Biometrika (2013), **xx**, x, *pp.* 1–9

Printed in Great Britain

Supplementary Material for 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 jmgt@umich.edu jdkalbfl@umich.edu

APPENDIX 3

Generalized pool-adjacent-violators Algorithm for the simple ordering case in section 2·5

We present an algorithm for the case $T_1 \geq_{st} \cdots \geq_{st} T_G$. Let J be a partition of $\{1, \ldots, G\}$, so that $J = \{B_1, B_2, \ldots\}$. Each member of J is called a block. An optimal solution from Algorithm 1 only contains blocks with consecutive integers. Let $B = \{a, \ldots, b\}$ $(1 \leq a \leq b \leq G)$, then B^- is the block that contains $a-1$ or $\varnothing$ if $a = 1$ and B^+ is the block that contains $b+1$ or $\varnothing$ if $b = G$. For a given block B, $S_B(x) = \exp(\hat{q}_B)$, where $\hat{q}_B$ maximizes the log profile likelihood, $\ell_B(q; x) = \sum_{i \in B} \ell_i(q; x)$. From Lemma 1, $\sum_{i \in B} d\ell_i(q; x)/dq = -\sum_{i \in B} K_i(q; x)$. Thus, the maximizer $\hat{q}_B$ is the root of the equation $\sum_{i \in B} K_i(q; x) = 0$, provided there is at least one failure in the block B prior to time t. Otherwise $\hat{q}_B = 0$.

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

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

Algorithm 1 yields the partition, $\hat{J} = \{\hat{B}_1, \ldots, \hat{B}_r\}$. If $i \in \hat{B}_j$, then $S_i(x) = S_{\hat{B}_j}(x)$, which is the pointwise constrained estimator at x. It can be seen that either $\hat{S}_g(x) = 1$ or $N_g(x) = 0$ for all $g < L$ and either $\hat{S}_g(x) = 0$ or $N_g(x) = 0$ for all $g > U$, where $L = \min\{g : M_g(x) > 0\}$ or $L = G + 1$ if none exists, and $U = \max\{g : N_g(x) > 0$ and $S^*_g(x) > 0\}$ or $U = 0$ if none exists.

APPENDIX 4

Proof of Theorem 2

Notation and characteristics of the pointwise constrained estimator

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

LEMMA 2. *A partition $J_\xi(x)$ with corresponding estimate $S_\xi(B; x)$ is the pointwise constrained estimator subject to the constraints E_ξ at time x if and only if*

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

Proof. Sufficiency. Since the joint log profile likelihood $\ell_\xi(q; x)$ for populations in ξ as shown in equation (6) is a separable concave function, if the condition (ii) in Lemma 2 is satisfied, the estimate will be the pointwise constrained estimator subject to constraints $\cup_{B \in J_\xi(x)} E_B$. It follows that if condition (i) in Lemma 2 is also satisfied, the estimate must be the pointwise constrained estimator subject to constraints E_ξ because $\cup_{B \in J_\xi(x)} E_B \subset E_\xi$ and adding more constraints can not increase the likelihood.

Necessity. Obviously condition (i) holds in Lemma 2 if $S_\xi(B; x)$ is the pointwise constrained estimator. If we write down the Karush–Kuhn–Tucker conditions (Kuhn & Tucker, 1951) needed for maximizing the log profile likelihood subject to the constraints E_ξ, the Lagrangian multipliers related to the constraint $(i, j) \in E_\xi$ for any i and j in different block of $J_\xi(x)$ will be zero in the solution because these constraints are inactive in the solution. Thus if we delete these zero valued Lagrangian multipliers, the remaining Karush–Kuhn–Tucker conditions of the populations in any $B \in J_\xi(x)$ are exactly the same as the pointwise constrained estimator subject to the constraints E_B. Since the constraints are linear and the joint log profile likelihood is concave, the Karush–Kuhn–Tucker conditions are also sufficient in our problem. Thus the condition (ii) in Lemma 2 must also hold. □

Lemma 2 is useful in later proofs because it enables us to consider blocks separately. If $\hat{B}$ is a block from the pointwise constrained estimator subject to constraint E_ξ for any subpopulation ξ at time x, $\hat{S}_\xi(\hat{B}; x) = \hat{S}_{\hat{B}}(\hat{B}; x)$ will remain the same for any subpopulation ξ for the same block $\hat{B}$. So we use $\hat{S}(\hat{B}; x)$ as the estimate of the pointwise constrained estimator at time x if $\hat{B} \in \hat{J}_\xi(x)$.

We give two more lemmas to characterize the pointwise constrained estimator and $K_g(q; x)$.

LEMMA 3. *(a) For any $x_2 > x_1 > 0$ and $q \leq 0$, $K_g(q, x_2) \geq K_g(q, x_1)$;*
(b) For any $q_2 < q_1 \leq 0$ and $x > 0$, $K_g(q_1, x) \geq K_g(q_2, x)$, and the equality holds only when $K_g(q_1; x) = K_g(q_2; x) = -N_g(x)$.

Proof of Lemma 3(a). We consider separately two cases.

First, if there is no observed event before or at x_1, then $K_g(q, x_1) = -N_g(x_1) \leq -N_g(x_2) \leq K_g(q, x_2)$.

Second, if there is at least one observed event before or at x_1, let $\hat{k}_j$ be the solution of the equation

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

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

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

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

Proof of Lemma 3(b). Suppose there is at least one observed event before or at x, and let $\hat{k}_j$ be the solution of the equation

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

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

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

LEMMA 4. *For any $\hat{B} \in \hat{J}_\xi(x)$, (a) $\sum_{g\in\hat{B}} K_g\{\log \hat{S}(\hat{B}; x); x\} \le 0$, and the strict inequality holds only when $\hat{S}(\hat{B}; x) = 1$; (b) for any $\hat{S}(\hat{B}; x) < 1$ and $B_1 \subset \hat{B}$, the following two conditions will not hold simultaneously: (i) For all $i \in B_1$ and $j \in \hat{B}/B_1$, $(j, i) \notin E_\xi$; (ii) $\sum_{g\in B_1} K_g\{\log \hat{S}(\hat{B}; x); x\} < 0$.*

Proof of Lemma 4(a). Profile likelihood $\sum_{g\in\hat{B}} \ell_g(q; x)$ is a concave function of q and so the $\hat{S}(\hat{B}; x)$ must satisfy

$$\sum_{g\in\hat{B}} K_g\{\log \hat{S}(\hat{B}; x); x\} = -\sum_{g\in\hat{B}} \frac{d}{dq}\ell_g\{\hat{S}(\hat{B}; x); x\} = 0.$$

The only one exception is when there is no observed event time before or at x for all $g \in \hat{B}$, in this case $\sum_{g\in\hat{B}} K_g\{\log \hat{S}(\hat{B}; x); x\} = -\sum_{g\in\hat{B}} N_g(x) < 0$ and $\hat{S}(\hat{B}; x) = 1$. □

Proof of Lemma 4(b). Suppose we can find a block B_1 satisfying both conditions (i) and (ii), since

$$\frac{d}{dq}\sum_{g\in B_1} \ell_g\{\log \hat{S}_\xi(B; x); x\} = -\sum_{g\in B_1} K_g\{\log \hat{S}(\hat{B}; x); x\} > 0,$$

we can increase estimate $S_\xi(B_1; x)$ to increase the log profile likelihood without violating the constraints. This contradicts $\hat{J}_\xi(x)$ is the partition of the pointwise constrained estimator at time x. □

An algorithm to obtain the pointwise constrained estimator at a time $x_2 > x_1$

For any $x_2 > x_1$, it can be seen that $\hat{S}_g(x_1) = \hat{S}_g(x_2)$ $(g = 1, \ldots, G)$ if there is no observation between x_1 and x_2, nor a censoring at x_1, nor an event at x_2. Now we consider the situation when only one group g^* has observations between x_1 and x_2. In this case, Algorithm 2 defines a method to obtain $\hat{J}_\xi(x_2)$ and $\hat{S}_\xi(\hat{B}; x_2)$, where $\xi = \{1, \ldots, G\}$. The idea is to find the pointwise constrained estimator at x_2 using the estimate at x_1 as the starting point.

To illustrate the algorithm, we first show an example in Fig. 5. In this, $\hat{J}_\xi(x_1)$ has five blocks, $\hat{B}_1, \ldots, \hat{B}_5$ and $g^* \in \hat{B}_2$. At first, $r = 2$ and $A_2 = \hat{B}_2$. Then we find $\hat{J}_{A_2}(x_2)$, the partition of the pointwise constrained estimator subject to constraints E_{A_2} at time x_2 and assume that it has four blocks $\hat{B}_{2.1}, \ldots, \hat{B}_{2.4}$ where $\hat{S}(\hat{B}_{2.1}; x_2) > \hat{S}(\hat{B}_{2.2}; x_2) > \hat{S}(\hat{B}_3, x_1) \ge \hat{S}(\hat{B}_{2.3}; x_2) > \hat{S}(\hat{B}_{2.4}; x_2)$. The blocks $\hat{B}_{2.1}$ and $\hat{B}_{2.2}$ remain separate in the solution and blocks $\hat{B}_3$, $\hat{B}_{2.3}$ and $\hat{B}_{2.4}$ are combined into A_3. Then we again find $\hat{J}_{A_3}(x_2)$ and assume that it has two blocks $\hat{B}_{3.1}$ and $\hat{B}_{3.2}$ where $\hat{S}(\hat{B}_{3.1}; x_2) > \hat{S}(\hat{B}_{3.2}; x_2) > \hat{S}(\hat{B}_4; x_1)$. Blocks $\hat{B}_{3.1}$ and $\hat{B}_{3.2}$ remain separate in the solution and the algorithm ends. The final partition $J_\xi(x_2)$ contains blocks $\hat{B}_1, \hat{B}_{2.1}, \hat{B}_{2.2}, \hat{B}_{3.1}, \hat{B}_{3.2}, \hat{B}_4$ and $\hat{B}_5$.

LEMMA 5. *Algorithm 2 gives the pointwise constrained estimator at x_2 and the estimate for each group is nonincreasing over time.*

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

Algorithm 2: An algorithm to obtain the pointwise constrained estimator at time x_2 using the pointwise constrained estimator at time x_1 as the starting value, where $x_2 > x_1$ and only population g^* has observations between x_1 and x_2. Below $\xi = \{1, \ldots, G\}$.

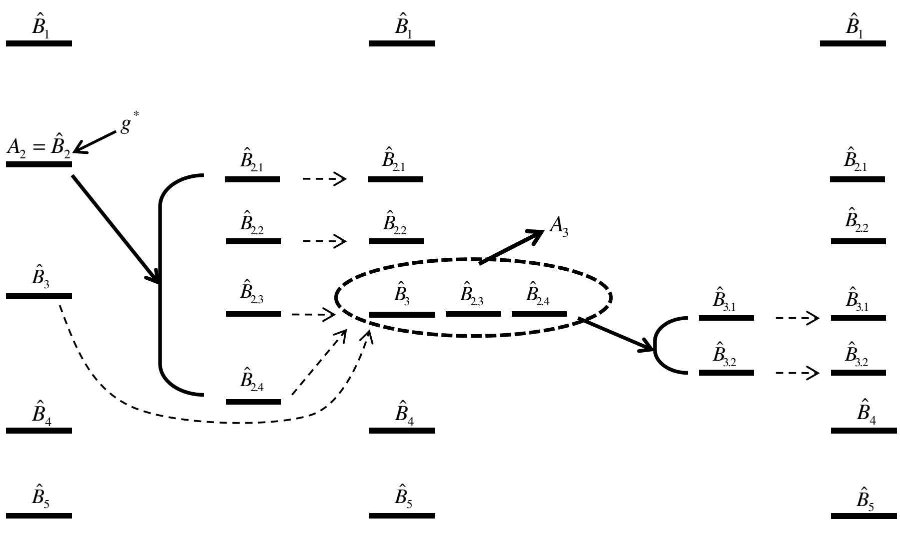

Fig. 5: An example of Algorithm 2 in Appendix 4.

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

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

Suppose $\hat{S}(\hat{B}_r; x_1) < \hat{S}(\hat{B}_{r.1}; x_2)$, then this will gives a contradiction. There are two cases to consider. First, we consider the first step in Algorithm 2. Here $g^* \in \hat{B}_r$ and $A_r = \hat{B}_r$ from line 2 in Algorithm 2. Then

$$\begin{aligned}
\sum_{g \in \hat{B}_{r.1}} K_g\{\log \hat{S}(\hat{B}_r; x_1); x_1\} &\leq \sum_{g \in \hat{B}_{r.1}} K_g\{\log \hat{S}(\hat{B}_r; x_1); x_2\} && \{\text{Lemma 3(a)}\} \\
&\leq \sum_{g \in \hat{B}_{r.1}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} && \{\text{Lemma 3(b)}\} \qquad \text{(A1)} \\
&\leq 0 && \{\text{Lemma 4(a)}\}.
\end{aligned}$$

From Lemma 3(b), equality holds in (A1) only when $K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} = -N_g(x_2)$ for all $g \in \hat{B}_{r.1}$. By our convention to set the estimate of a survivor function as low as possible when the number at risk is zero, $\hat{S}(\hat{B}_{r.1}; x_2) = 0$ if $N_g(x_2) = 0$ for all $g \in \hat{B}_{r.1}$. Since $\hat{S}(\hat{B}_{r.1}; x_2) > \hat{S}(\hat{B}_r; x_1) \geq 0$ by our assumption, we have $\sum_{g \in \hat{B}_{r.1}} N_g(x) > 0$. Hence we find that $\sum_{g \in \hat{B}_{r.1}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_1); x_1\} < 0$, which implies that $\hat{B}_{r.1} \subset \hat{B}_r$ and both conditions in Lemma 4(b) are satisfied. This contradicts that $\hat{B}_r \in \hat{J}_\xi(x_1)$.

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

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

we find that either there is at least one ℓ' that satisfies $\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} < 0$ or $\sum_{g \in B^*_\ell} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} = 0$ $(\ell = \ell_{r-1}, \ldots, L_{r-1})$.

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

$$\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_r; x_1); x_1\} \leq \sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_r; x_1); x_2\} \leq \sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} = 0; \tag{A2}$$

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

$$\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r-1.\ell'}; x_2); x_2\} \leq \sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} = 0. \tag{A3}$$

Neither the equality in equation (A2) nor the equality in equation (A3) can hold since otherwise $\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r.1}; x_2); x_2\} = \sum_{g \in B^*_{\ell'}} N_g(x_2) < 0$. Hence we find that $\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_r; x_1); x_1\} < 0$ in equation (A2) or $\sum_{g \in B^*_{\ell'}} K_g\{\log \hat{S}(\hat{B}_{r-1.\ell'}; x_2); x_2\} < 0$ in equation (A3), which contradicts $\hat{B}_r \in \hat{J}_\xi(x_1)$ or $\hat{B}_{r.\ell'} \in \hat{J}_{A_r}(x_2)$.

Thus, we established that $\hat{S}(\hat{B}_1; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{u-1}; x_2) \geq \hat{S}(\hat{B}_u; x_1) \geq \hat{S}(\hat{B}_{u.1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{u.L_u}; x_2) \geq \hat{S}(\hat{B}_{u+1}; x_1) \geq \hat{S}(\hat{B}_{(u+1).1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_{w.L_w}; x_2) \geq \hat{S}(\hat{B}_{w+1}; x_2) \geq \cdots \geq \hat{S}(\hat{B}_R; x_2)$. It is easy to see that the constraints are not violated in the solution $J_\xi(x_2)$ because $\hat{S}(\hat{B}_{u-1}; x_2) \geq \hat{S}(\hat{B}_{u.1}; x_2)$, $\hat{S}(\hat{B}_{r.L_r}; x_2) \geq \hat{S}(\hat{B}_{(r+1).1}; x_2)$ $(r = u, \ldots, w-1)$, and

$\hat{S}(\hat{B}_{w.L_w}; x_2) \geq S(\hat{B}_{w+1}; x_2)$. Therefore, the result from Algorithm 2 is the pointwise constrained estimator at time x_2. Furthermore, for any $g \in \hat{B}_r$ $(r = 1, \ldots, u-1, w+1, \ldots, R)$, $\hat{S}_g(x_2) = \hat{S}_g(x_1)$ since $\hat{B}_r \in J_\xi(x_2)$ and $\hat{S}(B_r; x_1) = \hat{S}(B_r; x_2)$, and for any $g \in \hat{B}_r$ $(r = u, \ldots, w)$, $\hat{S}_g(x_2) \leq \hat{S}_g(x_1)$, since $g \in \hat{B}_{r'.\ell}$ for an r' such that $r' \geq r$ and $\hat{S}(\hat{B}_r; x_1) \geq \hat{S}(\hat{B}_{r'.\ell}; x_2)$. □

To complete the proof of Theorem 2 when two or more groups have observations between x_1 and x_2, we can produce the pointwise constrained estimator by sequentially including observations from a group at a time. Since each time when we add more observations from a group, the pointwise constrained estimator will not increase compared to that before adding these observations, the pointwise constrained estimator will not increase over time.

APPENDIX 5

Proof of Theorem 3

To establish this, we first prove

LEMMA 6. $\max_{1\leq g\leq G} |S_g^*(t) - S_g(t)| \geq \max_{1\leq g\leq G} |\hat{S}_g(t) - S_g(t)|$.

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

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

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

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

A similar argument shows that

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

Thus,

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

So

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

This establishes Lemma 6. □

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

$$\lim_{n_g\to\infty} \mathrm{pr}\{\sup_{t<\tau} |\hat{S}_g(t) - S_g(t)| > \epsilon\} \leq \lim_{n_g\to\infty} \mathrm{pr}\{\sup_{t<\tau} \max_{1\leq k\leq G} |S_k^*(t) - S_k(t)| > \epsilon\} \leq \sum_{k=1}^{G} \lim_{n_k\to\infty} \mathrm{pr}\{\sup_{t<\tau} |S_k^*(t) - S_k(t)| > \epsilon\} = 0. \quad \text{(A4)}$$

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

LEMMA 7. *For a given k, let $E_k^+ = \{(k', k) \in E\}$, where $E = \{(g', g) : T_{g'} \geq_{st} T_g \ \ (g, g' = 1, \ldots, G)\}$. If $N_k(t) = 0$, then for any group g satisfying $N_g(t) > 0$, $\hat{S}_g(t) = \tilde{S}_g(t)$, where $\tilde{S}_g(t)$ is the pointwise constrained estimator subject to constraints defined by E/E_k^+, which denotes the set of con-*

straints in E excluding those in E_k^+.

Proof. The only possible situation that $\tilde{S}_g(t)$ is not the pointwise constrained estimator subject to the constraints defined by E is that there exist $(k', k) \in E_k^+$ and $\tilde{S}_{k'}(t) < \tilde{S}_k(t)$. Since $N_k(t) = 0$, the likelihood does not change if we lower the estimate for group k at time t. So set $\tilde{S}_k(t) = \min\{\tilde{S}_g(t) : (g, k) \in E_k^+\}$, then all constraints in E will be satisfied, hence $\tilde{S}_g(t), g \neq k$ is the pointwise constrained estimator. We have shown in Appendix 4 that $\hat{S}_g(t)$ is unique if $N_g(t) > 0$. Therefore $\hat{S}_g(t) = \tilde{S}_g(t)$ if $N_g(t) > 0$. □

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

LEMMA 8. *$Q_g^*(t)$ is uniformly consistent for $Q_g(t)$ on $[0, \infty)$.*

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

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

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

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

under the condition $S_g(\tau_g) = S_g(\tau_g-)$ {Corollary 1.2 in Stute & Wang 1993, page 1595}. □

Let $E(t) = E/\bigcup_{k:\tau_k<t} E_k^+$ and let $\hat{Q}_g(t)$ be the pointwise constrained estimator of $Q_g(t)$ subject to constraint $E(t)$, then the strong uniform consistency for $\hat{Q}_g(t)$ holds for all $t \geq 0$ using the same argument leading to the result in equation (A4). Since $\hat{Q}_g(t) = \hat{S}_g(t)$ by applying Lemma 7 multiple times and $Q_g(t) = S_g(t)$ for all $t < \tau_g$, the strong uniform consistency of $\hat{S}_g(t)$ for $S_g(t)$ is established on $[0, \tau_g)$. If $S_g(\tau_g-) = S_g(\tau_g)$, the strong uniform consistency of $S_g(t)$ for $S_g(t)$ holds on $[0, \tau_g]$.

This completes the proof of Theorem 3.

APPENDIX 6

Proof of Theorem 4

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

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

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

Thus,

$$n^{1/2}\{\hat{q} - \log S_g(x)\} = Z_g^L(x) + n^{1/2} \sum_{X_{gi}\leq x} K_g(\hat{q}; x) \frac{d_{gi}}{n_{gi}^2} \{1 + o_p(1)\} + O_p(n^{-1/2}). \quad \text{(A5)}$$

Since both $n^{1/2}\{\hat{q} - \log S_g(x)\}$ and $Z_g^L(x)$ are bounded in probability, $n^{1/2}\sum_{X_{gi}\le x} d_{gi}K_g(\hat{q};x)/n_{gi}^2$ must be bounded in probability. Thus equation (A5) becomes

$$n^{1/2}\{\hat{q} - \log S_g(x)\} = Z_g^L(x) + n^{1/2}K_g(\hat{q};x)\sum_{X_{gi}\le x}\frac{d_{gi}}{n_{gi}^2} + o_p(1). \quad \text{(A6)}$$

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

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

Thus $w_{gn}(x) \to c_g/\sigma_g^2(x) = w_g(x)$ as $n \to \infty$. Multiplying equation (A6) by $w_{gn}(x)$ gives

$$w_{gn}(x)n^{1/2}\{\hat{q} - \log S_g(x)\} = w_{gn}(x)Z_g^L(x) + \frac{n^{\frac{3}{2}}K_g(\hat{q};x)}{S_g^2(x)} + o_p(1). \quad \text{(A7)}$$

Since $\sum_{g=\ell}^u K_g(\hat{q};x) = 0$ for any n, summing equation (A7) over g from ℓ to u and dividing by $\sum_{g=\ell}^u w_{gn}(x)$ yields

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

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

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

in distribution.

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

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

where $L_k(x) = \min\{i : S_i(x) = S_k(x)\}$ and $U_k(x) = \max\{i : S_i(x) = S_k(x)\}$ as defined in Theorem 4. Thus

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

in distribution.

This completes the proof of Theorem 4. Now we discuss extensions and special cases of Theorem 4.

First consider the case when there exists g' such that $c_{g'} = 0$ while $c_g > 0$. The asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}$ will be the same as in equation 7 with the weight for group g' set to zero. This is because

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

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

Then we discuss the case when there are some groups for which the support of the censoring distribution is less than x. As discussed in Appendix 5, the asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}$, $x < \tau_g$, can be obtained by modifying the constraint set to $E(x)$. Ordering constraints $T_{k'} \ge_{st} T_k$ ($k' = 1, \ldots, k-1$) are removed if $x > \tau_k^*$. Also if $S_k(x) < S_k(\tau_k-)$, then constraints $T_k \ge_{st} T_{k'}$ ($k = k + 1, \ldots, G$) will be asymptotically irrelevant because $S_k^*(x)$ can always take value $S_k^*(\tau_k^*)$ and $S_k^*(\tau_k^*) > S_{k'}^*(x)$ ($k' = k+1, \ldots, G$) asymptotically. So group k can be removed from obtaining the asymptotic distribution of $n_g^{1/2}\{\hat{S}_g(x) - S_g(x)\}, x < \tau_g$ or equivalently we can set $w_k = 0$ in equation 7 at time x.

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

APPENDIX 7

Algorithm to calculate a_g $(g = 1, \ldots, G)$ in section 5·2

```
Initialization: A = {1, ..., G}, a_1 = ··· = a_G = 1;
while A ≠ ∅ do
    foreach i ∈ A do
        a*_i = 1;
        foreach j ∈ {1, ..., G} do
            if (i, j) ∈ E and S̃_i(t, a_i) < S̃_j(t, a_j) or (j, i) ∈ E and S̃_j(t, a_j) < S̃_i(t, a_i)
            then
                if j ∈ A then
                    ã = {a : S̃_i(t, a) = S̃_j(t, a)};
                else
                    ã = {a : S̃_i(t, a) = S̃_j(t, a_j)};
                end
                a*_i = min{a*_i, ã};
            end
        end
    end
    foreach i ∈ A and a*_i = min_g{a*_g} do
        a_i = a*_i;
        A = A/{i};
    end
end
```

Algorithm 3: Algorithm to calculate bias over-correction parameter a_g $(g = 1, \ldots, G)$ in section 5·2

REFERENCES

KUHN, H. W. & TUCKER, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, J. Neyman, ed. Berkeley: University of California Press, pp. 481–492.

ROBERTSON, T. & WALTMAN, P. (1968). On estimating monotone parameters. *The Annals of Mathematical Statistics* **39**, 1030–1039.

STUTE, W. & WANG, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics* **21**, 1591–1607.