On Estimation and Inference under Order Restrictions

by

Yong Seok Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2011

Doctoral Committee:

Professor John D. Kalbfleisch, Co-chair
Professor Jeremy M. G. Taylor, Co-chair
Professor Bin Nan
Associate Professor Ji Zhu

For my parents, my sister, my wife, my daughter and my son

ACKNOWLEDGMENTS

My deepest gratitude is to my dissertation co-chairs, Dr. Jeremy Taylor and Dr. Jack Kalbfleisch, for their guidance, patience and encouragement in this study. Their numerous helpful suggestions about how to study statistical methods, how to organize and present ideas, and even how to write a paper have made the completion of this dissertation possible. I would like to specially thank my academic advisor, also one of my dissertation co-chairs, Dr. Jeremy Taylor for his generous financial support for most of my stay in the university of Michigan and for many interesting research topics that he has introduced. I am also grateful to my other dissertation committee members, Dr. Bin Nan and Dr. Ji Zhu, for their willingness to be in my committee, for their careful review of my dissertation and their valuable comments to improve my dissertation.

I would like to think all faculty members, staffs, and students in the department of Biostatistics at the university of Michigan for making this department an enjoyable place to study and to conduct research.

Finally and most importantly, I would like to thank my family. I thank my parents and my sister's support and confidence in me. I thank my mother-in-law for her unbounded helps so that I have more time to focus on my work. I thank my wife for her unlimited support, encouragement and love. I thank my daughter Eleanor and son Evan for their bringing happiness and joy of the life and for their unending requests to finish my degree. All those have essentially driven me for the completion of this dissertation.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF FIGURES vii

LIST OF TABLES viii

LIST OF APPENDICES x

ABSTRACT xi

CHAPTER

I. Introduction 1

II. Nonparametric Maximum Likelihood Estimator of Stochastically Ordered Survivor Functions 6

2.1 Introduction 6
2.2 Estimation of Discrete Survivor Functions 10
  2.2.1 One-sample Case 10
  2.2.2 One-sample Case: Bounded Below Constraint 11
  2.2.3 One-sample Case: Bounded Above Constraint 13
  2.2.4 Two-sample Case 16
2.3 Extension to Non-discrete Case 19
  2.3.1 One-sample Case with $T_1 \geq_{st} T_2$ and $S_2(t)$ is Known . . 20
  2.3.2 One-sample Case with $T_1 \leq_{st} T_2$ and $S_2(t)$ is Known . . 20
  2.3.3 Two-sample Case With No Potential Event Times 22
2.4 Consistency 24
2.5 Example 24
2.6 Confidence Intervals 26
2.7 Simulation Study 27

2.7.1 Two-sample Case when Sample Size is Small 27
2.7.2 Properties of Confidence Intervals 28
2.8 Discussion 33

III. Pointwise Constrained Maximum Likelihood Estimator of the Survivor Function under Stochastic Ordering 35

3.1 Introduction 35
3.2 Estimation Methods 38
3.2.1 Notation and Likelihood 38
3.2.2 Linearly Constrained Convex Minimization 40
3.2.3 Maximum Likelihood Estimator of the Survivor Function Subject to a Single Constraint 41
3.2.4 The Profile Likelihood of $S(x)$ 42
3.2.5 Reformulation of the Problem Using Profile Likelihood . 43
3.2.6 Generalized Pool-adjacent-violators Algorithm in the Simple Ordering Case 45
3.3 Consistency and Asymptotic Distribution 46
3.4 Comparison with Kaplan-Meier Estimator when Sample Size is Large 48
3.4.1 Simple Ordering Case 48
3.4.2 The Two-sample Case, $G = 2$ 49
3.5 Confidence Intervals 50
3.5.1 Bootstrap 50
3.5.2 Confidence Interval Centered on Constrained Estimator . 51
3.6 Simulation Studies 52
3.6.1 Two-sample Case when Sample Size is Small 52
3.6.2 Two-sample Case: Asymptotic Properties 53
3.6.3 Simple Ordering Case 55
3.7 Example 57
3.8 Discussion 60

IV. Confidence Intervals under Order Restriction 63

4.1 Introduction 63
4.2 Confidence Intervals for μ_1 and μ_2 with Known Variances 65
4.2.1 Family of Confidence Intervals 65
4.2.2 Selection of γ 68
4.2.3 Properties of Proposed Confidence Intervals 70
4.3 Confidence Intervals for μ_1 and μ_2 with Unknown Variances 72
4.3.1 Restricted Confidence Interval when σ_1^2 and σ_2^2 are Unknown but σ_2^2/σ_1^2 is Known 72
4.3.2 Restricted Confidence Intervals when both σ_1^2 and σ_2^2 are Unknown 73
4.4 Confidence Intervals with Three or more Groups 76

4.4.1 Three-sample Case with Known Variances 76
4.4.2 Three-sample Case with Unknown Variances 78
4.4.3 Confidence Intervals for more than Three Groups 79
4.5 Other Restricted Confidence Intervals 80
4.6 Simulation Study 81
4.7 Example 84
4.8 Discussion 86

V. Future Research 89

APPENDICES 93

BIBLIOGRAPHY 125

LIST OF FIGURES

Figure

2.1 Estimates of survivor functions, discrete case. 17

2.2 Estimates of survivor functions, continuous case. 23

2.3 Estimates of survivor functions, two-sample case. 25

2.4 Comparison of RMSEs for different estimators. 29

3.1 Comparison of mean squared errors for different estimators. 54

3.2 Comparison of asymptotic relative efficiency under the constraint $T_1 \geq_{st} T_2$. 55

3.3 Comparison of Kaplan-Meier estimator and pointwise C-NPMLE in the three sample case. 56

3.4 Estimates in the University of Michigan prostate cancer example. 59

4.1 Sample space of (y_1, y_2). 67

4.2 Comparison of width of confidence interval using $\hat{\gamma}$ and minimum possible width. 70

4.3 Coverage probability and ratio of average width of restricted confidence interval for μ_1 compared to unrestricted confidence interval for μ_1. 72

4.4 Upper limits of the confidence interval in three or more sample case. . . . 78

E.1 An example of Algorithm E.1 in Appendix E.2. 115

LIST OF TABLES

Table

2.1 Results for Example 2.7. 16

2.2 Results for Example 2.12. 22

2.3 Survival times (in years) of male patients in example in section 2.5. 25

2.4 KM and C-NPMLE estimates in example in section 2.5. 26

2.5 Distributions and sample sizes used in simulation study. 28

2.6 MSE ($\times 10^4$) and Bias ($\times 10^2$) in a simulation study in section 2.7.2. . . . 31

2.7 Coverage rates ($\times 10^2$) and average widths ($\times 10^2$) of confidence intervals. 32

3.1 Underlying Distributions in a simulation study in section 3.6.1. 52

3.2 Coverage rate (%) and average width (%) of 95% confidence intervals. . . 57

3.3 Estimates and confidence intervals of survivor functions for some selected times in the prostate cancer example. 60

4.1 Theoretical maximum coverage rate of CI for μ_1 in the situations with different ratio of variances. 71

4.2 Different combinations of population means, variances and sample sizes used in simulation studies. 82

4.3 Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the variances are known. 83

4.4 Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the ratios of variances are known. 84

4.5 Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the variances are estimated separately. 85

4.6 Half-life of an antibiotic in rats. 86

4.7 Estimates and confidence intervals of half-lives in the example in section 4.7. 86

LIST OF APPENDICES

Appendix

A. Proof of Theorem 2.13 in Section 2.4 94

B. Proof for the Two-sample Case from Section 2.2 96
  B.1 Characteristics of Results from Theorem 2.8 and Algorithm 2.9 . . 96
  B.2 Optimization Problem for the Two-sample Case 100
  B.3 Uniqueness of the Solution 103

C. The Pool-adjacent-violators Algorithm for the Linear Ordering Case 106

D. Proof of Theorem 3.2 108

E. Proof of Theorem 3.4: $\hat{S}_g(t)$ is Nonincreasing in t 110
  E.1 Notation and Characteristics of the Pointwise C-NPMLE 110
  E.2 An Algorithm to Obtain the Pointwise C-NPMLE at a Time $x_2 > x_1$ 113
  E.3 Completion of the Proof 117

F. Proof of Theorem 3.5 118

G. Proof of Theorem 3.6 121

ABSTRACT

The aim of statistical analysis and inference is to draw meaningful conclusions. In the case where there is prior knowledge of stochastic orderings or inequalities, it is desirable to incorporate this information in the estimation. This avoids possible unrealistic estimates, and may also lead to gain in efficiency.

In this dissertation we first present the constrained nonparametric maximum likelihood estimator (C-NPMLE) of the survivor functions in one- and two-sample settings. Dykstra (1982) also considered C-NPMLE for such problems, however, as we show, Dykstra's method has an error and does not always give the C-NPMLE. We corrected this error and simulation shows improvement in efficiency compared to Dykstra's estimator. Confidence intervals based on bootstrap methods are proposed. Uniqueness and consistency of the proposed estimators is established.

Second, we propose a new estimator, the pointwise C-NPMLE, which is defined at each time t by the estimates of the survivor functions subject to constraints at t only. The estimator is shown to be non-increasing in t, and the consistency and the asymptotic distribution of the estimators are presented. In the development of this estimator and the characterization of its properties, we transform the problem into one that uses the profile likelihood; we adapt the pool-adjacent-violators algorithm, in which pooling is defined in a special way. Different methods to construct confidence intervals are also proposed. The estimator is shown to have good properties compared to other potential estimators.

Finally, we propose a new method to construct confidence intervals (CIs) for G independent normal means under the linear ordering constraint. The method is based on defining intermediate random variables that are related to the original observations and using the CIs of the means of these intermediate random variables to restrict the original CIs from the separate groups. This method is extended to the case with three or more groups and the simulation studies show that the proposed CIs have coverage rates close to nominal levels with reduced average widths.

CHAPTER I

Introduction

The aim of statistical analysis and inference is to draw meaningful conclusions about the real world. In the case where there is prior knowledge of stochastic orderings or inequalities, it is desirable to incorporate this information in estimation method to avoid unrealistic estimates, and it may also lead to gain in efficiency. For example, if μ_g is the average toxicity rate of a drug for dose level g in a clinical trial, the parameters should satisfy the restriction $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_G$; we should then require that estimates also satisfy this constraint. In a cancer study, we expect patients with a higher stage of a certain cancer at diagnosis to have lower survival probabilities at all times than patients with a lower stage of the same cancer at diagnosis; we will seek estimators of the survivor probabilities that satisfy this ordering constraint.

Let $S_1(t)$ and $S_2(t)$ be the survivor functions of random variables T_1 and T_2. Then T_1 is stochastically larger (smaller) than T_2, denoted by $T_1 \geq_{st} T_2$ ($T_1 \leq_{st} T_2$), if $S_1(t) \geq S_2(t)$ ($S_1(t) \leq S_2(t)$) for all t (Lehmann, 1955). This type of stochastic order is called the usual stochastic order. There are many other types of stochastic orders, such as the hazard rate order, the likelihood ratio order, and the residual time order (see Shaked and Shanthikumar (2007) for more details). The concept of stochastic order can be generalized to $G > 2$

groups and all the groups can be fully or partially ordered. Some special cases are the linear ordering or simple ordering, such as $T_1 \geq_{st} \cdots \geq_{st}, T_G$; the simple tree ordering, such as $T_1 \geq_{st} T_2, T_1 \geq_{st} T_3, \ldots, T_1 \geq_{st} T_G$; the factorial ordering, such as $T_1 \geq_{st} T_2 \geq_{st} T_4$, $T_1 \geq_{st} T_3 \geq_{st} T_4$; and the umbrella ordering, such as $T_1 \geq_{st} \cdots \geq_{st} T_u \leq_{st} \cdots \leq_{st} T_G$. In this dissertation, we consider the usual stochastic order in Chapters II and III for survival distributions. In Chapter IV, we consider linear ordering of the means (as shown in the toxicity rate of a drug example in the previous paragraph) when the underlying distributions are normals.

Since the early 1950s, estimation problems in a restricted parameter space including ordered parameters have been studied. One commonly used method is restricted maximum likelihood estimation. Ayer et al. (1955) studied the binomial case when the event probabilities satisfy linear ordering constraint and first gave a minimax form for the solution. Many other methods such as isotonic regression have also been proposed as well as methods of testing hypotheses in restricted parameter spaces. While there is considerable literature in point estimations and hypothesis tests (see Robertson et al. (1988), Silvapulle and Sen (2005) for more details), there is much less on confidence intervals. Bayesian methods in which the restriction is incorporated into the prior can also be used (Dunson and Peddada, 2008; Taylor et al., 2007; Marchand and Strawderman, 2006; Zhang and Woodroofe, 2003; Roe and Woodroofe, 2000). Marchand and Strawderman (2004) and van Eeden (2006) reviewed estimation methods that have been developed in the past and discussed the "good" properties of restricted estimators, such as dominance, minimax and admissibility. Cohen and Sackrowitz (2004) discussed some inference issues and pointed out that traditional inference methods, such as likelihood based methods, can lead to some undesirable properties in restricted parameter problems. In addition, Andrews (2000) pointed out that the bootstrap method, which has been very useful for constructing confidence intervals for complicated situations, will fail when a parameter is on the boundary or close to the boundary of the parameter space. Thus, it is of interest to develop an inference procedure without depending on traditional inference methods.

In terms of estimating survivor or distribution functions, the nonparametric approach is generally used to avoid strong unrealistic assumptions about the underlying distributions. For independent right censored data, one commonly used estimator is the constrained nonparametric maximum likelihood estimator (C-NPMLE), which is the estimator that maximizes the generalized likelihood (Johansen, 1978) subject to the stochastic ordering constraint. Brunk et al. (1966) studied the C-NPMLE in the two-sample case without censoring, and Dykstra (1982) extended this result to right censored data in the one- and two-sample cases. For the case with three or more groups, Feltz and Dykstra (1985), Dykstra and Feltz (1989), Dardanoni and Forcina (1998), Hoff (2003b) and Lim et al. (2009) proposed various method for the C-NPMLE. Other alternative estimators include the swapping method in which estimates of Kaplan-Meier estimators are swapped when the constraint is violated (Lo, 1987), the averaging method in which two Kaplan-Meier estimators are averaged when the constraint is violated with weights based on initial sample sizes (Rojo, 2004), and isotonic regression of Kaplan-Meier estimators with the weights based on initial sample sizes under linear ordering constraint (El Barmi and Mukerjee, 2005). Recent works also considered estimating the survivor functions from the posterior distributions by incorporating the information of stochastic order through the prior distribution (Dunson and Peddada, 2008; Karabatsos and Walker, 2007; Hoff, 2003a; Evans et al., 1997; Arjas and Gasbarra, 1996).

In studying the C-NPMLE of the survivor function, we found that some conditions were not appropriately considered in the algorithms and proofs in Dykstra (1982). As a consequence, his result is incorrect for some data configurations. Specifically, Dykstra's estimator is the C-NPMLE in the one-sample case when estimating $S_1(t)$ subject to the constraint $T_1 \geq_{st} T_2$. However, for other types of constraints, Dykstra's method may fail to find the C-NPMLE. In Chapter II, we present a correct method to obtain the C-NPMLE of the survivor functions in one- and two-sample settings where the survivor distributions could be discrete or continuous and discuss the non-uniqueness of the estimators. We also present a computationally efficient algorithm to obtain the C-NPMLE. To address the non-uniqueness of the C-NPMLE of $S_1(t)$ when $S_1(t) \leq S_2(t)$, we consider the maximum C-NPMLE (MC-NPMLE) of $S_1(t)$. In the one-sample case with arbitrary upper bound survivor function $S_2(t)$, we present a novel and efficient algorithm for finding the MC-NPMLE of $S_1(t)$. We study consistency of the the C-NPMLE and consider methods for constructing confidence intervals. In simulation studies, we compare the C-NPMLE with Dykstra's estimator and other alternatives.

A number of authors (Rojo, 2004; Gangnon and King, 2002; El Barmi and Mukerjee, 2005) have noted that the Dykstra's estimator as the C-NPMLE of the stochastically ordered survivor functions can have relatively large pointwise bias and mean squared error (MSE) at a fixed time, and they have suggested alternatives that can have smaller MSE of estimator for underlying distribution. A corrected version of the C-NPMLE presented in Chapter II does lead to improved properties. However, this corrected C-NPMLE still often yields poorer pointwise properties than some of the other estimators that have been proposed. When we consider finite sample properties of an estimator $\hat{S}(t)$, we usually use pointwise criteria, such as pointwise bias or pointwise MSE, at each fixed t. In contrast to pointwise estimators such as described in Rojo (2004) and Lo (1987), the C-NPMLE is focused on estimating the whole survival curve. So it is perhaps not surprising that Rojo's estimator typically has better properties when evaluated using metrics such as pointwise MSE. On the other hand, these estimators do not adapt well to unequal censoring distributions between groups, whereas the C-NPMLE does. This motivated us to propose a new constrained estimator in Chapter III, a pointwise C-NPMLE, which is defined at each time t by the estimates of the survivor functions subject to constraints at time t only. We also propose an efficient algorithm to obtain the estimators. The estimator is shown to be non-increasing in t and consistent. The asymptotic distribution of the estimators are presented. In the development of this estimator and the characterization of its properties, we transform the problem into one that uses the profile likelihood; we adapt the pool-adjacent-violators algorithm, in which pooling is defined in a special way. Different methods to construct confidence intervals including fixed width confidence intervals and bootstrap confidence intervals are proposed and the properties of these intervals are compared. In simulation studies, we also compare the properties of the pointwise C-NPMLE with those of other potential alternatives.

In Chapter IV, we consider the problem of constructing confidence intervals for G independent normal population means under linear ordering constraints. For this problem, confidence intervals based on asymptotic distributions, likelihood ratio tests and bootstrap methods do not have good properties, particularly when some of the population means are close to each other. We propose a new method based on defining intermediate random variables that are related to the original observations and using the confidence intervals of the means of these intermediate random variables to restrict the original confidence intervals from the separate groups. In the two-sample case with known variances or known ratio of the variances, we present a theorem about the coverage rates of the confidence intervals. We also proposed methods to construct confidence intervals when the variances of the populations need to be estimated separately. The method is also extended to the case with three or more groups. In simulation studies, we compare coverage rates and average widths of the new method with those of other methods in various different scenarios.

CHAPTER II

Nonparametric Maximum Likelihood Estimator of Stochastically Ordered Survivor Functions

2.1 Introduction

Suppose that the random variable $T > 0$ is the time until some specified event, such as death or recurrence of a disease. Our interest centers on estimating the survivor function of T, $S(t) = P(T > t), t > 0$. With right censored data, the Kaplan-Meier estimator (KM, Kaplan and Meier, 1958) is commonly used. In some instances with two or more groups, we may have prior knowledge that the survivor function of one group is greater than or equal to that of another group at all times. This type of constraint is called *stochastic ordering* and can arise in many contexts; for example, with time from diagnosis to death of cancer patients where the survival probability for a lower tumor stage group can be reasonably assumed to be larger than that in a higher stage group. As well as wanting an estimator to be consistent with this prior knowledge, it can be expected that an estimator that satisfies the constraint will be more precise, with lower sampling variability, than one that does not utilize this knowledge, particularly in small sample size settings. As an obvious example of the potential for substantial improvements in efficiency, consider three groups with the middle group bounded both above and below. If the middle group has small sample size compared to the other two, the efficiency of the constrained estimator will be substantially better than that of the simple KM estimator for that middle group.

The cancer application mentioned above is just one example where distributions will be ordered, there are numerous other examples in biomedical and other areas of research where there is a strong rationale for an ordering of distributions. One approach to imposing ordering is through parametric modeling, an alternative approach that imposes less assumptions is through non-parametric estimation subject to an ordering constraint. In view of the frequency of situations where ordering constraints are natural, the potential benefit by using these constraints, and the mild nature of the assumptions, it is surprising to us that ordered constrained estimation is not used more in applications.

Let T_1 and T_2 have survivor functions $S_1(t)$ and $S_2(t)$ respectively then T_1 is stochastically less than T_2 ($T_1 \leq_{st} T_2$) if $S_1(t) \leq S_2(t)$ for all t. There are many possible definitions of ordering of survival functions (Shaked and Shanthikumar, 1994, 2007). The above one is called usual stochastic ordering. Other possible definitions that make stronger assumptions are hazard rate ordering, reverse hazard ordering and likelihood ratio ordering. There are also weaker forms of ordering, such as second-order dominance (Rojo and El Barmi, 2003) and stochastic precedence (Arcones, Kvam, and Samaniego, 2002). Which form of ordering is most appropriate in any application will depend on the context. In this paper we focus on the usual stochastic ordering. First we consider a one-sample problem in which data are available from $S_1(t)$ and $S_2(t)$ is known. Then we consider the more applicable two-sample problem in which $S_1(t)$ and $S_2(t)$ are unknown and data are available on both.

A constrained nonparametric maximum likelihood estimator (C-NPMLE) is a nonparametric estimator that maximizes the likelihood subject to the constraint. Since the initial work of Brunk, Franck, Hanson, and Hogg (1966), many methods and algorithms have been proposed to obtain the C-NPMLE for different situations, including the one-sample case, the two-sample case, the linear ordering case, in which the constraint takes the form $T_1 \leq_{st}, \ldots, \leq_{st} T_G$, and a general partial ordering case, which includes constraints such as $T_1 \leq_{st} T_2, T_1 \leq_{st} T_3$. Brunk, Franck, Hanson, and Hogg (1966) studied the C-NPMLE in the two-sample case without censoring. Dykstra (1982) extended this work to accommodate right censored data in the one- and two-sample cases. According to his work, the C-NPMLE is a modified KM type estimator with an adjustment to the number of subjects in each risk set. In the case of linear ordering or general partial ordering, Feltz and Dykstra (1985), Dykstra and Feltz (1989), Hoff (2000, 2003b), and Lim et al. (2009) extended this work and proposed various methods to find the C-NPMLE. The NPMLE of an unconstrained survivor function has jumps only at observed event times and the C-NPMLE has been assumed to have jumps only at observed event times by many researchers. However, as can be seen in section 2.3, in some cases there is no C-NPMLE that jumps only at observed event times. This incorrect assumption that jumps only occur at observed event times has also been implicitly made in research on likelihood ratio tests (e.g. Thomas and Grunkemeier 1975; Li 1995; Murphy 1995).

Some data configurations were not appropriately considered in the main theorem and the proof in Dykstra (1982). As a consequence, the theorem that he stated is not correct and his algorithm does not always give the C-NPMLE. More specifically, Dykstra's estimator (D-estimator) is the C-NPMLE in the one-sample case when estimating $S_1(t)$ subject to the constraint $T_1 \geq_{st} T_2$. However, for other constraints, his method fails and the purpose of this article is to provide a correctly stated theorem and associated algorithm.

As an illustration, we give a simple example of a one-sample case that the D-estimator is not a C-NPMLE. Suppose that S_1 and S_2 are known to have probability mass only at times $1, 4, 5$ and the observed event times from S_1 are 1, 2+, 3+ and 5 (+ denotes censoring). The likelihood based on the data is $L = \{S_1(0) - S_1(1)\} \times S_1(2) \times S_1(3) \times \{S_1(4) -$

$S_1(5)\}$. If the constraint is $S_1(1) \leq S_2(1) = 0.8$ and $S_1(4) \leq S_2(4) = 0.4$. The D-estimator is $\tilde{S}_1(1) = \tilde{S}_1(2) = \tilde{S}_1(3) = \tilde{S}_1(4) = 0.4$ and $\tilde{S}_1(5) = 0$ with the corresponding likelihood $\tilde{L} = (1 - 0.4) \times 0.4^3 = 0.0384$. It is easily seen that another constrained estimate, $\hat{S}_1(1) = \hat{S}_1(2) = \hat{S}_1(3) = 2/3$, $\hat{S}_1(4) = 0.4$ and $\hat{S}_1(5) = 0$, gives a larger likelihood, $\hat{L} = (1 - 2/3) \times (2/3)^2 \times 0.4 \approx 0.0593$.

Some alternative estimators of constrained survivor functions have also been proposed. In the two-sample case, Lo (1987) proposed a simple estimator that swaps the estimates of survivor functions when the constraint is violated. Rojo (2004) and El Barmi and Mukerjee (2005) proposed estimators that use the weighted average of the two KM estimators at times when the constraint is violated with weights based on the initial sample sizes. In numerical work (Rojo and Ma, 1996; Rojo, 2004), these alternative estimators were found to be superior to the C-NPMLE in terms of pointwise mean squared error. However, these investigations used the incorrect C-NPMLE from Dykstra (1982) and did not consider unequal censoring patterns between the two groups.

In this paper, to develop the ideas and the notation, we start with the simplest one-sample case with discrete survivor function before considering the more important two-sample case. In section 2.2, we consider the discrete case, where we assume that T_g follows a discrete distribution and the potential death times are also given. In section 2.3, we extend to the case where $S_1(t)$ and $S_2(t)$ are not discrete functions. In section 2.4, we show the uniform consistency of the C-NPMLE in the two-sample case. In section 2.5, we analyze larynx cancer data in the two-sample case. In section 2.6, we propose methods to construct confidence intervals and in section 2.7, we conduct a simulation study to compare finite sample property of the C-NPMLE with the D-estimator, Lo's estimator and Rojo's estimator. Proofs of the theorems and derivations of the algorithms are given in the Appendix.

2.2 Estimation of Discrete Survivor Functions

2.2.1 One-sample Case

Consider a discrete failure time variable T_1 with potential failure times $a_1 < \cdots < a_m$ and let $a_0 = 0$ and $a_{m+1} = +\infty$. We are interested in estimating the discrete survivor function $S_1(t)$ based on a right censored sample of T_1. We further suppose that a discrete survivor function $S_2(t)$ with the same potential failure times is given. Our problem is to estimate $S_1(t)$ under the *bounded below constraint* ($S_1(t) \geq S_2(t)$) or the *bounded above constraint* ($S_1(t) \leq S_2(t)$).

The censoring mechanism is assumed independent and the right censored data are summarized by:

d_{1i} the number of events at a_i, $i = 1, \ldots, m$;

n_{1i} the number at risk just prior to a_i, $i = 1, \ldots, m$; and

c_{1i} the number of censored subjects in $[a_i, a_{i+1})$ $i = 0, \ldots, m$.

Let $h_{gi} = \log\left\{S_g(a_i)/S_g(a_{i-1})\right\}, i = 1, \ldots, m$, so that $1 - \exp(h_{gi})$ is the discrete hazard and $\log S_g(a_i) = \sum_{j=1}^{i} h_{gj}$ for $g = 1, 2$, $i =, \ldots, m$. The likelihood of $S_1(a_1), \ldots, S_1(a_m)$ is

$$L(S_1(\cdot)) = S_1(a_0)^{c_{10}} \prod_{i=1}^{m} \left[\left\{S_1(a_{i-1}) - S_1(a_i)\right\}^{d_{1i}} S_1(a_i)^{c_{1i}}\right],$$

and the log likelihood written as a function of $\boldsymbol{h}_1 = (h_{11}, \ldots, h_{1m})$, is

$$logL(\boldsymbol{h}_1) = \sum_{i=1}^{m} \left[d_{1i} \log\left\{1 - \exp(h_{1i})\right\} + (n_{1i} - d_{1i})\, h_{1i}\right].$$

The likelihood is maximized subject to $\sum_{j=1}^{i} h_{1j} \geq \sum_{j=1}^{i} h_{2j}$ or $\sum_{j=1}^{i} h_{1j} \leq \sum_{j=1}^{i} h_{2j}$, $i = 1, \ldots, m$ under bounded below or bounded above constraint respectively. Consider now a particular vector $\boldsymbol{h}_1^* = (h_{11}^*, \ldots, h_{1m}^*)$. In the bounded below case, the i^{th} constraint is said to be *inactive* if $\sum_{j=1}^{i} h_{1j}^* > \sum_{j=1}^{i} h_{2j}$, *active* if $\sum_{j=1}^{i} h_{1j}^* = \sum_{j=1}^{i} h_{2j}$ or *violated* if $\sum_{j=1}^{i} h_{1j}^* < \sum_{j=1}^{i} h_{2j}$. There is a similar definition in the bounded above case.

2.2.2 One-sample Case: Bounded Below Constraint

Dykstra (1982) first proposed a method to obtain the C-NPMLE in the bounded below case, and we next describe the associated theorem and algorithm. In preparation for this, we define a function of k,

$$H(a,b,k) = \sum_{j=a}^{b} \log\Big(1 - \frac{d_{1j}}{n_{1j}+k}\Big) - \sum_{j=a}^{b} h_{2j} \tag{2.1}$$

for a, b integer with $1 \leq a \leq b \leq m$. In (2.1) and elsewhere, if both d_{1j} and $(n_{1j}+k)$ equal to 0, then 0/0 is interpreted as 0. Let $D(a,b) = \max_{a \leq i \leq b} d_{1i}$ and $V(a,b) = \min_{a \leq i \leq b} h_{2i}$. For $a \leq i \leq b$, let $K(a,b)$ be the unique solution of the equation $H(a,b,k) = 0$ if $D(a,b) > 0$ and $V(a,b) < 0$; $K(a,b) = +\infty$ if $D(a,b) > 0$ and $V(a,b) = 0$; $K(a,b) = -\infty$ if $D(a,b) = 0$ and $V(a,b) < 0$; and otherwise $K(a,b) = 0$. Further, let $K^+(a,b) = \max\{K(a,b), 0\}$ and $K^-(a,b) = \max\{-K(a,b), 0\}$.

Theorem 2.1 (Bounded Below Constraint (modified from Dykstra, 1982))**.**
Let $m' = \max(i : n_{1i} > 0)$ *and* $\tau = a_{m'+1}$. *For each* a, b *with* $1 \leq a \leq b \leq m'$, *let* $\hat{k}^i = \min_{a \leq i} \max_{b \geq i} K^+(a,b)$, *and* $\hat{h}_{1i} = \log\{1 - d_{1i}/(n_{1i} + \hat{k}^i)\}$. *Then, the C-NPMLE of* $S_1(\cdot)$ *is* $\hat{S}_1(t) = \exp(\sum_{i:a_i \leq t} \hat{h}_{1i})$, $t < \tau$. *The C-NPMLE can be defined arbitrarily for* $t \geq \tau$ *subject to the constraint.*

In this theorem, $\hat{k}^i$ is a nonnegative real number. It can be shown that $\hat{k}^1 \geq \hat{k}^2 \geq \cdots$. This theorem gives a method of obtaining the MLE, however, the $\hat{\boldsymbol{h}}_1$ can be obtained more easily using an algorithm that was presented by Dykstra (1982) (see Algorithm 2.2). While this algorithm does give the C-NPMLE, we developed (see Algorithm 2.3) a computationally more efficient version of it.

Algorithm 2.2 (Bounded Below Constraint (modified from Dykstra, 1982))**.**

1. *Set* $i_0 = 0$*,* $\ell = 1$ *and* $m' = \max\{i : n_{1i} > 0\}$*.*
2. *If there exists* $b > i_{\ell-1}$ *such that* $K(i_{\ell-1}+1, b) > 0$*, then let* $i_\ell = \min\left\{\arg\max_{b>i_{\ell-1}} K(i_{\ell-1}+1, b)\right\}$ *and* $k_\ell = K(i_{\ell-1}+1, i_\ell)$*. Otherwise, let* $i_\ell = m'$*,* $k_\ell = 0$*.*
3. *Let* $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} + k_\ell)\}, i_{\ell-1} + 1 \leq j \leq i_\ell$*.*
4. *If* $i_\ell = m'$*, stop. Otherwise, set* $\ell = \ell + 1$ *and go to step 2.*

Note that this algorithm gives a KM type estimator in which the number at risk is potentially modified at each potential failure time. It can be shown that $k_1 \geq k_2 \geq \cdots$, so that this estimate is essentially a KM estimate based on modified data where k_1 more subjects are placed at risk at time 0, and at time a_{i_ℓ}, $k_\ell - k_{\ell+1}, \ell = 1, 2, \ldots$ of these additional subjects are censored.

Step 2 in Algorithm 2.2 is looking for the next active constraint in the solution. A root finding procedure is needed to calculate $K(i_{\ell-1} + 1, b)$. To find a root with high precision is computationally intensive, so it is inefficient to calculate $K(i_{\ell-1} + 1, b)$ for all $b > i_{\ell-1}$ to find the index of the next active constraint. Instead we propose another algorithm that is equivalent to Algorithm 2.2 but only calculates $K(i_{\ell-1} + 1, b)$ when necessary.

Algorithm 2.3 (Bounded Below Constraint (a more efficient algorithm))**.**

1. *Set* $i_0 = 0$*,* $\ell = 1$ *and* $m' = \max(i : n_{1i} > 0)$*.*
2. *Let* $i_\ell = \min_{b>i_{\ell-1}} \{b : H(i_{\ell-1} + 1, b, 0) < 0\}$*, then set* $k_\ell = K(i_{\ell-1} + 1, i_\ell)$*. If no such* i_ℓ *exists, set* $i_\ell = m'$ *and* $k_\ell = 0$ *and go to step 4.*

3. *Let* $I = \min_{b>i_\ell}\{b : H(i_\ell+1, b, k_\ell) < 0\}$. *If no such* I *exists, then go to step 4. Otherwise, set* $i_\ell = I$ *and repeat step 3.*

4. *Let* $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j}+k_\ell)\}, i_{\ell-1}+1 \le j \le i_\ell$.

5. *If* $i_\ell = m'$*, stop. Otherwise, set* $\ell = \ell + 1$ *and go to step 2.*

The two algorithms are equivalent because steps 2 and 3 in Algorithm 2.3 are looking for $\min\{\arg\max_{b>i_{\ell-1}} K(i_{\ell-1}+1, b)\}$ as in Algorithm 2.2. However, Algorithm 2.3 implements a root finding procedure only when it finds a position b, where $K(i_{\ell-1}+1, b)$ is larger than the previously found maximum $K(i_{\ell-1}+1, i_\ell)$. This significantly improves the efficiency of the calculations.

2.2.3 One-sample Case: Bounded Above Constraint

For the bounded above constraint, Dykstra (1982) presented a theorem to obtain the C-NPMLE that is similar to the theorem of the bounded below constraint, except for an adjustment for an "exception" that may happen before the first event time. For more details, see Dykstra (1982). The proof, however, did not consider some data configurations in which the "exception" can occur at later times and the method does not always yield a C-NPMLE. In this section, we present a correct theorem and algorithm for this type of constraint in the discrete case. Then in the next section we discuss cases where $S_1(t)$ and $S_2(t)$ are not necessarily discrete functions.

The C-NPMLE may not be unique even before the last observed time. To circumvent this, we define the maximum C-NPMLE (MC-NPMLE) and then present a theorem and an algorithm to obtain the MC-NPMLE under bounded above constraint.

Definition 2.4. The maximum C-NPMLE (MC-NPMLE) is the C-NPMLE that maximizes the estimate of the survivor function in the class of all C-NPMLEs.

We demonstrate in Appendix B that the MC-NPMLE exists and is unique.

Theorem 2.5 (Bounded Above Constraint)**.** *Let* $m' = \max(i : n_{1i} > 0)$. *For each* a, b *with* $1 \le a \le b \le m'$, *let* $\hat{k}^i = \min_{a \le i} \max_{b \ge i} \min(K^-(a,b), n_{1b})$, *and sequentially define*

$$\hat{h}_{1i} = \begin{cases} \log\{1 - d_{1i}/(n_{1i} - \hat{k}^i)\} & d_{1i} > 0 \\ \min\left\{0,\ \sum_{j=1}^{i} h_{2j} - \sum_{j=1}^{i-1} \hat{h}_{1j}\right\} & d_{1i} = 0 \end{cases}$$

then, the MC-NPMLE of $S_1(\cdot)$ *is* $\hat{S}_1(t) = \exp\left(\sum_{i:a_i \le t} \hat{h}_{1i}\right)$, $t < \tau$, *where* $\tau = a_{m'+1}$.

Even though the C-NPMLE may not be unique, it can be shown that the difference between the MC-NPMLE and any other C-NPMLE can only be at times where $d_{1i} = 0$ and $n_{1i} = \hat{k}^i$.

The major difference between this theorem and Dykstra's incorrect result is in the definition of $\hat{k}^i$. Dykstra defined $\hat{k}^i = \min_{a \le i} \max_{b \ge i} K^-(a,b)$. However, $\hat{k}^i$ cannot be larger than the number at risk at any position between a and b in the solution of a C-NPMLE. For the right censored case, $\hat{k}^i = \min_{a \le i} \max_{b \ge i} \min\{K^-(a,b), n_{1b}\}$ because n_{1b} is the smallest in this range.

A computationally efficient algorithm that obtains $\hat{k}^i$ easily is given by:

Algorithm 2.6 (Bounded Above Constraint)**.**

1. *Set* $i_0 = 0$, $\ell = 1$, $m' = \max(i : n_{1i} > 0)$.

2. *Let* $i_\ell = \min_{b > i_{\ell-1}}\{b : H(i_{\ell-1} + 1, b, 0) > 0\}$. *If no such* i_ℓ *exists, go to step 7, otherwise go to step 3.*

3. *If* $d_{1i_\ell} = 0$ *and* $H(i_{\ell-1} + 1, i_\ell, -n_{1i_\ell}) \ge 0$, *then set* $k_\ell = n_{1i_\ell}$ *and go to step 5, otherwise set* $k_\ell = -K(i_{\ell-1} + 1, i_\ell)$ *and go to step 4.*

4. *Let* $I = \min_{b>i_\ell}\{b : n_{1b} > k_\ell \text{ and } H(i_\ell + 1, b, -k_\ell) > 0\}$. *If no such* I *exists, then go to step 5. Otherwise, set* $i_\ell = I$ *and go to step 3.*

5. *Let* $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} - k_\ell)\}, i_{\ell-1} + 1 \le j \le i_\ell - 1$

   $\hat{h}_{1i_\ell} = \sum_{j=i_{\ell-1}+1}^{i_\ell} h_{2j} - \sum_{j=i_{\ell-1}+1}^{i_\ell - 1} \hat{h}_{1j}.$

6. *If* $i_\ell = m'$*, stop. Otherwise, set* $\ell = \ell + 1$ *and go to step 2.*

7. *Let* $\hat{h}_{1j} = \log(1 - d_{1j}/n_{1j}), i_{\ell-1} + 1 \le j \le m'$*, stop.*

Heuristically, the solution is a Kaplan-Meier type estimator of modified data with right censoring and left truncation. For i with $i_{\ell-1} < i \le i_\ell$, the number at risk is modified to $n_{1i} - k_\ell$. Since $k_1 \ge k_2 \ge \cdots$, the modified data can be described as k_1 subjects being removed at time 0 and replaced over time as left truncated data. In particular, $k_\ell - k_{\ell+1}$ are added (left truncated) at $a_{i_\ell}+, \ell = 1, 2, \cdots$. Note that the number at risk in the modified data can be zero at some times, and when this occurs, the C-NPMLE may not be unique. This corresponds to the result that, for left truncated data, the MLE is not unique when the number at risk is zero at an intermediate point.

Example 2.7 (One-sample case with a bounded above constraint)**.** Suppose we have observed survival times 1, 2, 2.5+, 3, 3.5+, 4.5+, 5.5+, 6.5+, 9, 11.5+ (+ denotes censoring). We assume the potential event times are integers from 1 to 12, $S_1(t) \le S_2(t), t = 1, \ldots, 12$, and the values for $S_2(t)$ are given in Table 2.1.

Table 2.1 gives KM estimate, $S_1^*(t)$, and the MC-NPMLE, $\hat{S}_1(t)$. The values of $\hat{k}^i$ are from Theorem 2.5 and those of k_ℓ are from Algorithm 2.6. The active constraints are at times 6, 7 and 8. As indicated above, if $i = i_{\ell-1} + 1, \ldots, i_\ell$, then $\hat{k}^i = k_\ell$. For example, $\hat{k}^1 = \cdots = \hat{k}^6 = k_1$. As can be seen in Figure 2.1, the major difference between the MC-NPMLE and the D-estimator is that the former has jumps in the estimate for population 2

Table 2.1: Results for Example 2.7. a_i is potential event time with corresponding number of event d_{1i} and number at risk n_{1i}, S_2 is constraint, S_1^* is KM estimate, $\hat{S}_1$, $\hat{h}_{1i}$ and k_ℓ are results from Algorithm 2.6, $\hat{k}^i$ is from Theorem 2.5, S_1^D is D-estimates (Dykstra, 1982), and the last observed time $\tau = 11.5$.

a_i	1	2	3	4	5	6	7	8	9	10	11	12
d_{1i}	1	1	1	0	0	0	0	0	1	0	0	0
n_{1i}	10	9	7	5	4	4	2	2	2	1	1	0
S_2	0.94	0.92	0.86	0.68	0.54	0.40	0.36	0.32	0.30	0.26	0.22	0.20
S_1^*	0.90	0.80	0.69	0.69	0.69	0.69	0.69	0.69	0.34	0.34	0.34	0.34
$\hat{S}_1$	0.86	0.71	0.54	0.54	0.54	0.40	0.36	0.32	0.16	0.16	0.16	0.16
$\hat{h}_{1i}$	-0.15	-0.18	-0.29	0	0	-0.29	-0.11	-0.12	-0.69	0	0	0
k_ℓ	$\longleftarrow k_1 = 3 \longrightarrow$						$k_2 = 2$	$k_3 = 2$	$\longleftarrow k_4 = 0 \longrightarrow$			
$\hat{k}^i$	3	3	3	3	3	3	2	2	0	0	0	0
S_1^D	0.804	0.609	0.32	0.32	0.32	0.32	0.32	0.32	0.16	0.16	0.16	0.16

at time 5, 6,7 and 8, where there is no observed event for that population. As expected, the MC-NPMLE has larger log likelihood value than the D-estimator (-12.4 versus -13.22) and is closer to the KM estimator.

2.2.4 Two-sample Case

The notation is similar to that in the one-sample case except that $S_2(a_i)$ also needs to be estimated so that d_{2i}, n_{2i} and $c_{2i}, 1 \leq i \leq m$ are also observed.

The likelihood of $S_1(a_1), \ldots, S_1(a_m), S_2(a_1), \ldots, S_2(a_m)$ is

$$L(S_1(\cdot), S_2(\cdot)) = \prod_{g=1}^{2} \Big[S_g(a_0)^{c_{g0}} \prod_{j=1}^{m} \{S_g(a_{j-1}) - S_g(a_j)\}^{d_{gj}} \, S_g(a_j)^{c_{gj}} \Big],$$

and the corresponding log likelihood of $\boldsymbol{h}_g = (h_{g1}, \ldots, h_{gm}), g = 1, 2$ is

$$\log L(\boldsymbol{h}_1, \boldsymbol{h}_2) = \sum_{g=1}^{2} \sum_{i=1}^{m} [d_{gi} \log \{1 - \exp(h_{gi})\} + (n_{gi} - d_{gi})\, h_{gi}]. \tag{2.2}$$

A C-NPMLE in the two-sample case is an estimator that maximizes the log likelihood (2.2) subject to the constraints, $\sum_{j=1}^{i} h_{1j} \geq \sum_{j=1}^{i} h_{2j}$, and $h_{1i}, h_{2i} \leq 0, i = 1, \ldots, m.$

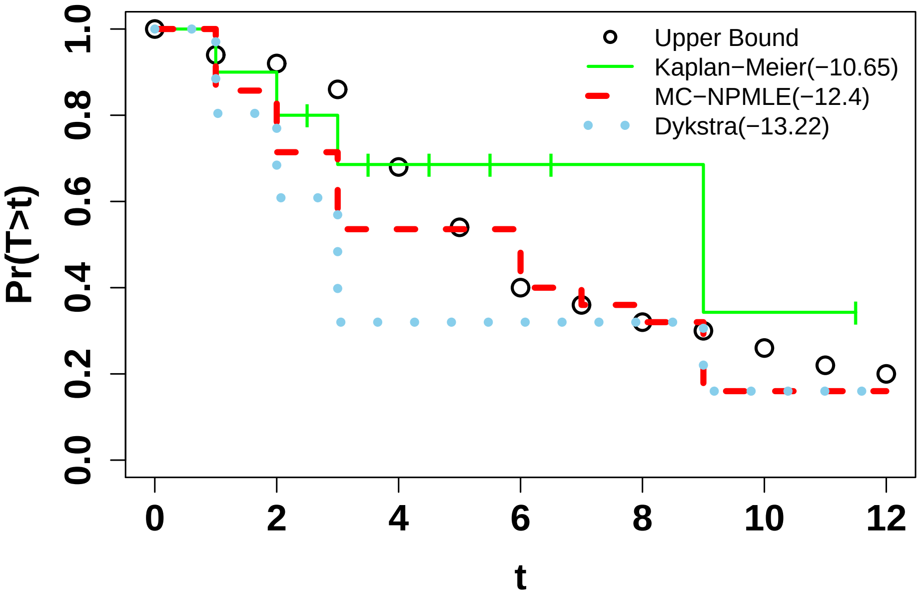


Figure 2.1: Estimates of survivor functions under bounded above constraint in discrete situation. Log likelihood values shown in parentheses.

A method for the two-sample case was described by Dykstra (1982), but has the same problem as for the bounded above constraint. Here again the C-NPMLE of the lower survivor function may not be unique. We propose a theorem and an algorithm to estimate the C-NPMLE of $S_1(t)$ and the MC-NPMLE of $S_2(t)$. In preparation for this, we define

$$H_2(a,b,k) = \sum_{j=a}^{b} \log\left(1 - \frac{d_{1j}}{n_{1j}+k}\right) - \sum_{j=a}^{b} \log\left(1 - \frac{d_{2j}}{n_{2j}-k}\right), \ 1 \le a \le b \le m.$$

Let $D_g(a,b) = \max_{a \le j \le b} d_{gj}, g = 1, 2$. For $a \le i \le b$, let $K_2(a,b)$ be the unique solution of the equation $H_2(a,b,k) = 0$ if $D_1(a,b) > 0$ and $D_2(a,b) > 0$; $K_2(a,b) = +\infty$ if $D_1(a,b) > 0$ and $D_2(a,b) = 0$; $K_2(a,b) = -\infty$ if $D_1(a,b) = 0$ and $D_2(a,b) > 0$; and otherwise $K_2(a,b) = 0$. Further, let $K_2^+(a,b) = \max\{K_2(a,b), 0\}$.

Theorem 2.8 (Two-sample case). *Let* $m' = \max\{i : n_{1i} > 0, n_{2i} > 0\}$. *For each* a, b *with* $1 \le a \le b \le m'$, *let* $\hat{k}^i = \min_{a \le i} \max_{b \ge i} \min\{K_2^+(a,b), n_{2b}\}$, $i \le m'$ *and* $\hat{k}^i = 0$ *if* $i > m'$. *Let*

$$\hat{h}_{1i} = \log\{1 - d_{1j}/(n_{1j} + \hat{k}^i)\}$$

$$\hat{h}_{2i} = \begin{cases} \log\{1 - d_{2j}/(n_{2j} - \hat{k}^i)\} & d_{2i} > 0 \textit{ or } i > m' \\ \min(0, \sum_{j=1}^{i} \hat{h}_{1j} - \sum_{j=1}^{i-1} \hat{h}_{2j}) & d_{2i} = 0 \textit{ and } i \leq m'. \end{cases}$$

The C-NPMLE of $S_1(\cdot)$ *is* $\hat{S}_1(t) = \exp(\sum_{i:a_i \leq t} \hat{h}_{1i}), t < \tau_1$*, and the MC-NPMLE of* $S_2(\cdot)$ *is* $\hat{S}_2(t) = \exp(\sum_{i:a_i \leq t} \hat{h}_{2i}), t < \tau_2$*, where* $\tau_g = \max(a_{i+1} : n_{gi} > 0)$*.* $S_g(t), g = 1, 2$ *can be defined arbitrarily for* $t > \tau_g$ *subject to the constraint,* $g = 1, 2$*.*

We also describe an efficient algorithm to obtain the solution.

Algorithm 2.9 (Two-Sample Case)**.**

```
1. $i_0 = 0, \ell = 1$, $m_g = \max(i : n_{gi} > 0)$, $g = 1, 2$ and $m' = \min(m_1, m_2)$
2. Let $i_\ell = \min\{b : H_2(i_{\ell-1} + 1, b, 0) < 0\}$. If no such $i_\ell$ exists, go to step 7, otherwise go to step 3.
3. If $d_{2i_\ell} = 0$ and $H_2(i_{\ell-1} + 1, i_\ell, n_{2i_\ell}) \leq 0$, then set $k_\ell = n_{2i_\ell}$ and go to step 5, otherwise set $k_\ell = K_2(i_{\ell-1} + 1, i_\ell)$ and go to step 4.
4. Let $I = \min_{b > i_\ell}\{b : n_{2b} > k_\ell$ and $H_2(i_\ell + 1, b, k_\ell) < 0\}$. If no such $I$ exists, then go to step 5. Otherwise, set $i_\ell = I$ and go to step 3.
5. Let $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} + k_\ell)\},\ i_{\ell-1} + 1 \leq j \leq i_\ell$
       $\hat{h}_{2j} = \log\{1 - d_{2j}/(n_{2j} - k_\ell)\},\ i_{\ell-1} + 1 \leq j \leq i_\ell - 1$
       $\hat{h}_{2i_\ell} = \sum_{j=i_{\ell-1}+1}^{i_\ell} \hat{h}_{1j} - \sum_{j=i_{\ell-1}+1}^{i_\ell - 1} \hat{h}_{2j}$.
6. If $i_\ell = m'$, go to step 7. Otherwise, set $\ell = \ell + 1$ and go to step 2.
7. For $g$, such that $m_g > m'$, set $\hat{h}_{gj} = \log(1 - d_{gj}/n_{gj}), i_{\ell-1} < j \leq m_g$, stop.
```

As Dykstra (1982) described, the solution for population 1 can be obtained from a modified data set in which $\hat{k}^i$ of the observations from population 2 that are at risk at time a_i are taken as observations from population 1. From this point of view, $\hat{k}_i$ must be less than or equal to n_{2i}, because population 2 in the modified data can not have a negative at risk number at any time. The solution of population 2, however, may not be obtained from its own modified data and depends on the solution for population 1.

2.3 Extension to Non-discrete Case

In this section, we extend the results of section 2 so as not to assume known potential event times. Thus, in the one-sample case, suppose $S_2(t)$ is a known survivor function defined on $[0, +\infty)$ and we seek the C-NPMLE under bounded below and bounded above constraints. In the two-sample case, we estimate both $S_1(t)$ and $S_2(t)$ from data subject to $T_1 \geq_{st} T_2$.

Let $\{X_1, \ldots, X_N\}$ be the union of all distinct observed event times from the populations 1 and 2, and set $X_0 = 0, X_{N+1} = \infty$ for convenience. The number of events at X_i in population g is d_{gi}. Let $C_{gi}^{(j)}$, $g = 1, 2$,$i = 1, \ldots, N$, $j = 1, \ldots, m_{gi}$ be the censoring times of population g in $[X_i, X_{i+1})$, $i = 0, \ldots, N$, and let (Y_{gi}, Δ_{gi}) be observations, where Y_{gi} is the observed time and Δ_{gi} is the event indicator ($\Delta_{gi} = 1$ if event occurred or $\Delta_{gi} = 0$ if right censored).

Proceeding as in Johansen (1978), the likelihood to maximize, subject to constraint, is

$$\begin{aligned} L(S_1(\cdot), S_G(\cdot)) &= \prod_{g=1}^{G} \prod_{i=1}^{n_g} \{S_g(Y_{gi}-) - S_g(Y_{gi})\}^{\Delta_{gi}} S_g(Y_{gi})^{1-\Delta_{gi}} \\ &= \prod_{g=1}^{G} \Bigg(\prod_{j=1}^{m_{g0}} S_g(C_{g0}^{(j)}) \prod_{i=1}^{N} \Big[\{S_g(X_i-) - S_g(X_i)\}^{d_{gi}} \prod_{j=1}^{m_{gi}} S_1(C_{gi}^{(j)}) \Big] \Bigg) \end{aligned} \tag{2.3}$$

where $G = 1$ for the one-sample case and $G = 2$ for the two-sample case.

2.3.1 One-sample Case with $T_1 \geq_{st} T_2$ and $S_2(t)$ is Known

If any survivor function $\bar{S}(t)$ satisfying the stochastic ordering constraint is replaced by a discrete $\hat{S}(t)$ having possible jumps only at observed event times and $\hat{S}(X_i) = \bar{S}(X_i)$ for all i, the likelihood will not decrease and the constraint is not violated since $\hat{S}(t) \geq \bar{S}(t)$ for all t. Thus, the C-NPMLE can be obtained within the class of survivor functions with jumps only at observed event times. So Algorithm 2.3 can be directly used by setting the observed event times to potential event times and estimating $\hat{S}_1(t)$ for $t \leq \max(Y_{1i})$, the last observed time.

2.3.2 One-sample Case with $T_1 \leq_{st} T_2$ and $S_2(t)$ is Known

This is the most complicated case. The discrete method is not easy to apply because the constraint may be relevant at all times if $S_2(t)$ is not a step function.

One way to obtain a C-NPMLE of $S_1(t)$ is the "limit method", in which we use the limit of a discrete function to approach a continuous one. For example, we choose R evenly spaced times between 0 and $\max(Y_{1i})$ as potential event times, apply the bounded above constraint at these R times, and obtain the limiting estimate of $\hat{S}_1(t)$ with Algorithm 2.6 as R goes to infinity. However, this method is computationally intensive. We propose instead a method that, through judicious selection of a finite number of appropriate potential event times, yields the MC-NPMLE as described in Algorithm 2.10.

Algorithm 2.10. *Let $C_i, i = 1, \cdots, n_c$ be all distinct observed censoring times and let X_i^- be the time just before observed event time X_i.*

1. *Let $X_i', i = 1, 2, \cdots$ be the distinct ordered times from the union of X_i, X_i^- nd C_i.*

2. *Estimate $\hat{S}_1(t)$, which is the MC-NPMLE with potential event times at all X_i', using Algorithm 2.6.*

3. $\tilde{S}_1(t) = \min\{\hat{S}_1(t), S_2(t)\}$.

In practice, we set $X_i^- = X_i$ but order X_i^- before X_i.

Theorem 2.11. *$\tilde{S}_1(t)$ from Algorithm 2.10 is the MC-NPMLE of $S_1(t)$ subject to $T_1 \le_{st} T_2$.*

Proof. First, we show that $\tilde{S}_1(t)$ is a C-NPMLE. Let $\bar{S}_1(t)$ be a C-NPMLE subject to $T_1 \le_{st} T_2$. Note that $\hat{S}_1(t)$ is the maximum likelihood estimator subject to fewer constraints (only at times X_i') compared to $\bar{S}_1(t)$, we have that $L(\hat{S}_1(t)) \ge L(\bar{S}_1(t))$. Further $L(\bar{S}_1(t)) \ge L(\tilde{S}_1(t))$ since $\tilde{S}_1(t) = \min(\hat{S}_1(t), S_2(t)) \le S_2(t)$. Note that at every time X_i', $i = 1, \cdots, n_{tot}$, $\hat{S}_1(X_i') \le S_2(X_i')$, the difference between $\hat{S}_1(t)$ and $\tilde{S}_1(t)$ may only occur in time interval (X_i', X_{i+1}') for some i. The five possible time intervals are (C, C), (C, X^-), (X, C) (X^-, X) and (X, X^-), where C represents censoring time, X event time and X^- time just before X. None of these intervals includes C, X or X^-, the three elements that determine likelihood (2.3). So $L(\tilde{S}_1(t)) = L(\hat{S}_1(t)) = L(\bar{S}_1(t))$, which implies that $\tilde{S}_1(t)$ is a C-NPMLE subject to $T_1 \le_{st} T_2$.

Then, we show that $\tilde{S}_1(t)$ is the MC-NPMLE. Suppose it is not, we must be able to find a time x^* where $\bar{S}_1(x^*) > \tilde{S}_1(x^*) = \min\{\hat{S}_1(x^*), S_2(x^*)\}$. Then $\bar{S}_1(x^*) > \hat{S}_1(x^*)$ since $\bar{S}_1(x^*) \le S_2(x^*)$. Consider another survivor function $S_1'(t)$ with jumps only at the times X_i' and $S_1'(X_i') = \bar{S}_1(X_i')$ for all i, $S_1'(t)$ is constrained estimator of $S_1(t)$ subject to discrete constraint at all X_i's. Since $S_1'(x^*) = S_1'(\max(X_i' : X_i' \le x^*)) = \bar{S}_1(\max(X_i' : X_i' \le x^*)) \ge \bar{S}_1(x^*) > \hat{S}_1(x^*)$ and $\hat{S}_1(t)$ is the MC-NPMLE with discrete constraint, $S_1'(t)$ is not a C-NPMLE subject to the discrete constraint. So $L(\bar{S}_1(t)) = L(S_1'(t)) < L(\hat{S}_1(t)) = L(\tilde{S}_1(t))$, which is a contradiction. Thus, $\tilde{S}_1(t)$ is the MC-NPMLE. □

Example 2.12 (One-sample case with a bounded above constraint, continuous case)**.** In Example 2.7, suppose we take $S_2(t)$ to be piecewise linear with knots at the discrete points in Table 2.1.

Table 2.2: Results for Example 2.12. The X_i', d_{1i} and n_{1i} are defined in Algorithm 2.10, S_2 is constraint, S_1^* is KM estimate, and $\hat{S}_1$ and $\hat{h}_{1i}$ are results from Algorithm 2.6.

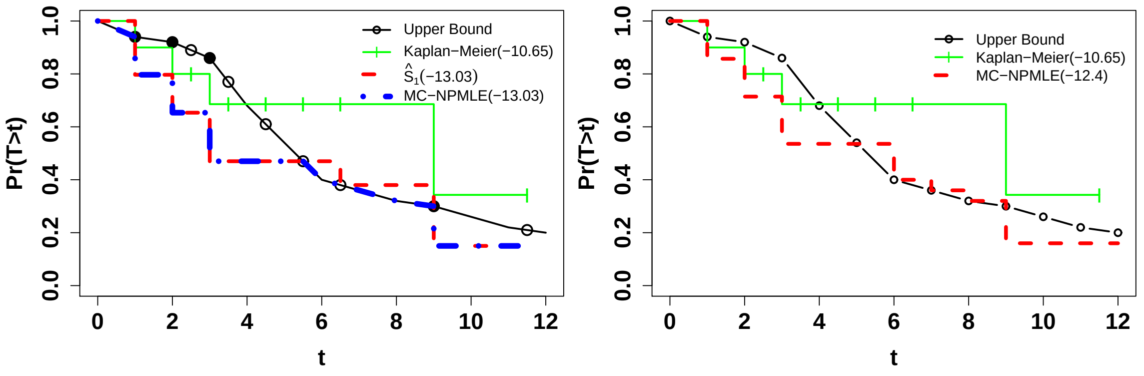
X_i'	1-	1	2-	2	2.5	3-	3	3.5	4.5	5.5	6.5	9-	9	11.5
d_{1i}	0	1	0	1	0	0	1	0	0	0	0	0	1	0
n_{1i}	10	10	9	9	8	7	7	6	5	4	3	2	2	1
S_1^*	1	0.90	0.90	0.80	0.80	0.80	0.69	0.69	0.69	0.69	0.69	0.69	0.34	0.34
S_2	0.94	0.94	0.92	0.92	0.89	0.86	0.86	0.77	0.61	0.38	0.34	0.30	0.30	0.21
$\hat{S}_1$	0.94	0.80	0.80	0.65	0.65	0.65	0.47	0.47	0.47	0.47	0.38	0.30	0.15	0.15
$\hat{h}_{1i}$	-0.06	-0.18	0	-0.22	0	0	-0.41	0	0	-0.10	-0.11	-0.13	-0.69	0

The procedure for calculating the MC-NPMLE in this case with Algorithm 2.10 is illustrated in Table 2.2 and Figure 2.2a. First, choose X_i' as in the first row of Table 2.2 and find d_{1i} and n_{1i}, the number of events and the number at risk at time X_i'. Algorithm 2.6 with potential event times X_i' is used to calculate $\hat{S}_1(t)$ as shown in the sixth row in Table 2.2 and the dash plot in figure 2.2a. Finally, calculate the MC-NPMLE as $\min\{\hat{S}_1(t), S_2(t)\}$ as shown with the thick dot-dash curve in figure 2.2a. Note that the MC-NPMLE is not a step function in this example.

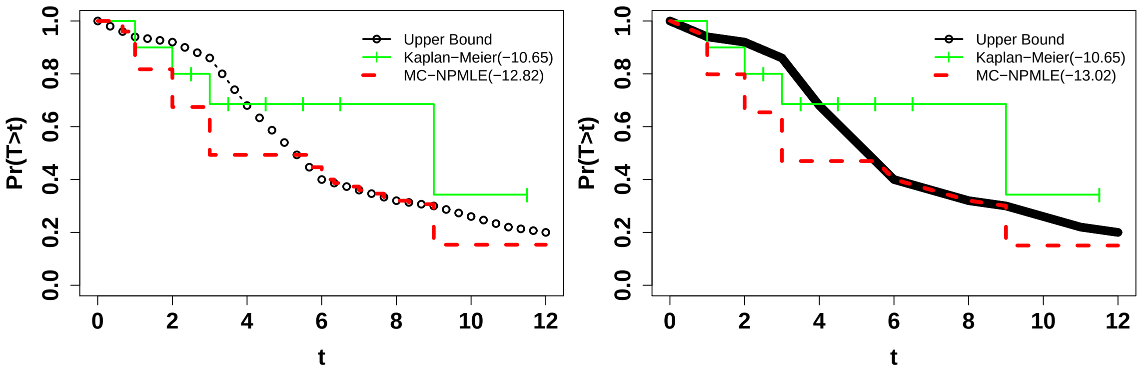
Figure 2.2b, 2.2c, and 2.2d are the plots using the “limit method” with 12, 36 and 360 potential event times respectively. As the number of potential points increases, the estimate becomes closer to the MC-NPMLE obtained through Algorithm 2.10. The log likelihood is -12.4 with 12 potential event times and decreases to -13.02 with 360 potential event times. This will approach -13.03, the same as that from Algorithm 2.10, as the number of potential event times goes to infinity.

2.3.3 Two-sample Case With No Potential Event Times

Consider any pair of survivor functions $\bar{S}_1(t)$ and $\bar{S}_2(t)$ satisfying the stochastic ordering constraint $S_1(t) \geq S_2(t)$ for all t. If we replace these by discrete survivor functions $\hat{S}_1(t)$ and $\hat{S}_2(t)$ with possible jumps at observed event times, $X_1, \ldots, X_N$, and $\hat{S}_g(X_i) =$



(a) Algorithm 2.10 (b) Limit approach 12 potential points



(c) Limit approach with 36 potential points (d) Limit approach 360 potential points

Figure 2.2: Estimates of survivor functions under continuous bounded above constraint. Log likelihood values shown in parentheses.

$\bar{S}_g(X_i)$, $g = 1, 2$, $i = 1, \ldots, N$, the likelihood cannot decrease and the constraint is not violated anywhere. Thus, the C-NPMLE can be obtained in the class of survivor functions with jumps only at observed event times, which is the same as obtaining the C-NPMLE in the discrete case. Theorem 2.8 and Algorithm 2.9 can be directly used to obtain $\hat{S}_1(t)$ and $\hat{S}_2(t)$ with the observed event times as the potential event times and we estimate $S_g(t)$ for $t \leq \max_i(Y_{gi}), g = 1, 2$. Note that $\hat{S}_1(t)$ is the unique C-NPMLE of $S_1(t)$ and $\hat{S}_2(t)$ is the unique MC-NPMLE of $S_2(t)$. As expected as the sample size of one group (n_1 or n_2) becomes very large, the two-sample case estimator for the other group approaches the corresponding one-sample estimator in section 2.3.1 and 2.3.2.

2.4 Consistency

Dykstra (1982) presented a proof of consistency of the D-estimator. Here, we give a proof of consistency of the corrected C-NPMLE.

Suppose that $(C_{gi}, T_{gi}), g = 1, 2, i = 1, \cdots, n_g$ are independent and identically distributed with $C_{gi} \perp T_{gi}$. Suppose that $\tau > 0$ satisfies $P(\min\{C_{gi}, T_{gi}\} > \tau) = 0$. Let n_1 and n_2 be the respective sample size of sample 1 and 2. Meier (1975) has shown that the product limit estimator $S_g^*(t)$ of the survivor function $S_g(t)$ satisfies

$$\lim_{n_g \to \infty} P\Big\{\sup_{x \le \tau} |S_g^*(x) - S_g(x)| > \epsilon\Big\} = 0 \tag{2.4}$$

for any given $\epsilon > 0$.

Theorem 2.13. *Suppose the data are from the iid case as described above, where $S_1(t) \geq S_2(t)$ for all t. Suppose further that $S_g(\tau) > 0, g = 1, 2$ for given $\tau > 0$. Then*

$$\lim_{n_g \to \infty} P(\sup_{t \le \tau} |\hat{S}_g(t) - S_g(t)| > \epsilon) = 0 \tag{2.5}$$

for any given $\epsilon > 0$.

Proof. See Appendix A. □

2.5 Example

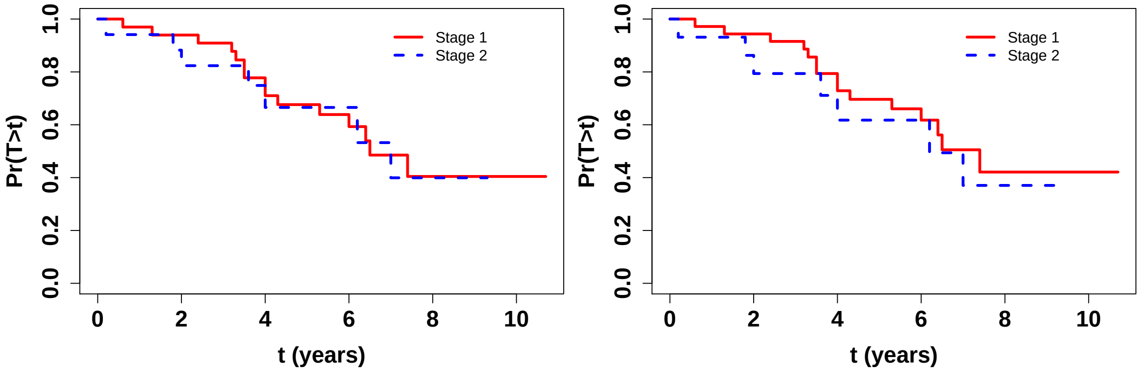
This example is a case study of survival times from diagnosis of male larynx cancer patients (Kardaun, 1983). We analyze the data from the patients with larynx cancer stages 1 and 2, which are shown in Table 2.3.

Table 2.4 summarizes the results. $S_1^*(t)$ and $S_2^*(t)$ are KM estimates for stage 1 and stage 2 patients respectively and are plotted in Figure 2.3a. The two plots cross each other, indicating that there exist violations of the stochastic ordering constraint $T_1 \geq_{st} T_2$. Times

Table 2.3: Survival times (in Years) of male patients with larynx cancer stage 1 and stage 2 in example in section 2.5. + represents censoring.

stage 1							stage 2			
0.6	3.2+	4.0	5.3	6.1+	6.7+	8.1+	0.2	3.3+	4.3+	7.6+
1.3	3.3	4.0	5.5+	6.2+	7.0+	9.6+	1.8	3.6	5.0+	9.3+
2.4	3.3+	4.3	5.9+	6.4	7.4	10.7+	2.0	3.6+	6.2	
2.5+	3.5	4.5+	5.9+	6.5	7.4+		2.2+	4.0	7.0	
3.2	3.5	4.5+	6.0	6.5+	8.1+		2.6+	4.3+	7.5+	

$0.2, 0.6, \ldots, 7.4$ in Table 2.4 (first and sixth rows) are distinct observed event times. We set these as potential event times and calculate $\hat{S}_1(t)$ and $\hat{S}_2(t)$ from Algorithm 2.9. The remaining times 9.3 and 10.7 are the last observed censoring times of population 2 and 1. Figure 2.3b shows the C-NPMLE of survivor functions subject to $S_1(t) \geq S_2(t)$ for each group. The D-estimator is the same as the C-NPMLE in this case. Compared to the plots in Figure 2.3a, we can see that the effect of the constraint is to make $\hat{S}_1(t)$ larger than $S_1^*(t)$ and $\hat{S}_2(t)$ smaller than $S_2^*(t)$ for all $t > 0$. The estimates of median life times of stage 1 and stage 2 patients are 6.5 and 7.0 months from KM estimators respectively, which contradicts our belief about cancer stages; the corresponding estimates of 7.4 and 6.2 months from C-NPMLEs, are more realistic.



(a) KM Estimator

(b) C-NPMLE

Figure 2.3: Estimates of survivor functions, two-sample case.

Table 2.4: KM and C-NPMLE estimates of survivor functions for male patients with larynx cancer in example in section 2.5. S_1^* and S_2^* are KM estimates and $\hat{S}_1$ and $\hat{S}_2$ are estimates from the C-NPMLE. ** denotes last observed time for each group.

t	.2	.6	1.3	1.8	2.0	2.4	3.2	3.3	3.5	3.6
$\hat{S}_1(t)$	1	0.972	0.944	0.944	0.944	0.915	0.886	0.856	0.794	0.794
$\hat{S}_2(t)$	0.931	0.931	0.931	0.863	0.794	0.794	0.794	0.794	0.794	0.711
$S_1^*(t)$	1	0.970	0.939	0.939	0.939	0.909	0.878	0.845	0.778	0.778
$S_2^*(t)$	0.941	0.941	0.941	0.882	0.824	0.824	0.824	0.824	0.824	0.749

t	4.0	4.3	5.3	6.0	6.2	6.4	6.5	7.0	7.4	9.3	10.7
$\hat{S}_1(t)$	0.729	0.696	0.660	0.617	0.617	0.561	0.505	0.505	0.421	0.421	0.421**
$\hat{S}_2(t)$	0.617	0.617	0.617	0.617	0.494	0.494	0.494	0.370	0.370	0.370**	
$S_1^*(t)$	0.710	0.676	0.639	0.593	0.593	0.539	0.485	0.485	0.404	0.404	0.404**
$S_2^*(t)$	0.665	0.665	0.665	0.665	0.532	0.532	0.532	0.399	0.399	0.399**	

2.6 Confidence Intervals

If the true survivor functions $S_1(t)$ and $S_2(t)$ are never equal, then the asymptotic variances of $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are the same as those of KM estimators. This motivate one possible method to obtain confidence intervals using asymptotic variances. In this method the confidence intervals for unrestricted estimators are shifted and centered on the constrained estimators (Hwang and Peddada, 1994). We apply this idea on a log transformed scale in this paper and consider the following approximate confidence interval,

$$\hat{S}_g(x)\exp\{\pm 1.96\sigma_g^*(x)\},$$

where $\sigma_g^*(x)$ is standard error estimate of $\log S_g^*(x)$ (see Kalbfleisch and Prentice 2002, page 17).

Another possible approach to construct confidence intervals is to use the bootstrap methods. In this paper, we consider two sampling schemes, a 'standard' and a 'restricted' scheme. In the standard scheme, survival time and censoring indicator pairs are sampled with replacement within each group whereas in the restricted scheme (denoted R-Roobstrap), event times are drawn from the distribution $1 - \hat{S}_g(t)$ and censoring times are drawn from $1 - S_g^{c*}(t)$, where $S_g^{c*}(t)$ is the KM estimate of censoring survivor function for group g. For each bootstrap sample, a bootstrap estimate $\hat{S}_g^b(t), b = 1, \ldots, B$ is obtained by applying the C-NPMLE or the MC-NPMLE. Confidence intervals based on these bootstrap estimates can be constructed using percentiles (Efron, 1981) or a bias corrected percentiles (Efron, 1987). In our simulation study, we conduct the bias corrected (BC) bootstrap on the $\arcsin(\sqrt{s})$ transformed scale for $s = S_g(t)$.

2.7 Simulation Study

2.7.1 Two-sample Case when Sample Size is Small

We conducted a Monte Carlo simulation study to compare finite sample properties of four different estimators – Dykstra (1982), Lo (1987), Rojo (2004) and the C-NPMLE for the two-sample case where $S_1(t) \geq S_2(t)$ for all t. The root mean square error (RMSE) of the estimates of the survivor functions over a range of values of t are shown. Each simulation consists of 10,000 replications. The upper (lower) plot of each sub-figure shows the RMSE of estimates of $S_1(t)$ ($S_2(t)$). The distributions and sample sizes for the simulations are shown in Table 2.5.

As before, let $S_1^*(t)$ and $S_2^*(t)$ be the KM estimates. Lo's estimators are defined as $\hat{S}_1^L(t) = \max\{S_1^*(t), S_2^*(t)\}$ and $\hat{S}_2^L(t) = \min\{S_1^*(t), S_2^*(t)\}$. Rojo's estimators are defined as $\hat{S}_1^R(t) = \max[S_1^*(t), \{n_1 S_1^*(t) + n_2 S_2^*(t)\}/(n_1 + n_2)]$ and $\hat{S}_2^R(t) = \min[\{n_1 S_1^*(t) + n_2 S_2^*(t)\}/(n_1 + n_2), S_2^*(t))]$, where n_1 and n_2 are sample size of population 1 and 2. In order to minimize different effects from estimates beyond the last observed time in each population for different estimators, we set the estimates of survivor functions as low as

Table 2.5: Distributions and sample sizes used in simulation study.

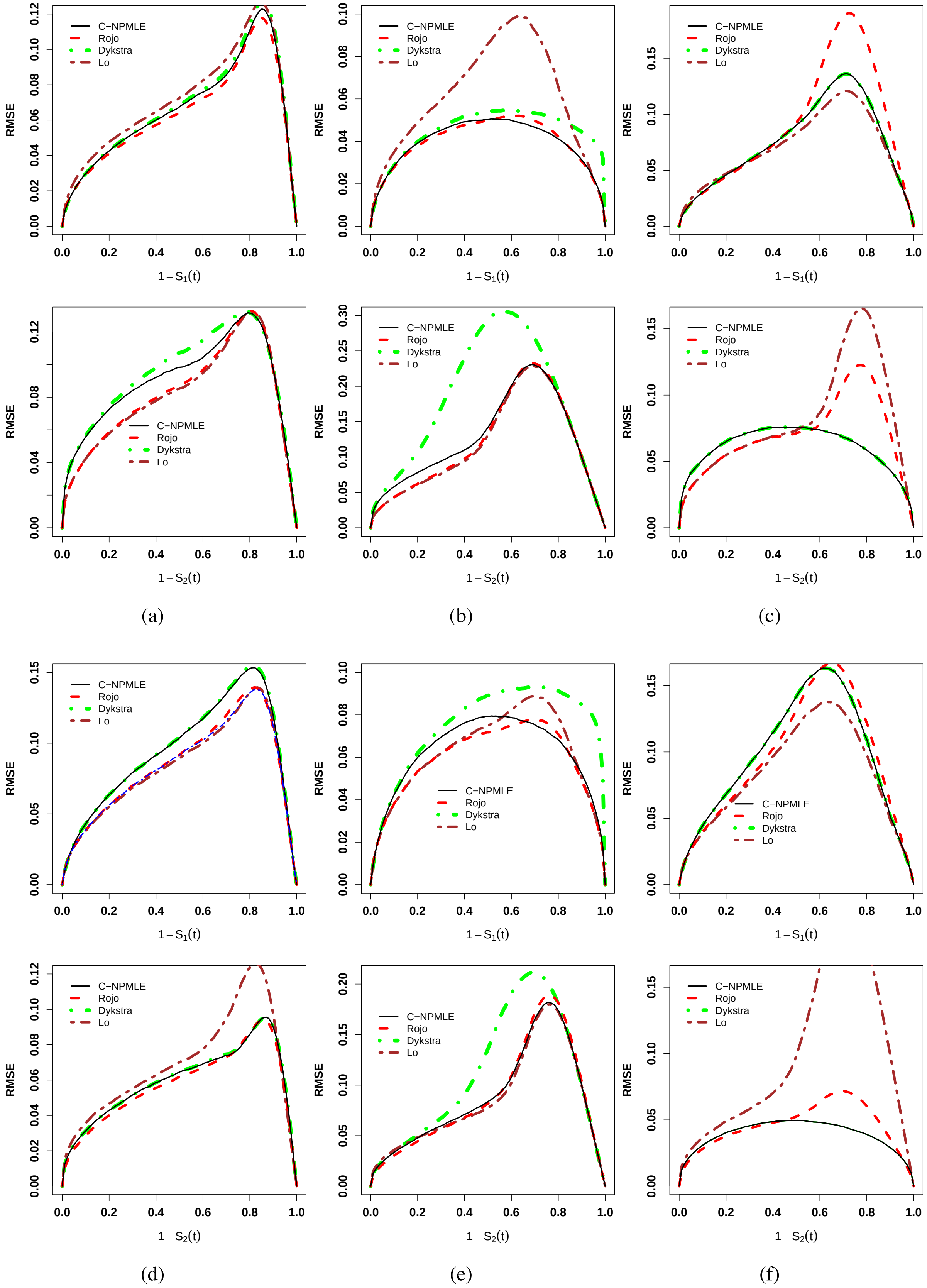
	Event Distributions		Censoring Distributions		Sample size	
	$\log S_1(t)$	$\log S_2(t)$	$\log S_1^c(t)$	$\log S_2^c(t)$	n_1	n_2
Figure 3.1a	$-t$	$-1.2t$	$-1.5t$	$-1.5t$	100	40
Figure 3.1b	$-t$	$-1.2t$	No censoring	$-3.0t$	100	40
Figure 3.1c	$-t$	$-1.2t$	$-3.0t$	No censoring	100	40
Figure 3.1d	$-t$	$-1.2t$	$-1.5t$	$-1.5t$	40	100
Figure 3.1e	$-t$	$-1.2t$	No censoring	$-3.0t$	40	100
Figure 3.1f	$-t$	$-1.2t$	$-3.0t$	No censoring	40	100

possible after the last observed time for each population.

Dykstra's estimator has similar efficiency compared to the C-NPMLE when population 2 is significantly less censored (Figure 3.1c and 3.1f), but in other cases, the C-NPMLE has smaller MSE compared to Dykstra's estimator. In cases with the same censoring distributions (Figure 3.1a and 3.1d), Rojo's estimator behaves better than other estimators. The intuitive reason that the C-NPMLE is not the best, despite maximizing the likelihood, is because the C-NPMLE is focussed on estimating the whole distribution, whereas the Rojo's and Lo's estimators are pointwise estimators, and the RMSE is a pointwise criterion. However, if population 1 and 2 have significantly different censoring distributions, the C-NPMLE is the preferred estimator. Specifically when population 1 is excessively censored (Figure 3.1c and 3.1f), Rojo's estimator has large RMSE compared to the C-NPMLE where the true survivor functions are small (< 0.4). Lo's estimator in general does not behave well when the two populations have different censoring distributions (Figure 3.1b, 3.1c and 3.1f).

2.7.2 Properties of Confidence Intervals

We conducted a simulation study to evaluate finite sample properties of the CIs proposed in section 3.5. The sample sizes and the underlying distributions are shown in Table



(a) (b) (c)

(d) (e) (f)

Figure 2.4: Comparison of RMSEs for different estimators.

2.6. Each simulation consists of 1,000 replications and bootstrap CIs are based on 199 bootstrap samples. We construct bootstrap CIs using percentiles and bias corrected percentiles. For the bias correction, we used $\arcsin(\sqrt{s})$ transformation at times 0.29 and 0.69, where the survival probabilities of group 1 are 0.75 and 0.50 respectively. We also evaluated the bias corrected intervals on the original scale and on the $\log(s)$ scale, and found that the binomial variance stabilizing $\arcsin(\sqrt{s})$ transformation gave CIs with slightly better properties than other two transformations and hence we only report the results under this transformation.

Table 2.6 shows the mean squared errors (MSEs) and biases of the KM estimator and the C-NPMLE at times 0.29 and 0.69. As expected, $\hat{S}_1(t)$ is positively biased and $\hat{S}_2(t)$ is negatively biased. The closer the two distributions or the smaller the sample sizes, the larger the bias and the MSE of the C-NPMLE. The MSE of the C-NPMLE is generally smaller than that of the KM estimator unless the two survival probabilities are very close to each other as in cases a, b and e.

The coverage rates and average widths of the CIs described in section 3.5 are shown in Table 2.7. Bootstrap methods using the percentiles (columns 5 and 7 in Table 2.7) do not work well with fairly low coverage rates when the two survivor probabilities are not far from each other (a, b, c, e and f). However, we found that the bias corrected bootstrap CIs (Bootstrap (BC_{asin}) and R-Bootstrap (BC_{asin}) in Table 2.7) produce narrower intervals with reasonable coverage rates, especially when using the restricted sampling scheme (R-Bootstrap (BC_{asin}) in Table 2.7). In comparison, the shifted constant width CI that is centered on the C-NPMLE (last column in Table 2.7) has the same width as the KM estimator on the $\log$ scale. The results suggest that the asymptotic variance gives CIs with reasonable coverage rate except that in some cases (e.g. case d at time 0.29 and case c at time 0.69), the CIs have too high a coverage rate. These intervals, however, tend to be

Table 2.6: MSE ($\times 10^4$) and Bias ($\times 10^2$). Event time distribution is $\exp(\lambda_g), g = 1, 2$ and censoring distribution is $U(0, 1.5)$ for both groups in all scenarios.

Event time	Sample	MSE		Bias		MSE		Bias	
distributions	sizes	KM	C-NPMLE	KM	C-NPMLE	KM	C-NPMLE	KM	C-NPMLE
		$t = 0.29$				$t = 0.69$			
a. $\lambda_1 = 1$	100	19.1	19.9	-0.01	2.45	36.1	39.3	0.21	3.41
$\lambda_2 = 1.01$	50	41.3	53.7	0.16	-5.16	68.3	76.5	0.14	-5.95
b. $\lambda_1 = 1$	100	19.4	19.5	0.18	2.22	33.4	34.4	0.16	2.71
$\lambda_2 = 1.1$	50	44.9	49.1	-0.17	-4.51	75.4	65.7	0.23	-4.60
c. $\lambda_1 = 1$	100	22.6	21.1	-0.09	1.59	35.7	33.7	-0.05	1.92
$\lambda_2 = 1.2$	50	48.4	46.0	-0.05	-3.55	72.5	61.5	-0.32	-3.95
d. $\lambda_1 = 1$	100	20.9	18.9	-0.08	1.20	33.7	30.8	-0.11	1.18
$\lambda_2 = 1.4$	50	49.2	40.3	0.26	-2.25	71.5	57.3	-0.02	-2.23
e. $\lambda_1 = 1$	400	5.52	6.3	0.05	1.37	8.92	10.3	0.12	1.79
$\lambda_2 = 1.01$	200	10.7	12.9	0.20	-2.54	18.3	20.9	0.10	-3.18
f. $\lambda_1 = 1$	400	5.31	5.24	0.10	1.01	8.81	8.59	0.13	1.22
$\lambda_2 = 1.1$	200	10.7	9.99	0.17	-1.69	17.4	14.9	0.24	-1.89
g. $\lambda_1 = 1$	400	5.33	4.98	0.01	.611	8.73	8.15	0.11	0.73
$\lambda_2 = 1.2$	200	11.9	10.2	0.00	-1.19	18.2	15.1	-0.06	-1.22
h. $\lambda_1 = 1$	400	5.06	4.84	-0.02	0.27	8.43	8.24	-0.03	0.20
$\lambda_2 = 1.4$	200	12.4	11.4	-0.03	-0.56	16.1	15.1	-0.01	-0.39

wider than those from the bias corrected bootstrap. Based on our simulation results, bias corrected bootstrap approach using the restricted scheme (R-Bootstrap ($\mathrm{BC}_{\mathrm{asin}}$) in Table 2.7) is preferable.

It should be noted that the estimate and the standard error is on appropriate data summary when the information is symmetric about the estimate. But here, especially when $S_1(t)$ is close to $S_2(t)$, the most appropriate CI are asymmetric. Thus simple repeating of the standard error for $\hat{S}_g(t)$ based on an asymptotic variance is not recommended in this situation.

Table 2.7: Coverage rates ($\times 10^2$) and average widths ($\times 10^2$) of confidence intervals.

Sampling scheme			Bootstrap			R-Bootstrap			
Estimator			KM	C-NPMLE				K-M	C-NPMLE
Confidence Interval			Percentile		BC$_{\text{asin}}$	Percentile	BC$_{\text{asin}}$	$S_g(t)\exp(\pm 1.96\sigma_g^*)$	
$t = 0.29$	a.	$S_1(t)$	95(17.5)	85(14.7)	94(15.2)	84(15.2)	95(16.0)	95(17.9)	93(18.5)
		$S_2(t)$	94(24.6)	73(20.5)	91(19.7)	68(22.0)	94(20.7)	93(25.2)	95(23.5)
	b.	$S_1(t)$	94(17.4)	85(14.9)	95(15.4)	84(15.2)	96(16.0)	95(17.9)	93(18.4)
		$S_2(t)$	93(25.2)	78(21.1)	92(20.4)	75(22.5)	94(21.3)	93(26.0)	97(24.5)
	c.	$S_1(t)$	93(17.4)	87(15.2)	91(15.7)	85(15.4)	93(16.2)	94(17.9)	93(18.3)
		$S_2(t)$	92(25.7)	85(21.6)	92(20.9)	84(22.9)	92(21.7)	93(26.5)	96(25.3)
	d.	$S_1(t)$	94(17.4)	90(15.5)	94(16.0)	90(15.8)	94(16.5)	95(17.9)	94(18.2)
		$S_2(t)$	94(26.6)	91(22.9)	94(22.3)	91(23.7)	93(22.8)	95(27.5)	98(26.5)
	e.	$S_1(t)$	93 (8.8)	82 (7.6)	91 (7.7)	80 (7.8)	92 (8.0)	94 (9.0)	91 (9.1)
		$S_2(t)$	94(12.3)	73(10.0)	90 (9.7)	68(10.7)	95(10.3)	95(12.7)	95(12.2)
	f.	$S_1(t)$	94 (8.8)	87 (7.8)	93 (7.9)	86 (7.9)	93 (8.1)	94 (9.0)	94 (9.1)
		$S_2(t)$	94(12.7)	84(10.6)	92(10.3)	82(11.0)	93(10.7)	96(13.0)	97(12.7)
	g.	$S_1(t)$	94 (8.7)	90 (7.9)	93 (8.1)	90 (8.1)	94 (8.2)	95 (9.0)	95 (9.1)
		$S_2(t)$	94(13.0)	89(11.1)	91(10.9)	89(11.5)	90(11.2)	94(13.3)	97(13.1)
	h.	$S_1(t)$	95 (8.8)	93 (8.3)	94 (8.4)	93 (8.3)	94 (8.4)	96 (9.0)	96(9.0)
		$S_2(t)$	94(13.5)	93(12.3)	89(12.1)	93(12.4)	90(12.3)	96(13.8)	97(13.7)
$t = 0.69$	a.	$S_1(t)$	94(22.4)	83(19.8)	92(19.8)	82(20.5)	94(20.6)	94(23.1)	92(24.7)
		$S_2(t)$	95(31.7)	74(23.8)	91(24.2)	70(25.8)	95(26.4)	96(32.9)	96(29.1)
	b.	$S_1(t)$	94(22.4)	88(20.0)	93(20.1)	86(20.6)	94(20.7)	94(23.2)	93(24.4)
		$S_2(t)$	93(31.7)	81(24.3)	92(24.8)	78(26.0)	94(26.7)	94(32.9)	97(29.7)
	c.	$S_1(t)$	94(22.5)	90(20.5)	92(20.5)	89(20.7)	94(20.7)	94(23.1)	94(24.1)
		$S_2(t)$	93(31.5)	86(25.0)	93(25.6)	84(26.0)	94(27.0)	95(32.8)	98(30.3)
	d.	$S_1(t)$	94(22.4)	93(20.9)	94(20.9)	93(21.1)	94(21.1)	95(23.1)	96(23.8)
		$S_2(t)$	92(31.0)	89(25.7)	91(26.7)	90(26.4)	92(27.6)	94(32.5)	97(30.8)
	e.	$S_1(t)$	94(11.3)	83(10.0)	91(10.0)	81(10.3)	92(10.3)	95(11.6)	91(11.9)
		$S_2(t)$	93(15.9)	72(12.2)	90(12.2)	66(13.0)	93(13.1)	94(16.4)	93(15.3)
	f.	$S_1(t)$	94(11.3)	89(10.3)	93(10.3)	88(10.5)	93(10.5)	95(11.6)	95(11.8)
		$S_2(t)$	95(15.9)	85(12.7)	92(12.8)	83(13.3)	93(13.4)	95(16.4)	97(15.7)
	g.	$S_1(t)$	94(11.3)	91(10.5)	94(10.5)	91(10.6)	94(10.6)	95(11.6)	95(11.7)
		$S_2(t)$	94(15.8)	91(13.4)	91(13.5)	90(13.7)	91(13.8)	94(16.3)	97(15.9)
	h.	$S_1(t)$	94(11.2)	94(10.8)	93(10.8)	94(10.9)	92(10.9)	96(11.6)	96(11.6)
		$S_2(t)$	95(15.6)	93(14.3)	91(14.4)	94(14.4)	90(14.6)	96(16.0)	97(15.9)

2.8 Discussion

Dykstra's estimator is the same as the C-NPMLE when the stochastically smaller population has no censoring. When censoring exists in that population, our experience suggests that Dykstra's estimators and the estimators developed in this paper are usually different for small sample settings. In large sample cases, they are frequently identical at early times but then differ in the tail. A fundamental difference between Dykstra's estimator and the C-NPMLE developed in this paper is that the C-NPMLE can have probability mass at non-event times throughout the follow-up period whereas, for Dykstra's estimator, this can happen only prior to the first failure in the stochastically smaller population.

The results from the simulation suggests that the C-NPMLE has a smaller or equal RMSE when compared to Dykstra's estimator. Rojo's estimator has good properties under the pointwise criteria because it is based on pointwise estimates. It does not, however, adapt well to unequal censoring. There is the potential to develop an improved approach that might adapt the NPMLE to the pointwise case. The simulation results also suggest the bias correct Bootstrap confidence interval using the restricted sampling scheme is preferable in practice.

Præstgaard and Huang (1996) established the asymptotic distribution of Dykstra's estimator. If, as seems likely, the Dykstra's estimator and the C-NPMLE are asymptotically equivalent, the asymptotic distribution should also apply to the C-NPMLE. However, in our opinion, the asymptotic distribution is not very useful for finite sample inference. If $S_1(t)$ is strictly greater than $S_2(t)$, then the asymptotic distributions of the C-NPMLE of $\hat{S}_1(t)$ and $\hat{S}_2(t)$ will be identical to that of KM estimators, i.e. the constraint becomes irrelevant asymptotically while the constraint is still relevant everywhere in the finite sample case. If $S_1(t) = S_2(t)$, then the asymptotic distribution is mathematically interesting, but probably not appropriate to use in this context where all are willing to assume that $S_1(t) \geq S_2(t)$.

We are currently investigating extensions of these estimators and algorithms to settings with more than two samples with linear ordering or partial ordering constraints. Hoff (2000, 2003b) and Lim, Kim, and Wang (2009) have proposed different methods to obtain the C-NPMLE under partial ordering constraints. However, we believe that the extension of the methods in this paper will provide additional insights and have the potential to improve the methods through reducing the number of parameters to compute and the number of constraints to apply.

CHAPTER III

Pointwise Constrained Maximum Likelihood Estimator of the Survivor Function under Stochastic Ordering

3.1 Introduction

We often encounter situations where there is prior knowledge of stochastic ordering among distributions. For example, in a cancer study, we expect patients with a lower stage of cancer at diagnosis to have lower death rates at all times than those with a higher stage. Stochastic ordering is an important concept and has a wide range of applications, in such fields as biomedical research, economics and finance, and system reliability in engineering science. In addition to the natural desire for estimators of the distributions to satisfy the same ordering restrictions as the underlying distributions, there is the potential for improved efficiency by applying the constraints in the estimation method as compared to ignoring them.

For random variables T_1 and T_2 with corresponding survivor functions $S_1(t)$ and $S_2(t)$, T_1 is stochastically larger than T_2 ($T_1 \geq_{st} T_2$) if $S_1(t) \geq S_2(t)$ for all t (Lehmann, 1955). For G groups, the concept can be generalized to partial ordering; specifically, we say that $T_g, g = 1, \ldots, G$ satisfy the partial ordering constraints defined by the constraint set $E \subset \{1, \ldots, G\}^2$ if for any $(i, j) \in E$, $T_i \geq_{st} T_j$. Special cases of this are *simple ordering*, in which $T_1 \geq_{st} \cdots \geq_{st} T_G$, for which $E = \{(1,2),(2,3),\ldots,(G-1,G)\}$; *tree ordering*, in which $T_1 \geq_{st} T_2, T_1 \geq_{st} T_3, \ldots, T_1 \geq_{st} T_G$ for which $E = \{(1,2),(1,3),\ldots,(1,G)\}$; *umbrella ordering*, in which $T_1 \geq_{st} \cdots \geq_{st} T_i \leq_{st} T_{i+1} \leq_{st} \cdots \leq_{st} T_G$ for which $E = \{(1,2),(2,3),\ldots,(i-1,i),(i+1,i),(i+2,i+1),\ldots,(G,G-1)\}$, and *factorial ordering* such as $T_1 \geq_{st} T_2 \geq_{st} T_4, T_1 \geq_{st} T_3 \geq_{st} T_4$, for which $E = \{(1,2),(2,4),(1,3),(3,4)\}$. The constraint set E is said to be *non-redundant* if, for any $(i,k) \in E, (k,j) \in E$, and $(i,j) \in E$ implies $i = k$ or $k = j$. Any set E has an equivalent non-redundant set, and we henceforth assume that E is non-redundant.

We consider independent right censored samples of the form (Y_{gi}, Δ_{gi}), $g = 1,\ldots,G$, $i = 1,\ldots,n_g$, where Y_{gi} is the observed time and Δ_{gi} is the event indicator ($\Delta_{gi} = 1$ if the event occurred or $\Delta_{gi} = 0$ if right censored). We assume that the censoring mechanism is independent, so that the generalized likelihood as a function of survivor functions is

$$L\{S_1(\cdot),\ldots,S_G(\cdot)\} = \prod_{g=1}^{G}\prod_{i=1}^{n_g}\Big\{S_g(Y_{gi}-) - S_g(Y_{gi})\Big\}^{\Delta_{gi}} S_g(Y_{gi})^{1-\Delta_{gi}}, \tag{3.1}$$

under the condition that there are no common jumps between the life time and censoring distributions. The E-constrained nonparametric maximum likelihood estimator (C-NPMLE) is the estimator $\hat{S}_1, ..., \hat{S}_G$ that maximizes (3.1) subject to the partial ordering constraint E. Brunk et al. (1966) studied the constrained nonparametric maximum likelihood estimator in the two-sample case without censoring. Dykstra (1982) and we in Chapter II extended this result to right censored data. In the case of three or more populations with a general partial ordering, Hoff (2003b) and Lim et al. (2009) proposed different computational methods for obtaining the C-NPMLE. The C-NPMLE has the properties that a violation of a constraint in the Kaplan-Meier estimators (Kaplan and Meier, 1958) at an earlier time affects the estimator at a later time, even if there is no violation at this later time. This could be considered as undesirable. A number of authors have noted that the

C-NPMLE of $S_1, \ldots, S_G$ can have relatively large pointwise bias and mean squared error at a fixed t and have suggested alternatives (Rojo and Ma, 1996; Rojo, 2004; Gangnon and King, 2002; El Barmi and Mukerjee, 2005) that can have better mean squared error properties. In Chapter II, we noted a correction to the C-NPMLE presented by Dykstra (1982), which did lead to improved properties, however, this corrected C-NPMLE still often has poorer pointwise properties than some other estimators that have been proposed. Some of these alternative estimators are relatively simple to define. In the two sample problem, Lo (1987) suggested swapping the Kaplan-Meier estimates of the survivor functions when the constraint is violated. Gangnon and King (2002) proposed constrained estimators of the distribution functions for a simple ordering case using a minimum distance estimation method and also extended Lo's estimator to the simple ordering case. For the two sample problem, Rojo (2004) proposed estimating both survivor functions as the weighted average of the two Kaplan-Meier estimators at times when the constraint is violated, where the weights are based on the initial sample sizes. El Barmi and Mukerjee (2005) extended Rojo's estimators to the simple ordering situation using isotonic regression method. The simulation study in Chapter II shows that some of these estimators have smaller mean squared error than the C-NPMLE when the censoring distributions are equal, however, when the censoring distributions are substantially different between the groups, they may have larger mean squared error than the C-NPMLE. Moreover, these alternative estimators have not been explicitly extended to a general partial ordering case.

When we consider finite sample properties of an estimator $\hat{S}(t)$, we usually use pointwise criteria, such as pointwise bias or pointwise mean squared error at each fixed t. In contrast to pointwise estimators such as described in Rojo (2004) and Lo (1987), the C-NPMLE is focused on estimating the whole survival curve. So it is perhaps not surprising that Rojo's estimators typically have better properties when evaluated using metrics such as pointwise mean squared error. On the other hand, these estimators do not adapt well to unequal censoring distributions between groups whereas the C-NPMLE does. This motivated us to propose a new constrained estimator, a pointwise constrained nonparametric maximum likelihood estimator.

Definition 3.1 (Pointwise C-NPMLE)**.** For each specified time x, let $\tilde{S}_g(t;x)$ be the maximum likelihood estimator of $S_g(t)$ under the constraint $S_i(x) \geq S_j(x)$ for all $(i,j) \in E$. Then $\hat{S}_g(t) = \tilde{S}_g(t;t)$, $g = 1, \ldots, G$ for all t is the pointwise constrained nonparametric maximum likelihood estimator of the survivor function S_g under the partial stochastic ordering constraint E.

The rest of the chapter is organized as follows. In section 3.2 we develop the estimator. We consider consistency and asymptotic distributions in sections 3.3, and in section 3.4, we compare asymptotic properties of the pointwise C-NPMLE with the Kaplan-Meier estimator. In section 3.5, we consider methods to construct confidence intervals. We conduct a simulation study to compare small and large sample properties of different estimators in section 3.6, and apply the method to a real data example in section 3.7.

3.2 Estimation Methods

3.2.1 Notation and Likelihood

To obtain the pointwise C-NPMLE as given in Definition 3.1, it is required to maximize the likelihood (3.1) subject to the constraints $S_i(x) \geq S_j(x)$ for all $(i,j) \in E$ for a fixed time x. This will give the estimates of $\tilde{S}_1(t;x), \ldots, \tilde{S}_G(t;x)$ and the constrained maximization will be repeated for all times x of interest.

Let $X_{gj}, j = 1, 2, \cdots, m_g$ be the distinct event times in group g and define $X_{g0} = 0$

and $X_{g(m_g+1)} = \infty, g = 1, \ldots, G$. Let $N_g(t)$ be the number at risk at time t in group g and $M_g(t)$ be the number of distinct events in $(0, t]$ in group g. Let d_{gj} and n_{gj} be respectively the number of events and the number at risk in group g at time X_{gj}, where $g = 1, \ldots, G, j = 1, \ldots, m_g$.

It is convenient to redefine the problem in terms of hazards. Let $h_g(t) = \log\{S_g(t)/S_g(t-)\}$, so that $1 - \exp h_g(t)$ is the discrete hazard in group g at time t. The log likelihood of (3.1) can then be rewritten as

$$\begin{aligned}&\log L(h_1, \cdots, h_G)\\ &\quad= \sum_{g=1}^{G}\Big\{\sum_{i=1}^{m_g}\big[d_{gi}\log\{1 - \exp h_g(X_{gi})\} + (n_{gi} - d_{gi})\, h_g(X_{gi})\big] + N_g(x)h_g^{\delta}(x)\Big\},\end{aligned} \tag{3.2}$$

where $h_g = \{h_g(X_p), \ldots, h_g(X_{gm_g}), h_g^{\delta}(x)\}, g = 1, \ldots, G$. The corresponding constraints are $\sum_{j=1}^{M_p(x)} h_p(X_{pj}) + h_p^{\delta}(x) \le \sum_{j=1}^{M_r(x)} h_r(X_{rj}) + h_r^{\delta}(x)$, for all $(p, r) \in E$, and $h_g^{\delta}(x) \le 0, g = 1, \ldots, G$. In this, $h_g^{\delta}(x) = I(x \ne X_{gM_g(x)})h_g(x)$; this accounts for the fact that if $x = X_{gM_g(x)}$, we do not have the extra term $N_g(x)h_g(x)$ in the log likelihood (3.2). Note that $h_g(X_{gj})$ must be negative because of the term $\log\{1 - \exp h_g(X_{gj})\}$ in the log likelihood (3.2). Thus, the constraint $h_g(X_{gj}) \le 0$ is not necessary.

To maximize the log likelihood subject to the constraints, we introduce Lagrange multipliers $\lambda_{ij} \ge 0$ and $\nu_g \ge 0$, giving the corresponding Lagrangian function as

$$\begin{aligned} lagrL(h, \lambda, \nu) =& \sum_{g=1}^{G}\sum_{i=1}^{m_g}\big[d_{gi}\log\{1 - \exp h_g(X_{gi})\} + (n_{gi} - d_{gi})\, h_g(X_{gi})\big]\\ &+ \sum_{g=1}^{G}\{N_g(x)h_g^{\delta}(x) - \nu_g h_g^{\delta}(x)\}\\ &- \sum_{(p,r)\in E}\lambda_{pr}\Big\{\sum_{j=1}^{M_p(x)} h_p(X_{pj}) + h_p(x) - \sum_{j=1}^{M_r(x)} h_r(X_{rj}) - h_r(x)\Big\}.\end{aligned}$$

3.2.2 Linearly Constrained Convex Minimization

There is a large literature on general approaches to linearly constrained convex minimization problems. Essentially, there are three types of algorithms. One is the interior point method that targets complementary slackness while maintaining primal and dual feasibility. Another is the primal active set method that targets dual feasibility while maintaining primal feasibility and complementary slackness. A third is the dual active set method that targets primal feasibility while maintaining dual feasibility and complementary slackness.

In general our data contain a much larger number of observed event times than the number of groups. Interior point methods such as geometric programming (Lim et al., 2009) and primal active set methods such as some gradient based methods (Snyman, 2005) simultaneously optimize over the large number of quantities $h_g^{\delta}(x)$ and $h_g(X_{gi}), g = 1, \ldots, G, i = 1, \ldots, m_g$ at each time x of interest, so these two types of methods are not computationally efficient in our setting. Dual active set methods that maximize the dual problem $L_D(\lambda, \nu) = \sup_h \{-lagrL(h, \lambda, \nu)\}$ with constraints $\lambda_{ij} \geq 0$ and $\nu_g \geq 0$ involve many fewer parameters, however, the dual function $L_D(\lambda, \nu)$ is difficult to express as a function of λ, ν and the feasible range of λ, ν is difficult to specify in our problem. So dual active set methods are also difficult to implement in our context.

In section 3.2.5, we transform the problem of maximizing the log likelihood (3.2) subject to the linear constraints to another simple concave maximization problem subject to linear constraints by using the profile likelihood. In preparation for this, we first discuss the constrained maximum likelihood estimator of the survivor function in the one sample case.

3.2.3 Maximum Likelihood Estimator of the Survivor Function Subject to a Single Constraint

We consider a one-sample case and for ease of presentation, we omit the group indicator g. Without constraint, the maximum likelihood estimator has probability mass only at the observed event times up to and including the last observed time. The log likelihood analogous to (3.2) is

$$\log L(h) = \sum_{j=1}^{m} \left[d_j \log\left\{1 - \exp h(X_j)\right\} + (n_j - d_j)\, h(X_j) \right], \tag{3.3}$$

where $h = \{h(X_1), \ldots, h(X_m)\}$ and (3.3) is maximized at $h(X_j) = \log(1 - d_j/n_j), j = 1, \ldots, m$, which corresponds to the Kaplan-Meier estimator.

Consider now the maximum likelihood estimator subject to the constraint $S(x) = \exp(q)$. In this case, the maximum likelihood estimator of the survivor function can be obtained in the class of survivor functions with positive probability mass at event times X_i and nonnegative probability mass at time x. The optimization problem is to maximize the log likelihood of $h = \{h(X_1), \ldots, h(X_m), h^{\delta}(x)\}$,

$$\log L(h) = \sum_{i=1}^{m} \left[d_i \log\left\{1 - \exp h(X_i)\right\} + (n_i - d_i)\, h(X_i) \right] + N(x) h^{\delta}(x),$$

subject to $\sum_{j=1}^{M(x)} h(X_j) + h^{\delta}(x) = q$ and $h^{\delta}(x) \le 0$.

Let $K(q;x) = -N(x)$ if $M(x) = 0$, otherwise let $K(q;x) = \max(-N(x), \hat{k})$, where $\hat{k}$ is the unique solution of the equation $\sum_{j=1}^{M(x)} \log\{1 - d_j/(n_j + k)\} = q$. Note that $\hat{k} = \infty$ if $q = 0$ and $\hat{k} = d_{M(x)} - n_{M(x)}$ if $q = -\infty$. Let

$$\hat{h}(q;X_i) = \begin{cases} \log\left\{1 - \dfrac{d_i}{n_i + K(q;x)}\right\} & i \le M(x) \\ \log(1 - \dfrac{d_i}{n_i}) & i > M(x) \end{cases}, \tag{3.4}$$

$$\hat{h}^{\delta}(q;x) = q - \sum_{j=1}^{M(x)} \hat{h}(q;X_j).$$

Theorem 3.2. *The maximum likelihood estimator of* $S(t)$ *subject to constraint* $S(x) = \exp(q)$ *at a given* x *is* $\hat{S}(t) = \exp\{\sum_{X_j \le t} \hat{h}(q; X_j) + I(t \ge x) h^{\delta}(q; x)\}, t \le \tau$, *where* τ *is the last observed time.*

Proof. See Appendix D. □

Thomas and Grunkemeier (1975) and Li (1995) considered the same maximization problem as described above. However, Thomas and Grunkemeier (1975) solved the problem with the equality constraint $\sum_{j=1}^{M(x)} h(X_j) = q$, which implicitly assumes that $\hat{h}(x) = 0$ if x is not an observed event time, whereas Li (1995) mistakenly "proved" that $\hat{h}(x) = 0$ unless x is an observed event time. In fact, the maximization problem described above involves two constraints: $\sum_{j=1}^{M(x)} h(X_j) + h^{\delta}(x) = q$ and $h^{\delta}(x) \le 0$. It is possible that $\hat{h}^{\delta}(x) < 0$ if $K(q; x) = -N(x)$. The inequality constraint of a parameter such as $h^{\delta}(x) \le 0$ has been neglected in these approaches. It is necessary, however, to apply the Karush-Kuhn-Tucker conditions (Karush 1939; Kuhn and Tucker 1951) to all possible inequality constraints, including the bounds of parameters and then cautiously remove unnecessary constraints that are guaranteed by other constraints.

3.2.4 The Profile Likelihood of $S(x)$

From Theorem 3.2 and (3.4), the profile log likelihood of $S(x) = \exp(q)$ at a given x is

$$\begin{aligned} pl_h(q; x) =& \sup_{h \in R} \log L(h) \\ =& \sum_{i=1}^{m} \left[d_i \log \left\{ 1 - \exp \hat{h}(q; X_i) \right\} + (n_i - d_i) \hat{h}(q; X_i) \right] + N(x) \hat{h}^{\delta}(q; x), \end{aligned} \tag{3.5}$$

where $R = \{h : \sum_{i=1}^{M(x)} h(X_i) + h^{\delta}(x) = q\}$, and $\hat{h}(q; X_i)$ and $\hat{h}^{\delta}(q; x)$ are defined in equation (3.4).

Lemma 3.3. *The derivative of the profile log likelihood* (3.5) *with respect to* q *is* $-K(q; x)$.

Proof. We consider separately the cases i) $K(q;x) > -N(x)$ and ii) $K(q;x) = -N(x)$.

i) If $K(q;x) > -N(x)$, then $\hat{h}^\delta(x) = 0$ and $\hat{h}(q;X_i) = \log(1 - d_i/n_i)$ for $i > M(x)$, which does not depend on q. For any $i \leq M(x)$,

$$\begin{aligned}
&\frac{d}{d\hat{h}(q;X_i)} \left[d_i \log \left\{ 1 - \exp \hat{h}(q;X_i) \right\} + (n_i - d_i)\, \hat{h}(q;X_i) \right] \\
&\qquad = \frac{-d_i \exp\{\hat{h}(q;X_i)\}}{1 - \exp\{\hat{h}(q;X_i)\}} + n_i - d_i = n_i - \frac{d_i}{1 - \exp \hat{h}(q;X_i)} \\
&\qquad = n_i - \{n_i + K(q;x)\} = -K(q;x).
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{d}{dq} pl_h(q;x) &= \frac{d}{dq} \left\{ \sum_{i=1}^{m} \left[d_i \log \left\{ 1 - \exp \hat{h}(q;X_i) \right\} + (n_i - d_i)\, \hat{h}(q;X_i) \right] + N(x) \hat{h}^\delta(x) \right\} \\
&= \sum_{i=1}^{M(x)} -K(q;x) \frac{d\hat{h}(q;X_i)}{dq} = -K(q;x) \frac{d}{dq} \sum_{i=1}^{M(x)} \hat{h}(q;X_i) = -K(q;x).
\end{aligned}$$

ii) If $K(q;x) = -N(x)$, then $\hat{h}(q;X_i) = \log\{1 - d_i/[n_i - N(x)]\}$ for $i \leq M(x)$ and $\hat{h}(q;X_i) = \log(1 - d_i/n_i)$ for $i > M(x)$ are not functions of q. So

$$\begin{aligned}
&\frac{d}{dq} pl_h(q;x) \\
&\quad = \frac{d}{dq} \left\{ \sum_{i=1}^{m} \left[d_i \log \left\{ 1 - \exp \hat{h}(q;X_i) \right\} + (n_i - d_i)\, \hat{h}(q;X_i) \right] + N(x) \hat{h}^\delta(q;x) \right\} \\
&\quad = N(x) = -K(q;x). \qquad \square
\end{aligned}$$

3.2.5 Reformulation of the Problem Using Profile Likelihood

For given x, maximizing the log likelihood (3.2) subject to the constraints in E can be redefined as maximizing the log profile likelihood

$$
\begin{aligned}
pl_h(q_1,\ldots,q_G;x) &= \sum_{g=1}^{G} pl_{h_g}(q_g;x)\\
&= \sum_{g=1}^{G}\left\{\sum_{i=1}^{M_g(x)}\left[(n_{gi}-d_{gi})\log\{n_{gi}+K_g(q_g;x)-d_{gi}\}-n_{gi}\log\{n_{gi}+K_g(q_g;x)\}\right]\right.\\
&\quad\left.+I\Big\{K_g(q_g;x)=N_g(x)\Big\}N_g(x)\left[q_g-\sum_{j=1}^{M_g(x)}\log\left\{1-\frac{d_{gi}}{n_{gi}+K_g(q_g;x)}\right\}\right]\right\}+C,
\end{aligned}
\tag{3.6}
$$

subject to constraints $q_i \geq q_j$, for all $(i,j) \in E$ and $q_g \leq 0, g = 1,\cdots,G$. Note that by using the profile likelihood, only G parameters $q = (q_1,\ldots,q_G)$ need to be estimated, and that $\hat{S}_g(x) = \exp(\hat{q}_g)$, where $\hat{q} = (\hat{q}_1,\ldots,\hat{q}_G)$ is the maximum likelihood estimator of q.

Any of the general methods described in section 3.2.2 can be used to maximize the log profile likelihood (3.6) under the corresponding linear constraints. Note that the log profile likelihood (3.6) is easy to calculate and the derivative is

$$\frac{d}{dq^T} ph_h(q;x) = \{-K_1(q_1;x),\ldots,-K_G(q_G;x)\}^T,$$

which is needed in some constrained optimization algorithms.

To obtain the pointwise C-NPMLE $\hat{S}_g(t), g = 1,\ldots,G$ for all t, it is not necessary to maximize the profile likelihood at every t. It can be seen that the pointwise C-NPMLE may only jump at observed event times and at times just after observed censoring times. Let $\{X'_j\}$ be the union of all distinct times Y_{gi} if $\Delta_{gi} > 0$ and Y^+_{gi} if $\Delta_{gi} = 0$. Note that Y^+_{gi} can be taken as $Y_{gi} + \epsilon$ for a small $\epsilon > 0$. We calculate $\hat{S}_g(X'_j)$, and then $\hat{S}_g(t)$ is a step function with jumps at only X'_j, i.e. $\hat{S}_g(t) = \hat{S}_g(\max\{X'_j : X'_j \leq t\})$ for all $t \leq \tau_g, g = 1,\cdots,G$.

The following Theorem shows that the pointwise C-NPMLE is a valid survivor function for every group until the last observed time of that group.

Theorem 3.4. *The pointwise C-NPMLE $\hat{S}_g(t), g = 1,\cdots,G$ obtained from maximizing the profile likelihood (3.6) is a non-increasing function in t. That is, for any $0 \leq x < y \leq \tau_g$,*

$\hat{S}_g(x) \geq \hat{S}_g(y)$.

Proof. See Appendix E. $\square$

3.2.6 Generalized Pool-adjacent-violators Algorithm in the Simple Ordering Case

Suppose that G survivor functions satisfy the simple stochastic ordering constraint $T_1 \geq_{st} \cdots \geq_{st} T_G$, and we aim to estimate the pointwise C-NPMLE at time x. A generalized pool-adjacent-violators algorithm can be used, where the set of violating groups to be "pooled" is called B. The common value $\hat{q}$ for all groups in B is the one that maximizes the log profile likelihood

$$pl_B(q;x) = \sum_{i\in B} pl_{h_i}(q;x) = \sum_{i\in B} \log L_i(\hat{h}_i).$$

Since from Lemma 3.3, $\frac{d}{dq}\sum_{i\in B} pl_{h_i}(q;x) = -\sum_{i\in B} K_i(q;x)$, the maximizer $\hat{q}$ is the root of the equation

$$\sum_{i\in B} K_i(q;x) = 0, \tag{3.7}$$

when the maximizer $\hat{q} < 0$. Here $B \subset \{1, \cdots, G\}$ is a set of indices of survivor functions that we are interested in "pooling". In the simple ordering setting, B will be a set of adjacent integers, for example $\{2\}$ or $\{2, 3, 4\}$.

Best, Chakravarti, and Ubhaya (1999) studied minimization of the sum of convex functions subject to linear ordering constraint using the pool-adjacent-violators algorithm method. Our log profile likelihood is the sum of concave functions, so maximizing the log profile likelihood is the same as minimizing separable convex functions as considered by Best et al. (1999). We adapt their algorithm, details of which can be found in Appendix C.

3.3 Consistency and Asymptotic Distribution

Let $S_g^*(t)$ be the Kaplan-Meier estimator of $S_g(t)$ and let $S_g^c(t)$ be the censoring survivor function for group g. Further, let $\tau_g = \inf\{t : S_g(t)S_g^c(t) = 0\}, g = 1, \ldots, G$. Under the condition that there are no common jumps between the life time and censoring distributions, Stute and Wang (1993) showed that the Kaplan-Meier estimator $S_g^*(t)$ is uniformly consistent for $S_g(t)$ on $[0, \tau_g)$. A similar result holds for the pointwise C-NPMLE. The following theorem is proved in Appendix F.

Theorem 3.5. *Let $\hat{S}_k(t)$ be the pointwise C-NPMLE (see Definition 3.1). Under the condition of no common jumps of $S_k(t)$ and $S_k^c(t)$, as $n_k \to \infty, k = 1, \ldots, G$,*

$$\sup_{t<\tau_g} |\hat{S}_g(t) - S_g(t)| \to 0 \qquad \textit{with probability } 1.$$

Let $W_g(V)$ be a Brownian motion on $[0, \infty)$ with variance function $V_g(t)$. As shown in Gill (1983),

$${n_g}^{1/2}\left(\frac{S_g^* - S_g}{S_g}\right) \to W(V_g) \text{ in distribution on } [0, \tau_g] \text{ as } n_g \to \infty,$$

where

$$V_g(t) = -\int_0^t \frac{dS_g(x)}{S_g^2(x-)S_g^c(x-)}.$$

For a fixed time x, ${n_g}^{1/2}\{S_g^*(x) - S_g(x)\}\to N\{0, \sigma_g^2(x)\}$ in distribution where $\sigma_g^2(x) = V_g(x)S_g^2(x)$.

Let $n = \sum_{g=1}^G n_g$ and assume that $\lim_{n\to\infty} n_g/n = c_g > 0$ and let $Z_g^*(x) = n^{1/2}\{S_g^*(x) - S_g(x)\}, g = 1, \ldots, G$. Then $\{Z_1^*(x), \ldots, Z_G^*(x)\}^T \to \{Z_1(x), \ldots, Z_G(x)\}^T$ in distribution, where $Z_g(x) \sim N(0, \sigma_g^2(x)/c_g)$ and $Z_g(x)'s$ are independent.

Theorem 3.6. *For a fixed time $x < \min\{\tau_k : L_g(x) \leq k \leq U_g(x)\}$ and under the simple* *ordering constraint* $T_1 \geq_{st} \cdots \geq_{st} T_G$*,*

$$n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\} \;\to\; c_g^{1/2} \min_{L_g(x)\leq \ell \leq g} \max_{g \leq u \leq U_g(x)} \frac{\sum_{k=\ell}^{u}\{Z_k(x)w_k(x)\}}{\sum_{k=\ell}^{u} w_k(x)} \text{ in distribution}, \tag{3.8}$$

where $w_g(x) = c_g/\sigma_g^2(x)$*,* $L_g(x) = \min\{i : S_i(x) = S_g(x)\}$ *and* $U_g(x) = \max\{i : S_i(x) = S_g(x)\}$.

Proof. See Appendix G. □

In Appendix G, we discuss extensions and special cases of Theorem 3.6 to situations where the number at risk in one group is zero.

Let $\check{S}_g(x)$ be the estimate of $S_g(x)$ by applying the isotonic regression algorithm to $S_g^*(x)$ with weights $w_g(x), g = 1, \ldots, G$, subject to constraint $S_1(x) \geq_{st} \cdots, \geq_{st} S_G(x)$. Then, $\check{S}_g(x)$ has a min-max form (Barlow et al., 1972)

$$\check{S}_g(x) = \min_{1\leq \ell \leq g} \max_{g \leq u \leq G} \frac{\sum_{k=\ell}^{u}\{S_k^*(x)w_k(x)\}}{\sum_{k=\ell}^{u} w_k(x)}.$$

From Theorem 2 in El Barmi and Mukerjee (2005), it can be seen that

$$n_g^{1/2}\{\check{S}_g(x) - S_g(x)\} \;\overset{D}{\to}\; c_g^{1/2} \min_{L_g(x)\leq \ell \leq g} \max_{g \leq u \leq U_g(x)} \frac{\sum_{k=\ell}^{u}\{Z_k(x)w_k(x)\}}{\sum_{k=\ell}^{u} w_k(x)}.$$

From (3.8), it follows that $\hat{S}_g(x)$ and $\check{S}_g(x)$ are asymptotically equivalent. We hypothesize that this equivalence to the isotonic regression will also hold under the partial ordering constraint. This yields the following conjecture for the asymptotic distribution of the pointwise C-NPMLE.

Conjecture 3.7. *For a fixed time* x*,*

$$n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\} \;\overset{D}{\to}\; c_g^{1/2}\; f_g(Z_1(x), \ldots, Z_G(x); \frac{c_1}{\sigma_1^2}, \ldots, \frac{c_G}{\sigma_G^2}, x),$$

as $n \to \infty$ *for any* x *given* $S_g(x)S_g^c(x) > 0$*. Here* $\Psi_g(x) = \{i : S_i(x) = S_g(x)\}$*,* $E_g(x) = \{(i,j) : i, j \in \Psi_g(x) \text{ and } (i,j) \in E\}$ *and* $f_g(z_1, \ldots, z_G; w_1, \ldots, w_G, x)$ *is the* *solution function for* μ_g *that minimizes* $\sum_{i=1}^{G} w_i(z_i - \mu_i)^2$ *under constraints* $\mu_i \geq \mu_j$ *for all* $(i,j) \in E_g(x)$.

If this conjecture is correct, inference methods developed for the isotonic regression could also be useful for the pointwise C-NPMLE.

3.4 Comparison with Kaplan-Meier Estimator when Sample Size is Large

3.4.1 Simple Ordering Case

In the simple ordering case with no censoring, El Barmi and Mukerjee (2005) showed that their isotonic regression estimator has smaller asymptotic mean squared error than the unrestricted Kaplan-Meier estimator. A similar result holds for the pointwise C-NPMLE compared to the Kaplan-Meier estimator when there is right censoring.

Theorem 3.8. *Consider the simple ordering constraint* $T_1 \geq_{st} T_2 \cdots \geq_{st} T_G$. *For a fixed* x *with* $S_k^c(x)S_k(x) > 0$ *for all* $k = 1, \ldots, G$, *let* $n_k^{1/2}\{\hat{S}_k(x) - S_k(x)\} \xrightarrow{D} \hat{Z}_k$ *and* $n_k^{1/2}\{S_k^*(x) - S_k(x)\} \xrightarrow{D} Z_k$, $k = 1, \ldots, G$. *If there exists at least one* g' *with* $S_{g'}(x) = S_g(x)$, *then*

$$E(\hat{Z}_g^2) < E(Z_g^2).$$

If no such g' *exists, then* $\hat{S}_g(x)$ *and* $S_g^*(x)$ *are asymptotically equivalent.*

Thus, the pointwise C-NPMLE has smaller asymptotic mean square error than -the Kaplan-Meier estimator. In fact, a stronger inequality relation holds. Namely $pr[|\hat{Z}_g| \leq \epsilon] > pr[|Z_g| \leq \epsilon]$ for all $\epsilon > 0$. In the two sample case, we calculate asymptotic bias and asymptotic mean squared error of the C-NPMLE in section 3.4.2.

3.4.2 The Two-sample Case, $G = 2$

If $S_1(x) > S_2(x)$, then asymptotically, the constraint is irrelevant and $n_1^{1/2}\{\hat{S}_1(x) - S_1(x)\} \to \sigma_1(x)\bar{Z}_1$ and $n_2^{1/2}\{\hat{S}_2(x) - S_2(x)\} \to \sigma_2(x)\bar{Z}_2$ in distribution, as $n_1, n_2 \to \infty$, where $\bar{Z}_1$ and $\bar{Z}_2$ are independent standard normal random variables.

If $S_1(x) = S_2(x)$, from Theorem 3.6, we consider $n_1 \to \infty$ and $n_2 \to \infty$ such that $n_1/n_2 \to c_1/c_2$ as before. In this case,

$$\begin{aligned} & n_1^{1/2}\{\hat{S}_1(x) - S_1(x)\} \xrightarrow{D} \sigma_1(x) \max\left\{\bar{Z}_1, \frac{\bar{Z}_1 + c(x)^{1/2}\bar{Z}_2}{1 + c(x)}\right\} \\ \text{and} \qquad & n_2^{1/2}\{\hat{S}_2(x) - S_2(x)\} \xrightarrow{D} \sigma_2(x) \min\left\{\bar{Z}_2, \frac{c(x)\bar{Z}_2 + c(x)^{1/2}\bar{Z}_1}{1 + c(x)}\right\} \end{aligned} \tag{3.9}$$

where $c(x) = \lim_{n_1, n_2 \to \infty} n_2\sigma_1^2(x)/\{n_1\sigma_2^2(x)\} = c_2\sigma_1^2(x)/\{c_1\sigma_2^2(x)\}$.

When $S_1(x) = S_2(x)$, direct calculation from (3.9) shows that the asymptotic mean squared errors are

$$\begin{aligned} & \lim_{n_1 \to \infty} E[n_1\{\hat{S}_1(x) - S_1(x)\}^2] = \frac{\{2 + c(x)\}\sigma_1^2(x)}{2\{1 + c(x)\}} \\ \text{and} \qquad & \lim_{n_2 \to \infty} E[n_2\{\hat{S}_2(x) - S_2(x)\}^2] = \frac{\{1 + 2\,c(x)\}\sigma_2^2(x)}{2\{1 + c(x)\}}. \end{aligned} \tag{3.10}$$

These two asymptotic mean squared errors are always smaller than the unrestricted counterparts $\sigma_1^2(x)$ and $\sigma_2^2(x)$.

Let $\tilde{S}_1(x)$ and $\tilde{S}_2(x)$ be the estimates of Rojo (2004) or El Barmi and Mukerjee (2005). Based on definitions of their estimators, when $S_1(x) = S_2(x)$, the asymptotic mean squared errors are

$$\begin{aligned} & E[n_1\{\tilde{S}_1(x) - S_1(x)\}^2] = \sigma_1^2(x) + \frac{c\{\sigma_2^2(x) - (2 + c)\sigma_1^2(x)\}}{2\{1 + c\}^2} \\ \text{and} \qquad & E[n_2\{\tilde{S}_2(x) - S_2(x)\}^2] = \sigma_2^2(x) + \frac{c\{\sigma_1^2(x) - (1 + 2c)\sigma_2^2(x)\}}{2\{1 + c\}^2}. \end{aligned} \tag{3.11}$$

where $c = c_2/c_1$.

It can be shown that the asymptotic mean squared error of $\hat{S}_g(x)$ is less than or equal to that of $\tilde{S}_g(x), g = 1, 2$ and the equality holds only when $\sigma_1^2(x) = \sigma_2^2(x)$, in which case $\tilde{S}_g(x)$ and $\hat{S}_g(x), g = 1, 2$ are asymptotically equivalent. From equation (3.11) we see that when $\sigma_2^2(x)/\sigma_1^2(x) > c_2/c_1 + 2$, the Rojo's estimator $\tilde{S}_1(x)$ is asymptotically less efficient than the Kaplan-Meier estimator $S_1^*(x)$ and when $\sigma_1^2(x)/\sigma_2^2(x) > c_1/c_2 + 2$, $\tilde{S}_2(x)$ is asymptotically less efficient compared to $S_2^*(x)$.

From equation (3.9), the asymptotic bias of $\hat{S}_g(x), g = 1, 2$ can be calculated as follows

$$\begin{aligned}\lim_{n_1\to\infty} E[n_1^{1/2}(\hat{S}_1(x) - S_1(x))] &= \sigma_1(x)\int_{-\infty}^{\infty}\int_{c^{1/2}(x)}^{\infty}\frac{c^{1/2}z_2 - c(x)z_1}{1 + c(x)} f_{\bar{Z}_2}(z_2) f_{\bar{Z}_1}(z_1) dz_2 dz_1 \\ &= \sigma_1(x)\left[\frac{c(x)}{2\pi\{1 + c(x)\}}\right]^{\frac{1}{2}} < \sigma_1(x)\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \approx 0.4\sigma_1(x).\end{aligned}$$

Similarly,

$$\lim_{n_2\to\infty} E[n_2^{1/2}\{\hat{S}_2(x) - S_2(x)\}] = \sigma_2(x)\left[\frac{1}{2\pi\{1 + c(x)\}}\right]^{\frac{1}{2}} < \sigma_2(x)\left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \approx 0.4\sigma_2(x).$$

This asymptotic bias is less than 0.4 times the standard deviation of the Kaplan-Meier estimator. As shown in Theorem 3.8, the mean squared error of the C-NPMLE is less than that of the Kaplan-Meier estimate.

3.5 Confidence Intervals

3.5.1 Bootstrap

While there is a rich literature on the point estimation of a survivor functions under stochastic ordering constraint, there has been little discussion on constructing confidence intervals of the survivor functions under stochastic ordering. Rojo (2004) discussed weak convergence to a Gaussian process of his estimator from which confidence band could be constructed. Here we introduce several other possible ways to construct confidence intervals, and compare their properties in section 3.6.3.

Two sampling schemes for bootstrapping are presented. The first is the standard scheme, in which survival time and censoring indicator pairs are sampled with replacement from the data set. The second is a scheme based on simulating bootstrap samples from the constrained estimator; that is, event times are sampled from $\hat{S}_g(t)$ and censoring times are sampled from $S_g^{c*}(t)$, the Kaplan-Meier estimate of censoring survivor function in group g. For each bootstrap sample, a bootstrap estimate $\hat{S}_g^b(t), b = 1, \ldots, B$ is obtained by applying the pointwise C-NPMLE. Confidence intervals based on these bootstrap estimates can be constructed using percentiles (Efron, 1981), pivotal or percentile-t tests (Hall, 1992) or bias corrected and accelerated tests (Efron, 1987).

As a comparison, we also tried standard bootstrap and then applied Kaplan-Meier estimator to obtain bootstrap estimates, which is bootstrapping the unrestricted Kaplan-Meier estimator.

3.5.2 Confidence Interval Centered on Constrained Estimator

Hwang and Peddada (1994) suggested a method in which a confidence interval is computed for the unrestricted estimator and then shifted and centered on the constrained estimator. They showed that, under fairly general conditions, for elliptically symmetrically distributed random variables, the coverage probability for the shifted interval will exceed the nominal level. For the survivor function, we apply the intervals on a log transformed scale and consider the following approximate confidence interval,

$$\hat{S}_g(x) \exp\{\pm 1.96 \sigma_g^*(x)\},$$

where $\sigma_g^*(x)$ is standard error estimate of $\log S_g^*(x)$.

Another method to obtain confidence intervals centered on $\hat{S}_g(x)$ is by using the bootstrap, which is to sample event times from $\hat{S}_g(t)$ and censoring times from $S_g^{c*}(t)$ to obtain a censored bootstrap sample, then fit the Kaplan-Meier estimator to obtain bootstrap estimates $S_g^{*b}(x), b = 1, \ldots, B$ for each group, from which confidence intervals are obtained using percentiles.

We also introduce a reduced width confidence interval centered on the constrained estimator under the simple ordering constraint. Let $[\hat{l}_i, \hat{u}_i]$ be an unrestricted confidence interval for $S_i(x)$. We define the reduced width CI for $S_g(x)$ as $[\hat{\ell}'_g, \hat{u}'_g]$, where $\hat{u}'_g = \min_{1 \le i \le g} \hat{u}_i$ and $\hat{l}'_g = \max_{g \le i \le G} \hat{l}_i$, where $g = 1, \ldots, G$.

3.6 Simulation Studies

3.6.1 Two-sample Case when Sample Size is Small

We have conducted numerous simulation studies to compare finite sample properties of three different constrained estimators, Rojo (2004), the C-NPMLE in Chapter II and the pointwise C-NPMLE, and compare them to the unconstrained Kaplan-Meier estimator in the two-sample case. In this paper, we show results for the scenarios as described in Table 3.1 where the constraint is $S_1(t) \ge S_2(t)$ for all t.

Table 3.1: Underlying Distributions in a simulation study in section 3.6.1. When event random variable follows a $U(0,1)$ distribution, the censoring distributions $\exp(3.2)$, $\exp(1.6)$ and $\exp(0.67)$ give approximately 70%, 50% and 30% censoring rates.

	Underlying Distributions		Censoring Distributions		Sample size	
	$S_1(t)$	$S_2(t)$	$S_1^c(t)$	$S_2^c(t)$	n_1	n_2
Figure 3.1a	$\exp(1)$	$\exp(1.1)$	$\exp(1)$	$\exp(1)$	80	80
Figure 3.1b	$\exp(1)$	$\exp(1.1)$	$\exp(1)$	No censoring	80	80
Figure 3.1c	$\exp(1)$	$\exp(1.1)$	$\exp(2)$	No censoring	80	80
Figure 3.1d	$\exp(1)$	$\exp(1.1)$	$U(0,1.6)$	No censoring	80	80
Figure 3.1e	$U(0,1)$	$U^{1.1}(0,1)$	$\exp(3.2)$	$\exp(1.6)$	50	50
Figure 3.1f	$U(0,1)$	$U^{1.1}(0,1)$	$\exp(0.67)$	$\exp(3.2)$	50	50

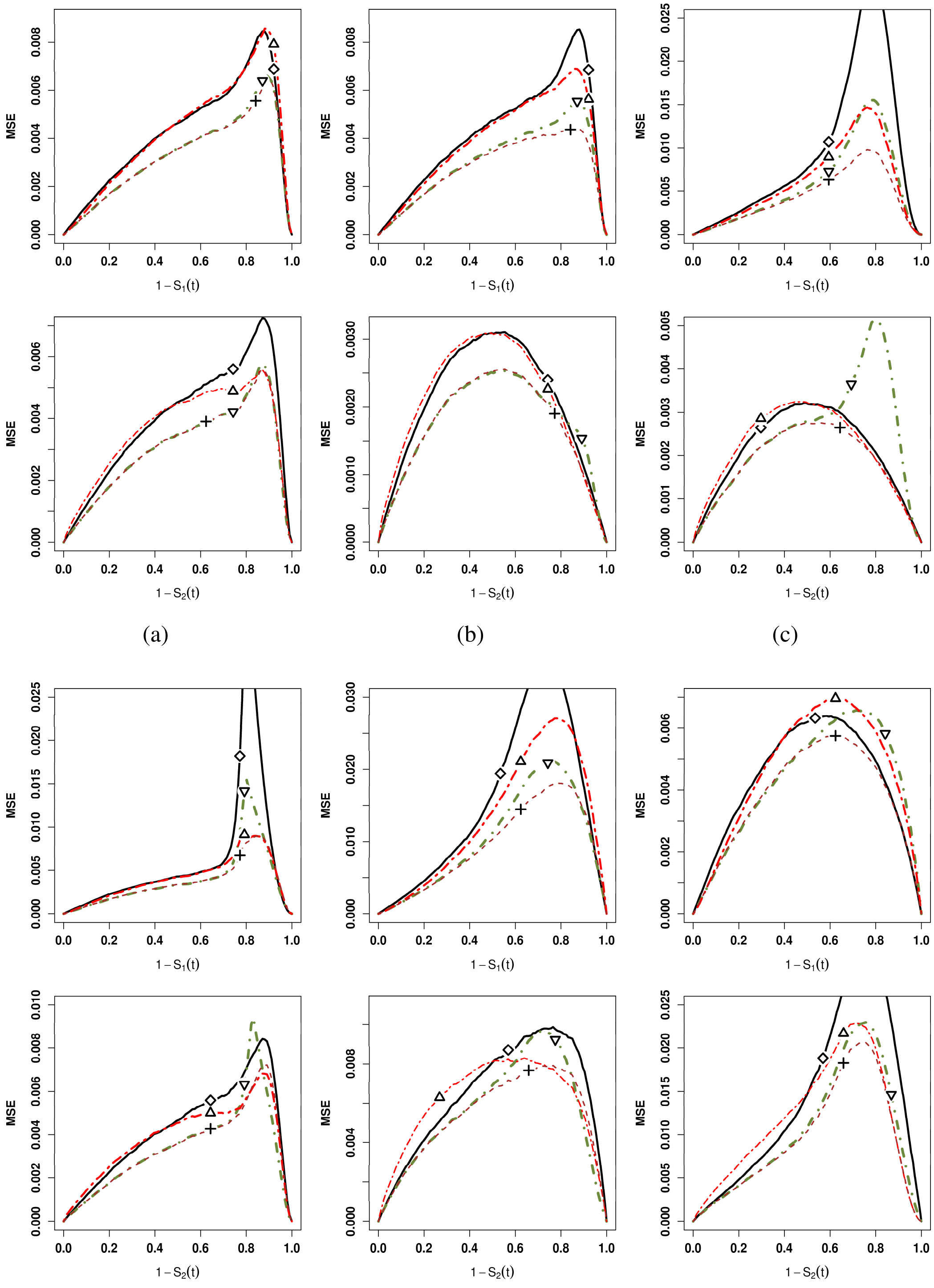
The mean squared error of the estimates of the survivor functions over a range of values of t are shown. Each simulation consists of 10,000 replications. The upper (lower) plot of each sub-figure in Figure 3.1 shows the mean squared error of estimates of $S_1(t)$ ($S_2(t)$).

In cases with the same censoring distributions (Figure 3.1a), Rojo's estimator and the pointwise C-NPMLE have smaller mean squared error compared to other estimators. We observed the Rojo's estimator is almost as efficient as the pointwise C-NPMLE when censoring distributions are the same or very close to each other. However, if population 1 and 2 have different censoring distributions, the pointwise C-NPMLE in general has smaller mean squared error compared to all other estimators. Rojo's estimator does not adjust well to the unequal censoring distributions (Figure 3.1b – 3.1f) even when the censoring rates are close to each other (Figure 3.1d). Note that the pointwise C-NPMLE is the only estimator that dominates the Kaplan-Meier estimator at any time in any situation considered.

In summary, we found that the pointwise C-NPMLE has smaller mean squared error almost everywhere and is a suggested estimator when estimating survivor functions under stochastic ordering constraint.

3.6.2 Two-sample Case: Asymptotic Properties

We define the asymptotic relative efficiency as the inverse ratio of the mean squared errors and compare the asymptotic relative efficiency of the three constrained estimators to the Kaplan-Meier estimator in the two sample case in Figure 3.2. The underlying distributions are $S_1(t) = S_2(t) = \exp(-t)$, $S_1^c(t) = 1$ and $S_2^c(t) = \exp(-2t)$. Note that the constraint is asymptotically relevant everywhere. We set $\lim_{n_1,n_2\to\infty} n_1/n_2 = 1$. The asymptotic relative efficiency of the full constrained nonparametric maximum likelihood estimator is based on simulated data with very large sample size. Asymptotic relative effi-



(a) (b) (c)

(d) (e) (f)

Figure 3.1: Comparison of mean squared errors for different estimators. ◇: Kaplan-Meier estimator; △: C-NPMLE; ▽: Rojo's estimator; +: pointwise C-NPMLE.

ciencies of the pointwise C-NPMLE and the Rojo's estimator are calculated using equations (3.10) and (3.11).

The pointwise C-NPMLE dominates all other estimators for all t while Rojo's estimator could be inefficient for some t (left graph in Figure 3.2). Note that compared to the Kaplan-Meier estimator, the full constrained nonparametric maximum likelihood estimator is less efficient everywhere in this setting.

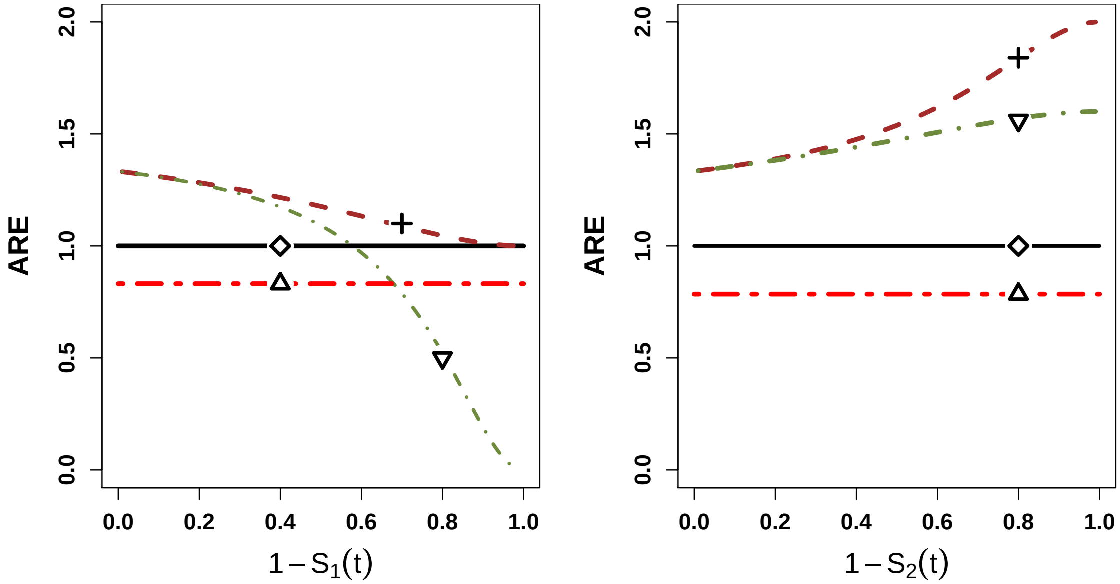


Figure 3.2: Comparison of asymptotic relative efficiency under the constraint $T_1 \geq_{st} T_2$. The underling distributions are: $S_1(t) = S_2(t) = \exp(-t)$, $S_1^c(t) = 1$ and $S_2^c(t) = \exp(-2t)$. $\Diamond$: Kaplan-Meier estimator; $\triangle$: C-NPMLE; $\triangledown$: Rojo's estimator; $+$: pointwise C-NPMLE.

3.6.3 Simple Ordering Case

In this section, we compare finite sample properties of the pointwise C-NPMLE with the Kaplan-Meier estimator in the simple ordering case and investigate the CIs described in section 3.5. We consider three groups with underlying distributions $T_1 \sim \exp(1)$, $T_2 \sim \exp(1.1)$ and $T_3 \sim \exp(1.4)$ and a uniform censoring distribution $C \sim U(0, 4.3)$, which yields an overall censoring rate of about 20%. Sample sizes are $n_1 = n_3 = 40$ and $n_2 = 20$.

The simulation is based on 10,000 replicates and 200 bootstrap estimates. We construct the CIs using percentiles (Efron, 1981) at time 0.26 and 0.63, where the survival rates of group 2 are 0.75 and 0.5 respectively.

Figure 3.3 shows the mean squared error of the pointwise C-NPMLEs and Kaplan-Meier estimators and we see efficiency gains at all times for all group. The largest gains are in the estimation of $S_2(t)$ where the mean squared error of the pointwise C-NPMLE is less than half of the mean squared error of the Kaplan-Meier estimator at almost all the times.

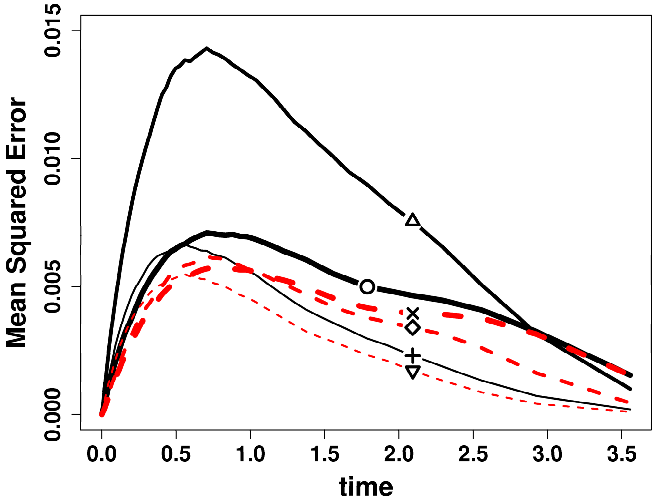


Figure 3.3: Comparison of Kaplan-Meier estimator and pointwise C-NPMLE in the three sample case. $\bigcirc$, $\triangle$ or $+$: Kaplan-Meier estimates for group 1, 2 or 3; $\times$, $\Diamond$ or $\bigtriangledown$: pointwise C-NPMLE for group 1, 2 or 3.

The coverage rates and average widths of the confidence intervals described in Section 3.5 are shown in Table 3.2. As expected, the confidence intervals centered on the pointwise C-NPMLE have higher coverage rate (V vs. I and VI vs. II), but these confidence intervals tend to be too conservative while the average widths are about the same as the unrestricted counterparts. The bootstrap methods (III and IV) give the confidence intervals with significantly reduced width, but the coverage rates tend to be lower for some groups (III and IV at time 0.26, group 1). The average widths of reduced width confidence intervals for group 2 are significantly smaller than the original confidence intervals while the coverage rates are about the same (V' vs. V and VI' vs. VI). Bootstrap methods III and IV are not ideal but seem best among these considered and have quite reasonable coverage properties.

Table 3.2: Coverage rate (%) and average width (%) of 95% confidence intervals. The distribution functions are $T_1 \sim \exp(1)$, $T_2 \sim \exp(1.1)$, $T_3 \sim \exp(1.4)$ and $C \sim \text{uniform}(0, 4.3)$ for all groups. Total censoring rate is 20%. Sample sizes are $n_1 = 40$, $n_2 = 20$ and $n_3 = 40$. Bootstrapping confidence interval is based on percentiles (Efron, 1981). Results are based on 10,000 simulation samples.

		$t = 0.26$			$t = 0.63$	
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
I	92.8 (25.3)	92.8 (35.6)	92.8 (27.8)	93.7 (31.2)	93.1 (43.5)	93.2 (30.8)
II	92.1 (26.1)	89.2 (37.5)	93.2 (28.8)	94.5 (32.4)	93.3 (46.3)	94.5 (32.3)
III	90.1 (21.0)	97.5 (23.9)	94.7 (25.0)	92.9 (27.3)	97.5 (29.2)	93.9 (27.2)
IV	89.9 (20.4)	94.8 (22.5)	93.3 (24.4)	92.6 (26.6)	94.5 (27.7)	93.0 (26.9)
V	94.0 (24.7)	99.4 (37.1)	96.5 (28.2)	95.8 (31.2)	99.3 (44.3)	95.4 (30.7)
V'	94.0 (24.6)	99.3 (31.2)	96.5 (28.1)	95.8 (31.1)	99.2 (37.9)	95.4 (30.6)
VI	92.4 (26.8)	97.5 (37.9)	96.8 (28.3)	95.0 (33.6)	98.8 (46.6)	96.9 (31.7)
VI'	91.6 (26.1)	97.4 (31.5)	96.8 (28.3)	94.8 (33.2)	98.8 (39.6)	96.9 (31.7)

I: Standard bootstrap then Kaplan-Meier estimator; II: Log transformation, i.e. $S_g^*(x)\exp\{\pm 1.96\sigma_g^*(x)\}$; III: Sampling from $\hat{S}$ (the pointwise C-NPMLE) then $\hat{S}$; IV: standard bootstrap then $\hat{S}$; V: Sampling from $\hat{S}$ then Kaplan-Meier estimator; V': V with width reduced; VI: $\hat{S}_g(x)\exp\{\pm 1.96\sigma_g^*(x)\}$; VI': VI with width reduced.

3.7 Example

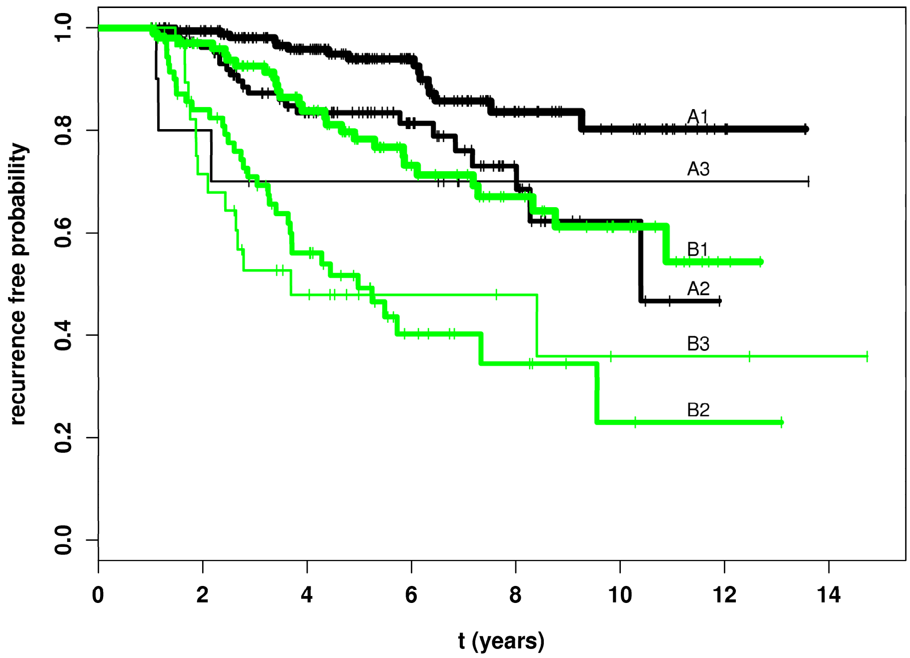
The data are from prostate cancer patients who received radiation therapy at the University of Michigan Hospital, a portion of data used in Proust-Lima and Taylor (2009). Five-hundred and three patients without planned hormonal therapy are included in this example to estimate the survivor function of time to first recurrence of prostate cancer. For this analysis recurrence is defined as the first occurrence of local recurrence, distant metastasis or initiation of salvage hormone therapy.

It is reasonable to expect that the patients with higher baseline prostate-specific antigen value have a higher recurrence rate of prostate cancer than those with lower baseline prostate-specific antigen values. The Gleason grade is a measure of the aggressiveness of the tumor cells obtained from microscopic inspection of a biopsy prior to the treatment. It is reasonable to expect that the patients with lower Gleason grade have a lower recurrence rate. In this example, we divided the patients into 6 groups based on baseline prostate-specific antigen is less than 10 or larger than or equal to 10 (A or B) and Gleason ≤ 6, $= 7$ or ≥ 8 (1, 2 or 3). Thus A1 represents the patient group with baseline prostate-specific antigen < 10 and Gleason ≤ 6 and B3 the patient group with baseline prostate-specific antigen ≥ 10 and Gleason ≥ 8. The natural set of constraints for the survivor functions are $A1 \geq A2 \geq A3, B1 \geq B2 \geq B3, A1 \geq B1, A2 \geq B2$ and $A3 \geq B3$.

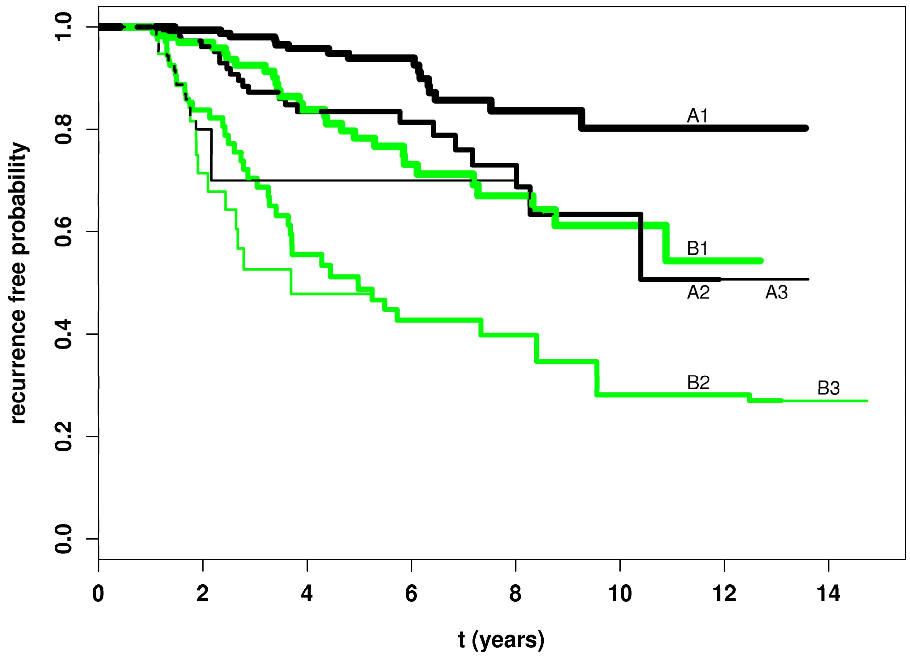
The Kaplan-Meier estimates of each groups are shown in Figure 3.4a. The unrestricted Kaplan-Meier estimates do not satisfy the stochastic ordering. Specifically we notice that between 1 year and 2.5 years, the groups $A2$, $B2$ and $B3$ do not satisfy ordering constraint and after 5 years the orderings of $A2$ and $A3$, and $B2$ and $B3$ are also violated.

The pointwise C-NPMLEs are shown in Figure 3.4b. The estimates satisfy the stochastic ordering constraint at all times. Between 1 year and 2.5 years, we notice that the survivor functions take a common value in groups $A2$, $B2$ and $B3$ and after 5 years, groups $A2$ and $A3$, $B2$ and $B3$ are combined together respectively. Interestingly at around 12.5 years, there is a jump of the survivor function in groups $B2$ and $B3$, even though there is no observed event.

Detailed results of point estimates and corresponding confidence intervals for some selected times are shown in Table 3.3.



(a) Kaplan-Meier estimates



(b) the pointwise C-NPMLE

Figure 3.4: Estimates in the University of Michigan prostate cancer example. Here A or B represent baseline prostate-specific antigen is $<$ or ≥ 10, and 1, 2 or 3 represents Gleason score is $<, =$ or > 7. The vertical bars in Kaplan-Meier estimator are observed censoring times.

Table 3.3: Estimates and confidence intervals of survivor functions for some selected times in the prostate cancer example. Confidence interval using standard bootstrap method are shown in parenthesis. Bootstrap confidence interval is based on percentiles. The unit of the results is percentage (%).

Time (years)		1.5	5	8.5
A1	Kaplan-Meier	99.4 (98.1, 100)	93.9 (89.5, 97.8)	83.6 (75.1, 91.1)
	PC-NPMLE	99.4 (98.5, 100)	93.9 (89.5, 97.8)	83.6 (75.3, 91.2)
A2	Kaplan-Meier	99.1 (97.4, 100)	83.4 (75.1, 90.3)	62.2 (43.4, 79.0)
	PC-NPMLE	99.1 (97.4, 100)	83.4 (75.5, 90.7)	63.4 (45.6, 78.9)
A3	Kaplan-Meier	80.0 (60.0, 100)	70.0 (40.0, 100)	70.0 (0.0, 90.0)
	PC-NPMLE	88.7 (82.1, 99.6)	70.0 (44.7, 87.4)	63.4 (20.0, 75.1)
B1	Kaplan-Meier	98.0 (95.0, 100)	78.3 (69.0, 86.6)	64.2 (51.7, 75.5)
	PC-NPMLE	98.0 (95.0, 100)	78.3 (69.0, 86.6)	64.2 (51.7, 75.5)
B2	Kaplan-Meier	86.8 (79.2, 94.2)	48.8 (35.7, 62.3)	34.2 (14.6, 50.8)
	PC-NPMLE	88.7 (82.9, 94.4)	48.8 (38.4, 62.4)	34.7 (22.1, 51.0)
B3	Kaplan-Meier	96.4 (88.9, 100)	47.8 (28.2, 66.2)	35.9 (10.0, 60.7)
	PC-NPMLE	88.7 (81.8, 94.2)	47.9 (29.1, 57.6)	34.7 (10.0, 46.9)

A or B: baseline prostate-specific antigen < 10 or ≥ 10; 1,2 or 3: Gleason $<, =$ or > 7; PC-NPMLE: Pointwise C-NPMLE.

3.8 Discussion

The pointwise C-NPMLE is a likelihood based pointwise estimator. Unlike the full constrained nonparametric estimator maximum likelihood estimator, the violation of a constraint at one time does not affect the estimates at other times. The "weights" for each group, which are based on maximizing the likelihood when the constrains are violated, are natural and give better results than using weights based on initial sample size of each group as used by Rojo (2004) and El Barmi and Mukerjee (2005).

Rojo's estimator in the two-sample case and El Barmi and Mukerjee's estimator in the simple ordering case are the pointwise C-NPMLE without censoring. However, if censoring exists, these estimators are quite different in some situations, especially when the censoring distributions are significantly different between groups. Another feature of the El Barmi and Mukerjee's estimator is the range of times for which the estimator is defined. Specifically, it is defined only until the minimum time of the last observations in all groups. If the last observed time in one group is much earlier than others, which could happen when the sample size of a group is much smaller than others, then all the estimators are subsequently undefined at times even though there may be a large number of observations at risk in the other groups. On the other hand, the pointwise C-NPMLE for a group is defined up to the last observed time of that group and is not affected by short valid ranges of other groups.

Because the pointwise C-NPMLE can have jumps at non-event times, the likelihood ratio tests of survivor functions, first introduced by Thomas and Grunkemeier (1975) and discussed by Li (1995) and Murphy (1995) may be problematic, because they assume the jumps only occur at event times. Thus the likelihood ratio test and confidence interval and confidence band based on the likelihood ratio test may need to be revised.

The methods we introduced to construct confidence interval are far from perfect. The bootstrap methods can have lower coverage rate than the nominal level for some populations while confidence intervals centered on the pointwise C-NPMLE all tend to have coverage rates that are too high without reducing the width of the confidence interval. Investigating other methods of constructing confidence interval is an interesting area of study.

We have observed that the pointwise C-NPMLE may be less efficient than the unconstrained Kaplan-Meier estimator under certain partial ordering constraints, particularly under tree ordering constraints, $T_1 \leq_{st} T_2, T_1 \leq_{st} T_3, \ldots, T_1 \leq_{st} T_G$. The estimation problem of normal means under partial ordering constraint is discussed by Lee (1988); Hwang and Peddada (1994); Fernandez et al. (1999); Cohen and Sackrowitz (2002); Chaudhuri and Perlman (2005, 2007). The strategy discussed in Hwang and Peddada (1994) is to reorder and switch some of the estimators and then apply fixed width confidence intervals. They showed that this approach did improve the coverage probabilities of the confidence intervals. It is an open question to see if their strategy can also be shown to improve the properties in the constrained survival data setting.

CHAPTER IV

Confidence Intervals under Order Restriction

4.1 Introduction

Consider a G-sample problem where the observations $X_{gi}, g = 1, \ldots, G, i = 1, \ldots, n_g$ are independent random variables with distribution function $F_g(x; \mu_g)$. When estimating $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_G)$, there often exists information about the inequality orderings and bounds of these parameters. For example, if μ_g is the average height of children of age g or μ_g is the toxicity rate of a drug for dose level g in a clinical trial, the parameters should satisfy the restriction:

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_G. \tag{4.1}$$

This type of ordering is called *simple ordering* or *linear ordering*. The natural estimator for order restricted parameters is the restricted maximum likelihood estimator (MLE). For the case where $F_g(x; \mu_g), g = 1, \ldots, G$ is a normal distribution function with mean μ_g and variance σ^2, the MLE of $\hat{\mu}_g$ under restriction (4.1) is the isotonic regression estimator (Barlow et al., 1972; Robertson et al., 1988; Silvapulle and Sen, 2005). The restricted MLE

$\hat{\mu}_g$ has been shown to dominate the unrestricted MLE $\bar{X}_g$ in the sense that

$$P(|\hat{\mu}_g - \mu_g| \leq c) \geq P(|\bar{X}_g - \mu_g| \leq c), g = 1, \ldots, G, \tag{4.2}$$

for all $c > 0$ (Kelly, 1989; Lee, 1981). In this paper, we focus on constructing confidence intervals for the parameters μ_g under the linear ordering constraints.

Estimation problems in a restricted parameter space have been studied since the 1950s. Marchand and Strawderman (2004) and van Eeden (2006) reviewed estimation methods that have been developed in the past and discussed the "good" properties of restricted estimators, such as dominance, minimax and admissibility. Cohen and Sackrowitz (2004) discussed some inference issues and pointed out that traditional inference methods, such as likelihood based method, can lead to some undesirable properties in restricted parameter problems. Andrews (2000) also pointed out that the bootstrap method, which has been very useful for constructing confidence intervals of complicated parameters, will fail when a parameter is on the boundary or close to the boundary of the parameter space. Thus, it is of interest to develop an inference procedure without depending on traditional inference methods.

Specialized methods for constructing confidence intervals under order restrictions have been suggested. Schoenfeld (1986) proposed a method for one-sided intervals based on inverting the likelihood ratio test for the ordered means from a normal distribution. Hwang and Peddada (1994) proposed constant length confidence intervals, in which the confidence interval, derived without the order restriction assumption, is shifted and centered at an improved estimator, e.g. centered at the restricted MLE in the linear ordering case. From the dominance properties described in equation (4.2), coverage rates of these restricted methods exceed the nominal levels obtained from unrestricted intervals. In addition, bootstrap and other resampling procedures are discussed by Peddada (1997) and Li et al. (2010).

In this paper, we propose a novel method to construct confidence intervals under a linear ordering constraint. In section 4.2, we consider a two-sample case of ordered normal means with known variances and obtain some theoretical results about the coverage rate and width of the confidence interval. In section 4.3, we propose methods for the case when the population variances are unknown. We extend the methods to the case with three or more samples in section 4.4. In section 4.5, we describe some other confidence intervals that have been proposed in the literature and in section 4.6, we conduct simulation studies to compare those confidence intervals with our approach. In section 4.7, we illustrate the method using data on half-lives of an antibiotic in an animal study.

4.2 Confidence Intervals for μ_1 and μ_2 with Known Variances

4.2.1 Family of Confidence Intervals

Let $X_g \sim N(\mu_g, \sigma_g^2), g = 1, 2$, where σ_g^2 is known. Our goal is to construct $1 - \alpha$ confidence intervals for μ_1 and μ_2 when it is known that $\mu_1 \leq \mu_2$. Let $X(\gamma) = \gamma X_1 + (1 - \gamma)X_2$, where $\gamma \in [0, 1]$. The mean and variance of $X(\gamma)$ are $\mu(\gamma) = EX(\gamma) = \gamma\mu_1 + (1 - \gamma)\mu_2$ and $\sigma^2(\gamma) = \text{var}\{X(\gamma)\} = \gamma^2\sigma_1^2 + (1 - \gamma)^2\sigma_2^2$. Let $z_{1-\alpha/2}$ be the upper $\alpha/2$ quantile of a standard normal distribution and let $t_{1-\alpha/2,\nu}$ be the upper $\alpha/2$ quantile of a standard t distribution with degree of freedom ν, which we denote for convenience by z and t_ν. The unrestricted confidence intervals for μ_1, μ_2 and $\mu(\gamma)$ are $\mu_g \in [X_g - z\sigma_g, X_g + z\sigma_g]$, $g = 1, 2$ and $\mu(\gamma) \in [X(\gamma) - z\sigma(\gamma), X(\gamma) + z\sigma(\gamma)]$. Since $\mu_1 \leq \mu(\gamma) \leq \mu_2$, it is sensible to consider modifying the limits of the confidence intervals for μ_1 and μ_2, based on the limits of the confidence interval $\mu(\gamma)$. We here propose a family of confidence interval $[L_1(\gamma), U_1(\gamma)]$

for μ_1 and $[L_2(\gamma), U_2(\gamma)]$ for μ_2 as follows:

$$\begin{aligned} L_1(\gamma) &= \min\{X_1 - z\sigma_1, X(\gamma) - z\sigma(\gamma)\}, \\ U_1(\gamma) &= \min\{X_1 + z\sigma_1, X(\gamma) + z\sigma(\gamma)\}, \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} L_2(\gamma) &= \max\{X_2 - z\sigma_2, X(\gamma) - z\sigma(\gamma)\}, \\ U_2(\gamma) &= \max\{X_2 + z\sigma_2, X(\gamma) + z\sigma(\gamma)\}. \end{aligned} \tag{4.4}$$

Henceforth in sections 2 and 3, we only consider upper and lower limits for μ_1. Those for μ_2 will be of the same form except for changing $\min$ to $\max$ and changing the subscript from 1 to 2. In this section, we develop the method and theory for the case of one observation per group. The results apply to multiple observations per groups by simply replacing X_g by the group mean and replacing σ_g^2 by σ_g^2/n_g.

The following is preliminary to a discussion of the coverage probability for μ_1 using $L_1(\gamma)$ and $U_1(\gamma)$.

Definition 4.1. $\boldsymbol{X} = (X_1, \ldots, X_k)$ has an elliptical unimodal distribution with location $\boldsymbol{\mu}$ and positive-definite matrix $\boldsymbol{\Sigma}$ if its probability density function is

$$f(\boldsymbol{x}) = C\, h\{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\},$$

where $h(t)$ is a nonincreasing function in t.

Theorem 4.2. *Suppose $\boldsymbol{Y} = (Y_1, Y_2)^T$ has a bivariate elliptical unimodal distribution with location $\boldsymbol{\mu} = (0, \Delta)$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\Delta \geq 0$. Define c_α so that $pr(|Y_1| \leq c_\alpha) = 1 - \alpha$. Then*

$$Q = P\{\min(Y_1 - c_\alpha, Y_2 - c_\alpha) \leq 0 \leq \min(Y_1 + c_\alpha, Y_2 + c_\alpha)\} \geq 1 - \alpha. \tag{4.5}$$

Proof. The joint probability density function of $\boldsymbol{Y}$ is

$$f(y_1, y_2) = C\,h\left\{y_1^2 + (y_2 - \Delta)^2 - 2\rho y_1(y_2 - \Delta)\right\}.$$

Then, from (4.5),

$$\begin{aligned} Q &= P(Y_1 \geq -c_\alpha, Y_2 \geq -c_\alpha) - P(Y_1 \geq c_\alpha, Y_2 \geq c_\alpha) \\ &= \int\limits_{\mathbb{D}\cup\mathbb{E}\cup\mathbb{H}} f(y_1, y_2)dy_1dy_2 \\ &= \int\limits_{\mathbb{D}\cup\mathbb{E}\cup\mathbb{F}} f(y_1, y_2)dy_1dy_2 + \int\limits_{\mathbb{H}} f(y_1, y_2)dy_1dy_2 - \int\limits_{\mathbb{F}} f(y_1, y_2)dy_1dy_2, \end{aligned} \tag{4.6}$$

where $\mathbb{D}$, $\mathbb{E}$, $\mathbb{F}$ and $\mathbb{H}$ are defined in Figure 4.1. It is clear that

$$\int\limits_{\mathbb{D}\cup\mathbb{E}\cup\mathbb{F}} f(y_1, y_2)dy_1dy_2 = P\{|Y_1| \leq c_\alpha\} = 1 - \alpha. \tag{4.7}$$

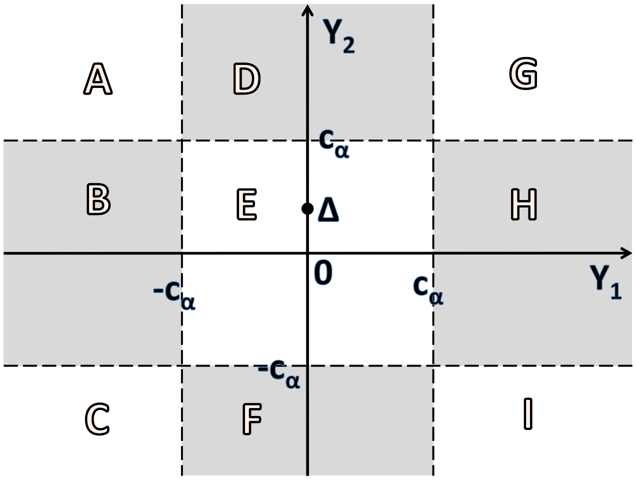


Figure 4.1: Sample space of (y_1, y_2).

Let $\boldsymbol{y}_H = (y_1, y_2)$ be a point in $\mathbb{H}$ with $\boldsymbol{y}_F = (-y_2, -y_1)$ a corresponding point in $\mathbb{F}$. It can be seen that

$$(\boldsymbol{y}_H - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_H - \boldsymbol{\mu}) - (\boldsymbol{y}_F - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_F - \boldsymbol{\mu}) = -2\Delta(1-\rho)(y_1 + y_2) \leq 0$$

for all $\Delta \geq 0$ and $(y_1, y_2) \in \mathbb{H}$. Since $h(t)$ is nonincreasing, we have $f(y_1, y_2) \geq f(-y_2, -y_1)$ if $\Delta \geq 0$ and $(y_1, y_2) \in \mathbb{H}$. Thus the density at each point in $\mathbb{H}$ is greater than or equal to the density at the corresponding point in $\mathbb{F}$. Since $(-y_2, -y_1)$ varies over all of $\mathbb{F}$ as (y_1, y_2) varies over $\mathbb{H}$,

$$\int\limits_{\mathbb{H}} f(y_1, y_2) dy_1 dy_2 - \int\limits_{\mathbb{F}} f(y_1, y_2) dy_1 dy_2 \geq 0. \tag{4.8}$$

The result (4.5) follows immediately from (4.6), (4.7) and (4.8). □

Corollary 4.3. *If* $X_1 \sim N(\mu_1, \sigma_1^2)$ *and* $X_2 \sim N(\mu_2, \sigma_2^2)$ *are independent with* $\mu_1 \leq \mu_2$*, then* $P\{L_1(\gamma) \leq \mu_1 \leq U_1(\gamma)\} \geq 1 - \alpha$ *for all* $\gamma \in [0, 1]$.

Proof. This follows by setting $Y_1 = (X_1 - \mu_1)/\sigma_1$, $Y_2 = \{X(\gamma) - \mu_1\}/\sigma(\gamma)$ and $h(t) = \exp(-t/2)$ in Theorem 4.2. The corresponding $\Delta = \{\mu(\gamma) - \mu_1\}/\sigma(\gamma) = (1 - \gamma)(\mu_2 - \mu_1)/\sigma(\gamma) \geq 0$, $\rho = \{1 + (1 - \gamma)^2 \sigma_2^2/(\gamma^2 \sigma_1^2)\}^{-1/2}$ and $C = (2\pi)^{-1}(1 - \rho^2)^{-1/2}$. □

This corollary shows that the coverage rate of the interval $[L_1(\gamma), U_1(\gamma)]$ always exceeds the nominal level when the variances of X_1 and X_2 are known.

4.2.2 Selection of γ

We would like to choose a value of γ to make the width of the confidence intervals for μ_1 defined in equation (4.3) as small as possible. One possible choice of γ is the value that minimizes $W_1(\gamma) = E\{U_1(\gamma) - L_1(\gamma)\}$, but $W_1(\gamma)$ depends on the unknown mean difference $\mu_2 - \mu_1$. There is a $\gamma \in (0, 1)$ for which $\sigma^2(\gamma) < \sigma_1^2$, and it can be seen that when $\sigma^2(\gamma) < \sigma_1^2$, $U_1(\gamma) - L_1(\gamma) \leq 2z\sigma_1$ for any observations X_1 and X_2, thus the width of the interval can be reduced by suitable choice of γ. Another intuitive choice of γ is the value that minimizes $\sigma^2(\gamma)$. It is easy to see that $\sigma^2(\gamma)$ is minimized at $\hat{\gamma} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. In Theorem 4.4, we show that this $\hat{\gamma}$ also minimizes $W_1(\gamma)$ if $\mu_1 = \mu_2$.

Theorem 4.4. *$W_1(\gamma)$ is minimized at $\gamma = \hat{\gamma}$ if $\mu_1 = \mu_2$.*

Proof. Consider the case $\gamma \geq 1 - 2\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, for which $\sigma^2(\gamma) \leq \sigma_1^2$.

Let $c(\gamma) = z\{\sigma_1 - \sigma(\gamma)\}/(1-\gamma)$. Then we have

$$U_1(\gamma) - L_1(\gamma) = \begin{cases} 2z\sigma_1 & \text{if } X_2 - X_1 > c(\gamma) \\ (1-\gamma)(X_2 - X_1) + z\sigma_1 + z\sigma(\gamma) & \text{if } -c(\gamma) < X_2 - X_1 \leq c(\gamma) \\ 2z\sigma(\gamma) & \text{if } X_2 - X_1 \leq -c(\gamma). \end{cases}$$

So,

$$W_1(\gamma) = \int_{c(\gamma)}^{\infty} 2z\sigma_1 f_{X_2-X_1}(x)dx + \int_{-c(\gamma)}^{c(\gamma)} \{(1-\gamma)x + z\sigma_1 + z\sigma(\gamma)\} f_{X_2-X_1}(x)dx$$
$$+ \int_{-\infty}^{-c(\gamma)} 2z\sigma(\gamma) f_{X_2-X_1}(x)dx = z\{\sigma_1 + \sigma(\gamma)\},$$

because $X_2 - X_1 \sim N(0, \sigma_1^2 + \sigma_2^2)$, $\int_{-c(\gamma)}^{c(\gamma)} x f_{X_2-X_1}(x)dx = 0$ and $\int_c^{\infty} f_{X_2-X_1}(x)dx = \int_{-\infty}^{-c} f_{X_2-X_1}(x)dx$.

Similarly we can show that if $\gamma < 1 - 2\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, $W_1(\gamma) = z\{\sigma_1 + \sigma(\gamma)\}$.

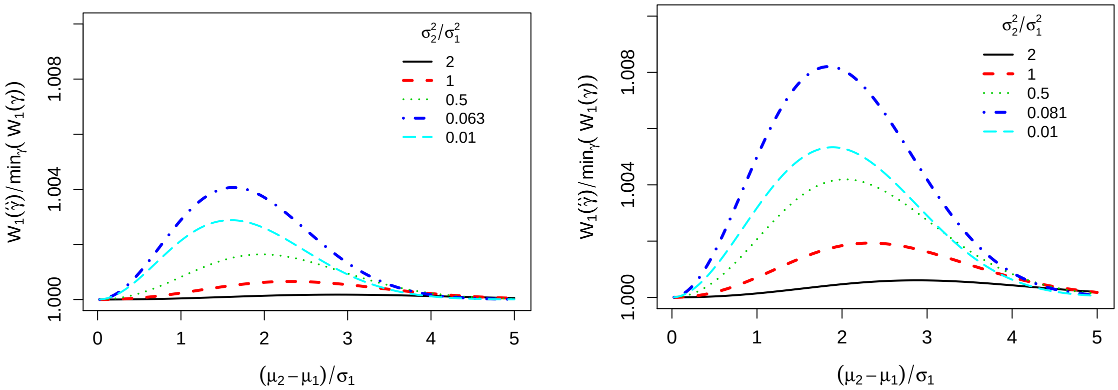
Thus minimizing $W_1(\gamma)$ is the same as minimizing $\sigma(\gamma)$, which happens at $\gamma = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. □

Using this $\hat{\gamma}$, the proposed confidence interval for μ_1 is

$$\hat{L}_1 = \min(X_1 - z\sigma_1, \hat{X} - z\hat{\sigma}), \ \ \hat{U}_1 = \min(X_1 + z\sigma_1, \hat{X} + z\hat{\sigma}). \tag{4.9}$$

where $\hat{X} = (X_1\sigma_2^2 + X_2\sigma_1^2)/(\sigma_1^2 + \sigma_2^2)$ and $\hat{\sigma}^2 = \sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)$.

An interesting question is how much wider this CI is compared to the narrowest possible CI when $\mu_1 \neq \mu_2$. To investigate this, we compute $\min_\gamma W_1(\gamma)$ through numerical minimization over γ for a given $\mu_2 - \mu_1$ and compare it with $W_1(\hat{\gamma})$. The results are shown in Figure 4.2. The largest possible average width for the confidence interval using $\hat{\gamma}$, compared to the optimal, occurs when $\sigma_2^2/\sigma_1^2 = 0.063$ for nominal level of 95% and when $\sigma_2^2/\sigma_1^2 = 0.081$ for nominal level of 99%. Compared to the minimum possible $W_1(\gamma)$, $W_1(\hat{\gamma})$ is only at most 0.4% wider at nominal level of 95% and at most 0.8% wider at nominal level of 99%. This indicates that the CI using $\hat{\gamma}$ is almost as efficient as the most efficient CI in this family.



(a) 95% CI

(b) 99% CI

Figure 4.2: Comparison of width of confidence interval using $\hat{\gamma}$ and minimum possible width.

4.2.3 Properties of Proposed Confidence Intervals

4.2.3.1 Maximum Coverage Rate

Let $\rho = \sigma_2/\sqrt{\sigma_1^2 + \sigma_2^2}$ and let $\Delta = (\mu_2 - \mu_1)(1 - \rho^2)/(\rho\sigma_1)$. Let $Y_1 = (X_1 - \mu_1)/\sigma_1$ and let $Y_2 = (\hat{X} - \mu_1)/\hat{\sigma}$. Then the joint distribution of Y_1 and Y_2 is

$$f(y_1, y_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{y_1^2 + (y_2 - \Delta)^2 - 2\rho y_1(y_2 - \Delta)}{2(1-\rho^2)} \right\}.$$

Table 4.1: Theoretical maximum coverage rate of CI for μ_1 in the situations with different ratio of variances.

$1-\alpha$	σ_2^2/σ_1^2								
	10^{-9}	10^{-3}	0.01	0.1	0.2	0.5	1	2	10
.95	.969	.969	.968	.965	.962	.959	.956	.953	.950
.90	.933	.932	.930	.924	.920	.913	.909	.905	.901
.80	.852	.850	.846	.835	.829	.819	.812	.806	.801
.70	.761	.759	.753	.740	.732	.721	.713	.707	.701

The coverage probability is

$$P = 1 - \alpha + \int_{z}^{\infty}\int_{-\infty}^{z} f(y_1, y_2)dy_2dy_1 - \int_{-z}^{\infty}\int_{-\infty}^{-z} f(y_1, y_2)dy_2dy_1.$$

Setting $dP/d\Delta = 0$, we find that maximum coverage probability of the proposed confidence interval for fixed ρ occurs at $\hat{\Delta}$ that solves the equation

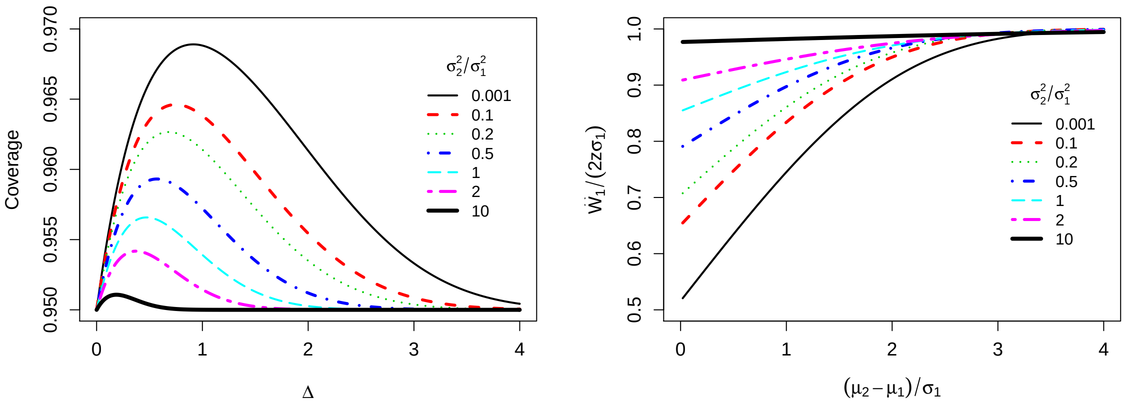
$$\Phi\left\{(z\rho - z - \Delta\rho)/\sqrt{1-\rho^2}\right\} - \Phi\left\{(z - z\rho - \Delta\rho)/\sqrt{1-\rho^2}\right\}\exp(-2z\Delta) = 0, \tag{4.10}$$

where Φ is the cumulative distribution function of standard normal random variable. It can be shown that equation (4.10) has one and only one positive solution for Δ for any $0 < \rho < 1$. As can be seen in Table 4.1 and Figure 4.3a, the theoretical maximum coverage rate increases as σ_2^2/σ_1^2 (or equivalently ρ) decreases and approaches 0.969 for nominal level of 95% when σ_2^2/σ_1^2 goes to 0.

4.2.3.2 Confidence Interval Width

The reduction of average width compared to the unrestricted confidence interval depends on σ_2^2/σ_1^2 and $\mu_2 - \mu_1$ as can be seen in Figure 4.3b. For the CI of μ_1, the smaller the σ_2^2/σ_1^2, the smaller the average width, and the closer the means, the smaller the average width. The average width can be half the width of the unrestricted CI when $\sigma_2^2/\sigma_1^2 \to 0$ and

$\mu_1 = \mu_2$. If the variance of X_1 and X_2 are similar, the average width will be about 85% of the unrestricted one when $\mu_1 = \mu_2$.



(a) Coverage Probability

(b) proportion of average width

Figure 4.3: Coverage probability and ratio of average width of restricted confidence interval for μ_1 compared to unrestricted confidence interval for μ_1.

4.3 Confidence Intervals for μ_1 and μ_2 with Unknown Variances

Suppose we observed $X_{gi} \sim N(\mu_g, \sigma_g^2)$, $g = 1, 2, i = 1, 2, \cdots, n_g$. Let $\bar{X}_g = \sum_{i=1}^{n_g} X_{gi}/n_g, g = 1, 2$. We consider two cases.

4.3.1 Restricted Confidence Interval when σ_1^2 and σ_2^2 are Unknown but σ_2^2/σ_1^2 is Known

Let $\hat{s}_1^2 = \left\{\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_{2i} - \bar{X}_2)^2/p\right\}/\nu$ and let $\tilde{X} = (\bar{X}_1 n_1 p + \bar{X}_2 n_2)/(n_2 + n_1 p)$, where $p = \sigma_2^2/\sigma_1^2$ and $\nu = n_1 + n_2 - 2$. Then $\tilde{\mu} = E\tilde{X} = (\mu_1 n_1 p + \mu_2 n_2)/(n_2 + n_1 p)$ and $\tilde{\sigma}^2 = \text{var}(\tilde{X}) = \sigma_1^2 p/(n_2 + n_1 p)$. Let $\tilde{s}^2 = p\hat{s}_1^2/(n_2 + n_1 p)$, then $(\tilde{X} - \tilde{\mu})/\tilde{s}$ and $\sqrt{n_1}(\bar{X}_1 - \mu_1)/\hat{s}_1$ follow standard T distributions with degree of freedom

ν. The unrestricted CIs for μ_1 and $\tilde{\mu}$ are:

$$\mu_1 \in [\bar{X}_1 - t_\nu \ \hat{s}_1/\sqrt{n_1}, \bar{X}_1 + t_\nu \ \hat{s}_1/\sqrt{n_1}],$$

$$\tilde{\mu} \in [\tilde{X} - t_\nu \ \tilde{s}, \tilde{X} + t_\nu \ \tilde{s}].$$

We propose a restricted CI for μ_1 as follows:

$$\begin{aligned} \hat{L}_1 &= \min(\bar{X}_1 - t_\nu \ \hat{s}_1/\sqrt{n_1}, \tilde{X} - t_\nu \ \tilde{s}), \\ \hat{U}_1 &= \min(\bar{X}_1 + t_\nu \ \hat{s}_1/\sqrt{n_1}, \tilde{X} + t_\nu \ \tilde{s}). \end{aligned} \tag{4.11}$$

When σ_2^2/σ_1^2 is known, the pivotal random variables $(X_1 - \mu_1)/\hat{s}_1$ and $(\tilde{X} - \tilde{\mu})/\tilde{s}$ follow a bivariate T distribution. Since the multivariate T both belongs to the elliptical distribution family, the result in Theorem 4.2 concerning coverage rates of CIs is applicable.

Corollary 4.5. *The confidence interval defined in (4.11) satisfies* $P(\mu_1 \in [\hat{L}_1, \hat{U}_1]) \geq 1 - \alpha$.

Proof. This follows by setting $Y_1 = (\bar{X}_1 - \mu_1)/\hat{s}_1$, $Y_2 = \{\tilde{X} - \mu_1\}/\tilde{s}$ and $h(t) = \{1 + t/\nu\}^{-(\nu+2)/2}$ in Theorem 4.2. The corresponding $\Delta = (\tilde{\mu} - \mu_1)E(1/\tilde{s}) = (\mu_2 - \mu_1)n_2 E(1/\tilde{s})/(n_2 + n_1 p) \geq 0$, $\rho = \sqrt{pn_1/(pn_1 + n_2)}$ and $C = 1/(2\pi\sqrt{1-\rho^2})$. □

4.3.2 Restricted Confidence Intervals when both σ_1^2 and σ_2^2 are Unknown

Let $\hat{s}_g^2 = \sum_{i=1}^{n_g}(X_{gi} - \bar{X}_g)^2/(n_g - 1)$, $g = 1, 2$. Unrestricted CIs for μ_1 and μ_2 are:

$$\mu_1 \in [\bar{X}_1 - t_{n_1-1}\hat{s}_1/\sqrt{n_1}, \bar{X}_1 + t_{n_1-1}\hat{s}_1/\sqrt{n_1}],$$

$$\mu_2 \in [\bar{X}_2 - t_{n_2-1}\hat{s}_2/\sqrt{n_2}, \bar{X}_2 + t_{n_2-1}\hat{s}_2/\sqrt{n_2}].$$

Similar to the approaches in earlier sections, we consider an intermediate random variables $\tilde{X}$, with mean $\tilde{\mu}$, obtain a confidence interval $[\tilde{L}, \tilde{U}]$ for $\tilde{\mu}$ and then define the restricted CI for μ_1 as follows:

$$\hat{L}_1 = \min(\bar{X}_1 - t_{n_1-1}\hat{s}_1/\sqrt{n_1}, \ \tilde{L}), \ \ \hat{U}_1 = \min(\bar{X}_1 + t_{n_1-1}\hat{s}_1/\sqrt{n_1}, \ \tilde{U}). \tag{4.12}$$

In this case, it is not possible to find an intermediate random variable $\tilde{X} = \gamma\bar{X}_1 + (1-\gamma)\bar{X}_2$ with exactly appropriate properties. Even for a fixed γ, the distribution of $\tilde{X}$ does not have a simple form, and obtaining the distribution of $\tilde{X}$ for a constant γ is in fact a variant of the Behrens-Fisher problem. Instead, we propose three methods of defining $\tilde{X}$ and approximating its distribution.

Method 1: For $\tilde{X} = \gamma\bar{X}_1+(1-\gamma)\bar{X}_2$, var$(\tilde{X}) = \gamma^2\sigma_1^2/n_1+(1-\gamma)^2\sigma_2^2/n_2$ is minimized at $\gamma = n_1\sigma_2^2/(n_1\sigma_2^2 + n_2\sigma_1^2)$, giving $\tilde{X} = (n_1\sigma_2^2\bar{X}_1 + n_2\sigma_1^2\bar{X}_2)/(n_1\sigma_1^2 + n_2\sigma_1^2)$ and the minimum variance $\sigma_1^2\sigma_2^2/(n_2\sigma_1^2 + n_1\sigma_2^2)$. However, we do not know σ_1^2 and σ_2^2, so we use unbiased estimator $\hat{s}_1^2$ and $\hat{s}_2^2$ to approximate $\tilde{X}$, giving

$$\tilde{X} = (n_1\hat{s}_2^2\bar{X}_1 + n_2\hat{s}_1^2\bar{X}_2)/(n_1\hat{s}_2^2 + n_2\hat{s}_1^2).$$

The coverage rate for $\tilde{\mu}$ $(= \gamma\mu_1 + (1-\gamma)\mu_2)$ using variance estimate $\hat{s}_1^2\hat{s}_2^2/(n_2\hat{s}_1^2 + n_1\hat{s}_2^2)$ will be too low , because it does not incorporate the uncertainty in the estimation of σ_1^2 and σ_2^2. One approach to allow for this is to modify the estimated variance based on thresholds of t_v distributions. Since $P(\bar{X}_g - \mu_g > t_\nu\hat{s}_g/\sqrt{n_g}) = P[\bar{X}_g - \mu_g > \{\hat{s}_g t_\nu/(z\sqrt{n_g})\}z] = \alpha/2$, we propose to approximate the distribution of $\bar{X}_g - \mu_g$ with a $N(0, \hat{s}_g^2 t_\nu^2/(z^2 n_g))$ distribution. This gives exactly the same $1-\alpha$ confidence interval for $\mu_g, g = 1, 2$ as using a t distribution. Thus, we use $\hat{s}_g^2 t_v^2/z^2$ as the estimate of σ_g^2 to adjust for the uncertainty of $\hat{s}_g^2$ and propose the following variance estimate of $\tilde{\sigma}^2$:

$$\tilde{\sigma}^2 = \frac{t_{n_1-1}^2 t_{n_2-1}^2 \hat{s}_1^2\hat{s}_2^2}{n_2 t_{n_1-1}^2\hat{s}_1^2 + n_1 t_{n_2-1}^2\hat{s}_2^2} \times \frac{1}{z^2}. \tag{4.13}$$

The approximate confidence interval for $\tilde{\mu}$ is then $\tilde{X} \pm z\tilde{\sigma}$.

Method 2: Since $(\bar{X}_g - \mu_g)\sqrt{n_g}/\hat{s}_g \sim T_g$, conditional on $\bar{X}_g$ and $\hat{s}_g^2$, $\mu_g \sim_f \bar{X}_g + (\hat{s}_g/\sqrt{n_g})T_g$, $g = 1, 2$, where T_g is a standard T random variable with degree of freedom $n_g - 1$ and $\sim_f$ represents the fiducial distribution, which is equivalent to a Bayesian posterior distribution under the usual noninformative priors. The variance of $\tilde{\mu}$ is minimized at $\hat{\gamma} = n_1^* \hat{s}_2^2/(n_2^* \hat{s}_1^2 + n_1^* \hat{s}_2^2)$, where $n_g^* = n_g(n_g - 3)/(n_g - 1), g = 1, 2$. This suggests defining

$$\tilde{X} = (n_1^* \hat{s}_2^2 \bar{X}_1 + n_2^* \hat{s}_1^2 \bar{X}_2)/(n_1^* \hat{s}_2^2 + n_2^* \hat{s}_1^2). \tag{4.14}$$

We still suggest using the variance estimate in (4.13), giving the CI

$$\tilde{L} = \tilde{X} - z\tilde{\sigma}, \;\; \tilde{U} = \tilde{X} + z\tilde{\sigma}. \tag{4.15}$$

The use of $\tilde{\sigma}^2$ is desirable because the CI for μ_1 derived from $(\tilde{X} - z\tilde{\sigma}, \tilde{X} + z\tilde{\sigma})$ using $\tilde{\sigma}^2$ from (4.13) always gives smaller or at least equal length interval compared to the unrestricted interval $(\bar{X}_1 - t_{n_1-1}\hat{s}_1, \bar{X}_1 + t_{n_1-1}\hat{s}_1)$, whereas this does not hold if we replace n_g by n_g^* in (4.13).

Another way to calculate $\tilde{L}$ and $\tilde{U}$ is to use the exact fiducial distribution of $\tilde{\mu}$, which is defined as

$$\tilde{\mu} \sim \tilde{X} + \hat{\gamma}(\hat{s}_1/\sqrt{n_1})T_1 + (1 - \hat{\gamma})(\hat{s}_2/\sqrt{n_2})T_2, \tag{4.16}$$

and then numerically calculate the percentiles of this distribution.

Simulations show that the restricted CI using the exact fiducial distribution of $\tilde{\mu}$ in (4.16) gives similar results to the CI defined by (4.15) using $\tilde{\sigma}^2$ when comparing average width and coverage rate.

Method 3: Define $\tilde{X}$ as the maximum likelihood estimator subject to the constraint $\mu_1 = \mu_2$. The log likelihood is

$$\log L = -\frac{n_1 + n_2}{2}\log(2\pi) - \frac{n_1}{2}\log\sigma_1^2 - \frac{n_2}{2}\log\sigma_2^2 - \frac{\sum_{i=1}^{n_1}(X_{1i} - \mu)^2}{2\sigma_1^2} - \frac{\sum_{i=1}^{n_2}(X_{2i} - \mu)^2}{2\sigma_2^2}.$$

The log profile likelihood is given by

$$\log pl\{\mu, \hat{\sigma}_1^2(\mu), \hat{\sigma}_2^2(\mu)\} = -\frac{n_1 + n_2}{2}\{\log(2\pi) + 1\} - \frac{n_1}{2}\log\hat{\sigma}_1^2(\mu) - \frac{n_2}{2}\log\hat{\sigma}_2^2(\mu),$$

where $\hat{\sigma}_g^2(\mu) = \sum_{i=1}^{n_g}(X_{gi} - \mu)^2/n_g, g = 1, 2$. Then $\tilde{X}$ will be the solution of the following equation

$$\frac{n_1^2(\bar{X}_1 - \mu)}{\sum(X_{1i} - \bar{X}_1)^2 + n_1(\bar{X}_1 - \mu)^2} + \frac{n_2^2(\bar{X}_2 - \mu)}{\sum(X_{2i} - \bar{X}_2)^2 + n_2(\bar{X}_2 - \mu)^2} = 0,$$

and the CI for $\tilde{\mu}$ that we propose is $(\tilde{X} - z\tilde{\sigma}, \tilde{X} + z\tilde{\sigma})$, where $\tilde{\sigma}^2$ is given by (4.13).

4.4 Confidence Intervals with Three or more Groups

Suppose $X_{gi} \sim N(\mu_g, \sigma_g^2), i = 1, \ldots, n_g, g = 1, \ldots, G$ and assume that $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_G$. Let $\bar{X}_g = \sum_{i=1}^{n_g} X_{gi}/n_g, g = 1, \ldots, G$.

4.4.1 Three-sample Case with Known Variances

We first consider joining neighboring groups or sets of neighboring groups. Let $\bar{X}_{12} = (\bar{X}_1 n_1 \sigma_1^{-2} + \bar{X}_2 n_2 \sigma_2^{-2})/(n_1\sigma_1^{-2} + n_2\sigma_2^{-2})$, $\bar{X}_{23} = (\bar{X}_2 n_2 \sigma_2^{-2} + \bar{X}_3 n_3 \sigma_3^{-2})/(n_2\sigma_2^{-2} + n_3\sigma_3^{-2})$ and $\bar{X}_{123} = \sum_{g=1}^{3}(\bar{X}_g n_g \sigma_g^{-2})/\sum_{g=1}^{3}(n_g\sigma_g^{-2})$ with $\bar{\mu}_{12}, \bar{\mu}_{23}$ and $\bar{\mu}_{123}$ denoting their means. Let $\bar{\sigma}_{12}^2 = 1/(n_1\sigma_1^{-2} + n_2\sigma_2^{-2})$, $\bar{\sigma}_{23}^2 = 1/(n_2\sigma_2^{-2} + n_3\sigma_3^{-2})$ and $\bar{\sigma}_{123}^2 = 1/\sum_{g=1}^{3}(n_g\sigma_g^{-2})$.

The unrestricted CIs for μ_g are $\bar{X}_g \pm z\sigma_g/\sqrt{n_g}$, $g = 1, 2, 3$, where L_g and U_g denote the lower and upper limits of these CIs. The unrestricted CIs for $\bar{\mu}_{12}$, $\bar{\mu}_{23}$ and $\bar{\mu}_{123}$ are $\bar{X}_{12} \pm z\bar{\sigma}_{12}$, $\bar{X}_{23} \pm z\bar{\sigma}_{23}$, $\bar{X}_{123} \pm z\bar{\sigma}_{123}$ and let $\bar{L}_{12}$, $\bar{U}_{12}$, $\bar{L}_{23}$, $\bar{U}_{23}$, $\bar{L}_{123}$, and $\bar{U}_{123}$ denote these lower and upper limits.

We reduce the problem to one of comparing two groups. For group 1, we can construct the confidence interval for μ_1 based on the comparison of group 1 and group 2. Thus the confidence interval for μ_1 would be

$$\hat{L}_1 = \min(L_1, \bar{L}_{12}), \quad \hat{U}_1 = \min(U_1, \bar{U}_{12}). \tag{4.17}$$

An alternative that may be worth considering in some circumstances would be to base the confidence interval of the comparison of group 1 and the combined group 2 and group 3. If it is believed that μ_2 and μ_3 are close to each other, then this may be a good choice. Then the restricted confidence interval for μ_1 would be

$$\hat{L}_1 = \min(L_1, \bar{L}_{123}), \ \ \hat{U}_1 = \min(U_1, \bar{U}_{123}), \tag{4.18}$$

The confidence interval for μ_3 is based on the comparison of groups 2 and 3, and is given by

$$\hat{L}_3 = \min(L_3, \bar{L}_{23}), \ \ \hat{U}_3 = \min(U_3, \bar{U}_{23}), \tag{4.19}$$

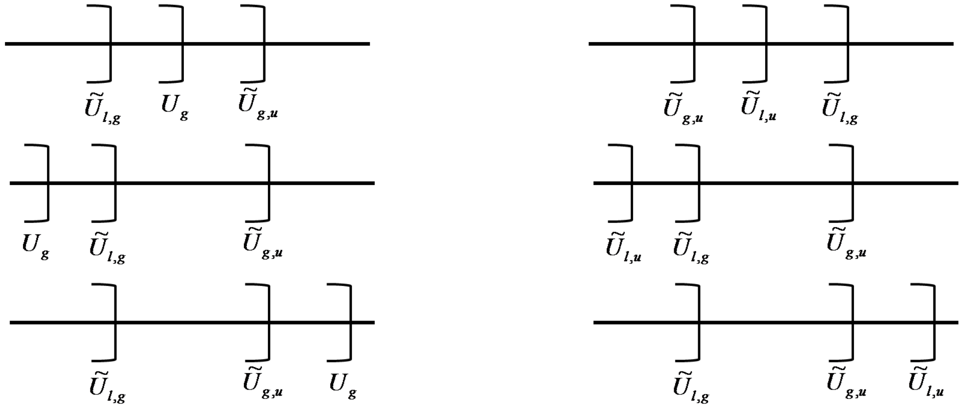
Again an alternative that might be worth considering in some circumatances, if it is believed that μ_1 and μ_2 are close to each other, is to combine groups 1 and 2 to give the restricted confidence interval for μ_3 as

$$\hat{L}_3 = \min(L_3, \bar{L}_{123}), \ \ \hat{U}_3 = \min(U_3, \bar{U}_{123}), \tag{4.20}$$

In practice, we recommend the default of not combining other groups and using equations (4.17) and (4.19), while recognizing that there may be advantages for using equations (4.18) and (4.20) instead. The decision of whether to combine groups could be based on either prior knowledge, or potentially a pre-test could be performed from the available data. We will briefly discuss a possible form for such a pre-test in the discussion.

Now we consider the confidence interval for μ_2. This includes two two-sample problems. For the upper bound of the confidence interval, if $\bar{U}_{23} \geq \bar{U}_{12}$ then $\hat{U}_2 = \min\{\bar{U}_{23}, \max(U_2, \bar{U}_{12})\}$ and $\hat{U}_2 = \max\{\bar{U}_{12}, \min(U_2, \bar{U}_{23})\}$ are both possible upper bounds, however, both are equal and equal to median$\{U_2, \bar{U}_{12}, \bar{U}_{23}\}$ (see Figure 4.4a). If $\bar{U}_{23} < \bar{U}_{12}$, it is not straightforward how to pick a value for $\hat{U}_2$. In the two-sample case for group 1 and 2, $\hat{U}_2 = \max(U_2, \bar{U}_{12})$ implies that $\hat{U}_2 \geq \bar{U}_{12}$, while in the two-sample case for group 2 and 3, $\hat{U}_2 = \min(U_2, \bar{U}_{23})$ implies that $\hat{U}_2 \leq \bar{U}_{23}$. Since $\bar{U}_{23} < \bar{U}_{12}$, a good value for $\hat{U}_2$ should be between $\bar{U}_{23}$ and $\bar{U}_{12}$. The true means are ordered as $\bar{\mu}_{12} \leq \bar{\mu}_{123} \leq \bar{\mu}_{23}$. This suggests one possible choice for $\hat{U}_g$ will be $\bar{U}_{123}$. However, $\bar{U}_{123}$ may not be between $\bar{U}_{23}$ and $\bar{U}_{12}$, therefore we propose $\hat{U}_2 = \text{median}(\bar{U}_{12}, \bar{U}_{123}, \bar{U}_{23})$ (see Figure 4.4b), and note that $\hat{U}_2 = \bar{U}_{123}$ in most, but not all cases. Thus the proposed restricted confidence interval for μ_2 is

$$
\begin{aligned}
\hat{U}_2 &= \begin{cases} \text{median}(U_2, \bar{U}_{12}, \bar{U}_{23}) & \text{if } \bar{U}_{23} \geq \bar{U}_{12} \\ \text{median}(\bar{U}_{12}, \bar{U}_{123}, \bar{U}_{23}) & \text{otherwise.} \end{cases} \\
\hat{L}_2 &= \begin{cases} \text{median}(L_2, \bar{L}_{12}, \bar{L}_{23}) & \text{if } \bar{L}_{23} \geq \bar{L}_{12} \\ \text{median}(\bar{L}_{12}, \bar{L}_{123}, \bar{L}_{23}) & \text{otherwise.} \end{cases}
\end{aligned}
\tag{4.21}
$$



(a) $\bar{U}_{12} \leq \bar{U}_{23}$ (b) $\bar{U}_{12} > \bar{U}_{23}$

Figure 4.4: Upper limits of the confidence interval in three or more sample case.

4.4.2 Three-sample Case with Unknown Variances

In this case, the restricted confidence intervals can also be defined using (4.17)-(4.21). Here, we discuss how to define the limits of unrestricted confidence intervals L_g, U_g, g=1,2,

$\bar{L}_{12}$, $\bar{U}_{12}$, $\bar{L}_{123}$, $\bar{U}_{123}$, $\bar{L}_{23}$ and $\bar{U}_{23}$.

If $w_g = \sigma^2/\sigma_g^2, g = 1, 2, 3$ is known, for some unknown σ^2, let $\hat{s}^2 = \sum_{g=1}^{3}\sum_{i=1}^{n_g} w_g(X_{gi} - \bar{X}_g)^2/(\sum_{g=1}^{3} n_g - 3)$, then the unrestricted confidence interval for μ_g is $L_g, U_g = \bar{X}_g \pm t_\nu \hat{s}(n_g w_g)^{-1/2}$, where $\nu = n_1+n_2+n_3-3$. Let $\bar{X}_{12} = (\bar{X}_1 n_1 w_1 + \bar{X}_2 n_2 w_2)/(n_1 w_1 + n_2 w_2)$, $\bar{X}_{12} = (\bar{X}_2 n_2 w_2 + \bar{X}_3 n_3 w_3)/(n_2 w_2 + n_3 w_3)$ and $\bar{X}_{123} = \sum_{g=1}^{3}(\bar{X}_g n_g w_g)/\sum_{g=1}^{3}(n_g w_g)$, then the unrestricted confidence intervals for $\bar{\mu}_{12}, \bar{\mu}_{123}$ and $\bar{\mu}_{23}$ will be $(\bar{L}_{12}, \bar{U}_{12}) = \bar{X}_{12} \pm t_\nu \hat{s}(n_1 w_1 + n_2 w_2)^{-1/2}$, $(\bar{L}_{123}, \bar{U}_{123}) = \bar{X}_{123} \pm t_\nu \hat{s}(n_1 w_1 + n_2 w_2 + n_3 w_3)^{-1/2}$ and $(\bar{L}_{23}, \bar{U}_{23}) = \bar{X}_{23} \pm t_\nu \hat{s}(n_2 w_2 + n_3 w_3)^{-1/2}$.

If $\sigma_g^2, g = 1, 2, 3$ needs to be estimated separately, let $\hat{s}_g = \sum_{i=1}^{n_g}(X_{gi} - \bar{X}_g)^2/(n_g - 1)$, then the unrestricted confidence interval for μ_g is $(L_g, U_g) = \bar{X}_g \pm t_{n_g - 1}\hat{s}_g/\sqrt{n_g}$. In this situation, we use method 2 in section 4.3.2 to obtain the means and confidence intervals for the combined groups. Let $n_g^* = n_g(n_g - 3)/(n_g - 1), g = 1, 2, 3$, then the mean estimates are $\bar{X}_{12} = \left(\sum_{g=1}^{2} n_g^* \hat{s}_g^{-2} \bar{X}_g\right) / \left(\sum_{g=1}^{2} n_g^* \hat{s}_g^{-2}\right)^{-1}$, $\bar{X}_{123} = \left(\sum_{g=1}^{3} n_g^* \hat{s}_g^{-2} \bar{X}_g\right) / \left(\sum_{g=1}^{3} n_g^* \hat{s}_g^{-2}\right)^{-1}$ and $\bar{X}_{23} = \left(\sum_{g=2}^{3} n_g^* \hat{s}_g^{-2} \bar{X}_g\right) / \left(\sum_{g=2}^{3} n_g^* \hat{s}_g^{-2}\right)^{-1}$. The variance estimates for these means are $\bar{\sigma}_{12} = \left(\sum_{g=1}^{2} n_g \hat{s}_g^{-2} t_{n_g-1}^{-2} z^2\right)^{-1}$, $\bar{\sigma}_{23} = \left(\sum_{g=2}^{3} 2 n_g \hat{s}_g^{-2} t_{n_g-1}^{-2} z^2\right)^{-1}$ and $\bar{\sigma}_{123} = \left(\sum_{g=1}^{3} n_g \hat{s}_g^{-2} t_{n_g-1}^{-2} z^2\right)^{-1}$. The unrestricted confidence intervals for these means are defined as $(\bar{L}_{12}, \bar{U}_{12}) = \bar{X}_{12} \pm z\bar{\sigma}_{12}$, $(\bar{L}_{123}, \bar{U}_{123}) = \bar{X}_{123} \pm z\bar{\sigma}_{123}$ and $(\bar{L}_{23}, \bar{U}_{23}) = \bar{X}_{23} \pm z\bar{\sigma}_{23}$.

4.4.3 Confidence Intervals for more than Three Groups

The CI for μ_1 can be constructed just by considering this as a 2 sample problem with groups 1 and 2. An alternative, that may lead to a gain in efficiency, is to consider combining other groups. However, there are many possible two-sample problems that could to be used to construct the restricted CI. The key here would be to decide which set of $\bar{X}_g$'s is used to form a two-sample problem with $\bar{X}_1$. If it is believed that $\mu_2, \ldots, \mu_u$ are close to each other, the restricted CI for μ_1 will depend on the limits of the CI for $\bar{\mu}_{1,u}$, which is the mean of group 1 to u, specifically a convex combination of μ_1 to μ_u with the weight proportional to the inverse of the variance. Similarly, if it is believed that $\mu_\ell, \ldots, \mu_{G-1}$ are close to each other, the restricted CI for μ_G will depend on the limits of the CI for $\bar{\mu}_{\ell,G-1}$.

The problem of constructing the CI for $\mu_g, 1 < g < G$ can be treated as a three-sample problem with groups $g-1$, g and $g+1$. Again, it may be possible to gain some efficiency by combining other groups. If it is believed that $\mu_\ell, \ldots, \mu_{g-1}$ are close to each other and $\mu_{g+1}, \ldots, \mu_u$ are close to each other, groups ℓ to $g-1$ can be considered as one group with a smaller mean and groups $g+1$ to u as the other group with a larger mean. Then the method for the middle group of the three-sample case can be used to construct a restricted CI for μ_g.

4.5 Other Restricted Confidence Intervals

There are a number of other possible approaches for constructing restricted confidence intervals, including bootstrap based confidence intervals and constant length confidence intervals (Hwang and Peddada, 1994).

Two sampling schemes for the bootstrap based on pivotal distributions are considered. The first scheme is based on the unrestricted MLE, in which $X_{gi}^b, b = 1, \ldots, B, i = 1, \ldots, n_g, g = 1, \ldots, G$ is sampled from $N(\bar{X}_g, \sigma_g^2)$ if σ_g^2 is known, X_{gi}^b is sampled from $\bar{X}_g + \hat{s}\, T_g / \sqrt{w_g}$, where T_g is standard T random variable with degree of freedom $\sum_{g=1}^{G}(n_g - 1)$ if w_g is known, or X_{gi}^b is sampled from $\bar{X}_g + \hat{s}_g T_g$, where T_g is standard T random variable with degree of freedom $n_g - 1$ if σ_g^2 is estimated. The second scheme is based on the restricted MLE $\hat{\mu}_g$, in which X_{gi}^b is sampled in the three different ways as described above except that the mean is $\hat{\mu}_g$ instead of $\bar{X}_g$, where $\hat{\mu}_g$ is the restricted MLE of μ_g. For each bootstrap sample, a bootstrap estimate $\hat{\mu}_g^b, b = 1, \ldots, B$ is obtained by applying the restricted maximum likelihood method. CIs are based on the percentiles of the bootstrap distribution of $\hat{\mu}_g^b$.

Hwang and Peddada (1994) proposed a constant length confidence interval in which the center of the CI is shifted from the unrestricted MLE to the restricted MLE. They showed that, under fairly general conditions, the coverage probability of the CI centered on the restricted MLE will exceed the nominal level. In our setting, the constant length confidence interval is defined as $\hat{\mu}_g \pm z\sigma_g/\sqrt{n_g}$ if σ_g^2 is known or as $\hat{\mu}_g \pm t_v \hat{s}_g/\sqrt{n_g}$ if σ_g^2 is unknown, where $v = \sum_{g=1}^{G}(n_g - 1)$ for known w_g or $v = n_g - 1$ for the case when σ_g^2s are estimated separately.

4.6 Simulation Study

We have undertaken numerous simulation studies for the two-sample case. We considered many different scenarios by varying $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, n_1$ and n_2. We found that the proposed method gives excellent coverage rates close to the nominal level even in small sample sizes and that the widths are narrower than those of unrestricted intervals and can be substantially narrower. The three methods in section 4.3.2 give similar results with very slightly better properties for methods 2. In this paper, we present results only for the more interesting and challenging three-sample case.

Let the population means of the three groups be ordered as $\mu_1 \leq \mu_2 \leq \mu_3$. Coverage probabilities and the average width of CIs are calculated based on 5,000 simulated datasets and each bootstrap confidence interval is based on 999 bootstrap estimates. The distributions and sample sizes for the simulations are listed in Table 4.2. We included in the comparison the CI based on the unrestricted estimates (Unres), the shifted constant length CI (Shifted(Const)), the bootstrap methods and the new method of section 4.4. The results using equations (4.18) and (4.20) in which we assume all groups are close (New(combined)) and using equations (4.17) and (4.19) in which we assume all groups are separated (New(separate)) are shown. The three parametric bootstrap methods are, completely unrestricted (Bootstrap(Unres)), the method where the bootstrap samples are simulated from a distribution centered at $\bar{X}_g$ and the restricted MLE is estimated (Bootstrap(RMLE), and the method where the bootstrap samples are simulated from a distribution centered at $\hat{\mu}_g$ and the restricted MLE is estimated (Bootstrap-R(RMLE)).

Table 4.2: Different combinations of population means, variances and sample sizes used in simulation studies.

	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}^2$	n
(a)	-0.1, 0, 0.1	5, 5, 10	5, 5, 10
(b)	-1.0, 0, 1.0	5, 5, 10	5, 5 ,10
(c)	-0.1, 0, 0.1	5, 5, 5	5, 5, 50
(d)	-0.1, 0, 0.1	10, 10, 10	5, 50, 10
(e)	-0.1, 0, 0.1	5, 50, 10	10, 10, 10
(f)	-0.1, 0, 2.0	10, 10, 10	5, 50, 10
(g)	-0.1, 0, 2.0	10, 10, 10	5, 10, 50

We present the results for coverage rates and average confidence interval widths in Table 4.3 for σ_g^2 known, in Table 4.4 for known ratios of variances and in Table 4.5 for the case where all variances are estimated.

As expected, the shifted constant length confidence interval centered on the restricted MLE has higher coverage probability than the nominal level. However, the coverage rate tends to be extremely high for μ_2 when all three population means are close to one another (cases (a), (c), (e) in Table 3, 4 and 5). Even though the bootstrap method works well when all the population means are well separated(case (b)), the coverage rates for some popula-

Table 4.3: Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the variances are known.

Method		Unres	Shifted Const	Bootstrap Unres	Bootstrap RMLE	Bootstrap-R RMLE	New separate	New combined
(a)	μ_1	94.9(3.92)	97.3(3.92)	94.8(3.90)	92.0(2.99)	91.3(3.16)	95.3(3.37)	95.7(3.16)
	μ_2	95.2(3.92)	99.7(3.92)	95.1(3.90)	96.7(2.68)	98.3(2.81)	97.2(2.84)	97.2(2.84)
	μ_3	94.9(3.92)	97.2(3.92)	94.7(3.90)	91.9(3.00)	91.4(3.16)	95.1(3.38)	95.5(3.16)
(b)	μ_1	95.3(3.92)	97.2(3.92)	95.2(3.90)	95.1(3.37)	95.9(3.42)	95.9(3.61)	95.6(3.67)
	μ_2	94.9(3.92)	98.4(3.92)	94.5(3.90)	95.7(3.03)	97.4(3.10)	96.4(3.30)	96.4(3.30)
	μ_3	94.8(3.92)	97.1(3.92)	94.8(3.91)	95.2(3.38)	96.2(3.43)	95.6(3.61)	95.0(3.67)
(c)	μ_1	95.2(3.92)	97.5(3.92)	95.0(3.90)	91.4(2.57)	90.3(2.82)	95.3(3.38)	96.7(2.67)
	μ_2	94.7(3.92)	99.7(3.92)	94.7(3.90)	96.2(1.87)	98.0(2.10)	97.5(2.06)	97.5(2.06)
	μ_3	95.2(1.24)	95.8(1.24)	95.0(1.23)	94.0(1.17)	94.2(1.18)	95.2(1.21)	95.0(1.19)
(d)	μ_1	95.0(5.54)	97.5(5.54)	94.8(5.52)	90.3(3.43)	89.5(3.85)	95.9(3.69)	95.8(3.63)
	μ_2	94.5(1.75)	96.8(1.75)	94.4(1.74)	94.5(1.61)	95.2(1.63)	95.2(1.64)	95.2(1.64)
	μ_3	95.3(3.92)	97.8(3.92)	94.9(3.90)	91.5(2.62)	91.4(2.88)	96.0(2.81)	96.2(2.79)
(e)	μ_1	95.5(2.77)	96.7(2.77)	95.3(2.76)	93.8(2.46)	93.6(2.51)	95.4(2.71)	95.4(2.52)
	μ_2	95.9(8.77)	100 (8.77)	95.7(8.73)	97.2(2.96)	98.8(3.25)	97.9(3.30)	97.9(3.30)
	μ_3	94.8(3.92)	97.0(3.92)	94.9(3.90)	91.8(3.05)	91.6(3.22)	94.8(3.75)	95.2(3.15)
(f)	μ_1	95.0(5.54)	97.3(5.54)	94.9(5.52)	91.6(3.54)	91.5(3.94)	95.3(3.72)	96.2(3.91)
	μ_2	94.5(1.75)	95.1(1.75)	94.3(1.75)	94.1(1.68)	94.3(1.68)	94.6(1.71)	94.6(1.71)
	μ_3	94.9(3.92)	95.9(3.92)	94.9(3.90)	94.8(3.57)	95.8(3.58)	94.9(3.75)	94.9(3.75)
(g)	μ_1	94.8(5.54)	97.3(5.54)	94.6(5.52)	93.3(4.13)	93.2(4.40)	95.4(4.41)	94.8(4.79)
	μ_2	94.7(3.92)	98.0(3.92)	94.7(3.90)	94.0(3.17)	95.0(3.22)	95.1(3.40)	95.1(3.40)
	μ_3	95.2(1.75)	95.3(1.75)	95.1(1.74)	95.1(1.72)	95.2(1.72)	95.2(1.75)	95.2(1.75)

tion means can be well below the nominal level in some situations (μ_1 for cases (a), (c), (d) and (f)). There are no noticeable improvements from using the bootstrap method with sampling centered on the restricted MLE compared to the bootstrap method with sampling centered on the unrestricted MLE. The method proposed in this paper gives fairly accurate coverage rate with reduced width of the intervals in all the situations considered in this simulation study and is the recommended method. Although the properties of confidence intervals are improved by using only adjacent group's information (NEW separate), the efficiency can be improved by combining the groups with similar means as seen in scenarios (a), (c) and (e) (NEW(combined) vs NEW(separate), comparing width of confidence

Table 4.4: Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the ratios of variances are known.

Method		Unres	Shifted Const	Bootstrap Unres	Bootstrap RMLE	Bootstrap-R RMLE	New separate	New combined
(a)	μ_1	95.0(4.16)	97.2(4.16)	94.8(4.14)	91.8(3.16)	91.2(3.33)	95.4(3.58)	95.7(3.36)
	μ_2	95.3(4.16)	99.3(4.16)	95.0(4.14)	96.1(2.81)	97.7(2.94)	97.0(3.02)	97.0(3.02)
	μ_3	94.7(4.16)	97.0(4.16)	94.4(4.14)	91.8(3.16)	91.5(3.33)	95.1(3.58)	95.3(3.35)
(b)	μ_1	95.1(4.16)	96.6(4.16)	94.9(4.14)	94.9(3.55)	95.3(3.61)	95.6(3.83)	95.3(3.88)
	μ_2	94.7(4.16)	98.2(4.16)	94.9(4.14)	95.9(3.18)	97.3(3.25)	95.9(3.50)	95.9(3.50)
	μ_3	94.9(4.16)	96.8(4.16)	94.6(4.14)	95.0(3.56)	96.0(3.61)	95.5(3.83)	95.1(3.88)
(c)	μ_1	95.3(3.99)	97.4(3.99)	95.1(3.97)	91.2(2.61)	90.2(2.86)	95.2(3.43)	96.6(2.72)
	μ_2	95.0(3.99)	99.7(3.99)	94.9(3.97)	96.3(1.90)	98.0(2.12)	97.5(2.09)	97.5(2.09)
	μ_3	95.2(1.26)	95.9(1.26)	94.9(1.25)	94.2(1.19)	94.1(1.20)	95.2(1.23)	95.2(1.21)
(d)	μ_1	95.0(5.64)	97.6(5.64)	94.8(5.62)	90.8(3.48)	89.9(3.91)	96.0(3.76)	95.9(3.70)
	μ_2	95.0(1.78)	96.5(1.78)	94.7(1.78)	94.9(1.64)	95.4(1.66)	95.4(1.67)	95.4(1.67)
	μ_3	95.1(3.99)	97.6(3.99)	94.8(3.97)	91.7(2.66)	91.6(2.92)	95.7(2.86)	95.7(2.84)
(e)	μ_1	95.3(2.87)	96.5(2.87)	95.2(2.86)	93.6(2.54)	93.5(2.60)	95.5(2.81)	95.2(2.61)
	μ_2	95.6(9.08)	100 (9.08)	95.5(9.04)	96.9(3.03)	98.8(3.31)	97.6(3.41)	97.6(3.41)
	μ_3	95.1(4.06)	96.9(4.06)	95.1(4.04)	92.2(3.15)	92.1(3.31)	95.2(3.89)	95.5(3.26)
(f)	μ_1	94.7(5.63)	97.0(5.63)	94.7(5.60)	91.5(3.58)	91.6(3.98)	95.1(3.77)	96.0(3.97)
	μ_2	94.6(1.78)	95.3(1.78)	94.3(1.77)	94.1(1.70)	94.3(1.71)	94.7(1.74)	94.7(1.74)
	μ_3	94.9(3.98)	95.7(3.98)	94.7(3.96)	94.7(3.62)	95.5(3.62)	94.9(3.81)	94.9(3.80)
(g)	μ_1	94.9(5.64)	97.4(5.64)	94.7(5.61)	93.4(4.19)	93.1(4.46)	95.5(4.49)	95.0(4.86)
	μ_2	94.6(3.98)	97.6(3.98)	94.6(3.97)	94.1(3.21)	95.2(3.25)	94.8(3.45)	94.8(3.45)
	μ_3	95.3(1.78)	95.4(1.78)	95.4(1.77)	95.4(1.74)	95.0(1.74)	95.3(1.78)	95.3(1.78)

interval for μ_1 or μ_3). However, the effect of combining groups that are well separated can lead to wider intervals, as seen for μ_1 in cases (f) and (g). In practice, if we believe that all the population means are well separated as in case (b), the bootstrap method is a good alternative, and sometimes yield confidence intervals of smaller average width.

4.7 Example

The half-life of a drug is the time needed to halve the concentration of the drug in the body of a human or an animal. The half-life may vary with the concentration of the drug,

Table 4.5: Empirical coverage rates and average widths of 95% confidence intervals for μ_1, μ_2 and μ_3 when the variances are estimated separately.

Method		Unres	Shifted Const	Bootstrap Unres	Bootstrap RMLE	Bootstrap-R RMLE	New separate	New combined
(a)	μ_1	95.3(5.28)	96.6(5.28)	95.3(5.25)	91.3(3.89)	89.6(4.00)	95.0(4.65)	94.8(4.32)
	μ_2	95.5(5.17)	97.1(5.17)	95.5(5.15)	94.4(3.16)	93.8(3.19)	94.7(3.35)	94.7(3.35)
	μ_3	94.9(4.40)	96.6(4.40)	94.7(4.37)	90.3(3.38)	87.6(3.46)	95.4(4.07)	94.8(3.85)
(b)	μ_1	94.5(5.21)	95.4(5.21)	94.6(5.19)	94.5(4.27)	94.4(4.30)	95.1(4.75)	95.3(4.76)
	μ_2	94.6(5.22)	96.0(5.22)	94.5(5.18)	94.7(3.55)	95.1(3.58)	94.8(4.01)	94.8(4.01)
	μ_3	94.7(4.41)	96.3(4.41)	94.5(4.39)	94.0(3.73)	94.9(3.76)	96.1(4.16)	96.3(4.17)
(c)	μ_1	95.3(5.24)	97.4(5.24)	95.1(5.21)	90.0(3.31)	88.9(3.57)	95.8(4.66)	96.9(3.74)
	μ_2	94.8(5.21)	98.3(5.21)	94.5(5.18)	95.0(2.14)	96.0(2.30)	95.8(2.40)	95.8(2.4.0)
	μ_3	95.1(1.26)	95.3(1.26)	94.9(1.26)	93.6(1.21)	92.5(1.20)	94.9(1.26)	94.6(1.25)
(d)	μ_1	95.2(7.43)	97.4(7.43)	95.2(7.40)	91.1(4.39)	90.2(4.80)	96.2(5.15)	96.0(5.09)
	μ_2	94.9(1.79)	96.5(1.79)	94.6(1.78)	94.5(1.65)	95.0(1.65)	95.3(1.68)	95.3(1.68)
	μ_3	95.2(4.41)	97.8(4.41)	95.2(4.39)	92.4(2.89)	91.4(3.14)	96.5(3.40)	96.4(3.38)
(e)	μ_1	95.2(3.11)	96.0(3.11)	95.1(3.09)	93.8(2.72)	93.3(2.76)	95.7(3.07)	95.2(2.90)
	μ_2	95.1(9.82)	99.9(9.82)	94.9(9.77)	96.3(3.12)	97.6(3.33)	96.8(3.60)	96.8(3.60)
	μ_3	95.1(4.40)	96.7(4.40)	95.1(4.38)	91.7(3.33)	91.1(3.49)	96.0(4.29)	95.6(3.76)
(f)	μ_1	94.7(7.34)	97.3(7.34)	94.3(7.29)	92.3(4.44)	92.0(4.84)	95.7(5.08)	97.0(5.30)
	μ_2	94.6(1.79)	94.8(1.79)	94.5(1.78)	94.3(1.72)	93.7(1.71)	94.2(1.75)	94.2(1.75)
	μ_3	94.6(4.38)	95.2(4.38)	94.4(4.35)	94.4(3.88)	95.1(3.87)	95.2(4.12)	95.2(4.11)
(g)	μ_1	94.7(7.41)	96.1(7.41)	94.2(7.37)	92.2(5.18)	91.9(5.41)	94.6(5.89)	96.1(6.28)
	μ_2	94.4(4.39)	96.1(4.39)	94.2(4.37)	93.1(3.46)	92.5(3.40)	94.0(3.72)	94.0(3.72)
	μ_3	95.2(1.79)	95.4(1.79)	95.0(1.78)	95.1(1.75)	95.0(1.75)	95.4(1.79)	95.4(1.79)

and usually is longer for higher concentration levels. Table 4.6 shows data from Hirotsu (2005). It contains the half-lives in hours of an antibiotic at four different doses that is injected into rats. The higher dose level should result in a higher concentration and hence it is reasonable to assume the half life is shorter for the lower dose level.

The analysis is based on two scenarios. First, we assume all the observation are from normal distributions with means that depend on the doses but with the same variance. Second, we assume that the variances of different dose levels may not be equal. The results are shown in Table 4.7. It can be seen that widths of restricted CIs are narrower compared to their unrestricted counterparts. Both limits of the confidence interval for doses of 10 mg/kg and 50 mg/kg are modified. The most noticeable reduction in the width of the restricted confidence interval is for the dose of 50 mg/kg when we estimate the variances separately, where the width of the restricted confidence interval is about 35% of the unrestricted one. We also considered the impact of combining some neighboring groups as a way to obtain shorter intervals. Specifically we combined the 5 mg/kg and 10 mg/kg groups and the 25 mg/kg and 50 mg/kg groups. The results were quite similar to those in Table 4.7 and are not shown.

Table 4.6: Half-life of an antibiotic in rats.

Dose (mg/kg)	Data (h)					Average
5	1.17	1.12	1.07	0.98	1.04	1.08
10	1.00	1.21	1.24	1.14	1.34	1.19
25	1.55	1.63	1.49	1.53		1.55
50	1.21	1.63	1.37	1.50	1.81	1.50
200	1.78	1.93	1.80	2.07	1.70	1.86

Table 4.7: Estimates and confidence intervals of half-lives in the example in section 4.7.

Dose	Equal variances		Different variances	
(mg/kg)	MLE(CI)	RMLE (CI)	MLE (CI)	RMLE (CI)
5	1.08 (0.78, 1.38)	1.08 (0.78, 1.34)	1.08 (0.87, 1.28)	1.08 (0.87, 1.18)
10	1.19 (0.89, 1.49)	1.19 (1.04 ,1.45)	1.19 (0.84, 1.54)	1.19 (1.03, 1.53)
25	1.55 (1.25, 1.85)	1.52 (1.25, 1.62)	1.55 (1.36, 1.74)	1.55 (1.36, 1.63)
50	1.50 (1.20, 1.81)	1.52 (1.42, 1.78)	1.50 (0.86, 2.15)	1.55 (1.46, 1.91)
200	1.86 (1.56, 2.16)	1.86 (1.58, 2.16)	1.86 (1.45, 2.26)	1.86 (1.60, 2.26)

4.8 Discussion

In this paper, we developed a novel method for constructing confidence intervals under linear ordering constraints for normal population means. In the two sample case, we showed that the coverage rate is at least the nominal level when the variances or the ratio of the variances is known. Simulation studies show that the coverage rates are also close to the nominal level even when the variances of populations are estimated separately.

The methods developed in section 4.2, 4.3 and 4.4 are applicable to normal observations. On account of the central limit theorem, we expect the coverage rates of the restricted CIs for the means to be close to the nominal level in nonnormal populations if the sample size is fairly large. We found this to be empirically true in simulations (not shown), except when the distribution is very skewed and the sample size is relatively small. Even in this case, the coverage rates were not substantially below the nominal level, and showed much better coverage rates than the alternative bootstrap CI's.

The method proposed in this paper can be generalized to other distributions by using transformations. For example, if T is a monotone transformation and $T(\hat{\mu}_g)$ is approximately normally distributed, then it is possible to apply our method to estimate the CI of $T(\mu_g)$ and then apply T^{-1} to obtain the CI of μ_g. For example, for the binomial case with success probability μ_g, the variance stabilizing transformation $\sin^{-1}(\sqrt{\mu_g})$ could be used.

The strategy that we developed in this paper can be broadly described using two stages: in the first stage, obtain an unrestricted CI, (L_g, U_g) for μ_g using a normal or t distribution; in the second stage, modify these bounds based on the order restrictions, using for example, equations (4.17), (4.19) and (4.21). A modification of this method is to use the bootstrap in the first stages to obtain (L_g, U_g).

The methodology described in this paper is applicable in the linear ordering case. Hwang and Peddada (1994) discussed a constant length confidence intervals centered on the restricted estimator under more general ordering constraints. It will be interesting to develop a method using strategies similar to those in this paper for the more general ordering situation.

As discussed in Section 4.4 the method for three or more samples could potentially be made more efficient by combining groups rather than just considering the closest group. Whether it is beneficial to combine groups depends on the closeness of the means of neighboring groups. It could be useful to develop a method that can automatically decide which sets of means are close to each other and so are better to be combined when constructing the restricted confidence intervals. For example, one possible approach for the three-group situation would be to test: $H_0 : \mu_1 \leq \mu_2 = \mu_3$ vs $H_a : \mu_1 \leq \mu_2 < \mu_3$ at a certain significance level to decide whether μ_2 and μ_3 are close to each other, and so to decide whether or not to combine group 2 and group 3 to construct the restricted confidence intervals for μ_1.

CHAPTER V

Future Research

The method discussed in Chapter II and the concept of the MC-NPMLE can be extended to settings with more than two samples with linear ordering or partial ordering constraints. Hoff (2000, 2003b) proposed an estimation method via mixtures, in which the constrained measures are represented with unconstrained mixtures of simple and known extreme measures, probability measures over the points of the extreme set. The extreme set is the set of all extreme points, where an extreme point of a convex set is a point in the set that can not be written as a convex combinations of other points in the set. His method is appealing because it can be used in Bayesian analysis to achieve uncertainty estimation (Hoff, 2003a). However, his method focused on the case where the measures are distribution functions, and did not consider censored data. It would be interesting to develop a method using mixtures to obtain the C-NPMLE and the MC-NPMLE of survivor functions with right censored or interval censored data. Lim et al. (2009) also proposed a different method to obtain the C-NPMLE under partial ordering constraints using geometric programming. However, we believe that the extension of the methods in Chapter II will provide additional insights and have the potential to improve the computational efficiency through reducing the number of parameters to compute and the number of constraints to apply.

While the nature of my dissertation research has been quite mathematical, the area of order restricted estimation is broadly applicable. There are many situations where there is a strong rationale for monotonicity or order restrictions, and incorporating this information via a pointwise C-NPMLE is an appealing approach that does not require the use of strong parametric assumptions. It will also be interesting to find more cases and areas to apply our methods. For example, one possible application is estimating the probability of toxicity of each dose in a Phase I clinical trial. The toxicity at a fixed time is of interest, yet there may exist noninformative censoring such as early drop-out. In this situation, the traditional nonparametric method based on the binomial distribution does not apply and the pointwise C-NPMLE is a natural solution. Another possible area is a stratified Cox model when the stochastic ordering constraint to the baseline distribution function across different strata. The efficiency of the relative risk estimation might be improved by appropriately applying a pointwise C-NPMLE to obtain the baseline survivor functions for different strata.

Another possible future research area is the general area of Bayesian methodology in order restricted inference where the order restrictions can be incorporated through the prior distributions. Taylor et al. (2007), Marchand and Strawderman (2006), Zhang and Woodroofe (2003) and Roe and Woodroofe (2000) have studied Bayesian methods by applying truncated noninformative priors in the restricted parameter space. Dunson and Peddada (2008), Karabatsos and Walker (2007), Hoff (2003a), Gelfand and Kottas (2001), and Arjas and Gasbarra (1996) considered various Bayesian methods based on full nonparametric likelihoods to study stochastically ordered survivor functions. One feature of the Bayesian approach is that the posterior distributions are usually strictly ordered (i.e. $S_1(t) < S_2(t)$), however, here may be advantage to assign prior probability mass to the boundaries of the order-restricted parameter space (i.e. to $S_1(t) = S_2(t)$). It would be also of interest to develop a Bayesian method based on the pointwise likelihood to conduct hypothesis testings and construct confidence intervals pointwisely.

Further future research also includes proof or disproof of Conjecture 3.7, and developing a method that can automatically decide which sets of means are better to be combined when constructing the restricted confidence intervals as discussed in Chapter IV.

Another possible direction for the future research is to consider smoothing estimates as well as order restrictions. Specifically to consider restricted estimation of smooth survival curves under stochastic ordering constraint. Because smoothness is a common feature of most survival distributions, smooth estimators of survivor functions is preferable. In the past, numerous methods have been proposed to obtain smooth estimates of density or distribution functions, such as a piecewise exponential estimator (Kim and Proschan, 1991), Kernel estimator, penalized spline estimator, and spline-based estimators (see Simonoff (1996)). However, there has been little discussion of smooth estimation under stochastic ordering constraint in the literature; an exception was the work of Chaubey and Kochar (2000) who proposed a method of smooth estimation of constrained survivor functions in the one- and two-sample cases based on Hille's theorem analysis. Some possible approaches are (1) smoothing the discrete constrained estimators, such as the C-NPMLE or the pointwise C-NPMLE, by using Gamma Kernel (Bouezmarni and Rombouts, 2010; Chen, 2000, 2002), Beta Kernel (Chen, 1999) or Nonnegative Boundary Corrected Estimator (Jones, 1993; Jones and Foster, 1996); (2) restricting smoothed survivor functions, and (3) using a basis of spline functions such as monotone regression spline (Abrahamowicz et al., 1989; Ramsay, 1988) to obtain a smooth estimator using a semi-infinite programming method (see Weber (2003)).

In the Chapter IV, we constructed a method of obtaining confidence intervals that was applied to normal distributions. It will be worthwhile to generalize and investigate this approach for other distributions such binomial distribution and gamma distribution or families of distributions such as exponential family and location parameter family.

APPENDICES

APPENDIX A

Proof of Theorem 2.13 in Section 2.4

To fix notation, let $a_0 = 0, a_1, a_2, \cdots a_m$ be the complete ordered observed event times of any given data in the two sample case, $i_0 = 0, i_1, i_2, \cdots i_L$ the index of active constraint times, and $k_1, k_2, \cdots, k_L$ be the corresponding k values from Algorithm 2.9.

The last active constraint time from Algorithm 2.9 satisfies $a_{i_L} \leq \tau$. $\hat{S}_1(t)/S_1^*(t)$ is non-decreasing and $\hat{S}_2(t)/S_2^*(t)$ is non-increasing in t in any sample. At the last active constraint a_{i_L}, $S_1^*(a_{i_L}) \leq \hat{S}_1(a_{i_L}) = \hat{S}_2(a_{i_L}) \leq S_2^*(a_{i_L})$.

So for any $x \leq a_{i_L}$,

$$\begin{aligned} S_1^*(x) \leq \hat{S}_1(x) \leq S_1^*(x)\frac{\hat{S}_1(a_{i_L})}{S_1^*(a_{i_L})} &\leq S_1^*(x)\frac{S_2^*(a_{i_L})}{S_1^*(a_{i_L})} \\ &\leq \left[\sup_{t \leq a_{i_L}} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \leq \left[\sup_{t \leq \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \end{aligned}$$

Similarly,

$$S_2^*(x) \geq \hat{S}_2(x) \geq \left[\inf_{t \leq \tau} \frac{S_1^*(t)}{S_2^*(t)}\right] S_2^*(x).$$

For any $x > a_{i_L}$ in the same sample,

$$S_1^*(x) \le \hat{S}_1(x) = \hat{S}_1(a_{i_L}) \times \frac{S_1^*(x)}{S_1^*(a_{i_L})}$$
$$\le \left[\sup_{t \le \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(a_{i_L}) \times \frac{S_1^*(x)}{S_1^*(a_{i_L})} = \left[\sup_{t \le \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x).$$

So, regardless of where a_{i_L} is, in any sample, for any $x \le \tau$, we always have

$$S_1^*(x) \le \hat{S}_1(x) \ge \left[\sup_{t \le \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \text{ and } S_2^*(x) \ge \hat{S}_2(x) \ge \left[\inf_{t \le \tau} \frac{S_1^*(t)}{S_2^*(t)}\right] S_2^*(x)$$

As n_1, n_2 go to ∞, from (2.4), for any $x \le \tau$, $S_1^*(x) \to S_1(x)$ and $S_2^*(x) \to S_2(x)$. So

$$\frac{S_2^*(x)}{S_1^*(x)} \to \frac{S_2(x)}{S_1(x)} \Rightarrow \sup_{x \le \tau} \frac{S_2^*(x)}{S_1^*(x)} \to \sup_{x \le \tau} \frac{S_2(x)}{S_1(x)} = 1,$$

in probability if indeed $S_1(t) \ge S_2(t)$ for all t. Thus $P\left\{\sup_{x \le \tau} |\hat{S}_1(x) - S_1^*(x)| > \epsilon\right\} \to 0$ for every $\epsilon > 0$. Then using Meier's result (2.4), we obtain the desired result for $\hat{S}_1(t)$.

Similarly, we can show that $P\left\{\sup_{x \le \tau} |\hat{S}_2(x) - S_2^*(x)| > \epsilon\right\} \to 0$ for every $\epsilon > 0$ and hence we obtain the desired result for $\hat{S}_2(t)$.

To simplify the proof, we only show consistency of the C-NPMLE in the case of iid data. However, it can be shown that the estimators are consistent in the more general situation as discussed in Dykstra (1982).

APPENDIX B

Proof for the Two-sample Case from Section 2.2

Recall that $m_g = \max\{i : n_{gi} > 0\}, g = 1, 2$ and $m' = \min(m_1, m_2)$. In this section, we prove that the $\hat{h}_{gi}, g = 1, 2, i = 1, \ldots, m_g$ in Theorem 2.8 and Algorithm 2.9 is the C-NPMLE. We use $\hat{h}^A_{gi}$ and $\hat{h}^T_{gi}$ to distinguish the results from Algorithm 2.9 and Theorem 2.8 before we prove that they are the same. In Lemma B.2, we show that $\hat{h}^A_{gi} = \hat{h}^T_{gi}$, and in Lemma B.3, we show that $\hat{h}_{1i}$ is the C-NPMLE and $\hat{h}_{2i}$ is the MC-NPMLE.

B.1 Characteristics of Results from Theorem 2.8 and Algorithm 2.9

Steps 2,3 and 4 in Algorithm 2.9 are used to calculate k_ℓ. Suppose the algorithm has R_ℓ iterations with initial step 2 to 3 (0^{th} iteration) and step 4 to 3 ($1^{st}, \ldots, R_\ell^{th}$ iteration) before it finds i_ℓ. Here, $i_\ell^{(r)}$ and $k_\ell^{(r)}$, $r = 0, \ldots, R_\ell$ are results from iteration r. Let $i_\ell^{(-1)} = i_{\ell-1}$ and $k_\ell^{(-1)} = 0$ for convenience. Note that $H_2(a, b, k)$ is a non-decreasing function in k and strictly increasing if there is at least one observed event between a and b.

Lemma B.1. *The result for any data from Algorithm 2.9 has the following properties:*

(a). $\sum_{j=1}^{i_\ell} (\hat{h}^A_{1j} - \hat{h}^A_{2j}) = 0, \ \ell = 1, \ldots, L;$

(b). $k_\ell = \min\{K_2(i_{\ell-1}+1, i_\ell), n_{2i_\ell}\}$ *and* $H_2(i_{\ell-1}+1, i_\ell, k_\ell) \le 0$*;*

(c). $k_\ell^{(r)} > k_\ell^{(r-1)},\ r = 1, \ldots, R_\ell$*;*

(d). $\sum_{j=1}^{i} \hat{h}_{1j}^{A} \ge \sum_{j=1}^{i} \hat{h}_{2j}^{A},\ i = 1, \ldots, m'$*;*

(e). $k_1 \ge k_2 \ge \cdots \ge k_L > 0$.

Proof. (a) It is obvious since $\sum_{j=i_{\ell-1}+1}^{i_\ell} \hat{h}_{2j}^{A} = \sum_{j=i_{\ell-1}+1}^{i_\ell} \hat{h}_{2j}^{A},\ \ell = 1, \ldots, L$.

(b) From step 3 in Algorithm 2.9, k_ℓ can take two values: $k_\ell = n_{2i_\ell}$ if $H_2(i_{\ell-1} + 1, i_\ell, n_{2i_\ell}) \le 0$ (in this case $K_2(i_{\ell-1}+1, i_\ell) \ge n_{2i_\ell}$), or $k_\ell = K_2(i_{\ell-1}+1, i_\ell)$.

(c) Show by contradiction. Suppose there exists r' such that $k_\ell^{(r')} \le k_\ell^{(r'-1)}$. It follows that

$$\begin{aligned} &H_2(i_{\ell-1}+1, i_\ell^{(r')}, k_\ell^{(r')}) \le H_2(i_{\ell-1}+1, i_\ell^{(r')}, k_\ell^{(r'-1)}) \\ &= H_2(i_{\ell-1}+1, i_\ell^{(r'-1)}, k_\ell^{(r'-1)}) + H_2(i_\ell^{(r'-1)}+1, i_\ell^{(r')}, k_\ell^{(r'-1)}) \\ &= H_2(i_\ell^{(r'-1)}+1, i_\ell^{(r')}, k_\ell^{(r'-1)}) < 0 \qquad \text{(step 2 or 4 in Algorithm 2.9).} \end{aligned} \tag{B.1}$$

However, from step 3 in Algorithm 2.9, $k_\ell^{(r')}$ must either satisfy:

$K_2(i_{\ell-1}, i_\ell^{(r')})$. Then $H_2(i_{\ell-1}+1, i_\ell^{(r')}, k_\ell^{(r')}) = 0$, which contradicts (B.1); or

$n_{2i_\ell^{(r')}}$. Then $k_\ell^{(r'-1)} \ge k_\ell^{(r')} = n_{2i_\ell^{(r')}}$, which contradicts the condition $n_{2i_\ell^{(r')}} > k_\ell^{(r'-1)}$ that is required to reach iteration r' in step 4 of Algorithm 2.9.

(d) Suppose there exists i' such that $\sum_{j=1}^{i'} \hat{h}_{1j}^{A} < \sum_{j=1}^{i'} \hat{h}_{2j}^{A}$, equivalently $\sum_{j=i_{\ell'-1}+1}^{i'} \hat{h}_{1j}^{A} < \sum_{j=i_{\ell'-1}+1}^{i'} \hat{h}_{2j}^{A}$. Then each of the three possible valid ranges of i' leads to a contradiction. That is either:

$i' = i_\ell$,$\ell \le L$. Then $\sum_{j=1}^{i'} \hat{h}_{1j}^{A} = \sum_{j=1}^{i'} \hat{h}_{2j}^{A}$, which contradicts Lemma B.1 (a); or

$i' > i_L$. Then $H_2(i_L+1, i', 0) = \sum_{j=i_L+1}^{i'} (\hat{h}_{1j} - \hat{h}_{2j}) < 0$, which contradicts the condition $H_2(i_L+1, b, 0) \ge 0$ for all $b > i_L$ in step 2 of Algorithm 2.9; or

$i_{\ell-1} < i' < i_\ell$ for a $\ell \leq L$. Then for $r' = \max\{r : i_\ell^{(r)} < i'\}$,

$$H_2(i_{\ell-1}+1, i', k_\ell) = \sum_{j=i_{\ell-1}+1}^{i'} (\hat{h}_{1j} - \hat{h}_{2j}) < 0.$$

It follows that $H_2(i_{\ell-1}+1, i', k_\ell^{(r')}) < 0$ since $k_\ell^{(r')} < k_\ell^{(R_\ell)} = k_\ell$ from (c) and $H_2(i_{\ell-1}+1, i_\ell^{(r')}, k_\ell^{(r')}) + H_2(i_\ell^{(r')}+1, i', k_\ell^{(r')}) < 0$. Thus $H_2(i_\ell^{(r')}+1, i', k_\ell^{(r')}) < 0$, which implies that there exists r^* with $r^* > r'$ and $i_\ell^{(r^*)} \leq i'$. This is impossible because if so, then $i' = i_\ell^{(r'+1)}$ and for any $r'+1 < R_\ell$, we have

$$0 = H_2(i_{\ell-1}+1, i_\ell^{(r'+1)}, k_\ell^{(r'+1)}) < H_2(i_{\ell-1}+1, i_\ell^{(r'+1)}, k_\ell) = \sum_{j=i_{\ell-1}+1}^{i'} (\hat{h}_{1j}^A - \hat{h}_{2j}^A) < 0.$$

(e) Suppose there exists ℓ such that $k_{\ell+1} > k_\ell > 0$. Then $k_\ell < k_{\ell+1} \leq n_{2i_{\ell+1}}$. Moreover, $H_2(i_\ell+1, i_{\ell+1}, k_\ell) < H_2(i_\ell+1, i_{\ell+1}, k_{\ell+1}) \leq 0$. It follows that the algorithm must not have stopped at R_ℓ^{th} iteration in step 4 of Algorithm 2.9, which is a contradiction. □

Lemma B.2. *Based on the same data, the results from Algorithm 2.9 and Theorem 2.8 satisfy:*

(a). $\hat{k}^i = k_\ell$ *if* $i = i_{\ell-1}+1, \ldots, i_\ell$, $\ell = 1, \ldots, L$ *and* $\hat{k}^i = 0$ *if* $i > i_L$*;*

(b). $\hat{h}_{gi}^T = \hat{h}_{gi}^A, g = 1, 2, i = 1, \ldots, m_g$.

Proof. (a) If $i_{\ell-1} < i \leq i_\ell$ for $\ell \leq L$, then for any $a \leq i$, there exists $\ell' \leq \ell$ such that $i_{\ell'-1} < a \leq i_{\ell'}$. Then from Lemma B.1 (b),

$$H_2(i_{\ell'-1}, i_{\ell'}, k_{\ell'}) \leq 0 = H_2(a, i_{\ell'}, K_2(a, i_{\ell'})).$$

It follows that $H_2(a, i_{\ell'}, k_{\ell'}) \leq H_2(a, i_{\ell'}, K_2(a, i_{\ell'}))$, since $H_2(i_{\ell'-1}, a-1, k_{\ell'}) \geq 0$ from Lemma B.1 (d) and so $K_2(a, i_{\ell'}) \geq k_{\ell'} \geq \cdots \geq k_\ell$ from Lemma B.1 (e). Thus

$$H_2(a, i_\ell, K_2(a, i_\ell)) = 0 = H_2(a, i_{\ell'}, K_2(a, i_{\ell'})) + \sum_{j=\ell'+1}^{\ell} H_2(i_{j-1}, i_j, k_j)$$

$$\geq H_2(a, i_{\ell'}, k_\ell) + \sum_{j=\ell'+1}^{\ell} H_2(i_{j-1}, i_j, k_\ell) = H_2(a, i_\ell, k_\ell).$$

It follows $K_2(a, i_\ell) \geq k_\ell > 0$ and $\min(K_2^+(a, i_\ell), n_{2i_\ell}) \geq k_\ell$ since $k_\ell \leq n_{2i_\ell}$ from Lemma B.1 (b). Therefore,

$$\hat{k}^i = \min_{a \leq i} \max_{b \geq i} \min\{K_2^+(a, b), n_{2b}\} \geq \min_{a \leq i} \min\{K_2^+(a, i_\ell), n_{2i_\ell}\} \geq k_\ell.$$

However, obtaining $\hat{k}_i$ is a minimization problem and its lower bound can be reached when $a = i_{\ell-1} + 1$ and $b = i_\ell$. Thus $\hat{k}^i = k_\ell$.

If $i > i_L$, then $K_2(i_L + 1, b) \leq 0$ for all $b > i_L$ because $H_2(i_L + 1, b, 0) \geq 0$ from step 2 of Algorithm 2.9. So $K_2^+(i_L + 1, b) = 0$ for all $b > i_L$. Hence, $0 \leq \hat{k}^i \leq \max_{b \geq i} \min\{K_2^+(i_L + 1, b), n_{2b}\} = 0$, i.e. $\hat{k}^i = 0$.

(b) For population 1, $\hat{h}_{1i}^T = \log[1 - d_{1i}/(n_{1i} + \hat{k}^i)]$, $\hat{h}_{1i}^A = \log[1 - d_{1i}/(n_{1i} + k_\ell)]$ and $\hat{k}^i = k_\ell$ if $i_{\ell-1} < i \leq i_\ell$, $i = 1, \ldots, L$, so $\hat{h}_{1i}^T = \hat{h}_{1i}^A$ for all $i \leq i_L$. If $i > i_L$, then $\hat{h}_{1i}^T = \log[1 - d_{1i}/n_{1i}] = \hat{h}_{1i}^A$, $i_L < i \leq m_1$.

For population 2, we use induction.

Let $\hat{h}_{20}^A = \hat{h}_{20}^T = 0$, then the result holds for $j = 0$;

Assume for all $i \leq j$, $\hat{h}_{2i}^A = \hat{h}_{2i}^T$;

For the next index $j + 1$, there are possible cases (i), (ii), (iii):

(i). $d_{2(j+1)} > 0$. Then

$$\begin{aligned} \hat{h}_{2(j+1)}^T &= \log[1 - d_{2(j+1)}/(n_{2(j+1)} - \hat{k}^i)] \\ &= \log[1 - d_{2(j+1)}/(n_{2(j+1)} - k_\ell)] = \hat{h}_{2(j+1)}^A, \quad \text{if } i_{\ell-1} < j+1 \leq i_\ell, i \leq L \\ \hat{h}_{2(j+1)}^T &= \log[1 - d_{2(j+1)}/n_{2(j+1)}] = \hat{h}_{2(j+1)}^A, \quad \text{if } i_L < j+1 \leq m_2. \end{aligned}$$

(ii). $d_{2(j+1)} = 0$ and $j + 1 \neq i_\ell, \ell = 1, \cdots, L$. Then $\hat{h}_{(j+1)}^A = 0$. So

$\sum_{i=1}^{(j+1)} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A = \sum_{i=1}^{j+1}(\hat{h}_{1i}^A - \hat{h}_{2i}^A) \geq 0$. It follows that

$$\hat{h}_{(j+1)}^T = \min\{\sum_{i=1}^{j+1} \hat{h}_{1i}^T - \sum_{i=1}^{j} \hat{h}_{2i}^T, 0\} = \min\{\sum_{i=1}^{j+1} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A, 0\} = 0 = \hat{h}_{(j+1)}^A;$$

(iii). $d_{2(j+1)} = 0$ and $j+1 = i_\ell$. Then $\sum_{i=1}^{j+1}(\hat{h}_{1i}^A - \hat{h}_{2i}^A) = 0$. So

$$\hat{h}_{(j+1)}^T = \min\{\sum_{i=1}^{j+1} \hat{h}_{1i}^T - \sum_{i=1}^{j} \hat{h}_{2i}^T, 0\} = \min\{\sum_{i=1}^{j+1} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A, 0\} = \hat{h}_{2(j+1)}^A. \square$$

B.2 Optimization Problem for the Two-sample Case

Consider a general nonlinear optimization problem with inequality constraint

$$\text{minimize } f(\boldsymbol{\theta})$$

$$\text{subject to } g_j(\boldsymbol{\theta}) \leq 0,\ j = 1, 2, \ldots, m,$$

for $\boldsymbol{\theta} \in \mathbb{R}^n$. Define the Lagrangian as

$$Lagr(\boldsymbol{\theta}, \boldsymbol{\lambda}) = f(\boldsymbol{\theta}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{\theta}).$$

For continuously differentiable functions f and g_j, Karush (1939) and Kuhn and Tucker (1951) independently derived the necessary conditions at the solution $\boldsymbol{\theta}^*$. Assume the existence of Lagrange multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, then at the solution $\boldsymbol{\theta}^*$, the following conditions must be satisfied:

$$\frac{\partial f}{\partial \theta_i}(\boldsymbol{\theta}^*) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\boldsymbol{\theta}^*)}{\partial \theta_i} = 0, i = 1, \ldots, n$$

$$g_i(\boldsymbol{\theta}^*) \leq 0, i = 1, \ldots, m$$

$$\lambda_i^* g_i(\boldsymbol{\theta}^*) = 0, i = 1, \ldots, m$$

$$\lambda_i^* \geq 0, i = 1, \ldots, m.$$

These conditions, known as KKT conditions, also constitute sufficient conditions if $f(\boldsymbol{\theta})$ and $g_i(\boldsymbol{\theta})$ are also convex functions. For more details, see Snyman (2005).

In the two-sample problem, we maximize the log likelihood (2.2) subject to a stochastic ordering constraint. As seen in Theorem 2.8 and Algorithm 2.9, we estimate $\hat{h}_{gi}$ for $i \leq m_g$, $g = 1, 2$, since for $i > m_g$, there is no data available in population g. Further, if we set $\hat{h}_{1i} = 0$ for all $i > m_1$ and $\hat{h}_{2i} = -\infty$ for all $i > m_2$, the stochastic ordering constraint is automatically satisfied given $\sum_{j=1}^{m'}(\hat{h}_{1j} - \hat{h}_{2j}) \geq 0$ for all $i > m'$. So the log likelihood (2.2) can be maximized separately for $i \leq m'$ and $i > m'$. Like the KM estimator, the log likelihood is maximized by $\hat{h}_{gi} = \log(1 - d_{gi}/n_{gi})$, $m' + 1 \leq i \leq m_g$. So in Lemma B.3, we only consider maximizing $\sum_{g=1}^{2} \sum_{i=1}^{m'} \left\{d_{gi}\log\left(1 - e^{h_{gi}}\right) + (n_{gi} - d_{gi})\, h_{gi}\right\}$ under the stochastic ordering constraint.

Lemma B.3. *The result* $\{\hat{h}_{gi}, g = 1, 2, i = 1, \ldots, m_g\}$ *from Theorem 2.8 and Algorithm 2.9 is the solution of maximizing the log likelihood* (2.2) *under stochastic ordering constraint* $\sum_{j=1}^{i}(h_{2j} - h_{1j}) \leq 0$, $i = 1, \ldots, m'$, *and* $h_{1i}, h_{2i} \leq 0$, $i = 1, \ldots, m_g$.

Proof. The optimization problem here is:

$$\text{minimize } -\sum_{g=1}^{2}\sum_{i=1}^{m'} \left\{d_{gi}\log\left(1 - e^{h_{gi}}\right) + (n_{gi} - d_{gi})\, h_{gi}\right\} \tag{B.2}$$

$$\text{subject to } \begin{cases} g_i(\boldsymbol{h}_1, \boldsymbol{h}_2) = \sum_{j=1}^{i}(h_{2j} - h_{1j}) \leq 0 \\ g_{m'+i}(\boldsymbol{h}_1, \boldsymbol{h}_2) = h_{2i} \leq 0 \\ g_{2m'+i}(\boldsymbol{h}_1, \boldsymbol{h}_2) = h_{1i} \leq 0 \end{cases}$$

and the corresponding Lagrangian is

$$Lagr(\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{\lambda}) = -\sum_{g=1}^{2}\sum_{i=1}^{m'} \left\{d_{gi}\log\left(1 - e^{h_{gi}}\right) + (n_{gi} - d_{gi})\, h_{gi}\right\} + \sum_{j=1}^{3m'} \lambda_j g_j(\boldsymbol{h}_1, \boldsymbol{h}_2).$$

Thus the KKT conditions are:

$$\frac{d_{1i}e^{\hat{h}_{1i}}}{1-e^{\hat{h}_{1i}}} - (n_{1i} - d_{1i}) - \sum_{j=i}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+2m'} = 0 \tag{B.3a}$$

$$\frac{d_{2i}e^{\hat{h}_{2i}}}{1-e^{\hat{h}_{2i}}} - (n_{2i} - d_{2i}) + \sum_{j=i}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+m'} = 0 \tag{B.3b}$$

$$\sum_{j=1}^{i} (\hat{h}_{2j} - \hat{h}_{1j}) \leq 0 \tag{B.3c}$$

$$\hat{\lambda}_i \sum_{j=1}^{i} (\hat{h}_{2j} - \hat{h}_{1j}) = 0 \tag{B.3d}$$

$$\hat{\lambda}_i, \lambda_{i+m'}, \lambda_{i+2m'} \geq 0 \tag{B.3e}$$

$$\hat{h}_{1i} \leq 0, h_{2i} \leq 0 \tag{B.3f}$$

$$\hat{\lambda}_{i+m}\hat{h}_{2i} = 0 \tag{B.3g}$$

$$\hat{\lambda}_{i+2m}\hat{h}_{1i} = 0 \tag{B.3h}$$

We define $\hat{\lambda}_i$, $\hat{\lambda}_{i+m'}$ and $\hat{\lambda}_{i+2m'}$, $i = 1, \cdots, m'$ as follows :

$$\hat{\lambda}_i = \begin{cases} k_L & \text{if } i = i_L \\ k_\ell - k_{\ell+1} & \text{if } i = i_\ell,\ \ell = 1, \ldots, L-1 \\ 0 & \text{otherwise} \end{cases} \tag{B.4}$$

$$\hat{\lambda}_{i+m'} = \begin{cases} 0 & \text{if } d_{2i} > 0, \\ n_{2i} - k_\ell & \text{if } d_{2i} = 0 \text{ and } i_{\ell-1} < i \leq i_\ell,\ \ell = 1, \ldots, L \\ n_{2i} & \text{if } d_{2i} = 0 \text{ and } i > i_L \end{cases} \tag{B.5}$$

$$\hat{\lambda}_{i+2m'} = \begin{cases} 0 & \text{if } d_{1i} > 0 \\ n_{1i} + \sum_{j=i}^{m'} \hat{\lambda}_j & \text{if } d_{1i} = 0 \end{cases} \tag{B.6}$$

Conditions (B.3c) and (B.3f) are satisfied by Algorithm 2.9. Condition (B.3e) is also satisfied since $k_1 \geq \cdots \geq k_L > 0$ (Lemma B.1 (e)) and $k_\ell \leq n_{2i_\ell} \leq n_{2i}$ (Lemma B.1 (b)) if $i_{\ell-1} < i \leq i_\ell$. Algorithm 2.9 always gives $\hat{h}_{1i} = 0$ if $d_{1i} = 0$ and $\hat{\lambda}_{i+2m'} = 0$ if $d_{1i} > 0$ from (B.6), so condition (B.3h) is satisfied. If $d_{2i} = 0$, $\hat{h}_{2i} < 0$ only when $i = i_\ell$ for some ℓ and $k_\ell = n_{2i_\ell}$. So in this case, $\hat{\lambda}_{i+m'} = 0$ from (B.5), which can lead to (B.3g). From (B.4), $\hat{\lambda}_i \neq 0$ when $i = i_\ell$. However, $\sum_{j=1}^{i_\ell}(\hat{h}_{2j} - \hat{h}_{1j}) = 0$, so condition (B.3d) is also satisfied. From (B.4), we also know that $\sum_{j=i}^{m'} \hat{\lambda}_j = k_\ell$ and $\sum_{j=i}^{m'} \hat{\lambda}_j = 0$ if $i \geq i_L$. If $d_{1i} = 0$, then the condition (B.3a) is $-n_{1i} - \sum_{j=i}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+2m'} = 0$, which is satisfied with definition of $\hat{\lambda}_{i+2m'}$ in (B.6). If $d_{1i} > 0$, then $\hat{\lambda}_{i+2m'} = 0$ from (B.6), so the condition (B.3a) is also satisfied. Similarly, the condition (B.3b) is satisfied.

All KKT conditions are satisfied at the solution from Algorithm 2.9, and (B.2) reaches the global minimum since the optimization function and all constraints are convex. □

B.3 Uniqueness of the Solution

Lemma B.4. *From Theorem 2.8 and Algorithm 2.9,* $\{\hat{h}_{1i}, i = 1, \ldots, m_1\}$ *is the unique C-NPMLE of* h_{1i} *and* $\{\hat{h}_{2i}, i = 1, \ldots, m_2\}$ *is the unique MC-NPMLE of* h_{2i} *under the stochastic ordering constraint.*

Proof. In this proof, we first remove some unnecessary stochastic ordering constraints; then we show that $\hat{k}^i = \sum_{j=i}^{m'} \hat{\lambda}_j$ are unique; last we discuss the uniqueness of the C-NPMLE of h_{1i} and the MC-NPMLE of h_{2i}.

For any C-NPMLE, $\hat{h}_{1i} = 0$ if $d_{1i} = 0$, because $\hat{\lambda}_{i+m'} = n_{1i} + \sum_{j=i}^{m'} \hat{\lambda}_j > 0$ if $i \leq m_1$. So $\sum_{j=1}^{i}(\hat{h}_{2j} - \hat{h}_{1j}) \leq 0$ for $d_{1i} = 0$ will be automatically satisfied given $\sum_{j=1}^{i}(\hat{h}_{2j} - \hat{h}_{1j}) \leq 0$ for $d_{1i} > 0$ and $\hat{h}_{1i}, \hat{h}_{2i} \leq 0$.

Thus for $d_{1i} = 0$, the condition (B.3d) is not necessary, or we can simply set $\hat{\lambda}_i = 0$. Based on this setting of $\hat{\lambda}_i$, we show that $\hat{k}^i = \sum_{j=i}^{m'} \hat{\lambda}_j$ is unique.

Suppose $\hat{k}_i$ is not unique, then we can find two sets of $\{\hat{k}^i\}$ and $\{\tilde{k}^i\}$ from $\{\hat{\lambda}_i\}$ and $\{\tilde{\lambda}_i\}$ that satisfy the KKT conditions with corresponding solutions $\{\hat{h}_{gi}\}$ and $\{\tilde{h}_{gi}\}$. Let $i^* = \min(i : \hat{k}^i \neq \tilde{k}^i)$. Without loss of generality, assume $\hat{k}^{i^*} > \tilde{k}^{i^*} \geq 0$. Then

$$\sum_{j=1}^{i^*-1} (\tilde{h}_{2j} - \tilde{h}_{1j}) = 0. \tag{B.7}$$

Because $\tilde{\lambda}_{i^*-1} = \tilde{k}^{(i^*-1)} - \tilde{k}^{i^*} = \hat{k}^{(i^*-1)} - \tilde{k}^{i^*} > \hat{k}^{(i^*-1)} - \hat{k}^{i^*} = \tilde{\lambda}_{i^*-1} \geq 0$.

Let $\hat{\imath} = \min\{i \geq i^* : \sum_{j=1}^{i} (\hat{h}_2 - \hat{h}_1) = 0, d_{1i} > 0\}$. Then $\hat{\lambda}_j = 0$ for $j = i^*, \ldots, \hat{\imath}$ from condition (B.3d). So

$$\hat{k}^{i^*} = \cdots = \hat{k}^{\hat{\imath}} > \tilde{k}^{i^*} \geq \cdots \geq \tilde{k}^{\hat{\imath}} \tag{B.8}$$

and $\tilde{\lambda}_{i+m'} = n_{2i} - \tilde{k}^i > n_{2\hat{\imath}} - \hat{k}^{\hat{\imath}} \geq 0$ from condition (B.3b) if $d_{2i} = 0$, $i = i^*, \ldots, \hat{\imath}$. So

$$\tilde{h}_{2i} = 0 \qquad \text{if } d_{2i} = 0, i = i^*, \ldots, \hat{\imath} \tag{B.9}$$

from condition (B.3g). Therefore,

$$\begin{aligned}
\sum_{j=1}^{\hat{\imath}} (\tilde{h}_{2j} - \tilde{h}_{1j}) &= \sum_{j=1}^{i^*-1} (\tilde{h}_{2j} - \tilde{h}_{1j}) + \sum_{j=i^*}^{\hat{\imath}} (\tilde{h}_{2j} - \tilde{h}_{1j}) = \sum_{j=i^*}^{\hat{\imath}} (\tilde{h}_{2j} - \tilde{h}_{1j}) && \text{(from (B.7))} \\
&= \sum_{j=i^*}^{\hat{\imath}} \{\log(1 - d_{2j}/(n_{2j} - \tilde{k}^j) - \log(1 - d_{1j}/(n_{1j} + \tilde{k}^j)\} && \text{(from (B.9))} \\
&> \sum_{j=i^*}^{\hat{\imath}} \{\log(1 - d_{2j}/(n_{2j} - \hat{k}^j) - \log(1 - d_{1j}/(n_{1j} + \hat{k}^j)\} && \text{(from (B.8))} \\
&\geq \sum_{j=i^*}^{\hat{\imath}} (\hat{h}_{2j} - \hat{h}_{1j}) \geq \sum_{j=1}^{i^*-1} (\hat{h}_{2j} - \hat{h}_{1j}) + \sum_{j=i^*}^{\hat{\imath}} (\hat{h}_{2j} - \hat{h}_{1j}) = 0.
\end{aligned}$$

which contradicts condition (B.3c). Thus $\hat{k}^i$ is unique and $\hat{h}_{1i}$ is unique, because $\hat{h}_{1i} = \log\{1 - d_{1i}/(n_{1i} + \hat{k}^i)\}$ if $d_{1i} > 0$ and $\hat{h}_{1i} = 0$ if $d_{1i} = 0$.

Also $\hat{h}_{2i} = \log\{1 - d_{2i}/(n_{2i} - \hat{k}^i)\}$ if $d_{2i} > 0$, and $\hat{h}_{2i} = 0$ if $d_{2i} = 0$ and $\hat{k}^i < n_{2i}$ (because $\hat{\lambda}_{i+m'} = n_{2i} - \hat{k}^i > 0$). Therefore, all C-NPMLEs may only differ from each other in population 2 when $d_{2i} = 0$ and $\hat{k}^i = n_{2i}$.

If we sequentially set $\hat{h}_{2i} = \min(\sum_{j=1}^{i} \hat{h}_{1j} - \sum_{j=1}^{i-1} \hat{h}_{2j}, 0)$ if $d_{2i} = 0$ as in Theorem 2.8, then $\sum_{j=1}^{i} \hat{h}_{2j}$ is maximized. Because if $\sum_{j=1}^{i-1} \hat{h}_{2j}$ is maximized, the maximum possible value of $\sum_{j=1}^{i} \hat{h}_{2j}$ is $\min(\sum_{j=1}^{i-1} \hat{h}_{2j}, \sum_{j=1}^{i} \hat{h}_{1j})$, which can be obtained by setting $\hat{h}_{2i} = \min(\sum_{j=1}^{i} \hat{h}_{1j} - \sum_{j=1}^{i-1} \hat{h}_{2j}, 0)$ if $d_{2i} = 0$. Since $\hat{h}_{2i}$ sequentially takes a unique value, the MC-NPMLE is also unique. $\square$

Proofs of the results in section 2.2 for the one-sample case are much simpler and are not presented here.

APPENDIX C

The Pool-adjacent-violators Algorithm for the Linear Ordering Case

We present an algorithm for the case $T_1 \geq_{st} T_2 \cdots \geq_{st} T_G$. Let J be a partition of $\{1, 2, \cdots, G\}$, so that $J = \{B_1, B_2, \cdots\}$. Each member of J is called a block, denoted generally by B. An optimal solution from Algorithm C.1 only contains blocks with consecutive integers. Let $B = \{a, a+1, \cdots, b\}, 1 \leq a \leq b \leq G$, then B^- is the block that contains $a-1$ or $\varnothing$ if $a = 1$ and B^+ is the block that contains $b+1$ or $\varnothing$ if $b = G$. For a given block B, $S_B(x) = \exp(\hat{q}_B)$, where $\hat{q}_B$ is the solution to $\sum_{g \in B}(q; x) = 0$ or 0 if $M_g(x) = 0$ for all $g \in B$.

Algorithm C.1 yields the partition $\hat{J} = \{\hat{B}_1, \ldots, \hat{B}_r\}$ and if $i \in \hat{B}_j$, then $S_i(x) = S_{\hat{B}_j}(x)$ is the pointwise C-NPMLE at x. Let $L = \min\{g : M_g(x) > 0\}$ or $G+1$ if none exists and $U = \max\{g : N_g(x) > 0$ and $S^*_g(x) > 0\}$ or 0 if none exists. Then it is easily seen that $\hat{S}_g(x) = 1$ or $N_g(x) = 0$ for all $g < L$ and $\hat{S}_g(x) = 0$ or $N_g(x) = 0$ for all $g > U$.

Algorithm C.1: Pool adjacent violators algorithm to calculate the pointwise nonparametric maximum likelihood estimator under the simple ordering constraint at time x.

```
Initialization: $J = \{\{1\}, \{2\}, \ldots, \{G\}\}$;
                $B = \{1\}, B^+ = \{2\}$ and $B^- = \varnothing$;
while $B^+ \neq \varnothing$ do
    if $S_B(x) \leq S_{B^+}(x)$ then
        $J \leftarrow J/\{B, B^+\} \cup \{B \cup B^+\}$ (replace $B, B^+$ in $J$ with their union) ;
        $B \leftarrow B \cup B^+$ (replace $B$ with $B^+ \cup B$) ;
        Set new $B^+$ ;
        while $B^- \neq \varnothing$ and $S_B(x) \geq S_{B^-}(x)$ do
            $J \leftarrow J/\{B, B^-\} \cup \{B \cup B^-\}$;
            $B \leftarrow B \cup B^-$ ;
            Set new $B^-$;
        end
    else
        $B^- = B, B = B^+$;
        Set new $B^+$;
    end
end
```

APPENDIX D

Proof of Theorem 3.2

Let λ_1 and λ_2 be Lagrange multipliers. The corresponding Lagrangian function is

$$\begin{aligned} lagrL(\boldsymbol{h}, \boldsymbol{\lambda}) = & \sum_{i=1}^{m} \Big[d_i \log\Big\{1 - \exp h(X_i)\Big\} + (n_i - d_i) h(X_i) \Big] + N(x) h^{\delta}(x) \\ & + \lambda_1 \Big\{ \sum_{j=1}^{M(x)} h(X_j) + h^{\delta}(x) - q \Big\} - \lambda_2 h^{\delta}(x). \end{aligned} \tag{D.1}$$

The Karush-Kuhn-Tucker conditions that must be satisfied at the solution $\hat{\boldsymbol{h}}$ are:

$$-\frac{d_i \exp \hat{h}(X_i)}{1 - \exp \hat{h}(X_i)} + (n_i - d_i) + \hat{\lambda}_1 = 0, \qquad i \le M(x) \tag{D.2a}$$

$$-\frac{d_i \exp \hat{h}(X_i)}{1 - \exp \hat{h}(X_i)} + (n_i - d_i) = 0, \qquad i > M(x) \tag{D.2b}$$

$$N(x) + \hat{\lambda}_1 - \hat{\lambda}_2 = 0, \tag{D.2c}$$

$$\sum_{j=1}^{M(x)} \hat{h}(X_j) + \hat{h}^{\delta}(x) - q = 0, \tag{D.2d}$$

$$\hat{h}^{\delta}(x) \le 0, \tag{D.2e}$$

$$\hat{\lambda}_2 \hat{h}^{\delta}(x) = 0, \tag{D.2f}$$

$$\hat{\lambda}_2 \ge 0. \tag{D.2g}$$

From (D.2a), we have $\hat{h}_1(X_i) = \log\{1 - d_i/(n_i + \hat{\lambda}_1)\}, i \leq M_1(x)$. Note that either $\hat{\lambda}_2 = 0$ or $\hat{h}^\delta(x) = 0$ from (D.2f), so

(1). if $\hat{\lambda}_2 = 0$, then $\hat{\lambda}_1 = -N(x)$ from (D.2c), which is only valid when $\hat{h}^\delta(x) = q - \sum_{j=1}^{M(x)} \log[1 - d_i/\{n_i - N(x)\}] \leq 0$;

(2). if $\hat{h}^\delta(x) = 0$, then $\hat{\lambda}_1$ is the solution of the equation $q - \sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + \lambda)\} = 0$ from (D.2d), which is only valid when $\hat{\lambda}_1 \geq -N(x)$ from (D.2c).

Since $\sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + k)\}$ is an increasing function in k, we can find that $\hat{\lambda}_1 = \max\{\hat{k}, -N(x)\}$, where $\hat{k}$ is the solution of the equation $f(k) = \sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + k)\} - q = 0$. It follows that $\hat{\lambda}_1$ is exactly the same as $K(q; x)$ defined in Theorem 3.2. Therefore, the unique solution from solving (D.2) is the same as in equation (3.4).

APPENDIX E

Proof of Theorem 3.4: $\hat{S}_g(t)$ is Nonincreasing in t

E.1 Notation and Characteristics of the Pointwise C-NPMLE

Let $J_\xi(x)$ be a partition of $\xi \subset \{1, \ldots, G\}$ at time x. For example, if $\xi = \{1, 2, 5\}$, $J_\xi(x)$ might be $\{\{1\}, \{2\}, \{5\}\}$ or $\{\{1, 2\}, \{5\}\}$. Each element B of $J_\xi(x)$ is called a block. Let $E_\xi = \{(i, j) : i, j \in \xi$ and $(i, j) \in E\}$. The pointwise C-NPMLE in ξ subject to constraints E_ξ can be represented as the partition $\hat{J}_\xi(x)$ where every group in each block $B \in \hat{J}_\xi(x)$ has the same estimated survivor function $\hat{S}_\xi(B; x)$ and for $B_1, B_2 \in \hat{J}_\xi(x)$, $\hat{S}_\xi(B_1; x) \neq \hat{S}_\xi(B_2; x)$ if $B_1 \neq B_2$. In Lemma E.1, we give a characterization of the pointwise C-NPMLE. Note that the pointwise C-NPMLE may not be unique after the last observed time for each group. To circumvent this, we set the estimates as low as possible subject to not violating constraints.

Lemma E.1. *A partition $J_\xi(x)$ with corresponding estimate $S_\xi(B; x)$ is the pointwise C-NPMLE subject to the constraints E_ξ at time x if and only if*

1. *Constraints are not violated. That is, for any* $i \in B_1 \subset J_\xi(x)$ *and* $j \in B_2 \subset J_\xi(x)$, $(j, i) \notin E_\xi$ when $S_\xi(B_1; x) > S_\xi(B_2; x)$; and
2. *For any* $B \in J_\xi(x)$, *the estimate* $S_B(B; x) = S_\xi(B; x)$ *where* $J_B(x) = \{B\}$ *is the pointwise C-NPMLE subject to the constraints* E_B.

Proof. Sufficiency. Since the joint log profile likelihood $pl_\xi(q; x)$ for populations in ξ as shown in section 2.6 is a separable concave function, if the condition 2 in Lemma E.1 is satisfied, the estimate will be the pointwise C-NPMLE subject to constraints $\cup_{B \in J_\xi(x)} E_B$. It follows that if condition 1 in Lemma E.1 is also satisfied, the estimate must be the pointwise C-NPMLE subject to constraints E_ξ because $\cup_{B \in J_\xi(x)} E_B \subset E_\xi$ and adding more constraints can not increase the likelihood.

Necessity. Obviously condition 1 holds in Lemma E.1 if $S_\xi(B; x)$ is the pointwise C-NPMLE. If we write down the KKT conditions needed for maximizing the log profile likelihood subject to the constraints E_ξ, the Lagrangian multipliers related to the constraint $(i, j) \in E_\xi$ for any i and j in different block of $J_\xi(x)$ will be zero in the solution because these constraints are inactive in the solution. Thus if we delete these zero valued Lagrangian multipliers, the remaining KKT conditions of the populations in any $B \in J_\xi(x)$ are exactly the same as the pointwise C-NPMLE subject to the constraints E_B. Since the constraints are linear and the joint log profile likelihood is concave, the KKT conditions are also sufficient in our problem. Thus the condition 2 in Lemma E.1 must also hold. □

Lemma E.1 is useful in later proofs because it enables us to consider blocks separately. If $\hat{B}$ is a block from the pointwise C-NPMLE subject to constraint E_ξ for any subpopulation ξ at time x, $\hat{S}_\xi(\hat{B}; x) = \hat{S}_{\hat{B}}(\hat{B}; x)$ will remain the same for any subpopulation ξ for the same block $\hat{B}$. So we use $\hat{S}(\hat{B}; x)$ as the estimate of the pointwise C-NPMLE at time x if $\hat{B} \in \hat{J}_\xi(x)$.

We give two more lemmas to characterize the pointwise C-NPMLE and $K_g(q; x)$.

Lemma E.2. *(a) For any* $x_2 > x_1 > 0$ *and* $q \leq 0$, $K_g(q, x_2) \geq K_g(q, x_1)$*;*

(b) For any $q_2 < q_1 \leq 0$ *and* $x > 0$, $K_g(q_1, x) \geq K_g(q_2, x)$*, and the equality holds only when* $K_g(q_1; x) = K_g(q_2; x) = -N_g(x)$*.*

Proof. (a) We consider separately the cases i) there is no observed event before or at x_1 and ii) there is at least one observed event before or at x_1.

i) In this case, $K_g(q, x_1) = -N_g(x_1) \leq -N_g(x_2) \leq K_g(q, x_2)$.

ii) Let $\hat{k}_j$ be the solution of the equation

$$\sum_{i:X_{gi} \leq x_j} \log\left(1 - \frac{d_{gi}}{n_{gi} + k}\right) = q,$$

then

$$\sum_{i:X_{gi} \leq x_2} \log\left(1 - \frac{d_{gi}}{n_{gi} + \hat{k}_2}\right) = q = \sum_{i:X_{gi} \leq x_1} \log\left(1 - \frac{d_{gi}}{n_{gi} + \hat{k}_1}\right) \geq \sum_{i:X_{gi} \leq x_2} \log\left(1 - \frac{d_{gi}}{n_{gi} + \hat{k}_1}\right).$$

It follows that $\hat{k}_1 \leq \hat{k}_2$ and hence

$$K_g(q, x_2) = \max\{\hat{k}_2, -N_g(x_2)\} \geq \max\{\hat{k}_1, -N_g(x_1)\} = K_g(q, x_1).$$

(b) Suppose there is at least one observed event before or at x, and let $\hat{k}_j$ be the solution of the equation

$$\sum_{i:X_{gi} \leq x} \log\left(1 - \frac{d_{gi}}{n_{gi} + k}\right) = q_j,$$

then $\hat{k}_2 < \hat{k}_1$. Since $K_g(q_j; x) = \max\{\hat{k}_j, -N_g(x)\}$, it can be seen that $K_g(q_1, x) > K_g(q_2, x)$ except when both $\hat{k}_1$ and $\hat{k}_2$ are less than or equal to $-N_g(x)$, in which case $K_g(q_1; x) = K_g(q_2; x) = -N_g(x)$.

If there is no observed event before or at x, then $K_g(q_1; x) = K_g(q_2; x) = -N_g(x)$ by definition. □

Lemma E.3. *For any $\hat{B} \in \hat{J}_\xi(x)$,*

(a) $\sum_{g\in\hat{B}} K_g\{\log \hat{S}(\hat{B};x);x\} \leq 0$, and the strict inequality holds only when $\hat{S}(\hat{B};x) = 1$;

(b) for any $\hat{S}(\hat{B};x) < 1$ and $B_1 \subset \hat{B}$, the following two conditions will not hold simultaneously:

  i) *For all $i \in B_1$ and $j \in \hat{B}/B_1$, $(j,i) \notin E_\xi$;*

  ii) $\sum_{g\in B_1} K_g\{\log \hat{S}(\hat{B};x);x\} < 0$.

Proof. (a) Profile likelihood $\sum_{g\in\hat{B}} pl_{h_g}(q;x)$ is a concave function of q and so the $\hat{S}(\hat{B};x)$ must satisfy $\sum_{g\in\hat{B}} K_g(\log \hat{S}(\hat{B};x);x) = -\sum_{g\in\hat{B}} \frac{d}{dq} pl_{h_g}(\hat{S}(\hat{B};x);x) = 0$. The only one exception is when there is no observed event time before or at x for all $g \in \hat{B}$, in this case $\sum_{g\in\hat{B}} K_g(\log \hat{S}(\hat{B};x);x) = -\sum_{g\in\hat{B}} N_g(x) < 0$ and $\hat{S}(\hat{B};x) = 1$.

(b) Note that $\frac{d}{dq}\sum_{g\in B_1} pl_{h_g}(\log \hat{S}_\xi(B;x);x) = -\sum_{g\in B_1} K_g(\log \hat{S}(\hat{B};x);x) > 0$ and suppose that we can find such a block B_1 that satisfies both i) and ii). Then we can increase estimate $S_\xi(B_1;x)$ to increase the log profile likelihood without violating the constraints. This contradicts $\hat{J}_\xi(x)$ is the partition of the pointwise C-NPMLE at time x. □

E.2 An Algorithm to Obtain the Pointwise C-NPMLE at a Time $x_2 > x_1$

For any $x_2 > x_1$, it can be seen that $\hat{S}_g(x_1) = \hat{S}_g(x_2)$, for all $g = 1, \ldots, G$ if there is no observation between x_1 and x_2, nor a censoring at x_1, nor an event at x_2. Now we consider the situation when only one group g^* has observations between x_1 and x_2. In this case, Algorithm E.1 defines a method to obtain $\hat{J}_\xi(x_2)$ and $\hat{S}_\xi(\hat{B};x_2)$, where $\xi = \{1,\ldots,G\}$. The idea is to find the pointwise C-NPMLE at x_2 using the estimate at x_1 as the starting point.

Algorithm E.1: An algorithm to obtain the pointwise C-NPMLE at time x_2 using the pointwise C-NPMLE at time x_1 as the starting value, where $x_2 > x_1$ and only population g^* has observations between x_1 and x_2. Below $\xi = \{1, \ldots, G\}$.

```
$J_\xi(x_2) = \hat{J}_\xi(x_1) = \{\hat{B}_1, \ldots, \hat{B}_R\}$, where $\hat{S}(\hat{B}_1; x_1) > \cdots > \hat{S}(\hat{B}_R; x_1)$;
Find $r$ such that $g^* \in \hat{B}_r$ and let $A_r = \hat{B}_r$;
while $r \leq R$ do
    Find $\hat{J}_{A_r}(x_2) = \{\hat{B}_{r.1}, \ldots, \hat{B}_{r.L_r}\}$, where $\hat{S}(\hat{B}_{r.1}; x_2) > \cdots > \hat{S}(\hat{B}_{r.L_r}; x_2)$
    (This is the partition of the pointwise C-NPMLE at time $x_2$ subject to constraint
    $E_{A_r}$ for groups in $A_r$ ) ;
    if $r = R$ or $\hat{S}(\hat{B}_{r.L_r}; x_2) > \hat{S}(\hat{B}_{r+1}; x_1)$ then
        $J_\xi(x_2) = J_\xi(x_2)/\{A_r\} \cup \hat{J}_{A_r}(x_2)$    ( replace $\{A_r\}$ with $\hat{J}_{A_r}(x_2)$ );
        stop ;
    else
        $\ell_r = \max\{\ell^* : \hat{S}(\hat{B}_{r.\ell^*}; x_2) > \hat{S}(\hat{B}_{r+1}; x_1)\}$ ;
        $A_{r+1} = \hat{B}_{r+1} \cup \hat{B}_{r.(\ell_r+1)} \cup \cdots \cup \hat{B}_{r.L_r}$ ;
        $J_\xi(x_2) = J_\xi(x_2)/\{A_r, \hat{B}_{r+1}\} \cup \{\hat{B}_{r.1}, \ldots, \hat{B}_{r.\ell_r}\} \cup \{A_{r+1}\}$ ;
        $r = r + 1$ ;
    end
end
```

To illustrate the algorithm, we first show an example in Figure E.1. In this, $\hat{J}_\xi(x_1)$ has five blocks, $\hat{B}_1, \cdots, \hat{B}_5$ and $g^* \in \hat{B}_2$. At first, $r = 2$ and $A_2 = \hat{B}_2$. Then we find $\hat{J}_{A_2}(x_2)$, the partition of the pointwise C-NPMLE subject to constraints E_{A_2} at time x_2 and assume that it has four blocks $\hat{B}_{2.1}, \cdots, \hat{B}_{2.4}$ where $\hat{S}(\hat{B}_{2.1}; x_2) > \hat{S}(\hat{B}_{2.2}; x_2) > \hat{S}(\hat{B}_3, x_1) \geq \hat{S}(\hat{B}_{2.3}; x_2) > \hat{S}(\hat{B}_{2.4}; x_2)$. The blocks $\hat{B}_{2.1}$ and $\hat{B}_{2.2}$ remain separate in the solution and blocks $\hat{B}_3$, $\hat{B}_{2.3}$ and $\hat{B}_{2.4}$ are combined into A_3. Then we again find $\hat{J}_{A_3}(x_2)$ and assume that it has two blocks $\hat{B}_{3.1}$ and $\hat{B}_{3.2}$ where $\hat{S}(\hat{B}_{3.1}; x_2) > \hat{S}(\hat{B}_{3.2}; x_2) > \hat{S}(\hat{B}_4; x_1)$. Blocks $\hat{B}_{3.1}$ and $\hat{B}_{3.2}$ remain separate in the solution and the algorithm ends. The final partition $J_\xi(x_2)$ contains blocks $\hat{B}_1, \hat{B}_{2.1}, \hat{B}_{2.2}, \hat{B}_{3.1}, \hat{B}_{3.2}, \hat{B}_4$ and $\hat{B}_5$.

Lemma E.4. *Algorithm E.1 gives the pointwise C-NPMLE at x_2 and the estimate for each group is nonincreasing over time.*

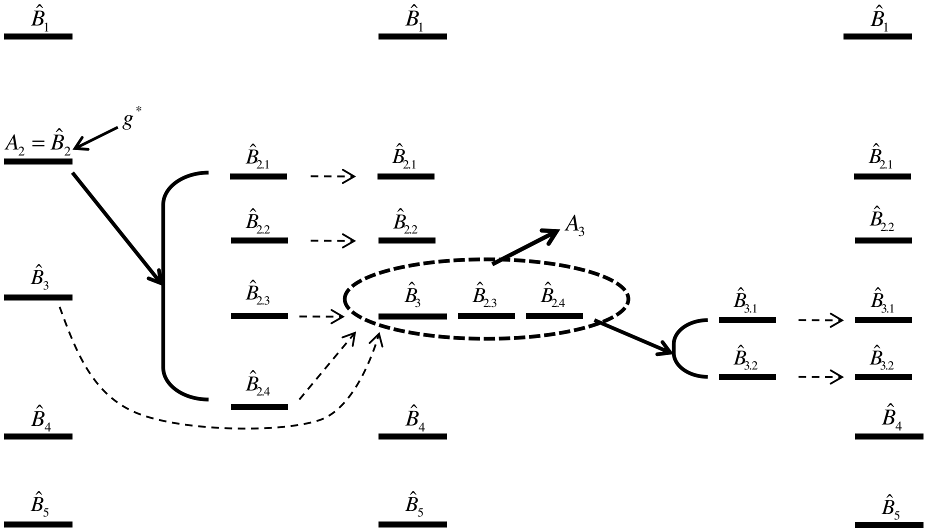


Figure E.1: An example of Algorithm E.1 in Appendix E.2.

Proof. Let $J_\xi(x_2) = \{\hat{B}_1, \ldots, \hat{B}_{u-1}, \hat{B}_{u.1}, \ldots, \hat{B}_{w.\ell_w}, \hat{B}_{w+1}, \ldots, \hat{B}_R\}$ be the result from Algorithm E.1. Then, $\hat{S}(\hat{B}_r; x_2) = \hat{S}(\hat{B}_r; x_1), r = 1, \ldots, u-1, w+1, \ldots, R$ because there is no observation for the groups in $\hat{B}_r$ between x_1 and x_2. Thus, for all $B \in J_\xi(x_2)$, the pointwise C-NPMLE of groups in B has the common estimate of survivor functions $\hat{S}_B(B; x_2)$, which implies that condition 2 in Lemma E.1 must be satisfied.

Next, we prove $\hat{S}(\hat{B}_r; x_1) \geq \hat{S}(\hat{B}_{r.1}; x_2)$ for all $r = u, \cdots, w$.

Suppose $\hat{S}(\hat{B}_r; x_1) < \hat{S}(\hat{B}_{r.1}; x_2)$, then this will gives a contradiction. There are two cases to consider:

Case 1: First step in Algorithm E.1. In this case, $g^* \in \hat{B}_r$ and $A_r = \hat{B}_r$ (line 2 in Algorithm E.1). Then

$$\begin{aligned}\sum_{g \in \hat{B}_{r.1}} K_g(\log \hat{S}(\hat{B}_r; x_1); x_1) &\leq \sum_{g \in \hat{B}_{r.1}} K_g(\log \hat{S}(\hat{B}_r; x_1); x_2) \qquad \text{(Lemma E.2(a))} \\ &\leq \sum_{g \in \hat{B}_{r.1}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) \qquad \text{(Lemma E.2(b))} \quad \text{(E.1)}\end{aligned}$$

$$\leq 0 \qquad \text{(Lemma E.3(a)).}$$

From Lemma E.2(b), equality holds in equation (E.1) only when $K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = -N_g(x_2)$ for all $g \in \hat{B}_{r.1}$. By our convention to set the estimate of a survivor function as low as possible when the number at risk is zero, $\hat{S}(\hat{B}_{r.1}; x_2) = 0$ if $N_g(x_2) = 0$ for all $g \in \hat{B}_{r.1}$. Since $\hat{S}(\hat{B}_{r.1}; x_2) > \hat{S}(\hat{B}_r; x_1) \geq 0$ by our assumption, we have $\sum_{g \in \hat{B}_{r.1}} N_g(x) > 0$. Hence we find that $\sum_{g \in \hat{B}_{r.1}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_1); x_1) < 0$, which implies that $\hat{B}_{r.1} \subset \hat{B}_r$ and both conditions in Lemma E.3(b) are satisfied. This contradicts that $\hat{B}_r \in \hat{J}_\xi(x_1)$.

Case 2: Subsequent steps in Algorithm E.1. In this case, $A_r = \hat{B}_r \cup \hat{B}_{(r-1).(\ell_{r-1}+1)} \cup \cdots \cup \hat{B}_{(r-1).L_{(r-1)}}$ (line 10 in Algorithm E.1 from previous step) and block $\hat{B}_{r.1}$ can be divided into blocks $B^*_{\ell_{r-1}}, \ldots, B^*_{L_{r-1}}$ such that $B^*_{\ell_{r-1}} \subset \hat{B}_r$ and $B^*_\ell \subset \hat{B}_{(r-1).\ell}, \ell = \ell_{r-1} + 1, \ldots, L_{r-1}$. Since

$$\sum_{\ell=\ell_{r-1}}^{L_{r-1}} \sum_{g \in B^*_\ell} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = \sum_{g \in \hat{B}_{r.1}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) \leq 0,$$

we find that either

a) there is at least one ℓ' that satisfies $\sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) < 0$; or

b) $\sum_{g \in B^*_\ell} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = 0$ for all $\ell = \ell_{r-1}, \ldots, L_{r-1}$.

In case b), we pick ℓ' such that $\sum_{g \in B^*_{\ell'}} N_g(x_2) > 0$. Note that $\hat{S}(\hat{B}_{r.1}; x_2) > \hat{S}(\hat{B}_r; x_1) \geq \hat{S}(\hat{B}_{(r-1).\ell}; x_2)$, for all $\ell = \ell_{r-1} + 1, \ldots, L_{r-1}$. If $\ell' = \ell_{r-1}$, then we have

$$\begin{aligned} \sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_r; x_1); x_1) &\leq \sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_r; x_1); x_2) \\ &\leq \sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = 0, \end{aligned} \tag{E.2}$$

or if $\ell' > \ell_{r-1}$, then we have

$$\sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r-1.\ell'}; x_2); x_2) \leq \sum_{g \in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = 0. \tag{E.3}$$

Neither the equality in equation (E.2) nor the equality in equation (E.3) can hold since otherwise

$\sum_{g\in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r.1}; x_2); x_2) = \sum_{g\in B^*_{\ell'}} N_g(x_2) < 0$. Hence we find that

$\sum_{g\in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_r; x_1); x_1) < 0$ in equation (E.2) or $\sum_{g\in B^*_{\ell'}} K_g(\log \hat{S}(\hat{B}_{r-1.\ell'}; x_2); x_2) < 0$ in equation (E.3), which contradicts $\hat{B}_r \in \hat{J}_\xi(x_1)$ or $\hat{B}_{r.\ell'} \in \hat{J}_{A_r}(x_2)$.

Thus, we established that $\hat{S}(\hat{B}_1; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{u-1}; x_2) \geq \hat{S}(\hat{B}_u; x_1) \geq \hat{S}(\hat{B}_{u.1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{u.L_u}; x_2) \geq \hat{S}(\hat{B}_{u+1}; x_1) \geq \hat{S}(\hat{B}_{(u+1).1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{w.L_w}; x_2) \geq \hat{S}(\hat{B}_{w+1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_R; x_2)$. It is easy to see that the constraints are not violated in the solution $J_\xi(x_2)$ because $\hat{S}(\hat{B}_{u-1}; x_2) \geq \hat{S}(\hat{B}_{u.1}; x_2)$, $\hat{S}(\hat{B}_{r.L_r}; x_2) \geq \hat{S}(\hat{B}_{(r+1).1}; x_2), r = u, \ldots, w-1$, and $\hat{S}(\hat{B}_{w.L_w}; x_2) \geq S(\hat{B}_{w+1}; x_2)$. Therefore, the result from Algorithm E.1 is the pointwise C-NPMLE at time x_2. Furthermore, for any $g \in \hat{B}_r, r = 1, \ldots, u-1, w+1, \ldots, R$, $\hat{S}_g(x_2) = \hat{S}_g(x_1)$ since $\hat{B}_r \in J_\xi(x_2)$ and $\hat{S}(B_r; x_1) = \hat{S}(B_r; x_2)$, and for any $g \in \hat{B}_r, r = u, \ldots, w$, $\hat{S}_g(x_2) \leq \hat{S}_g(x_1)$, since $g \in \hat{B}_{r'.\ell}$ for an r' such that $r' \geq r$ and $\hat{S}(\hat{B}_r; x_1) \geq \hat{S}(\hat{B}_{r'.\ell}; x_2)$. □

E.3 Completion of the Proof

To complete the proof of Theorem 3.4 when two or more groups have observations between x_1 and x_2, we can produce the pointwise C-NPMLE by sequentially including observations from a group at a time. Since each time when we add more observations from a group, the pointwise C-NPMLE will not increase compared to that before adding these observations, the pointwise C-NPMLE will not increase over time.

APPENDIX F

Proof of Theorem 3.5

To establish this, we first prove

Lemma F.1. $\max_{1 \leq g \leq G} |S_g^*(t) - S_g(t)| \geq \max_{1 \leq g \leq G} |\hat{S}_g(t) - S_g(t)|$.

of Lemma. At a fixed time t, we first prove for any k,

$$\hat{S}_k(t) - S_k(t) \leq \max_{1 \leq g \leq G} \{S_g^*(t) - S_g(t)\}.$$

If $\hat{S}_k(t) \leq S_k^*(t)$, then $\hat{S}_k(t) - S_k(t) \leq S_k^*(t) - S_k(t) \leq \max_{1 \leq g \leq G}\{S_g^*(t) - S_g(t)\}$.

If $\hat{S}_k(t) > S_k^*(t)$, then there must be at least one r in the same "pooled" group such that $S_r(t) \leq S_k(t)$ and $S_r^*(t) \geq \hat{S}_r(t) = \hat{S}_k(t)$. Otherwise, if we divide this "pooled" group B into two blocks $B_1 = \{g : g \in B, S_g(t) \leq S_k(t)\}$ and $B - B_1$, then the likelihood will increase if we lower the common estimate of groups in block B_1 at time t since all estimates of survivor functions for the groups in B_1 change towards the unrestricted maximum likelihood estimators, and the constraint will not be violated, which contradicts that $\hat{S}_g(t)$ is the pointwise constrained nonparametric maximum likelihood estimator. It follows that $\hat{S}_k(t) - S_k(t) \leq \hat{S}_r(t) - S_r(t) \leq S_r^*(t) - S_r(t) \leq \max_{1 \leq g \leq G}\{S_g^*(t) - S_g(t)\}$.

A similar argument shows that

$$\hat{S}_k(t) - S_k(t) \geq \min_{1 \leq g \leq G} \{S_g^*(t) - S_g(t)\}.$$

Thus,

$$\begin{aligned} -\max_{1 \leq g \leq G} |S_g^*(t) - S_g(t)| \leq \min_{1 \leq g \leq G} \{S_g^*(t) - S_g(t)\} \leq \hat{S}_k(t) - S_k(t) \\ \leq \max_{1 \leq g \leq G} \{S_g^*(t) - S_g(t)\} \leq \max_{1 \leq g \leq G} |S_g^*(t) - S_g(t)|. \end{aligned}$$

So

$$|\hat{S}_k(t) - S_k(t)| \leq \max_{1 \leq g \leq G} |S_g^*(t) - S_g(t)|.$$

This establishes Lemma F.1. □

For the case when $t \leq \tau = \min\{\tau_1, \ldots, \tau_G\}$ and $n_g \to \infty, g = 1, \ldots, G$,

$$\begin{aligned} \lim_{n_g \to \infty} pr\{\sup_{t<\tau} |\hat{S}_g(t) - S_g(t)| > \epsilon\} \leq \lim_{n_g \to \infty} pr\{\sup_{t<\tau} \max_{1 \leq k \leq G} |S_k^*(t) - S_k(t)| > \epsilon\} \\ \leq \sum_{k=1}^{G} \lim_{n_k \to \infty} pr\{\sup_{t<\tau} |S_k^*(t) - S_k(t)| > \epsilon\} = 0. \end{aligned} \tag{F.1}$$

Next we consider the case when $t > \tau$ and $n_g \to \infty, g = 1, \ldots, G$.

Lemma F.2. *For a given k, let $E_k^+ = \{(k', k) \in E\}$, where $E = \{(g', g) : T_{g'} \geq_{st} T_g, g, g' = 1, \ldots, G\}$. If $N_k(t) = 0$, then for any group g satisfying $N_g(t) > 0$, $\hat{S}_g(t) = \tilde{S}_g(t)$, where $\tilde{S}_g(t)$ is the pointwise constrained nonparametric maximum likelihood estimator subject to constraints defined by E/E_k^+ (which denotes the set of constraints in E excluding those in E_k^+).*

Proof. The only possible situation that $\tilde{S}_g(t)$ is not the pointwise constrained nonparametric maximum likelihood estimator subject to the constraints defined by E is that there exist $(k', k) \in E_k^+$ and $\tilde{S}_{k'}(t) < \tilde{S}_k(t)$. Since $N_k(t) = 0$, the likelihood does not change if we lower the estimate for group k at time t. So set $\tilde{S}_k(t) = \min\{\tilde{S}_g(t) : (g, k) \in E_k^+\}$, then all constraints in E will be satisfied, hence $\tilde{S}_g(t), g \neq k$ is the pointwise constrained nonparametric maximum likelihood estimator. We have shown in Appendix D that $\hat{S}_g(t)$ is unique if $N_g(t) > 0$. Therefore $\hat{S}_g(t) = \tilde{S}_g(t)$ if $N_g(t) > 0$. $\square$

Let $Q_g^*(t) = S_g^*(\min\{t, \tau_g^*\})$ and $Q_g(t) = S_g(\min\{t, \tau_g-\})$, where τ_g^* is the last observed time in group g. Then

Lemma F.3. *$Q_g^*(t)$ is uniformly consistent for $Q_g(t)$ on $[0, \infty)$.*

Proof. If $S_g^c(\tau_g-) = 0$, then $\tau_g^* \to \tau_g$ as n_g goes to infinity,

$$\sup_{t<\infty} |Q_g^*(t) - Q_g(t)| = \sup_{t<\tau_g} |S_g^*(t) - S_g(t)| \to 0 \text{ with probability 1.}$$

If $S_g^c(\tau_g-) > 0$, then $S_g(\tau_g) = S_g(\tau_g-)$ by the condition of no common jumps of $S_g(t)$ and $S_g^c(t)$, and $\tau_g^* = \tau_g$ as n_g goes to infinity. So

$$\sup_{t<\infty} |Q_g^*(t) - Q_g(t)| = \sup_{t\leq\tau_g} |S_g^*(t) - S_g(t)| \to 0 \text{ with probability 1,}$$

under the condition $S_g(\tau_g) = S_g(\tau_g-)$. (see Corollary 1.2 in Stute and Wang 1993, p 1595). $\square$

Let $E(t) = E/\bigcup_{k:\tau_k<t} E_k^+$ and let $\hat{Q}_g(t)$ be the pointwise nonparametric maximum likelihood estimator of $Q_g(t)$ subject to constraint $E(t)$, then the strong uniform consistency for $\hat{Q}_g(t)$ holds for all $t \geq 0$ using the same argument leading to the result in equation (F.1). Since $\hat{Q}_g(t) = \hat{S}_g(t)$ by applying Lemma F.2 multiple times and $Q_g(t) = S_g(t)$ for all $t < \tau_g$, the strong uniform consistency of $\hat{S}_g(t)$ for $S_g(t)$ is established on $[0, \tau_g)$. If $S_g(\tau_g-) = S_g(\tau_g)$, the strong uniform consistency of $S_g(t)$ for $S_g(t)$ holds on $[0, \tau_g]$.

This completes the proof of Theorem 3.5.

APPENDIX G

Proof of Theorem 3.6

Let $Z_g^L(x) = n^{1/2}\{\log S_g^*(x) - \log S_g(x)\}$, then by the delta method, $Z_g^L(x) \xrightarrow{D} Z_g/S_g(x)$, $g = 1, \ldots, G$, where $Z_g, g = 1, \ldots, G$ is defined in Section 3. For a fixed x, since $S_g^*(x)$ is a consistent estimator of $S_g(x)$, $g = 1, \ldots, G$, if $(i, j) \in E$ and $S_i(x) > S_j(x)$, $pr\{S_i^*(x) - S_j^*(x) \leq 0\} \to 0$ as $n_i, n_j \to \infty$, i.e. the constraint between group i and j will be asymptotically inactive with arbitrary large probability at time x. So the asymptotic distribution of $\hat{S}_g(x)$ is only determined by the groups with the same true survivor function at time x.

For any group g, $N_g(x)/n_g \to S_g(x)S_g^c(x)$ in probability as $n_g \to \infty$. So $1/N_g(x) = O_p(1/n)$ for all x where $S_g(x)S_g^c(x) > 0$. Let $\hat{q} = Av_n(\ell, u, x)$ be the common value of the survivor function when combining groups ℓ to u at time x and assume that $S_\ell(x) = \cdots = S_u(x)$. Then from Theorem 3.2 and using the fact that $K_g(\hat{q}; x)/n \to 0$ in probability as $n \to \infty$, it follows that for each g, $\ell \leq g \leq u$,

$$\begin{aligned}
\hat{q} &= \sum_{X_{gi} \leq x} \log\left\{1 - \frac{d_{gi}}{n_{gi} + K_g(\hat{q}; x)}\right\} = -\sum_{X_{gi} \leq x} \frac{d_{gi}}{n_{gi} + K_g(\hat{q}; x)}\left\{1 + O_p\left(\frac{1}{n}\right)\right\} \\
&= \sum_{X_{gi} \leq x} \log(1 - \frac{d_{gi}}{n_{gi}}) + \sum_{X_{gi} \leq x} \frac{d_{gi}}{n_{gi}} \frac{K_g(\hat{q}; x)}{n_{gi}} \{1 + o_p(1)\} + O_p\left(\frac{1}{n}\right).
\end{aligned}$$

Thus,

$$n^{1/2}\{\hat{q} - \log S_g(x)\} = Z_g^L(x) + n^{1/2} \sum_{X_{gi} \le x} K_g(\hat{q}; x) \, \frac{d_{gi}}{n_{gi}^2} \, \{1 + o_p(1)\} + O_p(n^{-1/2}) \quad \text{(G.1)}$$

Since both $n^{1/2}\{\hat{q} - \log S_g(x)\}$ and $Z_g^L(x)$ are bounded in probability, $n^{1/2} \sum_{X_{gi} \le x} d_{gi} K_g(\hat{q}; x)/n_{gi}^2$ must be bounded in probability. Thus equation (G.1) becomes

$$n^{1/2}\{\hat{q} - \log S_g(x)\} = Z_g^L(x) + n^{1/2} K_g(\hat{q}; x) \sum_{X_{gi} \le x} \frac{d_{gi}}{n_{gi}^2} + o_p(1) \quad \text{(G.2)}$$

Let $w_{gn}(x) = n/\{S_g^2(x) \sum_{X_{gi} \le x} d_{gi}/n_{gi}^2\}$. It is well known that

$$\frac{1}{n_g} \sum_{X_{gi} \le x} \frac{d_{gi}}{n_{gi}^2} \xrightarrow{p} V_g(x) \text{ as } n_g \to \infty.$$

Thus $w_{gn}(x) \to c_g/\sigma_g^2(x) = w_g(x)$ as $n \to \infty$. Multiplying equation (G.2) by $w_{gn}(x)$ gives

$$w_{gn}(x) n^{1/2}\{\hat{q} - \log S_g(x)\} = w_{gn}(x) Z_g^L(x) + \frac{n^{\frac{3}{2}} K_g(\hat{q}; x)}{S_g^2(x)} + o_p(1) \quad \text{(G.3)}$$

Since $\sum_{g=\ell}^{u} K_g(\hat{q}; x) = 0$ as shown in equation (3.7) for any n, summing equation (G.3) over g from ℓ to u and dividing by $\sum_{g=\ell}^{u} w_{gn}(x)$ yields

$$n^{1/2}\{\hat{q} - \log S_k(x)\} = \frac{\sum_{g=\ell}^{u} Z_g^L(x) w_{gn}(x)}{\sum_{g=\ell}^{u} w_{gn}(x)} + o_p(1) \xrightarrow{D} \frac{\sum_{g=\ell}^{u} Z_g(x) w_g(x)}{S_k(x) \sum_{g=\ell}^{u} w_g(x)},$$

for any k such that $\ell \le k \le u$ because all $S_g(x)'s$ are equal for $\ell \le g \le u$. Thus by the delta method, we have

$$n^{1/2}\{Av_n(\ell, u, x) - S_k(x)\} \xrightarrow{D} \frac{\sum_{g=\ell}^{u} Z_g(x) w_g(x)}{\sum_{g=\ell}^{u} w_g(x)}.$$

Robertson and Waltman (1968) showed that the maximum likelihood estimator under the simple ordering constraint is

$$\hat{S}_k(x) = \min_{L_k(x) \le \ell \le k} \max_{k \le u \le U_k(x)} Av_n(\ell, u, x),$$

where $L_k(x) = \min\{i : S_i(x) = S_k(x)\}$ and $U_k(x) = \max\{i : S_i(x) = S_k(x)\}$ as defined in Theorem 3.6. Thus

$$\begin{aligned} n_k^{1/2}\{\hat{S}_k(x) - S_k(x)\} &= c_k^{1/2} \min_{L_k(x) \le \ell \le k} \max_{k \le u \le U_k(x)} n^{1/2}\{Av_n(\ell, u, x) - S_k(x)\} \\ &\xrightarrow{D} c_k^{1/2} \min_{L_k(x) \le \ell \le k} \max_{k \le u \le U_k(x)} \frac{\sum_{g=\ell}^{u} Z_g(x) w_g(x)}{\sum_{g=\ell}^{u} w_g(x)}. \end{aligned}$$

This completes the proof of Theorem 3.6. Now we discuss extensions and special cases of Theorem 3.6.

First consider the case when there exists g' such that $c_{g'} = 0$ while $c_g > 0$. The asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}$ will be the same as in equation (3.8) with the weight for group g' set to zero. This is because

$$\lim_{c_{g'} \to 0} \frac{Z_{g'}(x) w_{g'}(x)}{\sum_{g=\ell}^{u} w_g(x)} = \lim_{c_{g'} \to 0} \frac{N(0, 1/w_{g'}) w_{g'}(x)}{\sum_{g=\ell}^{u} w_g(x)} = \lim_{c_{g'} \to 0} \frac{N(0,1) w_{g'}^{1/2}(x)}{\sum_{g=\ell}^{u} w_g(x)} \xrightarrow{p} 0.$$

This result might indirectly show that the finite samples can be ingored in the asymptotic properties in our setting.

Then we discuss the case when there are some groups for which the support of the censoring distribution is less than x. As discussed in Appendix E, the asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}$, $x < \tau_g$, can be obtained by modifying the constraint set to $E(x)$. Note that the ordering constraints $T_{k'} \ge_{st} T_k$, $k' = 1, \ldots, k-1$ are removed if $x > \tau_k^*$. Also if $S_k(x) < S_k(\tau_k -)$, then constraints $T_k \ge_{st} T_{k'}, k = k+1, \ldots, G$ will be asymptotically irrelevant because $S_k^*(x)$ can always take value $S_k^*(\tau_k^*)$ and $S_k^*(\tau_k^*) > S_{k'}^*(x), k' = k+1, \ldots, G$ asymptotically. So group k can be removed from obtaining the asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}, x < \tau_g$ or equivalently we can set $w_k = 0$ in equation

(3.8) at time x. If $S_k(x) = S_k(\tau_k-)$, the problem will be changed to the partial ordering case and then we can appeal to the Conjecture in section 3 to give the asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}$. For example, in the case where $T_1 \geq_{st} T_2 \geq_{st} T_3 \geq_{st} T_4$ and $\tau_1 < \tau_2 < \tau_3 < \tau_4$, we consider the asymptotic distribution of $n_4^{1/2}\{\hat{S}_4(x) - S_4(x)\}$ at time $x \in [\tau_3, \tau_4)$. If $S_1(x) = S_1(\tau_1-) = \cdots = S_4(x) = S_4(\tau_4-)$, the constraints at time x are changed to $T_1 \geq_{st} T_4$, $T_2 \geq_{st} T_4$ and $T_3 \geq_{st} T_4$.

BIBLIOGRAPHY

REFERENCES

Abrahamowicz, M., Clampl, A., and Ramsay, J. O. (1989), "Nonparametric Density Estimation for Censored Survival Data: Regression-spline Approach," *The Canadian Journal of Statistics*, 20, 171–185.

Andrews, D. W. K. (2000), "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405.

Arcones, M. A., Kvam, P. H., and Samaniego, F. J. (2002), "Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint," *Journal of the American Statistical Association*, 97, 170–182.

Arjas, E. and Gasbarra, D. (1996), "Bayesian Inference of Survival Probabilities, Under Stochastic Ordering Constraints," *Journal of the American Statistical Association*, 91, 1101–1109.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," *The Annals of Mathematical Statistics*, 26, pp. 641–647.

Barlow, R. E., Bartholomew, D., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference under Order Restrictions: the Theory and Application of Isotonic Regression*, New York: Wiley.

Best, M. J., Chakravarti, N., and Ubhaya, V. A. (1999), "Minimizing Separable Convex

Functions Subject to Simple Chain Constraints," *SIAM Journal on Optimization*, 10, 658–672.

Bouezmarni, T. and Rombouts, J. V. K. (2010), "Nonparametric Density Estimation for Positive Time Series," *Computational Statistics & Data Analysis*, 54, 245–261.

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*, 61, 1067–1080.

Chaubey, Y. P. and Kochar, S. C. (2000), "Smooth Estimation of Stochastically Ordered Survival Functions," *Journal of the Indian Statistical Association*, 38, 209–225.

Chaudhuri, S. and Perlman, M. D. (2005), "On the Bias and Mean-square Error of Order-restricted Maximum Likelihood Estimators," *Journal of Statistical Planning and Inference*, 130, 229 – 250.

— (2007), "Consistent Estimation of the Minimum Normal Mean under the Tree-order Restriction," *Journal of Statistical Planning and Inference*, 137, 3317 – 3335.

Chen, S. (2000), "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

— (2002), "Local Linear Smoothers Using Asymmetric Kernels," *Annals of the Institute of Statistical Mathematics*, 54, 312–323.

Chen, S. X. (1999), "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131–145.

Cohen, A. and Sackrowitz, H. B. (2002), "Inference for the Model of Several Treatments and a Control," *Journal of Statistical Planning and Inference*, 107, 89–101.

— (2004), "A Discussion of Some Inference Issues in Order Restricted Models," *The Canadian Journal of Statistics*, 32, 199–205.

Dardanoni, V. and Forcina, A. (1998), "A Unified Approach to Likelihood Inference on

Stochastic Orderings in a Nonparametric Context," *Journal of the American Statistical Association*, 93, 1112–1123.

Dunson, D. B. and Peddada, S. D. (2008), "Bayesian Nonparametric Inference on Stochastic Ordering," *Biometrika*, 95, 859–874.

Dykstra, R. L. (1982), "Maximum Likelihood Estimations of the Survival Functions of Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, 77, 621–628.

Dykstra, R. L. and Feltz, C. J. (1989), "Nonparametric Maximum Likelihood Estimation of Survival Functions with a General Stochastic Ordering and its Dual," *Biometrika*, 76, 331–341.

Efron, B. (1981), "Censored Data and the Bootstrap," *Journal of the American Statistical Association*, 76, 312–319.

— (1987), "Better Bootstrap Confidence Intervals (with Comments)," *Journal of the American Statistical Association*, 82, 171–197.

El Barmi, H. and Mukerjee, H. (2005), "Inferences Under a Stochastic Ordering Constraint: The K-Sample Case," *Journal of the American Statistical Association*, 100, 252–261.

Evans, M., Gilula, Z., Guttman, I., and Swartz, T. (1997), "Bayesian Analysis of Stochastically Ordered Distributions of Categorical Variables," *Journal of the American Statistical Association*, 92, pp. 208–214.

Feltz, C. 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, 1012–1019.

Fernandez, M. A., Rueda, C., and Salvador, B. (1999), "The Loss of Efficiency Estimating Linear Functions under Restrictions," *Scandinavian Journal of Statistics*, 26, 579–592.

Gangnon, R. E. and King, W. N. (2002), "Minimum Distance Estimation of the Distribution

Functions of Stochastically Ordered Random Variables," *Journal of the Royal Statistical Society*, 51, 485–492.

Gelfand, A. E. and Kottas, A. (2001), "Nonparametric Bayesian Modeling for Stochastic Order," *Annals of the Institute of Statistical Mathematics*, 53, 865–876.

Gill, R. (1983), "Large Sample Behaviour of the Product-limit Estimator on the Whole Line," *The Annals of Statistics*, 11, 49–58.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York:Springer-Verlag.

Hirotsu, C. (2005), "Isotonic Inference," in *Encyclopedia of Biostatistics*, eds. Armitage, P. and Colton, T., John Wiley & Sons, Ltd.

Hoff, P. D. (2000), "Constrained Nonparametric Maximum Likelihood via Mixtures," *Journal of Computational and Graphical Statistics*, 9, 633–641.

— (2003a), "Bayesian Methods for Partial Stochastic Orderings," *Biometrika*, 90, 303–317.

— (2003b), "Nonparametric Estimation of Convex Models via Mixtures," *The Annals of Statistics*, 31, 174–200.

Hwang, J. T. G. and Peddada, S. D. (1994), "Confidence Interval Estimation Subject to Order Restrictions," *The Annals of Statistics*, 22, 67–93.

Johansen, S. (1978), "The Product Limit Estimator as Maximum Likelihood Estimator," *Scandinavian Journal of Statistics*, 5, 195–199.

Jones, M. C. (1993), "Simple Boundary Correction for Kernel Density Estimation," *Statistical Computing*, 3, 135–146.

Jones, M. C. and Foster, P. J. (1996), "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation," *Statistica Sinica*, 6, 1005–1013.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, Inc., 2nd ed.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

Karabatsos, G. and Walker, S. G. (2007), "Bayesian Nonparametric Inference of Stochastically Ordered Distributions, with Pólya Trees and Bernstein Polynomials," *Statistics & Probability Letters*, 77, 907–913.

Kardaun, O. (1983), "Statistical Survival Analysis of Male Larynx-Cancer Patients - a Case Study -," *Statistica Neerlandica*, 37, 103–126.

Karush, W. (1939), "Minima of Functions of Several Variables with Inequalities as Side Constraints," M.Sc. dissertation, University of Chicago, Chicago, Illinois, Department of Mathematics.

Kelly, R. E. (1989), "Stochastic Reduction of Loss in Estimating Normal Means by Isotonic Regression," *The Annals of Statistics*, 17, 937–940.

Kim, J. S. and Proschan, F. (1991), "Piecewise Exponential Estimator of the Survivor Function," *Reliability, IEEE Transactions on*, 40, 134–139.

Kuhn, H. W. and Tucker, A. W. (1951), "Nonlinear Programming," in *Proceedings of 2nd Berkeley Symposium*, ed. Neyman, J., Berkeley: University of California Press, pp. 481–492.

Lee, C.-I. C. (1981), "The Quadratic Loss of Isotonic Regression under Normality," *The Annals of Statistics*, 9, pp. 686–688.

— (1988), "Quadratic Loss of Order Restricted Estimators for Treatment Means with a Control," *The Annals of Statistics*, 16, pp. 751–758.

Lehmann, E. L. (1955), "Ordered Families of Distributions," *The Annals of Mathematical Statistics*, 26, 399–419.

Li, G. (1995), "On Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored Data," *The Annals of Statistics*, 25, 95–104.

Li, Z., Taylor, J. M. G., and Nan, B. (2010), "Construction of Confidence Intervals and

Regions for Ordered Binomial Probabilities," *The American Statistician*, 64, 291–298.

Lim, J., Kim, S. J., and Wang, X. (2009), "Estimating Stochastically Ordered Survival Functions via Geometric Programming," *Journal of Computational and Graphical Statistics*, 18, 978–994.

Lo, S. H. (1987), "Estimation of Distribution Functions under Order Restrictions," *Statistics and Decisions*, 5, 251–262.

Marchand, E. and Strawderman, W. E. (2004), "Estimation in Restricted Parameter Spaces: A Review," *Lecture Notes-Monograph Series*, 45, 21–44.

Marchand, r. and Strawderman, W. E. (2006), "On the Behavior of Bayesian Credible Intervals for Some Restricted Parameter Space Problems," *Lecture Notes-Monograph Series*, 50, pp. 112–126.

Meier, P. (1975), "Estimation of a Distribution Function from Incomplete Observations," in *Perspectives in Probability and Statistics*, ed. Gani, J., Sheffield, England: Applied Probability Trust, pp. 67–87.

Murphy, S. A. (1995), "Likelihood Ratio-based Confidence Intervals in Survival Analysis," *Journal of the American Statistical Association*, 90, 1399–1405.

Peddada, S. D. (1997), "Confidence Interval Estimation of Population Means Subject to Order Restrictions Using Resampling Procedures," *Statistics & Probability Letters*, 31, 255–265.

Præstgaard, J. T. and Huang, J. (1996), "Asymptotic Theory for Nonparametric Estimation of Survival Curves under Order Restrictions," *The Annals of Statistics*, 24, 1679–1716.

Proust-Lima, C. and Taylor, J. M. G. (2009), "Development and Validation of a Dynamic Prognostic Tool for Prostate Cancer Recurrence Using Repeated Measures of Posttreatment PSA: a Joint Modeling Approach," *Biostatistics*, 10, 535–549.

Ramsay, J. O. (1988), "Monotone Regression Splines in Action(with Comments)," *Statis-*

tical Science, 3, 425–461.

Robertson, T. and Waltman, P. (1968), "On Estimating Monotone Parameters," *The Annals of Mathematical Statistics*, 39, 1030–1039.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, John Wiley & Sons Ltd.

Roe, B. P. and Woodroofe, M. B. (2000), "Setting Confidence Belts," *Physical Review D*, 63, 013009.

Rojo, J. (2004), "On the Estimation of Survival Functions under a Stochastic Order Constraint," in *The First Erich L. Lehmann Symposium Optimality. IMS Lecture Notes-Monograph Series*, ed. Pérez-Abreu, V., Beachwood, OH: Institute of Mathematical Statistics, vol. 44, pp. 37–61.

Rojo, J. and El Barmi, H. (2003), "Estimation of Distribution Functions Under Second Order Stochastic Dominance," *Statistica Sinica*, 13, 903–926.

Rojo, J. and Ma, Z. (1996), "On the Estimation of Stochastically Ordered Survival Functions," *Journal of Statistical Computation and Simulation*, 55, 1–21.

Schoenfeld, D. A. (1986), "Confidence Bounds for Normal Means Under Order Restrictions, With Application to Dose-Response Curves, Toxicology Experiments, and Low-Dose Extrapolation," *Journal of the American Statistical Association*, 81, 186–195.

Shaked, M. and Shanthikumar, J. G. (1994), *Stochastic Orders and Their Applications*, Associated Press.

— (2007), *Stochastic Orders*, Springer New York.

Silvapulle, M. J. and Sen, P. K. (2005), *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*, Wiley-Interscience.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, Springer-Verlag New York, Inc.

Snyman, J. A. (2005), *Practical Mathematical Optimization: An Introduction to Basic*

Optimization Theory and Classical and New Gradient-Based Algorithms, New York: Springer.

Stute, W. and Wang, J.-L. (1993), “The Strong Law under Random Censorship,” *The Annals of Statistics*, 21, pp. 1591–1607.

Taylor, J. M. G., Wang, L., and Li, Z. (2007), “Analysis on Binary Responses with Ordered Covariates and Missing Data.” *Statistics in Medicine*, 26, 3443–3458.

Thomas, D. R. and Grunkemeier, G. L. (1975), “Confidence Interval Estimation of Survival Probabilities for Censored Data,” *Journal of the American Statistical Association*, 70, 865–871.

van Eeden, C. (2006), *Restricted Parameter Space Estimation Problems: Admissibility and Minimaxity Properties*, Springer.

Weber, G.-W. (2003), *Generalized Semi-Infinite Optimization and Related Topics*, Langer Graben 17, 32657 Lemgo, Germany: Heldermann Verlag.

Zhang, T. and Woodroofe, M. (2003), “Credible and Confidence Sets for Restricted Parameter Spaces,” *Journal of Statistical Planning and Inference*, 115, 479 – 490.