

## Pointwise nonparametric maximum likelihood estimator of stochastically ordered survivor functions

BY YONGSEOK PARK, JEREMY M. G. TAYLOR AND JOHN D. KALBFLEISCH

*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.*  
yongpark@umich.edu jmg@umich.edu jdkalbf@umich.edu

### SUMMARY

In this paper, we consider estimation of survivor functions from groups of observations with right-censored data when the groups are subject to a stochastic ordering constraint. Many methods and algorithms have been proposed to estimate distribution functions under such restrictions, but none have completely satisfactory properties when the observations are censored. We propose a pointwise constrained nonparametric maximum likelihood estimator, which is defined at each time  $t$  by the estimates of the survivor functions subject to constraints applied at time  $t$  only. We also propose an efficient method to obtain the estimator. The estimator of each constrained survivor function is shown to be nonincreasing in  $t$ , and its consistency and asymptotic distribution are established. A simulation study suggests better small and large sample properties than for alternative estimators. An example using prostate cancer data illustrates the method.

*Some key words:* Censored data; Constrained nonparametric maximum likelihood estimator; Kaplan–Meier estimator; Maximum likelihood estimator; Order restriction.

### 1. INTRODUCTION

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

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

We consider independent right-censored samples of the form  $(Y_{gi}, \Delta_{gi})$  ( $g = 1, \dots, G$ ;  $i = 1, \dots, n_g$ ), where  $Y_{gi}$  is the observed time and  $\Delta_{gi}$  is the event indicator. We assume that the

censoring mechanism is independent, so that the generalized likelihood is

$$L\{S_1(\cdot), \dots, S_G(\cdot)\} = \prod_{g=1}^G \prod_{i=1}^{n_g} \{S_g(Y_{gi}-) - S_g(Y_{gi})\}^{\Delta_{gi}} S_g(Y_{gi})^{1-\Delta_{gi}}. \quad (1)$$

The  $E$ -constrained nonparametric maximum likelihood estimator maximizes (1) subject to the partial-ordering constraint  $E$ . Brunk et al. (1966) studied the constrained nonparametric maximum likelihood estimator in the two-sample case without censoring. Dykstra (1982), as corrected by Park et al. (2012), extended this result to right-censored data. In the case of three or more populations with general partial-ordering constraints, Hoff (2003) and Lim et al. (2009) proposed different computational methods for obtaining the constrained nonparametric maximum likelihood estimator.

This estimator has the undesirable property that a violation of a constraint in the Kaplan–Meier estimators (Kaplan & Meier, 1958) at an earlier time affects the estimator at a later time, even if there is no violation at this later time. A number of authors have noted that the constrained nonparametric maximum likelihood estimator can have relatively large pointwise bias and mean squared error at a fixed  $t$  and have suggested alternatives (Rojo & Ma, 1996; Rojo, 2004; El Barmi & Mukerjee, 2005) that can have better mean squared error properties. Park et al. (2012) noted a correction to the constrained nonparametric maximum likelihood estimator presented by Dykstra (1982), which led to improved properties, but this corrected estimator still often has poorer pointwise properties than other estimators, some of which are relatively simple to define. In the two-sample problem, Lo (1987) suggested swapping the Kaplan–Meier estimates of the survivor functions when the constraint is violated. Rojo (2004) proposed estimating both survivor functions as the weighted average of the two Kaplan–Meier estimators at times when the constraint is violated, where the weights are based on the initial sample sizes. El Barmi & Mukerjee (2005) extended Rojo’s estimators to the simple ordering situation using isotonic regression. The simulation study in Park et al. (2012) shows that some of these estimators have smaller mean squared error than the constrained nonparametric maximum likelihood estimator when the censoring distributions are equal, but when the censoring distributions differ substantially between groups, the alternative estimators may have larger mean squared error than the constrained nonparametric maximum likelihood estimator. Moreover, these alternative estimators have not been explicitly extended to a general partial-ordering case.

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

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

2. ESTIMATION METHODS

2.1. Notation and likelihood

To obtain the pointwise constrained estimator as given in Definition 1, it is required to maximize the likelihood (1) subject to the constraints  $S_i(x) \geq S_j(x)$  for all  $(i, j) \in E$  for a fixed time  $x$ . This will give the estimates of  $\tilde{S}_1(t; x), \dots, \tilde{S}_G(t; x)$  and the constrained maximization will be repeated for all times  $x$  of interest.

Let  $X_{gj}$  ( $j = 1, \dots, m_g$ ) be the distinct event times in group  $g$  and define  $X_{g0} = 0$  and  $X_{g(m_g+1)} = \infty$  ( $g = 1, \dots, G$ ). Let  $N_g(t)$  be the number at risk at time  $t$  in group  $g$  and let  $M_g(t)$  be the number of distinct events in  $(0, t]$  in group  $g$ . Let  $d_{gj}$  and  $n_{gj}$  be, respectively, the number of events and the number at risk in group  $g$  at time  $X_{gj}$ .

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

$$\log L(h_1, \dots, h_G) = \sum_{g=1}^G \left\{ \sum_{i=1}^{m_g} (d_{gi} \log[1 - \exp\{h_g(X_{gi})\}] + (n_{gi} - d_{gi})h_g(X_{gi})) + N_g(x)h_g^\delta(x) \right\}, \tag{2}$$

where  $h_g = \{h_g(X_{g1}), \dots, h_g(X_{gm_g}), h_g^\delta(x)\}$  ( $g = 1, \dots, G$ ). The corresponding constraints are  $\sum_{j=1}^{M_p(x)} h_p(X_{pj}) + h_p^\delta(x) \geq \sum_{j=1}^{M_r(x)} h_r(X_{rj}) + h_r^\delta(x)$ , for all  $(p, r) \in E$ , and  $h_g^\delta(x) \leq 0$ . In this,  $h_g^\delta(x) = I(x \neq X_{gM_g(x)})h_g(x)$ , which is included to account for the fact that if  $x = X_{gM_g(x)}$ , we do not have the extra term  $N_g(x)h_g(x)$  in the loglikelihood (2).

2.2. Linearly constrained convex minimization

There is a large literature on general approaches to linearly constrained convex minimization problems. There are essentially three types of algorithms: interior point, primal active set and dual active set methods. In general, our data contain many more observed event times than groups. Interior point and primal active set methods simultaneously optimize over the large number of quantities  $h_g^\delta(x)$  and  $h_g(X_{gi})$  ( $g = 1, \dots, G; i = 1, \dots, m_g$ ) at each time  $x$  of interest, and so are not computationally efficient in our setting. Dual active set methods may involve many fewer parameters, but the dual function itself is difficult to express as a function of Lagrange multipliers and the feasible range of these multipliers is difficult to specify in our problem. So the dual active set method is also difficult to implement in our context.

In § 2.4, we transform the problem of maximizing the loglikelihood (2) subject to the linear constraints to another simple concave maximization problem subject to linear constraints by using the profile likelihood. In preparation for this, we first discuss the constrained maximum likelihood estimator of the survivor function in the one-sample case.

2.3. Maximum likelihood estimator of the survivor function subject to a single constraint

In the one-sample case without constraints, the maximum likelihood estimator has probability mass only at the observed event times. The loglikelihood analogous to (2) is

$$\log L(h) = \sum_{j=1}^m [d_j \log\{1 - \exp h(X_j)\} + (n_j - d_j)h(X_j)], \tag{3}$$

where  $h = \{h(X_1), \dots, h(X_m)\}$  and (3) is maximized at  $h(X_j) = \log(1 - d_j/n_j)$  ( $j = 1, \dots, m$ ), which corresponds to the Kaplan–Meier estimator.

Consider now the maximum likelihood estimator subject to the constraint  $S(x) = \exp(q)$ . The maximum likelihood estimator of the survivor function will have positive probability mass at event times  $X_i$  and nonnegative probability mass at time  $x$ . The optimization problem is to maximize the loglikelihood of  $h = \{h(X_1), \dots, h(X_m), h^\delta(x)\}$ ,

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

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

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

$$\hat{h}(q; X_i) = \begin{cases} \log \left\{ 1 - \frac{d_i}{n_i + K(q; x)} \right\}, & i \leq M(x), \\ \log \left( 1 - \frac{d_i}{n_i} \right), & i > M(x). \end{cases} \tag{4}$$

**THEOREM 1.** *The maximum likelihood estimator of  $S(t)$  subject to constraint  $S(x) = \exp(q)$  at a given  $x$  is  $\hat{S}(t) = \exp \{ \sum_{X_j \leq t} \hat{h}(q; X_j) + I(t \geq x)h^\delta(q; x) \}$  ( $t \leq \tau$ ), where  $\tau$  is the last observed time.*

*Proof.* See the Appendix. □

Thomas & Grunkemeier (1975) and Li (1995) considered the maximization problem described above. However, Thomas & Grunkemeier (1975) solved the problem with the equality constraint  $\sum_{j=1}^{M(x)} h(X_j) = q$ , which implicitly assumes that  $\hat{h}(x) = 0$  if  $x$  is not an observed event time, whereas Li (1995) mistakenly proved that  $\hat{h}(x) = 0$  unless  $x$  is an observed event time. In fact, the maximization problem described above involves two constraints:  $\sum_{j=1}^{M(x)} h(X_j) + h^\delta(x) = q$  and  $h^\delta(x) \leq 0$ . It is possible that  $\hat{h}^\delta(x) < 0$  if  $K(q; x) = -N(x)$ . The inequality constraint,  $h^\delta(x) \leq 0$ , has been neglected in these approaches. It is necessary, however, to apply the Karush–Kuhn–Tucker conditions (Kuhn & Tucker, 1951) to all possible inequality constraints, including the bounds on the parameters, and only omit the redundant constraints.

#### 2.4. Reformulation of the problem using profile likelihood

The profile loglikelihood of  $S(x) = \exp(q)$  at a given  $x$  is

$$\begin{aligned} \ell(q; x) &= \sup_{h \in R} \log L(h) \\ &= \sum_{i=1}^m \{d_i \log[1 - \exp\{\hat{h}(q; X_i)\}] + (n_i - d_i)\hat{h}(q; X_i)\} + N(x)\hat{h}^\delta(q; x), \end{aligned} \tag{5}$$

where  $R = \{h : \sum_{i=1}^{M(x)} h(X_i) + h^\delta(x) = q\}$ , and  $\hat{h}(q; X_i)$  and  $\hat{h}^\delta(q; x)$  are defined in (4).

LEMMA 1. *The derivative of the profile loglikelihood (5) with respect to  $q$  is  $-K(q; x)$ .*

*Proof.* See the Appendix. □

For given  $x$ , maximizing the loglikelihood (2) subject to the constraints in  $E$  can be redefined as maximizing the profile loglikelihood  $\ell(q_1, \dots, q_G; x)$ , which equals

$$\begin{aligned} \sum_{g=1}^G \ell_g(q_g; x) &= \sum_{g=1}^G \left( \sum_{i=1}^{M_g(x)} [(n_{gi} - d_{gi}) \log\{n_{gi} + K_g(q_g; x) - d_{gi}\} \right. \\ &\quad \left. - n_{gi} \log\{n_{gi} + K_g(q_g; x)\}] + I\{K_g(q_g; x) = N_g(x)\} N_g(x) \right. \\ &\quad \left. \times \left[ q_g - \sum_{j=1}^{M_g(x)} \log \left\{ 1 - \frac{d_{gi}}{n_{gi} + K_g(q_g; x)} \right\} \right] \right), \end{aligned} \tag{6}$$

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

Any of the general methods described in §2.2 can be used to maximize the profile loglikelihood (6) under the corresponding linear constraints. The profile loglikelihood (6) and its derivative,  $d\ell(q; x)/dq^T = \{-K_1(q_1; x), \dots, -K_G(q_G; x)\}^T$ , are easily calculated.

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

The following theorem shows that  $\hat{S}_g(t)$  ( $g = 1, \dots, G$ ) is a valid survivor function.

THEOREM 2. *The pointwise constrained estimator  $\hat{S}_g(t)$  obtained from maximizing the profile likelihood (6) is a nonincreasing function in  $t$  for each  $g = 1, \dots, G$ . That is, for any  $0 \leq x < y \leq \tau_g$ ,  $\hat{S}_g(x) \geq \hat{S}_g(y)$ .*

*Proof.* See the Supplementary Material. □

### 2.5. Generalized pool-adjacent-violators algorithm in the simple ordering case

Suppose that  $G$  survivor functions satisfy the simple stochastic ordering constraint  $T_1 \geq_{st} \dots \geq_{st} T_G$ , and we aim to estimate the pointwise constrained estimator at time  $x$ . A generalized pool-adjacent-violators algorithm can be used as developed in Best et al. (1999), because the profile loglikelihood is the sum of concave functions. The results of the generalized pool-adjacent-violators algorithm lead to a set of blocks,  $B_1, \dots, B_r$ , where  $r \geq 1$ ,  $B_r = \{u_{r-1} + 1, \dots, u_r\}$  and  $0 = u_0 < \dots < u_r = G$ . This is described in an algorithm in the Supplementary Material. The final estimate of the survivor function for each group in a block  $B_r$  takes a common value  $\exp(\hat{q}_r)$ , where  $\hat{q}_r$  maximizes the profile loglikelihood,  $\ell_{B_r}(q; x) = \sum_{i \in B_r} \ell_i(q; x)$  and  $0 \geq \hat{q}_1 > \dots > \hat{q}_r$ .

3. CONSISTENCY AND ASYMPTOTIC DISTRIBUTION

Let  $S_g^*(t)$  be the Kaplan–Meier estimator of  $S_g(t)$  and let  $S_g^c(t)$  be the censoring survivor function for group  $g$ . Further, let  $\tau_g = \inf\{t : S_g(t)S_g^c(t) = 0\}$  ( $g = 1, \dots, G$ ). Under the condition that there are no common jumps between the event and censoring distributions, [Stute & Wang \(1993\)](#) showed that the Kaplan–Meier estimator  $S_g^*(t)$  is uniformly consistent for  $S_g(t)$  on  $[0, \tau_g)$ . A similar result holds for the pointwise constrained estimator. The following theorem is proved in the Supplementary Material.

**THEOREM 3.** *Let  $\hat{S}_g(t)$  be the pointwise constrained estimator given in Definition 1. Under the condition of no common jumps of  $S_g(t)$  and  $S_g^c(t)$ ,  $\sup_{t < \tau_g} |\hat{S}_g(t) - S_g(t)| \rightarrow 0$  with probability 1 as  $n_g \rightarrow \infty$  ( $g = 1, \dots, G$ ).*

Let  $W_g(V_g)$  be a Brownian motion on  $[0, \infty)$  with variance function  $V_g(t)$ . As shown in [Gill \(1983\)](#),  $n_g^{1/2}(S_g^* - S_g)S_g^{-1} \rightarrow W(V_g)$  in distribution on  $[0, \tau_g]$  as  $n_g \rightarrow \infty$ , where  $V_g(t) = -\int_0^t \{S_g^2(x-)S_g^c(x-)\}^{-1} dS_g(x)$ . For a fixed time  $x$ ,  $n_g^{1/2}\{S_g^*(x) - S_g(x)\} \rightarrow N\{0, \sigma_g^2(x)\}$  in distribution, where  $\sigma_g^2(x) = V_g(x)S_g^2(x)$ .

Let  $n = \sum_{g=1}^G n_g$  and assume that  $\lim_{n \rightarrow \infty} n_g/n = c_g > 0$  and let  $Z_g^*(x) = n^{1/2}\{S_g^*(x) - S_g(x)\}$  ( $g = 1, \dots, G$ ). Then  $\{Z_1^*(x), \dots, Z_G^*(x)\}^T \rightarrow \{Z_1(x), \dots, Z_G(x)\}^T$  in distribution, where  $Z_g(x) \sim N\{0, \sigma_g^2(x)/c_g\}$  and  $Z_1(x), \dots, Z_G(x)$  are independent.

**THEOREM 4.** *For a fixed time  $x < \min\{\tau_k : L_g(x) \leq k \leq U_g(x)\}$  and under the simple ordering constraint  $T_1 \geq_{st} \dots \geq_{st} T_G$ ,*

$$n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\} \rightarrow c_g^{1/2} \min_{L_g(x) \leq \ell \leq g} \max_{g \leq u \leq U_g(x)} \frac{\sum_{k=\ell}^u \{Z_k(x)w_k(x)\}}{\sum_{k=\ell}^u w_k(x)} \tag{7}$$

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

*Proof.* See the Supplementary Material. □

In the Supplementary Material, the asymptotic distribution of the pointwise constrained estimator is discussed for situations where the number at risk in some groups is zero.

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

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

From [El Barmi & Mukerjee \(2005, Theorem 2\)](#), it can be seen that

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

in distribution. From (7), it follows that  $\hat{S}_g(x)$  and  $\check{S}_g(x)$  are asymptotically equivalent. We hypothesize that this equivalence to isotonic regression will also hold under the

partial-ordering constraint. This yields the following conjecture for the asymptotic distribution of the pointwise constrained estimator.

CONJECTURE 1. For a fixed time  $x$ ,

$$n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\} \rightarrow c_g^{1/2} f_g \left\{ Z_1(x), \dots, Z_G(x); \frac{c_1}{\sigma_1^2}, \dots, \frac{c_G}{\sigma_G^2}, x \right\}$$

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

If this conjecture is correct, inference methods developed for isotonic regression could also be useful for the pointwise constrained estimator.

#### 4. COMPARISON WITH THE KAPLAN–MEIER ESTIMATOR WHEN SAMPLE SIZE IS LARGE

##### 4.1. Simple ordering case

In the simple ordering case with no censoring, [El Barmi & Mukerjee \(2005\)](#) showed that their isotonic regression estimator has smaller asymptotic mean squared error than the unrestricted Kaplan–Meier estimator. A similar result holds for the pointwise constrained estimator compared with the Kaplan–Meier estimator when there is right censoring.

THEOREM 5. Consider the simple ordering constraint  $T_1 \geq_{st} \dots \geq_{st} T_G$ . For a fixed  $x$  with  $S_k^c(x)S_k(x) > 0$  for all  $k = 1, \dots, G$ , let  $n_k^{1/2}\{\hat{S}_k(x) - S_k(x)\} \rightarrow \hat{Z}_k$  and  $n_k^{1/2}\{S_k^*(x) - S_k(x)\} \rightarrow Z_k$  in distribution. If there exists at least one  $g'$  with  $S_{g'}(x) = S_g(x)$ , then  $E(\hat{Z}_g^2) < E(Z_g^2)$ . If no such  $g'$  exists, then  $\hat{S}_g(x)$  and  $S_g^*(x)$  are asymptotically equivalent.

Thus, the pointwise constrained estimator has smaller asymptotic mean squared error than the Kaplan–Meier estimator. In fact, a stronger inequality relation holds. Namely  $\text{pr}(|\hat{Z}_g| \leq \epsilon) > \text{pr}(|Z_g| \leq \epsilon)$  for all  $\epsilon > 0$ . In § 4.2, we calculate the asymptotic bias and asymptotic mean squared error of the pointwise constrained estimator in the two-sample case.

##### 4.2. The two sample case, $G = 2$

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

Let  $n_2/n_1 \rightarrow c$  as  $n_1, n_2 \rightarrow \infty$ . We consider asymptotic properties when  $S_1(x) = S_2(x)$ . From Theorem 4, we can show that

$$\begin{aligned} n_1^{1/2}\{\hat{S}_1(x) - S_1(x)\} &\rightarrow \sigma_1(x) \max \left\{ \bar{Z}_1, \frac{\bar{Z}_1 + c(x)^{1/2}\bar{Z}_2}{1 + c(x)} \right\}, \\ n_2^{1/2}\{\hat{S}_2(x) - S_2(x)\} &\rightarrow \sigma_2(x) \min \left\{ \bar{Z}_2, \frac{c(x)\bar{Z}_2 + c(x)^{1/2}\bar{Z}_1}{1 + c(x)} \right\}, \end{aligned} \tag{8}$$

in distribution, where  $c(x) = c \sigma_1^2(x)/\sigma_2^2(x)$ . Direct calculation from (8) shows that the asymptotic mean squared errors are

$$\begin{aligned} \lim_{n_1 \rightarrow \infty} E[n_1\{\hat{S}_1(x) - S_1(x)\}^2] &= \frac{\{2 + c(x)\}\sigma_1^2(x)}{2\{1 + c(x)\}}, \\ \lim_{n_2 \rightarrow \infty} E[n_2\{\hat{S}_2(x) - S_2(x)\}^2] &= \frac{\{1 + 2c(x)\}\sigma_2^2(x)}{2\{1 + c(x)\}}. \end{aligned} \tag{9}$$

These are always smaller than the unrestricted counterparts  $\sigma_1^2(x)$  and  $\sigma_2^2(x)$ .

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

$$\begin{aligned} E[n_1\{\tilde{S}_1(x) - S_1(x)\}^2] &= \sigma_1^2(x) + \frac{c\{\sigma_2^2(x) - (2 + c)\sigma_1^2(x)\}}{2\{1 + c\}^2}, \\ E[n_2\{\tilde{S}_2(x) - S_2(x)\}^2] &= \sigma_2^2(x) + \frac{c\{\sigma_1^2(x) - (1 + 2c)\sigma_2^2(x)\}}{2\{1 + c\}^2}. \end{aligned} \tag{10}$$

It can be shown that the asymptotic mean squared error of  $\hat{S}_g(x)$  is less than or equal to that of  $\tilde{S}_g(x)$  ( $g = 1, 2$ ), with equality only when  $\sigma_1^2(x) = \sigma_2^2(x)$ , in which case  $\tilde{S}_g(x)$  and  $\hat{S}_g(x)$  are asymptotically equivalent. From (10) we see that when  $\sigma_2^2(x)/\sigma_1^2(x) > c_2/c_1 + 2$ , Rojo’s estimator  $\tilde{S}_1(x)$  is asymptotically less efficient than the Kaplan–Meier estimator  $S_1^*(x)$  and when  $\sigma_1^2(x)/\sigma_2^2(x) > c_1/c_2 + 2$ ,  $\tilde{S}_2(x)$  is asymptotically less efficient than  $S_2^*(x)$ .

From (8), the asymptotic biases of  $\hat{S}_1(x)$  and  $\hat{S}_2(x)$  are

$$\begin{aligned} \lim_{n_1 \rightarrow \infty} E[n_1^{1/2}\{\hat{S}_1(x) - S_1(x)\}] &= \sigma_1(x) \int_{-\infty}^{\infty} \int_{c^{1/2}(x)}^{\infty} \frac{c^{1/2}z_2 - c(x)z_1}{1 + c(x)} f_{\bar{Z}_2}(z_2) f_{\bar{Z}_1}(z_1) dz_2 dz_1 \\ &= \sigma_1(x) \left[ \frac{c(x)}{2\pi\{1 + c(x)\}} \right]^{1/2} < \sigma_1(x) \left( \frac{1}{2\pi} \right)^{1/2}, \\ \lim_{n_2 \rightarrow \infty} E[n_2^{1/2}\{\hat{S}_2(x) - S_2(x)\}] &= \sigma_2(x) \left[ \frac{1}{2\pi\{1 + c(x)\}} \right]^{1/2} < \sigma_2(x) \left( \frac{1}{2\pi} \right)^{1/2}. \end{aligned}$$

## 5. CONFIDENCE INTERVALS

### 5.1. Asymptotic approaches

While there is a substantial literature on the estimation of survivor functions under stochastic ordering constraints, there has been little discussion of constructing confidence intervals for ordered survivor functions. Rojo (2004) demonstrated weak convergence to a Gaussian process of his estimator, from which confidence bands could be constructed. For the most part, however, asymptotic results are not particularly useful since, if the true inequalities at time  $t$  are strict, then the asymptotic distribution of  $\hat{S}_g(t)$  is the same as that of the Kaplan–Meier estimator and the corresponding approximate confidence interval would be unaffected by the restrictions. In our opinion, the most promising approach to constructing confidence intervals in these problems is through resampling methods that reflect the finite sample aspects. We consider some such approaches in the next section.



## 5.2. Bootstrap methods

We used a nonparametric resampling scheme, in which survival time and censoring indicator pairs are drawn with replacement from the data separately for each group. For each bootstrap sample, a bootstrap estimate  $\hat{S}_g^b(t)$  ( $b = 1, \dots, B$ ) is obtained by applying the pointwise constrained estimator. Simple confidence intervals based on these bootstrap estimates can be constructed using the percentile or the basic bootstrap method (Efron & Tibshirani, 1993; Davison & Hinkley, 1997). For a nominal level of  $(1 - 2\alpha)$ , the percentile confidence interval for  $S_g(t)$  is  $\{\hat{S}_{g,\alpha}^B(t), \hat{S}_{g,1-\alpha}^B(t)\}$ , where  $\hat{S}_{g,\alpha}^B(t)$  is  $\alpha$ th percentile of the bootstrap distribution. The basic bootstrap method utilizes ideas of pivotal statistics, and can also be improved by use of transformations such as  $h(s) = \arcsin(s^{1/2})$ . The confidence interval for the basic bootstrap method is given by  $(h^{-1}[2h\{\hat{S}_g(t)\} - h\{\hat{S}_{g,1-\alpha}^B(t)\}], h^{-1}[2h\{\hat{S}_g(t)\} - h\{\hat{S}_{g,\alpha}^B(t)\}])$ .

While these simple methods are easy to apply, a number of different methods have been developed which have improved properties. The work in Andrews (2000) suggests that the use of the bootstrap for inference problems with order restrictions on the parameters may be particularly challenging. We investigated a number of different alternatives to the two simple bootstrap methods and present below a method which had reasonably good properties for the cases considered.

For the restricted estimation problem, the distribution of  $\hat{S}_g(t) - S_g(t)$  will generally not be symmetric or centred around zero and will differ from one group  $g$  to the next. It is to be expected that the bootstrap distribution  $\hat{S}_g^b(t) - \hat{S}_g(t)$  will be similarly biased. The method we propose uses the bootstrap distribution to correct the bias, but is adjusted so as not to overcorrect. Consider pointwise estimators  $\hat{S}_1(t)$  and  $\hat{S}_2(t)$ , where  $S_1(t) \geq S_2(t)$ . Let  $\bar{S}_g^B(t)$  be the mean of the bootstrap estimates  $\hat{S}_g^b(t)$ . The basic bootstrap method considers a pseudo estimator given by  $[2h\{\hat{S}_g(t)\} - h\{\hat{S}_{g,\alpha}^B(t)\}]$ . While  $\bar{S}_1^B(t) \geq \bar{S}_2^B(t)$ , the mean of the pseudo estimator may not satisfy the order constraint, i.e., it is possible that  $2h\{\hat{S}_1(t)\} - h\{\bar{S}_1^B(t)\} < 2h\{\hat{S}_2(t)\} - h\{\bar{S}_2^B(t)\}$ . This can be considered as an overcorrection and it might be expected that the properties of the confidence interval could be improved if this overcorrection is modified. Let  $\tilde{S}_g^B(t, a_g) = h\{\hat{S}_g(t)\} + a_g[h\{\hat{S}_g(t)\} - h\{\bar{S}_g^B(t)\}]$ , where  $0 \leq a_g \leq 1$ . Although  $\tilde{S}_g^B(t, a_g)$  will satisfy the ordering constraint for  $a_g = 0$ , for  $a_g = 1$ , this may not be true. Given a set of  $a_g$  such that the  $\tilde{S}_g^B(t, a_g)$  ( $g = 1, \dots, G$ ) satisfy the ordering constraints, the proposed adjusted basic bootstrap confidence interval is  $(h^{-1}[2h\{\hat{S}_g(t)\} - h\{\hat{S}_{g,1-\alpha}^B(t)\} + \delta_g], h^{-1}[2h\{\hat{S}_g(t)\} - h\{\hat{S}_{g,\alpha}^B(t)\} + \delta_g])$ , where  $\delta_g = (a_g - 1)[h\{\hat{S}_g(t)\} - h\{\bar{S}_g^B(t)\}]$ .

We propose the following method to obtain a set of  $a_g$  that satisfy the constraints. Let  $a_1 = a_2 = \dots = a_G = a$  and find the largest  $a$  that does not result in a violation of an order restriction. Use this value of  $a$  for the groups  $i$  and  $j$  for which  $\tilde{S}_i^B(t, a_i) = \tilde{S}_j^B(t, a_j)$ . For the remaining groups increase  $a$  until a new violation is about to occur, and use the new value of  $a$  for the groups that have the active constraint for  $\tilde{S}_i^B(t, a)$  and have not already had a fixed value of  $a_i$ . Continue in this way, gradually increasing  $a$ , using the value of  $a$  when constraints become active, until all values of  $a_g$  have been set or  $a = 1$ . An algorithm to obtain  $a_g$  ( $g = 1, \dots, G$ ) is given in the Supplementary Material.

## 5.3. Confidence interval centred on a constrained estimator

Hwang & Peddada (1994) suggested a method in which a confidence interval is computed for the unrestricted estimator and then shifted and centred on the constrained estimator. They showed that, under fairly general conditions, the coverage probability for the shifted interval will exceed the nominal level. For the survivor function, we apply this to intervals on a log transformed scale and consider the approximate  $100(1 - 2\alpha)\%$  confidence interval,  $\hat{S}_g(x) \exp\{\pm z_\alpha \sigma_g^*(x)\}$ , where

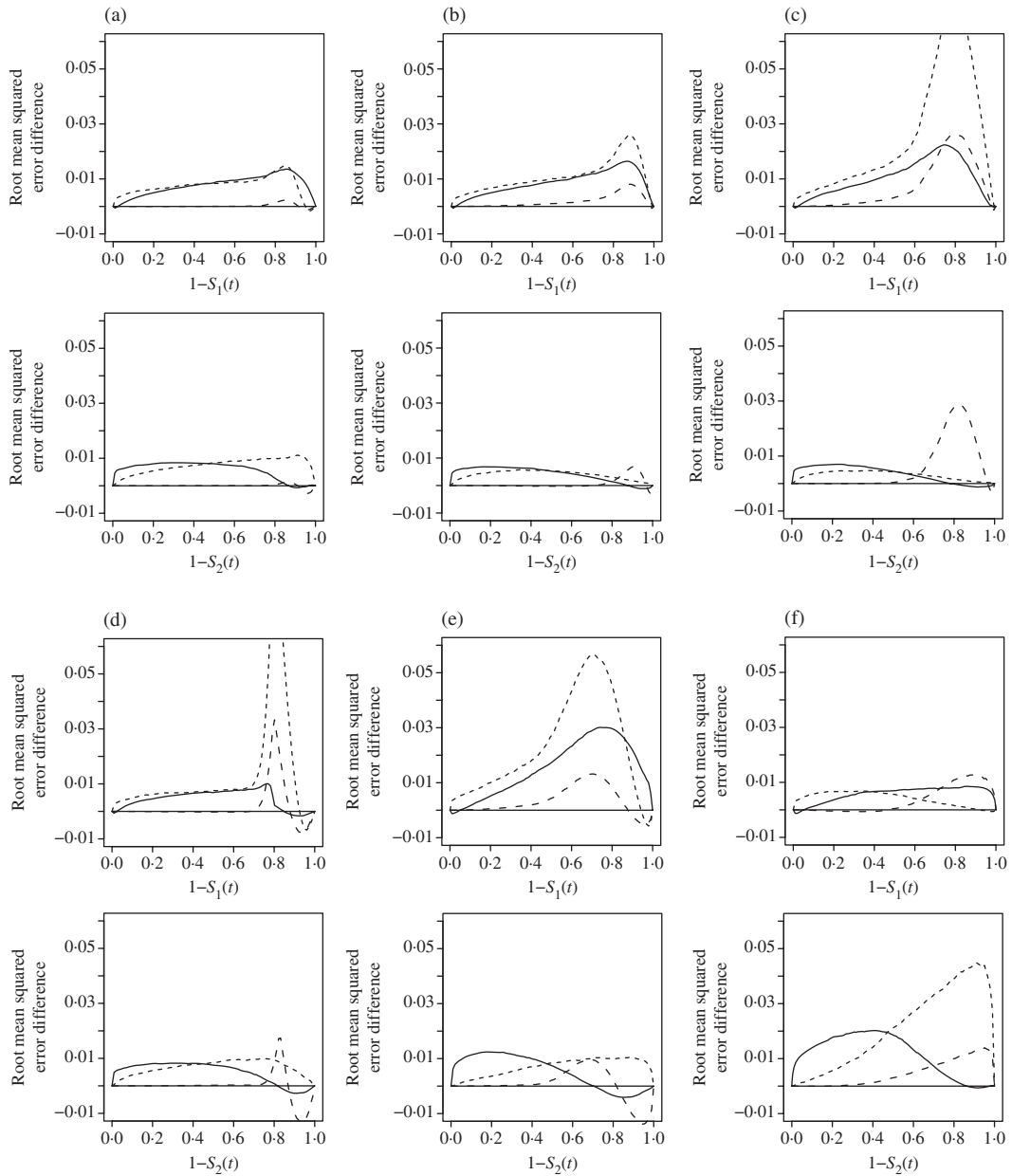


Fig. 1. Difference of root mean squared errors of estimators compared with the pointwise constrained estimator. Kaplan–Meier estimator (thin dashed); constrained nonparametric maximum likelihood estimator (thick solid); Rojo’s estimator (thick dashed); pointwise constrained estimator (thin solid). Sample sizes are 80 in cases (a)–(d) and 50 in cases (e) and (f) for both groups. Event distributions are  $\exp(1)$  for group 1 and  $\exp(1.1)$  for group 2 in cases (a)–(d), and  $U(0, 1)$  for group 1 and  $U^{1.1}(0, 1)$  for group 2 in cases (e) and (f). The censoring distributions are  $\exp(1)$  and  $\exp(1)$  in case (a),  $\exp(1)$  and no censoring in case (b),  $\exp(2)$  and no censoring in case (c),  $U(0, 1.6)$  and no censoring in case (d),  $\exp(3.2)$  and  $\exp(1.6)$  in case (e) and  $\exp(0.67)$  and  $\exp(3.2)$  in case (f) for groups 1 and 2, respectively. When event random variable follows a  $U(0, 1)$  distribution, the censoring distributions  $\exp(3.2)$ ,  $\exp(1.6)$  and  $\exp(0.67)$  give approximately 70, 50 and 30% censoring rates.

$\sigma_{g^*}(x)$  is the standard error estimate of  $\log S_g^*(x)$  (Kalbfleisch & Prentice 2002, p. 17), and  $z_\alpha$  is the  $\alpha$ th percentile of the standard normal distribution.

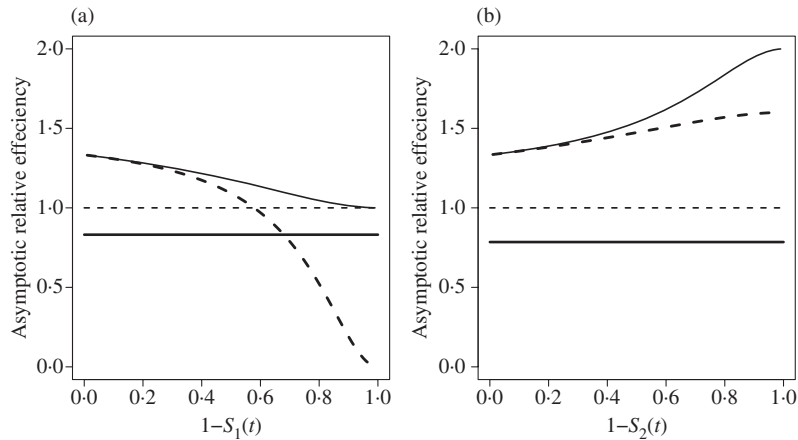


Fig. 2. Comparison of asymptotic relative efficiencies under the constraint  $T_1 \geq_{st} T_2$ . The underlying distributions are:  $S_1(t) = S_2(t) = \exp(-t)$ ,  $S_1^c(t) = 1$  and  $S_2^c(t) = \exp(-2t)$ . Kaplan–Meier estimator (thin dashed); constrained nonparametric maximum likelihood estimator (thick solid); Rojo’s estimator (thick dashed); pointwise constrained estimator (thin solid).

## 6. SIMULATION STUDIES

### 6.1. Two-sample case when sample size is small

We have conducted numerous simulation studies to compare the finite sample properties of three different constrained estimators, Rojo’s estimator (Rojo, 2004), the constrained nonparametric maximum likelihood estimator (Park et al., 2012) and the pointwise constrained estimator, and compared them to the unconstrained Kaplan–Meier estimator in the two-sample case. In this paper, we show results for scenarios where  $G = 2$  and  $S_1(t) \geq S_2(t)$  for all  $t$ .

The upper and lower plots of each panel in Fig. 1 show differences of root mean squared errors of estimators of  $S_1(t)$  and  $S_2(t)$  over a range of values of  $t$ , compared with the pointwise constrained estimator. In cases with the same censoring distributions, Fig 1(a), Rojo’s estimator and the pointwise constrained estimator have smaller root mean squared error than the other estimators. However, if populations 1 and 2 have different censoring distributions, the pointwise constrained estimator has smallest root mean squared error among all estimators at almost all times. Rojo’s estimator does not adjust well to the unequal censoring distributions, Figs. 1(b)–(f), even when the censoring rates are close to each other, Fig. 1(d). The pointwise constrained estimator is the only estimator that dominates the Kaplan–Meier estimator at almost all times in all situations considered. Each simulation is based on 10 000 replications.

### 6.2. Two-sample case: asymptotic properties

We define the asymptotic relative efficiency as the inverse ratio of the mean squared errors and compare the asymptotic relative efficiency of the three constrained estimators to the Kaplan–Meier estimator in the two sample case in Fig. 2. The underlying distributions are  $S_1(t) = S_2(t) = \exp(-t)$ ,  $S_1^c(t) = 1$  and  $S_2^c(t) = \exp(-2t)$ . The constraint is asymptotically relevant at all times. We set  $\lim_{n_1, n_2 \rightarrow \infty} n_1/n_2 = 1$ . The asymptotic relative efficiency of the full constrained nonparametric maximum likelihood estimator is based on simulated data with a very large sample size. Asymptotic relative efficiencies of the pointwise constrained estimator and Rojo’s estimator are calculated using (9) and (10).

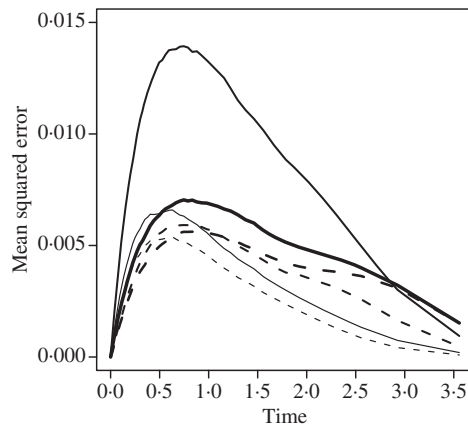


Fig. 3. Comparison of the Kaplan–Meier estimator and pointwise constrained estimator in the three sample case. Comparison of the Kaplan–Meier estimates for groups 1 (black solid), 2 (thick solid) and 3 (thin solid) and pointwise constrained estimator for group 1 (black dashed), 2 (thick dashed) and 3 (thin dashed) case.

The pointwise constrained estimator dominates all other estimators for all  $t$ , whereas Rojo’s estimator could be inefficient for some  $t$ , as seen in Fig. 2(a). Compared with the Kaplan–Meier estimator, the full constrained nonparametric maximum likelihood estimator is less efficient at all times in this setting.

### 6.3. Simple ordering case

In this section, we compare finite sample properties of the pointwise constrained estimator with the Kaplan–Meier estimator in the simple ordering case and investigate the confidence intervals described in § 5. We consider three groups with underlying distributions  $T_1 \sim \exp(1)$ ,  $T_2 \sim \exp(1.1)$  and  $T_3 \sim \exp(1.4)$  and a uniform censoring distribution  $C \sim U(0, 4.3)$ , which gives an overall censoring rate of about 20%. Sample sizes are  $n_1 = n_3 = 40$  and  $n_2 = 20$ . The simulation is based on 10 000 replicates.

Figure 3 shows the mean squared error of the pointwise constrained estimator and the Kaplan–Meier estimator. The figure shows efficiency gains for the pointwise constrained estimator at all times for all groups, with the largest gains for the estimation of  $S_2(t)$ , where the mean squared error of the pointwise constrained estimator is less than half of the mean squared error of the Kaplan–Meier estimator at almost all times.

Bootstrap intervals are based on 1999 bootstrap estimates. We evaluate confidence intervals at time 0.26 and 0.63, where the survival rates of group 2 are 0.75 and 0.5, respectively. In addition, we also conducted a simulation study for additional two cases with different distributions; see Table 1. The coverage rates and average widths of the confidence intervals described in § 5 are shown in Table 1. As expected, the confidence interval centred on the pointwise constrained estimator,  $\hat{S}_g \exp(\pm 1.96\sigma_g^*)$ , is overly conservative with large average width and has higher coverage rate. The bootstrap methods give confidence intervals with significantly reduced widths, but the coverage rates can be somewhat low for some groups, especially when using the percentile or the basic bootstrap methods. The transformation and the adjusted methods described in § 5 both give slightly better coverage rates. The overall best results are obtained with the combination of the basic bootstrap with  $\arcsin(s^{1/2})$  transformation and controlling for bias overcorrection.

## 7. EXAMPLE

The data are from prostate cancer patients who received radiation therapy at the University of Michigan Hospital, a portion of the data used in Proust-Lima & Taylor (2009). Five hundred

Table 1. Percent coverage, average width, of nominal 95% confidence intervals

	$t = 0.26$			$t = 0.63$		
Distribution	exp(1)	exp(1.1)	exp(1.4)	exp(1)	exp(1.1)	exp(1.4)
Percentile	91 (20.8)	95 (22.9)	95 (24.9)	94 (27.2)	95 (28.2)	94 (27.4)
Basic	89 (20.8)	90 (22.9)	91 (24.9)	91 (27.2)	88 (28.2)	90 (27.4)
With adjustment	90 (20.8)	92 (22.9)	93 (24.9)	92 (27.2)	91 (28.2)	91 (27.4)
$\arcsin(s^{1/2})$	94 (22.5)	93 (23.1)	93 (24.4)	93 (27.3)	90 (28.0)	92 (27.8)
With adjustment	95 (22.4)	94 (23.2)	94 (24.5)	94 (27.3)	93 (28.1)	94 (27.8)
$\hat{S}_g \exp\{\pm 1.96\sigma_g^*\}$	92 (26.7)	98 (37.8)	97 (28.3)	95 (33.6)	99 (46.5)	97 (31.7)
Distribution	exp(1)	exp(1.05)	exp(1.2)	exp(1)	exp(1.05)	exp(1.2)
Percentile	90 (20.2)	96 (21.5)	94 (23.4)	92 (26.2)	96 (26.5)	94 (26.6)
Basic	90 (20.2)	92 (21.5)	93 (23.4)	92 (26.2)	91 (26.5)	92 (26.6)
With adjustment	91 (20.2)	94 (21.5)	94 (23.4)	92 (26.2)	93 (26.5)	93 (26.6)
$\arcsin(s^{1/2})$	95 (21.9)	94 (21.8)	94 (22.7)	94 (26.4)	93 (26.4)	94 (26.8)
With adjustment	96 (21.7)	95 (21.9)	95 (22.9)	95 (26.4)	95 (26.4)	95 (26.8)
$\hat{S}_g \exp\{\pm 1.96\sigma_g^*\}$	93 (26.9)	98 (37.6)	98 (27.1)	95 (34)	99 (47.2)	98 (31.5)
Distribution	exp(1)	exp(1.2)	exp(1.6)	exp(1)	exp(1.2)	exp(1.6)
Percentile	92 (21.6)	95 (24.6)	95 (26.2)	94 (28.1)	95 (29.8)	94 (27.5)
Basic	90 (21.6)	87 (24.6)	90 (26.2)	91 (28.1)	86 (29.8)	89 (27.5)
With adjustment	91 (21.6)	90 (24.6)	92 (26.2)	92 (28.1)	89 (29.8)	91 (27.5)
$\arcsin(s^{1/2})$	93 (23.2)	90 (24.7)	92 (25.8)	93 (28.2)	88 (29.6)	92 (28.0)
With adjustment	95 (23.1)	93 (24.8)	94 (25.9)	94 (28.2)	92 (29.7)	94 (28.0)
$\hat{S}_g \exp\{\pm 1.96\sigma_g^*\}$	93 (26.6)	98 (38.9)	96 (29.4)	95 (33.2)	99 (47.0)	96 (31.3)

Sample sizes are  $n_1 = 40$ ,  $n_2 = 20$  and  $n_3 = 40$  and censoring distribution is  $Un(0, 4.3)$ . The five bootstrap confidence intervals are the percentile method, and the basic bootstrap method with or without  $\arcsin(s^{1/2})$  transformation and with or without an adjustment for bias overcorrection.  $\hat{S}_g \exp\{\pm 1.96\sigma_g^*\}$  is the centred method of [Hwang & Peddada \(1994\)](#). Results are based on 10 000 simulation samples.

and three patients without planned hormonal therapy are used to estimate the survivor function of time to first recurrence of prostate cancer. For this analysis, recurrence is defined as the first of local recurrence, distant metastasis or initiation of salvage hormone therapy.

It is expected that patients with higher baseline prostate-specific antigen levels have a higher recurrence rate than those with lower baseline prostate-specific antigen values. The Gleason grade is a measure of the aggressiveness of the tumour cells obtained from microscopic inspection of a biopsy prior to the treatment. It is also expected that patients with a lower Gleason grade will have a lower recurrence rate. In this example, we divided the patients into six groups labelled A1, A2, A3, B1, B2 and B3 based on whether or not their baseline prostate-specific antigen is less than 10, and whether their Gleason grade is  $\leq 6$ ,  $= 7$  or  $\geq 8$ . Patients with baseline prostate-specific antigen  $< 10$  and Gleason  $\leq 6$  are labelled as A1, patients with baseline prostate-specific antigen  $< 10$  and Gleason  $= 7$  as A2 etc. The natural set of constraints for the survivor functions are  $A1 \geq A2 \geq A3$ ,  $B1 \geq B2 \geq B3$ ,  $A1 \geq B1$ ,  $A2 \geq B2$  and  $A3 \geq B3$ .

The Kaplan–Meier estimates of each group are shown in Fig. 4(a). The unrestricted Kaplan–Meier estimates do not satisfy the stochastic ordering constraints. Specifically, between 1 and 2.5 years, the groups A2, B2 and B3 do not satisfy the ordering constraints and after 5 years the orderings of A2 and A3, and B2 and B3 are incorrect.

The pointwise constrained estimates, shown in Fig. 4(b), satisfy the stochastic ordering constraints at all times. Between 1 and 2.5 years, the survivor functions take a common value in groups A2, B2 and B3 and after 5 years, groups A2 and A3 and groups B2 and B3 have common estimates. At around 12.5 years, there is a jump in the survivor function estimate for groups B2

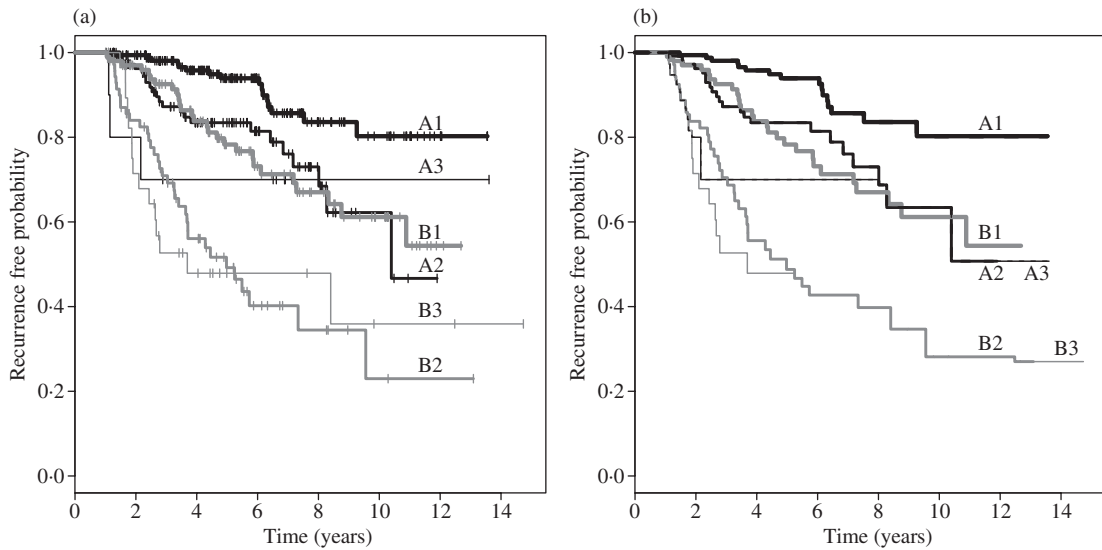


Fig. 4. Estimates in the University of Michigan prostate cancer example. The vertical bars in (a) are observed censoring times. For the labels of the lines, A and B indicate, respectively, low and high baseline prostate-specific antigen; and 1, 2 and 3 indicate, respectively, low, medium and high Gleason grade. (a) Kaplan–Meier estimator and (b) pointwise constrained estimator.

Table 2. *Estimates and confidence intervals (%) of survivor functions for some selected times in the prostate cancer example*

	Time (years)	1.5	5	8
A1	Kaplan–Meier estimator	99.4 (97.6, 100)	93.9 (90.0, 97.2)	83.6 (76.4, 91.1)
	Pointwise constrained estimator	99.4 (97.6, 99.9)	93.9 (90.0, 97.2)	83.6 (76.4, 90.2)
A2	Kaplan–Meier estimator	99.1 (96.6, 99.9)	83.4 (75.6, 90.2)	73.0 (63.2, 82.3)
	Pointwise constrained estimator	99.1 (96.6, 99.9)	83.4 (76.6, 89.9)	73.0 (64.0, 83.0)
A3	Kaplan–Meier estimator	80.0 (36.0, 98.0)	70.0 (44.8, 92.7)	70.0 (44.8, 100)
	Pointwise constrained estimator	88.7 (70.9, 94.9)	70.0 (53.2, 91.6)	70.0 (52.6, 93.6)
B1	Kaplan–Meier estimator	98.0 (92.2, 99.7)	78.3 (71.2, 86.1)	67.0 (57.8, 77.8)
	Pointwise constrained estimator	98.0 (95.2, 99.7)	78.3 (71.2, 86.1)	67.0 (57.8, 77.8)
B2	Kaplan–Meier estimator	86.8 (79.6, 93.6)	48.8 (38.9, 62.2)	34.2 (22.9, 52.6)
	Pointwise constrained estimator	88.7 (83.0, 93.9)	48.8 (39.6, 59.9)	39.8 (30.3, 54.0)
B3	Kaplan–Meier estimator	96.4 (86.2, 99.8)	47.9 (33.4, 67.5)	47.9 (33.5, 69.4)
	Pointwise constrained estimator	88.7 (83.5, 93.8)	47.9 (38.4, 69.2)	39.8 (29.9, 54.2)

Nominal 95% bootstrap confidence intervals using  $\arcsin(s^{1/2})$  transformation and controlling bias overcorrection are shown in parentheses.

and B3, even though there are no observed events at that time. This happens because the number of individuals at risk in the stochastically smaller group B3 at time  $t = 12.5$  changes, which results in  $\hat{h}^\delta(t) < 0$ , as discussed in § 2.3.

Detailed results of point estimates and corresponding confidence intervals for some selected times are shown in Table 2.

### 8. DISCUSSION

The pointwise constrained estimator is a likelihood based pointwise estimator. Unlike the full constrained nonparametric maximum likelihood estimator, the violation of a constraint at one time does not affect the estimates at other times. The pointwise constrained estimator gives a

common estimate based on maximizing the likelihood when the constraints are violated and compared with other estimators that use averaging based on initial sample sizes (Rojo, 2004; El Barmi & Mukerjee, 2005), it has better properties when censoring exists.

When there is no censoring, Rojo's estimator in the two-sample case and El Barmi and Mukerjee's estimator in the simple ordering case are identical to the pointwise constrained estimator. However, if censoring exists, these estimators can be quite different, especially when the censoring distributions differ significantly between groups. Another feature of El Barmi and Mukerjee's estimator is the range of times for which the estimator is defined. Specifically, it is defined only until the minimum of the times of the last observations in all groups. Thus, if the last observed time in one group is much earlier than in other groups, then estimates in all other groups are undefined at subsequent times even though there may be a large number of observations at risk. On the other hand, the pointwise constrained estimator for a group is defined up to the last observed time of that group.

The pointwise constrained estimator can have jumps at nonevent times. Thus, the likelihood ratio statistics of the restricted survivor function, first introduced by Thomas & Grunkemeier (1975) and discussed by Li (1995) and Murphy (1995) are not exactly correct, because they assume the jumps can only occur at event times. Thus, the likelihood ratio test and confidence interval based on the likelihood ratio test may need to be revised.

Methods to construct confidence intervals in order restricted problems are not well developed. Bootstrap methods generally work better when the distributions are approximately normal after some transformations. When constraints are present, it is not clear whether there exists any such transformation. We proposed a method to control overcorrection of bias when using the basic bootstrap methods and found improved properties of confidence intervals. Further investigation of this approach on other applications could be useful.

#### ACKNOWLEDGEMENT

This research was partially supported by the National Institutes of Health, U.S.A. We thank the editor, associate editor and two referees for their many valuable comments.

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the generalized pool-adjacent-violators algorithm for the simple ordering case in § 2.5, proofs of Theorems 2–4, and an algorithm to calculate  $a_g$  described in § 5.2.

#### APPENDIX

*Proof of Theorem 1.* Let  $\lambda_1$  and  $\lambda_2$  be Lagrange multipliers. The corresponding Lagrangian function is

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

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

$$-\frac{d_i \exp \hat{h}(X_i)}{1 - \exp \hat{h}(X_i)} + (n_i - d_i) + \hat{\lambda}_1 = 0, \quad i \leq M(x), \quad (\text{A1})$$

$$-\frac{d_i \exp \hat{h}(X_i)}{1 - \exp \hat{h}(X_i)} + (n_i - d_i) = 0, \quad i > M(x), \tag{A2}$$

$$N(x) + \hat{\lambda}_1 - \hat{\lambda}_2 = 0, \tag{A3}$$

$$\sum_{j=1}^{M(x)} \hat{h}(X_j) + \hat{h}^\delta(x) - q = 0, \tag{A4}$$

$$\hat{h}^\delta(x) \leq 0, \tag{A5}$$

$$\hat{\lambda}_2 \hat{h}^\delta(x) = 0, \tag{A6}$$

$$\hat{\lambda}_2 \geq 0. \tag{A7}$$

From (A1), we have  $\hat{h}_1(X_i) = \log\{1 - d_i/(n_i + \hat{\lambda}_1)\}$  for  $i \leq M_1(x)$ . Either  $\hat{\lambda}_2 = 0$  or  $\hat{h}^\delta(x) = 0$  from (A6). If  $\hat{\lambda}_2 = 0$ , then  $\hat{\lambda}_1 = -N(x)$  from (A3), which is only valid when  $\hat{h}^\delta(x) = q - \sum_{j=1}^{M(x)} \log[1 - d_i/\{n_i - N(x)\}] \leq 0$ ; otherwise  $\hat{h}^\delta(x) = 0$  and  $\hat{\lambda}_1$  is the solution of the equation,  $q - \sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + \lambda)\} = 0$ , from (A4), which is only valid when  $\hat{\lambda}_1 \geq -N(x)$  from (A3). Since  $\sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + k)\}$  is an increasing function in  $k$ , we can see that  $\hat{\lambda}_1 = \max\{\hat{k}, -N(x)\}$ , where  $\hat{k}$  is the solution of the equation  $\sum_{j=1}^{M(x)} \log\{1 - d_i/(n_i + k)\} - q = 0$ . It follows that  $\hat{\lambda}_1$  is exactly the same as  $K(q; x)$  defined in Theorem 1. Therefore, the unique solution from solving (A2)–(A7) is as given in Equation (4).  $\square$

*Proof of Lemma 1.* We consider separately two cases where  $K(q; x) > -N(x)$  and  $K(q; x) = -N(x)$ .

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

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

Thus,

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

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

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

$\square$



## REFERENCES

- ANDREWS, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68**, 399–405.
- BARLOW, R. E., BARTHOLOMEW, D., BREMNER, J. M. & BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: Wiley.
- BEST, M. J., CHAKRAVARTI, N. & UBHAYA, V. A. (1999). Minimizing separable convex functions subject to simple chain constraints. *SIAM J. Optimiz.* **10**, 658–72.
- 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. *J. Am. Statist. Assoc.* **61**, 1067–80.
- DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.
- DYKSTRA, R. L. (1982). Maximum likelihood estimations of the survival functions of stochastically ordered random variables. *J. Am. Statist. Assoc.* **77**, 621–8.
- EFRON, B. & TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- EL BARMÍ, H. & MUKERJEE, H. (2005). Inferences under a stochastic ordering constraint: the  $k$ -sample case. *J. Am. Statist. Assoc.* **100**, 252–61.
- GILL, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49–58.
- HOFF, P. D. (2003). Nonparametric estimation of convex models via mixtures. *Ann. Statist.* **31**, 174–200.
- HWANG, J. T. G. & PEDDADA, S. D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.* **22**, 67–93.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. New York : Wiley.
- KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457–81.
- KUHN, H. W. & TUCKER, A. W. (1951). Nonlinear programming. In *Proc. 2nd Berkeley Symp.*, Ed. J. Neyman, pp. 481–92. Berkeley: University of California Press.
- LEHMANN, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26**, 399–419.
- LI, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Ann. Statist.* **25**, 95–104.
- LIM, J., KIM, S. J. & WANG, X. (2009). Estimating stochastically ordered survival functions via geometric programming. *J. Comp. Graph. Statist.* **18**, 978–94.
- LO, S. H. (1987). Estimation of distribution functions under order restrictions. *Statist. Dec.* **5**, 251–62.
- MURPHY, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *J. Am. Statist. Assoc.* **90**, 1399–405.
- PARK, Y., KALBFLEISCH, J. D. & TAYLOR, J. M. G. (2012). Constrained nonparametric maximum likelihood estimation of stochastically ordered survivor functions. *Can. J. Statist.* **40**, 22–39.
- PROUST-LIMA, C. & TAYLOR, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* **10**, 535–49.
- ROJO, J. (2004). On the estimation of survival functions under a stochastic order constraint. In *The First Erich L. Lehmann Symposium: Optimality*, Ed. V. Pérez-Abreu and J. Rojo, vol. 44, pp. 37–61. Beachwood, OH: Institute of Mathematical Statistics.
- ROJO, J. & MA, Z. (1996). On the estimation of stochastically ordered survival functions. *J. Statist. Comp. Simul.* **55**, 1–21.
- STUTE, W. & WANG, J.-L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591–607.
- THOMAS, D. R. & GRUNKEMEIER, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Am. Statist. Assoc.* **70**, 865–71.

[Received January 2011. Revised December 2011]