# Estimating Stochastically Ordered Survival Functions via Geometric Programming

Johan Lim[*]       Seung Jean Kim[†]       Xinlei Wang [‡]

January 2009

## Abstract

Many procedures have been proposed to compute nonparametric maximum likelihood estimators (NPMLEs) of survival functions under stochastic ordering constraints. However, each of them is only applicable to a specific type of stochastic ordering constraint and censoring, and is often hard to implement. In this paper, we describe a general and flexible method based on geometric programming for computing the NPMLEs from right or interval censored data. To this end, we show that the monotonicity properties of the likelihood function and the stochastic ordering constraints considered in the literature allow us to reformulate the estimation problem as a geometric program (GP), a special type of mathematical optimization problem, which can be transformed to a convex optimization problem, and then solved globally and efficiently. We apply this GP-based method to real data examples to illustrate its generality in handling different types of ordering constraints and censoring. We also conduct simulation studies to examine its numerical performance for various sample sizes.

**Key words:**   Interval Censoring; Interior-point Algorithms; Nonparametric Maximum Likelihood Estimation; Right Censoring; Simple Stochastic Ordering; Uniform Stochastic Ordering.

[*]Department of Statistics, Seoul National University, Seoul, 151-747, Korea. (`johanlim@snu.ac.kr`)

[†]Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305. (`sjkim@stanford.edu`)

[‡]All correspondence to Xinlei Wang, Department of Statistical Science, Southern Methodist University, Dallas, TX 75275-0332 (`swang@mail.smu.edu`)

# 1 Introduction

The problem of estimating survival functions under stochastic ordering constraints arises naturally in many contexts. For example, one wants to examine the relationship between severity of cancer status and survival time. A patient with severe cancer is less likely to survive as long as one with a milder case. The survival functions for patients with different severity status therefore involve stochastic ordering constraints.

Two types of stochastic ordering have been widely studied in the literature: simple and uniform stochastic ordering. Let $S_1, \ldots, S_N$ be the survival functions of populations of interest. The functions are *simply stochastically ordered* if $S_1(t) \leq S_2(t) \leq \cdots \leq S_N(t)$ for every $t$ (Lehmann 1955), denoted by $S_1 \preceq_{\mathrm{S}} S_2 \preceq_{\mathrm{S}} \cdots \preceq_{\mathrm{S}} S_N$. A stronger version of simple stochastic ordering is uniform stochastic ordering (e.g., Dykstra, Kochar and Robertson, 1991; Rojo and Samaniego, 1991; Rojo and Samaniego, 1993; Mukerjee, 1996). The functions $S_1, \ldots, S_N$ are *uniformly stochastically ordered* if $S_i(t)/S_{i+1}(t)$ is nondecreasing in $t$ for every $i = 1, \ldots, N-1$, denoted by $S_1 \preceq_{\mathrm{U}} S_2 \preceq_{\mathrm{U}} \cdots \preceq_{\mathrm{U}} S_N$. Uniform ordering is equivalent to ordering in failure rates because $S_1 \preceq_{\mathrm{U}} S_2$ is equivalent to

$$\mathrm{P}\big(T_1 > s + t \big| T_1 > s\big) \leq \mathrm{P}\big(T_2 > s + t \big| T_2 > s\big), \quad \text{for every } s, t > 0,$$

where $T_1$ and $T_2$ are survival times from $S_1$ and $S_2$. An example given in Dykstra, Kochar and Robertson (1991) illustrates the practical meaning of uniform stochastic ordering, in which they considered survival times for two different medical treatments. If the better of the two treatments is administered initially (better in the sense of simple stochastic ordering), it might not be the better treatment if patients are examined at a later time point. However, if the survival times are uniformly ordered, then there is no doubt which treatment is preferred at any time point. Applications of uniform stochastic ordering can be easily found in reliability studies, clinical trials, and queues and other stochastic models (see e.g., Ross 1983, Stoyan 1983, and Capéraá 1988).

Without ordering constraints, nonparametric maximum likelihood estimators (NPMLEs) of survival functions are well studied. For uncensored data, the NPMLEs are simply the empirical survival functions. For right censored data, they are the Kaplan-Meier product estimators (Kaplan and Meier, 1958). For interval censored data, the NPMLEs under "case 1" interval censoring have a closed form expression derived from generalized isotonic regression

(Robertson et al. 1988); under "case 2" or the general interval censoring, there is no closed form available and several algorithms were proposed to compute the NPMLEs numerically (e.g., Efron, 1967; Wellner and Zhan, 1997; Jongbloed, 1998).

In the presence of stochastic ordering constraints, the NPMLEs are hard to compute generally. Substantial work in obtaining the constrained NPMLEs has been done for uncensored or right censored data while the previous work is rather scarce for interval censored data.

For simply stochastically ordered survival functions, Brunk, Franck, Handon, and Hogg (1966) studied the NPMLEs of two distributions from uncensored samples. Dykstra (1982) considered the same problem, but with right censored data, by reformulating it as a convex optimization problem, and then solving the Karush-Kuhn-Tucker (KKT) optimality conditions. His procedures cannot be extended directly to problems with more than two populations. Feltz and Dykstra (1985) proposed iterative procedures to solve the KKT conditions for the case of multiple populations. Dykstra and Feltz (1989) further considered the estimation problem with arbitrary simple partial ordering using the Fenchel duality theorem. Præstgaard and Huang (1996) showed that the asymptotic distribution of the NPMLEs for two sample problems is a limiting process related to the concave majorant of Brownian motion. The asymptotic distribution of the NPMLEs for more than two samples remains unknown.

For uniformly stochastically ordered survival functions, Dykstra, Kochar and Robertson (1991) considered the NPMLEs and the likelihood ratio testing with right censored observations from multiple populations. They obtained asymptotic properties including the consistency of the NPMLEs when the number of events at each time point increases. On the other hand, several authors pointed out possible inconsistency of the NPMLEs in other cases. Rojo and Samaniego (1991) and Mukerjee (1996) gave examples of one-sample and two-sample problems, in which the NPMLEs are inconsistent under uniform stochastic ordering. In these cases, as mentioned by Mukerjee (1996), nonparametric maximum likelihood estimation still provides a way of finding (rough) estimates, just like the method of moments in the parametric case.

Despite significant progress in constrained estimation of survival functions (for uncensored and right-censored data), obtaining actual estimates still remains a difficult task (Bami and Mukerjee, 2004). In fact, each of the existing procedures is only applicable to a specific

type of stochastic ordering constraint and censoring, and often hard to implement.

In this paper, we describe a new method for computing the NPMLEs of survival functions with right or interval censored observations from multiple populations, which can readily handle the two types of stochastic ordering constraints mentioned above. The method can also handle a mix of these constraints and partial ordering. To this end, we show that the monotonicity properties of the likelihood function and stochastic ordering constraints allow us to reformulate the estimation problem as a *geometric program* (GP), a special type of mathematical optimization problem, which can be transformed to a convex optimization problem, and then solved globally and efficiently by using interior-point methods. This GP-based estimation method has several important computational merits including high efficiency, great reliability and robustness, as will be discussed in Section 2.

It should be mentioned that GP-based methods had been used to estimate multinomial probabilities under constraints in several specific problems. Alldredge and Armstrong (1974) considered the problem of estimating overlap sizes created by interlocking sampling schemes. Mazumdar and Jefferson (1983) considered estimation of Bernoulli probabilities when sums of $k$ Bernoulli random variables are observed. Briker, Kortanek, and Xu (1997) considered constraints on local odds ratios when frequencies are observed along with a rectangular array.

The paper is organized as follows. Section 2 gives a brief introduction to GP. In Sections 3 and 4, we discuss the use of GP with right and interval censored data for finding the NPMLEs of stochastically ordered survival functions, respectively. To illustrate the application of our proposed method and examine its performance, we provide data examples as well as simulation studies. Section 5 concludes the paper.

## 2  An Overview of Geometric Programming

Here, we give a brief description of geometric programming. We refer the reader to the paper *A Tutorial on Geometric Programming* (Boyd, Kim, Vandenberghe and Hassibi, 2007) for more on geometric programming.

A geometric program is defined through two classes of functions: *monomials* and *posynomials*. Let $x_1, \ldots, x_n$ denote $n$ real positive variables, and $x = (x_1, \ldots, x_n)$ be a vector with

components $x_i$. A real valued function $f$ of $x$, with the form

$$f(x) = cx_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

where $c > 0$ and $a_i \in \mathbb{R}$, is called a *monomial function*, or informally, a *monomial* (of the variables $x_1, \ldots, x_n$). The sum of one or more monomials, namely

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

where each $c_k$ is positive, is called a *posynomial function* or simply a *posynomial*.

A *geometric program* (in posynomial form) is an optimization problem of the form

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \le 1, \quad i = 1, \ldots, m, \\
& g_j(x) = 1, \quad j = 1, \ldots, p,
\end{array}
\tag{1}
$$

where $f_0$ and $f_i$s are posynomials, $g_j$s are monomials, and the optimization variables $x_1,\ldots, x_n$ are all positive.

In general, GPs in posynomial form are not convex optimization problems, but they can be reformulated as convex problems (with convex objective and inequality constraint functions, and linear equality constraints) by change of variables and transformation of the objective and constraint functions. We use the logarithms $y_i = \log x_i$ of the original variables $x_i$, and minimize the logarithm $\log f_0$ of the original objective. We also replace the inequality constraints $f_i \le 1$ with $\log f_i \le 0$, and the equality constraints $g_j = 1$ with $\log g_j = 0$, to obtain the equivalent formulation

$$
\begin{array}{ll}
\text{minimize} & \log f_0(e^{y_1}, \ldots, e^{y_n}) \\
\text{subject to} & \log f_i(e^{y_1}, \ldots, e^{y_n}) \le 0, \quad i = 1, \ldots, m, \\
& \log g_j(e^{y_1}, \ldots, e^{y_n}) = 0, \quad j = 1, \ldots, p,
\end{array}
\tag{2}
$$

with variables $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Unlike the original GP, this transformed problem is convex, in which the equality constraints are linear.

Efficient solution methods for general convex optimization problems are well developed (Boyd and Vandenberghe, 2004; Nesterov and Nemirovsky, 1994; Norcedal and Wright, 1999; Wright, 1997; Ye, 1997). In particular, interior-point methods are very efficient. As stated in Boyd et al. (2007), standard interior-point algorithms can solve a GP with $1,000$ variables and $10,000$ constraints within a minute, on a small desktop computer; for sparse

problems each of whose constraints depends on only a modest number of the variables, far larger problems can be readily solved. Interior-point methods for GPs require essentially no algorithm parameter tuning and they require no starting point or initial guess of the optimal solution. Moreover, they always find the (globally) optimal solution, and when the problem is infeasible (i.e., the constraints are mutually inconsistent), they provide a certificate showing that no feasible point exists. We provide a brief description of a primal-dual interior-point method for GPs as supplemental material on the JCGS website.

Once a problem is formulated as a GP, it can be solved by a solver for convex problems, as described above. We can recover the solution to the original GP by taking the exponential of the solution to the equivalent convex problem. Standard solvers for convex problems require an user to describe the problem in a specific format and do not support flexibility in problem description. The user needs to transform the problem into a form suitable for the solvers. It is often convenient to use a parser that automates the transformation from a text description of the GP into a form suitable for a solver that is already available. Examples are CVX (Grant, Boyd and Ye, 2005), GGPLAB (Mutapcic, Koh, Kim and Boyd, 2006), and YALMIP (Löfberg, 2003), which have a simple interface to the MathWorks' MATLAB that recognizes and solves GPs. We will illustrate the merits of GP-based methods in our numerical studies, using GGPLAB.

# 3 The Case of Right Censoring

## 3.1 The Problem

Consider independent random samples from $N$ populations, with possibly right censored observations. Let $Y_{ik}$ be the failure time of the $k$th subject from the $i$th population, where $i = 1, \ldots, N$, $k = 1, \ldots, n_i$, and $n_i$ is the sample size for the $i$th population. We assume that $Y_{i1}, \ldots, Y_{in_i}$ are i.i.d. and have density $f_i(\cdot)$; and that each $Y_{ik}$ is independent of $C_{ik}$, the corresponding censoring time. The observed data from the $i$th population can be given in the form of $(x_{i1}, \delta_{i1}), \ldots, (x_{in_i}, \delta_{in_i})$, where $x_{ik} = \min(y_{ik}, c_{ik})$, and $\delta_{ik} = 1$ if $y_{ik} \leq c_{ik}$ and 0 otherwise. Let $\{t_1, \ldots, t_m\}$ be the union of all observed failure times pooled from the $N$ populations, with $t_1 < \cdots < t_m$. For the $i$th population, let $d_{ij}$ denote the number of failures that occur at time $t_j$, and $l_{ij}$ denote the number of losses (censored observations) within

the interval $[t_j, t_{j+1})$. Then, $n_{ij} = \sum_{r=j}^{m}(d_{ir} + l_{ir})$ is the number of observations in the $i$th population who survive just prior to $t_j$. We further assume those $l_{ij}$ losses occur at times $c_r^{ij}$, $r = 1, 2, \ldots, l_{ij}$.

Let $S_i$ denote the survival function of the $i$th population. We assume that $S_1, \ldots, S_N$ are stochastically ordered in an appropriate sense. Let $J \subset \{1, \ldots, N\} \times \{1, \ldots, N\}$ denote the set of pairs of indexes satisfying $S_i \preceq S_{i'}$, $(i, i') \in J$. Here, for $(i, i') \in J$, $S_i \preceq S_{i'}$ means a stochastic ordering constraint (e.g., simple or uniform) between $S_i$ and $S_{i'}$.

Given the observed data, the (true) likelihood function (up to multiplicative terms depending on the distributions of censoring times) can be written as

$$\mathcal{L}_T(S_1, \ldots, S_N) \propto \prod_{i=1}^{N} \prod_{k=1}^{n_i} f_i(x_{ik})^{\delta_{ik}} S_i(x_{ik})^{1-\delta_{ik}}. \tag{3}$$

Generally, it is infeasible to find the MLE by maximizing (3) over the class of continuous survival functions. To overcome this difficulty, we follow the previous work (Feltz and Dykstra, 1985; Dykstra, Kochar, and Robertson, 1991) and find the NPMLE by replacing the likelihood in (3) by an empirical version, where the density $f_i(x)$ is replaced by the jump of the survival functions $S_i(x-) - S_i(x)$:

$$\overline{\mathcal{L}}(S_1, \ldots, S_N) = \prod_{i=1}^{N} \prod_{j=1}^{m} \left\{ \left( S_i(t_j-) - S_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} S_i(c_r^{ij}) \right\}. \tag{4}$$

When there are no ordering constraints (i.e., $J$ is empty), to maximize $\overline{\mathcal{L}}$, it suffices to restrict attention to survival functions that are piecewise constant with jumps only at the observed failure times (see, e.g., Turnbull 1976, Wong and Yu 1999, Gentleman and Vandal 2001, Maathuis 2005). Then (4) can be reduced to

$$\mathcal{L}(S_1, \ldots, S_N) = \prod_{i=1}^{N} \prod_{j=1}^{m} \left( S_i(t_{j-1}) - S_i(t_j) \right)^{d_{ij}} S_i(t_j)^{l_{ij}}. \tag{5}$$

Under stochastic ordering constraints, the NPMLEs of survival curves are not necessarily piecewise constant so that a reduction from $\overline{\mathcal{L}}$ to $\mathcal{L}$ is not generally available. An example is given in Rojo and Samaniego (1991), which involves a survival function that is stochastically smaller than a fixed known survival function. In this example, due to the ordering constraints, the NPMLE should be strictly monotone over all $t$; thus the NPMLE is not

piecewise constant and the maximization should be done over the class of strictly mono-tone survival functions. However, even in this case or many other cases, the NPMLE over piecewise constant functions can often provide a good approximation to the true NPMLE, when the sample size is not small, in the sense that the distance between the two estimates converges to 0 almost surely (see Appendix A).

As a result, we find the estimates of the survival functions by solving the following optimization problem:

$$
\begin{aligned}
\text{maximize} \quad & \mathcal{L}(S_1, \ldots, S_N) \\
\text{subject to} \quad & S_i \preceq S_{i'}, \quad (i, i') \in J, \\
& S_i \in \mathcal{C}(t_1, \ldots, t_m), \quad i = 1, \ldots, N.
\end{aligned}
\tag{6}
$$

where $\mathcal{L}(S_1, \ldots, S_N)$ is given in (5), and $C(t_1, \ldots, t_m)$ denotes the set of functions that are constant on the intervals $[t_j, t_{j+1})$.

## 3.2 The GP-Based Method

We describe the GP formulation of (6) below. We define new variables

$$
p_{ij} = S_i(t_j)/S_i(t_{j-1}), \qquad q_{ij} = 1 - p_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, m,
$$

where $S_i(t_0) = 1$ for all $i = 1, \ldots, N$. Noting that $S_i(t_j) = \prod_{r=1}^{j} p_{ir}$, the objective of (6) becomes

$$
\mathcal{L}(S_1, \ldots, S_N) = \prod_{i=1}^{N} \prod_{j=1}^{m} q_{ij}^{d_{ij}} p_{ij}^{l_{ij}} \prod_{r<j} p_{ir}^{d_{ij}+l_{ij}}.
$$

In terms of the optimization variables $p_{ij}$s and $q_{ij}$s, (6) is equivalent to

$$
\begin{aligned}
\text{minimize} \quad & \prod_{i=1}^{N} \prod_{j=1}^{m} q_{ij}^{-d_{ij}} p_{ij}^{-l_{ij}} \prod_{r<j} p_{ir}^{-d_{ij}-l_{ij}} \\
\text{subject to} \quad & p_{ij} + q_{ij} = 1, \quad i = 1, \ldots, N, \quad j = 1, \ldots, m, \\
& f_k(p_{11}, \ldots, p_{Nm}) \leq 1, \quad k = 1, \ldots |J|,
\end{aligned}
\tag{7}
$$

where $f_k$s correspond to the constraint functions expressed in terms of $p_{ij}$s. Here $|J|$ denotes the cardinality of $J$, the index set of order constraints introduced in Section 3.1.

The objective of (7) is a monomial of $p_{ij}$s and $q_{ij}$s. As will be shown later in this section, the inequality constraint functions $f_k$s are posynomials of the variables. In a GP, the only equality constraints allowed involve monomials. Since the equality constraints of

(7) are posynomial equalities, (7) is not a GP. We replace those posynomial equalities with inequalities to obtain the GP

$$
\begin{aligned}
\text{minimize} \quad & \prod_{i=1}^{N} \prod_{j=1}^{m} q_{ij}^{-d_{ij}} p_{ij}^{-l_{ij}} \prod_{r<j} p_{ir}^{-d_{ij}-l_{ij}} \\
\text{subject to} \quad & p_{ij} + q_{ij} \leq 1, \quad i = 1, \ldots, N, \; j = 1, \ldots, m, \\
& f_k(p_{11}, \ldots, p_{Nm}) \leq 1, \quad k = 1, \ldots |J|.
\end{aligned}
\tag{8}
$$

We now establish the equivalence between (7) and (8). Since the posynomials $p_{ij} + q_{ij}$ are monotone increasing in the variables, the objective is decreasing in the variables, and $q_{ij}$s do not appear in any of the posynomial inequality constraint functions $f_k$, we can see that at the optimal point, the inequality constraints $p_{ij} + q_{ij} \leq 1$ must be tight. In other words, for any $\overline{p}_{ij}, \overline{q}_{ij}$, if at least one of the inequalities $\overline{p}_{ij} + \overline{q}_{ij} < 1$ is strict, then they cannot be optimal since $\hat{p}_{ij}, \hat{q}_{ij}$ with $\hat{p}_{ij} = \overline{p}_{ij}$ and $\hat{q}_{ij} = \overline{q}_{ij} + (1 - p_{ij} - q_{ij})$ are feasible for (8) but

$$
\prod_{i=1}^{N} \prod_{j=1}^{m} \overline{q}_{ij}^{-d_{ij}} \; \overline{p}_{ij}^{-l_{ij}} \prod_{r<j} \overline{p}_{ir}^{-d_{ij}-l_{ij}} > \prod_{i=1}^{N} \prod_{j=1}^{m} \hat{q}_{ij}^{-d_{ij}} \hat{p}_{ij}^{-l_{ij}} \prod_{r<j} \hat{p}_{ir}^{-d_{ij}-l_{ij}}.
$$

In summary, we can solve the original problem (6), through the tractable GP (8).

We show that the stochastic ordering constraints widely studied in the literature are compatible with GP.

- *Simple ordering.* The constraint $S_1 \preceq_{\text{S}} \cdots \preceq_{\text{S}} S_N$ can be written as $S_i(t_j) < S_{i+1}(t_j)$ for $i = 1, \ldots, N-1$ and $j = 1, \ldots, m$. In terms of $p_{ij}$s, the ordering constraint can be expressed as $\prod_{r=1}^{j} p_{ir} \leq \prod_{r=1}^{j} p_{i+1,r}$, or equivalently as the monomial inequality $\prod_{r=1}^{j} p_{ir} p_{i+1,r}^{-1} \leq 1$, for $i = 1, \ldots, N-1$, and $j = 1, \ldots, m$.

- *Uniform ordering.* The constraint $S_1 \preceq_{\text{U}} \cdots \preceq_{\text{U}} S_N$ can be written as $S_i(t_j)/S_{i+1}(t_j) \leq S_i(t_{j+1})/S_{i+1}(t_{j+1})$, for $i = 1, \ldots, N-1$ and $j = 1, \ldots, m-1$. Again with $p_{ij}$s, the ordering constraint can be expressed as $p_{ij} \leq p_{i+1,j}$, or equivalently the monomial inequality $p_{ij} p_{i+1,j}^{-1} \leq 1$, for $i = 1, \ldots, N-1$ and $j = 1, \ldots, m-1$.

Besides the above two types of ordering constraints, the GP-based method can handle any GP-compatible constraints, such as upper or lower bounds on the survival function at some time points and different types of ordering constraints in multiple sample problems.

9

## 3.3 Examples

We first illustrate the generality and flexibility of our GP-based method with examples of right-censored data from the previous literature. We applied it to the oropharynx cancer data in Kalbfleisch and Prentice (1980). We used GGPLAB (Mutapcic et al., 2006) to find the NPMLEs on a 1.80GHz Pentium IV computer. The GGPLAB codes are available from the JCGS website, which can be readily modified to handle similar problems.

The data set involves a large clinical trial carried out by the Radiation Therapy Oncology Group in the United States, where patients' survival (or censoring) times were recorded in days. It also contains variables that are potentially related to the survival time, including sex, T staging and N staging. Here, T and N staging measure the extent of the tumor at the primary site and at regional lymph nodes, respectively. Specifically, T = 1 indicates a small primary tumor, T = 4 indicates a massive tumor, and T = 2 and T = 3 indicate two intermediate sizes. Similarly, N = 0 indicates no clinical evidence of lymph node metastasis and N = 1, 2, 3 indicate increasing magnitudes of lymph node deterioration. Following Feltz and Dykstra (1985) and (1989), we only used a small part of the data in our analysis, where all female patients and patients with (T,N) = (1, 0), (1, 1), (2, 0) and (2, 1) were excluded.

### Simple stochastic ordering

Our first example considers the problem from Feltz and Dykstra (1985), which assumes simple stochastic ordering among the groups (T,N) = (3, 1), (3, 2), and (3, 3), denoted as population 1, 2 and 3, respectively. Due to the nature of the disease, it would be reasonable to have the ordering with respect to lymphatic involvement: $S_3 \preceq_S S_2 \preceq_S S_1$. However, the unrestricted NPMLEs (Kaplan-Meier estimates) of survival functions, given in Figure 1(a), reveal some reverse ordering between $S_1$ and $S_2$ at times 347 and 1092, and also between $S_2$ and $S_3$ at time 105.

The NPMLEs subject to the above constraint obtained by the GP method are shown in Figure 1(b). The duality gap is set to $10^{-6}$, which means that the maximum likelihood value we found is within the gap $10^{-6}$ from the true value. In this example, the total number of observed failure times is $m = 41$. The number of Newton steps required to solve the GP is 44 and the CPU time is about 1 second.

Although we solved the same optimization problem as in Feltz and Dykstra (1985), our

estimates from the GP method show obvious differences from those of Feltz and Dykstra (1985). First, the GP estimates provide a higher likelihood value than the Feltz-Dykstra estimates; that is, $-136.91$ versus $-137.19$. Secondly, the GP estimates are closer to the Kaplan-Meier estimates than the Feltz-Dykstra estimates. For example, the GP estimate of $S_1(929)$ is 0.3845, which is closer to the Kaplan-Meier estimate 0.364 than the estimate 0.4881 by Feltz and Dykstra. The same phenomenon can also be observed at several other time points, as shown in Figure 1(c) which plots the three estimates of $S_2$ obtained by the Kaplan-Meier method, the Feltz-Dykstra method, and the GP method.

## Partial stochastic ordering

The second example is from Dykstra and Feltz (1989), which involves partial ordering among groups $(T,N) = (3,0)$, $(3,3)$, $(4,0)$, and $(4,3)$, denoted by population $1 \sim 4$, respectively. For the same T stage groups, the survival curve of $N = 3$ would be expected to be smaller than that of $N = 0$. Similarly, for the same N stage groups, the survival curve of $T = 4$ would be expected to be smaller than that of $T = 3$. Hence, $S_1, S_2, S_3$ and $S_4$ would satisfy $S_4 \preceq_S S_2 \preceq_S S_1$, and $S_4 \preceq_S S_3 \preceq_S S_1$. Figure 2(a) shows the unrestricted NPMLEs. As pointed out in Dykstra and Feltz (1989), the above order assumptions are violated at various points.

Using the GP formulation discussed in Section 3.2, the NPMLEs in this example are the solution to the GP

$$
\begin{aligned}
\text{minimize} \quad & \prod_{i=1}^{4}\prod_{j=1}^{75} q_{ij}^{-d_{ij}} p_{ij}^{-l_{ij}} \prod_{r=1}^{j} p_{ir}^{-d_{ir}-l_{ir}} \\
\text{subject to} \quad & p_{ij} + q_{ij} \leq 1, \quad i = 1,2,3,4, \quad j = 1,\ldots,75, \\
& \prod_{r=1}^{j} p_{4r} p_{3r}^{-1} \leq 1, \quad \prod_{r=1}^{j} p_{2r} p_{1r}^{-1} \leq 1, \quad j = 1,\ldots,75, \\
& \prod_{r=1}^{j} p_{3r} p_{1r}^{-1} \leq 1, \quad \prod_{r=1}^{j} p_{4r} p_{2r}^{-1} \leq 1, \quad j = 1,\ldots,75.
\end{aligned}
$$

It took around 2.1 seconds and 49 Newton steps to solve this GP.

Figure 2(b) plots the estimated survival functions via the GP method and they are similar to the results of Dykstra and Feltz (1989).

**Mix of two types of stochastic ordering**

To further illustrate the flexibility in ordering constraints that the GP method can handle, we consider both simple and uniform stochastic ordering constraints simultaneously, using the same data from Dykstra and Feltz (1989). We assume the uniform stochastic ordering between the first, $(T, N) = (3, 0)$, and the third population, $(T, N) = (4, 0)$, but assume the simple stochastic ordering between the third $(T, N) = (4, 0)$ and the fourth population $(T, N) = (4, 3)$; that is, $S_4 \preceq_S S_3 \preceq_U S_1$.

In this example, the NPMLEs of $S_1$, $S_3$ and $S_4$ are the solution to the following GP:

$$\text{minimize} \quad \prod_{i=1,3,4} \prod_{j=1}^{49} q_{ij}^{-d_{ij}} p_{ij}^{-l_{ij}} \prod_{r=1}^{j} p_{ir}^{-d_{ir}-l_{ir}}$$
$$\text{subject to} \quad p_{ij} + q_{ij} \leq 1, \quad i = 1, 3, 4, \quad j = 1, \ldots, 49,$$
$$\prod_{r=1}^{j} p_{4r} p_{3r}^{-1} \leq 1, \quad p_{3j} p_{1j}^{-1} \leq 1, \quad j = 1, \ldots, 49.$$

It took 1.8 seconds and 49 Newton iterations to solve this GP.

Figure 2(c) plots the survival curves estimated by the GP method. The figure illustrates the difference between simple and uniform stochastic ordering. The difference between the NPMLEs of $S_1(t)$ and $S_3(t)$ (uniform stochastic ordering) increases as $t$ increases, whereas the difference between $S_3(t)$ and $S_4(t)$ (simple stochastic ordering) does not.

## 3.4  Numerical Performance

We conducted two simulation studies to examine the numerical performance of the GP method.

We first compare the computing time of the GP method to that of the algorithm by Feltz and Dykstra (1985) in estimating simply stochastically ordered survival functions, to shed light on the potential gain in computational efficiency. We set the sample size $n = 100, 200, 300$ and 400 and for each size, we randomly generated 100 date sets. Each data set involves four populations with equal sample size $n/4$ whose survival times were generated from exponential distributions with means $10, 20, 30$ and 40.

Figure 3(a) plots the mean CPU time ($\pm 2$ s.d.) versus sample size for the two methods. Clearly, our GP method is much faster than the Feltz-Dykstra method, and the difference

becomes larger when $n$ gets larger. This is because for $N$-sample problems $(N > 2)$, the Feltz-Dykstra method iteratively applies the pairwise algorithm designed for two-sample problems by Dykstra (1982) until it converges. The algorithm for two samples solves a number of nonlinear equations that is proportional to the number of violations of order constraints. Due to its large number of iterations, the Feltz-Dykstra method needs to solve many nonlinear equations iteratively, especially for large $N$ or $n$. In contrast, the GP method only needs to solve a single optimization problem.

The sizes of the problems considered above are modest. In our second experiment, to study the performance of the GP method for relatively large problems, we chose $n$ to be 200, 400, 600, 800, and 1000. Again, 100 data sets were generated for each $n$. Each data set involves two populations with equal sample size $n/2$ whose survival times were generated from exponential distributions with means 20 and 50. Thus, the two survival curves are stochastically ordered, both simply and uniformly.

We used the GP method to compute the NPMLEs under simple and uniform stochastic ordering constraints, respectively. Figure 3(b) shows the mean CPU time ($\pm 2$ s.d.) for each sample size. We can see that the computing speed for estimating simply ordered survival functions is much slower and the time grows much faster than that for uniformly ordered survival functions. This is not surprising since uniform ordering yields a sparse GP while simple ordering yields a much less sparse GP.

Before we end this section, we should mention that in our simulation setting, we did not consider right censored observations. In fact, right censoring reduces the number of optimization variables as well as the number of constraints (recall in our GP formulation, $m$ is the number of unique failure times observed). Thus, the setting without censored observations is, in general, the worst case, in terms of computing complexity.

# 4   The Case of Interval Censoring

This section shows how the GP method can be used to estimate survival functions under various constraints from interval censored data. Such data arise when a failure time $Y$ can not be observed, but can only be determined to be within an interval $(L, R)$ obtained from a sequence of examination times. There are several types of interval censoring data in

accordance with the number of observation times per individual: "case k" interval censoring refers to the case with $k$ observation times per individual (Groeneboom and Wellner, 1992; Wellner, 1995). In this section, we focus on estimation of survival functions with "case 1" interval censored data, and give a brief discussion about "case 2" data.

## 4.1   The Problem and the GP-Based Method

Consider "case 1" interval censored data from $N$ populations. Let $Y_{ik}$ be the unobserved failure time and $C_{ik}$ be the examination (or observation) time of the $k$th subject from the $i$th population, for $i = 1, \ldots, N$ and $k = 1, \ldots, n_i$. Let $S_i(t)$ be the survival function of $Y_{ik}$ and $g_i(c)$ be the density of $C_{ik}$. We assume all $Y_{ik}$s and $C_{ik}$s are independent. The only knowledge about the failure time $Y_{ik}$ is whether it has occurred before $C_{ik}$ or not. So the observed data are of the form $\{(c_{ik}, \delta_{ik}), \ i = 1, \ldots, N; \ k = 1, \ldots, n_i\}$, where $\delta_{ik} = \mathrm{I}(y_{ik} \leq c_{ik})$, and $\mathrm{I}(\cdot)$ is the indicator function.

We further let $\{t_1, \ldots, t_m\}$ be the unique ordered observations of $\{0, c_{ik}, \ i = 1, \ldots, N; \ k = 1, \ldots, n_i\}$. Let $n_{ij} = \sum_{k=1}^{n_i} \mathrm{I}(c_{ik} = t_j)$ be the number of subjects observed at $t_j$ in population $i$, and $r_{ij} = \sum_{k=1}^{n_i} \delta_{ik} \mathrm{I}(c_{ik} = t_j)$ be the number of subjects who have failed before time $t_j$ among the $n_{ij}$ subjects, for $i = 1, 2, \ldots, N$ and $j = 1, \ldots, m$. Noting that $r_{ij} \sim \mathrm{Binomial}\big(n_{ij}, 1 - S_i(t_j)\big)$, the likelihood function (up to a normalizing constant) is

$$\mathcal{L}\big(S_1, S_2, \ldots, S_N\big) \;=\; \prod_{i=1}^{N}\prod_{j=1}^{m} \big[S_i(t_j)\big]^{n_{ij}-r_{ij}} \big[1 - S_i(t_j)\big]^{r_{ij}} \big[g_i(t_j)\big]^{n_{ij}}, \tag{9}$$

where $S_i(t_1) \geq S_i(t_2) \geq \cdots \geq S_i(t_m)$ for each $i$. Again, (9) depends on $S_i$s only through their values at the observation times. So we follow the convention that the NPMLEs are assumed to be piecewise constant with jumps only at $t_j$s. The NPMLEs of $S_i$s (without any other order constraints) can be computed by solving a generalized isotonic regression problem via the pool-adjacent-violators algorithm (Robertson, et al. 1988; Sun 2006).

As in the case of right censoring, the maximization of the likelihood (9) over piecewise constant survival functions can be formulated into a GP. Let $p_{ij} = 1 - S_i(t_j)$ and $q_{ij} = S_i(t_j)$. Following a similar argument given in Section 3.2, we can again show that the equality constraints $p_{ij} + q_{ij} = 1$ can be relaxed to $p_{ij} + q_{ij} \leq 1$ so that we only need to solve the

following GP to compute the NPMLEs:

$$
\begin{aligned}
\text{maximize} \quad & \prod_{i=1}^{N} \prod_{j=1}^{m} p_{ij}^{r_{ij}} q_{ij}^{n_{ij}-r_{ij}} \\
\text{subject to} \quad & q_{i1} \geq q_{i2} \geq \cdots \geq q_{im}, \quad \text{for } i = 1, 2, \ldots, N, \\
& p_{ij} + q_{ij} \leq 1, \quad \text{for } i = 1, 2, \ldots, N, \ j = 1, 2, \ldots, m.
\end{aligned} \tag{10}
$$

The advantage of our GP-based method, again, lies in its flexibility of handling a wide class of GP-compatible constraints on survival functions. For example, the simple stochastic ordering constraint between populations $i$ and $i'$, $S_i \preceq_{\mathrm{S}} S_{i'}$, can be written as $q_{ij} q_{i'j}^{-1} \leq 1$ for $j = 1, \ldots, m$; and the uniform stochastic ordering constraint $S_i \preceq_{\mathrm{U}} S_{i'}$ can be written as $q_{ij} q_{i',j+1}^{-1} q_{i'j}^{-1} q_{i,j+1}^{-1} \leq 1$ for $j = 1, \ldots, m-1$. Other GP-feasible examples include $\alpha_{ij} \leq S_i(t_{j+1})/S_i(t_j) \leq \beta_{ij}$ (i.e., $\alpha_{ij} \leq q_{i,j+1} q_{ij}^{-1} \leq \beta_{ij}$) for constants $\alpha_{ij}$s and $\beta_{ij}$s, and $S_i(t_{j+1})/S_i(t_j) \leq S_i(t_j)/S_i(t_{j-1})$ (i.e., $q_{i,j-1} q_{ij}^{-2} q_{i,j+1} \leq 1$) for some $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, m\}$. All the above examples involve monomials of the optimization variables $q_{ij}$s, but not $p_{ij}$s so that the relaxation from $p_{ij} + q_{ij} = 1$ to $p_{ij} + q_{ij} \leq 1$ can be done as before.

To summarize, the NPMLEs of the survival functions under GP-compatible constraints can be obtained by solving the following GP,

$$
\begin{aligned}
\text{maximize} \quad & \prod_{i=1}^{N} \prod_{j=1}^{m} p_{ij}^{r_{ij}} q_{ij}^{n_{ij}-r_{ij}} \\
\text{subject to} \quad & q_{i1} \geq q_{i2} \geq \cdots \geq q_{im}, \quad \text{for } i = 1, 2, \ldots, N, \\
& p_{ij} + q_{ij} \leq 1, \quad \text{for } i = 1, 2, \ldots, N, \ j = 1, 2, \ldots, m, \\
& f_k(q_{11}, \ldots, q_{Nm}) \leq 1, k = 1, 2, \ldots, K,
\end{aligned} \tag{11}
$$

where $f_k$s correspond to the constraint functions expressed in terms of $q_{ij}$s, and $K$ denotes the number of such constraints. The computational cost of (11) is not much different from that of the unconstrained problem (10). In contrast, the classical isotonic regression method can only handle (10).

We now discuss the use of GP with "case 2" interval censored data briefly. Unlike "case 1" interval censoring, computing the NPMLEs from "case 2" data is quite involved even without imposing any order constraints among survival functions, for which several estimation approaches have been proposed in the literature. Among the earliest was the expectation-maximization (EM) algorithm proposed by Efron (1967). The maximization step in the EM algorithm estimates multinomial probabilities, which has been shown to be equivalent to solving a GP (Mazumdar and Jefferson, 1983). Although it is generally slower than the iterative convex minorant (ICM) algorithm by Jongbloed (1998) and a hybrid algorithm of ICM and EM by Wellner and Zhan (1997), the EM algorithm, when combined

15

with GP in the maximization step, can allow GP-compatible constraints easily, whereas the other algorithms do not have such flexibility. We omit details for brevity.

## 4.2 An Example

We consider the lung tumor data given in Hoel and Walberg (1972) and Section 1.2.1 in Sun (2006). The data consist of 144 male mice from two treatment groups, conventional environment (population 1) and germ free environment (population 2). For each animal, the death time and the existence of lung tumor were recorded when it was found dead. So the data are "case 1" interval censored. It is natural to assume the time to lung tumor of population 1 is stochastically earlier than that of population 2, that is $S_1(t) \leq S_2(t)$ for every $t$. Our goal is to obtain the NPMLEs of $S_1$ and $S_2$ under the simple stochastic ordering constraint.

We applied the proposed method and plotted the estimated survival functions with and without the ordering constraint in Figure 4(a) and Figure 4(b). We can see that they show obvious difference; violations to the constraint of simple stochastic ordering disappear when it was imposed. To our best knowledge, no existing method can compute the constrained NPMLEs from interval censored data.

## 5 Discussion

In this paper, we have described the GP-based approach for estimating survival functions under stochastic ordering constraints. We have illustrated the proposed approach with several data examples and simulation studies. We have shown that, as opposed to the existing approaches in the literature, the GP-based approach is flexible and general; it can be used with right-censored data or case 1 and 2 interval-censored data; it can readily handle the simple and uniform ordering constraints and moreover can handle a mix of the two constraints. Another merit is that it can rely on interior-point methods for GPs, which are very robust, in addition to being fast. Finally, it can rely on a GP solver and parser such as GGPLAB, which makes straightforward the job of translating the problem of survival function estimation into a standard GP format.

Finally, we mention that for interval censoring, the GP formulations for case 1 and 2

data do not work for the general mixed case. Hence, it would be of interest to investigate it as a topic of future research.

# Acknowledgment

# References

Alldredge, J. and Armstrong, D. (1974), Maximum likelihood estimation for multinomial distribution using geometric programming. *Technometrics*, **16**, 585-587.

Barmi, H. and Mukerjee, H. (2005), Inferences under a stochastic ordering constraint: the k-sample case. *Journal of Applied Mathematics and Decision Sciences*, **100**, 252-261

Borwein, J. and Lewis, A. (2000), *Convex Analysis and Nonlinear Optimization*, Springer.

Boyd, S., Kim, S., Vandenberghe, L., and Hassibi, A. (2007), A tutorial on geometric programming. *Optimization and Engineering*, **8**, 67-127.

Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*. Cambridge University Press.

Briker, D., Kortanek, K., and Xu, L. (1997), Maximum likelihood estimates with order restrictions on probabilities and odds ratios: a geometric programming approach. *Journal of Applied Mathematics and Decision Sciences*, **1**, 53-65.

Brunk, H., Franck, W., Hanson, D,. L., and Hogg, R. (1966), Maximum likelihood estimation of the distribution functions of two stochastically ordered random variables. *Journal of the American Statistical Association*, **61**, 1067-1080.

Capéraá. P. (1988). Tail ordering and asymptotic efficiency of rank tests. *Annals of Statistics*, **16**, 470-478.

Dykstra, R. (1982), Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association*, **77**, 621-628.

Dykstra, R. and Feltz, C. (1989), Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. *Biometrika*, **76**, 331-341.

Dykstra, R., Kochar, S., and Robertson, T. (1991), Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics*, **19**, 870-888.

Dykstra, R., Kochar, S., and Robertson, T. (1995), Inference for likelihood ratio ordering in the two-sample problem. *Journal of the American Statistical Association*, **90**, 1034-1040.

Efron, B. (1967), The two sample problem with censored data. In *Proceeding of 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 831-853.

Feltz, C. and Dykstra, R. (1985), Maximum likelihood estimator of the survival functions of $N$ stochastically ordered distributions. *Journal of the American Statistical Association*, **80**, 1012-1019.

Gentleman, R. and Vandal, A. C. (2001). Computational algorithms for censored-data problems using intersection graphs. *Journal of Computational and Graphical Statistics*. **10**, 403-421.

Grant, M., Boyd S., and Ye, Y. (2005), cvx: Matlab software for disciplined convex programming. Available from `www.stanford.edu/~boyd/cvx/`.

Groeneboom, P. and Wellner, J. A. (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Basel.

Hoel, D.G. and Walburg, H.E. (1972). Statistical analysis of survival experiments. *Journal of National Cancer Institute*, **49**, 361-372.

Johansen, S. (1978), The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics*, **5**, 195-199.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, **7**, 310-321.

Kalbfleisch, J. and Prentice, R. (1958), *The statistical analysis of failure rate data*, New York: Wiley.

Kaplan, E. and Meier, P. (1958), Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, **27**, 887-906.

Lehmann, E. (1955), Ordered families of distributions. *Annals of Mathematical Statistics*, **26**, 399-419.

Löfberg, J. (2003), YALMIP. Yet Another LMI Parser. Version 2.4. Available from `http://control.ee.ethz.ch/~joloef/yalmip.php`.

Maathuis, M. H. (2005). Reduction algorithm for the MLE for the distribution function of bivariate interval censored data. *Journal of Computational and Graphical Statistics*. **14**, 352C362.

Mazumdar, M. and Jefferson, T. (1983). Maximum likelihood estimates for multinomial probabilities via geometric programming. *Biometrika*, **70**, 257-261.

MOSEK ApS (2002), *The MOSEK Optimization Tools Version 2.5. User's Manual and Reference*. Available from `www.mosek.com`.

Mukerjee, H. (1996), Estimation of survival functions under uniform stochastic ordering, *Journal of the American Statistical Association*, **91**, 1694-1689.

Mutapcic A., Koh K., Kim S., and S. Boyd (2006), GGPLAB: A Matlab toolbox for geometric programming. Available from `www.stanford.edu/~boyd/ggplab/`.

Nesterov, Y. and Nemirovsky, A. (1994), *Interior-Point Polynomial Methods in Convex Programming*, volume 13 of *Studies in Applied Mathematics*. SIAM Philadelphia, PA.

Norcedal, J. and Wright, S. (1999), *Numerical Optimization*, Springer Series in Operations Research. Springer.

Præstgaard, J. and Huang, J. (1996), Asymptotic theory of nonparametric estimation of survival curves under order restrictions. *Annals of Statistics*, **24**, 1679-1716.

Robertson, T., Wright, F., and Dykstra, R. (1988), *Order Restricted Statistical Inference*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons.

Rojo, J. and Samaniego, F. (1991), On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statistica and probability Letters*, **11**, 267-271.

Rojo, J. and Samaniego, F. (1993), On estimating a survival curve subject to a uniform stochastic ordering constraint. *Journal of the American Statistical Association*, **88**, 566-572.

Ross, S.M. (1983). *Stochastic Processes*. Wiley, New York.

Schick, A. and Yu. Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics*, **27**, 45-55.

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, New York.

Sun, J. (2006), *The statistical analysis of interval-censored failure time data*. Springer, New York.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored, and truncated data. *Journal of the Royal Statistical Society B*. **38**, 290C29.5

Wellner, J.A. (1995). Interval censoring, case 2: alternative hypothesis. In *Analysis of Censored Data*, vol 27 of *IMS Lecture Notes - Monograph Series*. IMS, Hayward, 271-291.

Wellner, J.A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from cnesored data. *Journal of the American Statistical Association*, **92**, 945-959.

Wong, G. Y. and Yu, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *Journal of Multivariate Analysis.* **69**, 155-166.

# A Piecewise constant approximation to the NPMLE

In this section, we show that the NPMLE over the family of piecewise constant functions ( say $\overline{\mathbf{S}}_m^{\mathrm{pc}}$) provides a good approximation to the true NPMLE (say $\overline{\mathbf{S}}_m$) asymptotically. We assume that both survival times and censoring times of each population have positive densities on a compact set so that the maximum space between two nearest observations, $\max_j |t_{j+1} - t_j|$, converges to 0 a.s., as $m$ increases.

Let $\mathcal{C}$ be the space of all survival functions $\mathbf{S} = (S_1, S_2, \ldots, S_N)$, and $\mathcal{C}(t_1, t_2, \ldots, t_m)$ be the space of piecewise constant survival functions with break points at $t_1, t_2, \ldots, t_m$. Here, we do not consider trivial survival functions which have all the mass at a single point. We equip these spaces $\mathcal{C}$ and $\mathcal{C}(t_1, t_2, \ldots, t_m)$ with the $\mathcal{L}_\infty$ metric defined as

$$d(\mathbf{S}, \mathbf{T}) = \sum_{i=1}^{N} \sup_x |S_i(x) - T_i(x)|.$$

We make two more assumptions on the likelihood function and the NPMLE. We assume that the NPMLE, $\overline{\mathbf{S}}_m$, is unique on the given space $\mathcal{C}$. As in previous literature (Feltz and Dykstra, 1985; Dykstra, Kochar, and Robertson, 1991), we assume the likelihood function is discrete only on observation points in the generalized maximum likelihood framework developed by Kiefer and Wolfowitz (1956). Thus, the likelihood function to be maximized has the form (4).

We will establish the following:

(i) The likelihood function (4) is continuous in $\mathbf{S}$.

(ii) As $m$ increases,

$$d\left(\overline{\mathbf{S}}_m, \mathcal{C}(t_1, t_2, \ldots, t_m)\right) = \min_{\mathbf{T} \in \mathcal{C}(t_1, t_2, \ldots, t_m)} d\left(\overline{\mathbf{S}}_m, \mathbf{T}\right)$$

converges to zero almost surely.

These two results show that $d(\overline{\mathbf{S}}_m^{\mathrm{pc}}, \overline{\mathbf{S}}_m)$ converges to 0 a.s., as $m$ increases. ¿From (ii), we can take a sequence $\overline{\mathbf{T}}_m$ in $\mathcal{C}(t_1, t_2, \ldots, t_m)$ which converges to $\overline{\mathbf{S}}_m$. By the continuity of the likelihood function, their likelihood values also converge to that of $\overline{\mathbf{S}}_m$. Thus, if $d(\overline{\mathbf{S}}_m^{\mathrm{pc}}, \overline{\mathbf{S}}_m)$ does not converge to 0 a.s., the likelihood value of $\overline{\mathbf{S}}_m^{\mathrm{pc}}$ becomes smaller than that of $\overline{\mathbf{T}}_m$ for

22

sufficiently large $m$. It contradicts the definition of $\overline{\mathbf{S}}_m^{\mathrm{pc}}$ as the maximizer of the likelihood over $\mathcal{C}(t_1, t_2, \ldots, t_m)$.

We first show that the likelihood function is continuous in $\mathbf{S}$. Let

$$
\theta = \max_{i,j} \max \left\{ \left( S_i(t_j-) - S_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} S_i(c_r^{ij}), \left( T_i(t_j-) - T_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} T_i(c_r^{ij}) \right\},
$$

which is smaller than 1 for non-trivial survival functions $\mathbf{S}$. Then, we have

$$
\left| \overline{\mathcal{L}}(\mathbf{S}) - \overline{\mathcal{L}}(\mathbf{T}) \right|
$$

$$
= \left| \prod_{i=1}^{N} \prod_{j=1}^{m} \left\{ \left( S_i(t_j-) - S_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} S_i(c_r^{ij}) \right\} - \prod_{i=1}^{N} \prod_{j=1}^{m} \left\{ \left( T_i(t_j-) - T_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} T_i(c_r^{ij}) \right\} \right|
$$

$$
\leq \theta^{Nm-1} \sum_{i=1}^{N} \sum_{j=1}^{m} \left| \left( S_i(t_j-) - S_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} S_i(c_r^{ij}) - \left( T_i(t_j-) - T_i(t_j) \right)^{d_{ij}} \prod_{r=1}^{l_{ij}} T_i(c_r^{ij}) \right|
$$

$$
\leq \theta^{Nm-1} \cdot 3m \cdot d(\mathbf{S}, \mathbf{T})
$$

$$
\leq d(\mathbf{S}, \mathbf{T}),
$$

for sufficiently large $m$. Thus, $\mathcal{L}(\mathbf{S})$ is continuous in $\mathbf{S}$.

We now show that

$$
\lim_{m \to \infty} d\left( \overline{\mathbf{S}}_m, \mathcal{C}(t_1, t_2, \ldots, t_m) \right) = 0 \quad a.s.
$$

Recall that $\overline{\mathbf{S}}_m = \left( \overline{S}_{1,m}, \overline{S}_{2,m}, \ldots, \overline{S}_{N,m} \right)$ can have discontinuity points only at $t_1, t_2, \ldots, t_m$. We define a function $\mathbf{T}_m = \left( T_{1,m}, T_{2,m}, \ldots, T_{N,m} \right)$ in $\mathcal{C}(t_1, t_2, \ldots, t_m)$ as

$$
T_{i,m}(t) = \overline{S}_{i,m}(t_j), \quad t \in [t_j, t_{j+1}),
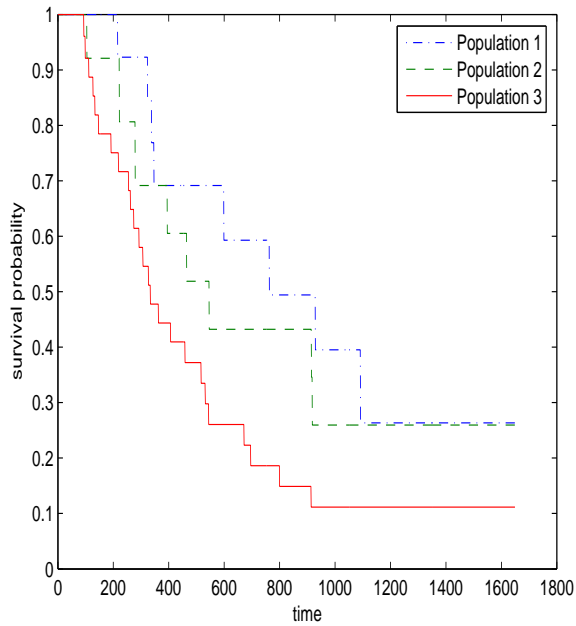$$

for $j = 0, 1, \ldots, m$ with $t_0 = 0$ and $t_{m+1} = \infty$. Then, we know that

$$
d\left( \overline{\mathbf{S}}_m, \mathcal{C}(t_1, t_2, \ldots, t_m) \right) \leq d\left( \overline{\mathbf{S}}_m, \overline{\mathbf{T}}_m \right) \leq \max_i \max_{j=1}^{m} \left| \overline{S}_{i,m}(t_j) - \overline{S}_{i,m}(t_{j+1}-) \right|, \tag{12}
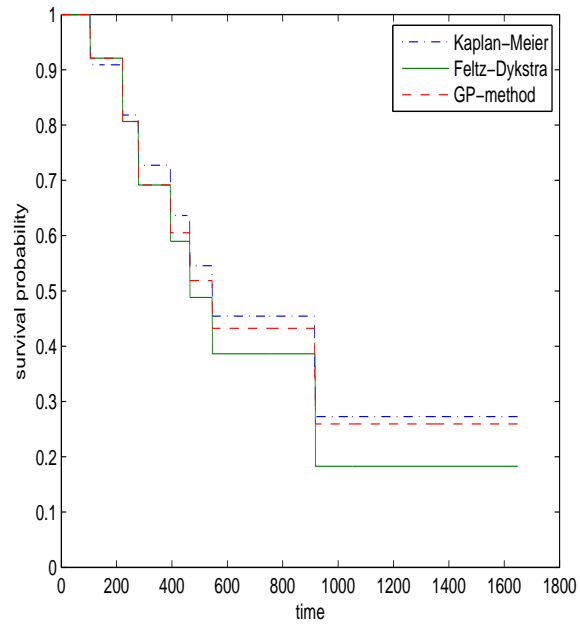$$

where $\overline{S}_{i,m}(t_{m+1}-)$ is set to be 0 for every $i$. The rightmost quantity in (12) converges to 0 a.s., as $m$ increases, which can be seen from the fact that $\mathbf{S}_\infty$ only has discontinuous points at $t_1, \ldots, t_m$ and $\max_j \left| t_{j+1} - t_j \right|$ decreases to 0 a.s.
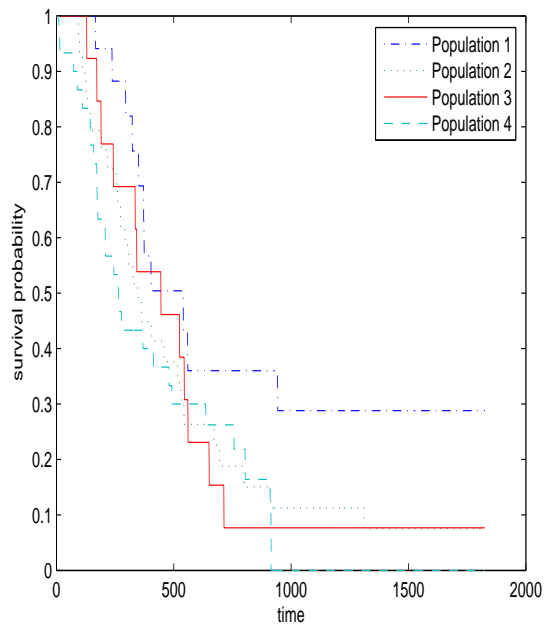
23
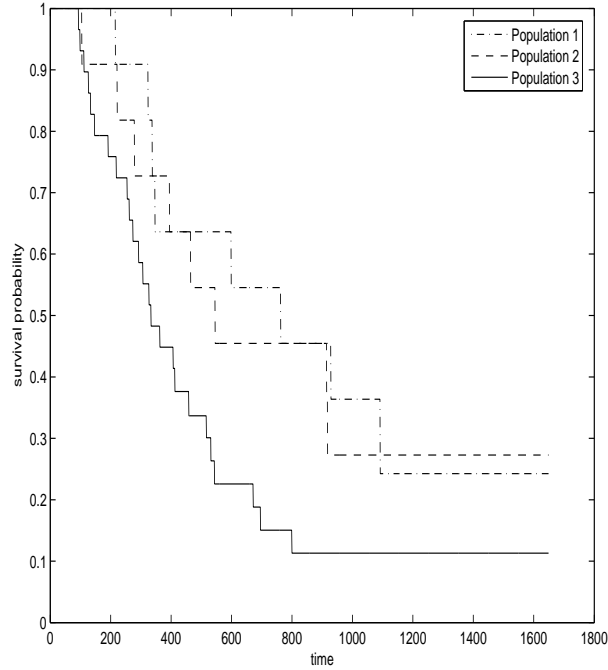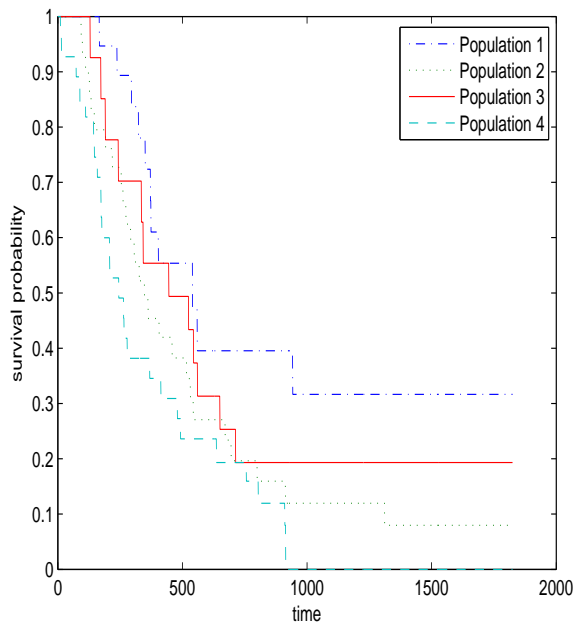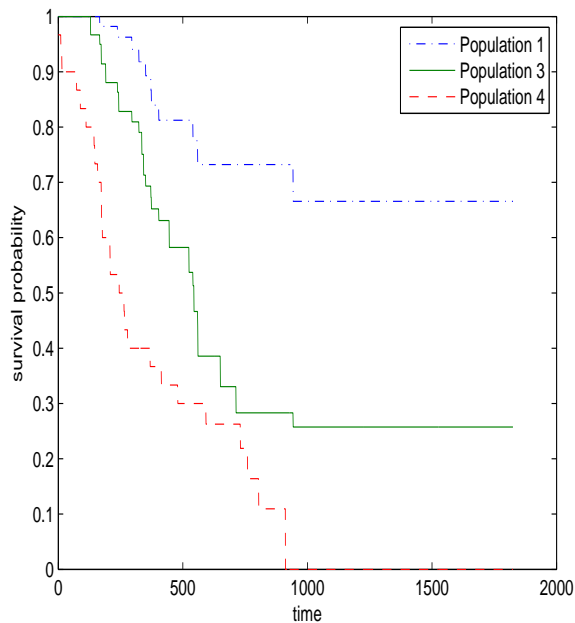
Figure 1: Results for right censored data from Feltz and Dykstra (1985): (a) the Kaplan-Meier estimates (NPMLEs without ordering constraints); (b) the GP estimates under simple stochastic ordering, where $S_1$ is simply larger than $S_2$, and $S_2$ is simply larger than $S_3$; (c) the estimated $S_2$ by the Kaplan-Meier method (NPMLE without order constraints), and by the Feltz-Dykstra method and the GP method under simple stochastic ordering.

Figure 2: Results for right censored data from Dykstra and Feltz (1989): (a) the Kaplan-Meier estimates (NPMLEs without ordering constraints) of the four populations; (b) the GP estimates under partial simple stochastic ordering, where $S_1$ is simply larger than both $S_2$ and $S_3$ that are simply larger than $S_4$; (c) the GP estimates under mixed simple and uniform stochastic ordering constraints, where $S_1$ is uniformly larger than $S_3$, and $S_3$ is simply larger than $S_4$.
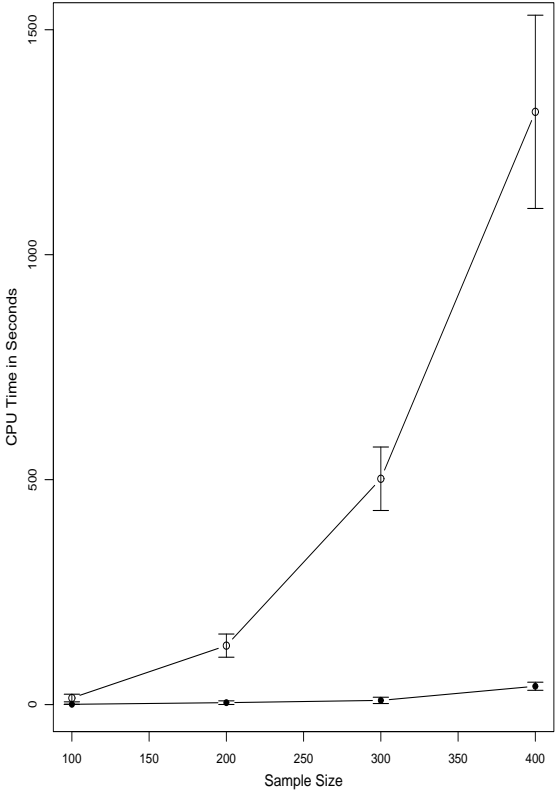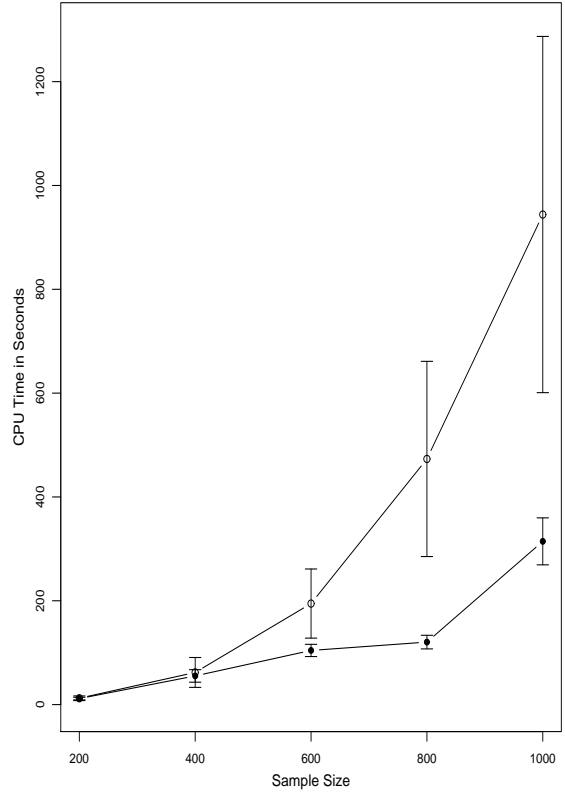
(a)                                                (b)

Figure 3: Simulation Results: (a) comparison of computing time: "○" represents the Feltz-Dykstra method and ● represents the GP method; (b) scalability of the GP method: "○" is for the case of simple stochastic ordering and "●" for the case of uniform stochastic ordering.
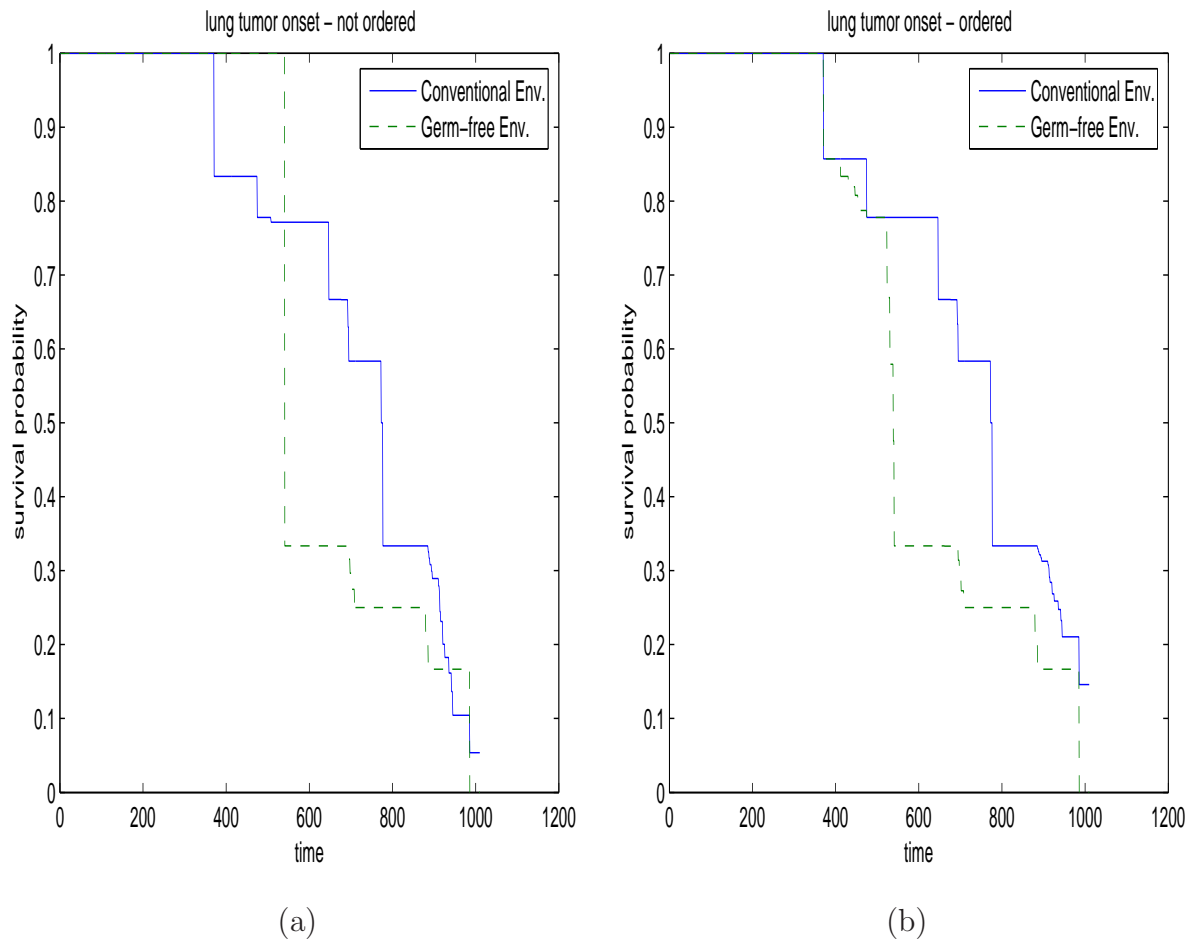
Figure 4: Results for "case 1" interval censored data: (a) unconstrained estimates of survival functions of time to lung tumor on set; (b) constrained estimates of survival functions of time to lung tumor on set.