quantifications such as increasing on average and increasing on split average are also often considered in applications.

These quantifications of functions correspond to closed convex cones in a R^k space. The concept of duals of closed convex cones and the associated duality theorems in finite dimensional Euclidean space have proven to be useful in order restricted problems. Several authors have made extensive use of the concept of convex cones and their duals in R^k. Among these are Rockfellar (1970), Barlow and Brunk (1972), Robertson and Wright (1981), and Dykstra (1984). See Robertson, Wright and Dykstra (1988) for more details on the theory and applications of this subject.

The idea of ordering random variables with respect to the considered property is not very old. The (usual) stochastic ordering was first introduced by Mann and Whitney (1947) and Lehmann (1955). Since then many new notions have been introduced in the literature to characterize orders of random variables, such as the uniform stochastic ordering and the likelihood ratio ordering. The (usual) stochastic ordering, the uniform stochastic ordering and the likelihood ratio ordering are three of the most well studied orderings in the literature and can be expressed conveniently in terms of total positivity (TP) of probability functions. Stochastic orderings between random variables can arise in numerous settings and have many useful applications. For example, the simplest way of comparing two random variables is by comparing their means. However, such a comparison is based on only two single numbers (the means), and sometimes it is not very informative, especially in nonparametric statistical inference. Stochastic orders can also be used to deduce probability inequalities which are useful to obtain bounds for probabilities that are tedious to compute or analytically impossible to handle. For example, Lehmann (1959, P.112, Problem 11) showed that X is stochastically smaller than Y if and only if $Eu(X) \leq Eu(Y)$ for all increasing functions u. The reader is referred to the newly published book by Shaked and Shanthikumar (1994) for an overview on stochastic orderings and their applications.

Quantifications of real-valued functions and quantifications of random variable are closely related and statistical inferences with regard to these two classes of quantifications behave similarly. A class of bivariate quantifications are defined based on these two orders. This bivariate notion has direct interpretation and appealing properties. More important, it characterizes a degree of positive dependence among random variables and therefore makes it possible to study the positive dependence of random variables by using the theories of the orders of real-valued functions and the orders of random variables.

In Chapter 2 we introduce some basic results on least squares regression and particularly, the isotonic regression. We will introduce three algorithms that have been used extensively in studying and computing the isotonic regressions, namely, the pool-adjacent-violators algorithm, the minimum-lower-sets algorithm and the min-max formula. We will also give a simplified proof of the correct convergence of an iterative procedure which was first proposed by Dykstra and Robertson (1982a) for a matrix partial order and then extended by Dykstra (1983) and Dykstra and Boyle (1987) to a very general setting.

In Chapter 3 we extend the notions of orders of real-valued vectors in R^k space to real-valued functions in a measurable space and calculate the corresponding dual cones. We exhibit an important property of duality in the problem of hypothesis testing and demonstrate that some seemingly different problems have actually the same likelihood ratio test statistics and critical regions.

In Chapter 4 we introduce the order of random variables in terms of total positivity of probability functions. The definition of total positivity given in this chapter is an extension of the usual one and can be readily used to define the quantification of a sequence of random variables. We observe that the orders of real-valued functions and the orders of random variables are closely related and statistical inference regarding these two orders behave similarly.

In Chapter 5, we define the quantification of bivariate random variables based on the quantification of real-valued functions and the quantifications of random variables. We show that this quantification is closely related to the positive dependence of random variables which has important applications in reliability analysis, life sciences and many other fields. More specifically, we will show that the definitions of positive dependence of random variables in reliability analysis (Barlow and Proschan 1975) and positive associations for ordinal random variables (Agresti 1984 and Grove 1984) are special cases of our bivariate notions. But the bivariate notions defined in this chapter have direct interpretations and nice properties and the relations among them are readily revealed. In addition, it allows us to study the positive dependence of random variables by using the theories of quantifications of real-valued functions and random variables. Some aspects of estimation problem are also considered in this chapter.

The remaining chapters are some applications of the above theories.

In Chapter 6, we consider the problem of estimating a multinomial parameter under various ordering constraints for a general multinomial estimation procedure defined by Cressie and Read (1984).

In Chapter 7, we consider the problem of estimating a survival function that is stochastically bounded both from below and from above, with right-censored data. We extend the one-sided problems considered by Dykstra (1982) and propose an efficient iterative algorithm to find bounded estimates in finite steps, usually two or three steps. The proposed algorithm is an iterative procedure such that at each step one needs only to solve several non-overlapping one-sided problems. An example involving survival times for heart transplant patients which appeared in Crowley and Hu (1977) is given to illustrate the proposed algorithm. We also conduct a simulation study to investigate the increase in efficiency obtained by using the stochastically bounded constraints. Simulation study shows that in general, utilizing the prior knowledge of a lower bound and an upper bound may reduce the point-wise MSE's and the amount of reduction in MSE's could be substantial for small and moderate sample sizes for a pair of sharp bounds.

In Chapter 8, we consider the problem of simulating tail probabilities with a known stochastic bound. The proposed procedure may increase the efficiency of simulation significantly.

Chapter 2

Isotonic Regression and Least Squares Problems

2.1 Introduction

Isotonic regression problem arises from the maximum likelihood estimation of normal means under an order restriction and it plays a very important role in the order restricted inference. Its usefulness is greatly enhanced by the fact that it solves a wide variety of restricted estimation problems in which the objective function may take many different forms other than the sum of squares. Its application includes maximum likelihood estimation of ordered normal variances, ordered binomial parameters (bioassay), ordered Poisson means, ordered multinomial parameters as well as a variety of problems from other areas, such as reliability theory and density estimation, (cf. sec 1.5 of Robertson, Wright and Dykstra (1988)). In addition, solutions to many other optimization problems can be expressed in terms of the isotonic regression, see Barlow and Brunk (1972), Dykstra and Lee (1991), and Dykstra, Lee and Yan (1995).

The application of isotonic regression can be readily extended to some other important problems by the theory of Fenchel's duality. Duality is an important concept in order restricted inference. For one thing, it provides an alternative approach to a problem that may be more tractable, or provides additional insight into the problem. It is also possible to use duality concepts to expand the collection of problems for which one has solutions. The reader is referred to Robertson, Wright and Dykstra (1988) for more details on this subject.

The problem of developing algorithms for the isotonic regressions has received a great deal of attention, see Barlow *et al.* (1972). In fact, isotonic regression is a quadratic programming problem and there is an extensive literature on the methods of computing solutions. The problem of computing the isotonic regression is a special case and a number of efficient algorithms have been proposed.

The most widely used algorithm for a simple order is the pool-adjacent-violators algorithm (PAVA) first published by Ayer, Brunk, Ewing, Reid and Silverman (1955). PAVA is a very efficient algorithm but it does not apply in general to partially ordered isotonic regression. For general partially ordered isotonic regression the most well known algorithm is the minimum-lower-sets algorithm of Brunk (1955). Several other algorithms have been developed for quasi or partial orders to increase the efficiency of the computation, such as the minimum violator algorithm due to Thompson (1962), an algorithm due to Eeden (1958) and its improvement due to Gebhardt (1970), and the min-max algorithm due to Lee (1983), among others.

An iterative algorithm for the matrix partial order is developed by Dykstra and Robertson (1982a). This type of iterative algorithm has been extended to a large number of restricted optimization problems by Dykstra (1983) and Dykstra and

Boyle (1987).

In Section 2.2 we first review some concepts and preliminary results of the least squares regressions. Concepts of quasi-orders and isotonic regressions are given in Section 2.3. In Section 2.4 we introduce three extensively used algorithms for isotonic regressions, namely PAVA, minimum-low-sets algorithm and the min-max formula. Most of the contents of Section 2.3 and 2.4 can be found in Robertson, Wright and Dykstra (1988). In section 2.5 we give a simplified proof of the correct convergence of the iterative procedure proposed by Dykstra and Boyle (1987) for a general least squares problem.

2.2 Basic Concepts and Least Squares Regression

2.2.1 Convex Sets, Cones and Dual Cones

Let R^k be a k-dimensional Euclidean space with the inner product defined by

$$(f, g) = \sum_{i=1}^{k} f_i g_i w_i, \quad \forall f, g \in R^k, \tag{2.1}$$

where $w = (w_1, \ldots, w_k)$ is a vector of weights such that $w_i > 0$, $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$.

A subset C of R^k is said to be *convex set* if $(1 - \lambda)f + \lambda g \in C$ whenever $f \in C, g \in C$ and $0 \leq \lambda \leq 1$. It is well known that the intersection of an arbitrary number of convex sets is still convex.

A subset $\mathcal{C}$ of R^k is called a *cone* if it is closed under nonnegative scalar multiplication, i.e. $\lambda f \in \mathcal{C}$ when $f \in \mathcal{C}$ and $\lambda > 0$. Note that a cone is not necessarily "pointed." For example, subspaces of R^k are special cones. So are the open and closed half-spaces corresponding to a hyperplane containing the origin.

For a convex cone C, the subset C^* of R^k defined by

$$C^* = \{g \in R^k : (g, f) = \sum_{i=1}^{m} g_i f_i w_i \leq 0, \quad \forall f \in C\}. \tag{2.2}$$

is called the *Fenchel dual* or *polar* of C. In particular, if $C = S$ is a subspace of R^k, then

$$S^* = S^\perp = \{g \in R^k : (g, f) = 0, \forall f \in S\}. \tag{2.3}$$

It can be shown that C^* is also a convex cone and furthermore, it is closed. For any two subsets A, B of R^k, denote $A + B$ the direct sum of sets A, B, i.e., $A + B = \{f + g | f \in A,\ g \in B\}$. Let C, C_1 and C_2 be convex cones. We have the following results,

$$\text{(a)}\ \ C \subset (C^*)^*, \text{ and } C = (C^*)^* \ \text{ if } C \text{ is closed;} \tag{2.4}$$

$$\text{(b)}\ \ (-C)^* = -C^*; \tag{2.5}$$

$$\text{(c)}\ \ C_1^* \subset C_2^* \ \text{ if } C_1 \supset C_2 \tag{2.6}$$

$$\text{(d)}\ \ (C_1 + C_2)^* = C_1^* \cap C_2^*; \tag{2.7}$$

$$\text{(e)}\ \ (C_1 \cap C_2)^* = C_1^* + C_2^* \ \text{ if the latter is closed,} \tag{2.8}$$

see Rockafellar (1970, p.146).

2.2.2 Least Squares Regression and Projection

In the least squares regression, we are interested in the problem of

$$\text{Minimize}_{\,f \in C} \sum_{i=1}^{k} (g_i - f_i)^2 w_i. \tag{2.9}$$

where $g \in R^k$ is a given vector and $C \subset R^k$ is a closed convex set. The solution to the problem (2.9) exists and it is unique. This unique solution, denoted by $E(g|C)$, is called the least squares projection of **g** onto C.

Brunk (1965) showed that a vector $g^* \in C$ is the solution if and only if

$$(g - g^*, f - g^*) \leq 0, \quad \forall f \in C. \tag{2.10}$$

Furthermore, if $C = \mathcal{C}$ is a closed convex cone, then $g^* \in \mathcal{C}$ is the solution if and only if

$$(g - g^*, g^*) = 0, \tag{2.11}$$

and

$$(g - g^*, f) \leq 0, \quad \forall f \in \mathcal{C}. \tag{2.12}$$

Barlow and Brunk (1972) showed that

$$E(g|\mathcal{C}) + E(g|\mathcal{C}^*) = g. \tag{2.13}$$

It follows that

$$(E(g|\mathcal{C}), E(g|\mathcal{C}^*)) = 0. \tag{2.14}$$

An affine transformation of a set C by $\alpha \in R^m$ is defined to be the set

$$C + \alpha = \{f + \alpha : f \in C\}. \tag{2.15}$$

Lemma 2.2.1 *Let $C \subset R^k$ be a closed convex set. Then*

$$E(g \mid C + \alpha) = \alpha + E(g - \alpha \mid C). \tag{2.16}$$

Proof. Let $g^* = E(g|C + \alpha)$. By (2.10), $\forall f \in C$,

$$\begin{aligned} & (g - g^*, (f + \alpha) - g^*) \\ = \; & ((g - \alpha) - (g^* - \alpha), f - (g^* - \alpha)) \leq 0. \end{aligned}$$

Therefore, by (2.10) again, $g^* - \alpha = E(g - \alpha|C)$. The proof is complete. $\square$

Corollary 2.2.1 *Let $\mathcal{C} \subset R^k$ be a closed convex cone. Then*

$$E(g \mid \mathcal{C} + \alpha) = g - E(g - \alpha|\mathcal{C}^*). \tag{2.17}$$

Proof. By (2.16) and (2.13),

$$\begin{aligned} E(g|\mathcal{C} + \alpha) &= \alpha + E(g - \alpha|\mathcal{C}) \\ &= \alpha + (g - \alpha) - E(g - \alpha|\mathcal{C}^*) \\ &= g - E(g - \alpha|\mathcal{C}^*). \end{aligned}$$

□

Let $g^* = E(g \mid \mathcal{C} + \alpha)$, where $\mathcal{C}$ is a closed convex cone. By (2.16), $g^* - \alpha = E(g - \alpha|\mathcal{C}) \in \mathcal{C}$. By (2.17), $g - g^* = E(g - \alpha|\mathcal{C}^*)$. It follows from (2.14) that

$$(g^* - \alpha, g - g^*) = 0, \tag{2.18}$$

and

$$(g^* - \alpha, x) \leq 0, \quad \forall\, x \in \mathcal{C}^*. \tag{2.19}$$

2.3 Quasi-order of Finite Sets and Isotonic Regressions

2.3.1 Quasi-order

Let X be a finite set $\{x_1, x_2, \ldots, x_k\}$. A binary relation $\prec$ on X is a *simple order* if

1. it is *reflexive*: $x \prec x$ for $x \in X$;

2. it is *transitive*: $x, y, z \in X$, $x \prec y$ and $y \prec z$ imply $x \prec z$;

3. it is *antisymmetric*: $x, y \in X$, $x \prec y$ and $y \prec x$ imply $x = y$;

4. it is *comparable*: $x, y \in X$ implies that either $x \prec y$ or $y \prec x$.

A binary relation $\prec$ on X is called a *partial order* if it is reflexive, transitive, and antisymmetric, but there may be noncomparable elements. A binary relation $\prec$ is called a *quasi-order* if it is reflexive and transitive, but it need not be antisymmetric and it may admit noncomparable elements. A partial order usually arises when vector comparison are involved. The following examples are some partial orders that are frequently encountered in applications.

Example 2.2.1 (*Simple order*): $x_1 \prec x_2 \prec \ldots \prec x_k$.

Example 2.2.2 (*Simple loop order*): $x_0 \prec x_i \prec x_{k+1}$, $i = 1, 2, \ldots, k$.

Example 2.2.3 (*Simple tree order*): $x_0 \prec x_i$, $i = 1, 2, \ldots, k$.

Example 2.2.4 (*Umbrella order*): $x_1 \prec x_2 \prec \cdots \prec x_{i_0} \succ x_{i_0+1} \succ \cdots \succ x_k$.

Simple order is one of the most important orders and has many useful applications. This will be evident throughout this article. The simple tree order, the simple loop order and the umbrella order are three partial orders that have found many useful applications. The simple tree order is a subset of the simple loop order. These two orders arise in sampling situations where one wishes to compare several treatments with one or two extreme controls. For example, in a drug analysis, several drugs may be compared to a zero-dose control and a most effective but expensive drug control. The umbrella order is closely related to the unimodal property and has found useful applications in estimating density functions, (see Robertson *et al.* 1988 for some more details on this subject).

2.3.2 Isotonic Regression

A real-valued function, f, on X is said to be *isotonic* with respect to the quasi-ordering $\prec$ on X if $x, y \in X$ and $x \prec y$ imply $f(x) \le f(y)$.

Let g be a given function on X and w a given positive weight function on X. An isotonic function g^* on X is called an *isotonic regression* of g with weight w if it minimizes

$$\sum_{x \in X} [g(x) - f(x)]^2 w(x)$$

in the class of all isotonic functions on X.

A real-valued function on a finite set X can be considered as a point of a Euclidean space which has as its dimension the number of points in X. In this setting, the collection, $\mathcal{I}$, of all isotonic functions on X with respect to a given quasi-order is a closed convex cone and the isotonic regression g^* is the closest point of $\mathcal{I}$ to g with distance induced by the inner product

$$(f, g) = \sum_{i=1}^{k} f_i g_i w_i.$$

The existence and uniqueness then follow from the general theory of least squares problem described earlier in this chapter.

2.3.3 Properties of Isotonic Regression

The isotonic regression has a number of important properties. Some of them are given below.

Theorem 2.3.1 *Suppose g_1 and g_2 are isotonic functions on X such that $g_1(x) \le g(x) \le g_2(x)$ for all $x \in X$, and if g^* is an isotonic regression of g, then also*

$g_1(x) \le g^*(x) \le g_2(x)$ for all $x \in X$. In particular, if a and b are constants such that $a \le g(x) \le b$ for all $x \in X$, then also $a \le g^*(x) \le b$ for $x \in X$. (Th. 1.3.4 RWD)

Suppose g and w are functions on X, set

$$\mathrm{Av}(A) = \frac{\sum_{x \in A} w(x)g(x)}{\sum_{x \in A} w(x)}$$

for those A nonempty subsets of X. While $\mathrm{Av}(A)$ depends on g, this will not be made explicit in the notation. Let $[g^* = c]$ denote $\{x \in X : g^*(x) = c\}$.

Theorem 2.3.2 *If c is any real number and if the set $[g^* = c]$ is nonempty then $c = Av([g^* = c])$. (Th 1.3.5, RWD)*

Theorem 2.3.3 *For an arbitrary real-valued function, Ψ, defined on the reals,*

$$(g - g^*, \Psi(g^*)) = 0.$$

(Th 1.3.6 RWD)

Theorem 2.3.2 reduces the problem of computing g^* to finding the sets on which g^* is constant (i.e. its *level sets*). There are a number of algorithms in computing isotonic regressions and we will introduce three of them in the next section that have been extensively used, namely the *pool-adjacent-violators algorithm* (PAVA), the *minimum-lower-sets algorithm* and the *min-max formula*.

2.4 Algorithms for Isotonic Regression

2.4.1 Pool-Adjacent-Violators Algorithm for the Simple Order

Let X be a finite set $\{x_1, x_2, \ldots, x_k\}$ with a simple order $x_1 \prec x_2 \prec \ldots \prec x_k$. Then a real valued function f on X is isotonic if and only if $f(x_1) \leq f(x_2) \leq \ldots \leq f(x_k)$. Let g be a given function on X and w a given positive weight function on X. By definition, the isotonic regression of g is an isotonic function that minimizes in the class of isotonic functions f on X the sum of squares

$$\sum_{x \in X} [g(x) - f(x)]^2 w(x).$$

The PAVA starts with g. If g is isotonic, then $g^* = g$. Otherwise, there must exist an index i such that $g(x_{i-1}) > g(x_i)$. These two values are then replaced by their weighted average, namely $\mathrm{Av}(\{i-1, i\})$ and the two weights $w(x_{i-1})$ and $w(x_i)$ are replaced by $w(x_{i-1}) + w(x_i)$. If this new set of $k-1$ values is isotonic, then $g^*(x_{i-1}) = g^*(x_i) = \mathrm{Av}(\{i-1, i\})$ and $g^*(x_j) = g(x_j)$ otherwise. If this new set is not isotonic then this process is repeated using the new values and weights until an isotonic set of values is obtained.

2.4.2 Max-min Formulas

Let $\prec$ be a given quasi-order on X. A subset L of X is called a *lower set* with respect to the quasi-order $\prec$ if $y \in L$ and $x \prec y$ imply $x \in L$. A subset U of X is called an upper set if $x \in U$ and $x \prec y$ imply $y \in U$. We denote the class of all lower sets by $\mathcal{L}$ and the class of all the upper sets by $\mathcal{U}$. A subset B of X is called a *level set* if there exists a lower set L and an upper set U such that $B = L \cap U$.

Theorem 2.4.1 *The isotonic regression of g is given by*

$$g^*(x) = \max_{U:x\in U} \min_{L:x\in L} Av(L\cap U) \tag{2.20}$$

$$= \min_{L:x\in L} \max_{U:x\in U} Av(L\cap U) \tag{2.21}$$

(Th. 1.4.4 RWD).

For illustration, let us consider the simple order defined on X by $x_1 \prec x_2 \prec \cdots \prec x_k$. The nonempty lower sets are of the form $\{x_1, x_2, \ldots, x_i\}$; $i = 1, 2, \ldots, k$, and the nonempty upper sets are of the form $\{x_i, x_{i+1}, \ldots, x_k\}$; $i = 1, \ldots, k$. For the simple order, the max-min formula can be expressed by

$$g^*(x_i) = \max_{j\le i} \min_{h:h\ge i} Av(\{x_j, x_{j+1}, \ldots, x_h\}) \tag{2.22}$$

$$= \min_{h\ge i} \max_{j\le i} Av(\{x_j, x_{j+1}, \ldots, x_h\}). \tag{2.23}$$

2.4.3 The Minimum-lower-sets Algorithm

Let B_1 denote the union of all lower sets of minimum average. B_1 is the level set on which g^* assumes its smallest value:

$$g^*(x) = Av(B_1) = \min\{Av(L) : L \in \mathcal{L}\} \quad \text{for } x \in B_1.$$

Now consider the averages of level sets of the form $L \cap B_1^c$, level sets consisting of lower sets with B_1 subtracted. Select again the union of these level sets of minimum average, say B_2. The level set B_2 is the set on which g^* assumes its next smallest value:

$$g^*(x) = Av(B_2) = \min\{Av(L \cap B_1^c) : L \in \mathcal{L}\} \quad \text{for } x \in B_2.$$

This process is continued until X is exhausted.

2.5 A Proof of the Convergence of Dykstra's Algorithm for Restricted Least Squares Regression

Many important least squares problems can be expressed as

$$\text{Minimize}_{\ x\in\cap_1^r(K_i+\alpha_i)} \parallel g - x \parallel . \tag{2.24}$$

where $K_1, K_2, \ldots, K_r$ are closed convex cones in R^m and $\alpha_1, \alpha_2, \ldots, \alpha_r \in R^m$. Dykstra and Robertson (1982a) and Dykstra (1983) proposed an iterative method for the case $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$, and showed that their procedure converges correctly. Later, Dykstra and Boyle (1987) extended this algorithm for arbitrary α_i's and showed that the procedure also converges to the desired solution as long as the feasible set is nonempty. In this section we consider the same problem of Dykstra and Boyle (1987) and give a simpler proof of the convergence of the algorithm. First, we rewrite their procedure as following.

Step 0. Initial settings: let $g_{0,r} = g$, $I_{0,i} = 0$, $i = 1, 2, ..., r$, and $n = 1$.

Step 1. Compute

$$\begin{aligned} g_{n,1} &= E(g_{n-1,r} - I_{n-1,1} | K_1 + \alpha_1), \\ I_{n,1} &= g_{n,1} - (g_{n-1,r} - I_{n-1,1}). \end{aligned} \tag{2.25}$$

$$\begin{aligned} g_{n,s} &= E(g_{n,s-1} - I_{n-1,s} | K_s + \alpha_s), \\ I_{n,s} &= g_{n,s} - (g_{n,s-1} - I_{n-1,s}), \\ & \quad s = 2, 3, ..., r. \end{aligned} \tag{2.26}$$

Step 2. Replace n by $n+1$ and go to step 1.

Note that as we cyclically project onto one of the convex sets, the last increment for that set is removed prior to that projection and a new increment for that set is always formed. It follows from the algorithm that

$$g_{n,s} = g + \sum_{t=1}^{s} I_{n,t} + \sum_{t=s+1}^{r} I_{n-1,t}, \quad s = 1, 2, \ldots, r, \tag{2.27}$$

where the second summand is 0 if $s = r$. The utility of the algorithm is based on the following theorem which has been proved by Dykstra and Boyle's (1987). The following is a simplified proof of the same theorem. The difference between the two proofs is that we will use directly the basic property (2.10) of least squares regressions to show the correct convergence of each convergent subsequence while Dykstra and Boyle's proof is not so straightforward.

Theorem 2.5.1 *If* $\cap_{i=1}^{r}(K_i + \alpha_i) \neq \emptyset$, *then*

$$\lim_{n\to\infty} g_{(n,l)} = E(g | \cap_1^r (K_i + \alpha_i))$$

for every $l = 1, 2, \ldots, r$.

Proof. Since $K_i + \alpha_i$ are closed convex sets, so is the nonempty set $\cap_{s=1}^{r}(K_i + \alpha_i)$. Therefore the projection of g onto $\cap_{s=1}^{r}(K_i + \alpha_i)$ exists, say g^* and this projection is unique. Note that the key relationships:

$$g_{n,i-1} - g_{n,i} = I_{n-1,i} - I_{n,i}, \quad i = 2, \ldots, r, \tag{2.28}$$

$$g_{n-1,r} - g_{n,1} = I_{n-1,1} - I_{n,1}, \tag{2.29}$$

hold among the projections and increments. It follows that

$$\| g_{n,i-1} - g^* \|^2 = \| (g_{n,i} - g^*) + (I_{n-1,i} - I_{n,i}) \|^2$$

$$= \| g_{n,i} - g^* \|^2 + \| I_{n-1,i} - I_{n,i} \|^2 + 2(g^* - \alpha_i, I_{n,i} - I_{n-1,i}) + 2(g_{n,i} - \alpha_i, I_{n-1,i} - I_{n,i})$$

for $i \geq 2$. Note that the last term is nonnegative since by (2.11) and (2.12), $(g_{n,i} - \alpha_i, I_{n,i}) = 0$ and $(g_{n,i} - \alpha_i, I_{n-1,i}) \geq 0$. It follows that

$$\| g_{n,i-1} - g^* \|^2 \geq \| g_{n,i} - g^* \|^2 + \| I_{n-1,i} - I_{n,i} \|^2 + 2(g^* - \alpha_i, I_{n,i} - I_{n-1,i})$$

for $i \geq 2$. In a similar fashion,

$$\| g_{n-1,r} - g^* \|^2 \geq \| g_{n,1} - g^* \|^2 + \| I_{n-1,1} - I_{n,1} \|^2 + 2(g^* - \alpha_1, I_{n,1} - I_{n-1,1}).$$

Noting the "telescoping property" of the term $(g^* - \alpha_i, I_{n,i} - I_{n-1,i})$, we may write

$$\| g - g^* \|^2 \geq \| g_{n,r} - g^* \|^2 + \sum_{k=1}^{n} \sum_{l=1}^{r} \| I_{k-1,l} - I_{k,l} \|^2 + 2 \sum_{l=1}^{r} (g^* - \alpha_l, I_{n,l}). \quad (2.30)$$

Since $g^* - \alpha_i \in K_i$ and $I_{n,i} \in -K_i^*$, the last term is nonnegative. Therefore,

$$\sum_{k=1}^{\infty} \sum_{l=1}^{r} \| I_{k-1,l} - I_{k,l} \|^2 < \infty. \quad (2.31)$$

Thus,

$$\| I_{n-1,l} - I_{n,l} \| = \| g_{n,l-1} - g_{n,l} \| \to 0, \quad (l \geq 2)$$

and

$$\| I_{n-1,1} - I_{n,1} \| = \| g_{n-1,r} - g_{n,1} \| \to 0, \quad \text{as } n \to \infty. \quad (2.32)$$

By (2.30), $g_{n,r}$ are uniformly bounded. So there exists a convergent subsequence, say $g_{n_j,r}$ with the limit h. Of course, by (2.32), $g_{n_j,i}$ also converges to h for each $i = 1, 2, \ldots, r-1$. By the closeness of $K_i + \alpha_i$, $h \in \cap_1^r (K_i + \alpha_i)$. Now, for any

$x \in \cap_1^r (K_i + \alpha_i)$,

$$\begin{aligned}
(g - h, x - h) &= \lim_{j\to\infty} (g - g_{n_j,r}, x - h) \\
&= \lim_{j\to\infty} (-\sum_{l=1}^{r} I_{n_j,l}, x - h) \\
&= -\sum_{l=1}^{r} \lim_{j\to\infty} (I_{n_j,l}, x - g_{n_j,l}) \\
&= -\sum_{l=1}^{r} \lim_{j\to\infty} (I_{n_j,l}, x - \alpha_l) + \sum_{l=1}^{r} \lim_{j\to\infty} (I_{n_j,l}, g_{n_j,l} - \alpha_l). \quad (2.33)
\end{aligned}$$

The first term of (2.33) is nonnegative because $x - \alpha_l \in K_l$ and $I_{n_j,l} \in -K_l^*$ by (2.25). The last term of (2.33) is zero by (2.25) and (2.18). It follows that $(g - h, h - x) \leq 0$. Thus by (2.10), $h = g^*$. By symmetry, one can show that any convergent subsequence $\{g_{n_i,s}\}$, $s = 1, 2, \ldots, r$ will have the same limit g^*. The proof is complete. □

Chapter 3

Quantifications of Real-Valued Functions and Their Duals

3.1 Introduction

In the previous chapter we have reviewed some basic results regarding isotonic regression. While isotonicity is one of the most important quantifications for real-valued functions, a number of other quantifications are also of great importance both in theory and application. In this chapter we will introduce some notions that are closely related to the notions for random variables in the next chapter.

These quantifications of functions correspond to convex cones in l_2 or L_2 space . The concept of convex cones and their duals in finite dimensional Euclidean space has proven to be useful in order restricted problems. Several authors have made extensive use of the concept of convex cones and their duals in R^k. Among these are Rockfellar (1970), Barlow and Brunk (1972), Robertson and Wright (1981), and Dykstra (1984). In this chapter we investigate the dual cones of quantifications of general real-valued functions. We also consider some applications of the concept of dual cones in the problems of hypothesis testings.

3.2 Quantifications of Discrete Functions

3.2.1 Definition and Properties

We first consider the simple case of discrete functions on a countable set of real numbers X. Without loss of generality, we assume that $X = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Let $w(\cdot)$ be a given nonnegative function on X such that $\sum_{x \in X} w(x) < \infty$.

Definition 3.2.1 *Let f be a real-valued function on X such that $\sum_{x \in X} |f(x)| w(x) < +\infty$. The f is said to be (with respect to the weight function w)*

monotonic increasing, or in the order of $\ll_m$, if

$$f(x) \le f(y), \quad \text{for any } x < y \in X \text{ with } w(x) > 0, w(y) > 0;$$

increasing on left average, or in the order of $\ll_{(-)}$, if

$$\frac{\sum_{i \le x} w(i) f(i)}{\sum_{i \le x} w(i)} \le \frac{\sum_{i \le y} w(i) f(i)}{\sum_{i \le y} w(i)}, \quad \text{for any } x < y \in X \text{ with } \sum_{i \le x} w(i) > 0;$$

increasing on right average, or in the order of $\ll_{(+)}$, if

$$\frac{\sum_{i > x} w(i) f(i)}{\sum_{i > x} w(i)} \le \frac{\sum_{i > y} w(i) f(i)}{\sum_{i > y} w(i)}, \quad \text{for any } x < y \in X \text{ with } \sum_{i > y} w(i) > 0;$$

increasing on split average, or in the order of $\ll_s$, if

$$\frac{\sum_{i \le x} w(i) f(i)}{\sum_{i \le x} w(i)} \le \frac{\sum_{i > x} w(i) f(i)}{\sum_{i > x} w(i)}, \quad \text{for any } x \in X \text{ with } \sum_{i \le x} w(i), \sum_{i > x} w(x) > 0.$$

There are several equivalence properties for the order of $\ll_{(-)}$ and $\ll_{(+)}$ that can be conveniently used in applications.

Lemma 3.2.1 *The following statements are equivalent, (assuming the denominators are not zero):*

(a) $f(x)$ *is in the order of* $\ll_{(-)}$;

(b) $\frac{\sum_{i\leq x_1} f(i)w(i)}{\sum_{i\leq x_1} w(i)} \leq \frac{\sum_{x_1<i\leq x_2} f(i)w(i)}{\sum_{x_1<i\leq x_2} w(i)}$, *for any* $x_1 < x_2 \in X$;

(c) $\frac{\sum_{i\leq x} f(i)w(i)}{\sum_{i\leq x} w(i)} \leq \frac{\sum_{i\leq x+1} f(i)w(i)}{\sum_{i\leq x+1} w(i)}$, *for any* $x \in X$;

(d) $\frac{\sum_{i\leq x} f(i)w(i)}{\sum_{i\leq x} w(i)} \leq f(x+1)$, *for any* $x \in X$ *and* $w(x+1) > 0$;

(e) $\frac{\sum_{i\leq x} f(i)w(i)}{\sum_{i\leq x} w(i)} \leq f(x)$, *for any* $x \in X$ *and* $w(x) > 0$.

Proof. The equivalence of (a) and (b) and the equivalence of (c), (d), (e) are straight forward. We now prove the equivalence of (a) and (c). It is trivial that (a) implies (c). Conversely, suppose that (c) is true. Then for any x_1, we have

$$\frac{\sum_{i\leq x_1} f(i)w(i)}{\sum_{i\leq x_1} w(i)} \leq \frac{\sum_{i\leq x_1+1} f(i)w(i)}{\sum_{i\leq x_1+1} w(i)}, \quad \text{for any } x_1 \in X. \tag{3.1}$$

By induction, one obtains

$$\frac{\sum_{i\leq x_1} f(i)w(i)}{\sum_{i\leq x_1} w(i)} \leq \frac{\sum_{i\leq x_2} f(i)w(i)}{\sum_{i\leq x_2} w(i)}, \quad \text{for any } x_1 \leq x_2 \in X, \tag{3.2}$$

i.e., $f(x)$ is in the order of $\ll_{(-)}$. The proof is complete. □

It is easy to obtain the following analogs of the result regarding the order $\ll_{(+)}$.

Lemma 3.2.2 *The following statements are equivalent, (assuming the denominators are not zero):*

(a) $f(x)$ *is in the order of* $\ll_{(+)}$;

(b) $\frac{\sum_{x_1<i\leq x_2} f(i)w(i)}{\sum_{x_1<i\leq x_2} w(i)} \leq \frac{\sum_{i>x_2} f(i)w(i)}{\sum_{i>x_2} w(i)}$ *for any* $x_1 < x_2 \in X$;

(c) $\frac{\sum_{i\geq x} f(i)w(i)}{\sum_{i>x} w(i)} \leq \frac{\sum_{i\geq x+1} f(i)w(i)}{\sum_{i>x+1} w(i)}$ *for any* $x \in X$;

(d) $f(x) \leq \frac{\sum_{i\geq x} f(i)w(i)}{\sum_{i>x} w(i)}$ *for any* $x \in X$ *and* $w(x) > 0$;

(e) $f(x) \leq \frac{\sum_{i\geq x} f(i)}{\sum_{i\geq x} w(i)}$ *for any* $x \in X$ *and* $w(x) > 0$.

If there are finitely many positive $w(i)$, then the notions in Definition 3.2.1 are reduced to the quantification of vectors in R^k which have been well studied in the literature. The relationship among these orders are given by the following theorem.

Theorem 3.2.1 *(a)* $\ll_m$ *implies* $\ll_{(-)}$ *and* $\ll_{(+)}$; *(b)* $\ll_{(-)}$ *or* $\ll_{(+)}$ *implies* $\ll_s$.

Proof. By symmetry, it suffices to prove that (a) $\ll_m$ implies $\ll_{(-)}$, and (b) $\ll_{(-)}$ implies $\ll_s$.

(a) Suppose $f(x)$ is in the order of $\ll_m$, then by definition, $f(i) \leq f(x)$, for all $i \leq x \in X$, $w(i), w(x) > 0$. So

$$f(i)w(i) \leq f(x)w(i), \quad \text{for all } i \leq x. \tag{3.3}$$

By summing both sides of (3.3) with respect i over $i \leq x$, one obtains

$$\frac{\sum_{i\leq x} f(i)w(i)}{\sum_{i\leq x} w(i)} \leq f(x). \tag{3.4}$$

By Lemma 3.2.1, $f(x)$ is in the order of $\ll_{(-)}$.

(b) Suppose $f(x)$ is in the order of $\ll_{(-)}$. By Lemma 3.2.1, we have

$$\frac{\sum_{i\leq x_1} f(i)w(i)}{\sum_{i\leq x_1} w(i)} \leq \frac{\sum_{x_1<i\leq x_2} f(i)w(i)}{\sum_{x_1<i\leq x_2} w(i)}. \quad \text{for any } x_1 \leq x_2. \tag{3.5}$$

By letting $x_2 \to +\infty$ in (3.5) one obtains that $f(x)$ is in the order of $\ll_s$. □

3.2.2 Dual Cones of Quantifications of Discrete Functions

In the following discussion we will consider functions on X that lie in the space

$$l_2 = \{f : \sum_{x\in X} f^2(x)w(x) < +\infty\}.$$

For any two functions $f(\cdot)$ and $g(\cdot) \in l_2$, we define their inner product by

$$(f, g)_w = \sum_{x\in X} f(x)g(x)w(x).$$

Then by the Cauchy-Schwarz inequality, we have

$$|(f, g)| \leq \sum_{x\in X} |f(x)g(x)|w(x) \leq (\sum_{x\in X} f^2(x)w(x))^{1/2}(\sum_{x\in X} g^2(x)w(x))^{1/2} < +\infty,$$

for any $f, g \in l_2$. The corresponding norm of $f \in l_2$ is defined as

$$\| f \| = (f, f)^{1/2} = (\sum_{x\in X} f^2(x)w(x))^{1/2}.$$

The dual of a convex cone A in l_2 is defined to be the set

$$A^* = \{g \in l_2 : \sum_{x\in X} f(x)g(x)w(x) \leq 0, \text{ for all } f \in A\}. \tag{3.6}$$

Let

$$\begin{aligned}
A_m &= \{f \in l_2 : f \text{ is in the order of } \ll_m\};\\
A_{(-)} &= \{f \in l_2 : f \text{ is in the order of } \ll_{(-)}\};\\
A_{(+)} &= \{f \in l_2 : f \text{ is in the order of } \ll_{(+)}\};\\
A_s &= \{f \in l_2 : f \text{ is in the order of } \ll_s\}.
\end{aligned}$$

Lemma 3.2.3 *A_m, $A_{(+)}$, $A_{(-)}$, and A_s are convex cones in l_2.*

Proof. We shall first prove that A_m is a convex cone. Suppose $f, g \in A_m$. Then for any $x < y \in X$ with $w_x > 0$ and $w_y > 0$, we have $f(x) \leq f(y)$ and $g(x) \leq g(y)$. Therefore,

$$\alpha f(x) \leq \alpha f(y)$$

for any $\alpha \geq 0$ and

$$f(x) + g(x) \leq f(y) + g(y).$$

It follows that A_m is a convex cone. By a similar argument one can prove that $A_{(-)}$, $A_{(+)}$ and A_s are also convex cone. □

We shall next find the duals of those convex cones. When there are only a finite number of $w(i)$ such that $w(i) > 0$, our problem is equivalent to the ones in R^k spaces which have been studied by Barlow and Brunk (1972) and Dykstra (1984). Define

$$\mathcal{S} = \{f \in l_2 : \sum_{x \in X} f(x)w(x) = 0.\} \tag{3.7}$$

Lemma 3.2.4 *If $A \subset l_2$ is a convex cone that contains all the constant functions, then $A^* \subset \mathcal{S}$.*

Proof. Suppose $g \in A^*$. Then for any $f \in A$, we have $(g, f) \leq 0$. Since the constant functions with values 1 and -1 are in A_m, it follows that $g \in \mathcal{S}$. □

Theorem 3.2.2 *(a) $A_m^* = (-A_s) \cap \mathcal{S}$; (b) $A_s^* = (-A_m) \cap \mathcal{S}$.*

Proof. (a) We first prove that $A_m^\bullet \subset (-A_s) \cap \mathcal{S}$. Let $g \in A_m^\bullet$. Then

$$(f, g) \leq 0 \quad \text{for any } f \in A_m. \tag{3.8}$$

Particularly, for $x \in X$, define

$$e_x(y) = \begin{cases} -1/\sum_{i \leq x} w(i), & \text{if } y \leq x, \\ 1/\sum_{i > x} w(i), & \text{if } y > x. \end{cases}$$

It is clear that $e_x \in A_m$ and hence

$$(g, e_x) = -\frac{\sum_{i \leq x} g(i)w(i)}{\sum_{i \leq x} w(i)} + \frac{\sum_{i > x} g(i)w(i)}{\sum_{i > x} w(i)} \leq 0, \quad \text{for all } x \in X.$$

It follows that $g \in -A_s$. In addition, since A_m contains all the constant functions, by Lemma 3.2.4, $A_m^\bullet \subset (-A_s) \cap \mathcal{S}$.

Conversely, suppose $g \in (-A_s) \cap \mathcal{S}$. For a given $f \in A_m$, define $f^+(x) = f(x) \vee 0$ and $f^-(x) = f(x) \wedge 0$. Clearly, $f = f^+ + f^-$ and $f^+, f^- \in A_m$. Now, for any $x \in X$,

$$\begin{aligned} & \sum_{i \geq x} f^+(i)g(i)w(i) \\ = & \sum_{i \geq x} [\sum_{x \leq j \leq i} (f^+(j) - f^+(j-1)) + f^+(x-1)]g(i)w(i) \\ = & \sum_{j \geq x} [\sum_{i \geq j} g(i)w(i)](f^+(j) - f^+(j-1)) + f^+(x-1) \sum_{i \geq x} g(i)w(i). \end{aligned}$$

Since $g \in (-A_s) \cap \mathcal{S}$, we have $\sum_{i \geq j} g(i)w(i) \leq 0$ and therefore,

$$\sum_{i \geq x} f^+(i)g(i)w(i) \leq f^+(x-1) \sum_{i \geq x} g(i)w(i).$$

By taking the limit $x \to -\infty$ one obtains $(f^+, g) \leq 0$. By symmetry, one obtains $(f^-, g) \leq 0$. It follows that $(f, g) = (f^+, g) + (f^-, g) \leq 0$ and so $g \in A_m^\bullet$. Therefore, $(-A_s) \cap \mathcal{S} \subset A_m^\bullet$. It follows that $A_m^\bullet = (-A_s) \cap \mathcal{S}$.

We now prove (b). For any pairs $x, y \in X$ such that $x < y$, $w(x) > 0$ and $w(y) > 0$, define $e_{xy}(x) = -1$, $e_{xy}(y) = 1$, and $e_{xy}(z) = 0$ if $z \neq= x, y$. Clearly, $e_{xy} \in A_s$. If $g \in A_s^*$, then

$$(e_{xy}, g) = -g(x) + g(y) \leq 0,$$

and hence $g \in -A_m$. In addition, since A_s contains all the constant functions, by Lemma 3.2.4, $A_s^* \subset \mathcal{S}$. It follows that $A_s^* \subset (-A_m) \cap \mathcal{S}$.

On the other hand, by (a), (2.4) and (2.8), we have

$$A_m \subset (A_m^*)^* = ((-A_s) \cap \mathcal{S})^* = (-A_s^*) + \mathcal{S}^*.$$

Consequently,

$$A_m \cap \mathcal{S} \subset ((-A_s^*) + \mathcal{S}^*) \cap \mathcal{S} = (-A_s^*) \cap \mathcal{S} = -A_s^*.$$

Therefore, $(-A_m) \cap \mathcal{S} \subset A_s^*$. It follows that $A_s^* = (-A_m) \cap \mathcal{S}$. The proof is complete. □

Corollary 3.2.1 *(a)* $A_m = -A_s^* + \mathcal{S}^\perp$; *(b)* $A_s = -A_m^* + \mathcal{S}^\perp$.

Proof. It is trivial that $A_m = A_m \cap \mathcal{S} + \mathcal{S}^\perp$ and $A_s = A_s \cap \mathcal{S} + \mathcal{S}^\perp$. By Theorem 3.2.2, the proof is complete. □

Corollary 3.2.2 *If* $\sum_{x \in X} w(x) = 1$, *then for any* $f \in A_m$, $g \in A_s$,

$$\sum_{x \in X} f(x)g(x)w(x) \geq \sum_{x \in X} f(x)w(x) \sum_{x \in X} g(x)w(x).$$

Proof. Since $g \in A_s$, we have $g - \bar{g} \in A_s \cap \mathcal{S}$, where $\bar{g}(x) = \sum_{i \in X} g(i)w(i)$ for all $x \in X$. Therefore, by Theorem 3.2.2,

$$(f, g - \bar{g}) = \sum_{x \in X} f(x)g(x)w(x) - \sum_{x \in X} f(x)w(x) \sum_{x \in X} g(x)w(x) \geq 0.$$

□

The cones of $A_{(-)}$ and $A_{(+)}$ are closely related to the positive orthant. For convenience, we shall assume that $w(x) > 0$ for all $x \in X$ and define

$$e_x^{(-)}(y) = \begin{cases} -1/W(x), & \text{if } y \leq x; \\ 1/w(x+1), & \text{if } y = x+1; \\ 0 & \text{if } y > x+1 \end{cases} \tag{3.9}$$

and

$$e_x^{(+)}(y) = \begin{cases} 1/(W(+\infty) - W(x-1)), & \text{if } y \geq x; \\ -1/w(x-1), & \text{if } y = x-1; \\ 0 & \text{if } y < x-1 \end{cases} \tag{3.10}$$

Lemma 3.2.5 *(a) $e_x^{(-)} \in A_{(-)}$ for all $x \in X$ and $(e_x^{(-)}, e_y^{(-)}) = 0$ if $x \neq y$.*

(b) $e_x^{(+)} \in A_{(+)}$ for all $x \in X$ and $(e_x^{(+)}, e_y^{(+)}) = 0$ if $x \neq y$.

Lemma 3.2.6 *For any real-valued function $f \in l_2$, we have*

$$f = \bar{f} + \sum_{i \in X} \frac{(f, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} e_i^{(-)}. \tag{3.11}$$

$$= \bar{f} + \sum_{i \in X} \frac{(f, e_i^{(+)})}{(e_i^{(+)}, e_i^{(+)})} e_i^{(+)}, \tag{3.12}$$

where $\bar{f} = \sum_{i \in X} f(i)w(i) / \sum_{i \in X} w(i)$.

Proof. It suffices to prove (3.11). For any $x \in X$,

$$\bar{f} + \sum_{i \in X} \frac{(f, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} e_i^{(-)}(x) \tag{3.13}$$

$$= \bar{f} + \frac{(f, e^{(-)}_{x-1})}{(e^{(-)}_{x-1}, e^{(-)}_{x-1})} e^{(-)}_{x-1}(x) + \sum_{i \geq x} \frac{(f, e^{(-)}_i)}{(e^{(-)}_i, e^{(-)}_i)} e^{(-)}_i(x).$$

$$= \bar{f} + \left[- \sum_{j \leq x-1} f(j)w(j)/W(x-1) + f(x) \right] \frac{W(x-1)}{W(x)}$$

$$- \sum_{i \geq x} \left[- \sum_{j \leq i} f(j)w(j)/W(i) + f(i+1) \right] \frac{w(i+1)}{W(i+1)}. \qquad (3.14)$$

Since $\sum_{j \leq i} f(j)w(j) = \bar{f} \sum_{j \in X} w(j) - \sum_{j > i} f(j)w(j)$, we have

$$\left[- \sum_{j \leq x-1} f(j)w(j)/W(x-1) + f(x) \right] \frac{W(x-1)}{W(x)}$$

$$= \left\{ [\sum_{j \geq x} f(j)w(j) - \bar{f} \sum_{j \in X} w(j)]/W(x-1) + f(x) \right\} \frac{W(x-1)}{W(x)}$$

$$= f(x) - \bar{f} \sum_{j \in X} w(j)/W(x) + \frac{1}{W(x)} \sum_{j \geq x+1} f(j)w(j) \qquad (3.15)$$

and

$$\sum_{i \geq x} \left\{ - \sum_{j \leq i} f(j)w(j)/W(i) + f(i+1) \right\} \frac{w(i+1)}{W(i+1)}$$

$$= \sum_{i \geq x} \left\{ [\sum_{j \geq i+1} f(j)w(j) - \bar{f} \sum_{j \in X} w(j)]/W(i) + f(i+1) \right\} \frac{w(i+1)}{W(i+1)}$$

$$= -[\bar{f} \sum_{j \in X} w(j)] \sum_{i \geq x} w(i)/W(i)W(i+1) +$$

$$\sum_{i \geq x} \left[\sum_{j \geq i+2} \frac{w(i+1)}{W(i)W(i+1)} f(j)w(j) + \frac{w(i+1)}{W(i)} f(i+1) \right] \qquad (3.16)$$

It is trivial that the first term of (3.16) is equal to

$$-\bar{f}(\sum_{j \in X} w(j)/W(x) - 1). \qquad (3.17)$$

The second term of (3.16) can be written as

$$\sum_{j \geq x+2} \left[\sum_{x \leq i \leq j-2} \frac{w(i+1)}{W(i)W(i+1)} \right] f(j)w(j) + \sum_{i \geq x} \frac{w(i+1)}{W(i)} f(i+1)$$

$$= \sum_{j\geq x+2}\left[\frac{1}{W(x)} - \frac{1}{W(j-1)}\right] f(j)w(j) + \sum_{i\geq x}\frac{w(i+1)}{W(i)}f(i+1)$$
$$= \frac{1}{W(x)}\sum_{j\geq x+2} f(j)w(j) - \sum_{j\geq x+2}\frac{w(j)}{W(j-1)}f(j) + \sum_{i\geq x}\frac{w(i+1)}{W(i)}f(i+1)$$
$$= \frac{1}{W(x)}\sum_{j\geq x+1} f(j)w(j). \tag{3.18}$$

By combining equations (3.14) to (3.18), one obtains (3.11). □

Theorem 3.2.3

$$A_{(-)} = \{f \in l_2 : f = c + \sum_{i\in X} a_i e_i^{(-)} \quad , a_i \geq 0\}; \tag{3.19}$$
$$A_{(+)} = \{f \in l_2 : f = c + \sum_{i\in X} a_i e_i^{(+)} \quad , a_i \geq 0\} \tag{3.20}$$

where c is a constant function in l_2.

Proof. It suffices to prove (3.19). Denote

$$A = \{f \in l_2 : f = c + \sum_{i\in X} a_i e_i^{(-)}, \quad a_i \geq 0\}.$$

By the fact that $f \in l_2$, each component of f is a limit of an absolute convergent series. Since $A_{(-)}$ is a convex cone which contains all the real constant functions and $e_i^{(-)} \in A_{(-)}$, we have $A_{(-)} \supset A$. Conversely, by Lemma 3.10, for any $f \in A_{(-)}$,

$$f = \bar{f} + \sum_{i\in X}\frac{(f, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})}e_i^{(-)}.$$

It can be shown that $(f, e_i^{(-)}) \geq 0$ and hence, $A_{(-)} \subset A$. The proof is complete. □

Theorem 3.2.4 *(a) $A^*_{(-)} = (-A_{(-)}) \cap \mathcal{S}$; (b) $A^*_{(+)} = (-A_{(+)}) \cap \mathcal{S}$.*

Proof. By symmetry, it suffices to show (a). We shall first prove $A^{*}_{(-)} \subset (-A_{(-)}) \cap \mathcal{S}$. Suppose $g \in A^{*}_{(-)}$, then

$$(f, g) \leq 0 \quad \text{for any } f \in A_{(-)}. \tag{3.21}$$

Particularly, since $e_x \in A_{(-)}$, we have

$$(g, e_x) = -\frac{\sum_{i \leq x} g(i) w(i)}{\sum_{i \leq x} w(i)} + g(x+1) \leq 0, \quad \text{for all } x \in X.$$

By Lemma 3.2.1, $g \in -A_{(-)}$. In addition, since $A_{(-)}$ contains all the constant functions, by Lemma 3.2.4, $A^{*}_{(-)} \subset (-A_{(-)}) \cap \mathcal{S}$.

Conversely, suppose $g \in (-A_{(-)}) \cap \mathcal{S}$. Since $\bar{g} = 0$, by Lemma 3.2.6,

$$g = \sum_{i \in X} \frac{(g, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} e_i^{(-)}.$$

It is trivial that $(g, e_i^{(-)}) \leq 0$ and $(f, e_i^{(-)}) \geq 0$ for any $f \in A_{(-)}$. Therefore, $(g, f) \leq 0$ and hence, $g \in A^{*}_{(-)}$. Consequently, $(-A_{(-)}) \cap \mathcal{S} \subset A^{*}_{(-)}$. The proof is complete. □

The proof of the following result is similar to that of Corollary 3.2.2.

Corollary 3.2.3 *(a)* $A_{(-)} = -A^{*}_{(-)} + \mathcal{S}^{\perp}$; *(b)* $A_{(+)} = -A^{*}_{(+)} + \mathcal{S}^{\perp}$.

Corollary 3.2.4 *If* $\sum_{x \in X} w(x) = 1$, *then for any* $f, g \in A_{(-)}$ *(or* $A_{(+)}$*),*

$$\sum_{x \in X} f(x) g(x) w(x) \geq \left(\sum_{x \in X} f(x) w(x) \right) \left(\sum_{x \in X} g(x) w(x) \right).$$

3.2.3 Quantifications of Functions in a Restricted Space

In applications it is not uncommon that the functions of interest are restricted to some boundary constraint. Let K be a closed convex cone in l_2 and S is a subspace in l_2. By (2.8) and (2.3) one obtains

$$(K \cap S)^* = K^* + S^\perp \tag{3.22}$$

and conversely,

$$K^* = (K \cap S)^* \cap S, \quad \text{if } K^* \subset S. \tag{3.23}$$

Particularly, if

$$S = \mathcal{S} = \{f : \sum_{i \in X} f(i)w(i) = 0\}, \tag{3.24}$$

then

$$\mathcal{S}^* = \mathcal{S}^\perp = \{f : f(i) = f(i+1), \forall i \in X\}, \tag{3.25}$$

i.e., $\mathcal{S}^\perp$ is the subspace of all constant functions in l_2. In this case the two cones (3.22) can be easily obtained from one to another.

Theorem 3.2.5 *Let $\mathcal{S}$ be defined by (3.24). Then*

(a) $(A_m \cap \mathcal{S})^* = -A_s$;

(b) $(A_{(-)} \cap \mathcal{S})^* = -A_{(-)}$;

(c) $(A_{(+)} \cap \mathcal{S})^* = -A_{(+)}$;

(d) $(A_s \cap \mathcal{S})^* = -A_m$.

Proof. We first prove (a). By (3.22), Theorem 3.2.2, and Corollary 3.2.1,

$$\begin{aligned}(A_m \cap \mathcal{S})^* &= A_m^* + \mathcal{S}^\perp \\ &= ((-A_s) \cap \mathcal{S}) + \mathcal{S}^\perp \\ &= -A_s.\end{aligned}$$

By similar arguments, one can prove (b), (c) and (d). □

3.3 Algorithms for Convex Projections

We shall now consider the problem of

$$\min_{f \in \mathcal{C}} \| g - f \|^2$$

where $\mathcal{C} = A_m$, $A_{(-)}$, $A_{(+)}$ and A_s. The solution is called the projection onto $\mathcal{C}$ as we have defined in Chapter 2.

3.3.1 Projections onto A_m and A_s

It is trivial that the algorithms of the isotonic regression introduced in Section 2.4 can also be applied to find the projection $E(g|A_m)$. By (2.13) and Theorem 3.2.2, one obtains

$$E(g|A_s) = g - E(g|(-A_m) \cap \mathcal{S}) \tag{3.26}$$

$$= (g + \bar{g}) - E(g| - A_m). \tag{3.27}$$

3.3.2 Projections onto $A_{(-)}$

By Lemma 3.2.6, for any real-valued function $g \in l_2$,

$$g = \bar{g} + \sum_{i \in X} \frac{(g, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} e_i^{(-)}.$$

By Theorem 3.2.3, any real-valued function $f \in A_{(-)}$ can be written as

$$f = c + \sum_{i \in X} a_i e_i^{(-)}.$$

where c and a_i are real numbers with $a_i \geq 0$. Therefore, by Lemma 3.2.5 we have

$$\| g - f \|^2 = (\bar{g} - c)^2 \sum_{x \in X} w(x) + \sum_{i \in X} [(g, e_i^{(-)}) - (e_i^{(-)}, e_i^{(-)}) a_i]^2 / (e_i^{(-)}, e_i^{(-)}).$$

It follows that the optimal values of c and a_i are given by

$$c^* = \bar{g}, \quad \text{and} \quad a_i^* = \frac{(g, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} \vee 0,$$

where $x \vee 0 = \max\{x, 0\}$. Therefore,

$$E(g|A_{(-)}) = \bar{g} + \sum_{i \in X} \left(\frac{(g, e_i^{(-)})}{(e_i^{(-)}, e_i^{(-)})} \vee 0 \right) e_x. \tag{3.28}$$

3.3.3 Projections onto $A_{(+)}$

By a similar argument, one can show that

$$E(g|A_{(+)}) = \bar{g} + \sum_{i \in X} \left(\frac{(g, e_i^{(+)})}{(e_i^{(+)}, e_i^{(+)})} \vee 0 \right) e_x. \tag{3.29}$$

3.4 Quantifications of General Functions

Quantifications of discrete functions can also be extended to functions on measurable spaces. Let R be the whole real line and let $\mathcal{B}$ be the σ-algebra of Borel sets on R. Let W be a finite Lebesgue measure on $(R, \mathcal{B})$ with support X. Without loss of generality we may assume $W(X) = 1$. Denote

$$W(x) = W((-\infty, x]). \tag{3.30}$$

We denote by $L^p(W)$ the space of all measurable functions on X for which $\int |f|^p dW < \infty$.

Definition 3.4.1 *A function $f(x) \in L_1$ is said to be (with respect the measure W),* *monotone increasing, or in the order of* $\ll_m$, *if for any* $x_1 \le x_2 \in X$, $f(x_1) \le f(x_2)$;

increasing on left average, or in the order of $\ll_{(-)}$, *if for any* $x_1 \le x_2 \in X$ *with* $W(x_1) > 0$, $\int_{t \le x_1} f(t)dW(t)/W(x_1) \le \int_{t \le x_2} f(t)dW(t)/W(x_2)$;

increasing on right average, or in the order of $\ll_{(+)}$, *if for any* $x_1 \le x_2 \in X$ *with* $W(x_2) < 1$, $\int_{t > x_1} f(t)dW(t)/(1 - W(x_1)) \le \int_{t > x_2}^{+\infty} f(t)dW(t)/(1 - W(x_2))$;

increasing on split average, or in the order of $\ll_s$, *if for any* $x \in X$ *with* $0 < W(x) < 1$, $\int_{t \le x} f(t)dW(t)/W(x) \le \int_{t > x} f(t)dW(t)/(1 - W(x))$.

It is straight forward that $f(x)$ is in the order of $\ll_{(-)}$ if and only if

$$\frac{\int_{t \le x_1} f(t)dW(t)}{W(x_1)} \le \frac{\int_{x_1 < t \le x_2} f(t)dW(t)}{W(x_2) - W(x_1)} \tag{3.31}$$

for any $x_1 < x_2 \in X$ with $0 < W(x_1) < W(x_2)$. Similarly, $f(x)$ is in the order of $\ll_{(+)}$ if and only if

$$\frac{\int_{x_1 < t \le x_2} f(t)dW(t)}{W(x_2) - W(x_1)} \le \frac{\int_{t > x_2} f(t)dW(t)}{1 - W(x_2)}, \quad \text{for any } x_1 < x_2, W((x_1, x_2]) \ne 0. \tag{3.32}$$

for any $x_1 < x_2 \in X$ with $W(x_1) < W(x_2) < 1$. Similarly,

The following result is obvious.

Lemma 3.4.1 *Let* $\ll$ *denote one of the orders* $\ll_m$, $\ll_{(-)}$, $\ll_{(+)}$ *and* $\ll_s$. *Then* $f(x)$ *is in the order of* $\ll$ *if and only if* $f(x) + c$ *is in the same order of* $\ll$ *for any* $c \in R$.

Theorem 3.4.1 *(a)* $\ll_m$ *implies* $\ll_{(-)}$ *and* $\ll_{(+)}$; *(b)* $\ll_{(-)}$ *or* $\ll_{(+)}$ *implies* $\ll_s$.

Proof. By symmetry, it suffices to prove that $\ll_m$ implies $\ll_{(-)}$ which in turn implies $\ll_s$. (a) Suppose $f(x)$ is in the order of $\ll_m$, then by definition, for any $x_1 \in X$,

$$f(t) \leq f(x_1), \quad \text{for all } t \leq x_1, t \in X. \tag{3.33}$$

By integrating both sides of (3.33) over $(-\infty, x_1]$, we have

$$\int_{t \leq x_1} f(t)dW(t) \leq f(x_1)W(x_1) \tag{3.34}$$

By a similar argument one obtains

$$f(x_1)(W(x_2) - W(x_1)) \leq \int_{x_1 < x \leq x_2} f(t)dW(t) \quad \text{for any } x_2 > x_1 \in X \tag{3.35}$$

and hence, by (3.31), $f(x)$ is in the order of $\ll_{(-)}$.

(b) Now if $f(x)$ is in the order of $\ll_{(-)}$, then by (3.31),

$$\int_{t \leq x_1} f(t)dW(t)/W(x_1) \leq \int_{x_1 < t \leq x_2} f(t)dW(t)/(W(x_2) - W(x_1)), \tag{3.36}$$

By letting $x_2 \to +\infty$, one obtains that $f(x)$ is in the order of $\ll_s$. □

Similar to the discrete case we define the inner product of two functions $f, g \in L_2(W)$ by

$$(f, g) = \int_{x \in X} f(x)g(x)dW(x).$$

Then by the Cauchy-Schwarz inequality, we have

$$|(f,g)| \leq \int_{x \in X} |f(x)g(x)|dW(x) < (\int_{x \in X} f^2(x)dW(x))^{1/2}(\int_{x \in X} g^2(x)dW(x))^{1/2} < +\infty,$$

for any $f, g \in L_2$. The corresponding norm of $f \in L_2$ is defined as

$$\| f \| = (f, f)^{1/2} = (\int_{x \in X} f^2(x)dW(x))^{1/2}.$$

To avoid arguments in terms of measure theory, we will only consider continuous functions here and define

$$A_m = \{f \in L_2 : f \text{ is continuous and monotone increasing}\};$$
$$A_{(-)} = \{f \in L_2 : f \text{ is continuous and increasing on left average}\};$$
$$A_{(+)} = \{f \in L_2 : f \text{ is continuous and increasing on right average}\};$$
$$A_s = \{f \in L_2 : f \text{ is continuous and increasing on split average}\}$$

and

$$\mathcal{S} = \{f \in L_2 : \int_X f(x)dW(x) = 0.\}$$

Clearly, A_m, $A_{(+)}$, $A_{(-)}$, and A_s are convex cones in L_2 and $\mathcal{S}$ is a subspace of L_2.

Lemma 3.4.2 *If $f, g \in L_2$, then*

$$\lim_{x\to-\infty} \int_{t\le x} f(t)dW(t) \int_{t\le x} g(t)dW(t) / \int_{t\le x} dW(t) = 0; \tag{3.37}$$
$$\lim_{x\to+\infty} \int_{t>x} f(t)dW(t) \int_{t>x} g(t)dW(t) / \int_{t>x} dW(t) = 0. \tag{3.38}$$

Proof. It suffices to prove (3.37). By the Cauchy-Schwarz inequality,

$$\int_{t\le x} |f(t)|dW(t) \le (\int_{t\le x} f^2(t)dW(t))^{1/2} (\int_{t\le x} dW(t))^{1/2},$$
$$\int_{t\le x} |g(t)|dW(t) \le (\int_{t\le x} g^2(t)dW(t))^{1/2} (\int_{t\le x} dW(t))^{1/2}.$$

It follows that

$$\begin{aligned} & |\int_{t\le x} f(t)dW(t) \int_{t\le x} g(t)dW(t)| / \int_{t\le x} dW(t) \\ \le\ & (\int_{t\le x} f^2(t)dW(t) \int_{t\le x} g^2(t)dW(t))^{1/2}. \end{aligned}$$

Since $f, g \in L_2$, we have

$$\lim_{x\to-\infty}\int_{t\le x} f^2(t)dW(t) = 0, \quad \lim_{x\to-\infty}\int_{t\le x} g^2(t)dW(t) = 0.$$

The proof is complete. □

Theorem 3.4.2 *(a)* $A_m^* = (-A_s) \cap \mathcal{S}$;

(b) $A_{(-)}^* = (-A_{(-)}) \cap \mathcal{S}$;

(c) $A_{(+)}^* = (-A_{(+)}) \cap \mathcal{S}$;

(d) $A_s^* = (-A_m) \cap \mathcal{S}$.

Proof. (a) We first prove that $A_m^* \subset (-A_s) \cap \mathcal{S}$. Let $g \in A_m^*$. Then $(f, g) \le 0$ for any $f \in A_m$. Particularly, for each $x \in X$ with $0 < W(x) < 1$, define

$$e_x(y) = \begin{cases} -1/W(x), & \text{if } y \le x, \\ 1/(1 - W(x)), & \text{if } y > x. \end{cases}$$

It is clear that $e_x \in A_m$ and hence

$$(g, e_x) = -\frac{\int_{t\le x} g(t)dW(t)}{W(x)} + \frac{\int_{t>x} g(t)dW(t)}{1 - W(x)} \le 0, \quad \text{for all } x \in X.$$

It follows that $g \in -A_s$. Furthermore, since the constant functions with values 1 and -1 are also in A_m, by (3.8), $g \in \mathcal{S}$. Therefore, $A_m^* \subset (-A_s) \cap \mathcal{S}$.

Conversely, suppose $g \in (-A_s) \cap \mathcal{S}$. Let $M > 0$ be an arbitrary fixed real number and $n > 0$ be an integer. Denote

$$x_0 = -\infty; \quad x_i = -M + \frac{i-1}{n-1}(2M), \ i = 1, 2, \ldots, n; \quad x_{n+1} = +\infty$$

and

$$w_i = \int_{x_i < x \le x_{i+1}} dW(x);$$
$$a_i = \int_{x_i < x \le x_{i+1}} f(x)dW(x)/w_i;$$
$$b_i = \int_{x_i < x \le x_{i+1}} g(x)dW(x)/w_i, \quad i = 0, 1, \ldots, n+1,$$

where by convention, $0/0 = 0$. Since $f \in A_m$, we have

$$a_i \le a_{i+1}, \quad \text{if } w_i, w_{i+1} > 0, \; i = 0, 1, \ldots, n;$$

and since $g \in (-A_s) \cap \mathcal{S}$, we have

$$\frac{\sum_{i \le j} b_i w_i}{\sum_{i \le j} w_i} \ge \frac{\sum_{i > j} b_i w_i}{\sum_{i > j} w_i}, \quad j = 0, 1, \ldots, n \text{ and } \sum_{i=0}^{n+1} b_i w_i = 0.$$

Therefore, by Theorem 3.2.2.

$$\sum_{i=0}^{n+1} a_i b_i w_i \le 0. \tag{3.39}$$

Define $f_n(x) = a_i$ and $g_n(x) = b_i$, if $x_i < x \le x_{i+1}$. By the continuity of f and g, f_n and g_n are bounded functions with

$$\lim_{n \to +\infty} f_n(x) = \begin{cases} a_0, & \text{if } x \le -M; \\ f(x), & \text{if } -M < x \le M; \\ a_{n+1}, & \text{if } x > M \end{cases}$$

and

$$\lim_{n \to +\infty} g_n(x) = \begin{cases} b_0, & \text{if } x \le -M; \\ g(x), & \text{if } -M < x \le M; \\ b_{n+1}, & \text{if } x > M. \end{cases}$$

By the Lebesgue convergence theorem, we have

$$\lim_{n \to \infty} \int_{x \in X} f_n(x) g_n(x) dW(x) = \int_{-M < x \le M} f(x) g(x) dW(x) + a_0 b_0 + a_{n+1} b_{n+1}.$$

However, by (3.39), we have

$$\int_{x\in X} f_n(x)g_n(x)dW(x) = \sum_{i=0}^{n+1} a_i b_i w_i \geq 0.$$

Consequently, by Lemma 3.4.2,

$$\int_X f(x)g(x)dW(x) = \lim_{M\to+\infty} \int_{-M<x\leq M} f(x)g(x)dW(x) \leq 0.$$

Therefore, $A_m^* \supset (-A_s)\cap\mathcal{S}$. It follows that $A_m^* = (-A_s)\cap\mathcal{S}$. By a similar argument, one can prove (b),(c) and (d). □

Corollary 3.4.1 *If (a) $f \in A_m$, $g \in A_s$, or (b) $f, g \in A_{(-)}$ (or $A_{(+)}$), then*

$$\int_X f(x)g(x)dW(x) \geq \int_X f(x)dW(x) \int_X g(x)dW(x).$$

Proof. Similar to the proof of Corollary 3.2.2. □

3.5 Applications in Hypothesis Testing

3.5.1 Applications of Duality in Hypothesis Testing

The usefulness of the concept of duality in estimation is well demonstrated in the literature and the book by Robertson, Wright and Dykstra (1988) includes many important examples. Since $A_m^* = (-A_s)\cap\mathcal{S}$ and $A_s^* = (-A_m)\cap\mathcal{S}$, (Theorem 3.2.2), we shall see that statistical inferences regarding the orders of monotone increasing and increasing on split average are closely related.

Let $\mathbf{X}' = (X_1, X_2, \ldots, X_k)$ be a multivariate normal random variable of dimension k with mean $\boldsymbol{\mu}$ and known covariance matrix diag$(a_1, a_2, \ldots, a_k)$. We are interested in testing the hypothesis

$$H_0 : \boldsymbol{\mu} \in A_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \in A_1 - A_0 \tag{3.40}$$

where A_0 and A_1 are two nested closed convex cones in R^k such that $A_1 \subset A_2$. The likelihood ratio test (LRT) statistic for testing H_0 versus H_1 is given by

$$T = \| \mathbf{X} - E_{\mathbf{w}}(\mathbf{X}|A_0) \|^2 - \| \mathbf{X} - E_{\mathbf{w}}(\mathbf{X}|A_1) \|^2 \tag{3.41}$$

and one rejects H_0 for large values of T, where $w_i = 1/a_i$, and the metric $\| \cdot \|^2$ is induced by the inner product in R^k defined by

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{k} x_i y_i w_i.$$

T can be rewritten as

$$\begin{aligned} T &= \| E_{\mathbf{w}}(\mathbf{X}|A_1) - E_{\mathbf{w}}(\mathbf{X}|A_0) \|^2 + 2(E_{\mathbf{w}}(\mathbf{X}|A_1) - E_{\mathbf{w}}(\mathbf{X}|A_0), \mathbf{X} - E_{\mathbf{w}}(\mathbf{X}|A_1)) \\ &= \| E_{\mathbf{w}}(\mathbf{X}|A_1) - E_{\mathbf{w}}(\mathbf{X}|A_0) \|^2 + 2(E_{\mathbf{w}}(\mathbf{X}|A_1) - E_{\mathbf{w}}(\mathbf{X}|A_0), E_{\mathbf{w}}(\mathbf{X}|A_1^*)) \end{aligned}$$

where the last identity is obtained by (2.13). Therefore, by (2.14), one obtains

$$T = \| E_{\mathbf{w}}(\mathbf{X}|A_0) - E_{\mathbf{w}}(\mathbf{X}|A_1) \|^2 - 2(E_{\mathbf{w}}(\mathbf{X}|A_0), E_{\mathbf{w}}(\mathbf{X}|A_1^*)). \tag{3.42}$$

Theorem 3.5.1 *The LRT statistic of the hypotheses*

$$H_0' : \boldsymbol{\mu} \in A_1^* \quad \textit{versus} \quad H_1' : \boldsymbol{\mu} \in A_0^* - A_1^*. \tag{3.43}$$

is also given by (3.42).

Proof. It is known that A_1^* and A_0^* are also two closed convex cones and by (2.6), $A_1^* \subset A_0^*$. By (3.42), the LRT statistic of (3.43) is given by

$$T' = \| E_{\mathbf{w}}(\mathbf{X}|A_1^*) - E_{\mathbf{w}}(\mathbf{X}|A_0^*) \|^2 - 2(E_{\mathbf{w}}(\mathbf{X}|A_1^*), E_{\mathbf{w}}(\mathbf{X}|(A_0^*)^*)).$$

Consequently, by (2.13) and (2.4), we have

$$T' = \| E_{\mathbf{w}}(\mathbf{X}|A_1) - E_{\mathbf{w}}(\mathbf{X}|A_0) \|^2 - 2(E_{\mathbf{w}}(\mathbf{X}|A_1^*), E_{\mathbf{w}}(\mathbf{X}|A_0)) = T.$$

The proof is complete. □

Definition 3.5.1 *Let T be the LRT statistic of the hypotheses (3.40) given by (3.41). A vector $\boldsymbol{\mu}_0 \in A_0$ is said to be a least favorable configuration (LFC) of T if*

$$P_{\boldsymbol{\mu}_0}(T > c) = \sup_{\boldsymbol{\mu} \in A_0} P_{\boldsymbol{\mu}}(T > c), \quad \textit{for all } c \in R.$$

Denote by $\mathcal{L}_{A_0|A_1}(T)$ the collection of all such least favorable configurations.

Remark Even though (3.40) and (3.43) have the same LRT statistic, the null distributions of the LRT statistic are generally not the same. However, if the two tests have a common *least favorable configuration*, then problems (3.40) and (3.43) will have the same critical region for each significance level $\alpha \in (0, 1)$. In such a case we say that the two problems are (likelihood ratio) equivalent. The following result can be found in Hu and Wright (1994).

Theorem 3.5.2 *If $A_0 \subset A_1$ are closed convex cones and non-oblique, i.e., $P(P(\mathbf{x}|A_1)|A_0) = P(\mathbf{x}|A_0)$, then problems (3.40) and (3.43) and*

$$H_0'' : \boldsymbol{\mu} = \mathbf{0} \quad \textit{versus} \quad H_1'' : \boldsymbol{\mu} \in A_1 \cap A_0^* - \mathbf{0} \tag{3.44}$$

are equivalent.

Definition 3.5.2 *Let $Y_1, Y_2, \ldots, Y_k$ be independent normal variables with means 0 and variances w_i^{-1}. Let M be the number of level sets in $\mathbf{Y}^*$, the isotonic regression of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ with weight vector $\mathbf{w}$. The level probabilities are defined by*

$$P(l, k; \mathbf{w}) = P(M = l), \quad l = 1, 2, \ldots, k.$$

Some examples of equivalent testing problems are given below.

Example 3.5.1 Consider the case that

$$A_0 = \mathcal{S}^{\perp} = \{\mathbf{x} \in R^k : x_1 = x_2 = \cdots = x_k\};$$
$$A_1 = -A_m = \{\mathbf{x} \in R^k : x_1 \geq x_2 \geq \cdots \geq x_k\}.$$

It is trivial that A_0 and A_1 are non-oblique. In this case,

$$\begin{aligned} A_1^* &= A_s \cap \mathcal{S} = \{\mathbf{x} \in R^k : \frac{\sum_{j=1}^{i} x_j w_j}{\sum_{j=1}^{i} w_j} \leq \frac{\sum_{j=i+1}^{k} x_j w_j}{\sum_{j=i+1}^{k} w_j}, \\ &\qquad j = 1, 2, \ldots, k-1 \text{ and } \sum_{j=1}^{k} x_j w_j = 0\}; \\ A_0^* &= \mathcal{S} = \{\mathbf{x} \in R^k : \sum_{j=1}^{k} x_j w_j = 0\}. \end{aligned}$$

The first paper published on the test of $\mathcal{S}^{\perp}$ versus $A_m - \mathcal{S}^{\perp}$ was given by Bartholomew (1959) and the null distribution of the LRT statistic is given by

$$P(T \geq c) = \sum_{l=1}^{k} P(l, k; \mathbf{w}) P(\chi_{l-1}^2 \geq c), \tag{3.45}$$

a chi-bar-square distribution, where χ_i^2 is a standard chi-square variable with i degrees of freedom ($\chi_0^2 \equiv 0$).

Example 3.5.2 Consider the case

$$A_0 = -A_m = \{\boldsymbol{\mu} \in R^k : \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k\}$$

and $A_1 = R^k$. Then A_0 and A_1 are non-oblique and

$$\begin{aligned} A_0^* &= A_s \cap \mathcal{S} = \{\boldsymbol{\mu} \in R^k : \frac{\sum_{j=1}^{i} x_j w_j}{\sum_{j=1}^{i} w_j} \leq \frac{\sum_{j=i+1}^{k} x_j w_j}{\sum_{j=i+1}^{k} w_j}, \\ & \qquad j = 1, 2, \ldots, k-1 \text{ and } \sum_{j=1}^{k} x_j w_j = 0\} \end{aligned}$$

and $A_1^* = \{\mathbf{0}\}$. The null distribution of the LRT statistic for testing $\boldsymbol{\mu} = -A_m$ versus $\boldsymbol{\mu} \notin -A_m$ was obtained by Robertson and Wegman (1978) and has the form

$$\sup_{\boldsymbol{\mu} \in A_m} P(T \geq c) = \sum_{l=1}^{k} P(l, k; \mathbf{w}) P(\chi^2_{k-l} \geq c), \tag{3.46}$$

Example 3.5.3 The LRT of the null hypothesis $H_0 : \boldsymbol{\mu} \in \mathcal{S}^\perp$ versus the alternative hypothesis $H_1 : \boldsymbol{\mu} \in A_s - \mathcal{S}^\perp$ is equivalent to the LRT of the null hypothesis $H_0 : \boldsymbol{\mu} = 0$ versus the alternative hypothesis $H_1 : \boldsymbol{\mu} \in A_s \cap \mathcal{S} - \{\mathbf{0}\}$ of Example 3.5.2.

Example 3.5.4 The LRT of the null hypothesis $H_0 : \boldsymbol{\mu} \in A_s$ versus the alternative hypothesis $H_1 : \boldsymbol{\mu} \in R^k - A_s$ is equivalent to the LRT of the null hypothesis $H_0 : \boldsymbol{\mu} \in A_s \cap \mathcal{S}$ versus the alternative hypothesis $H_1 : \boldsymbol{\mu} \in \mathcal{S} - A_s$ of Example 3.5.1.

As is often the case, the procedures for normal means provide large sample approximations for nonnormal distributions as well as distribution-free procedures based on ranks. As an illustration, we consider the problem for testing a sequence of Poisson means. Our approach follows that of Dykstra and Robertson (1982b) for the multinomial analogue.

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$, $i = 1, 2, \ldots, k$ be independent samples from Poisson populations with means $\mu_1, \mu_2, \ldots, \mu_k$. Denote by $\hat{\mu}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ the unrestricted MLE of μ_i, $i = 1, 2, \ldots, k$. The LRT statistic for testing the null hypothesis A_0 versus the alternative hypothesis $A_1 - A_0$ is given by

$$T = -\sum_{i=1}^{k}(n_i\hat{\mu}_i \ln \mu_i^{(0)} - n_i\mu_i^{(0)}) + \sum_{i=1}^{k}(n_i\hat{\mu}_i \ln \mu_i^{(1)} - n_i\mu_i^{(1)}), \tag{3.47}$$

where $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\mu}^{(1)}$ are the MLE's of $\boldsymbol{\mu}$ when
$bfmu$ belongs to A_0 and A_1 respectively. Expanding $\ln \mu_i^{(0)}$ and $\ln \mu_i^{(1)}$ about the point $\hat{\mu}_i$, T can be expressed as follows:

$$T = \sum_{i=1}^{k} \hat{\mu}_i \alpha_i^{-2}[\sqrt{n_i}(\mu_i^{(0)} - \hat{\mu}_i)]^2 - \sum_{i=1}^{k} \hat{\mu}_i \beta_i^{-2}[\sqrt{n_i}(\mu_i^{(1)} - \hat{\mu}_i)]^2,$$

where α_i is between $\hat{\mu}_i$ and $\mu_i^{(0)}$ and β_i is between $\hat{\mu}_i$ and $\mu_i^{(1)}$. Under H_0, the random vector $\sqrt{\mathbf{n}}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ converges in distribution to $(U_1, U_2, \ldots, U_k)$ where $U_1, U_2, \ldots, U_k$ are independent normal variables with means 0 and variances $\mu_1, \mu_2, \ldots, \mu_k$. Using Theorem 4.4 of Billingsley (1968), it follows that, under H_0, T converges in law to

$$\| E_{\mathbf{w}}(\mathbf{U}|A_0) - U \|^2 - \| E_{\mathbf{w}}(\mathbf{U}|A_1) - U \|^2, \tag{3.48}$$

where $\mathbf{w} = 1/\boldsymbol{\mu}$. Statistic (3.48) is the same LRT statistic for the corresponding hypothesis for normal populations. If one is interested in the testing problem in Example 3.5.3, then the asymptotic distribution of the LRT statistic is given by

$$P(T \geq c) = \sum_{l=1}^{k} P(l, k; \mathbf{w}) P(\chi^2_{k-l} \geq c).$$

3.5.2 Increasing on Average

Since $A^*_{(-)} = (-A_{(-)}) \cap S$ and $A^*_{(+)} = (-A_{(+)}) \cap S$, (Theorem 3.2.4), problems associated with orders of increasing on average (from left or from right) and their dual problems are in fact equivalent. $A_{(-)}$ and $A_{(+)}$) are closely related to orthant cones in a R^k space which are sometimes more easily dealt with than A_m and A_s.

Let $\mathbf{X}' = (X_1, X_2, \ldots, X_k)$ be a multivariate normal random variable of dimension k with mean vector $\boldsymbol{\mu}$ and known covariance matrix $\boldsymbol{\Sigma}$. Consider the problem for testing the hypothesis

$$H_0 : \boldsymbol{\mu} \in A_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \in A_1 - A_0, \tag{3.49}$$

where $A_0 \subset A_1$ are two closed convex sets in R^k. It is trivial that the LRT statistic for testing the null hypothesis H_0 versus the alternative hypothesis H_1, rejects H_0 for large values of

$$T = (\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)}) + 2(\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(X - \boldsymbol{\mu}^{(1)}) \tag{3.50}$$

where $\mu^{(i)}$ is the solution to the problem

$$\min_{\boldsymbol{\mu} \in A_i} (X - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(X - \boldsymbol{\mu}),$$

a general quadratic program whose solution exits and is unique.

The problem of testing the hypotheses (3.49) can be simplified sometimes after making an appropriate transformation of $\mathbf{X}$. Let $\mathbf{Y} = \Gamma\mathbf{X}$, where Γ is a $k \times k$ nonsingular matrix. It is known that $\mathbf{Y}$ is a multivariate normal with mean $\boldsymbol{\nu} = \Gamma\boldsymbol{\mu}$ and covariance matrix $\Gamma\boldsymbol{\Sigma}^{-1}\Gamma'$. Define

$$\Gamma A = \{\Gamma x : x \in A\}. \tag{3.51}$$

Then the problem (3.49) is equivalent to testing the hypotheses

$$H_0 : \boldsymbol{\nu} \in \Gamma A_0 \quad \text{versus} \quad H_1 : \boldsymbol{\nu} \in \Gamma(A_1 - A_0). \tag{3.52}$$

Let Γ be a $k \times k$ nonsingular matrix and

$$A = \{\mathbf{x} \in R^k : D\mathbf{x} \geq \mathbf{c}\}$$

where D is a $r \times k$ matrix ($r \geq 1$) and $\mathbf{c}$ is a vector in R^k. Then

$$\Gamma A = \{\mathbf{y} \in R^k : D\Gamma^{-1}\mathbf{y} \geq \mathbf{c}\}$$

We now consider some testing problems associated with the orders of increasing on average. The orders of increasing on average are closely related to the *starshaped-orders* which are defined as follows.

A vector $\boldsymbol{\mu} \in R^k$ is called lower-starshaped if

$$\mu_1 \geq \frac{w_1\mu_1 + w_2\mu_2}{W_2} \geq \cdots \geq \frac{w_1\mu_1 + w_2\mu_2 + \cdots + w_k\mu_k}{W_k} \geq 0, \tag{3.53}$$

and upper-starshaped if

$$0 \leq \mu_1 \leq \frac{w_1\mu_1 + w_2\mu_2}{W_2} \leq \cdots \leq \frac{w_1\mu_1 + w_2\mu_2 + \cdots + w_k\mu_k}{W_k}, \tag{3.54}$$

where $\mathbf{w}$ is a weight vector and $W_i = \sum_{j=1}^{i} w_j$. Starshaped vectors arise in a variety of applications, see Shaked (1979). Shaked (1979) considered the estimation of a starshaped sequence of Poisson and normal means. Dykstra and Robertson (1982b) obtained the MLE of starshaped multinomial parameters and derived the asymptotic distribution of the LRT statistics. Theoretically, the upper-starshaped property is quite different from the lower-starshaped property, see Shaked (1979) and Dykstra (1984). In fact, the collection of lower-starshaped vectors is an orthogonal cone in a R^k space while the collection of upper-starshaped vectors is an oblique cone. However, in many applications like in the multinomial and Poisson cases, the variables are nonnegative and the constraint of $\mu_1 \geq 0$ is naturally satisfied by the estimates without this constraint. Therefore, the upper-starshaped restriction can be replaced by the order of increasing on left average, i.e.,

$$\mu_1 \leq \frac{w_1\mu_1 + w_2\mu_2}{W_2} \leq \cdots \leq \frac{w_1\mu_1 + w_2\mu_2 + \cdots + w_k\mu_k}{W_k}. \tag{3.55}$$

Let $\mathbf{X}' = (X_1, X_2, \ldots, X_k)$ be a multivariate normal random variable of dimension k with mean $\boldsymbol{\mu}$ and known covariance matrix $\text{diag}(a_1, a_2, \ldots, a_k)$. We will consider the following three hypotheses

$$H_0 : \boldsymbol{\mu} \in \mathcal{S}^{\perp}; \quad H_1 : \boldsymbol{\mu} \in \mathcal{A}_{(-)}; \quad H_2 : \boldsymbol{\mu} \in R^k \tag{3.56}$$

where $\mathbf{w} = (1/a_1, 1/a_2, \ldots, 1/a_k)'$. The hypotheses in (3.56) can be written as

$$H_0 : D\boldsymbol{\mu} = 0; \quad H_1 : D\boldsymbol{\mu} \geq 0; \quad H_2 : \boldsymbol{\mu} \in R^k \tag{3.57}$$

where

$$D = -\begin{pmatrix} \frac{w_1}{W_1} & -1 & 0 & \cdots & 0 & 0 \\ \frac{w_1}{W_2} & \frac{w_2}{W_2} & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{w_1}{W_{k-1}} & \frac{w_2}{W_{k-1}} & \frac{w_3}{W_{k-1}} & \cdots & \frac{w_{k-1}}{W_{k-1}} & -1 \end{pmatrix}$$

If we define

$$\Gamma = \begin{pmatrix} D \\ -\mathbf{w}'/W_k \end{pmatrix},$$

then

$$\Gamma^{-1} = -\begin{pmatrix} \frac{w_2}{W_2} & \frac{w_3}{W_3} & \frac{w_4}{W_4} & \cdots & \frac{w_k}{W_k} & 1 \\ -\frac{W_1}{W_2} & \frac{w_3}{W_3} & \frac{w_4}{W_4} & \cdots & \frac{w_k}{W_k} & 1 \\ 0 & -\frac{W_2}{W_3} & \frac{w_4}{W_4} & \cdots & \frac{w_k}{W_k} & 1 \\ 0 & 0 & -\frac{W_3}{W_4} & \cdots & \frac{w_k}{W_k} & 1 \\ \vdots & \vdots & \vdots & \ddots & \frac{w_k}{W_k} & 1 \\ 0 & 0 & 0 & \cdots & -\frac{W_{k-1}}{W_k} & 1 \end{pmatrix}.$$

So the three hypotheses in (3.56) reduce to

$$\begin{aligned} &H_0' : \nu_i = 0,\ i = 1, 2, \ldots, k-1; \\ &H_1' : \nu_i \geq 0,\ i = 1, 2, \ldots, k-1; \\ &H_2' : \nu \in R^k \end{aligned} \tag{3.58}$$

for a multivariate normal distribution with mean $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$ and covariance matrix $\mathrm{diag}(a_1', a_2', \ldots, a_k')$, where $a_i' = W_{i+1}/(W_i w_{i+1})$, $i = 1, 2, \ldots, k-1$ and $a_k' = 1/W_k$. By (3.50) it can be shown that LRT statistics of H_0' versus $H_1' - H_0'$ and H_1' versus $R^k - H_1'$ are both distributed as

$$L = \sum_{i=1}^{k-1} (Z_i \vee 0)^2, \tag{3.59}$$

where $Z_1, Z_2, \ldots, Z_k$ are independent standard normal variables.

The problem of testing hypotheses in (3.58) is a special case of the well known positive orthant problem and has received extensive attention in order restricted inference, see, e.g., Kudô (1963), Perlman (1969) and Tang, Gnecco and Geller (1989).

As an application, consider the Poisson problem as it appeared in Shaked (1979). Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$, $i = 1, 2, \ldots, k$ be independent samples from Poisson populations with means $\mu_1, \mu_2, \ldots, \mu_k$. Shaked (1979) obtained the MLE estimates of $\boldsymbol{\mu} \in -A_{(-)}$. By symmetry, the LRT statistics for testing equal means versus $\boldsymbol{\mu} \in -A_{(-)}$ but not all equal and $\boldsymbol{\mu} \in -A_{(-)}$ versus $\boldsymbol{\mu} \notin -A_{(-)}$ have the same asymptotic distribution given by (3.59).

Chapter 4

Quantifications of Random Variables

4.1 Introduction

Quantification, or order, of random variables is a very important concept in statistical inference and has many useful applications. For example, the simplest way of comparing two random variables is by comparing the two means. However, such a comparison is based on only two single numbers (the means), and therefore it is often not very informative, especially in nonparametric statistical inference. In addition, the means for some distributions do not exist, such as the Cauchy distribution.

Another application of stochastic orderings is that they can induce many important probability inequalities which play a fundamental role in probability and statistics. Inequalities are used to obtain bounds for probabilities that are more tedious to compute or analytically impossible to handle.

The idea of ordering distributions with respect to the considered property is not very old. The notion of the usual stochastic ordering was first introduced by Mann and Whitney (1947) to characterize the alternative when testing the equality of two distributions. Serious theoretical investigation on stochastic orderings seems to have been initiated by Lehmann (1955); this is the first frequently cited reference. The stochastic ordering, the uniform stochastic ordering and the likelihood ratio ordering are among the most important orderings that have been well studied. All these quantifications can be expressed most conveniently in terms of total positivity (TP) of probability functions which has been extensively applied in several domains of mathematics, statistics, economics, and mechanics.

In Section 4.2 we introduce the concept of total positivity and derive some preliminary results. In Section 4.3 we introduce the notion of quantification of random variables. We show that this notion can be expressed in terms of inequalities of cross-products of probabilities. Many other important properties are readily obtained from this result. In Section 4.4 we show that the quantification of random variables are closely related to the quantification of real-valued functions introduced in Chapter 3. This property plays an important role in the definition of bivariate quantification of random variables in the next chapter. In Section 4.5 the quantification of a series of random variables is defined and illustrated by some examples.

4.2 Total Positivity

For an excellent global view of the theory, the reader is referred to the classical book of Karlin (1968). This book represents a comprehensive, detailed treatment of the analytic structure of totally positive functions and conveys the breadth of the great variety of fields of its applications. A clear, systematic and detailed application of TP in reliability and life testing theory can be found in Chapter 3 to Chapter 5 of Barlow and Proschan (1975).

The theory of total positivity (TP) has been extensively applied in several domains of mathematics, statistics, economics, and mechanics. In statistics, totally positive functions are fundamental in permitting characterizations of best statistical procedures for decision theory. The scope and the power of this concept extend to ascertaining optimal policy for inventory and system supply problems, to clarifying the structure of stochastic processes with continuous path functions, to evaluating the reliability of coherent systems, and to understanding notions of statistical dependency. See Karlin (1968) and Barlow and Proschan (1975) for more details on its theory and application in reliability and life testing theory.

Definition Let D_1 and D_2 be two quasi-ordered sets and $f(x, y)$ a real valued function on $D_1 \times D_2$. $f(\cdot, \cdot)$ is said to be *totally positive of order k* with respect to the orders on D_1 and D_2, (TP_k) if for all $x_1 \prec x_2 \prec \cdots \prec x_m$, $y_1 \prec y_2 \prec \cdots \prec y_m$ ($x_i \in D_1$, $y_i \in D_2$), and all $1 \leq m \leq k$,

$$f\begin{pmatrix} x_1, x_2, \ldots, x_m \\ y_1, y_2, \ldots, y_m \end{pmatrix} = \begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) & \cdots & f(x_1, y_m) \\ f(x_2, y_1) & f(x_2, y_2) & \cdots & f(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_m, y_1) & f(x_m, y_2) & \cdots & f(x_m, y_m) \end{vmatrix} \geq 0.$$

where $|A|$ is the determinant of an $m \times m$ matrix A.

Remark 1. The definition given here is an extension on the usual one where both D_1 and D_2 are assumed to be linearly ordered one dimensional sets of real numbers. This extension will make it more convenient for us define the quantifications of a series of random variables in Section 4.5.

Remark 2. Typically, D_1 and D_2 are either intervals of the real line or countable sets of discrete values on the real line, such as the set of all the integers or the set of nonnegative integers. When D_1 or D_2 is a set of integers, the term "sequence" rather than "function" is usually used.

Many well known families of density functions (both continuous and discrete) are totally positive, see Karlin (1968, p. 19) for some important examples. In fact, every density is TP_1, (nonnegativeness), while TP_2 property is the monotone likelihood property. In addition, $f(x, y)$ is TP_∞ if it can be written as a product of a function of x alone and a function of y alone. So the joint probability density function of two independent random variables is TP_∞. TP_2 is the order of TP-ness which has been found to have a great applications. Higher order TP-ness has hardly been used in applications except for the occasional use of TP_3.

An important specialization occurs if a TP_k function may be written in the form $f(x, y) = g(x - y)$; $g(u)$ is then said to be a *Polya frequency function of order k*, (PF_k). Every PF_2 function is of the form $e^{-\psi(x)}$, where $\psi(x)$ is convex. It follows that probability density functions of the exponential, normal, Weibull and many other random variables are PF_2. Intriguing results in the structure theory of PF_k functions can be found in Karlin and Proschan (1960), Karlin, Proschan and Barlow (1961) and Barlow and Marshall (1964).

Theorem 4.2.1 *Let D_2 be an interval or a countable set on the real line with the usual σ-algebra $\mathcal{B}_2$ and the usual ordering. Let μ_2 be a finite Lebesgue measure on $(D_2, \mathcal{B}_2)$. If $f(x, y)$ is TP_k, then both $\int_{-\infty}^{y} f(x,t)\mu_2(dt)$, and $\int_{y}^{+\infty} f(x,t)\mu_2(dt)$, as functions on $D_1 \times D_2$ are also TP_k.*

Proof. It suffices to prove that $\int_{-\infty}^{y} f(x,t)\mu_2(dt)$, is TP_k. Let $1 \le m \le k$, $x_i \in D_1$, and $y_i \in D_2$, $i = 1, 2, \ldots, m$ be such that $x_1 \le x_2 \le \cdots \le x_m$, $y_1 \le y_2 \le \cdots \le y_m$.

Since $f(x, y)$ is TP_k, we have

$$\begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) & \cdots & f(x_1, y_m) \\ f(x_2, y_1) & f(x_2, y_2) & \cdots & f(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_m, y_1) & f(x_m, y_2) & \cdots & f(x_m, y_m) \end{vmatrix} \geq 0.$$

By integrating the $m \times m$ matrix column by column, one obtains

$$\begin{vmatrix} \int_{-\infty}^{y_1} f(x_1, t)\mu_2(dt) & \int_{y_1}^{y_2} f(x_1, t)\mu_2(dt) & \cdots & \int_{y_{m-1}}^{y_m} f(x_1, t)\mu_2(dt) \\ \int_{-\infty}^{y_1} f(x_2, t)\mu_2(dt) & \int_{y_1}^{y_2} f(x_2, t)\mu_2(dt) & \cdots & \int_{y_{m-1}}^{y_m} f(x_2, t)\mu_2(dt) \\ \vdots & \vdots & \ddots & \vdots \\ \int_{-\infty}^{y_1} f(x_m, t)\mu_2(dt) & \int_{y_1}^{y_2} f(x_2, m)\mu_2(dt) & \cdots & \int_{y_{m-1}}^{y_m} f(x_m, t)\mu_2(dt) \end{vmatrix} \geq 0.$$

Thus, by adding successively the first column to the second column, the second column to the third column, ..., and the $(m-1)$-th column to the m-th column,

$$\begin{vmatrix} \int_{-\infty}^{y_1} f(x_1, t)\mu_2(dt) & \int_{-\infty}^{y_2} f(x_1, t)\mu_2(dt) & \cdots & \int_{-\infty}^{y_m} f(x_1, t)\mu_2(dt) \\ \int_{-\infty}^{y_1} f(x_2, t)\mu_2(dt) & \int_{-\infty}^{y_2} f(x_2, t)\mu_2(dt) & \cdots & \int_{-\infty}^{y_m} f(x_2, t)\mu_2(dt) \\ \vdots & \vdots & \ddots & \vdots \\ \int_{-\infty}^{y_1} f(x_m, t)\mu_2(dt) & \int_{-\infty}^{y_2} f(x_2, m)\mu_2(dt) & \cdots & \int_{-\infty}^{y_m} f(x_m, t)\mu_2(dt) \end{vmatrix} \geq 0$$

and the proof is complete. □

Corollary 4.2.1 *Let $(D_2, \mathcal{B}_2, \mu_2)$ be defined as in Theorem 4.2.1. If $f(x, y)$ is TP_k, $(k \geq 2)$ and $\int_{-\infty}^{+\infty} f(x, t)\mu_2(dt)$ does not depend on x, $x \in D_1$, then*

(a) $\int_{-\infty}^{y} f(x, t)\mu_2(dt)$ as a function on D_1 is antitonic.

(b) $\int_{y}^{+\infty} f(x, t)\mu_2(dt)$ as a function on D_1 is isotonic.

Proof. It suffices to prove (a). Since $f(x, y)$ is TP_k, $(k \geq 2)$, by Theorem 4.2.1, $\int_{-\infty}^{y} f(x, t)\mu_2(dt)$ is TP_k and therefore, TP_2. If $\int_{-\infty}^{+\infty} f(x, t)\mu_2(dt) = 0$, then by the nonnegativeness of $f(x,y)$, $\int_{-\infty}^{y} f(x,t)\mu_2(dt) = 0$ for all $x \in D_1$. If $\int_{-\infty}^{+\infty} f(x,t)\mu_2(dt) > 0$, then since

$$\begin{vmatrix} \int_{-\infty}^{y} f(x_1,t)\mu_2(dt) & \int_{-\infty}^{+\infty} f(x_1,t)\mu_2(dt) \\ \int_{-\infty}^{y} f(x_2,t)\mu_2(dt) & \int_{-\infty}^{+\infty} f(x_2,t)\mu_2(dt) \end{vmatrix} \geq 0$$

for any $x_1 < x_2$, $(x_1, x_2 \in D_1)$ and $y \in D_2$, we have that

$$\int_{-\infty}^{y} f(x_1,t)\mu_2(dt) \geq \int_{-\infty}^{y} f(x_2,t)\mu_2(dt).$$

The proof is complete. □

Theorem 4.2.2 *Let both D_1 and D_2 be either intervals or countable sets on the real line with the usual ordering. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the usual σ-algebras on D_1 and D_2 with finite measures μ_1 and μ_2, respectively. If $f(x,y)$ is TP_k integrable function on $D_1 \times D_2$, then*

$$\int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(s,t)\mu_2(dt)\right] \mu_1(ds) = \int_{-\infty}^{y} \left[\int_{-\infty}^{x} f(s,t)\mu_1(ds)\right] \mu_2(dt)$$

and

$$\int_{x}^{+\infty} \left[\int_{y}^{+\infty} f(s,t)\mu_2(dt)\right] \mu_1(ds) = \int_{y}^{+\infty} \left[\int_{x}^{+\infty} f(s,t)\mu_1(ds)\right] \mu_2(dt)$$

as functions on $D_1 \times D_2$ are also TP_k.

Proof. Since $(D_1, \mathcal{B}_1, \mu_1)$ and $(D_2, \mathcal{B}_2, \mu_2)$ are complete measure space, the two identities follow from the well known Fubini Theorem, (see, e.g., p307 of Royden 1988). The TP property is proved by using Theorem 4.2.1 twice. □

4.3 Quantifications of Random Variables

Let X_1 and X_2 be two random variables with cumulative distribution functions (cdf) $F_1(\cdot)$ and $F_2(\cdot)$, and probability density functions (pdf) $f_1(\cdot)$ and $f_2(\cdot)$ (if they exist), respectively. We assume that F_1 and F_2 have the same domain $\mathcal{X}$. Denote $\bar{F}_i(x) = 1 - F_i(x)$, $i = 1, 2$.

Definition 4.3.1 *X_1 is said to be smaller than X_2*

(a) in the likelihood ratio ordering, denoted by $X_1 \preceq_m X_2$ if $f_2(x)/f_1(x)$ is nondecreasing in x over $\mathcal{X}$;

(b) in the uniform stochastic ordering from the left, denoted by $X_1 \preceq_{(-)} X_2$ if $F_2(x)/F_1(x)$ is nondecreasing in x over $\mathcal{X}$;

(c) in the uniform stochastic ordering from right, denoted by $X_1 \preceq_{(+)} X_2$ if $\bar{F}_2(x)/\bar{F}_1(x)$ is nondecreasing in x over $\mathcal{X}$;

(d) in the (usual) stochastic ordering, denoted by $X_1 \preceq_s X_2$ if $\bar{F}_1(x) \leq \bar{F}_2(x)$, for all $x \in \mathcal{X}$.

Let $I = \{1, 2\}$ be an ordered set with the natural order. Denote $L(i, x) = F_i(x)$, $\bar{L}(i, x) = 1 - F_i(x)$, and $l(i, x) = f_i(x)$ (when exist), for $i \in I$, $x \in D$. The following result can be derived directly from the definition.

Theorem 4.3.1 *(a) $X_1 \leq_m X_2$ if and only if $l(\cdot, \cdot)$ is TP_2.*

(b) $X_1 \leq_{(-)} X_2$ if and only if $L(\cdot, \cdot)$ is TP_2.

(c) $X_1 \leq_{(+)} X_2$ if and only if $\bar{L}(\cdot, \cdot)$ is TP_2.

(d) $X_1 \leq_s X_2$ if and only if $\bar{F}_i(x)$ is nondecreasing in i for each fixed $x \in \mathcal{X}$.

Theorem 4.3.2 *Let t_1, t_2, t_3 and t_4 be extended real numbers such that $t_1 \le t_2$ and $t_3 \le t_4$. Denote*

$$P(t_1, t_2; t_3, t_4) = P(t_1 < X \le t_2)P(t_3 < Y \le t_4).$$

Then

$$X \le_u Y \quad \text{if and only if} \quad P(t_1, t_2; t_3, t_4) \ge P(t_3, t_4; t_1, t_2)$$

for the following four cases,

(a) $u = m$: $t_1 < t_2 \le t_3 < t_4$ are arbitrary real numbers;

(b) $u = (-)$: $t_1 = -\infty$ and $t_2 = t_3 < t_4$ are arbitrary real numbers;

(c) $u = (+)$: $t_1 < t_2 = t_3$ are arbitrary real numbers and $t_4 = +\infty$;

(d) $u = s$: $t_1 = -\infty$, $t_4 = +\infty$ and $t_2 = t_3$ are arbitrary real numbers.

Proof. (a) We shall first prove the case $u = m$. Suppose that for any real numbers $t_1 < t_2 \le t_3 < t_4$,

$$P(t_1, t_2; t_3, t_4) \ge P(t_3, t_4; t_1, t_2),$$

i.e,

$$\int_{t_1}^{t_2} dF_1(t) \int_{t_3}^{t_4} dF_2(t) \ge \int_{t_3}^{t_4} dF_1(t) \int_{t_1}^{t_2} dF_2(t). \tag{4.1}$$

By dividing both sides of (4.1) by $t_2 - t_1$ and then taking limit $t_1 \to t_2$, we have

$$f_1(t_2) \int_{t_3}^{t_4} dF_2(t)dt \ge \int_{t_3}^{t_4} dF_1(t) f_2(t_2). \tag{4.2}$$

for any real numbers $t_2 \le t_3 < t_4$. By dividing both sides of (4.2) by $t_4 - t_3$ and then taking limit $t_4 \to t_3$, one obtains

$$f_1(t_2)f_2(t_3) \ge f_1(t_3)f_2(t_2) \tag{4.3}$$

for any $t_2 \leq t_3$. Therefore, $X \preceq_m Y$. Conversely, if $X \preceq_m Y$, then (4.3) holds for any $t_2 \leq t_3$. By integrating on both sides of (4.3) with respect to t_3 over $(t_3, t_4]$, one obtains (4.2) for any $t_3 < t_4$. By integrating f_1 on both sides of (4.2) with respect to t_3 over $(t_1, t_2]$, one obtains (4.1) for any $t_2 > t_1$.

(b) We shall next prove the case $u = (-)$. For $-\infty = t_1 < t_2 = t_3 < t_4$,

$$\begin{aligned}
& P(t_1, t_2; t_3, t_4) \geq P(t_3, t_4; t_1, t_2) \\
\Longleftrightarrow \quad & P(X \leq t_3)P(t_3 < Y \leq t_4) \geq P(t_3 < X \leq t_4)P(Y \leq t_3) \\
\Longleftrightarrow \quad & P(X \leq t_3)P(Y \leq t_4) \geq P(X \leq t_4)P(Y \leq t_3) \\
\Longleftrightarrow \quad & X \preceq_{(-)} Y
\end{aligned}$$

where the second " $\Longleftrightarrow$ " is obtained by adding or subtracting the term $P(X \leq t_3)P(Y \leq t_3)$ from both sides of the inequalities.

By a similar argument one can prove the case (c).

(d) Finally, we shall prove the case $u = s$. For $-\infty = t_1 < t_2 = t_3 < t_4 = +\infty$,

$$\begin{aligned}
& P(t_1, t_2; t_3, t_4) \geq P(t_3, t_4; t_1, t_2) \\
\Longleftrightarrow \quad & P(X \leq t_2)P(t_2 < Y) \geq P(t_2 < X)P(Y \leq t_2) \\
\Longleftrightarrow \quad & P(X > t_2) \leq P(Y > t_2) \\
\Longleftrightarrow \quad & X \preceq_s Y
\end{aligned}$$

where the second " $\Longleftrightarrow$ " is obtained by adding or subtracting the term $P(X > t_2)P(Y > t_2)$ from both sides of the inequalities. The proof is complete. □

The following well known relationship of these orders can be inferred from Theorem 4.3.2

Theorem 4.3.3 *Let X_1 and X_2 be two random variables.*

(a) $X_1 \leq_m X_2$ implies $X_1 \leq_{(-)} X_2$ and $X_1 \leq_{(+)} X_2$;

(b) $X_1 \leq_{(-)} X_2$ or $X_1 \leq_{(+)} X_2$ implies $X_1 \leq_s X_2$.

Theorem 4.3.4 *For the cases $u = m$, $(+)$, $(-)$ and s, $X \preceq_u Y$ if and only if $\phi(X) \preceq_u \phi(Y)$ for any strictly monotone increasing function ϕ on $\mathcal{X}$.*

Proof. Since ϕ is a strictly monotone increasing function on $\mathcal{X}$, we have

$$P(t_1 < X \leq t_2) = P(\phi(t_1) < \phi(X) \leq \phi(t_2))$$

for any $t_1, t_2 \in \mathcal{X}$. By Theorem 4.3.2 the proof is complete. □

Theorem 4.3.5 *Let X, Y, Z be three random variables with cdf's F, G, and $wF + (1-w)G$, respectively. If $X \preceq_u Y$, then $X \preceq_u Z \preceq_u Y$, for the cases $u = m$, $(+)$, $(-)$ and s.*

Proof. Let t_1, t_2, t_3 and t_4 be defined as in Theorem 4.3.2. By Theorem 4.3.2, $X \preceq_u Y$ if and only if

$$P(t_1 < X \leq t_2)P(t_3 < Y \leq t_4) \geq P(t_3 < X \leq t_4)P(t_1 < Y \leq t_2).$$

It follows that

$$\begin{aligned} & [wP(t_1 < X \leq t_2) + (1-w)P(t_1 < Y \leq t_2)]P(t_3 < Y \leq t_4) \\ \geq \; & [wP(t_3 < X \leq t_4) + (1-w)P(t_3 < Y \leq t_4)]P(t_1 < Y \leq t_2). \end{aligned}$$

By Theorem 4.3.2, $Z \preceq_u Y$. By symmetry one obtains $X \preceq_u Z$. □

One of the most important applications of orderings of random variables is that many important probability inequalities can be obtained from the orderings of distributions. For example, Lehmann (1959) showed that $X \leq_s Y$ if and only if $Eu(X) \leq Eu(Y)$ for all increasing functions $u(\cdot)$. An example in Ross (1983, p. 268) indicates that given two independent random variables, X and Y, $X \leq_m Y$ implies that $(2X+Y) \leq_s (X+2Y)$, and it is equivalent to $Eu(2X+Y) \leq Eu(X+2Y)$, for all increasing function $u(\cdot)$. The following extensions were obtained by Shanthikumar and Yao (1991).

(a) Let X and Y be independent random variables. Let

$$\mathcal{G}_m = \{\phi : R^2 \to R, \phi(x,y) \leq \phi(y,x) \text{ whenever } x \leq y\}.$$

Then $X \leq_m Y$ if and only if $\phi(X,Y) \leq_s \phi(Y,X)$ for all $\phi \in \mathcal{G}_m$.

(b) Let X and Y be independent random variables. Let

$$\mathcal{G}_{(+)} = \{\phi : R^2 \to R, \phi(x,y) \text{ is increasing in } x, \text{ for each } y, \text{ on } \{x \geq y\}$$
$$\text{and decreasing in } y, \text{ for each } y, \text{ on } \{y \geq x\}\}.$$

Then $X \leq_{(+)} Y$ if and only if $\phi(X,Y) \leq_s \phi(Y,X)$ for all $\phi \in \mathcal{G}_{(+)}$.

(c) Let X and Y be independent random variables. Let

$$\mathcal{G}_s = \{\phi : R^2 \to R, \phi(x,y) \text{ is increasing in } x \text{ and decreasing in } y\}.$$

Then $X \leq_s Y$ if and only if $\phi(X,Y) \leq_s \phi(Y,X)$ for all $\phi \in \mathcal{G}_s$.

When X and Y are not independent, the above properties define a class of orders of random variables by their joint distributions, (Shanthikumar and Yao 1991).

The (usual) stochastic order is the first one that appeared in the literature (Mann and Whitney 1947) and has received extensive attention. It arises in numerous settings and its existence can be easily identified in real situations. See Chapter 7 for more details. The uniform stochastic order from the right $\leq_{(+)}$ is well known under the terms *uniform stochastic order* in statistics and *hazard rate order* (when distributions are absolute continuous) in the reliability analysis. The term 'uniform stochastic order' comes from the fact that if $X \leq_{(+)} Y$, then $(X|X \geq t) \leq_s (Y|Y \geq t)$ for any given t. Many of the basic results regarding the uniform stochastic order can be found in Ross (1983). For an explanation and applications of uniform stochastic order in reliability analysis, see Barlow and Proschan (1975). Dykstra, Kochar and Robertson (1991) considered statistical inferences regarding the uniform stochastic order of several random variables. The likelihood ratio order has received relatively less attention in the literature and statistical inferences regarding this ordering were recently considered by Dykstra, Kochar and Robertson (1995).

It is trivial that all the four orders are equivalent for binary random variables. However, differences among these orders will increase as the dimension of the probability vectors increases.

Theorem 4.3.6 *Let X and Y be two discrete random variables with probability vectors $\mathbf{p} = (p_1, p_2, p_3)$ and $\mathbf{q} = (q_1, q_2, q_3)$, respectively. Then*

(a) $X \preceq_m Y$ if and only if $X \preceq_{(-)} Y$ and $X \preceq_{(+)} Y$.

(b) $X \preceq_s Y$ if and only if $X \preceq_{(-)} Y$ or $X \preceq_{(+)} Y$.

Proof. (a). By Theorem 4.3.3, it suffices to prove that if $X \preceq_{(-)} Y$ and $X \preceq_{(+)} Y$, then $X \preceq_m Y$. If $\mathbf{p} \leq_{(-)} \mathbf{q}$, by Theorem 4.3.2,

$$\begin{vmatrix} p_1 & p_1 + p_2 \\ q_1 & q_1 + q_2 \end{vmatrix} = \begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix} \geq 0$$

which is equivalent to $p_1/q_1 \geq p_2/q_2$. Similarly, if $\mathbf{p} \leq_{(+)} \mathbf{q}$, then

$$\begin{vmatrix} p_2+p_3 & p_3 \\ q_2+q_3 & q_3 \end{vmatrix} = \begin{vmatrix} p_2 & p_3 \\ q_2 & q_3 \end{vmatrix} \geq 0,$$

which is equivalent to $p_2/q_2 \geq p_3/q_3$. Combining the two inequalities we also have $p_1/q_1 \geq p_3/q_3$, and hence $X \preceq_m Y$.

(b). It suffices to prove that if $X \preceq_s Y$, then $X \preceq_{(-)} Y$ or $X \preceq_{(+)} Y$. If $X \preceq_s Y$, then

$$p_1 \geq q_1,\ p_1 + p_2 \geq q_1 + q_2, \text{ and } p_1 + p_2 + p_3 = q_1 + q_2 + q_3 = 1.$$

If $p_1/q_1 \geq (p_1 + p_2)/(q_1 + q_2)$, then

$$p_1/q_1 \geq (p_1 + p_2)/(q_1 + q_2) \geq 1 = (p_1 + p_2 + p_3)/(q_1 + q_2 + q_3).$$

Therefore, $X \preceq_{(-)} Y$. If on the other hand $p_1/q_1 < (p_1 + p_2)/(q_1 + q_2)$, then $p_1/q_1 < p_2/q_2$. Since $X \preceq_s Y$, it is trivial that $p_3 < q_3$, and $p_2 + p_3 < q_2 + q_3$. Therefore, $p_2/q_2 > 1 > p_3/q_3$, and hence

$$1 > (p_2 + p_3)/(q_2 + q_3) > p_3/q_3.$$

Therefore, $X \preceq_{(+)} Y$. The proof is complete. □

The following example shows that the property (a) of Theorem 4.3.6 for $k = 3$ does not hold for higher dimensions.

Example 4.3.1 Let X and Y be two discrete random variables with probability vectors $(4/10, 2/10, 3/10, 1/10)$ and $(1/4, 1/4, 1/4, 1/4)$. It is trivial that $X \preceq_{(-)} Y$ and $X \preceq_{(+)} Y$. However, it is not true that $X \preceq_m Y$.

4.4 Relationships between the Quantifications of Functions and the Quantifications of Random Variables

We have studied the quantifications of real-valued functions and the quantifications of random variables. We shall show that these two classes of quantifications are closely related.

Let X and Y be two random variables with cdf's $F(x)$ and $G(x)$ and pdf's $f(x)$ and $g(x)$ respectively. We assume that $F(x)$ and $G(x)$ have the same support $\mathcal{X}$.

Theorem 4.4.1

$$\textit{(a) } X \leq_m Y \iff \frac{g}{f} - 1 \in A_m \cap \mathcal{S};$$
$$\textit{(b) } X \leq_{(-)} Y \iff \frac{g}{f} - 1 \in A_{(-)} \cap \mathcal{S};$$
$$\textit{(c) } X \leq_{(+)} Y \iff \frac{g}{f} - 1 \in A_{(+)} \cap \mathcal{S};$$
$$\textit{(d) } X \leq_s Y \iff \frac{g}{f} - 1 \in A_s \cap \mathcal{S},$$

where A_m, $A_{(-)}$, $A_{(-)}$, A_s, and $\mathcal{S}$ are defined in Section 3.3 with $W = F$.

Proof. It suffices to prove (a), (b) and (d). First, since

$$\int_{\mathcal{X}} \left(\frac{g(x)}{f(x)} - 1 \right) dF(x) = 0,$$

we have

$$\frac{g}{f} - 1 \in \mathcal{S}. \tag{4.4}$$

Now,

$$\text{(a) } X \preceq_m Y \iff \begin{vmatrix} f(x_1) & f(x_2) \\ g(x_1) & g(x_2) \end{vmatrix} \geq 0, \text{ for any } x_1 \leq x_2 \in \mathcal{X}$$

$$\iff \frac{g(x_1)}{f(x_1)} \le \frac{g(x_2)}{g(x_2)}, \quad \text{for any } x_1 \le x_2 \in \mathcal{X}$$
$$\iff \frac{g}{f} \in A_m.$$

By Lemma 3.4.1 and (4.4) one obtains (a).

$$\text{(b)}\ X \preceq_{(-)} Y \iff \begin{vmatrix} F(x_1) & F(x_2) \\ G(x_1) & G(x_2) \end{vmatrix} \ge 0, \quad \text{for any } x_1 \le x_2 \in \mathcal{X}$$
$$\iff \frac{G(x_1)}{F(x_1)} \le \frac{G(x_2)}{F(x_2)}, \quad \text{for any } x_1 \le x_2 \in \mathcal{X}$$
$$\iff \int_{x \le x_1} \frac{g(x)}{f(x)} dF(x)/F(x_1) \le \int_{x \le x_2} \frac{g(x)}{f(x)} dF(x)/F(x_2), \quad \text{for any } x_1 \le x_2 \in \mathcal{X}$$
$$\iff \frac{g}{f} \in A_{(-)}.$$

By Lemma 3.4.1 and (4.4) one obtains (b).

$$\text{(c)}\ X \preceq_s Y \iff \bar{F}(x_1) \le \bar{G}(x_1) \quad \text{for any } x_1 \in \mathcal{X}$$
$$\iff G(x_1)\bar{F}(x_1) \le F(x_1)\bar{G}(x_1) \quad \text{for any } x_1 \in \mathcal{X}$$
$$\iff \frac{G(x_1)}{F(x_1)} \le \frac{\bar{G}(x_1)}{\bar{F}(x_1)}, \quad \text{for any } x_1 \in \mathcal{X}$$
$$\iff \int_{x \le x_1} \frac{g(x)}{f(x)} dF(x)/F(x_1) \le \int_{x > x_1} \frac{g(x)}{f(x)} dF(x)/\bar{F}(x_1) \quad \text{for any } x_1 \in \mathcal{X}$$
$$\iff \frac{g}{f} \in A_s.$$

By Lemma 3.4.1 and (4.4) one obtains (c). The proof is complete. □

Corollary 4.4.1 *If $X \sim Uniform(\mathcal{X})$, then*

(a) $X \preceq_m Y \iff g - 1 \in A_m \cap S;$

(b) $X \preceq_{(-)} Y \iff g - 1 \in \mathcal{A}_{(-)} \cap \mathcal{S}$;

(c) $X \preceq_{(+)} Y \iff g - 1 \in \mathcal{A}_{(+)} \cap \mathcal{S}$;

(d) $X \preceq_{s} Y \iff g - 1 \in \mathcal{A}_{s} \cap \mathcal{S}$

where the weight function W is the Lebesgue measure on $\mathcal{X}$.

Corollary 4.4.2 *Let h be the pdf of $F(Y)$. Then*

(a) $X \preceq_{m} Y \iff h - 1 \in \mathcal{A}_{m} \cap \mathcal{S}$;

(b) $X \preceq_{(-)} Y \iff h - 1 \in \mathcal{A}_{(-)} \cap \mathcal{S}$;

(c) $X \preceq_{(+)} Y \iff h - 1 \in \mathcal{A}_{(+)} \cap \mathcal{S}$;

(d) $X \preceq_{s} Y \iff h - 1 \in \mathcal{A}_{s} \cap \mathcal{S}$,

where the weight function W is the Lebesgue measure on $[0, 1]$.

Proof. By Theorem 4.3.4, $X \preceq_u Y$ if and only if $F(X) \preceq_u F(Y)$. Since $F(X)$ is a uniform distribution on $[0, 1]$, by Corollary 4.4.1 the proof is complete. □

Theorem 4.4.1 and its corollaries show that quantification of random variables and the quantification of real functions are closely related. An application of Theorem 4.4.1 is given in Chapter 6 where we consider the problem of estimating a multinomial parameter under various order constraints. In addition, this relationship plays an important role in the bivariate quantifications of random variables introduced Chapter 5.

4.5 Quantifications of a Series of Random Variables

Let $X_1, X_2, \ldots,$ and X_n be n random variables with cdf's $F_1, F_2, \ldots,$ and F_n, and pdf's $f_1, f_2, \cdots, f_n$, respectively. We assume that $F_1, F_2, \ldots,$ and F_n have the same support $\mathcal{X}$. Define $L(i,x) = F_i(x)$, $\bar{L}(i,x) = \bar{F}_i(x)$, and $l(i,x) = f_i(x)$, for $i = 1, 2, \ldots, n$ and $x \in \mathcal{X}$.

Let $I = \{1, 2, \ldots, n\}$ be a quasi-ordered set. A number of quantifications of $X_1, X_2, \ldots, X_n$ can be defined through the concept of total positivity and types of the quasi-order of I. Some examples are given below.

4.5.1 Linear Orderings

Let $\mathcal{I} = \{1, 2, \ldots, n\}$ be a linearly ordered set such that $1 \prec 2 \cdots \prec n$.

Example 4.5.1 $X_1, X_2, \ldots, X_n$ are said to be linearly likelihood ratio ordered (increasing) if I is a linear order set and $l(i,x)$ is TP_2. Recently Dykstra, Kochar and Robertson (1995) considered statistical inference regarding this ordering for $n = 2$. They obtained a closed form expressions for the maximum likelihood estimate and showed that the asymptotic distribution of the likelihood ratio statistic for testing the equality of the two populations against likelihood ratio ordering restriction is of the chi-bar-square type as discussed by Robertson, Wright and Dykstra (1988). Closed form expressions for the maximum likelihood estimates for more than two likelihood ratio ordered distributions have not been found.

Example 4.5.2 $X_1, X_2, \ldots, X_n$ are said to be linearly uniform stochastic ordered (increasing) if $\mathcal{I}$ is a linear order set and $\bar{L}(i,x)$ is TP_2. The uniform stochastic

stochastic ordering is the most tractable ordering. Dykstra, Kochar and Robertson (1991) have considered statistical inferences with respect to this quantification of the n distributions. They obtained a nice closed form expression for the nonparametric maximum likelihood estimates of the distributions and showed that the asymptotic distribution of the likelihood ratio statistic for testing the equality of the n populations against linear uniform stochastic ordering restriction is also of the chi-bar-square type. In fact, their result can be applied to a more general case that $\mathcal{I}$ is a quasi-order. An example involving data for survival times for carcinoma of the oropharynx is also given in Dykstra, Kochar and Robertson (1991).

Example 4.5.3 $X_1, X_2, \ldots, X_n$ are said to be linear stochastic ordered (increasing) if I is a linear order set and $\bar{L}(i, x)$ is isotonic for each fixed $x \in \mathcal{X}$. Stochastic ordering is the most extensively studied ordering, especially for $n = 2$, (see Chapter 6). When $n > 2$, closed form expressions for the MLE's do not exist and an iterative procedure for finding the MLE's was proposed by Feltz and Dykstra (1985). In Chapter 7 we will propose an algorithm that finds the nonparametric MLE of stochastically bounded survival functions in finite steps, usually two or three steps.

4.5.2 Partial Orderings

Some other orderings of $F_1, F_2, \ldots, F_n$ induced by a partial ordering on I may also be important in applications. For example, the simple tree ordering and the simple loop ordering are often encountered in the control studies. Statistical inference methods associated this kind of partial orderings have not been developed so far. However, the result obtained by Dykstra, Kochar and Robertson (1991) can be applied to any partial orders on I.

Chapter 5

Quantifications of Bivariate Random Variables

5.1 Introduction

In many applications the random variables of interest are dependent. For example, for two ordinal variables, high values of one variable may tend to be associated with high values of the other, and similar for low values. Such relationship of two random variables is known as positive dependence in reliability analysis. There are many ways in which positive dependence might be precisely defined, some based on single-valued measures and some on multiple inequality constraints. An example of the first type would be the requirement that the correlation coefficient of two random variables is positive. Examples of the second type were first considered by Lehmann (1966) and Esary, Prochan and Walkup (1967), among others. A number of its applications were considered in the papers mentioned above, Jogdeo (1968), Esary and Proschan (1970), Barlow and Proschan (1975) and Agresti (1980).

In this chapter we define a class of quantifications of bivariate random variables based on the quantifications of functions and random variables. These notions have direct interpretations and their relationships can be readily established. In addition, these notions are closely related to the concept of positive dependence. We will show that the notions of dependence of random variables in reliability analysis are special cases of these quantifications. But the bivariate quantification defined in this chapter presents a systematic definition and allows one to study bivariate dependence by using the result on quantifications of functions and random variables.

We will only discuss these notions in the bivariate case in this thesis because they are simpler and their relationships are more readily exposed. But all the the notions and results in this chapter can be readily extended to the multivariate case. Further more, for convenience, we will assume that each variate of the bivariate random variable is either discrete or continuous so that the joint density and marginal density functions exist, even though this requirement is not necessary in some cases.

In Section 5.2 quantifications of bivariate random variables are formulated. In Section 5.3 we derive an equivalence theorem of these notions. Hierarchical relations among these quantifications are established in Section 5.4. In Section 5.5 we show that these quantifications can be conveniently expressed in terms of the inequalities of cross product of probabilities over certain regions in the sample space. Some applications are given in Sections 5.6 and 5.7. In Section 5.6 we show that the notions of dependence of random variables in reliability analysis are special cases of these bivariate quantifications. In Section 5.7 we use the results developed in this chapter to analyze the association of ordinal variables.

5.2 Quantifications of Bivariate Random Variables

Suppose we have a bivariate random variable (X, Y) with the joint cumulative distribution function (cdf) $F(x, y)$ and the marginal cdf's $F_X(x)$ and $F_Y(y)$ of X and Y, respectively. Let $f(x, y)$, $f_X(x)$ and $f_Y(y)$ be the associated probability density functions (pdf). Let $\mathcal{X}$ and $\mathcal{Y}$ be the domains of F_X and F_Y respectively.

Let $(X|Y = y)$, $(X|Y \leq y)$ and $(X|Y > y)$ be random variables with cdf's, respectively,

$$\begin{aligned} &F_{X|Y}(x|y) = \left[\frac{\partial}{\partial y}F(x,y)\right] / f_Y(y), \\ &F(x,y)/F_y(y), \\ &(F_X(x) - F(x,y))/(\bar{F}_Y(y) \end{aligned} \tag{5.1}$$

where $\bar{F}(y) = 1 - F(y)$, and pdf's, respectively,

$$\begin{aligned} &f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \\ &\int_{t\leq y} f_{X|Y}(x|t)dF_Y(t)/F_Y(y) \\ &\int_{t>y} f_{X|Y}(x|t)dF_Y(t)/\bar{F}_Y(y). \end{aligned}$$

The origin of our definitions is based on the following observation. Consider $(X|Y = y)$. Clearly, $(X|Y = y)$ is a random process indexed by real-valued number $y \in \mathcal{Y}$. When $(X|Y = y)$ is considered as a function of y (in a general sense), one can introduce the quantifications of real-valued functions to characterize this function. However, since each value of such a function is no longer a real number but a random variable, the comparison of real numbers in the quantifications of real functions should naturally be replaced by the quantifications of random variables.

We will use the symbol "$\ll$" to denote a quantification of functions and "$\preceq$" to denote a quantification of random variables.

Definition 5.2.1 (X, Y) *is said to be in the order of*

$$(\preceq_m, \ll_m) \;\; \textit{if} \;\; (X|Y = y_1) \preceq_m (X|Y = y_2),$$
$$(\preceq_{(-)}, \ll_m) \;\; \textit{if} \;\; (X|Y = y_1) \preceq_{(-)} (X|Y = y_2),$$
$$(\preceq_{(+)}, \ll_m) \;\; \textit{if} \;\; (X|Y = y_1) \preceq_{(+)} (X|Y = y_2),$$
$$(\preceq_s, \ll_m) \;\; \textit{if} \;\; (X|Y = y_1) \preceq_s (X|Y = y_2)$$

for any $y_1, y_2 \in \mathcal{Y}$, $y_1 \leq y_2$.

Definition 5.2.2 (X, Y) *is said to be in the order of*

$$(\preceq_m, \ll_{(-)}) \;\; \textit{if} \;\; (X|Y \leq y_1) \preceq_m (X|Y \leq y_2),$$
$$(\preceq_{(-)}, \ll_{(-)}) \;\; \textit{if} \;\; (X|Y \leq y_1) \preceq_{(-)} (X|Y \leq y_2),$$
$$(\preceq_{(+)}, \ll_{(-)}) \;\; \textit{if} \;\; (X|Y \leq y_1) \preceq_{(+)} (X|Y \leq y_2),$$
$$(\preceq_s, \ll_{(-)}) \;\; \textit{if} \;\; (X|Y \leq y_1) \preceq_s (X|Y \leq y_2)$$

for any $y_1, y_2 \in \mathcal{Y}$, $y_1 \leq y_2$.

Definition 5.2.3 (X, Y) *is said to be in the order of*

$$(\preceq_m, \ll_{(+)}) \;\; \textit{if} \;\; (X|Y > y_1) \preceq_m (X|Y > y_2),$$
$$(\preceq_{(-)}, \ll_{(+)}) \;\; \textit{if} \;\; (X|Y > y_1) \preceq_{(-)} (X|Y > y_2),$$
$$(\preceq_{(+)}, \ll_{(+)}) \;\; \textit{if} \;\; (X|Y > y_1) \preceq_{(+)} (X|Y > y_2),$$
$$(\preceq_s, \ll_{(+)}) \;\; \textit{if} \;\; (X|Y > y_1) \preceq_s (X|Y > y_2)$$

for any $y_1, y_2 \in \mathcal{Y}$, $y_1 \leq y_2$.

Definition 5.2.4 *(X, Y) is said to be in the order of*

$$(\preceq_m, \ll_s) \;\; \text{if} \;\; (X|Y \le y) \preceq_m (X|Y > y),$$
$$(\preceq_{(-)}, \ll_s) \;\; \text{if} \;\; (X|Y \le y) \preceq_{(-)} (X|Y > y),$$
$$(\preceq_{(+)}, \ll_s) \;\; \text{if} \;\; (X|Y \le y) \preceq_{(+)} (X|Y > y),$$
$$(\preceq_s, \ll_s) \;\; \text{if} \;\; (X|Y \le y) \preceq_s (X|Y > y)$$

for any $y \in \mathcal{Y}$.

By symmetry, one can define the order of (X, Y) in the form of $(\ll, \preceq)$. For example, we say that (X, Y) is the order of $(\ll_m, \preceq_m)$ if

$$(Y|X = x_1) \preceq_m (Y|X = x_2), \quad \text{for any } x_1, x_2 \in \mathcal{X}, x_1 \le x_2.$$

And we say that (X, Y) is the order of $(\ll_s, \preceq_m)$ if

$$(Y|X \le x) \preceq_m (Y|X > x), \quad \text{for any } x \in \mathcal{X}.$$

However, the equivalence theorem in the next section implies that it suffices to consider only one of these two forms. This equivalent results is mainly due to the conjugate property of the quantifications of functions and those of random variables.

5.3 An Equivalence Theorem

Theorem 5.3.1 *(X, Y) is $(\preceq_u, \ll_v)$ if and only if (X, Y) is $(\ll_u, \preceq_v)$, where u and v stand for any of m,$(-)$, $(+)$ and s.*

Proof. We will prove the theorem case by case, following the sequential orders of the definitions in the previous section.

$u = m,\ v = m.$ By definition, (X,Y) is in the order of $(\preceq_m, \ll_m)$ if and only if

$$(X|Y = y_1) \preceq_m (X|Y = y_2), \text{ for any } y_1 \le y_2 \in \mathcal{Y}, \tag{5.2}$$

which is, by Theorem 4.3.1, equivalent to

$$\begin{vmatrix} f_{X|Y}(x_1|y_1) & f_{X|Y}(x_2|y_1) \\ f_{X|Y}(x_1|y_2) & f_{X|Y}(x_2|y_2) \end{vmatrix} \ge 0. \text{ for any } x_1 \le x_2 \in \mathcal{X},$$

which is, in turn, equivalent to

$$\begin{vmatrix} f(x_1, y_1) & f(x_2, y_1) \\ f(x_1, y_2) & f(x_2, y_2) \end{vmatrix} \ge 0, \quad \text{i.e. }, f(x,y) \text{ is TP}_2. \tag{5.3}$$

It follows that (X,Y) is $(\ll_m, \preceq_m)$ if and only if $f(x,y)$ is TP_2. By the same token, one can show that (X,Y) is $(\preceq_m, \ll_m)$ if and only if $f(x,y)$ is TP_2.

$u = (-),\ v = m.$ (X,Y) is in the order of $(\preceq_{(-)}, \ll_m)$ if and only if for any $y_1 \le y_2 \in \mathcal{Y}$,

$$(X|Y = y_1) \preceq_{(-)} (X|Y = y_2)$$

$$\Longleftrightarrow \begin{vmatrix} \int_{x\le x_1} f_{X|Y}(x|y_1)\frac{1}{f_X(x)}dF_X(x) & \int_{x\le x_1} f_{X|Y}(x|y_2)\frac{1}{f_X(x)}dF_X(x) \\ \int_{x\le x_2} f_{X|Y}(x|y_1)\frac{1}{f_X(x)}dF_X(x) & \int_{x\le x_2} f_{X|Y}(x|y_2)\frac{1}{f_X(x)}dF_X(x) \end{vmatrix} \ge 0$$

$$\Longleftrightarrow \begin{vmatrix} \int_{x\le x_1} f(x,y_1)\frac{1}{f_X(x)}dF_X(x) & \int_{x\le x_1} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) \\ \int_{x\le x_2} f(x,y_1)\frac{1}{f_X(x)}dF_X(x) & \int_{x\le x_2} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) \end{vmatrix} \ge 0 \tag{5.4}$$

$$\Longleftrightarrow \begin{vmatrix} \int_{x\le x_1} f_{Y|X}(y_1|x)dF_X(x) & \int_{x\le x_1} f_{Y|X}(y_2|x)dF_X(x) \\ \int_{x\le x_2} f_{Y|X}(y_1|x)dF_X(x) & \int_{x\le x_2} f_{Y|X}(y_2|x)dF_X(x) \end{vmatrix} \ge 0$$

$$\Longleftrightarrow \begin{vmatrix} \frac{\int_{x\le x_1} f_{Y|X}(y_1|x)dF_X(x)}{F(x_1)} & \frac{\int_{x\le x_1} f_{Y|X}(y_2|x)dF_X(x)}{F(x_1)} \\ \frac{\int_{x\le x_2} f_{Y|X}(y_1|x)dF_X(x)}{F(x_2)} & \frac{\int_{x\le x_2} f_{Y|X}(y_2|x)dF_X(x)}{F(x_2)} \end{vmatrix} \ge 0$$

$$\Longleftrightarrow \quad (Y|X \le x_1) \preceq_m (Y|X \le x_2),$$

for any $x_1 \leq x_2 \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_{(-)}, \preceq_m)$.

$u = (+)$, $v = m$. This part can be proved by its symmetry to the last case $u = (-)$ and $v = m$.

$u = s$, $v = m$. (X, Y) is in the order of $(\preceq_s, \ll_m)$ if and only if for any $y_1 \leq y_2 \in \mathcal{Y}$,

$$
\begin{aligned}
& (X|Y = y_1) \preceq_s (X|Y = y_2) \\
\iff & \int_{t>x} f_{X|Y}(t|y_1)\frac{1}{f_X(t)}dF_X(t) \leq \int_{t>x} f_{X|Y}(t|y_2)\frac{1}{f_X(t)}dF_X(t) \\
\iff & \begin{vmatrix} \int_{t>x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t>x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \\ f_Y(y_1) & f_Y(y_2) \end{vmatrix} \leq 0 \\
\iff & \begin{vmatrix} \int_{t>x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t>x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \\ \int_{\mathcal{X}} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{\mathcal{X}} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \end{vmatrix} \leq 0 \\
\iff & \begin{vmatrix} \int_{t>x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t>x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \\ \int_{t\leq x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t\leq x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \end{vmatrix} \leq 0 \qquad (5.5) \\
\iff & \begin{vmatrix} \int_{t>x} f_{Y|X}(y_1|t)dF_X(t)/\bar{F}_X(x) & \int_{t>x} f_{Y|X}(y_2|t)dF_X(t)/\bar{F}_X(x) \\ \int_{t\leq x} f_{Y|X}(y_1|t)dF_X(t)/F_X(x) & \int_{t\leq x} f_{Y|X}(y_2|t)dF_X(t)/F_X(x) \end{vmatrix} \leq 0 \qquad (5.6) \\
\iff & \begin{vmatrix} \int_{t\leq x} f_{Y|X}(y_1|t)dF_X(t)/F_X(x) & \int_{t>x} f_{Y|X}(y_1|t)dF_X(t)/\bar{F}_X(x) \\ \int_{t>x} f_{Y|X}(y_2|t)dF_X(t)/\bar{F}_X(x) & \int_{t\leq x} f_{Y|X}(y_2|t)dF_X(t)/F_X(x) \end{vmatrix} \leq 0 \qquad (5.7) \\
\iff & (Y|X \leq x) \preceq_m (Y|X > x),
\end{aligned}
$$

for any $x \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_s, \preceq_m)$.

$u = m$, $v = (-)$. (X, Y) is in the order of $(\preceq_m, \ll_{(-)})$ if and only if for any $y_1 \leq$

$y_2 \in \mathcal{Y}$,

$$(X|Y \le y_1) \preceq_m (X|Y \le y_2)$$

$$\iff \left| \begin{matrix} \frac{\int_{y \le y_1} f_{X|Y}(x_1|y) dF_Y(y)}{F_Y(y_1)} & \frac{\int_{y \le y_2} f_{X|Y}(x_1|y) dF_Y(y)}{F_Y(y_2)} \\ \frac{\int_{y \le y_1} f_{X|Y}(x_2|y) dF_Y(y)}{F_Y(y_1)} & \frac{\int_{y \le y_2} f_{X|Y}(x_2|y) dF_Y(y)}{F_Y(y_1)} \end{matrix} \right| \ge 0$$

$$\iff \left| \begin{matrix} \int_{y \le y_1} f(x_1, y) \frac{1}{f_Y(y)} dF_Y(y) & \int_{y \le y_2} f(x_1, y) \frac{1}{f_Y(y)} dF_Y(y) \\ \int_{y \le y_1} f(x_2, y) \frac{1}{f_Y(y)} dF_Y(y) & \int_{y \le y_2} f(x_2, y) \frac{1}{f_Y(y)} dF_Y(y) \end{matrix} \right| \ge 0 \tag{5.8}$$

$$\iff \left| \begin{matrix} \int_{y \le y_1} f_{Y|X}(y|x_1) \frac{1}{f_Y(y)} dF_Y(y) & \int_{y \le y_2} f_{Y|X}(y|x_1) \frac{1}{f_Y(y)} dF_Y(y) \\ \int_{y \le y_1} f_{Y|X}(y|x_2) \frac{1}{f_Y(y)} dF_Y(y) & \int_{y \le y_2} f_{Y|X}(y|x_2) \frac{1}{f_Y(y)} dF_Y(y) \end{matrix} \right| \ge 0$$

$$\iff (Y|X = x_1) \preceq_{(-)} (Y|X = x_2),$$

for any $x_1 \le x_2 \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_m, \preceq_{(-)})$.

<u>$u = (-)$, $v = (-)$.</u> (X, Y) is in the order of $(\preceq_{(-)}, \ll_{(-)})$ if and only if for any $y_1 \le y_2 \in \mathcal{Y}$,

$$(X|Y \le y_1) \preceq_{(-)} (X|Y \le y_2)$$

$$\iff \left| \begin{matrix} P(X \le x_1|Y \le y_1) & P(X \le x_1|Y \le y_2) \\ P(X \le x_2|Y \le y_1) & P(X \le x_2|Y \le y_2) \end{matrix} \right| \ge 0$$

$$\iff \left| \begin{matrix} P(X \le x_1, Y \le y_1) & P(X \le x_1, Y \le y_2) \\ P(X \le x_2, Y \le y_1) & P(X \le x_2, Y \le y_2) \end{matrix} \right| \ge 0 \tag{5.9}$$

$$\iff \left| \begin{matrix} P(Y \le y_1|X \le x_1) & P(Y \le y_2|X \le x_1) \\ P(Y \le y_1|X \le x_2) & P(Y \le y_2|X \le x_2) \end{matrix} \right| \ge 0$$

$$\iff (Y|X \le x_1) \preceq_{(-)} (Y|X \le x_2),$$

for any $x_1 \le x_2 \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_{(-)}, \preceq_{(-)})$.

$u = (+)$, $v = (-)$. This part can be proved by its symmetry to the last case $u = (-)$ and $v = (-)$.

$u = s$, $v = (-)$. (X, Y) is in the order of $(\preceq_s, \ll_{(-)})$ if and only if for any $y_1 \le y_2 \in \mathcal{Y}$,

$$(X|Y \le y_1) \preceq_s (X|Y \le y_2)$$

$$\iff P(X > x|Y \le y_1) \le P(X > x|Y \le y_2) \tag{5.10}$$

$$\iff \begin{vmatrix} P(X \le x, Y \le y_1) & P(X \le x, Y \le y_2) \\ P(Y \le y_1) & P(Y \le y_2) \end{vmatrix} \ge 0$$

$$\iff \begin{vmatrix} P(X \le x, Y \le y_1) & P(X \le x, Y \le y_2) \\ P(X > x, Y \le y_1) & P(X > x, Y \le y_2) \end{vmatrix} \ge 0 \tag{5.11}$$

$$\iff \begin{vmatrix} P(Y \le y_1|X \le x) & P(Y \le y_2|X \le x) \\ P(Y \le y_1|X > x) & P(Y \le y_2|X > x) \end{vmatrix} \ge 0$$

$$\iff (Y|X \le x) \preceq_{(-)} (Y|X > x),$$

for any $x \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_s, \preceq_{(-)})$.

The four cases when $v = (+)$ can be proved by their symmetry to the cases when $v = (-)$.

$u = m$, $v = s$. (X, Y) is in the order of $(\preceq_m, \ll_s)$ if and only if for any $y \in \mathcal{Y}$,

$$(X|Y \le y) \preceq_m (X|Y > y)$$

$$\iff \begin{vmatrix} \frac{\int_{t\le y} f_{X|Y}(x_1|t)dF_Y(t)}{F_Y(y)} & \frac{\int_{t>y} f_{X|Y}(x_1|t)dF_Y(t)}{F_Y(y)} \\ \frac{\int_{t\le y} f_{X|Y}(x_2|t)dF_Y(t)}{F_Y(y)} & \frac{\int_{t>y} f_{X|Y}(x_2|t)dF_Y(t)}{F_Y(y)} \end{vmatrix} \ge 0$$

$$\iff \begin{vmatrix} \int_{t\le y} f(x_1,t)\frac{1}{f_Y(t)}dF_Y(t) & \int_{t>y} f(x_1,t)\frac{1}{f_Y(t)}dF_Y(t) \\ \int_{t\le y} f(x_2,t)\frac{1}{f_Y(t)}dF_Y(t) & \int_{t>y} f(x_2,t)\frac{1}{f_Y(t)}dF_Y(t) \end{vmatrix} \ge 0 \qquad (5.12)$$

$$\iff \begin{vmatrix} \int_{t\le y} f_{Y|X}(t|x_1)\frac{1}{f_Y(t)}dF_Y(t) & \int_{t>y} f_{Y|X}(t|x_1)\frac{1}{f_Y(t)}dF_Y(t) \\ \int_{t\le y} f_{Y|X}(t|x_2)\frac{1}{f_Y(t)}dF_Y(t) & \int_{t>y} f_{Y|X}(t|x_2)\frac{1}{f_Y(t)}dF_Y(t) \end{vmatrix} \ge 0$$

$$\iff (Y|X = x_1) \preceq_s (Y|X = x_2),$$

for any $x_1 \le x_2 \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_m, \preceq_s)$.

<u>$u = (-)$, $v = s$.</u> (X, Y) is in the order of $(\preceq_{(-)}, \ll_s)$ if and only if for any $y \in \mathcal{Y}$,

$$(X|Y \le y) \preceq_{(-)} (X|Y > y)$$

$$\iff \begin{vmatrix} P(X \le x_1|Y \le y) & P(X \le x_1|Y > y) \\ P(X \le x_2|Y \le y) & P(X \le x_2|Y > y) \end{vmatrix} \ge 0$$

$$\iff \begin{vmatrix} P(X \le x_1, Y \le y) & P(X \le x_1, Y > y) \\ P(X \le x_2, Y \le y) & P(X \le x_2, Y > y) \end{vmatrix} \ge 0 \qquad (5.13)$$

$$\iff \begin{vmatrix} P(X \le x_1) & P(X \le x_1, Y > y) \\ P(X \le x_2) & P(X \le x_2, Y > y) \end{vmatrix} \ge 0 \qquad (5.14)$$

$$\iff P(Y > y|X \le x_1) \le P(Y > y|X \le x_2)$$

$$\iff (Y|X \le x_1) \preceq_s (Y|X \le x_2),$$

for any $x_1 \le x_2 \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_{(-)}, \preceq_s)$.

$u = (+),\ v = s.$ This part can be proved by its symmetry to the last case $u = (-)$ and $v = (-)$.

$u = s,\ v = s.$ (X, Y) is in the order of $(\preceq_s, \ll_s)$ if and only if for any $y \in \mathcal{Y}$,

$$
\begin{aligned}
& (X|Y \le y) \preceq_s (X|Y > y) \\
\iff\ & P(X > x|Y \le y) \le P(X > x|Y > y) && (5.15) \\
\iff\ & \begin{vmatrix} P(Y \le y) & P(Y > y) \\ P(X > x, Y \le y) & P(X > x, Y > y) \end{vmatrix} \ge 0 \\
\iff\ & \begin{vmatrix} P(X \le x, Y \le y) & P(X \le x, Y > y) \\ P(X > x, Y \le y) & P(X > x, Y > y) \end{vmatrix} \ge 0 && (5.16) \\
\iff\ & \begin{vmatrix} P(X \le x) & P(X \le x, Y > y) \\ P(X > x) & P(X > x, Y > y) \end{vmatrix} \ge 0 \\
\iff\ & P(Y > y|X \le x) \le P(Y > y|X > x) && (5.17) \\
\iff\ & (Y|X \le x) \preceq_s (Y|X > x),
\end{aligned}
$$

for any $x \in \mathcal{X}$, i.e., (X, Y) is in the order of $(\ll_s, \preceq_s)$. The proof is complete. □

By Theorem 5.3.1, the orders of $(\preceq_u, \ll_u)$ and $(\ll_u, \preceq_u)$ are equivalent. Even though the other orders do not possess such property of symmetry, they are, by their definitions, measures of degrees of congruence, or positive dependence of two random variables. Particular, the notions of positive dependence appeared in Barlow and Prochan (1975) are special cases of these quantifications, (see Section 5.6).

5.4 Hierarchical Relationship among Bivariate Quantifications

The relations among the bivariate orders defined in the previous section can be easily obtained by Theorem 5.3.1 and the well established results about the orders of random variables. Let $\preceq_{u_1}$ and $\preceq_{u_2}$ be two quantifications of random variables. We say that $\preceq_{u_1}$ is *stronger* than $\preceq_{u_2}$ if for any two random variables X and Y, $X \preceq_{u_1} Y$ implies $X \preceq_{u_2} Y$. Similarly, let $\ll_{v_1}$ and $\ll_{v_2}$ be two quantifications of real-valued functions. We say that $\ll_{v_1}$ is *stronger* than $\ll_{v_2}$ if for any real function $f(x)$, $f(x)$ is in the order of $\ll_{v_1}$ implies that $f(x)$ is in the order of $\ll_{v_2}$.

Theorem 5.4.1 *Let $\preceq u_1$ and $\preceq_{u_2}$ be two quantifications of random variables such that $\preceq_{u_1}$ is stronger than $\preceq_{u_2}$. Let $\ll_{v_1}$ and $\ll_{v_2}$ be two quantifications of real functions such that $\ll_{v_1}$ is stronger than $\ll_{v_2}$. Then if (X, Y) is $(\preceq_{u_1}, \ll_{v_1})$, (X, Y) is $(\preceq_{u_2}, \ll_{v_2})$.*

Proof. By Theorem 4.3.3, (X, Y) is the order of $(\preceq_{u_1}, \ll_{v_1})$ implies that (X, Y) is the order of $(\preceq_{u_2}, \ll_{v_1})$ which is, by Theorem 5.3.1, equivalent to $(\ll_{u_2}, \preceq_{v_1})$. Again, by Theorem 4.3.3, the latter implies $(\ll_{u_2}, \preceq_{v_2})$ which is, by Theorem 5.3.1, equivalent to $(\preceq_{u_2}, \ll_{v_2})$. The proof is complete. □

5.5 Bivariate Quantifications as Inequalities of the Cross Products of Probabilities

In this section we will show that the bivariate quantifications defined in Section 5.2 can be conveniently expressed in terms of inequalities of cross-products of probabilities over certain regions. Some important applications of this property is given in the next section.

Let x_1, x_2, y_1, y_2 be any extended real numbers such that $x_1 < x_2$ and $y_1 < y_2$. Denote

$$\begin{aligned} P(x_1, x_2; y_1, y_2) &= P(x_1 < X \le x_2, y_1 < Y \le y_2) \\ &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \frac{f(x,y)}{f_X(x) f_Y(y)} dF_X(x) dF_Y(y). \end{aligned}$$

Theorem 5.5.1 *(X, Y) is in the order of $(\preceq_u, \ll_v)$ if and only if*

$$\begin{vmatrix} P(x_1, x_2; y_1, y_2) & P(x_1, x_2; y_3, y_4) \\ P(x_3, x_4; y_1, y_2) & P(x_3, x_4; y_3, y_4) \end{vmatrix} \ge 0 \tag{5.18}$$

for the following cases

(a) $u = m$: $x_1 < x_2 \le x_3 < x_4$ are arbitrary real numbers;

(b) $u = (-)$: $x_1 = -\infty$ and $x_2 = x_3 < x_4$ are arbitrary real numbers;

(c) $u = (+)$: $x_1 < x_2 = x_3$ are arbitrary real numbers and $x_4 = +\infty$;

(d) $u = s$: $x_1 = -\infty$, $x_4 = +\infty$ and $x_2 = x_3$ are arbitrary real numbers

and

(a′) $v = m$: $y_1 < y_2 \le y_3 < y_4$ are arbitrary real numbers;

(b′) $v = (-)$: $y_1 = -\infty$ and $y_2 = y_3 < y_4$ are arbitrary real numbers;

(c′) $v = (+)$: $y_1 < y_2 = y_3$ are arbitrary real numbers and $y_4 = +\infty$;

(d′) $v = s$: $y_1 = -\infty$, $y_4 = +\infty$ and $y_2 = y_3$ are arbitrary real numbers.

Proof. We will prove this theorem case by case, following the same order as in the proof of Theorem 5.3.1.

<u>$u = m$, $v = m$.</u> Suppose (X, Y) is $(\preceq_m, \ll_m)$, by (5.3),

$$\begin{vmatrix} f(x_2, y_2) & f(x_2, y_3) \\ f(x_3, y_2) & f(x_3, y_3) \end{vmatrix} \ge 0, \tag{5.19}$$

for any $x_2 \le x_3 \in \mathcal{X}$ and $y_2 \le y_3 \in \mathcal{Y}$. Let $x_1, x_4 \in \mathcal{X}$ and $y_1, y_4 \in \mathcal{Y}$ be such that $x_1 < x_2 \le x_3 < x_4$ and $y_1 < y_2 \le y_3 < y_4$. Now by integrating with respect to the second argument of f, the first column of the matrix in (5.19) over $(y_1, y_2]$ and the second column over $(y_3, y_4]$, we have

$$\begin{vmatrix} \int_{y_1}^{y_2} f(x_2,t)\frac{1}{f_Y(t)}dF_Y(t) & \int_{y_3}^{y_4} f(x_2,t)\frac{1}{f_Y(t)}dF_Y(t) \\ \int_{y_1}^{y_2} f(x_3,t)\frac{1}{f_Y(t)}dF_Y(t) & \int_{y_3}^{y_4} f(x_3,t)\frac{1}{f_Y(t)}dF_Y(t) \end{vmatrix} \ge 0. \tag{5.20}$$

By integrating, with respect to the first argument of f, the first row of the matrix in (5.20) over $(x_1, x_2]$ and the second row over $(x_3, x_4]$, we have

$$\begin{vmatrix} P(x_1 < X \le x_2, y_1 < Y \le y_2) & P(x_1 < X \le x_2, y_3 < Y \le y_4) \\ P(x_3 < X \le x_4, y_1 < Y \le y_2) & P(x_3 < X \le x_4, y_3 < Y \le y_4) \end{vmatrix} \ge 0. \tag{5.21}$$

Therefore, (5.18) holds. Conversely, suppose (5.18), or equivalently, (5.21) holds for any $x_1 < x_2 \le x_3 < x_4 \in \mathcal{X}$ and $y_1 < y_2 \le y_3 < y_4 \in \mathcal{Y}$. By dividing the first row and second row of (5.21) by $x_2 - x_1$ and $x_4 - x_3$ respectively, and then taking the limits $x_1 \to x_2$, $x_4 \to x_3$, one obtains (5.20). By dividing the first column and second column of (5.20) by $y_2 - y_1$ and $y_4 - y_3$ respectively, and then taking the limits $y_1 \to y_2$, $y_4 \to y_3$, one obtains (5.19). Therefore, (X, Y) is $(\preceq_m, \ll_m)$.

$u = (-)$, $v = m$. By (5.4), (X, Y) is in the order of $(\preceq_{(-)}, \ll_m)$ if and only if for any $x_3 \le x_4 \in \mathcal{X}$ and $y_2 \le y_3 \in \mathcal{Y}$,

$$\begin{vmatrix} \int_{x \le x_3} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) & \int_{x \le x_3} f(x,y_3)\frac{1}{f_X(x)}dF_X(x) \\ \int_{x \le x_4} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) & \int_{x \le x_4} f(x,y_3)\frac{1}{f_X(x)}dF_X(x) \end{vmatrix} \ge 0,$$

or equivalent

$$\begin{vmatrix} \int_{x \le x_2} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) & \int_{x \le x_2} f(x,y_3)\frac{1}{f_X(x)}dF_X(x) \\ \int_{x_3 < x \le x_4} f(x,y_2)\frac{1}{f_X(x)}dF_X(x) & \int_{x_3 < x \le x_4} f(x,y_3)\frac{1}{f_X(x)}dF_X(x) \end{vmatrix} \ge 0 \tag{5.22}$$

where $x_2 = x_3$. By carrying on the same operations as in the case $u = m$ and $v = m$ for the second argument of f, one establishes the desired result.

$u = (+),\ v = m.$ This part can be proved by its symmetry to the last case $u = (-)$ and $v = m$.

$u = s,\ v = m.$ By (5.5), (X, Y) is in the order of $(\preceq_s, \ll_m)$ if and only if for any $x \in \mathcal{X}$ and $y_2 \leq y_3 \in \mathcal{Y}$,

$$\begin{vmatrix} \int_{t\leq x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t\leq x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \\ \int_{t>x} f(t, y_1)\frac{1}{f_X(t)}dF_X(t) & \int_{t>x} f(t, y_2)\frac{1}{f_X(t)}dF_X(t) \end{vmatrix} \geq 0. \tag{5.23}$$

By carrying on the same operations as in the case $u = m$ and $v = m$ for the second argument of f, one establishes the desired result.

$u = m,\ v = (-).$ By Theorem 5.3.1, (X, Y) is in the order of $(\preceq_m, \ll_{(-)})$ if and only if (X, Y) is the order of $(\ll_m, \prec_{(-)})$. We have proved for the case $(\prec_{(-)}, \ll_m)$. By symmetry, one obtains the desired result.

$u = (-),\ v = (-).$ By (5.9), (X, Y) is in the order of $(\preceq_{(-)}, \ll_{(-)})$ if and only if for any $x_3 < x_4 \in \mathcal{X}$ and $y_3 \leq y_4 \in \mathcal{Y}$,

$$\begin{vmatrix} P(X \leq x_3, Y \leq y_3) & P(X \leq x_3, Y \leq y_4) \\ P(X \leq x_4, Y \leq y_3) & P(X \leq x_4, Y \leq y_4) \end{vmatrix}$$

$$= \begin{vmatrix} P(X \leq x_3, Y \leq y_3) & P(X \leq x_3, y_3 < Y \leq y_4) \\ P(X \leq x_4, Y \leq y_3) & P(X \leq x_4, y_3 < Y \leq y_4) \end{vmatrix}$$

$$= \begin{vmatrix} P(X \leq x_2, Y \leq y_2) & P(X \leq x_2, y_3 < Y \leq y_4) \\ P(x_3 < X \leq x_4, Y \leq y_2) & P(x_3 < X \leq x_4, y_3 < Y \leq y_4) \end{vmatrix} \geq 0 \tag{5.24}$$

where $x_2 = x_3$ and $y_2 = y_3$.

$u = (+),\ v = (-).$ This part can be proved by its symmetry to the last case $u = (-)$ and $v = (-)$.

$u = s,\ v = (-).$ By (5.11), (X, Y) is in the order of $(\preceq_s, \ll_{(-)})$ if and only if for any

$x \in \mathcal{X}$ and $y_2 = y_3 < y_4 \in \mathcal{Y}$,

$$\begin{vmatrix} P(X \leq x, Y \leq y_3) & P(X \leq x, Y \leq y_4) \\ P(X > x, Y \leq y_3) & P(X > x, Y \leq y_4) \end{vmatrix}$$

$$= \begin{vmatrix} P(X \leq x, Y \leq y_2) & P(X \leq x, y_3 < Y \leq y_4) \\ P(X > x, Y \leq y_2) & P(X > x, y_3 < Y \leq y_4) \end{vmatrix} \geq 0. \qquad (5.25)$$

The four cases when $v = (+)$ can be proved by their symmetry to the cases when $v = (-)$.

The cases of $v = s$, $u = m$.or $(-)$ can be proved by the same argument as in the case $v = m$ or $(-)$ and $u = s$. The case $v = s$ and $u = (+)$ can be proved by symmetry. $\underline{u = s,\ v = s.}$ By (5.16), (X, Y) is in the order of $(\preceq_s, \ll_s)$ if and only if for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\begin{vmatrix} P(X \leq x, Y \leq y) & P(X \leq x, Y > y) \\ P(X > x, Y \leq y) & P(X > x, Y > y) \end{vmatrix} \geq 0.$$

The proof is complete. □

5.6 Positive Dependence of Random Variables

Positive dependence of random variables is an important concept and has many useful applications, especially in reliability analysis and life sciences; e.g., the life times of components in a system may be positively dependent because of the common environmental stress, shocks and common sources of power. The following definitions appeared in Barlow and Proschan (1975, p.142 and p.145).

Definition 5.6.1 *Given random variables X and Y, we say the following:*

(a) X and Y are positively quadrant dependent, denoted by PQD(X,Y), if

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y) \quad \textit{for all } x, y.$$

(b) X is left tail decreasing in Y, denoted by LTD(X|Y), if

$$P(X \leq x | Y \leq y) \quad \textit{is nonincreasing in } y \textit{ for all } x.$$

(c) X is right tail increasing in Y, denoted by RTI(X|Y), if

$$P(X > x | Y > y) \quad \textit{is nondecreasing in } y \textit{ for all } x.$$

(d) X is stochastically increasing in Y, denoted by SI(X|Y), if

$$P(X > x | Y = y) \quad \textit{is nondecreasing in } y \textit{ for all } x.$$

(e) X and Y are totally positive of order 2, denoted by $TP_2(X,Y)$, if the joint probability density $f(x,y)$ of X and Y is TP_2.

(f) X and Y are said to be right corner set increasing, denoted by $RCSI(X,Y)$, if

$$P(X > x, Y > y | X > x', Y > y')$$

is nondecreasing in x' and y' for each fixed x and y.

We shall now show that the above definitions are special cases of the bivariate quantifications defined in this chapter.

Theorem 5.6.1 *(a) (X,Y) is PQD(X,Y) if and only if (X,Y) is $(\preceq_s, \ll_s)$;*

(b) (X,Y) is LTD(X|Y) if and only if (X,Y) is $(\preceq_s, \ll_{(-)})$;

(c) (X,Y) *is RTI*$(X|Y)$ *if and only if* (X,Y) *is* $(\preceq_s, \ll_{(+)})$;

(d) (X,Y) *is SI*$(X|Y)$ *if and only if* (X,Y) *is* $(\preceq_s, \ll_m)$;

(e) (X,Y) *is* TP_2 (X,Y) *is* $(\preceq_m, \ll_m)$;

(f) (X,Y) *is* $RCSI(X,Y)$ *if and only if* (X,Y) *is* $(\preceq_{(+)}, \ll_{(+)})$.

Proof. (a). By Theorem (5.5.1), (X,Y) is $(\preceq_s, \ll_s)$ if and only if for any real numbers x and y,

$$\begin{vmatrix} P(X \le x, Y \le y) & P(X \le x, Y > y) \\ P(X > x, Y \le y) & P(X > x, Y > y) \end{vmatrix}$$

$$= \begin{vmatrix} P(Y \le y) & P(Y > y) \\ P(X > x, Y \le y) & P(X > x, Y > y) \end{vmatrix}$$

$$= \begin{vmatrix} 1 & P(Y > y) \\ P(X > x) & P(X > x, Y > y) \end{vmatrix} \ge 0$$

which is equivalent to (a).

(b). By Theorem (5.5.1), (X,Y) is $(\preceq_s, \ll_{(-)})$ if and only if for any real numbers x and $y_3 \le y_4$,

$$\begin{vmatrix} P(X \le x, Y \le y_3) & P(X \le x, y_3 < Y \le y_4) \\ P(X > x, Y \le y_3) & P(X > x, y_3 < Y \le y_4) \end{vmatrix}$$

$$= \begin{vmatrix} P(X \le x, Y \le y_3) & P(X \le x, Y \le y_4) \\ P(X > x, Y \le y_3) & P(X > x, Y \le y_4) \end{vmatrix}$$

$$= \begin{vmatrix} P(X \le x, Y \le y_3) & P(X \le x, Y \le y_4) \\ P(Y \le y_3) & P(Y \le y_4) \end{vmatrix} \ge 0,$$

which is equivalent $P(X \le x | Y \le y_3) \ge P(X \le x | Y \le y_4)$.

Similarly, one can prove (c).

(d) By definition, (X,Y) is $(\preceq_s, \ll_m)$ if and only if for any real numbers $y_1 \leq y_2$,

$$(X|Y = y_1) \preceq_s (X|Y = y_2),$$

i.e.,

$$P(X > x|Y = y_1) \leq P(X > x|Y = y_2)$$

for any real numbers x and $y_1 \leq y_2$. By definition, (X,Y) is $SI(X|Y)$.

(e) This is already proved in the proof of Theorem 5.5.1.

(f) This part is a little bit complicated. Note that

$$\begin{aligned} & P(X > x, Y > y|X > x', Y > y') \\ = & \frac{P(X > \max(x, x'), Y > \max(y, y')}{P(X > x', Y > y')} \\ = & \begin{cases} 1, & \text{if } x' > x,\ y' > y; \\ \frac{P(X>x',Y>y)}{P(X>x',y>y')} & \text{if } x' > x,\ y' \leq y; \\ \frac{P(X>x,Y>y')}{P(X>x',y>y')} & \text{if } x' \leq x,\ y' > y; \\ \frac{P(X>x,Y>y)}{P(X>x',y>y')} & \text{if } x' \leq x,\ y' \leq y \end{cases} \end{aligned} \tag{5.26}$$

It follows that (X,Y) is RCSI(X,Y) if and only if

$$\frac{P(X > x', Y > y)}{P(X > x', y > y')} \text{ is nondecreasing in } x' \text{ and } y' \text{ with } y' \leq y \tag{5.27}$$

and

$$\frac{P(X > x, Y > y')}{P(X > x', y > y')} \text{ is nondecreasing in } x' \text{ and } y' \text{ with } x' \leq x. \tag{5.28}$$

It can be shown that both (5.27) and (5.28) are equivalent to

$$\begin{aligned} & P(X > x_1, Y > y_1)P(X > x_2, Y > y_2) \\ \geq \ & P(X > x_2, Y > y_1)P(X > x_1, Y > y_2) \end{aligned} \tag{5.29}$$

for any $x_1 \leq x_2$, $y_1 \leq y_2$. Therefore, (X, Y) is RCSI(X, Y) if and only if (5.29) holds. Since

$$P(X > x_1, Y > y_1)P(X > x_2, Y > y_2)$$
$$-P(X > x_2, Y > y_1)P(X > x_1, Y > y_2)$$

$$= \begin{vmatrix} P(X > x_1, Y > y_1) & P(X > x_1, Y > y_2) \\ P(X > x_2, Y > y_1) & P(X > x_2, Y > y_2) \end{vmatrix}$$

$$= \begin{vmatrix} P(x_1 < X \leq x_2, Y > y_1) & P(x_1 < X \leq x_2, Y > y_2) \\ P(X > x_2, Y > y_1) & P(X > x_2, Y > y_2) \end{vmatrix}$$

$$= \begin{vmatrix} P(x_1 < X \leq x_2, Y > y_1) & P(x_1 < X \leq x_2, Y > y_2) \\ P(X > x_2, y_1 < Y \leq y_2) & P(X > x_2, Y > y_2) \end{vmatrix}$$

by Theorem 5.5.1, (X, Y) is $RCSI(X, Y)$ if and only if (X, Y) is $(\preceq_{(+)}, \ll_{(+)})$. □

By Theorems 5.4.1 and 5.6.1, one can easily obtain the chart of the implication among notions of bivariate dependence in Barlow and Proschan (1975, p.146). The proof of the following result can be found in Lehmann (1966).

Lemma 5.6.1 *(X, Y) is PQD if and only if*

$$E(g(X)h(Y)) \geq E(g(X))E(h(Y)) \tag{5.30}$$

for any nondecreasing functions g and h with finite expectation in (5.30). In addition, if (X,Y) is PQD and $E(XY) = E(X)E(Y)$, then X and Y are independent.

Consequently, if (X,Y) is PQD, then $Cov(X,Y) \geq 0$. However, the converse is not true as illustrated by the following example.

Example 5.6.1 *Suppose (X,Y) is a bivariate random which is uniformly distributed on $\{(0,0),(2,-1),(3,1)\}$. Then $Cov(X,Y) = 1/3 > 0$. Let$g(-1) = -1$ and $g(0) = g(1) = 0$. Clearly, g is a nondecreasing function on $\{-1,0,1\}$. However, $Cov(X, g(Y)) = -1/9 < 0$.*

5.7 Positive Associations of Ordinal Random Variables

In many studies variables are measured on ordinal scales. These scales consist of a collection of naturally ordered categories (e.g., stages of a disease, degree of recovery from an illness, ordinal preference scale). Ordinal scales also result when discrete measurement is used with inherently continuous variables such as age, education and degree of prejudice. There are many advantages to be gained from using ordinal methods of the standard nominal procedures. For example, ordinal methods have greater power for detecting important alternative to null hypotheses such as the one of independence. See Agresti (1984) for more details on the analysis of ordinal categorical data.

It is of great importance to study how ordinal variables interrelate with each other. For example, high values on one ordinal scale may tend to be associated with high values on the other, and similarly for low values. There are many ways that one can characterize such dependence of ordinal variables, some based on the single-valued measures and some on multiple inequality constraints. A well known example of the first type would be the requirement that the Kendall's τ be positive. In the following discussion we will consider some definitions of the second and relate them to the notions introduced in this chapter.

5.7.1 Odds Ratios of Cross-Classification Tables

Suppose that X and Y are ordinal variables with $X = 1, 2, \ldots, I$ and $Y = 1, 2, \ldots, J$. Denote $\pi_{ij} = P(X = i, Y = j)$. The following definitions can be found in Agresti (1984). We shall refer to X and Y as the row variable and the column variable respectively.

A basic set of $(I-1) \times (J-1)$ odds ratios is

$$\theta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}, \quad i = 1, \ldots, I-1; \; j = 1, \ldots, J-1. \tag{5.31}$$

These odds ratios are called *local ratios* and their values describe the relative magnitudes of "local" associations in the table.

The local odds ratios treat row and column alike. Another family of odds ratios, one that makes a distinction between row and column, is

$$\theta'_{ij} = \frac{(\sum_{b\le j}\pi_{ib})(\sum_{b>j}\pi_{i+1,b})}{(\sum_{b>j}\pi_{ib})(\sum_{b\le j}\pi_{i+1,b})}, \quad i = 1, \ldots, I-1; \; j = 1, \ldots, J-1. \tag{5.32}$$

These odds ratios are local in the row variable but "global" in the column variable, since all J levels of the column variables are used in each odds ratio.

A third family of odds ratios of ordinal variables is

$$\theta''_{ij} = \frac{(\sum_{a\le i}\sum_{b\le j}\pi_{ab})(\sum_{a>i}\sum_{b>j}\pi_{ab})}{(\sum_{a\le i}\sum_{b>j}\pi_{ab})(\sum_{a>i}\sum_{b\le j}\pi_{ab})}, \quad i = 1, \ldots, I-1; \; j = 1, \ldots, J-1. \tag{5.33}$$

These measures treat row and column alike and describe associations that are global in both variables.

For each set of the local, local-global, and global odds ratios, independence is equivalent to all odds ratios equaling 1. An association described by one of these measures is referred to as "positive" or "negative" in accordance with odds ratios greater or smaller than 1. By Theorem 5.5.1 it is easy to see that these three positive associations are equivalent to that (X, Y) is in the order of $(\preceq_m, \ll_m)$, $(\preceq_m, \ll_s)$ and $(\preceq_s, \ll_s)$, respectively.

A broad classes of odds ratios can be defined corresponding to the bivariate quantifications introduced in this chapter. These classes are listed below.

$$\theta_{ij}^{(m,m)} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}; \tag{5.34}$$

$$\theta_{ij}^{((-),m)} = \frac{(\sum_{a=1}^{i}\pi_{aj})\pi_{i+1,j+1}}{\pi_{i+1,j}(\sum_{a=1}^{i}\pi_{a,j+1})}; \tag{5.35}$$

$$\theta_{ij}^{((+),m)} = \frac{\pi_{ij}(\sum_{a=i+1}^{I}\pi_{a,j+1})}{(\sum_{a=i+1}^{I}\pi_{aj})\pi_{i,j+1}}; \tag{5.36}$$

$$\theta_{ij}^{(s,m)} = \frac{(\sum_{a=1}^{i}\pi_{aj})(\sum_{a=i+1}^{I}\pi_{a,j+1})}{(\sum_{a=i+1}^{I}\pi_{aj})(\sum_{a=1}^{i}\pi_{a,j+1})}; \tag{5.37}$$

$$\theta_{ij}^{(m,(-))} = \frac{(\sum_{b=1}^{j}\pi_{ib})\pi_{i+1,j+1}}{(\sum_{b=1}^{j}\pi_{i+1,b})\pi_{i,j+1}}; \tag{5.38}$$

$$\theta_{ij}^{((-),(-))} = \frac{(\sum_{a=1}^{i}\sum_{b=1}^{j}\pi_{ab})\pi_{i+1,j+1}}{(\sum_{b=1}^{j}\pi_{i+1,b})(\sum_{a=1}^{i}\pi_{a,j+1})}; \tag{5.39}$$

$$\theta_{ij}^{((+),(-))} = \frac{(\sum_{b=1}^{j}\pi_{ib})(\sum_{a=i+1}^{I}\pi_{a,j+1})}{(\sum_{a=i+1}^{I}\sum_{b=1}^{j}\pi_{ab})\pi_{i,j+1}}; \tag{5.40}$$

$$\theta_{ij}^{(s,(-))} = \frac{(\sum_{a=1}^{i}\sum_{b=1}^{j}\pi_{ab})(\sum_{a=i+1}^{I}\pi_{a,j+1})}{(\sum_{a=i+1}^{I}\sum_{b=1}^{j}\pi_{ab})(\sum_{a=1}^{i}\pi_{a,j+1})}; \tag{5.41}$$

$$\theta_{ij}^{(m,(+))} = \frac{\pi_{ij}(\sum_{b=j+1}^{J} \pi_{i+1,b})}{\pi_{i+1,j}(\sum_{b=j+1}^{J} \pi_{i+1,b})}; \tag{5.42}$$

$$\theta_{ij}^{((-),(+))} = \frac{(\sum_{a=1}^{i} \pi_{aj})(\sum_{b=j+1}^{J} \pi_{i+1,b})}{p_{i+1,j} \sum_{a=1}^{i} \sum_{b=j+1}^{J} \pi_{ab}}; \tag{5.43}$$

$$\theta_{ij}^{((+),(+))} = \frac{\pi_{ij}(\sum_{a=i+1}^{I} \sum_{b=j+1}^{J} \pi_{ab})}{(\sum_{a=i+1}^{I} \pi_{aj})(\sum_{b=j+1}^{J} \pi_{ib})}; \tag{5.44}$$

$$\theta_{ij}^{(s,(+))} = \frac{(\sum_{a=1}^{i} \pi_{aj})(\sum_{a=i+1}^{I} \sum_{b=j+1}^{J} \pi_{ab})}{(\sum_{a=i+1}^{I} \pi_{aj})(\sum_{a=1}^{i} \sum_{b=j+1}^{J} \pi_{ab})}; \tag{5.45}$$

$$\theta_{ij}^{(m,s)} = \frac{(\sum_{b=1}^{j} \pi_{ib})(\sum_{b=j+1}^{J} \pi_{i+1,b})}{(\sum_{b=1}^{j} \pi_{i+1,b})(\sum_{b=j+1}^{J} \pi_{ib})}; \tag{5.46}$$

$$\theta_{ij}^{((-),s)} = \frac{(\sum_{a=1}^{i} \sum_{b=1}^{j} \pi_{ab})(\sum_{b=j+1}^{J} \pi_{i+1,b})}{(\sum b = 1^{j} p_{i+1,b})(\sum_{a=1}^{i} \sum_{b=j+1}^{J} \pi_{ab})}; \tag{5.47}$$

$$\theta_{ij}^{((+),s)} = \frac{(\sum_{b=1}^{j} \pi_{ib})(\sum_{a=i+1}^{I} \sum_{b=j+1}^{J} \pi_{ab})}{(\sum_{a=i+1}^{I} \sum_{b=1}^{j} \pi_{ab})(\sum_{b=j+1}^{J} \pi_{ib})}; \tag{5.48}$$

$$\theta_{ij}^{(s,s)} = \frac{(\sum_{a=1}^{i} \sum_{b=1}^{j} \pi_{ab})(\sum_{a=i+1}^{I} \sum_{b=j+1}^{J} \pi_{ab})}{(\sum_{a=i+1}^{I} \sum_{b=1}^{j} \pi_{ab})(\sum_{a=1}^{i} \sum_{b=j+1}^{J} \pi_{ab})} \tag{5.49}$$

for $i = 1, \ldots, I-1$ and $j = 1, \ldots, J-1$.

Some of the above classes of odds ratios have also appeared in Grove (1984). For example, the odds ratios defined by (3.1), (3.2), (3.3) and (3.4) in Grove (1984) are equivalent to those defined by (5.42), (5.37), (5.43) and (5.47), respectively.

5.7.2 Sampling Schemes and Estimations

For a given random sample of size N, let X_{ij} denote the observed count of $(X, Y) = (i, j)$ and let $m_{ij} = N\pi_{ij}$ denote the corresponding expected count. Let

$$x_{i+} = \sum_{j=1}^{J} x_{ij}, \quad x_{+j} = \sum_{i=1}^{I} x_{ij}$$

and

$$m_{i+} = \sum_{j=1}^{J} m_{ij}, \quad m_{+j} = \sum_{i=1}^{I} m_{ij}.$$

It is clear that the cross-product of probabilities in Theorem 5.5.1 can be replaced by the corresponding cross-product of m_{ij}'s.

In the *simple sampling scheme* (SSS) where the total sample size N is fixed, $\{X_{ij}\}$ has a multinomial distribution, with pdf

$$f(\{x_{ij}\}) = \frac{N!}{\prod_{i,j} x_{ij}!} \prod_{i,j} \left(\frac{m_{ij}}{N}\right)^{x_{ij}}.$$

Therefore, the kernel of the likelihood function is given by

$$\prod_{i,j} m_{ij}{}^{x_{ij}}. \tag{5.50}$$

Although in observational studies only a single sample may be examined, in experimental situations it is more usual to have several groups, with the total number of individuals in each group determined by the sampling plan. The resulting distribution is a product of multinomials from these groups and the sampling scheme is called *product multinomial sampling scheme* (PMSS). For example, if the row totals are fixed by $n_1, n_2, \ldots, n_I$, then the sampling scheme is called PMSS with row totals fixed (PMSSR) and the resulting distribution is given by

$$f(\{x_{ij}\}) = \prod_i \left[\frac{x_{i+}!}{\prod_j x_{ij}!} \prod_j \left(\frac{m_{ij}}{m_{i+}}\right)^{x_{ij}}\right].$$

Clearly the parameters m_{ij} are not estimable in this scheme. A common approach to this problem is to consider the restricted parameter space with $m_{i+} = n_i$. Then the kernel of the likelihood function is also given by (5.50). Similar results hold for the product multinomial sampling scheme with column totals fixed (PMSSC).

Suppose that (X, Y) is in the order of $(\preceq_u, \ll_v)$. where u and v stand for any of m, $(-)$, $(-)$ and s. The restricted MLE's under the above sampling schemes are generally not the same unless the MLE under SSS is in the restricted parameter space of PMSS.

Theorem 5.7.1 *Let u and v denote one of the symbols m, $(-)$, $(+)$ and s.*

(a) Suppose that (X, Y) is in the order of $(\preceq_u, \ll_m)$. Then the restricted MLE's of m_{ij} under SSS and PMSSC are identical.

(b) Suppose that (X, Y) is in the order of $(\preceq_m, \ll_v)$. Then the restricted MLE's of m_{ij} under SSS and PMSSR are identical.

Proof. It suffices to prove (a). We shall first prove the case $u = m$. Under the simple sampling scheme, the MLE of m_{ij} minimizes

$$\prod_{i=1}^{I}\prod_{j=1}^{J} m_{ij}{}^{x_{ij}} \tag{5.51}$$

subject to the constraints $\sum_{i,j} m_{ij} = N$ and

$$m_{ij}m_{i+1,j+1} \geq m_{i,j+1}m_{i+1,j}, \quad i = 1, 2, \ldots, I-1, \; j = 1, 2, \ldots, J-1. \tag{5.52}$$

We can rewrite (5.51) as

$$(\prod_{j=1}^{J} m_{+j}^{x_{+j}}) \prod_{i=1}^{I}\prod_{j=1}^{J} p^{x_{ij}}, \tag{5.53}$$

where $p_{ij} = m_{ij}/m_{+j} = \pi_{ij}/\pi_{+j}$. Clearly the only constraint on m_{+j} is $\sum_{j=1}^{J} = N$. Therefore, the MLE of m_{+j} is given by x_{+j}. It follows that the MLE's under SSS and PMSSC are identical. Similarly, one can prove the cases $u = (-)$, $(+)$, and s. □

Corollary 5.7.1 *Suppose that (X, Y) is in the order of $(\preceq_m, \ll_m)$. Then the restricted MLE's of m_{ij} under SSS, PMSSR and PMSSC are identical.*

We shall now consider some applications of the above results.

Example 5.7.1 If $I = 2$, then the orders $(\preceq_u, \ll_v)$ are equivalent to $(\preceq_m, \ll_v)$, where $u, v = m, (-), (+)$, or s. By Theorem 5.7.1, the restricted MLE's of m_{ij} under SSS and PMSSR are identical.

Example 5.7.2 Suppose that (X, Y) is the order of $(\preceq_m, \ll_m)$. By Corollary 5.7.1, the MLE's of m_{ij} under SSS, PMSSR and PMSSC are identical.

Particularly, suppose $I = 2$. Then if we consider the order of (X, Y) as $(\preceq_m, \ll_m)$ and the sampling scheme as PMSSC, we will have the bioassay problem which was first considered by Ayer and coworkers (1956). Let $\mathbf{x}_1 = (x_{11}, x_{12}, \ldots, x_{1J})$ and $\mathbf{x}_2 = (x_{21}, x_{22}, \ldots, x_{2J})$. Then the MLE of m_{ij} is given by

$$\hat{m}_{1j} = (x_{1j} + x_{2j}) E_{(\mathbf{x}_1 + \mathbf{x}_2)} \left(\frac{\mathbf{x}_1}{\mathbf{x}_1 + \mathbf{x}_2} | \mathcal{A} \right)_j$$

and $\hat{m}_{2j} = x_{1j} + x_{2j} - \hat{m}_{1j}$, for $j = 1, 2, \ldots, J$, where $\mathcal{A} = \{(\theta_1, \ldots, \theta_J) : \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k\}$ is the cone of nonincreasing vectors. See Robertson, Wright, and Dykstra (1988, p32).

On the other hand, if we consider the order of (X, Y) as $(\preceq_m, \ll_m)$, which is equivalent to $(\ll_m, \preceq_m)$ by Theorem 5.3.1, and the sampling scheme as PMSSR, we will have a seemingly different problem: estimating m_{ij} is equivalent to estimating two multinomial parameters under the likelihood ratio ordering. This problem was recently considered by Dykstra, Kochar and Robertson (1995). In their paper, Dykstra, Kochar and Robertson obtained the MLE's of m_{ij} and derived the asymptotic distribution of the likelihood ratio statistics for testing the equality for two discrete distributions against the alternative that one distribution is smaller than the other in the likelihood ratio order. Even though the MLE's were obtained in the discrete setting, they provide generalized MLE's, in the sense of Kiefer and Wolfortz (1956), under the assumption that the family of interest is the collection of all pairs of univariate distributions. In addition, Dykstra, Kochar and Robertson (1995) showed that these estimates are strongly consistent.

Chapter 6

Multinomial Estimation Procedures under Order Restrictions

6.1 Introduction

Suppose $x_1, x_2, \ldots, x_k$ are the observed values of a random vector which possesses a multinomial distribution with parameter n and probability vector (PV) $\mathbf{p}$. Assume also that $\mathbf{p}$ is restricted to lie within a closed convex subset K of A where

$$A = \{(p_1, p_2, \ldots, p_k) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1\}$$

is the set of all probability vectors of length k.

Standard estimation procedures in a multinomial setting are the methods of maximum likelihood, Pearson minimum chi-square, Neyman modified minimum chi-square, minimum discrimination information, and the Freeman-Tukey criteria. These estimation techniques lead to optimization problems which can be phrased, respectively, in the following manner:

$$\min_{\mathbf{p} \in K} 2 \sum_{i=1}^{k} x_i \ln(x_i / np_i), \tag{6.1}$$

$$\min_{\mathbf{p}\in K} \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}, \tag{6.2}$$

$$\min_{\mathbf{p}\in K} \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{x_i}, \tag{6.3}$$

$$\min_{\mathbf{p}\in K} 2\sum_{i=1}^{k} np_i \ln(np_i/x_i), \tag{6.4}$$

$$\min_{\mathbf{p}\in K} 4\sum_{i=1}^{k} (\sqrt{x_i} - \sqrt{np_i})^2. \tag{6.5}$$

These problems may be difficult to solve, and indeed, this difficulty has often influenced the estimation procedure which is used. For example, if K is a subspace, the Neyman modified minimum chi-square procedure is essentially a weighted least squares problem which is well understood. All these procedures are asymptotically equivalent.

Cressie and Read (1984) define the directed divergence of the PV, $\mathbf{q}$, with respect to the PV, $\mathbf{p}$, of order λ as

$$I^{\lambda}(\mathbf{q}:\mathbf{p}) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{k} q_i \left[\left(\frac{q_i}{p_i} \right)^{\lambda} - 1 \right]. \tag{6.6}$$

In order to ensure that I^{λ} is defined for all PVs $\mathbf{p}$ and $\mathbf{q}$, for $\lambda \neq 0$ and -1, we evaluate $I^{\lambda}(\mathbf{q}:\mathbf{p})$ as

$$I^{\lambda}(\mathbf{q}:\mathbf{p}) = \frac{1}{\lambda(\lambda+1)} \left\{ \sum_{i=1}^{k} \frac{q_i^{\lambda+1}}{p_i^{\lambda}} - 1 \right\}. \tag{6.7}$$

with the convention that $0/0$ equals 0, and allow ∞ as a possible value. For $\lambda = 0$ or -1, we define $I^{\lambda}(\mathbf{q}:\mathbf{p})$ by continuity in λ. It is known that $I^{\lambda}(\mathbf{q}:\mathbf{p})$ is always non-negative and it is zero if and only if $\mathbf{p} = \mathbf{q}$. Furthermore, $I^{\lambda}(\mathbf{q}:\mathbf{p})$ is a strictly convex function of $\mathbf{p}$ over A if each $q_i > 0$. Thus $I^{\lambda}(\mathbf{q}:\mathbf{p})$ acts as a discrepancy measure between $\mathbf{p}$ and $\mathbf{q}$.

Cressie and Read (1984)'s beautiful observation is that all five of the aforementioned estimation criteria are special cases of the problem

$$\min_{\mathbf{p}\in K} 2nI^{\lambda}(\hat{\mathbf{p}} : \mathbf{p}). \tag{6.8}$$

In particular, if $\lambda = 1$, (6.7) reduces to the Pearson minimum chi-square expression (6.2). If $\lambda = -2$, it becomes the Neymann modified minimum chi-square expression (6.3). While not explicitly defined if $\lambda = 0$ or -1, we obtain the log-likelihood ratio expression (6.1) and the MDI expression (6.4), respectively, for these values of λ if we define matters by continuity. The Freeman-Tukey criterion (6.5) follows if $\lambda = -1/2$.

In this section, we consider the problem that K consists of probability vectors which satisfy some order restrictions. The results in Section 6.2 and 6.3 have been obtained by Dykstra and Lee (1991) and Dykstra, Lee and Yan (1996), respectively.

In the following discussion operations on vectors mean the corresponding operations on each element of the vectors. For example, $\mathbf{p} \leq \mathbf{q}$ means that $p_i \leq q_i$, $i = 1, 2, \ldots, k$.

6.2 Multinomial Estimation Procedures under Order Restrictions

6.2.1 Estimation under the Isotonic Constraint

Dykstra and Lee (1991) following earlier work of Dykstra (1985) and Lee (1987b) showed that if K is an isotonic cone, then all these procedures can be related and the corresponding estimates expressed in terms of equal weights, least squares projections. Specifically, if we let $\mathbf{p}^{(\lambda)}$ be the solution of (6.8) for an isotonic cone $K = I$, then

$$\mathbf{p}^{(\lambda)} = E\left(\hat{\mathbf{p}}^{\lambda+1}|I\right)^{\frac{1}{\lambda+1}} / \sum_{i=1}^{k}\left(\hat{\mathbf{p}}^{\lambda+1}|I\right)_i^{\frac{1}{\lambda+1}} \quad \text{for } \lambda > -1 \tag{6.9}$$

$$\mathbf{p}^{(\lambda)} = E\left(\hat{\mathbf{p}}^{\lambda+1}|A\right)^{\frac{1}{\lambda+1}} / \sum_{i=1}^{k}\left(\hat{\mathbf{p}}^{\lambda+1}|A\right)_i^{\frac{1}{\lambda+1}} \quad \text{for } \lambda < -1 \tag{6.10}$$

where A is the antitonic cone, and

$$\mathbf{p}^{(\lambda)} = \exp\{E(\ln\hat{\mathbf{p}}|I)\} / \sum_{i=1}^{k}\exp\{E(\ln\hat{\mathbf{p}}|I)_i\} \quad \text{for } \lambda = -1.$$

6.2.2 Estimation under the Stochastic Ordering Constraint

Recently, Dykstra, Lee and Yan (1996) considered the problem that K consists of a pair of probability vectors which are stochastically ordered. They showed that these procedures are also closely connected, behave similarly, and have elegant solutions in terms of a single least squares projection.

In the two-sample problem, let $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k)$ be the observed values of random vectors which possess independent multinomial distributions with parameters m and n and probability vectors $\mathbf{p}$ and $\mathbf{q}$. The estimate of $(\mathbf{p}, \mathbf{q})$ such that $\mathbf{p} \preceq_s \mathbf{q}$ are given by the solutions to the problem

$$\min_{\mathbf{p} \preceq_s \mathbf{q}} \left[2mI^{\lambda}(\hat{\mathbf{p}} : \mathbf{p}) + 2nI^{\lambda}(\hat{\mathbf{q}} : \mathbf{q})\right]. \tag{6.11}$$

where $\hat{\mathbf{p}} = \mathbf{x}/m$ and $\hat{\mathbf{q}} = \mathbf{y}/n$. In the event that criterion (6.1) (maximum likelihood) is used, Robertson and Wright (1981) have shown that the solution is given by

$$\begin{aligned} \mathbf{p}^{(0)} &= \hat{\mathbf{p}}\left\{\frac{m}{N} + \frac{n}{N}E_{\hat{\mathbf{p}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{D})\right\}, \\ \mathbf{q}^{(0)} &= \hat{\mathbf{q}}\left\{\frac{m}{N}E_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{I}) + \frac{n}{N}\right\} \end{aligned} \tag{6.12}$$

where $\mathcal{D} = \{\mathbf{x} : x_1 \geq x_2 \cdots \geq x_k\}$ and $\mathcal{I} = \{\mathbf{x} : x_1 \leq x_2 \cdots \leq x_k\}$. The following result can be found in Dykstra, Lee and Yan (1996).

Theorem 6.2.1 *The solution to (6.11) is given by*

$$\mathbf{p}^{(\lambda)} = \hat{\mathbf{p}}\left\{\frac{m}{N} + \frac{n}{N}E_{\hat{\mathbf{p}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{D})^{\lambda+1}\right\}^{1/(\lambda+1)}/c, \tag{6.13}$$

$$\mathbf{q}^{(\lambda)} = \hat{\mathbf{q}}\left\{\frac{m}{N}E_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{I})^{\lambda+1} + \frac{n}{N}\right\}^{1/(\lambda+1)}/c \tag{6.14}$$

where $N = m + n$ *and* c *is the normalizing constant*

$$c = \sum_{i=1}^{k}\hat{p}_i\left\{\frac{m}{N} + \frac{n}{N}E_{\hat{\mathbf{p}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{D})_i^{\lambda+1}\right\}^{1/(\lambda+1)} \tag{6.15}$$

$$= \sum_{i=1}^{k}\hat{q}_i\left\{\frac{m}{N}E_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{I})_i^{\lambda+1} + \frac{n}{N}\right\}^{1/(\lambda+1)} \tag{6.16}$$

if $\lambda \neq -1$; *if* $\lambda = -1$,

$$\mathbf{p}^{(-1)} = \hat{\mathbf{p}}E_{\hat{\mathbf{p}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{D})^{n/N}/c \tag{6.17}$$

$$\mathbf{q}^{(-1)} = \hat{\mathbf{q}}E_{\hat{\mathbf{q}}}(\hat{\mathbf{p}}/\hat{\mathbf{q}}|\mathcal{I})^{m/N}/c \tag{6.18}$$

where $c = \sum_{i=1}^{k}\hat{p}_iE_{\hat{\mathbf{p}}}(\hat{\mathbf{q}}/\hat{\mathbf{p}}|\mathcal{D})_i^{n/N} = \sum_{i=1}^{k}\hat{q}_iE_{\hat{\mathbf{q}}}(\hat{\mathbf{p}}/\hat{\mathbf{q}}|\mathcal{I})_i^{m/N}$.

Note that only one least squares projection is required for all λ in this case, while a different least squares projection is required for each value of λ for the isotone constraints, (see Section 6.2).

In the one sample problem, i.e., when one of the multinomial parameters $\mathbf{p}$ and $\mathbf{q}$ is known, Dykstra, Lee and Yan (1996) showed that the solution to the stochastic ordering estimation problem is (rather surprisingly) independent of λ. Specifically, if $\mathbf{q}$ is known, then the estimate of $\mathbf{p}$ is given by

$$\mathbf{p}^{(\lambda)} = \hat{\mathbf{p}}E_{\hat{\mathbf{p}}}(\mathbf{q}/\hat{\mathbf{p}}|\mathcal{D}) \text{ if } K = \{\mathbf{p} \in A : \mathbf{p} \preceq_s \mathbf{q}\}; \tag{6.19}$$

$$\mathbf{p}^{(\lambda)} = \hat{\mathbf{p}}E_{\hat{\mathbf{p}}}(\mathbf{q}/\hat{\mathbf{p}}|\mathcal{I}) \text{ if } K = \{\mathbf{p} \in A : \mathbf{q} \preceq_s \mathbf{p}\}. \tag{6.20}$$

See Robertson and Wright (1981).

6.2.3 Estimation under the Bound Constraint

Sometimes bounds for some p_i's may be obtained from other sources. Suppose one is interested in estimating $\mathbf{p}$ such that $\mathbf{p} \in K = \{\mathbf{p} \in A : \mathbf{p} \leq \mathbf{q}\}$, where $\mathbf{q}$ is a given, nonnegative, real vector. The solution exists if and only if $\sum_{i=1}^{k} q_i \geq 1$. Clearly, if $\mathbf{q}$ is a PV, then $\mathbf{q}$ is the only PV that satisfies the constraint and thus is the solution. When $\mathbf{q} = 1$, the problem reduces to the unrestricted one. We propose the following algorithm for the problem

$$\min_{\mathbf{p} \leq \mathbf{q}} 2nI^{\lambda}(\hat{\mathbf{p}} : \mathbf{p}).$$

Algorithm

Step 0. Let $s = 0$ and $V_0 = \emptyset$.

Step 1. For $i \notin V_s$, compute

$$p_{s,i} = \frac{1 - \sum_{i \in V_s} q_i}{\sum_{i \notin V_s} \hat{p}_i} \hat{p}_i.$$

Let $V_{s+1} = V_s \cup \{i : i \notin V_s, p_{s,i} > q_i\}$.

Step 2. If $A_{s+1} = A_s$, then the solution is given by $p_i^* = q_i$ for $i \in V_s$ and $p_i^* = p_{s,i}$ for $i \notin V_s$. Otherwise, replace s by $s + 1$ and go to Step 1.

The utility of the above algorithm lies in the following lemma.

Lemma 6.2.1 *Let $\mathbf{x}^*$ be the solution to the problem*

$$\min 2nI^{\lambda}(\hat{\mathbf{p}} : \mathbf{x}), \; s.t. \; \sum_{i=1}^{k} x_i = c, \; and \; \mathbf{x} \leq \mathbf{q}$$

where $c > 0$ is a given real number. Let $V = \{i : c\hat{p}_i \geq q_i, 1 \leq i \leq k\}$. Then $x_i = q_i$ if $i \in V$.

Proof. Clearly the solution exits and it is unique. Suppose $\mathbf{x}^*$ is the solution, then we have

$$x_i^* \leq q_i \leq c\hat{p}_i \quad \text{if } i \in V.$$

Suppose there exists an index $\alpha \in V$ such that

$$x_\alpha^* < q_\alpha \leq c\hat{p}_\alpha. \tag{6.21}$$

Since $\sum_{i=1}^k x_i^* = \sum_{i=1}^k c\hat{p}_i = c$, there must exist an index $\beta \notin V$ such that

$$c\hat{p}_\beta < x_\beta^* \leq q_\beta. \tag{6.22}$$

Let $c' = x_\alpha^* + x_\beta^*$. It can be shown that

$$\frac{2n}{\lambda(\lambda+1)}\left(\frac{\hat{p}_\alpha^{\lambda+1}}{x_\alpha^\lambda} + \frac{\hat{p}_\beta^{\lambda+1}}{(c'-x_\alpha)^\lambda}\right) \tag{6.23}$$

is a strict convex function of p_α which is minimized at $\bar{x}_\alpha = \hat{p}_\alpha c'/(\hat{p}_\alpha + \hat{p}_\beta)$. By (6.21) and (6.22), $\hat{p}_\alpha x_\beta^* > \hat{p}_\beta x_\alpha^*$. It follows that $\bar{x}_\alpha > x_\alpha^*$ and hence the function defined by (6.23) is strictly decreasing from $[x_\alpha^*, \bar{x}_\alpha)$. Therefore, while holding all x_i^*'s except x_α^* and x_β^* fixed, one can decrease $2nI^\lambda(\hat{\mathbf{p}} : \mathbf{x})$ by moving from x_α^* in the direction to $\bar{x}_\alpha$ ($x_\beta = c' - x_\alpha$)) without violating the constraints $\mathbf{x} \leq \mathbf{q}$ and $\sum_{i=1}^k x_i = c$. This yield a direct contradiction. The proof is complete. □

6.2.4 Estimation under the Uniform Stochastic Ordering Constraint

Here we are interested in the following problem

$$\min_{\mathbf{p} \preceq_{(-)} \mathbf{q}} 2nI^\lambda(\hat{\mathbf{p}} : \mathbf{p}). \tag{6.24}$$

where $\mathbf{q}$ is a given probability vector. We reparameterize by letting

$$\theta_i = \frac{p_{i+1}}{\sum_{j=1}^{i} p_j}, \ i = 1, 2, \ldots, k-1. \tag{6.25}$$

Then we have $(1 + \theta_i) = \sum_{j=1}^{i+1} p_j / \sum_{j=1}^{i} p_j$ and thus $\sum_{j=1}^{i} p_j = 1/\prod_{j=i}^{k}(1+\theta_{j-1})$. It follows that

$$p_i = \frac{\theta_{i-1}}{\prod_{j=i}^{k}(1+\theta_{j-1})}, \ \text{for } i = 2, 3, \ldots, k \tag{6.26}$$

and

$$p_1 = \frac{p_2}{\theta_1} = \frac{1}{\prod_{j=2}^{k}(1+\theta_{j-1})}. \tag{6.27}$$

By denoting $\theta_0 = +\infty$, the expression of (6.26) is also valid for $k = 1$. Similarly, one defines $\hat{\theta}_i$ and ϕ_i from $\hat{\mathbf{p}}$ and $\mathbf{q}$. Now,

$$\begin{aligned} I^\lambda &= \frac{1}{\lambda(\lambda+1)}\left[\sum_{i=1}^{k} \frac{\hat{p}_i^{\lambda+1}}{p_i^\lambda} - 1\right] \\ &= \frac{1}{\lambda(\lambda+1)}\left[\sum_{i=1}^{k} \frac{\hat{\theta}_{i-1}^{\lambda+1} \prod_{j=i}^{k}(1+\theta_{j-1})^\lambda}{\theta_{i-1}^{\lambda} \prod_{j=i}^{k}(1+\hat{\theta}_{j-1})^{\lambda+1}} - 1\right]. \end{aligned} \tag{6.28}$$

So the original problem (6.24) reduces to minimize (6.28) subject to

$$\theta_i \le \phi_i, \quad i = 1, 2, \ldots, k-1.$$

The first partial derivatives of I^λ with respect to θ_l are found to be

$$\begin{aligned} \frac{\partial I^\lambda}{\partial \theta_l} &= \frac{1}{\lambda+1}\left[\sum_{i=1}^{l+1} \frac{\hat{\theta}_{i-1}^{\lambda+1} \prod_{j=i}^{k}(1+\theta_{j-1})^\lambda}{\theta_{i-1}^{\lambda} \prod_{j=i}^{k}(1+\hat{\theta}_{j-1})^{\lambda+1}} \frac{1}{1+\theta_l}\right. \qquad (6.29) \\ &\qquad \left. - \frac{\hat{\theta}_l^{\lambda+1} \prod_{j=l+1}^{k}(1+\theta_{j-1})^\lambda}{\theta_l^{\lambda+1} \prod_{j=l+1}^{k}(1+\hat{\theta}_{j-1})^{\lambda+1}}\right] \\ &= \frac{1}{\lambda+1} \frac{\sum_{j=l+1}^{k}(1+\theta_{j-1})^\lambda}{\sum_{j=l+1}^{k}(1+\hat{\theta}_{j-1})^{\lambda+1}} \frac{1}{1+\theta_l} \qquad (6.30) \\ &\qquad \left[\sum_{i=1}^{l} \frac{\hat{\theta}_{i-1}^{\lambda+1} \prod_{j=i}^{l}(1+\theta_{j-1})^\lambda}{\theta_{i-1}^{\lambda} \prod_{j=i}^{l}(1+\hat{\theta}_{j-1})^{\lambda+1}} - \left(\frac{\hat{\theta}_l}{\theta_l}\right)^{\lambda+1}\right]. \qquad (6.31) \end{aligned}$$

$l = 1, 2, \ldots, k-1$. It can be shown by calculation that the all second partial derivatives are all nonnegative. Suppose θ_i^* is the solution to the problem (6.28). It follows that θ_l^* depends only on $\theta_i^*, \ldots, \theta_i^*$ by

$$\sum_{i=1}^{l} \frac{\hat{\theta}_{i-1}^{\lambda+1} \prod_{j=i}^{l}(1+\theta_{j-1}^*)^{\lambda}}{\theta_{i-1}^{*\ \lambda} \prod_{j=i}^{l}(1+\hat{\theta}_{j-1})^{\lambda+1}} - \left(\frac{\hat{\theta}_l}{\theta_l^*}\right)^{\lambda+1} = 0, \tag{6.32}$$

$l = 1, 2, \ldots, k-1$. The following result follows from the above argument.

Theorem 6.2.2 *The optimal value of θ_i is given by*

$$\theta_l^* = \min(a_l^{1/(\lambda+1)}\hat{\theta}_l, \phi_l). \tag{6.33}$$

where

$$a_l = \sum_{i=1}^{l} \frac{\hat{\theta}_{i-1}^{\lambda+1} \prod_{j=i}^{l}(1+\theta_{j-1}^*)^{\lambda}}{\theta_{i-1}^{*\ \lambda} \prod_{j=i}^{l}(1+\hat{\theta}_{j-1})^{\lambda+1}}. \tag{6.34}$$

It can be shown that $a_1 = 1$ and

$$a_{l+1} = \left(a_l + \frac{\hat{\theta}_l^{\lambda+1}}{\theta_l^{*\lambda}}\right) \frac{(1+\theta_l^*)^{\lambda}}{(1+\hat{\theta}_l)^{\lambda+1}}, \quad l = 2, 3, \ldots, k-1. \tag{6.35}$$

6.2.5 Estimation under the Likelihood Ratio Ordering Constraint

Here we are interested in the following problem

$$\min_{\mathbf{p} \preceq_m \mathbf{q}} 2nI^{\lambda}(\hat{\mathbf{p}} : \mathbf{p}). \tag{6.36}$$

where $\mathbf{q}$ is a given probability vector. By letting $x_i = p_i/q_i$ and $y_i = \hat{p}_i/q_i$, the problem (6.36) reduces to

$$\min \sum_{i=1}^{k} \frac{x_i^{-\lambda}}{\lambda(\lambda+1)}, \quad \text{s.t } x_1 \geq x_2 \geq \cdots \geq x_k, \text{ and } \sum_{i=1}^{k} x_i q_i = 1. \tag{6.37}$$

First, consider the case $\lambda \neq 0$ and -1. Define

$$\Phi(x) = \frac{x^{-\lambda}}{\lambda(\lambda+1)}.$$

Then

$$\phi(x) = \frac{d}{dx}\Phi(x) = -\frac{x^{-(\lambda+1)}}{\lambda+1}$$

and

$$\phi^{-1}(x) = [-(\lambda+1)x]^{-1/(\lambda+1)}.$$

By Theorem 3.1 of Dykstra and Lee (1991), the solution for $\mathbf{x}$ is given by

$$x^* = \left[-(\lambda+1)E_{qy^{\lambda+1}}\left(-\frac{\psi_0 q}{qy^{\lambda+1}}|\mathcal{D}\right)\right]^{-1/(\lambda+1)} \tag{6.38}$$

where ψ_0 is a constant such that

$$\sum_{i=1}^{k} q_i x_i^* = 1. \tag{6.39}$$

Suppose $\lambda+1 < 0$. Since $\mathbf{x}^* \geq 0$, the coefficient of $1/y^{\lambda+1}$ in the projection of (6.38) must be positive. By the identity

$$E_{\mathbf{w}}(\alpha\mathbf{f}|\mathcal{D}) = \alpha E_{\mathbf{w}}(\mathbf{f}|\mathcal{D}) \ \text{ if } \alpha > 0,$$

and a theorem of Robertson (1966),

$$E_{\mathbf{h}}(\mathbf{f}/\mathbf{h}|\mathcal{D}) = E_{\mathbf{f}}(\mathbf{h}/\mathbf{f}|\mathcal{A})^{-1}, \text{ for positive } \mathbf{f} \text{ and } \mathbf{h}$$

one obtains that

$$\begin{aligned} x^* &= \left[-\psi_0(\lambda+1)E_{qy^{\lambda+1}}\left(\frac{\mathbf{q}}{\mathbf{q}\mathbf{y}^{-(\lambda+1)}}|\mathcal{D}\right)\right]^{-1/(\lambda+1)} \\ &= [\psi_0(\lambda+1)]^{-1/(\lambda+1)} E_q\left(\mathbf{y}^{\lambda+1}|\mathcal{A}\right)^{1/(\lambda+1)}. \end{aligned} \tag{6.40}$$

Therefore, by (6.39), one obtains

$$[\psi_0(\lambda+1)]^{1/(\lambda+1)} = \sum_{i=1}^{k} q_i E_q \left(\mathbf{y}^{\lambda+1}|\mathcal{A}\right)_i^{1/(\lambda+1)}.$$

It follows that

$$\mathbf{p}^* = \mathbf{q}\frac{E_{\mathbf{q}}(\mathbf{y}^{\lambda+1}|\mathcal{A})^{1/(\lambda+1)}}{\sum_{i=1}^{k} q_i E_{\mathbf{q}}(\mathbf{y}^{\lambda+1}|\mathcal{A})_i^{1/(\lambda+1)}}. \tag{6.41}$$

By symmetry, one obtains

$$\mathbf{p}^* = \mathbf{q}\frac{E_{\mathbf{q}}(\mathbf{y}^{\lambda+1}|\mathcal{D})^{1/(\lambda+1)}}{\sum_{i=1}^{k} q_i E_{\mathbf{q}}(\mathbf{y}^{\lambda+1}|\mathcal{D})_i^{1/(\lambda+1)}}, \tag{6.42}$$

for $\lambda + 1 > 0$. The case $\lambda = 0$ can be easily handled by continuity argument in (6.41) or (6.42).

For the case $\lambda = -1$, the original problem is equivalent to minimizing

$$\sum_{i=1}^{k} \hat{p}_i \ln x_i$$

subject to $x_1 \geq x_2 \geq \cdots \geq x_k$ and $\sum_{i=1}^{k} x_i q_i = 1$. The solution can be obtained similarly as above and found to be

$$\mathbf{p}^{(\lambda)} = \mathbf{q}\frac{\exp\{E_{\mathbf{q}}(\ln \mathbf{y}|\mathcal{D})\}}{\sum_{i=1}^{k} q_i \exp\{E_{\mathbf{q}}(\ln \mathbf{y}|\mathcal{D})_i\}}. \tag{6.43}$$

Therefore, we have proved the following result.

Theorem 6.2.3 *The solution for* $\mathbf{p}$ *that minimizes (6.36) such that* $\mathbf{p} \preceq_m \mathbf{q}$ *is given by (6.41), (6.43) and (6.42) for* $\lambda < -1$, $\lambda = -1$, *and* $\lambda > -1$, *respectively.*

Remark: Clearly the above argument in the proof of Theorem 6.2.3 still holds for the more general case that $\mathbf{p}/\mathbf{q}$ lies in an isotonic cone. When $\mathbf{q}$ is the uniform multinomial parameter, the problem is reduced to the one solved by Dykstra and Lee (1991), (see Section 4.2).

6.3 Maximum Likelihood Estimates of Order Restricted Multinomial Parameters

The maximum likelihood estimates of order-restricted multinomial parameters can be obtained by letting $\lambda = 0$ in the previous section. Let $\mathcal{D} = \{\mathbf{x} : x_1 \geq x_2 \cdots \geq x_k\}$ be defined as before and $\mathbf{q}$ be a given multinomial parameter.

$\mathbf{p} \in \mathcal{D}$. By (6.9), the MLE of $\mathbf{p}$ is given by

$$\mathbf{p}^* = E(\hat{\mathbf{p}}|\mathcal{D}). \tag{6.44}$$

$\mathbf{p} \preceq_m \mathbf{q}$. Since

$$\sum_{i=1}^{k} q_i E_{\mathbf{q}}(\mathbf{y}|\mathcal{D})_i = \sum_{i=1}^{k} q_i E_{\mathbf{q}}(\hat{\mathbf{p}}/\mathbf{q}|\mathcal{D})_i = \sum_{i=1}^{k} q_i(\hat{\mathbf{p}}/\mathbf{q})_i = 1,$$

by (6.33) the MLE of $\mathbf{p}$ is given by

$$\mathbf{p}^* = \mathbf{q}E_{\mathbf{q}}(\hat{\mathbf{p}}/\mathbf{q}|\mathcal{D}). \tag{6.45}$$

$\mathbf{p} \preceq_{(+)} \mathbf{q}$. By inductive formula (6.35) one can show that $a_l = 1$, $l = 1, 2, \ldots, k$. Therefore, by (6.33) the MLE of $\boldsymbol{\theta}$ is given by

$$\theta_l^* = \min(\hat{\theta}_l, \phi_l), \tag{6.46}$$

where

$$\theta_i = \frac{p_{i+1}}{\sum_{j=1}^{i} p_j}; \quad \hat{\theta}_i = \frac{\hat{p}_{i+1}}{\sum_{j=1}^{i} p_j}; \quad \phi_i = \frac{q_{i+1}}{\sum_{j=1}^{i} q_j}, \tag{6.47}$$

$i = 1, 2, \ldots, k-1$.

$\mathbf{p} \preceq_s \mathbf{q}$. By (6.19), the MLE of $\mathbf{p}$ is given by

$$\mathbf{p}^* = \hat{\mathbf{p}}E_{\hat{\mathbf{p}}}(\mathbf{q}/\hat{\mathbf{p}}|\mathcal{D}) \tag{6.48}$$

Once again, by comparing the above results with those in Section 3.3 we see that the orders of real-valued functions $\ll_m$, $\ll_{(+)}$ ($\ll_{(-)}$) and $\ll_s$ are closely related with the orders of random variables $\preceq_m$, $\preceq_{(+)}$ ($\preceq_{(-)}$) and $\preceq_s$.

6.4 Estimation under Other Ordering Constraints

Estimates under some other constraints may be obtained by Theorem 4.4.1 as illustrated by the following examples.

Example 6.4.1 Let $\mathbf{q}$ be the uniform multinomial parameter. Then $\mathbf{p} \preceq_{(-)} \mathbf{q}$ if and only $(p_1 + p_2 + \ldots + p_i)/i \geq p_{i+1}$.

Example 6.4.2 Let $\mathbf{q}$ be the uniform multinomial parameter. Then $\mathbf{p} \preceq_s \mathbf{q}$ if and only $(p_1 + \cdots + p_i)/i \geq (p_{i+1} + \cdots + p_k)/(k-i)$.

Example 6.4.3 Let $\mathbf{q} = (1/2, 0, \ldots, 0, 1/2)$. Then $\mathbf{p} \leq \mathbf{q}$ if and only $p_1 + \cdots + p_i \geq p_i + \cdots + p_k$.

Chapter 7

Nonparametric Estimation of Bounded Survival Functions with Censored Observations

7.1 Introduction

Stochastic ordering between survival functions is a very important concept. It arises in numerous settings and has many useful applications. For example, Agresti (1974) and Bhattacharjee (1987) considered the problem of finding appropriate stochastic bounds for the time of extinction in some branching processes. Examples of the lower bounds and upper bounds for some test statistics in order restricted inferences can be found in Robertson, Wright and Dykstra (1988, p.141). When such orderings exist, it is desirable to recognize their occurrence and to model distributional structure under such orderings. Nevertheless, estimates of the survival functions may not bear out such properties because of the inherent variability of the observations. The literature on estimation problems involving stochastic ordering is extensive. Brunk, Franck, Hanson and Hogg (1966) obtained nonparametric maximum likelihood estimates (MLE) of two stochastically ordered distribution functions and studied their properties. Dykstra (1982) considered a similar problem with right-censored data. Feltz and Dykstra (1985) proposed an iterative algorithm to find NMLE's of more than two survival functions subject to linear stochastic ordering restrictions. Lee (1987) discussed the MLE's for stochastically ordered multinomial populations with fixed and random zeros. Robertson and Wright (1974), and Sampson and Whitaker (1989) considered stochastic orderings in higher dimensions.

In this chapter we consider the problem of estimating a survival function that is stochastically bounded both from below and from above, with right-censored data. In Section 7.2 we introduce some notations and extend the one-sided problems considered by Dykstra (1982). In Section 7.3 we derive the two-sided problem and propose an iterative algorithm to find estimates in finite steps, usually two or three steps. An example involving survival times for heart transplant patients which appeared in Crowley and Hu (1977) is given in Section 7.4 to illustrate the proposed algorithm. In Section 7.5 a simulation study is conducted to investigate the increase in efficiency obtained by using the stochastic bounded constraints.

7.2 Notation and the One-sided Problem

Suppose independent observations are taken from a distribution on the positive real line with survival function $P(t)$ and complete observations (deaths) occur on a subset of the times $S_1 < S_2 < \cdots < S_m$ ($S_0 = 0$ and $S_{m+1} = \infty$ for convenience). Let d_j denote the number of deaths at S_j and l_j denote the number of censored observations (losses) in the interval $[S_j, S_{j+1})$, assumed to occur at $L_i^{(j)}$, $i = 1, \ldots, l_j$. Let $n_j = \sum_{i=j}^{m}(d_i + l_i)$, the number of items surviving just prior to S_j. We assume that the censoring times are fixed, although the method also works with independent random censoring times, see Dykstra (1982).

Proceeding as in Johansen (1978), we can obtain the generalized MLE of the survival function (in the class of all univariate distributions) by finding the survival function $P(\cdot)$ that maximizes

$$\prod_{i=1}^{l_0} P(L_i^{(0)}) \times \prod_{j=1}^{m} \left\{ [P(S_j-) - P(S_j)]^{d_j} \prod_{i=1}^{l_j} P(L_i^{(j)}) \right\}. \tag{7.1}$$

This problem is equivalent to the one that maximizes

$$\prod_{j=1}^{m} [P(S_{j-1}) - P(S_j)]^{d_j} P(S_j)^{l_j} \tag{7.2}$$

if $P(t)$ is a right continuous step function. The unrestricted solution to (7.2) is given by

$$\hat{P}(S_j) = \prod_{i=1}^{j} (1 - \frac{d_i}{n_i}), \quad j = 1, 2, \ldots, m, \tag{7.3}$$

the well known Kaplan-Meier product limit estimates (K-M estimates).

Let $Q(\cdot)$ be a given survival function. We are interested in maximizing (7.2) under one of the following four constraints

(I) $P(S_j) \geq Q(S_j)$, $j = 1, 2, ..., m-1$, and $P(S_m) = Q(S_m)$;

$(I0)$ $P(S_j) \geq Q(S_j)$, $j = 1, 2, ..., m$;

(II) $P(S_j) \leq Q(S_j)$, $j = 1, 2, ..., m-1$, and $P(S_m) = Q(S_m)$;

$(II0)$ $P(S_j) \leq Q(S_j)$, $j = 1, 2, ..., m$.

As in Dykstra (1982), it is required to solve equations of the form

$$\prod_{j=a}^{b} \left(1 - \frac{d_j}{n_j + y}\right) = \frac{Q(S_b)}{Q(S_{a-1})}. \tag{7.4}$$

Let y_{ab} be the solution to (7.4) which lies between $\max_{a \leq j \leq b}(d_j - n_j)$ and the extended real number $+\infty$ so that each individual term on the left hand side of (7.4) is a value between 0 and 1. The following result is an extension of Dykstra's (1982) one-sided problem and its proof can be found in Section 7.6.

Theorem 7.2.1 *The solution to (7.2) is given by*

$$P^*(S_j) = \prod_{i=1}^{j}\left(1 - \frac{d_i}{n_i + y_i^*}\right), \quad j = 1, 2, \ldots, m \tag{7.5}$$

where

$y_i^* = \min_{a\leq i}\max_{b\geq i} y_{ab} = \max_{b\geq i}\min_{a\leq i} y_{ab}$, *for the constraint (I);*

$y_i^* = \min_{a\leq i}\max_{b\geq i} y_{ab}^+ = \max_{b\geq i}\min_{a\leq i} y_{ab}^+$, *for the constraint (I0);*

$y_i^* = \max_{a\leq i}\min_{b\geq i} y_{ab} = \min_{b\geq i}\max_{a\leq i} y_{ab}$, *for the constraint (II);*

$y_i^* = \max_{a\leq i}\min_{b\geq i} y_{ab}^- = \min_{b\geq i}\max_{a\leq i} y_{ab}^-$, *for the constraint (II0)*

with $y_{ab}^+ = \max\{y_{ab}, 0\}$ *and* $y_{ab}^- = \min\{y_{ab}, 0\}$.

The estimates subject to the one-sided constraints are still in the form of the K-M estimates and only require adjustment on n_i's. This remains the same for the two-sided problem considered in the next section. For a heuristic AI interpretation of the adjusting constants y_i^*'s, see Dykstra (1982). The values of y_i^*'s for the constraint (I) can be computed by the maximum lower sets type algorithm (MLSTA) as follows.

0. Set $r = 0$ and $i_0 = 0$.

1. Let i_{r+1} be the largest index j such that $y_{i_r+1,j} = \max_{i_r+1\leq i\leq m} y_{i_r+1,i}$. Set $y_i^* = y_{i_r+1,i_{r+1}}$, $i = i_r + 1, \ldots, i_{r+1}$.

2. Replace r by $r + 1$ and go to Step 1 if $i_r < m$.

To compute $\mathbf{y}^*$ for the constraint (II) we replace maximum in Step 1 by minimum.

Consider the case of no censored observations. Then $n_{i+1} = n_i - d_i$, $\hat{P}(S_m) = 0$, and by (7.4)

$$y_{ab} = \frac{n_a Q(S_b) - n_{b+1} Q(S_{a-1})}{Q(S_{a-1}) - Q(S_b)}. \tag{7.6}$$

Clearly, $y_{im} \geq 0$ and so the constraints (I) and $(I0)$ are equivalent. The equivalence of (II) and $(II0)$ can also be attained by setting $Q(S_m)$ to be 0. It suffices to consider the constraint (I).

Let $\mathbf{f} = (f_1, f_2, \ldots, f_m)$, $\mathbf{d} = (d_1, d_2, \ldots, d_m)$ and $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ such that $f_i = Q(S_{i-1}) - Q(S_i)$ and $z_i = f_i/d_i$ and let $\mathbf{z}^* = E_{\mathbf{d}}(\mathbf{z}|I)$ with $I = \{\mathbf{x} \in R^m : x_1 \leq x_2 \leq \cdots \leq x_m\}$. Then

$$z_i^* = \max_{a \leq i} \min_{b \geq i} \frac{Q(S_{a-1}) - Q(S_b)}{n_a - n_{b+1}}. \tag{7.7}$$

This closed form expression appeared in Robertson and Wright (1981). The values of z_i^*'s can be calculated by the standard algorithms of the isotonic regression, such as the pool-adjacent-violators algorithm (PAVA), maximum lower sets algorithm or min-max formula, see Robertson et al. (1988). Let $\alpha \leq \beta$ be two indices such that $y_{\alpha-1}^* > y_\alpha^* = \cdots = y_\beta^* > y_{\beta+1}^*$. Then from the proof of Theorem 7.2.1, one obtains $P^*(S_{\alpha-1}) = Q(S_{\alpha-1})$, $P^*(S_\beta) = Q(S_\beta)$ and $y_\alpha^* = \cdots = y_\beta^* = y_{\alpha\beta}$. It can be shown that

$$z_{\alpha-1}^* < z_\alpha^* = \cdots = z_\beta^* = \frac{Q(S_{\alpha-1})}{n_\alpha + y_{\alpha\beta}} = \frac{Q(S_\beta)}{n_{\beta+1} + y_{\alpha\beta}} = \frac{Q(S_{\alpha-1}) - Q(S_\beta)}{n_\alpha - n_{\beta+1}} < z_{\beta+1}^*. \tag{7.8}$$

The isotonic regression $\mathbf{z}^*$ under constraint (II) is obtained by using $\mathbf{z}^* = E_{\mathbf{d}}(\mathbf{z}|D)$ with $D = \{\mathbf{x} \in R^m : x_1 \geq x_2 \cdots \geq x_m\}$. The solution (7.5) can also be expressed by $P^*(S_j) = 1 - \sum_{j=1}^{i} z_j^* d_j$.

7.3 The Two-Sided Problem

We are interested in the problem of maximizing

$$\prod_{j=1}^{m} [P(S_{j-1}) - P(S_j)]^{d_j} P(S_j)^{l_j}$$

subject to the constraints

$$Q(S_j) \leq P(S_j) \leq R(S_j), \quad j = 1, 2, \ldots, m \tag{7.9}$$

where Q and R are two given survival functions. The problem can be solved normally in two or three steps. In the first step, we use the one that solves Dykstra's (1982) one-sided problem and the constraint ($I0$) is used. We then partition the problem according to the levels of $\mathbf{y}^*$ obtained in Step 1 and we readjust the upper bound R so that $R(S_\alpha) = Q(S_\alpha)$ if α is the last index of a level of y^*. For each partitioned problem in Step 2, the constraint (II), $P(S_j) \leq R(S_j)$, $j = \alpha_1 + 1, \ldots, \alpha_2$ and $P(S_{\alpha_2}) = R(S_{\alpha_2}) = Q(S_{\alpha_2})$, is used except for the last partitioned problem. The latter requires constraint ($II0$) instead. In Step 3 we repeat Step 2 for a lower bound $Q(S_j) \leq P(S_j)$ problem.

Algorithm: Iterative Partitioning Proportional Fitting (IPPF)

Denote $Q_i = Q(S_i)$, $P_i = P(S_i)$, $R_i = R(S_i)$.

Step 0 Set $r = 0$, $A^0 = \{0, m+1\}$, $R^0 = R$ and let $y^t_{m+1} = 0$ for any positive integer t.

Step 1 Let $Q^{2r+1}_\alpha = R^{2r}_\alpha$ for each $\alpha \in A^{2r}$. For two consecutive indices $\alpha_1 < \alpha_2$ in A^{2r}, let $Q^{2r+1}_i = Q_i \vee Q^{2r+1}_{\alpha_2}$ for $\alpha_1 < i < \alpha_2$. Let y^{2r+1}_{st} denote the constant y

which solves the equation

$$\prod_{j=s}^{t}\left(1-\frac{d_j}{n_j+y}\right)=\frac{Q_t^{2r+1}}{Q_{s-1}^{2r+1}}.$$

Apply the maximum lower sets type algorithm (MLSTA) to the subset $\{\alpha_1+1,\ldots,\alpha_2\}$ to calculate

$$y_i^{2r+1}=\min_{\{\alpha_1<s\leq i\}}\max_{\{i\leq t\leq\alpha_2\}}y_{st}^{2r+1}.$$

For the case $\alpha_2=m+1$, replace y_i^{2r+1} by 0 if $y_i^{2r+1}<0$. Set

$$A^{2r+1}=A^{2r}\cup\{\alpha:y_\alpha^{2r+1}>y_{\alpha+1}^{2r+1}\}.$$

Step 2 Let $R_\alpha^{2r+2}=Q_\alpha^{2r+1}$ for each $\alpha\in A^{2r+1}$. For two consecutive indices $\alpha_1<\alpha_2$ in A^{2r+1}, let $R_i^{2r+2}=R_i\wedge R_{\alpha_1}^{2r+2}$ for $\alpha_1<i<\alpha_2$. Let y_{st}^{2r+2} denote the constant y which solves the equation

$$\prod_{j=s}^{t}\left(1-\frac{d_j}{n_j+y}\right)=\frac{R_t^{2r+2}}{R_{s-1}^{2r+2}}.$$

Apply the minimum lower sets type algorithm (MLSTA) to the subset $\{\alpha_1+1,\ldots,\alpha_2\}$ to calculate

$$y_i^{2r+2}=\max_{\{\alpha_1<s\leq i\}}\min_{\{i\leq t\leq\alpha_2\}}y_{st}^{2r+2}.$$

For the case $\alpha_2=m+1$, replace y_i^{2r+2} by 0 if $y_i^{2r+2}>0$. Set

$$A^{2r+2}=A^{2r+1}\cup\{\alpha:y_\alpha^{2r+2}<y_{\alpha+1}^{2r+2}\}.$$

Step 3 Replace r by $r+1$ and go to step 1 until $A^{2r+2}=A^{2r+1}$.

Clearly, the IPPF algorithm converges in finite steps, usually two or three steps, but at most m. Let $P_j^t = \prod_{i=1}^{j}(1 - d_i/(n_i + y_i^t))$. They are the projections obtained at the step t and these values can provide us with information on the computations in the next step. The proof of the following theorem can be found in Section 7.6.

Theorem 7.3.1 *Let* $V^{2r+1} = A^{2r} \cup \{i : P_i^{2r} < Q_i^{2r+1}\}$ *and let* $V^{2r+2} = A^{2r+1} \cup \{i : P_i^{2r+1} > R_i^{2r+2}\}$. *Then* $A^t \subset V^t$.

Suppose that $\alpha_1 < \alpha_2$ are two consecutive indices in A^t. Then one computes $y_{\alpha+1,\beta}^{t+1}$ for each index β, $\alpha_1 < \beta < \alpha_2$, belonging to V^{t+1}. If there does not exist such an index β then P_i^t, $i = \alpha_1 + 1, \ldots, \alpha_2$ is the desired solution, and $y_i^l = y_i^t$ remain constant for $l \geq t$. The utility of the IPPF algorithm lies in the following theorem which is proved in Section 7.6.

Theorem 7.3.2 *Let* $\mathbf{y}^*$ *be the values obtained at the last step of the IPPF algorithm. Then the survival function*

$$P^*(S_j) = \prod_{i=1}^{j}\left(1 - \frac{d_i}{n_i + y_i^*}\right), \quad j = 1, 2, \ldots, m \tag{7.10}$$

is the solution to the problem (7.9).

An illustration of the IPPF algorithm is given in the next section.

7.4 Example

The IPPF algorithm is an iterative procedure such that at each step one needs only to solve several non-overlapping one-sided problems. For illustration, we consider the data which appeared in Crowley and Hu (1977). It consists of survival times for patients who had heart implants in the Stanford Heart Transplantation Program, and it includes censored observations of people who were still alive by the closing date for data collection, April 1, 1974. We wish to estimate the survival function of the post-transplant time T (in days).

Turnbull, Brown and Hu (1974) noticed that the accepted candidates into the program may come from a mixture of two populations, namely, "regular" and "hardy" patients. Suppose

$$P(t) = w\exp\{-t/\mu_1\} + (1-w)\exp\{-t/\mu_2\}$$

where $\mu_1 > \mu_2 > 0$ and $0 \le w \le 1$. The MLE's of these parameters based on 69 observations in Table 2 are found to be $\mu_1 = 1513$, $\mu_2 = 55.86$ and $w = 0.5626$. We consider the lower bound and upper bound of the unknown survival function to be

$$Q(t) = 0.45\exp\{-t/1500\} + 0.55\exp\{-t/55\}, \tag{7.11}$$

$$R(t) = 0.65\exp\{-t/1500\} + 0.35\exp\{-t/55\}. \tag{7.12}$$

We first illustrate the IPPF algorithm by the simple case of latent times, censored observations as well as uncensored observations, since in this case the computation of the solution at each step is simple. The latent times are grouped into nine classes as in Table 1. When P, Q and R are all having finite support by assuming $P(1792) = Q(1792) = R(1792) = 0$, A^0 in Step 0 of the IPPF algorithm is replaced by $A^0 = \{0, m\}$ and constraints ($I0$) and ($II0$) need not be used.

In the first step, the one-sided problem (I) is solved on the whole set $\{1, 2, \ldots, 9\}$. Let $f_i = Q_{i-1} - Q_i$. The values of z_i^1's, $i = 1, 2, \ldots, 9$, in (2.7) are obtained by the monotone increasing regression of $\mathbf{f}/\mathbf{d}$ with weight $\mathbf{d}$ using the minimum lower sets algorithm or the PAVA. Since $V^1 = \{0, 8, 9\}$, the monotone regression has two levels, $z_1^1 = \ldots = z_8^1 = .0125$ and $z_9^1 = .0275$, ($y_1^1 = \ldots = y_8^1 = 10.747$, $y_9^1 = 0.000$). In the second step, the upper bound is first adjusted at $i = 8$ by setting its values equal to the lower bound, $R_8^2 = Q_8^1 = .2476$ ($R_8^0 = .3577$). Then the one-sided problem (II) is applied to each subset of $\{1, 2, \ldots, 8\}$ and $\{9\}$. Let $f_i = R_{i-1}^2 - R_i^2$. The value of z_9^2 remains the same while the values of z_i^2's, $i = 1, 2, \ldots, 8$, are the monotone decreasing regression of $\mathbf{f}/\mathbf{d}$ with weight $\mathbf{d}$. Since $V^2 = \{0, 3, 8, 9\}$, the subset $\{1, 2, \ldots, 8\}$ is portioned into two sub-subsets, namely, $z_1^2 = z_2^2 = z_3^2 = .0126$ and $z_4^2 = \ldots = z_8^2 = .0125$, ($y_1^2 = y_2^2 = y_3^2 = 10.126$ and $y_4^2 = \ldots = y_8^2 = 10.786$). Similar procedures follow in the third step and it gives no new partition. Since the projection P_i^2 satisfies the restriction (7.9), it is the solution (7.10) in Theorem 7.3.2.

In the above case with no censored observations, the estimate can also be obtained by an algorithm proposed by Parnami, Singh and Puri (1993). However, not only is the IPPF algorithm much more efficient, but also it can be applied to problems with censored observations.

To apply the IPPF algorithm with censored observations, one needs to compute the values of y_i's by solving equation (7.4) and the min-max type formulas in Theorem 7.1. Table 2 contains a list of the original data, the adjusted bounds and the values of the projection at each step. The bound restricted estimate is obtained in three steps. In the first step, the one-sided problem ($I0$) is solved on the whole set $\{1, 2, \ldots, 42\}$. The set $V^1 = \{0, 1, 2, 3, 24, 25, 43\}$ and the values of y_i's are $y_1^1 = +\infty$, $y_2^1 = 29.95$, $y_3^1 = \ldots = y_{25}^1 = 0.055$, and $y_{26}^1 = \cdots = y_{42}^1 = 0$. It means y_{1j}'s are finite, $j = 2, 3, \ldots, 42$; $y_{2,2}$ is the largest among $y_{2,2}$, $y_{2,3}$, $y_{2,24}$ and $y_{2,25}$; $y_{3,25}$ is the largest among $y_{3,3}$, $y_{3,24}$ and $y_{3,25}$ while

$y_{26,j}$'s are non-positive. It also means that the left hand side of (7.4) with $y = y_{2,2}$ (say) is larger than the right hand side when $a = 2$, $b = 3, 24, 25$ (or $b = 3, 4, \ldots, 42$). In the second step, the upper bound is first adjusted by setting $R_1^2 = Q_1^1 = 1.0$ ($R_1^0 = 1.0$), $R_2^2 = Q_2^1 = .990$ ($R_2^0 = 0.933$), and $R_{25}^2 = \cdots = R_{29}^2 = Q_{25}^1 = .590$ ($R_{25}^0 = .732$, $R_{26}^0 = .632$, $R_{27}^0 = .623$, $R_{28}^0 = .613$, $R_{29}^0 = .603$), and then the one-sided problems (Π) and ($\Pi 0$) are applied to the subsets $\{3, \ldots, 25\}$ and $\{26, \ldots, 42\}$, respectively. For the subset $\{3, \ldots, 25\}$, its intersection with V^2 yields $\{4, 5, \ldots, 15\}$. The calculation yields $y_3^2 = y_4^2 = -34.86$, $y_5^2 = y_6^2 = y_7^2 = -13.13$, $y_8^2 = \ldots = y_{13}^2 = 4.04$ and $y_{14}^2 = \ldots = y_{25}^2 = 6.53$. It means $y_{3,4}$ is the smallest among $y_{3,4}, y_{3,5}, \ldots, y_{3,15}, y_{3,25}$; $y_{5,7}$ the smallest among $y_{5,5}, y_{5,6}, \ldots, y_{5,15}, y_{5,25}$; $y_{8,13}$ the smallest among $y_{8,8}, y_{8,9}, \ldots, y_{8,15}, y_{8,25}$; $y_{14,25}$ the smallest among $y_{14,14}, y_{14,15}, y_{14,25}$. It also means that the right hand side of (7.4)) with $y = y_{3,4}$ (say) is less than the right hand side when $a = 3$, $b = 5, 6, \ldots, 15, 25$ (or $b = 5, 6, \ldots, 25$). For the subset $\{26, \ldots, 42\}$, the values of y_i^2's remain the same since there are no violators in this partition. In the third step, the lower bound is first adjusted by setting $Q_4^3 = R_4^2 = .926$, $Q_7^3 = R_7^2 = .870$ and $Q_{13}^3 = R_{13}^2 = .788$, with possible adjustment of the other values to maintain the monotonicity of the survival functions. The one-sided problem (I) is then applied to the subsets $\{3, 4\}$, $\{5, 6, 7\}$, $\{8, \ldots, 13\}$ and $\{14, \ldots, 25\}$ and only the values on the subset $\{3, 4\}$ need to be considered. The new y_i's are $y_3^3 = -16.23$ and $y_4^3 = -42.80$. Since the projection P_i^3 satisfies the restriction (7.9), it is the solution (7.10) in Theorem 7.3.2.

The Kaplan-Meier product limit estimate (7.3) and the bounded estimate (7.10) are also plotted in Figure 1 along with the lower bound $Q(t)$ and the upper bound $R(t)$ in (7.11) and (7.12), respectively.

7.5 A Simulation Study

A simulation study was performed to investigate how the constraints affect the efficiency of the estimation. We consider four sampling survival functions with a series of stochastic bounds for each of the four cases. The root of the mean square errors (MSE's) of the restricted and unrestricted (K-M) estimates of selected right tail probabilities are calculated based on 10,000 iterations of the simulations with sample sizes 100 and 300.

In the Cases I and II, the sampling survival functions are the generalized maximum modulus introduced by Lee (1996a) in constructing Tukey-type confidence bands for monotone regressions. Let $Z_1, Z_2, \ldots, Z_k$ be independent standard normal variates. The generalized maximum modulus is defined to be

$$GM_k = \max_{1\leq i\leq j\leq k} \frac{|\sum_{h=i}^{j} Z_h|}{\sqrt{j-i+1}}$$

The survival function of GM_k is very complicated and selected percentiles can be found in Lee (1996a). Clearly, GM_k is larger than $M_k = \max_{1\leq i\leq j} |Z_i|$, the maximum modulus with survival function $1-(2\Phi(t)-1)^k$. On the other hand, by the Cauchy's Inequality, one obtains

$$GM_k \leq \sqrt{Z_1^2 + Z_2^2 + \cdots + Z_k^2},$$

where the right hand side is the square root of a chi-square random variable with degrees of freedom k. Therefore, $1-(2\Phi(t)-1)^k$ and $1-\chi_k^2(t^2)$ form a pair of lower and upper bands for the survival function $GM_k(t)$ for any positive real t. Sharper stochastic bounds for GM_k can also be found.

Case I. Sampling survival function: $GM_2(t)$; and its lower bound and upper bound given by

B1: $1-(2\Phi(t)-1)^2$, $1-0.5\chi_2^2(t^2)-0.5(2\Phi(t)-1)^2$;

B2: $1-(2\Phi(t)-1)^2$, $1-\chi_2^2(t^2)$;

B3: $1-(2\Phi(t)-1)^2$, no upper bound.

Case II. Sampling survival function: $GM_3(t)$; and its lower bound and upper bound given by

B1: $1-(2\Phi(t)-1)^3$, $1-0.5\chi_2^2(t^2)-0.25(2\Phi(t)-1)^3-0.25\chi_2^2(t^2)(2\Phi(t)-1)$;

B2: $1-(2\Phi(t)-1)^3$, $1-\chi_3^2(t^2)$;

B3: $1-(2\Phi(t)-1)^3$, no upper bound.

Case III. Sampling survival function: exp(-t); and its lower bound and upper bound given by

B1: $\exp(-t/0.8)$, $\exp(-t/1.2)$;

B2: $\exp(-t/0.8)$, $\exp(-t/1.5)$;

B3: $\exp(-t/0.8)$, no upper bound.

Case IV. Sampling survival function: $0.5\exp(-t)+0.5\exp(-t/10)$; and its lower bound and upper bound given by

B1: $0.7\exp(-t)+0.3\exp(-t/10)$, $0.3\exp(-t)+0.7\exp(-t/10)$;

B2: $0.7\exp(-t)+0.3\exp(-t/10)$, $0.1\exp(-t)+0.9\exp(-t/10)$;

B3: $0.7\exp(-t)+0.3\exp(-t/10)$, no upper bound.

The results of our simulation study are provided in Table 7.3. We are interested in estimating the survival function $P(t)$ at the four points, $t=P^{-1}(0.50)$, $P^{-1}(0.25)$, $P^{-1}(0.10)$ and $P^{-1}(0.05)$. In general, utilizing the prior knowledge of a lower bound and an upper bound may reduce the point-wise MSE's. The amount of reduction in

MSE's could be substantial for small and moderate sample sizes for a pair of sharp bounds $B1$. For example, when the sample size is 100 and the upper tail probability is 0.05, i.e., $P(t) = 0.05$, the root MSE's are reduced from .0219, .0218, .0219, .0220 (0.0218 for the exact standard deviation) to .0089, .0121, .0159, .0122 respectively for the four cases when the bounds $B1$ were used. In the Case I, the MSE for $B1$ is no more than 1/6 of that for the Kaplan-Meier product estimate when $n = 100$ and no more than 1/5 when $n = 300$. When only a lower bound or an upper bound is given and when the sample sizes are small, the MSE's of the restricted estimates may be larger than those of the unrestricted ones as the lower bound $B3$ is used in the Case I with sample size 100 and upper tail probabilities 0.25, 0.10 and 0.05. However, once sample size increases from 100 to 300 the MSE's of the restricted estimates become smaller than those of the unrestricted ones.

7.6 Discussion

The purpose of this chapter is to introduce an efficient algorithm to compute the bounded NMLE of survival functions. It normally takes two or three steps of computation for the IPPF algorithm to converge to the exact solution. When there are no censored data, a closed form expression for the projection is available at each step, see (7.7) for odd steps. The bounded NMLE should prove to be useful, both as a descriptive tool and a primitive technique for any procedure requiring estimation of a survival function or a distribution function. One may hypothesize a lower bound and an upper bound for a survival function. Its validity may be verified by the supremum of the distance between the bounded NMLE and the Kaplan-Meier product estimate.

The bounded NMLE of the survival functions can substantially reduce point-wise MSE's for small or moderate sample sizes compared to the Kaplan-Meier product estimates. The reduction is optimal when the lower bound and the upper bound are approximately the same distance from the underlying survival function. For the data in Table 2, the bounded NMLE with the lower bound (7.11) and the upper bound (7.11) may be a better estimate.

7.7 Proof of the Main Theorems

Proof of Theorem 7.2.1. It suffices to prove the result under the constraint (I). The proofs under the other constraints are similar. Using Dykstra's (1982) notations, we shall let $p_i = \ln P_i - \ln P_{i-1}$ and $q_i = \ln Q_i - \ln Q_{i-1}$. The original problem is equivalent to maximize a concave function

$$\sum_{j=1}^{m}[d_j \ln(1-e^{p_j}) + (n_j - d_j)p_j] \tag{7.13}$$

over a closed convex region

$$\{\mathbf{p} : \mathbf{p} \le 0; \sum_{j=1}^{i} p_j \ge \sum_{j=1}^{i} q_j, i = 1, 2, \ldots, m-1 \text{ and } \sum_{j=1}^{m} p_j = \sum_{j=1}^{m} q_j\}.$$

The solution exists and it is unique. Let

$$\Psi = \sum_{j=1}^{m}[d_j \ln(1-e^{p_j}) + (n_j - d_j)p_j] + \sum_{i=1}^{m} u_i \left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} q_j\right).$$

By the Equivalence Theorem (see Kuhn and Tucker, 1951), $\mathbf{p} \le 0$ is the solution if and only if there exist nonnegative real numbers $u_1, u_2, \ldots, u_{m-1}$ and a real number u_m such that

$$\frac{\partial \Psi}{\partial p_i} = -d_i/(1-e^{p_i}) + n_i + \sum_{j=i}^{m} u_j = 0, \quad i = 1, 2, \ldots, m; \tag{7.14}$$

$$u_i(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} q_j) = 0, \quad i = 1, 2, \ldots, m-1;$$

$$\sum_{j=1}^{i} p_j \geq \sum_{j=1}^{i} q_j, \quad i = 1, 2, \ldots, m-1, \text{ and } \sum_{j=1}^{m} p_j = \sum_{j=1}^{m} q_j.$$

By letting $y_i = \sum_{j=i}^{m} u_j$, and $f_i(y_i) = \ln(1 - d_i/(n_i + y_i)) - q_i$, $i = 1, 2, \ldots, m$, we have from (7.14)

$$p_i = \ln(1 - \frac{d_i}{n_i + y_i}) = q_i + f_i(y_i) \leq 0 \tag{7.15}$$

and

$$y_1 \geq y_2 \geq \cdots \geq y_m; \tag{7.16}$$

$$(y_i - y_{i+1}) \sum_{j=1}^{i} f_j(y_i) = 0, \quad i = 1, \ldots, m-1 \tag{7.17}$$

$$\sum_{j=1}^{i} f_j(y_i) \geq 0, \quad i = 1, \ldots, m-1 \text{ and } \sum_{j=1}^{m} f_j(y_i) = 0. \tag{7.18}$$

Let $\alpha \leq \beta$ be such that $y_{\alpha-1} > y_\alpha = \ldots = y_\beta > y_{\beta+1}$. By (7.17), $\sum_{j=\alpha}^{\beta} f_j(y_j) = 0$ and then by (7.4), $y_j = y_{\alpha\beta}$ for $j = \alpha, \ldots, \beta$. By (7.18), $\sum_{j=\alpha}^{b} f_j(y_j) \geq 0$ and $\sum_{j=a}^{\beta} f_j(y_j) \leq 0$, for any $b \geq \alpha$ and $a \leq \beta$. By the monotonicity of functions $f_j(y)$,

$$\sum_{j=\alpha}^{b} f_j(y_{\alpha b}) = 0 \leq \sum_{j=\alpha}^{b} f_j(y_j) \leq \sum_{j=\alpha}^{b} f_j(y_{\alpha\beta});$$

$$\sum_{j=a}^{\beta} f_j(y_{\alpha\beta}) \leq \sum_{j=a}^{\beta} f_j(y_j) \leq 0 = \sum_{j=a}^{\beta} f_j(y_{a\beta}).$$

It follows that $y_{\alpha b} \leq y_{\alpha\beta} \leq y_{a\beta}$. Therefore, for i between α and β we have that

$$y_{\alpha\beta} = \max_{b \geq i} y_{\alpha b} \geq \min_{a \leq i} \max_{b \geq i} y_{ab} \geq \max_{b \geq i} \min_{a \leq i} y_{ab} \geq \min_{a \leq i} y_{a\beta} = y_{\alpha\beta}$$

where it is trivial to establish the second inequality. It follows that

$$y_i = \min_{a \leq i} \max_{b \geq i} y_{ab} = \max_{b \geq i} \min_{a \leq i} y_{ab}.$$

The proof is complete. □

Proof of Theorem 7.3.1. It suffices to show that if $\alpha \in A^{2r+1} - A^{2r}$ then $P_\alpha^{2r} < Q_\alpha^{2r+1}$. Let $\beta_1 < \beta_2$ be two consecutive indices in A^{2r} such that $\beta_1 < \alpha < \beta_2$. By (7.16),

$$y_{\beta_1+1}^{2r+1} \geq y_{\beta_1+2}^{2r+1} \geq \cdots \geq y_\alpha^{2r+1} > y_{\alpha+1}^{2r+1} \geq \cdots \geq y_{\beta_2}^{2r+1}.$$

By (7.4), (7.5) and the IPPF algorithm we have that $P_\alpha^{2r+1} = Q_\alpha^{2r+1}$ and $Q_{\beta_i}^{2r+1} = R_{\beta_i}^{2r} = P_{\beta_i}^{2r}$, $i = 1, 2$. By Theorem 7.2.1,

$$y_{\beta_1+1}^{2r+1} > y_j^{2r} = y_{\beta_1+1,\beta_2}^{2r+1} > y_{\beta_2}^{2r+1}, j = \beta_1 + 1, \ldots, \beta_2.$$

If $y_\alpha^{2r+1} \geq y_\alpha^{2r}$ then

$$Q_\alpha^{2r+1} = P_\alpha^{2r+1} = Q_{\beta_1}^{2r+1} \prod_{j=\beta_1+1}^{\alpha} \left(1 - \frac{d_j}{n_j + y_j^{2r+1}}\right) > P_{\beta_1}^{2r} \prod_{j=\beta_1+1}^{\alpha} \left(1 - \frac{d_j}{n_j + y_j^{2r}}\right) = P_\alpha^{2r}.$$

If $y_\alpha^{2r+1} < y_\alpha^{2r}$, then

$$Q_\alpha^{2r+1} = P_\alpha^{2r+1} = Q_{\beta_2}^{2r+1} / \prod_{j=\alpha+1}^{\beta_2} \left(1 - \frac{d_j}{n_j + y_j^{2r+1}}\right) > P_{\beta_2}^{2r} / \prod_{j=\alpha+1}^{\beta_2} \left(1 - \frac{d_j}{n_j + y_j^{2r}}\right) = P_\alpha^{2r}.$$

This completes the proof. □

We shall derive two preliminary results that characterize the IPPF algorithm before presenting the proof of Theorem 7.3.2

Lemma 7.7.1 *(a) If $\alpha \in A^{2r+1} - A^{2r}$, then $P_\alpha^l = Q_\alpha$, for $l \geq 2r + 1$ and $Q_\alpha^{2l+1} = Q_\alpha$, for $l \geq 0$;*

(b) If $\alpha \in A^{2r+2} - A^{2r+1}$, then $P_\alpha^l = R_\alpha$, for $l \geq 2r + 2$ and $R_\alpha^{2l} = R_\alpha$, for $l \geq 0$.

Proof. It suffices to prove (*a*). Let $\alpha \in A^{2r+1} - A^{2r}$, and let $\alpha_1 < \alpha_2$ be two consecutive indices in A^{2r} such that α is generated by fitting $P_i \geq Q_i^{2r+1}$, $i = \alpha_1 + 1, ..., \alpha_2$ with $P_{\alpha_2} = Q_{\alpha_2}^{2r+1}$ if $\alpha_2 \leq m$. By (7.17) and (7.15) and Theorem 7.3.1,

$$P_\alpha^{2r+1} = Q_\alpha^{2r+1} > P_\alpha^{2r} \geq P_{\alpha_2}^{2r} = R_{\alpha_2}^{2r} = Q_{\alpha_2}^{2r+1}.$$

Since $Q_\alpha^{2r+1} = Q_\alpha \vee Q_{\alpha_2}^{2r+1}$, it follows that $P_\alpha^{2r+1} = Q_\alpha^{2r+1} = Q_\alpha$. It is clear from the algorithm that the lower bound Q_α^{2r+1} and the projection P_α^{2r+1} for $\alpha \in A^{2r+1}$ will be fixed throughout the remaining iterations of the computation process and hence, $P_\alpha^l = Q_\alpha^l = Q_\alpha$, for $l \geq 2r+1$. Since Q_α^{2l+1} is nondecreasing in l, it follows that $Q_\alpha^{2l+1} = Q_\alpha$, for $l \geq 0$. The proof is complete. □

Lemma 7.7.2 *(a) If $\alpha \in A^{2r+1} - A^{2r}$, then $y_\alpha^{2r+1} \leq y_\alpha^l$, and $y_{\alpha+1}^{2r+1} \geq y_{\alpha+1}^l$, for $l \geq 2r+1$.*

(b) If $\alpha \in A^{2r+2} - A^{2r+1}$, then $y_\alpha^{2r+2} \geq y_\alpha^l$, and $y_{\alpha+1}^{2r+2} \leq y_{\alpha+1}^l$, for $l \geq 2r+2$.

Proof. It suffices to prove the case that if $\alpha \in A^{2r+1} - A^{2r}$, then $y_\beta^{2l+1} \geq y_\beta^{2l+2}$ and $y_\beta^{2l+1} \geq y_\beta^{2l+3}$ where $l \geq r$ and $\beta = \alpha + 1$. The β is the leading index of all partitioned problems it belongs to after the $(2r+1)$st step. Define

$$\beta_i = \inf\{t : t \geq \beta, t \in A^{2l+i}\}, \quad i = 1, 2, 3.$$

Then $\beta_1 \geq \beta_2 \geq \beta_3$. In the $(2l+2)$nd step we solve a partitioned problem $\beta, \beta+1, \ldots, \beta_1$ and in the $(2l+3)$rd step we solve a partitioned problem $\beta, \beta+1, \ldots, \beta_2$. Therefore, $y_{\beta\beta_1}^{2l+2} = y_{\beta\beta_1}^{2l+1}$ and $y_{\beta\beta_2}^{2l+3} = y_{\beta\beta_2}^{2l+2}$. By the min-max formulas,

$$y_\beta^{2l+i} = y_{\beta\beta_i}^{2l+i} \geq y_{\beta t}^{2l+i}, \quad i = 1 \text{ or } 3$$

if $\beta \leq t \leq \beta_2$ and

$$y_\beta^{2l+2} = y_{\beta\beta_2}^{2l+2} \leq y_{\beta t}^{2l+2}$$

if $\beta \leq t \leq \beta_1$. For the case $\beta_3 = \beta_2$, we have that

$$y_\beta^{2l+3} = y_{\beta\beta_3}^{2l+3} = y_{\beta\beta_2}^{2l+3} = y_{\beta\beta_2}^{2l+2} = y_\beta^{2l+2} \leq y_{\beta\beta_1}^{2l+2} = y_{\beta\beta_1}^{2l+1} = y_\beta^{2l+1}.$$

For the case $\beta_3 < \beta_2$, we have that

$$y_\beta^{2l+3} = y_{\beta\beta_3}^{2l+3} \geq y_{\beta\beta_2}^{2l+3} = y_{\beta\beta_2}^{2l+2} = y_\beta^{2l+2}$$

and

$$y_{\beta\beta_3}^{2l+1} \leq y_{\beta\beta_1}^{2l+1} = y_\beta^{2l+1}.$$

However, in this case we have that $\beta_3 \in A^{2l+3} - A^{2l+2}$. By Lemma 7.7.1) we have that $Q_{\beta_3}^{2l+3} = Q_{\beta_3}^{2l+1} = Q_{\beta_3}$ and $Q_\alpha^{2l+3} = Q_\alpha^{2l+1} = Q_\alpha$. It follows that $y_{\beta\beta_3}^{2l+1} = y_{\beta\beta_3}^{2l+3}$ and

$$y_\beta^{2l+1} \geq y_\beta^{2l+3} \geq y_\beta^{2l+2}.$$

By induction the proof is complete. □

Proof of Theorem 7.3.2. Similar to the proof of Theorem 7.2.1, the original problem is equivalent to maximize a concave objective function

$$\sum_{i=1}^{m} [d_i \ln(1 - e^{p_i}) + (n_i - d_i)p_i]$$

over a closed and bounded convex region $\mathbf{p} \leq 0$ and

$$\sum_{j=1}^{i} q_j \leq \sum_{j=1}^{i} p_j \leq \sum_{j=1}^{i} r_j, \quad i = 1, 2, \ldots, m.$$

The solution exists and it is unique. Let

$$\Psi = \sum_{i=1}^{m}[d_i \ln(1-e^{p_i})+(n_i-d_i)p_i]+\sum_{i=1}^{m} u_i\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} q_j\right)+\sum_{i=1}^{m} v_i\left(\sum_{j=1}^{i} r_j - \sum_{j=1}^{i} p_j\right).$$

By the Equivalence Theorem (see Kuhn and Tucker, 1951), $\mathbf{p} \leq 0$ is the optimal solution if and only if there exist $\mathbf{u} \geq 0$ and $\mathbf{v} \geq 0$ such that

$$(I) \quad p_i = \ln\left(1 - \frac{d_i}{n_i + \sum_{j=i}^{m} u_j - \sum_{j=i}^{m} v_j}\right);$$

$$(II) \quad u_i(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} q_j) = 0, \quad v_i(\sum_{j=1}^{i} r_j - \sum_{j=1}^{i} p_j) = 0$$

$$(III) \quad \sum_{j=1}^{i} q_j \leq \sum_{j=1}^{i} p_j \leq \sum_{j=1}^{i} r_j,$$

$i = 1, 2, \ldots, m$. Initially we define $u_i^0 = v_i^0 = 0$, $i = 1, 2, \ldots, m$. If $i \in A^{2r+1} - A^{2r}$, let $u_i^l = y_i^l - y_{i+1}^l$ for each $l \geq 2r+1$; if $i \in A^{2r+2} - A^{2r+1}$, then let $v_i^l = y_{i+1}^l - y_i^l$ for each $l \geq 2r+2$. It follows that $y_i^l = \sum_{j=i}^{m} u_j^l - \sum_{j=i}^{m} v_j^l$ and hence (I) is satisfied at each step of the IPPF algorithm. Consider the case $i \in A^{2r+1} - A^{2r}$. Then $v_i^l = 0$ for all $l \geq 0$ and $u_i^l = 0$ for $l < 2r+1$. For $l \geq 2r+1$, by Lemma 7.7.2 one obtains

$$u_i^l = y_i^l - y_{i+1}^l \geq y_i^{2r+1} - y_{i+1}^{2r+1} > 0.$$

By Lemma A.1,

$$\sum_{j=1}^{i} p_j^l = \sum_{j=1}^{i} \ln(1 - \frac{d_j}{n_j + y_j^l}) = \ln Q_i = \sum_{j=1}^{i} q_j.$$

Therefore, the first equation of Condition (II) is satisfied at each step l, as is the second. Similarly, one can prove for the case $i \in A^{2r+2} - A^{2r+1}$. Our procedure terminates no later than m steps when the condition (III) is satisfied because at each step the index subset A^l will have one or more new elements. This completes the proof. □

Table 7.1: Post-transplant Survival Times of Heart Transplant Patients and Their Estimates in Grouped Data

i	S_i	d_i	n_i	$\hat{P}_i$	Q_i^1	P_i^1	R_i^2	P_i^2
1	7	4	69	.9420	.9322	.9498	.9551	.9494
2	14	3	65	.8986	.8722	.9122	.9153	.9115
3	28	5	62	.8261	.7723	.8495>	.8483*	.8483
4	56	12	57	.6522	.6322	.6990	.7526	.6982
5	112	8	45	.5362	.4894	.5987	.6489	.5980
6	224	5	37	.4638	.3969	.5360	.5658	.5355
7	448	12	32	.2899	.3340	.3856	.4823	.3853
8	896	11	20	.1304<	.2476*	.2476	.2476	.2476
9	1792	9	9	.0000	.0000	.0000	.0000	.0000

* The end of new partition.

Table 7.2: Post-transplant Survival Times of Heart Transplant Patients and Their Estimates

i	S_i	d_i	l_i	n_i	$\hat{P}_i$	Q_i^1	P_i^1	R_i^2	P_i^2	Q_i^3	P_i^3
1	0	1	0	69	.986<	1.00*	1.00	1.00	1.00	1.00	1.00
2	1	1	1	68	.971<	.990*	.990	.990	.990	.990	.990
3	3	1	C	66	.956<	.970	.975	.980	.968<	.970*	.970
4	12	1	1	65	.942	.889	.960>	.926*	.926	.926	.926
5	14	1	0	63	.927	.872	.945>	.915	.908	.872	.908
6	15	1	0	62	.912	.864	.929>	.910	.889	.870	.889
7	23	1	0	61	.897	.805	.914>	.870*	.870	.870	.870
8	25	1	0	60	.882	.792	.889>	.861	.857	.792	.857
9	26	1	0	59	.867	.785	.884>	.857	.843	.788	.843
10	27	1	0	58	.852	.779	.868>	.853	.830	.788	.830
11	29	1	1	57	.837	.766	.853>	.844	.816	.788	.816
12	39	1	0	55	.822	.709	.838>	.806	.802	.788	.802
13	44	1	0	54	.807	.684	.822>	.788*	.788	.788	.788
14	46	1	0	53	.791	.675	.807>	.782	.775	.675	.775
15	47	1	0	52	.776	.670	.791>	.779	.762	.670	.762
16	48	1	0	51	.761	.666	.776	.776	.749	.666	.749
17	50	1	0	50	.746	.657	.760	.770	.735	.657	.735
18	51	3	0	49	.700	.653	.714	.767	.696	.653	.696
19	54	1	0	46	.685	.640	.698	.758	.683	.640	.683
20	60	1	0	45	.670	.617	.683	.742	.669	.617	.669
21	63	1	0	44	.654	.606	.667	.735	.656	.606	.656
22	64	1	0	43	.639	.603	.652	.732	.643	.603	.643
23	65	2	0	42	.609	.600	.621	.730	.616	.600	.616
24	66	1	0	40	.593<	.596	.605	.727	.603	.596	.603
25	68	1	1	39	.578<	.590*	.590	.590	.590	.590	.590
26	127	1	0	37	.563	.468	.573	.590	.573	.468	.573
27	136	1	0	36	.547	.457	.558	.590	.558	.457	.558
28	147	1	0	35	.531	.446	.542	.590	.542	.446	.542
29	161	1	1	34	.516	.434	.526	.590	.526	.434	.526
30	228	1	1	32	.500	.395	.510	.564	.510	.395	.510

(to be continued)

Table 7.2 (continued)

31	253	1	0	30	.483	.386	.493	.553	.493	.386	.493
32	280	1	0	29	.466	.377	.476	.541	.476	.377	.476
33	297	1	1	28	.450	.372	.459	.535	.459	.372	.459
34	321	1	5	26	.432	.365	.441	.526	.441	.365	.441
35	551	1	2	20	.411	.312	.419	.450	.419	.312	.419
36	624	1	1	17	.387	.297	.394	.429	.394	.297	.394
37	730	1	1	15	.361	.277	.368	.400	.368	.277	.368
38	836	1	2	13	.333	.258	.340	.372	.340	.258	.340
39	896	1	0	10	.300	.248	.306	.358	.306	.248	.306
40	993	1	0	9	.266	.232	.272	.335	.272	.232	.272
41	1024	1	2	8	.233	.227	.238	.328	.238	.227	.238
42	1349	1	4	5	.187	.183	.190	.264	.190	.183	.190

* The end of new partition.

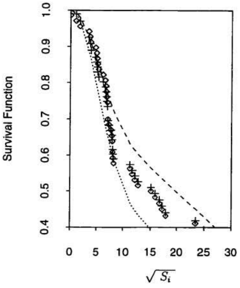


Figure 7.1: The K-M Estimate (◇), the Bounded MLE (+) and the Lower Bound and the Upper bound (dotted lines) the Post-Transplant Survival Function

Table 7.3: Root Mean Square Error of Bound Restricted Estimate of a Survival Function $P(t)$ under 10,000 Iterations

		Size 100				Size 300			
		$P(t)$				$P(t)$			
		0.50	0.25	0.10	0.05	0.50	0.25	0.10	0.05
	B1	.0130	.0119	.0092	.0089	.0123	.0111	.0065	.0042
Case	B2	.0245	.0234	.0169	.0134	.0194	.0180	.0121	.0084
I	B3	.0494	.0438	.0313	.0236	.0268	.0234	.0162	.0119
	K-M	.0501	.0436	.0303	.0219	.0288	.0251	.0174	.0126
	B1	.0255	.0236	.0162	.0121	.0217	.0195	.0126	.0086
Case	B2	.0411	.0369	.0250	.0180	.0265	.0221	.0145	.0101
II	B3	.0461	.0403	.0277	.0205	.0271	.0236	.0156	.0112
	K-M	.0503	.0438	.0300	.0218	.0289	.0252	.0172	.0127
	B1	.0334	.0315	.0226	.0159	.0256	.0233	.0163	.0116
Case	B2	.0435	.0391	.0282	.0201	.0277	.0245	.0171	.0123
III	B3	.0466	.0404	.0288	.0205	.0281	.0247	.0172	.0124
	K-M	.0506	.0432	.0307	.0219	.0286	.0249	.0174	.0127
	B1	.0451	.0372	.0207	.0122	.0277	.0239	.0151	.0094
Case	B2	.0472	.0398	.0239	.0148	.0281	.0244	.0161	.0106
IV	B3	.0486	.0422	.0286	.0205	.0284	.0248	.0169	.0119
	K-M	.0499	.0432	.0302	.0220	.0286	.0250	.0174	.0126
Exact S.D		.0500	.0433	.0300	.0218	.0289	.0250	.0173	.0126

Chapter 8

On Simulating Tail Probabilities with a Known Bound

The problem of estimating distribution functions is of great importance, particularly the upper tail probabilities. Simulations play a vital role in approximating probabilities of statistics with intractable distribution functions. In this chapter we will consider the problem of estimating tail probabilities of distribution functions with a known stochastic bound and having monotone decreasing density at the upper tails. Such prior knowledge may be utilized in the estimation problem to increase the efficiency.

8.1 The Problem

Let $F(\cdot)$ be the cumulative distribution function (cdf) of interest and $G(\cdot)$ be its stochastic bound such that $F(t) \leq G(t)$ for all t. We assume that $F(\cdot)$ has a probability density function $f(\cdot)$ which is monotone on the right tail $(a, +\infty)$, where a is a known real number. Suppose that one is interested in estimating $F(\xi)$ for a specific $\xi \in (a, +\infty)$. One can incorporate the prior knowledge of the stochastic bound and monotonicity to gain efficiency in estimation. The procedure is as follows.

Let m and n be positive integers and h be a positive real number such that $\xi - mh > a$. We first partition the real line into $k = m + n + 2$ intervals by $(-\infty, \xi_0 - mh]$, ...,$(\xi_0 - h, \xi_0]$, $(\xi_0, \xi_0 + h]$, ..., $(\xi_0 + nh, +\infty)$. Let $f_1, \ldots, f_k$ and $g_1, \ldots, g_k$ be the probabilities of F and G on these intervals, respectively. It is clear that $F(\xi) = \sum_{i=1}^{m+1} f_i$, $f_2 \geq \cdots f_{k-1}$, $\sum_{i=1}^{j} f_i \geq \sum_{i=1}^{j} g_i$, $j = 1, \ldots, k-1$, and $\sum_{i=1}^{k} f_i = 1$. A restricted estimate of f_i, denoted by BM-estimate, is then obtained by solving the problem

$$\min_{A \cap (\mathbf{g}+B)} \sum_{i=1}^{k} (f_i - \hat{f}_i)^2 \tag{8.1}$$

where $\hat{f}$ is the non-restricted MLE of f and

$$\begin{aligned} A &= \{\mathbf{f} = (f_1, f_2, \ldots, f_k) : f_2 \geq f_3 \geq \ldots \geq f_{k-1}\} \\ B &= \{\mathbf{f} = (f_1, f_2, \ldots, f_k) : \sum_{i=1}^{j} f_i \geq \sum_{i=1}^{j} g_i \quad, j = 1, \ldots, k-1, \sum_{i=1}^{k} f_i = \sum_{i=1}^{k} g_i = 1\}. \end{aligned}$$

Since A and B are closed convex cones, the algorithm proposed by Dykstra and Boyel (1987) can be applied to find the solution, (see Chapter 2 for the details of this algorithm and a simplified proof of the correct convergence of the algorithm).

8.2 A Simulation Study

It is well known that the simple isotonic regression always reduces the pointwise mean squares errors (MSE's) of estimates if the model is correct, (Lee, 1981). Pointwise MSE's of estimates obtained under the stochastic bound constraint can be significantly smaller than the MSE's of the unrestricted nonparametric MLE's when the bounds are properly imposed. But it could also increase the MSE's if only one sharp upper bound or one sharp lower bound is imposed and the sample size is small, see Lee, Yan and Shi (1996).

A simulation study is performed here to investigate the efficiency of our proposed procedure. The *relative efficiency* of two estimators is defined to be the reciprocal of the ratios of their MSE's.

Tables 1 and 2 present the relative efficiencies of restricted estimators to the unrestricted non-parametric MLE of the right tail probability 0.10 for the standard normal distribution and the standard exponential distribution. We observe that the combined constraints increase the efficiency of estimation substantially even when the improvement of individual constrains is not significant for small and moderate sample sizes. For example, in Table 8.1, the relative efficiencies of the estimators with sample size 50 are 1.11, 1.15, 1.18, 1.19 and 1.19 when only the monotone constraint is used; 1.60, 1.56, 1.46 1.34 and 1.20 when only the bound 2 constraint is used; and 2.33, 2.64, 2.56, 2.25 and 1.84 when both the monotone and bound 2 constraints are used for the listed lengths of intervals.

8.3 An Example

Generalized maximum modulus (GMM) was introduced by Lee (1996) in constructing Tukey-type confidence bands for monotone regression functions. Let Z_1, Z_2, $\ldots$, Z_n be independent standard normal variates. A GMM of order n is a random variable defined by

$$GM_n = \max_{1 \le i \le j \le n} \frac{|\sum_{h=i}^{j} Z_h|}{\sqrt{j-i+1}} \tag{8.2}$$

Table 8.1: Relative Efficiency of the Restricted Estimators to the Unrestricted MLE of Right Tail Probability of $F(t) = 1 - \exp(-t)$ at 2.30 with $k = 10$, (10,000 Iterations)

Sample Size	Constraint	Length of Intervals 0.08	0.16	0.24	0.32	0.40	MSE of MLE
50	Monotone only	1.11	1.15	1.18	1.19	1.19	18.59×10^{-4}
	Bound 1 only	1.30	1.32	1.32	1.30	1.27	
	Monotone & Bound 1	1.57	1.75	1.83	1.82	1.74	
	Bound 2 only	1.60	1.56	1.46	1.34	1.20	
	Monotone & Bound 2	2.33	2.64	2.56	2.25	1.84	
	Bound 3 only	1.67	1.52	1.34	1.18	1.02	
	Monotone & Bound 3	2.79	3.00	2.56	2.04	1.56	
100	Monotone only	1.09	1.14	1.15	1.14	1.13	9.39×10^{-4}
	Bound 1 only	1.15	1.16	1.17	1.17	1.13	
	Monotone & Bound 1	1.30	1.40	1.43	1.41	1.39	
	Bound 2 only	1.52	1.52	1.48	1.41	1.31	
	Monotone & Bound 2	1.99	2.23	2.17	1.86	1.71	
	Bound 3 only	1.71	1.59	1.44	1.29	1.13	
	Monotone & Bound 3	2.68	2.82	2.41	1.91	1.51	
500	Monotone only	1.06	1.08	1.07	1.05	1.03	1.80×10^{-4}
	Bound 1 only	1.00	1.00	1.00	1.01	1.01	
	Monotone & Bound 1	1.06	1.08	1.07	1.06	1.03	
	Bound 2 only	1.13	1.14	1.15	1.15	1.15	
	Monotone & Bound 2	1.22	1.26	1.26	1.24	1.21	
	Bound 3 only	1.54	1.54	1.51	1.44	1.33	
	Monotone & Bound 3	1.86	1.88	1.78	1.62	1.43	
1000	Monotone only	1.06	1.06	1.05	1.03	1.01	0.90×10^{-4}
	Bound 1 only	1.00	1.00	1.00	1.00	1.00	
	Monotone & Bound 1	1.06	1.06	1.13	1.03	1.01	
	Bound 2 only	1.02	1.03	1.03	1.03	1.05	
	Monotone & Bound 2	1.08	1.10	1.08	1.07	1.06	
	Bound 3 only	1.30	1.34	1.34	1.32	1.29	
	Monotone & Bound 3	1.45	1.48	1.45	1.38	1.28	

Note: The cumulative distribution functions of the three bounds are, respectively, $1 - \exp(-t/0.80)$, $1 - \exp(-t/0.90)$, and $1 - \exp(-t/0.95)$

Table 8.2: Relative Efficiency of the Restricted Estimators to the Unrestricted MLE of Right Tail Probability of $N(0,1)$ at 1.645 with $k = 10$, (10,000 Iterations)

Sample Size	Constraint	Length of Intervals 0.04	0.08	0.12	0.16	0.20	0.24	MSE of MLE
50	Monotone only	1.11	1.15	1.19	1.02	1.24	1.25	18.59×10^{-4}
	Bound 1 only	1.40	1.43	1.42	1.38	1.30	1.20	
	Monotone & Bound 1	1.74	1.88	2.26	2.33	2.24	2.01	
	Bound 2 only	1.58	1.57	1.49	1.40	1.28	1.15	
	Monotone & Bound 2	2.18	2.66	2.89	2.78	2.46	2.06	
	Bound 3 only	1.67	1.56	1.42	1.28	1.16	1.04	
	Monotone & Bound 3	2.75	3.24	3.26	2.83	2.34	1.88	
100	Monotone only	1.09	1.15	1.18	1.18	1.19	1.19	9.39×10^{-4}
	Bound 1 only	1.23	1.27	1.29	1.29	1.27	1.21	
	Monotone & Bound 1	1.42	1.60	1.73	1.78	1.77	1.67	
	Bound 2 only	1.49	1.51	1.49	1.43	1.34	1.22	
	Monotone & Bound 2	1.90	2.22	2.37	2.29	2.08	1.81	
	Bound 3 only	1.72	1.63	1.51	1.37	1.24	1.12	
	Monotone & Bound 3	2.65	3.07	2.96	2.52	2.07	1.72	
500	Monotone only	1.06	1.09	1.10	1.10	1.08	1.07	1.80×10^{-4}
	Bound 1 only	1.01	1.01	1.01	1.03	1.05	1.07	
	Monotone & Bound 1	1.07	1.10	1.02	1.13	1.14	1.15	
	Bound 2 only	1.10	1.12	1.17	1.18	1.18	1.15	
	Monotone & Bound 2	1.18	1.25	1.29	1.31	1.31	1.26	
	Bound 3 only	1.53	1.54	1.53	1.46	1.35	1.22	
	Monotone & Bound 3	1.84	2.00	1.96	1.82	1.61	1.37	
1000	Monotone only	1.06	1.07	1.07	1.06	1.05	1.03	0.90×10^{-4}
	Bound 1 only	1.00	1.00	1.00	1.00	1.01	1.03	
	Monotone & Bound 1	1.06	1.07	1.07	1.07	1.06	1.06	
	Bound 2 only	1.01	1.02	1.03	1.05	1.07	1.07	
	Monotone & Bound 2	1.07	1.10	1.11	1.11	1.13	1.11	
	Bound 3 only	1.29	1.34	1.36	1.34	1.30	1.20	
	Monotone & Bound 3	1.43	1.53	1.55	1.50	1.41	1.27	

Note: 1. The distributions of the three bounds are, respectively, $N(0,7)$, $N(0,0.8)$, and $N(0,0.9)$.

The distribution of GM_n is very complicated and its tail probabilities can be obtained by numerical integration of n dimensions (using e.g. NAG), but its precision for higher dimensional cases is still questionable. An alternative approximation can be obtained by simulation and selected percentiles of GMM random variables can be found in Lee (1996).

We now apply our proposed procedure to simulate tail probabilities of GMM. It is trivial that $GM_n \geq MM_n = \max_{1 \leq i \leq n} |Z_i|$. The latter is well known as the maximum modulus random variable. It follows that the cdf of MM_n is an upper bound for the cdf of GM_n. This bound will be used in the simulation for $n = 2$, 5, 10, 15 and 20. We shall assume that GM_n has a monotone decreasing density function at the right tails.

The number of intervals used in this simulation is 20 and the length of the interval is 0.05. Relative efficiency of the BM-estimator to the unrestricted MLE of the tail probabilities at the 90th, 95th and 99th percentiles (approximate) is listed in Table 8.3 for sample sizes 50, 100, 200 and 500. We observe that the improvement of the BM-estimator is very significant, especially for small sample size, small GMM order and at the 99th percentiles.

Table 8.3: Relative Efficiency of the BM-Estimator to the Unrestricted MLE of the GMM Tail Probabilities (10,000 Iterations)

Order of GMM	ξ	Sample Size 50	100	200	500
2	2.02	3.61	3.07	2.49	1.73
	2.32	3.75	3.21	2.77	2.03
	2.89	3.68	3.70	3.58	3.00
5	2.49	2.27	1.68	1.30	1.10
	2.76	2.80	1.99	1.49	1.21
	3.30	4.30	3.69	2.75	1.89
10	2.81	1.71	1.36	1.16	1.08
	3.06	2.16	1.58	1.28	1.14
	3.57	4.24	3.16	2.22	1.51
15	2.98	1.56	1.29	1.14	1.08
	3.23	1.98	1.47	1.25	1.11
	3.71	4.12	2.94	2.09	1.43
20	3.10	1.49	1.26	1.14	1.09
	3.34	1.87	1.42	1.23	1.11
	3.81	3.82	2.88	1.97	1.43

References

Agresti, A. (1974). "Bounds on the Extinction Time Distribution of a Branching Process," *Advances in Applied Probability* **6**, 322-335.

Agresti, A. (1980). "Generalized Odds Ratios for Ordinal Data," *Biometrics*,**36**, 59-67.

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, Inc.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman E. (1955). "An Empirical Distribution function with incomplete information," *Annals of Mathematical Statistics*, **26**, 641-7.

Barlow, R. E., Bartholomew, D.J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* New York: Wiley.

Barlow, R. E. and Brunk, H. D. (1972). "The Isotonic Regression Problem and Its Dual," *Journal of the American Statistical Association*, **67**, 140-7.

Barlow, R. E., and Marshall, A. W. (1964). "Bounds for Distributions with Monotone Hazard Rate, *I* and *II*," *Annals of Mathematical Statistics*, **35**, 1234-74.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, Inc.

Barlow, R. E. and Ubhaya, V. A. (1971). "Isotonic Approximation," in J. S. Rustagi (ed.), *Optimizing Methods in Statistics*, Academic Press, New York.

Bartholomew, D. J. (1959). "A Test of Homogeneity for Ordered Alternatives', *Biometrika*, **46**, 36-48.

Billingsley, P. (1968). *Convergence of Probabolity Measures*. Wiley, New York.

Brunk, H. D. (1955). "Maximum Likelihood Estimates of Monotone Parameters," *Annals of Mathematical Statistics*, **26**, 607-16.

Brunk, H. D. (1965). "Conditional Expectation given a σ-lattice and Applications," *Annals of Mathematical Statistics*, **36**, 1339-50.

Brunk, H. D., Franck, W. E., Hanson, D. L. and Hogg, R. V. (1966). "Maximum Likelihood Estimation of the Distributions of Two Stochastically Ordered Random Variables," *Journal of the American Statistical Association*,

Burger, H. U. (1993). "Dispersion orderings with applications to nonparametric tests," *Stat. and Prob. Lett.*,**16**, 1-9.

Bhattacharjee, M. C. (1987). "The Time to Extinction of Branching Processes and Log-convexity: *I*," *Probability in the Engineering and Informational Sciences* **1**, 265-278.

Crowley, John and Hu, M. (1977). "Covariance Analysis of Heart Transplant Survival Data," *Journal of the American Statistical Association*, **72**, 27-35.

Cressie, N. and T. R. C. Read (1984). "Multinomial Goodness of Fit Models." *Journal of the Royal Statistical Society*, **B46**, 440-464.

Dykstra, R. L. (1982). "Maximum Likelihood Estimation of the Survival Functions of Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, **77**, 621-628.

Dykstra, R. L. (1983). "An Algorithm for Restricted Least Squares Regressions," *Journal of the American Statistical Association*, **78**, 837-42.

Dykstra, R. L. (1984). "Dual Convex Cones for Order Restrictions with Applications." In Y. L. Tong (ed.), *Inequalities in Statistics and Probability*, Institute of Mathematical Statistics, Hayward, 228-35.

Dykstra, R. L. (1985). "Computational Aspects of I-projections," *J. Statist. Comput. Simul.*, **21**, 265-74.

Dykstra, R. L. and J. P. Boyle (1987). "An Algorithm for Least Squares Projection onto the Intersection of Translated, Convex Cones," *J. statist. Plann. Inf.*,**11**, *391-400.*

Dykstra, R. L., Feltz, C. J. (1989). "Nonparametric Maximum Likelihood Estimation of Survival Functions with a General Stochastic Ordering and Its Dual," Biometrika, **76**, *331-341.*

Dykstra, R. L., Kochar, S. and Robertson, T. (1991) "Statistical Inference for Uniform Stochastic Ordering in Several Populations Annals of Statistics, **19**,870-888.

Dykstra, R. L., Kochar, S. and Robertson, T. (1995) "Inference for Likelihood Ratio Ordering in the Two-Sample Problem," *Journal of the American Statistical*

Association, **90**, 1034-40.

Dykstra, R. L. and C. I. C. Lee (1991). "Multinomial Estimation Procedures for Isotonic Cones," *Statist. Probab. Lett.*, **11**, 155-60.

Dykstra, R. L. and C. I. C. Lee and X. Yan (1996). "Multinomial Estimation Procedures for Two Stochastically Ordered Distributions," to appear in *Statist. Probab. Lett.*

Dykstra, R. L. and Robertson, T. (1982a). "An Algorithm for Isotonic Regression for Two or More Independent Variables," *Annals of Statistics*, **10**, 708-11.

Dykstra, R. L. and Robertson, T. (1982b) "Order Restricted Statistical Tests on Multinomial and Poisson Parameters: The Starshaped Ordering," *Annals of Statistics*, **10**, 1246-1252.

Dykstra, R. L. and Robertson, T. (1983) "On Testing Monotone Tendencies," *Journal of the American Statistical Association*, **78**, 342-350.

Eeden, C. van (1958). "Testing and Estimating Ordered Parameters of Parameter of Probability Distributions,", *Doctoral Dissertation*, University of Amsterdam, Studentendrukkerij Poortper, Amsterdam.

Esary, J. D., Proschan, F. (1972). "Relationships among Some Concepts of Bivariate Dependence," *Annals of Mathematical Statistics*, **43**, 651-655.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). "Association of Random Variables, with Applications," *Annals of Mathematical Statistics*, **38**, 1466-74.

Feltz, C. J. and Dykstra, R. L. (1985). "Maximum Likelihood Estimation of the Survival Functions of N stochastically ordered Random Variables," *Journal of the American Statistical Association, 80*, 1012-9.

Gebhardt, F. (1970). "A Algorithm for monotone regression with one or more independent Variables," *BiometriKa*. **57**, 263-71.

Grove, D., M. (1984). "Positive association in a two way contigency table: likelihood ratio tests," *Commun. Statist. - Theor. Meth.*,**13** 931-945.

Hu, X. and Wright, F. T. (1994). "Likelihood Ratio Tests for a Class of Non-Oblique Hypotheses," *Ann. Inst. Statist. Math.*, **46**, 137-145.

Jogdeo, K. (1968). "Characterizations of Independence in Certain families of Bivariate and Multivariate Distributions," *Annals of Mathematical Statistics*, **39**, 433-441.

Johansen, S. (1978). "The Product Limit Estimators as Maximum Likelihood Estimator," *Scandinavian Journal of Statistics*, **5**, 195-199.

Karlin, S. (1968). *Total Positivity*. Standford University press, Standford.

Karlin, S., and Proschan, F. (1960). "Pólya Type Distributions of Convolutions," *Annals of Mathematical Statistics*, **31**, 721-736.

Karlin, S., Proschan, F., and Barlow, R. E. (1961). "Moment Inequalities of Pólya Frequency Functions," *Pacific J. Math.*, **11**, 1023-33.

Kaplan, E. L. and Meier, P. (1958). "Nonparametric Estimation from Incomplete Observations," *Journal of the American Association*, **53**, 457-481.

Kelly, R. E. (1989). "Stochastic Reduction of Loss in Estimating Normal Means by Isotonic Regression," *Annals of Statistics*, **17**, 937-940.

Kiefer, J., and Wolfowitz, J. (1956). "Consistency of the Maximum likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, **27**, 887-906.

Kudô, A. (1963). "A Multivariate Analogue of the One-sided Test", *Biometrika*, **50**, 403-418.

Lee, C. I. C. (1981). "The Quadratic Loss of Isotonic Regression under Normality," *Annals of Statistics*, **9**, 686-668.

Lee, C. I. C. (1983). "The Min-max Algorithm and Isotonic Regression," *Annals of Statistics*, **11**, 467-477.

Lee, C. I. C. (1987a). "Maximum Likelihood Estimation for Stochastically Ordered Multinomial Populations with Fixed and Random Zeros," *Proc. Symp. in Statistics and Festschrift in Honor of V. M. Joshi*, 189-197.

Lee, C. I. C. (1987b). "Chi-square Tests for and against an Order Restriction on Multinomial Parameters," *Journal of the American Statistical Association*, **82**, 611-618.

Lee, C. I. C. (1988). "Quadratic Loss of Order Restricted Estimators for Treatment Means with a Control," *Annals of Statistics*, **16**, 751-758.

Lee, C. I. C. (1996). "On Estimation for Monotone Dose-Response Curves," to appear in *Journal of the American Statistical Association*.

Lee, C. I. C. (1995b). "A One-Sided Generalized Studentized Range Test for Testing Against a Simple Ordered Alternative," *Technical Report*, Memorial University of Newfoundland.

Lehmann, E. l. (1955). "Ordered families of distributions," *Ann. Math. Stat.*, **26**, 399-419.

Lehmann, E. l. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

Lehmann, E. l. (1966). "Some Concepts of Dependence," *Ann. Math. Stat.*, **37**, 1137-1153.

Mann, H. B. and Whitney, D. R. (1947). "On Test of Whether One of Two Random Variables Is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, **18**, 50-60.

Parnami, J. C., Singh, H. and Puri, P. S. (1993). Solution of an Optimization Problem Arising in Maximum Likelihood Estimation of Ordered Distribution. *Sankhyā Series A*, **55**, 49-65.

Perlman, M. D. (1969). "One-sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, **40**, 549-567.

Robertson, T. (1966). "A Representation of Conditional Expections Given σ-lattices," *Annals of Mathematical Statistics*, **37**, 1279-83.

Robertson, T., and Wegman, E. J. (1978). "Likelihood Ratio Tests for Order Restrictions in Exponential Families," *Annals of Statistics*, **6**, 485-505.

Robertson, T., and Wright, F. T. (1974). "A Norm Reducing Property for Isotonized Cauchy Mean Value Functions," *Annals of Statistics*, **2**, 1302-7.

Robertson, T., and Wright, F. T. (1981). "Likelihood Ratio Tests for and against Stochastic Ordering between Multinomial Populations," *Annals of Statistics*, **9**, 1248-57.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Ordered Restricted Statistical Inference*, Wiley, New York.

Rockfellar, A., T. (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Ross, S. M. (1983). *Stochastic Processes*, Wiley, New York.

Shaked, M. (1979). "Estimation of Starshaped Sequences of Poisson and Normal Means," *Annals of Statistics*, **7**, 729-741.

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Inc.

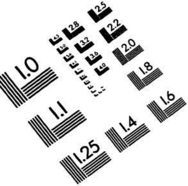
Shanthikumar, J. G. and Yao, D. D. (1991). "Bivariate Characterization of some stocahstic order realtions," *Adv. Appl. Prob.*, **23**, 642-659.

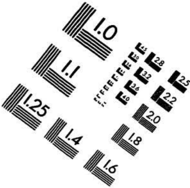
Tang, D.-I., Gnecco, C. and Geller, N. L. (1989). "An Approximate Likelihood Ratio Test for a Normal Mean Vector with Nonnegative Components with Application To Clinical Trials," *Biometrika*, **76**, 577-583.

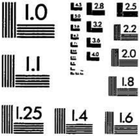
Thompson, W. A., Jt. (1962). "The Problem of Negative Estimates of Variance Component," *Annals of Mathematical Statistics*, **33**, 273-89.

Turnbull, B. W., Brown, B. Wm., and Hu, M. (1974). "Survivalship Analysis of Heart Transplant Data," *Journal of the American Association*, **53**, 457-481.

IMAGE EVALUATION
TEST TARGET (QA-3)







150mm

6"