

Constrained nonparametric maximum likelihood estimation of stochastically ordered survivor functions

Yongseok PARK*, John D. KALBFLEISCH and Jeremy M. G. TAYLOR

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, USA

Key words and phrases: Bootstrap; censored data; constrained NPMLE; Kaplan–Meier estimator; maximum constrained NPMLE; order restriction.

MSC 2010: Primary 62N02; secondary 62G30.

Abstract: This paper considers estimators of survivor functions subject to a stochastic ordering constraint based on right censored data. We present the constrained nonparametric maximum likelihood estimator (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 possibility of 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 $S_1(t)$. Dykstra (1982) also considered constrained nonparametric maximum likelihood estimation 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 and consistency of the estimators is proved. Data from a study on larynx cancer are analysed to illustrate the method. *The Canadian Journal of Statistics* 40: 22–39; 2012 © 2012 Statistical Society of Canada

Résumé: Cet article considère les estimateurs des fonctions de survie basés sur les données censurées à droite soumises à une contrainte d'ordonnement stochastique. Nous présentons un estimateur du maximum de vraisemblance non paramétrique contraint (C-NPMLE) des fonctions de survie dans le contexte d'un ou de deux échantillons lorsque la distribution de survie peut être discrète ou continue. De plus, nous discutons de la non-identifiabilité des estimateurs. Nous proposons aussi un algorithme, efficace au plan des calculs, pour obtenir le C-NPMLE. Pour considérer la possibilité de la non-identifiabilité du C-NPMLE de $S_1(t)$ lorsque $S_1(t) \leq S_2(t)$, nous considérons le C-NPMLE maximum (MC-NPMLE) de $S_1(t)$. Lorsque nous avons qu'un seul échantillon et une borne supérieure arbitraire $S_2(t)$ pour la fonction de survie, nous présentons un algorithme nouveau et efficace pour trouver le MC-NPMLE de $S_1(t)$. Dykstra (1982) a aussi considéré l'estimation par maximum de vraisemblance non paramétrique pour de tels problèmes, mais, comme nous le démontrons, il y a une erreur dans la méthode de Dykstra et elle ne conduit pas toujours à un C-NPMLE. Nous corrigeons cette erreur et des simulations montrent une amélioration de l'efficacité par rapport à l'estimateur de Dykstra. Des intervalles de confiance basés sur des méthodes de rééchantillonnage sont proposés et nous démontrons la cohérence de nos estimateurs. Notre méthode est illustrée à l'aide de données provenant d'une étude sur le cancer du larynx. *La revue canadienne de statistique* 40: 22–39; 2012 © 2012 Société statistique du Canada

Supporting Information can be found in the online version of this paper.

* Author to whom correspondence may be addressed.

E-mail: yongpark@umich.edu

1. INTRODUCTION

Suppose that random variable $T > 0$ is the time until some specified event, such as death or recurrence of a disease. Our interest centres on estimating the survivor function of T , $S(t) = P(T > t)$, $t > 0$. With right censored data, the Kaplan–Meier estimator (KM, Kaplan & 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 tumour 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 & Shanthikumar, 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 & El Barmi, 2003) and stochastic precedence (Arcones, Kvam & 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 MLE (C-NPMLE) is a nonparametric estimator that maximizes the likelihood subject to the constraint. Since the initial work of Brunk et al. (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 $T_1 \leq_{st}, \dots, \leq_{st} T_G$, and a general partial ordering case, which includes constraints such as $T_1 \leq_{st} T_2$, $T_1 \leq_{st} T_3$. Brunk et al. (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 & Dykstra (1985), Dykstra & Feltz (1989), Hoff (2000, 2003), and Lim, Kim & Wang (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 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 & 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 paper 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 $\hat{S}_1(1) = \hat{S}_1(2) = \hat{S}_1(3) = \hat{S}_1(4) = 0.4$ and $\hat{S}_1(5) = 0$ with the corresponding likelihood $\hat{L} = (1 - 0.4) \times 0.4^3 = 0.0384$. 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 & 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 & 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, 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 3, we extend to the case where $S_1(t)$ and $S_2(t)$ are not discrete functions. In Section 4, we show the uniform consistency of the C-NPMLE in the two-sample case. In Section 5, we analyse larynx cancer data in the two-sample case. In Section 6, we propose methods to construct confidence intervals and in Section 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 Supplementary Material.

2. ESTIMATION OF DISCRETE SURVIVOR FUNCTIONS

2.1. One-Sample Case

Consider a discrete failure time variable T_1 with potential failure times $a_1 < \dots < a_m$ 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 , $i = 1, \dots, m$;
- n_{1i} the number at risk just prior to a_i , $i = 1, \dots, m$; and
- c_{1i} the number of censored subjects in $[a_i, a_{i+1})$ $i = 0, \dots, m$.

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

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

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

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

The likelihood is maximized subject to $\sum_{j=1}^i h_{1j} \geq \sum_{j=1}^i h_{2j}$ or $\sum_{j=1}^i h_{1j} \leq \sum_{j=1}^i h_{2j}$, $i = 1, \dots, m$ under bounded below or bounded above constraint respectively. Consider now a particular vector $\mathbf{h}_1^* = (h_{11}^*, \dots, h_{1m}^*)$. 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. 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 ,

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

for a, b integer with $1 \leq a \leq b \leq m$. In (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 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 $\hat{k}^1 \geq \hat{k}^2 \geq \dots$. This theorem gives a method of obtaining the MLE, however, the $\hat{\mathbf{h}}_1$ can be obtained more easily using an algorithm that was presented by Dykstra (1982). A computationally more adaptation of this algorithm is given in the Supplementary Material.

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 constrained NPMLE (MC-NPMLE) and then present a theorem and an algorithm to obtain the MC-NPMLE under bounded above constraint.

Definition 1. *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 the Supplementary Material that the MC-NPMLE exists and is unique.

Theorem 2 (Bounded Above Constraint). *Let $m' = \max\{i : n_{1i} > 0\}$. For each a, b with $1 \leq a \leq b \leq m'$, let $\hat{k}^i = \min_{a \leq i} \max_{b \geq i} \min\{K^-(a, b), n_{1b}\}$, 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 $\hat{S}_1(t) = \exp(\sum_{i: a_i \leq t} \hat{h}_{1i})$, $t < \tau$, 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 $\hat{k}^i = \min_{a \leq i} \max_{b \geq i} K^-(a, b)$. 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, n_{1b} is the smallest in this range, so $\hat{k}^i = \min_{a \leq i} \max_{b \geq i} \min\{K^-(a, b), n_{1b}\}$.

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

Algorithm 1 (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 $H(i_{\ell-1} + 1, i_\ell, -n_{1i_\ell}) \geq 0$, 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 $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} - k_\ell)\}, i_{\ell-1} + 1 \leq j \leq i_\ell - 1$
 $\hat{h}_{1i_\ell} = \sum_{j=i_{\ell-1}+1}^{i_\ell} h_{2j} - \sum_{j=i_{\ell-1}+1}^{i_\ell-1} \hat{h}_{1j}$.
6. If $i_\ell = m'$, stop. Otherwise, set $\ell = \ell + 1$ and go to step 2.
7. Let $\hat{h}_{1j} = \log(1 - d_{1j}/n_{1j}), i_{\ell-1} + 1 \leq j \leq m'$, stop.

Heuristically, the solution is a KM type estimator of modified data with right censoring and left truncation. For i with $i_{\ell-1} < i \leq i_\ell$, the number at risk is modified to $n_{1i} - k_\ell$. Since $k_1 \geq k_2 \geq \dots$, 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 $a_{i_\ell} +$, $\ell = 1, 2, \dots$. 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.

TABLE 1: Results for Example 1.

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	3	2	2	2	1	1	0		
S_2	0.94	0.92	0.86	0.68	0.52	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	0.85	0.71	0.52	0.52	0.52	0.40	0.36	0.32	0.16	0.16	0.16	0.16		
\hat{h}_{1i}	-0.16	-0.19	-0.31	0	0	-0.26	-0.11	-0.12	-0.69	0	0	0		
k_ℓ	←		$k_1 = 3.2$			→		$k_2 = 3$	$k_3 = 2$	$k_4 = 2$	←		$k_5 = 0$	→
\hat{k}^i	3.2	3.2	3.2	3.2	3.2	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		

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 1, \hat{k}^i is from Theorem 2, S_1^D is D-estimates (Dykstra, 1982), and the last observed time $\tau = 11.5$.

Example 1 (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, $S_1(t) \leq S_2(t)$, $t = 1, \dots, 12$, and the values for $S_2(t)$ are given in Table 1.

Table 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 and those of k_ℓ are from Algorithm 1. The active constraints are at times 5, 6, 7 and 8. As indicated above, if $i = i_{\ell-1} + 1, \dots, i_\ell$, then $\hat{k}^i = k_\ell$. For example, $\hat{k}^1 = \dots = \hat{k}^5 = k_1$. As can be seen in Figure 1, the major difference between the MC-NPMLE and the D-estimator is that the former has jumps in the estimate for population 2 at times 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 (-12.41 vs. -13.22) and is closer to the KM estimator.

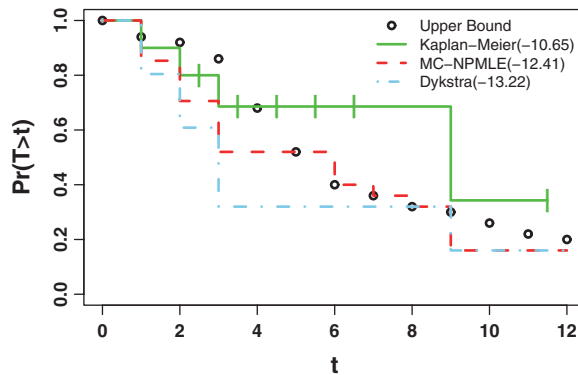


FIGURE 1: Estimates of survivor functions under bounded above constraint in discrete situation. Log likelihood values shown in parentheses. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

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 $S_1(a_1), \dots, S_1(a_m), S_2(a_1), \dots, S_2(a_m)$ is

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

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

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

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

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

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

Let $D_g(a, b) = \max_{a \leq j \leq b} d_{gj}, g = 1, 2$. For $a \leq i \leq b$, 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 3 (Two-sample case). *Let $m' = \max\{i : n_{1i} > 0, n_{2i} > 0\}$. For each a, b with $1 \leq a \leq b \leq m'$, let $\hat{k}^i = \min_{a \leq i} \max_{b \geq i} \min\{K_2^+(a, b), n_{2b}\}, i \leq m'$ and $\hat{k}^i = 0$ if $i > m'$. Let*

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

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 $t \geq \tau_g$ subject to the constraint, $g = 1, 2$.

We also describe an efficient algorithm to obtain the solution.

Algorithm 2 (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_ℓ 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 $I = \min_{b > i_\ell} \{b : n_{2b} > k_\ell \text{ and } H_2(i_\ell + 1, b, k_\ell) < 0\}$. 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 cannot 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.

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 $\{X_1, \dots, X_N\}$ 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 $C_{gi}^{(j)}$, $g = 1, 2, i = 1, \dots, N, j = 1, \dots, m_{gi}$ be the censoring times of population g in $[X_i, X_{i+1})$, $i = 0, \dots, N$, 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

$$\begin{aligned}
 L(S_1(\cdot), S_G(\cdot)) &= \prod_{g=1}^G \prod_{i=1}^{n_g} \{S_g(Y_{gi}-) - S_g(Y_{gi})\}^{\Delta_{gi}} S_g(Y_{gi})^{1-\Delta_{gi}} \\
 &= \prod_{g=1}^G \left(\prod_{j=1}^{m_{g0}} S_g(C_{g0}^{(j)}) \prod_{i=1}^N \left[\{S_g(X_i-) - S_g(X_i)\}^{d_{gi}} \prod_{j=1}^{m_{gi}} S_1(C_{gi}^{(j)}) \right] \right) \quad (3)
 \end{aligned}$$

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

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 the method for the discrete case can be directly used by setting the observed event times to potential event times and estimating $\hat{S}_1(t)$ for $t \leq \max(Y_{1j})$, the last observed time.

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 1 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 3.

Algorithm 3. Let $C_i, i = 1, \dots, n_c$ be all distinct observed censoring times and let X_i^- be the time just before observed event time X_i .

1. Let $X'_i, i = 1, 2, \dots$ be the distinct ordered times from the union of X_i, X_i^- and C_i .
2. Estimate $\hat{S}_1(t)$, which is the MC-NPMLE with potential event times at all X'_i , using Algorithm 1.
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 4. $\tilde{S}_1(t)$ from Algorithm 3 is the MC-NPMLE of $S_1(t)$ subject to $T_1 \leq_{st} T_2$.

Proof. See the Supplementary Material. ■

Example 2 (One-sample case with a bounded above constraint, continuous case). In Example 1, suppose we take $S_2(t)$ to be piecewise linear with knots at the discrete points in Table 1.

The procedure for calculating the MC-NPMLE in this case with Algorithm 3 is illustrated in Table 2 and Figure 2a. First, choose X'_i as in the first row of Table 2 and find d_{1i} and n_{1i} , the number of events and the number at risk at time X'_i . Algorithm 1 with potential event times X'_i is used to calculate $\hat{S}_1(t)$ as shown in the sixth row in Table 2 and the dash plot in Figure 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 2a. Note that the MC-NPMLE is not a step function in this example.

Figure 2b–d is 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 3. 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 3, as the number of potential event times goes to infinity.

TABLE 2: Results for Example 2.

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 X'_i, d_{1i} and n_{1i} are defined in Algorithm 3, S_2 is constraint, S_1^* is KM estimate, and \hat{S}_1 and \hat{h}_{1i} are results from Algorithm 1.

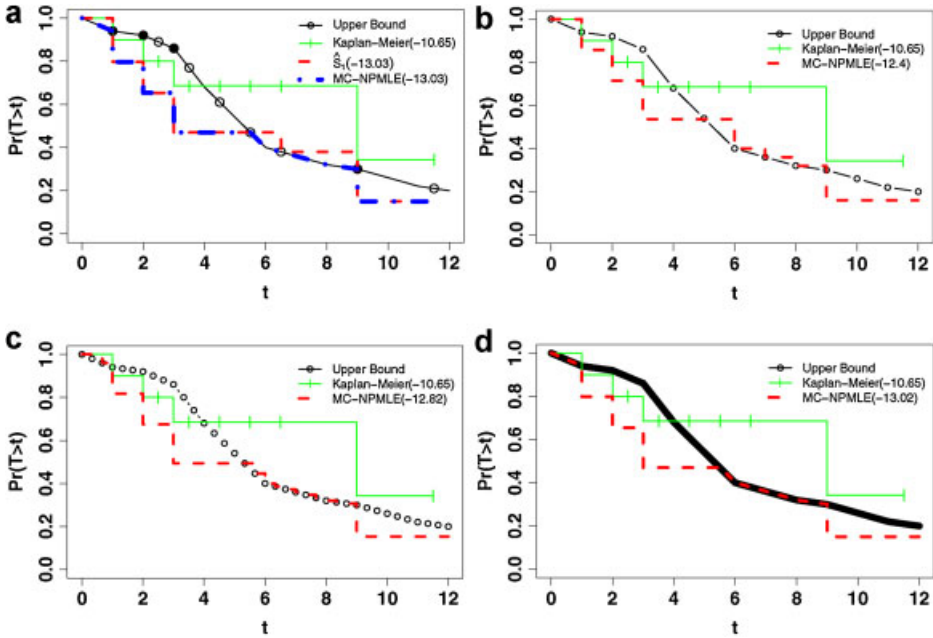


FIGURE 2: Estimates of survivor functions under continuous bounded above constraint. Log likelihood values shown in parentheses. (a) Algorithm 3; (b) Limit approach with 12 potential points; (c) Limit approach with 36 potential points; (d) Limit approach with 360 potential points. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

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, \dots, X_N , and $\hat{S}_g(X_i) = \bar{S}_g(X_i)$, $g = 1, 2, i = 1, \dots, N$, 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 3 and Algorithm 2 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 Sections 3.1 and 3.2.

4. CONSISTENCY

Dijkstra (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}) , $g = 1, 2, i = 1, \dots, n_g$ are independent and identically distributed (iid) 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

$$\lim_{n_g \rightarrow \infty} P \left\{ \sup_{x \leq \tau} |S_g^*(x) - S_g(x)| > \epsilon \right\} = 0 \tag{4}$$

for any given $\epsilon > 0$.

Theorem 5. *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*

$$\lim_{n_g \rightarrow \infty} P \left(\sup_{t \leq \tau} |\hat{S}_g(t) - S_g(t)| > \epsilon \right) = 0 \tag{5}$$

for any given $\epsilon > 0$.

Proof. See the Supplementary Material. ■

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 3.

Table 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 3a. The two plots cross each other, indicating that there exist violations of the stochastic ordering constraint $T_1 \geq_{st} T_2$. Times 0.2, 0.6, . . . , 7.4 in Table 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. The remaining times 9.3 and 10.7 are the last observed censoring times of population 2 and 1. Figure 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 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.

TABLE 3: Survival times (in years) of male patients with larynx cancer stage 1 and stage 2 in example in Section 5.

		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+		

+ represents censoring.

TABLE 4: KM and C-NPMLE estimates of survivor functions for male patients with larynx cancer in example in Section 5.

t	0.2	0.6	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**	

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.

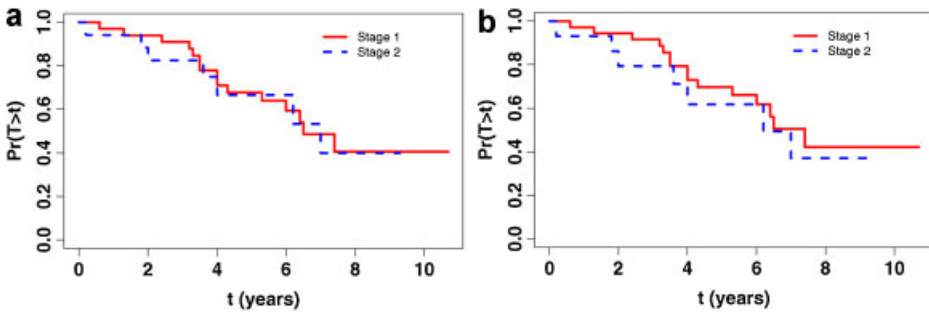


FIGURE 3: Estimates of survivor functions, two-sample case. (a) KM Estimator; (b) C-NPMLE. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

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 centred on the constrained estimators (Hwang & Peddada, 1994). We apply this idea on a log transformed scale in this paper and consider the following approximate $100(1 - 2\alpha)\%$ confidence interval,

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

where $\sigma_g^*(x)$ is standard error estimate of $\log S_g^*(x)$ and z_α is the α th percentile of the standard normal distribution (see Kalbfleisch & Prentice, 2002, page 17).

Another possible approach to construct confidence intervals is to use the bootstrap methods. 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. In the restricted scheme (denoted R-Bootstrap), 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 the censoring survivor function for group g . For each bootstrap sample, a bootstrap estimate $\hat{S}_g^b(t)$, $b = 1, \dots, B$ is obtained by applying the C-NPMLE or the MC-NPMLE. Confidence intervals can be constructed using the standard percentile method or the basic bootstrap method on the $h(s) = \arcsin(\sqrt{s})$ transformed scale (Davison & Hinkley, 1997, page 32). In the basic bootstrap method, the confidence interval limits are derived from the percentiles of $h^{-1}[2h\{\hat{S}_g(t)\} - h\{S_g^b(t)\}]$.

7. SIMULATION STUDY

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 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 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 4c and f), but in other cases, the C-NPMLE has smaller MSE compared to Dykstra’s estimator. In cases with the same censoring distributions (Figure 4a and d), 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

TABLE 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 4a	$-t$	$-1.2t$	$-1.5t$	$-1.5t$	100	40
Figure 4b	$-t$	$-1.2t$	No censoring	$-3.0t$	100	40
Figure 4c	$-t$	$-1.2t$	$-3.0t$	No censoring	100	40
Figure 4d	$-t$	$-1.2t$	$-1.5t$	$-1.5t$	40	100
Figure 4e	$-t$	$-1.2t$	No censoring	$-3.0t$	40	100
Figure 4f	$-t$	$-1.2t$	$-3.0t$	No censoring	40	100

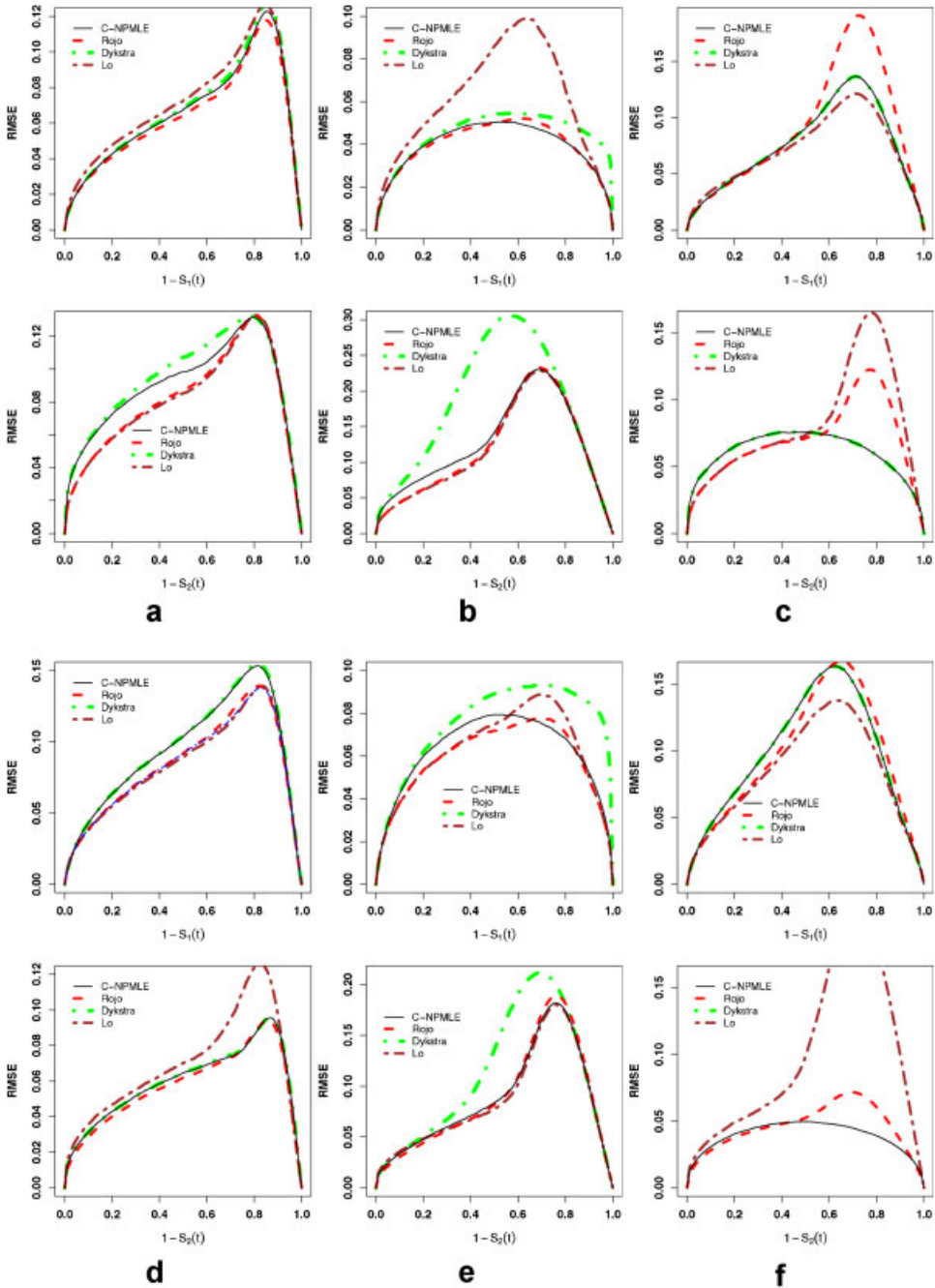


FIGURE 4: Comparison of RMSEs for different estimators. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

1 is excessively censored (Figure 4c and f), 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 4b, c, and f).

7.2. Properties of Confidence Intervals

We conducted a simulation study to evaluate finite sample properties of the CIs proposed in Section 6. The sample sizes and the underlying distributions are shown in Table 6. Each simulation consists of 1,000 replications and bootstrap CIs are based on 199 bootstrap samples. We construct bootstrap CIs using percentiles and basic bootstrap on $\arcsin(\sqrt{s})$ transformed scale at times 0.29 and 0.69, where the survival probabilities of group 1 are 0.75 and 0.50, respectively. We also evaluated basic bootstraps 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 6 shows the mean squared errors (MSEs) and biases of the KM estimator and the C-NPMLE. 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 (I, II, and V)

The coverage rates and average widths of the CIs described in Section 6 are shown in Table 7. Bootstrap methods using the percentiles (columns 5 and 7 in Table 7) do not work well with fairly low coverage rates when the two survivor functions are relatively close to one another (I, II, III, V, and VI). However, we found that the basic bootstrap CIs (Bootstrap $(\arcsin(\sqrt{s}))$ and R-Bootstrap $(\arcsin(\sqrt{s}))$ in Table 7) produce narrower intervals with reasonable coverage rates, especially when using the restricted sampling scheme (R-Bootstrap $(\arcsin(\sqrt{s}))$ in Table 7). In comparison,

TABLE 6: MSE ($\times 10^4$) and bias ($\times 10^2$).

Event time distributions	Sample sizes	MSE		Bias		MSE		Bias	
		KM	C-NPMLE	KM	C-NPMLE	KM	C-NPMLE	KM	C-NPMLE
		$t = 0.29$				$t = 0.69$			
I $\lambda_1 = 1$	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
II $\lambda_1 = 1$	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
III $\lambda_1 = 1$	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
IV $\lambda_1 = 1$	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
V $\lambda_1 = 1$	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
VI $\lambda_1 = 1$	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
VII $\lambda_1 = 1$	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
VIII $\lambda_1 = 1$	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

Event time distribution is $\exp(\lambda_g)$, $g = 1, 2$ and censoring distribution is $U(0, 1.5)$ for both groups in all scenarios.

TABLE 7: Coverage rates ($\times 10^2$) and average widths ($\times 10^2$) of 95% confidence intervals.

Sampling scheme		Bootstrap				R-Bootstrap			
		KM		C-NPMLE		K-M		C-NPMLE	
Estimator		Percentile		arcsin(\sqrt{s})		Percentile		arcsin(\sqrt{s})	
Confidence interval		Percentile		arcsin(\sqrt{s})		Percentile		arcsin(\sqrt{s})	
								$S_g(t) \exp(\pm 1.96\sigma_g^*)$	
$t = 0.29$	I	$S_1(t)$	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)
	II	$S_1(t)$	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)
	III	$S_1(t)$	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)
	IV	$S_1(t)$	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)
	V	$S_1(t)$	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)
	VI	$S_1(t)$	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)
	VII	$S_1(t)$	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)
	VIII	$S_1(t)$	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)
$t = 0.69$	I	$S_1(t)$	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)
	II	$S_1(t)$	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)
	III	$S_1(t)$	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)
	IV	$S_1(t)$	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)
	V	$S_1(t)$	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)
	VI	$S_1(t)$	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)
	VII	$S_1(t)$	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)
	VIII	$S_1(t)$	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)

the shifted constant width CI that is centred on the C-NPMLE (last column in Table 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., IV at time 0.29 and III at time 0.69), the CIs have too high a coverage rate. These intervals, however, tend to be wider than those from the basic bootstrap. Based on our simulation results, the basic bootstrap approach using the restricted sampling scheme (R-Bootstrap ($\arcsin(\sqrt{s})$) in Table 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.

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 suggest 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.

Præstgaard & 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, that is, 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 one is 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, 2003) and Lim, Kim & 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.

BIBLIOGRAPHY

- Arcones, M. A., Kvam, P. H., & Samaniego, F. J. (2002). Nonparametric estimation of a distribution subject to a stochastic precedence constraint. *Journal of the American Statistical Association*, 97, 170–182.
- Brunk, H. D., Franck, W. E., Hanson, D. L., & 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.

- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge.
- 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. & Feltz, C. J. (1989). Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. *Biometrika*, 76, 331–341.
- El Barmi, H. & Mukerjee, H. (2005). Inferences under a stochastic ordering constraint: The k-sample case. *Journal of the American Statistical Association*, 100, 252–261.
- Feltz, C. & 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.
- Hoff, P. D. (2000). Constrained nonparametric maximum likelihood via mixtures. *Journal of Computational and Graphical Statistics*, 9, 633–641.
- Hoff, P. D. (2003). Nonparametric estimation of convex models via mixtures. *The Annals of Statistics*, 31, 174–200.
- Hwang, J. T. G. & 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.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, New York.
- Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457–481.
- Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients—A case study. *Statistica Neerlandica*, 37, 103–126.
- Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *The Annals of Statistics*, 25, 95–104.
- Lim, J., Kim, S. J., & 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.
- Meier, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics*, Gani, J., editor. Applied Probability Trust, Sheffield, England, pp. 67–87.
- Murphy, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*, 90, 1399–1405.
- Præstgaard, J. T. & Huang, J. (1996). Asymptotic theory for nonparametric estimation of survival curves under order restrictions. *The Annals of Statistics*, 24, 1679–1716.
- Rojo, J. (2004). On the estimation of survival functions under a stochastic order constraint. In *The First Erich L. Lehmann Symposium: Optimality*, Vol. 44, Pérez-Abreu, V. & Rojo, J., editors. Institute of Mathematical Statistics, Beachwood, OH, pp. 37–61.
- Rojo, J. & El Barmi, H. (2003). Estimation of distribution functions under second order stochastic dominance. *Statistica Sinica*, 13, 903–926.
- Rojo, J. & Ma, Z. (1996). On the estimation of stochastically ordered survival functions. *Journal of Statistical Computation and Simulation*, 55, 1–21.
- Shaked, M. & Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer, New York.
- Thomas, D. R. & Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70, 865–871.

Received 19 January 2011

Accepted 18 October 2011