A LIKELIHOOD RATIO TEST FOR MONOTONE BASELINE HAZARD FUNCTIONS IN THE COX MODEL

Gabriela F. Nane

Delft University of Technology

Supplementary material

S1 PROOF LEMMA 1

Proof. The proof of (i) has been provided by Lemma 1 in Lopuhaä and Nane (2013). The NPMLE $\hat{\lambda}_n(x;\beta)$ is obtained by maximizing the (pseudo) loglikelihood function in (3.1) over all $0 \leqslant \lambda_0(T_{(1)}) \leqslant \ldots \leqslant \lambda_0(T_{(n)})$. As argued in Lopuhaä and Nane (2013), the estimator has to be a nondecreasing step function, that is zero for $x < T_{(1)}$, constant on the interval $[T_{(i)}, T_{(i+1)})$, for $i = 1, \ldots, n-1$, and can be chosen arbitrarily large for $x \geqslant T_{(n)}$. Then, for fixed $\beta \in \mathbb{R}^p$, the (pseudo) loglikelihood function in (3.1) reduces to

$$\begin{aligned} &\sum_{i=1}^{n-1} \Delta_{(i)} \log \lambda_0(T_{(i)}) - \sum_{i=2}^{n} \mathrm{e}^{\beta' Z_{(i)}} \sum_{j=1}^{i-1} \left[T_{(j+1)} - T_{(j)}\right] \lambda_0(T_{(j)}) \\ &\quad = \sum_{i=1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \end{aligned} \tag{S1.1}$$

Let $\lambda_i = \lambda_0(T_{(i)})$, for $i = 1, \ldots, n-1$, and $\lambda = (\lambda_1, \ldots, \lambda_{n-1})$. Then, finding the NPMLE reduces to maximizing

$$\phi(\lambda) = \sum_{i=1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_i - \lambda_i \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}, \tag{S1.2}$$

over the set $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{n-1}$. The NPMLE corresponds thus to a vector $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_{n-1})$ that maximizes ϕ over $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{n-1}$. To prove (ii), we first derive the Fenchel conditions of the estimator. Thus, we will show that the estimator $\hat{\lambda}_n(x;\beta)$ maximizes the (pseudo) loglikelihood function in (3.1) over the class of nondecreasing baseline hazard functions if and only if

$$\sum_{j\geqslant i}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\leqslant 0, \tag{S1.3}$$

for $i=1,2,\ldots,n-1$, and

$$\sum_{j=1}^{n-1}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat{\lambda}_j=0. \tag{S1.4}$$

The NPMLE $\hat{\lambda}_n(x;\beta)$ is thus uniquely determined by these Fenchel conditions. The rest of the proof focuses on deriving the Fenchel conditions (S1.3) and (S1.4) and on establishing (3.6).

First, note that the function ϕ in (S1.2) is concave and that the vector of partial derivatives $\nabla\phi(\lambda)=(\nabla_1\phi(\lambda),\ldots,\nabla_{n-1}\phi(\lambda))$ is given by

$$\nabla\phi(\lambda)=\left(\frac{\Delta_{(1)}}{\lambda_1}-\left[T_{(2)}-T_{(1)}\right]\sum_{l=2}^{n}\mathrm{e}^{\beta' Z_{(l)}},\ldots,\frac{\Delta_{(n-1)}}{\lambda_{n-1}}-\left[T_{(n)}-T_{(n-1)}\right]\mathrm{e}^{\beta' Z_{(n)}}\right).$$

Define now the functions $g_i(\lambda)=\lambda_{i-1}-\lambda_i$, for $i=1,\ldots,n-1$ and $\lambda_0=0$, and the vector $g(\lambda)=(g_1(\lambda),\ldots,g_{n-1}(\lambda))$. Moreover, define the matrix of partial derivatives by

$$G=\left(\frac{\partial g_i(\lambda)}{\partial\lambda_j}\right),\quad \text{for } i=1,\ldots,n-1;\ j=1,\ldots,n-1. \tag{S1.5}$$

Let $\tilde{\phi}(\lambda)=-\phi(\lambda)$. Then, maximizing (S1.2) over all $0\leqslant\lambda_1\leqslant\ldots\leqslant\lambda_{n-1}$ is equivalent with minimizing $\tilde{\phi}(\lambda)$ under the restriction that all components of the vector $g(\lambda)$ are negative. An adaptation of the Karush-Kuhn-Tucker theorem (e.g., see Theorem 8.1 in Groeneboom (1998)) states that $\hat{\lambda}$ minimizes $\tilde{\phi}$ over all vectors λ such that $g_i(\lambda)\leqslant 0$, for all $i=1,\ldots,n-1$, if and only if the following conditions hold

$$\nabla\tilde{\phi}(\hat{\lambda})+G^T\alpha=0, \tag{S1.6}$$

$$g(\hat{\lambda})+w=0, \tag{S1.7}$$

$$\langle\alpha,w\rangle=0, \tag{S1.8}$$

for $\alpha = (\alpha_1, \ldots, \alpha_{n-1})$, with $\alpha_i \geqslant 0$, $i = 1, \ldots, n-1$, and $w = (w_1, \ldots, w_{n-1})$, with $w_i \geqslant 0$, for $i = 1, \ldots, n-1$. The first condition (S1.6), yields that

$$\alpha_i = -\sum_{j \geqslant i} \nabla_j \phi(\hat{\lambda}) = -\sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \qquad \text{(S1.9)}$$

Since $\alpha_i \geqslant 0$, for all $i = 1, \ldots, n-1$, condition (S1.3) is immediate. From (S1.7), $w = -g(\hat{\lambda}) = (\hat{\lambda}_1 - \hat{\lambda}_0, \hat{\lambda}_2 - \hat{\lambda}_1, \ldots, \hat{\lambda}_{n-1} - \hat{\lambda}_{n-2})$, with $\hat{\lambda}_0 = 0$. Note that the condition $w_i \geqslant 0$ implies that $\hat{\lambda}_{i-1} \leqslant \hat{\lambda}_i$, for all $i = 1, \ldots, n-1$, which is trivially satisfied. Finally, by (S1.8),

$$\sum_{i=1}^{n-1} (\hat{\lambda}_i - \hat{\lambda}_{i-1}) \sum_{j \geqslant i} \nabla_j \phi(\hat{\lambda}) = 0,$$

which re-writes exactly to (S1.4).

To derive the expression in (3.6), we prove first that (S1.3) and (S1.4) imply that

$$\sum_{j=1}^{n-1} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} = 0. \qquad \text{(S1.10)}$$

Condition (S1.3) gives that $\sum_{j=1}^{n-1} \nabla_j \phi(\hat{\lambda}) \leqslant 0$. In addition, as the maximizer $\hat{\lambda}$ is nondecreasing,

$$\begin{aligned} \hat{\lambda}_1 \sum_{j=1}^{n-1} \nabla_j \phi(\hat{\lambda}) = &- \nabla_2 \phi(\hat{\lambda}) \hat{\lambda}_2 - \nabla_3 \phi(\hat{\lambda}) \hat{\lambda}_3 - \ldots - \nabla_{n-1} \phi(\hat{\lambda}) \hat{\lambda}_{n-1} \\ &+ \nabla_2 \phi(\hat{\lambda}) \hat{\lambda}_1 + \nabla_3 \phi(\hat{\lambda}) \hat{\lambda}_1 + \ldots + \nabla_{n-1} \phi(\hat{\lambda}) \hat{\lambda}_1 \\ = &\sum_{i=2}^{n-1} (\hat{\lambda}_{i-1} - \hat{\lambda}_i) \sum_{j \geqslant i} \nabla_j \phi(\hat{\lambda}) \geqslant 0. \end{aligned}$$

This shows (S1.10). Now let $B_1, \ldots, B_k$ be blocks of indices on which $\hat{\lambda}$ is constant such that $B_1 \cup \ldots \cup B_k = \{1, \ldots, n-1\}$ and let $v_{nj}(\beta)$ be the value of $\hat{\lambda}$ on the block B_j, with $j = 1, \ldots, k$. If $k = 1$, then the expression of v_{n1} is immediate from (S1.10). Moreover, observe that, by (S1.8), $\sum_{i=1}^{n-1} \alpha_i \left(\hat{\lambda}_i - \hat{\lambda}_{i-1} \right) = 0$, and since $\alpha_i \geqslant 0$ and $\hat{\lambda}_i \geqslant \hat{\lambda}_{i-1}$, for any $i = 1, \ldots, n-1$, it will follow that $\alpha_i = 0$, whenever $\hat{\lambda}_{i-1} < \hat{\lambda}_i$. Hence, for $k \geqslant 2$, there exist $k-1$ α's that are zero.

Then (3.6) follows by (S1.9) and (S1.10). For example, for $k \geqslant 3$, choose any two consecutive α_i that are zero. From (S1.9), we get that by subtracting these α_i's,

$$\sum_{i \in B_j} \nabla_i \phi(\hat{\lambda}) = \sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} = 0.$$

As $v_{nj}(\beta)$ is constant on B_j, this yields (3.6). □

S2 PROOF LEMMA 2

Proof. We will derive the Karush-Kuhn-Tucker (KKT) conditions, that uniquely determine the constrained NPMLE, and which implicitly provide the characterization in (ii). To prove the lemma, we will show that the estimator proposed in (i) satisfies these conditions.

The constrained NPMLE estimator is obtained by maximizing the objective function (3.1) over $0 \leqslant \lambda_0(T_{(1)}) \leqslant \ldots \leqslant \lambda_0(T_{(m)}) \leqslant \theta_0 \leqslant \lambda_0(T_{(m+1)}) \leqslant \ldots \leqslant \lambda_0(T_{(n-1)})$. In line with the reasoning for the unconstrained estimator, it can be argued that the constrained estimator has to be a nondecreasing step function that is zero for $x < T_{(1)}$, constant on $[T_{(i)}, T_{(i+1)})$, for $i = 1, \ldots, n-1$, is equal to θ_0 on the interval $[x_0, T_{(m+1)})$, and can be chosen arbitrarily large for $x \geqslant T_{(n)}$. Therefore, for a fixed $\beta \in \mathbb{R}$, the (pseudo) loglikelihood function in (3.1) reduces to

$$\begin{aligned} &\sum_{i=1}^{m-1} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \\ &\quad + \Delta_{(m)} \log \lambda_0(T_{(m)}) \\ &\quad - \left\{ \lambda_0(T_{(m)}) \left[x_0 - T_{(m)}\right] - \theta_0 \left[T_{(m+1)} - x_0\right] \right\} \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \\ &\quad + \sum_{i=m+1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \end{aligned} \tag{S2.1}$$

By letting $\lambda_i = \lambda_0(T_{(i)})$, for $i = 1, \ldots, n-1$, and $\lambda = (\lambda_1, \ldots, \lambda_{n-1})$, we then want to maximize

$$\phi^0(\lambda) = \sum_{i=1}^{m-1} \left\{ \Delta_{(i)} \log \lambda_i - \lambda_i \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}$$
$$+ \Delta_{(m)} \log \lambda_m - \lambda_m \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \tag{S2.2}$$
$$+ \sum_{i=m+1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_i - \lambda_i \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\},$$

over the set $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_m \leqslant \theta_0 \leqslant \lambda_{m+1} \leqslant \ldots \leqslant \lambda_{n-1}$. Let the vector $\hat{\lambda}^c = (\hat{\lambda}^c_1, \ldots, \hat{\lambda}^c_{n-1})$ denote the constrained NPMLE under the null hypothesis $H_0 : \lambda_0(x_0) = \theta_0$. We will show next that $\hat{\lambda}^c$ maximizes the objective function in (S2.2) over the class of nondecreasing baseline hazard functions, under the null hypothesis, if and only if the following conditions are satisfied

$$\sum_{j \leqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}^c_j} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0, \qquad \text{for } i = 1, \ldots, m-1, \tag{S2.3}$$

$$\sum_{j=1}^{m-1} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}^c_j} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}$$
$$+ \frac{\Delta_{(m)}}{\hat{\lambda}^c_m} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \geqslant 0, \tag{S2.4}$$

$$\sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}^c_j} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \leqslant 0, \text{ for } i = m+1, \ldots, n-1, \tag{S2.5}$$

and

$$\sum_{\substack{j=1 \\ j \neq m}}^{n-1} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}^c_j} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \left(\hat{\lambda}^c_j - \theta_0 \right)$$
$$+ \left\{ \frac{\Delta_{(m)}}{\hat{\lambda}^c_m} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \left(\hat{\lambda}^c_m - \theta_0 \right) = 0. \tag{S2.6}$$

The NPMLE $\hat{\lambda}^c$ is thus uniquely determined by these conditions. To prove (i), we will show that $\hat{\lambda}^0_n$ defined in (3.7) verifies the Karush-Kuhn-Tucker (KKT) conditions (S2.3)-(S2.6). Therefore, $\hat{\lambda}_n^0$ is the unique maximizer of $\phi^0(\lambda)$ in (S2.2), over the set $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_m \leqslant \theta_0 \leqslant \lambda_{m+1} \leqslant \ldots \leqslant \lambda_{n-1}$. As it will be seen further, despite bothersome calculations, the distinct form of the likelihood grants a unified framework for deriving the KKT conditions, that uses all the follow-up times, unlike the reasoning in Banerjee and Wellner (2001), where the (pseudo) loglikelihood is split and arguments are carried both to the left and to the right of x_0.

Similar to the unconstrained case, observe that the function ϕ^0 is concave and that the vector of partial derivatives is $\nabla\phi^0(\lambda) = (\nabla_1\phi^0(\lambda), \ldots, \nabla_{n-1}\phi^0(\lambda))$, with

$$\nabla_i \phi^0(\lambda) = \frac{\Delta_{(i)}}{\lambda_i} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}},$$

for $i = 1, \ldots, m-1, m+1, \ldots, n-1$, and

$$\nabla_m \phi^0(\lambda) = \frac{\Delta_{(m)}}{\lambda_m} - \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}}.$$

Moreover, define the vector $g(\lambda) = (g_1(\lambda), \ldots, g_{n-1}(\lambda))$, with

$$g_i(\lambda) = \begin{cases} \lambda_i - \lambda_{i+1} & i = 1, \ldots, m-1, \\ \lambda_m - \theta_0 & i = m, \\ \theta_0 - \lambda_{m+1} & i = m+1, \\ \lambda_{i-1} - \lambda_i & i = m+2, \ldots, n-1, \end{cases}$$

and consider the matrix of partial derivatives defined in (S1.5). Computations as in (S1.9) can be derived to show that condition (S1.6) yields (S2.3)-(S2.5), upon noting that

$$\alpha_i = \begin{cases} \sum_{j \leqslant i} \nabla_j \phi^0(\hat{\lambda}^c) & i = 1, \ldots, m, \\ -\sum_{j \geqslant i} \nabla_j \phi^0(\hat{\lambda}^c) & i = m+1, \ldots, n-1. \end{cases} \tag{S2.7}$$

Condition (S1.7) gives that $w = (\hat{\lambda}_2^c - \hat{\lambda}_1^c, \ldots, \theta_0 - \hat{\lambda}_m^c, \hat{\lambda}_{m+1}^c - \theta_0, \ldots, \hat{\lambda}_{n-1}^c - \hat{\lambda}_{n-2}^c)$, which together with (S1.8) and (S2.7), yields (S2.6). Moreover, (S1.8) gives that

$$\sum_{i=1}^{m-1} \alpha_i \left(\hat{\lambda}_{i+1}^c - \hat{\lambda}_i^c\right) + \alpha_m \left(\theta_0 - \hat{\lambda}_m^c\right) + \alpha_{m+1} \left(\hat{\lambda}_{m+1}^c - \theta_0\right) + \sum_{m+2}^{n-1} \alpha_i \left(\hat{\lambda}_i^c - \hat{\lambda}_{i-1}^c\right) = 0.$$

Obviously, $\alpha_i = 0$ if $\hat{\lambda}_i^c < \hat{\lambda}_{i+1}^c$, for $i = 1, \dots, m-1, m+1, \dots, n-1$, and (3.8) can be derived as in the proof of Lemma 1. For the block B_p^0 containing m, we get that

$$\sum_{i \in B_p^0 \setminus \{m\}} \left\{ \frac{\Delta_{(i)}}{v_{np}^0(\beta)} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} + \frac{\Delta_{(m)}}{v_{np}^0(\beta)} - \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} = 0,$$

which gives exactly (3.9). Therefore showing that the estimator $\hat{\lambda}_n^0$ defined in (3.7) satisfies the KKT conditions (S2.3)-(S2.6) also proves (ii).

Recall that $\hat{\lambda}_n^0$ is $\min(\hat{\lambda}_i^L, \theta_0)$, for $i = 1, \dots, m$, and that $\hat{\lambda}_i^L$ is the unconstrained estimator when considering only the follow-up times $T_{(1)}, \dots, T_{(m)}$. Moreover, $\hat{\lambda}_n^0$ is $\max(\hat{\lambda}_i^R, \theta_0)$, for $i = m+1, \dots, n-1$, where $\hat{\lambda}_i^R$ is the unconstrained estimator when considering only the follow-up times $T_{(m)}, \dots, T_{(n-1)}$. Note that (S1.10) together with (S1.3) imply that

$$\sum_{j \leqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0, \qquad \text{for } i = 1, \dots, n-1. \tag{S2.8}$$

The condition holds for $i = 1, \dots, m-1$, and, moreover,

$$\sum_{j \leqslant i} \left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \theta_0)} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant \sum_{j \leqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j^L} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0,$$

for $i = 1, \dots, m-1$. Therefore, $\min(\hat{\lambda}_i^L, \theta)$, for $i = 1, \dots, m-1$ satisfies (S2.3). Furthermore, (S2.8) holds for $i = m$, which implies that

$$\begin{aligned} \sum_{j=1}^{m-1} &\left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \theta_0)} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \\ &+ \left\{ \frac{\Delta_{(m)}}{\min(\hat{\lambda}_m^L, \theta_0)} - \left[x_0 - T_{(m)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \\ &\geqslant \sum_{j=1}^{m} \left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \theta_0)} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0, \end{aligned}$$

hence $\hat{\lambda}_n^0$ satisfies (S2.4) as well. It is straightforward that $\max(\hat{\lambda}_i^R, \theta_0)$, for $i = m+1, \dots, n-1$ satisfies (S2.5), since, by definition, $\hat{\lambda}_i^R$ satisfies (S1.3), for $i = m+1, \dots, n-1$, and

$$\sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\max(\hat{\lambda}_j^R, \theta_0)} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \leqslant \sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j^R} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \leqslant 0.$$

Finally, to check if $\hat{\lambda}_n^0$ verifies the condition (S2.6), we will argue on the blocks of indices on which $\hat{\lambda}_n$, and hence $\hat{\lambda}_i^L$ and $\hat{\lambda}_i^R$ are constant. By (3.6), for each block B_j, with $j = 1, \dots, k$, on which the unconstrained estimator has the constant value $v_{nj}(\beta)$,

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} v_{nj}(\beta) = 0,$$

and

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} = 0.$$

Then, on each block B_j that does not contain m, we can write

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \hat{\lambda}_i = \theta_0 \sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}, \tag{S2.9}$$

and this holds for $\hat{\lambda}_i^L$, as well as for $\hat{\lambda}_i^R$. It is straightforward that $\min(\hat{\lambda}_i^L, \theta_0)$, for $i = 1, \dots, m$, and $\max(\hat{\lambda}_i^R, \theta_0)$, for $i = m+1, \dots, n-1$ satisfy this relationship. For the block B_p that contains m, we have

$$\sum_{i\in B_p\backslash\{m\}}\left\{\frac{\Delta_{(i)}}{\hat\lambda_i^L}-\left[T_{(i+1)}-T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat\lambda_i^L$$
$$+\left\{\frac{\Delta_{(m)}}{\hat\lambda_m^L}-\left[T_{(m+1)}-x_0\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat\lambda_m^L$$
$$=\theta_0\sum_{i\in B_p\backslash\{m\}}\left\{\frac{\Delta_{(i)}}{\hat\lambda_i^L}-\left[T_{(i+1)}-T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}$$
$$+\theta_0\left\{\frac{\Delta_{(m)}}{\hat\lambda_m}-\left[T_{(m+1)}-x_0\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}.$$

Constraining $\hat\lambda_m^L$ to be θ_0 on the interval $[x_0, T_{(m+1)})$ yields

$$\begin{aligned}\sum_{i\in B_p\backslash\{m\}}&\left\{\frac{\Delta_{(i)}}{\hat\lambda_i^L}-\left[T_{(i+1)}-T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat\lambda_i^L\\&+\left\{\frac{\Delta_{(m)}}{\hat\lambda_m^L}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat\lambda_m^L\\&=\theta_0\sum_{i\in B_p\backslash\{m\}}\left\{\frac{\Delta_{(i)}}{\hat\lambda_i^L}-\left[T_{(i+1)}-T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\\&+\theta_0\left\{\frac{\Delta_{(m)}}{\hat\lambda_m^L}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}.\end{aligned}\tag{S2.10}$$

Once more, for $i\in B_p$, $\min(\hat\lambda_i^L,\theta_0)$ satisfies this relationship. Summing over all blocks in (S2.9) and (S2.10) completes the proof. □

S3 PROOF LEMMA 5

Proof. Note that the processes X_n and Y_n are monotone. By making use of Corollary 2 in Huang and Zhang (1994) and the remark above the corollary, it suffices to prove that the finite dimensional marginals of the process (X_n, Y_n) converge to the finite dimensional marginals of the process $(g_{a,b}, g_{a,b}^0)$, in order to prove the lemma.

For $x \geqslant T_{(1)}$, let

$$\widehat{W}_n(x)=W_n(\hat\beta_n,x)-W_n(\hat\beta_n,T_{(1)}),$$

where W_n is defined in (3.3), and where $\hat{\beta}_n$ is the maximum partial likelihood estimator. For fixed x_0 and $x \in [-k, k]$, with $0 < k < \infty$, define the process

$$\mathbb{Z}_n(x) = \frac{n^{2/3}}{\Phi(\beta_0, x_0)} \Big\{ V_n(x_0 + n^{-1/3}x) - V_n(x_0) - \lambda_0(x_0) \left[\widehat{W}_n(x_0 + n^{-1/3}x) - \widehat{W}_n(x_0) \right] \Big\}, \tag{S3.1}$$

where V_n is defined in (3.4). For a and b defined in (4.3), $\mathbb{Z}_n$ converges weakly to $X_{a,b}$, as processes in $B_{loc}(\mathbb{R})$, by Lemma 8 in Lopuhaä and Nane (2013). Define now

$$S_n(x) = \frac{n^{1/3}}{\Phi(\beta_0, x_0)} \left\{ \widehat{W}_n(x_0 + n^{-1/3}x) - \widehat{W}_n(x_0) \right\}. \tag{S3.2}$$

From the proof of Lemma 9 in Lopuhaä and Nane (2013), $S_n(x)$ converges almost surely to the deterministic function x, uniformly on every compact set.

Following the approach in Groeneboom (1985), Lopuhaä and Nane (2013) obtained the asymptotic distribution of the unconstrained maximum likelihood estimator $\hat{\lambda}_n$ by considering the inverse process

$$U_n(z) = \underset{x \in [T_{(1)}, T_{(n)}]}{\operatorname{argmin}} \left\{ V_n(x) - z\widehat{W}_n(x) \right\}, \tag{S3.3}$$

for $z > 0$, where the argmin function represents the supremum of times at which the minimum is attained. Since the argmin is invariant under addition of and multiplication with positive constants, it follows that

$$n^{1/3} \left[U_n(\theta_0 + n^{-1/3}z) - x_0 \right] = \underset{x \in I_n(x_0)}{\operatorname{argmin}} \left\{ \mathbb{Z}_n(x) - S_n(x)z \right\},$$

where $I_n(x_0) = [-n^{1/3}(x_0 - T_{(1)}), n^{1/3}(T_{(n)} - x_0)]$. For $z > 0$, the switching relationship $\hat{\lambda}_n(x) \leqslant z$ holds if and only if $U_n(z) \geqslant x$, with probability one. This translates, in the context of this lemma, to

$$n^{1/3} \left[\hat{\lambda}_n(x_0 + n^{-1/3}x) - \theta_0 \right] \leqslant z \Leftrightarrow n^{1/3} \left[U_n(\theta_0 + n^{-1/3}z) - x_0 \right] \geqslant x,$$

for $0 < x_0 < \tau_H$ and $\theta_0 > 0$, with probability one. The switching relationship is thus $X_n(x) \leqslant z \Leftrightarrow n^{1/3} \left[U_n(\theta_0 + n^{-1/3}z) - x_0 \right] \geqslant x$. Hence finding the limiting distribution of $X_n(x)$ resumes to finding the limiting distribution of $n^{1/3} \left[U_n(\theta_0 + n^{-1/3}z) - x_0 \right]$. By applying Theorem 2.7 in Kim and Pollard (1990), it follows that, for every $z > 0$,

$$n^{1/3}\left[U_n(\theta_0+n^{-1/3}z)-x_0\right]\xrightarrow{d}U(z),$$

as inferred in the proof of Theorem 2 in Lopuhaä and Nane (2013), where $U(z) = \sup\{t\in\mathbb{R} : X_{a,b}(t) - zt \text{ is minimal}\}$. It will result that, for every $x \in [-k, k]$,

$$\begin{aligned}P\left(X_n(x)\leqslant z\right) &= P\left(n^{1/3}\left[\hat\lambda_n(x_0+n^{-1/3}x)-\theta_0\right]\leqslant z\right)\\ &= P\left(n^{1/3}\left[U_n(\theta_0+n^{-1/3}z)-x_0\right]\geqslant x\right)\\ &\to P\left(U(z)\geqslant x\right).\end{aligned}$$

Using the switching relationship on the limiting process, it can be deduced that $U(z) \geqslant x \Leftrightarrow g_{a,b}(x) \leqslant z$, with probability one, and thus $X_n(x) \xrightarrow{d} g_{a,b}(x)$.

In order to prove the same type of result for $Y_n(x)$, consider first the following process

$$\widetilde{Y}_n(x) = n^{1/3}\left(\tilde\lambda_n(x_0+n^{-1/3}x)-\theta_0\right), \tag{S3.4}$$

where, for $x_0 \in (0, \tau_H)$, such that $T_{(m)} < x_0 < T_{(m+1)}$,

$$\tilde\lambda_n(x)=\begin{cases}0 & x<T_{(1)},\\ \hat\lambda_i^L & T_{(i)}\leqslant x<T_{(i+1)}, \text{ for } i=1,\dots,m-1\\ \hat\lambda_m^L & T_{(m)}\leqslant x<x_0,\\ 0 & x_0\leqslant x<T_{(m+1)},\\ \hat\lambda_i^R & T_{(i)}\leqslant x<T_{(i+1)}, \text{ for } i=m+1,\dots,n-1\\ \infty & x\geqslant T_{(n)},\end{cases}$$

with $\hat\lambda_i^L$ and $\hat\lambda_i^R$ defined in Lemma 2. For this, we have considered up to x_0 an unconstrained estimator which is constructed based on the sample points $T_{(1)},\dots,T_{(m+1)}$. Moreover, to the right of x_0, we have considered an unconstrained estimator based on the points $T_{(m+1)},\dots,T_{(n)}$. It is not difficult to see that

$$Y_n(x)=\begin{cases}\min\left(\widetilde{Y}_n(x),0\right) & x<0,\\ 0 & x=0,\\ \max\left(\widetilde{Y}_n(x),0\right) & x>0.\end{cases} \tag{S3.5}$$

For $z > 0$, define the inverse processes

$$U_n^L(z) = \underset{x\in[T_{(1)},T_{(m+1)}]}{\text{argmin}} \left\{V_n(x) - z\widehat{W}_n(x)\right\},$$
$$U_n^R(z) = \underset{x\in[T_{(m+1)},T_{(n)}]}{\text{argmin}} \left\{V_n(x) - z\widehat{W}_n(x)\right\}$$

Take $x < x_0$. The switching relationship for $\tilde{\lambda}_n$ is given by $\tilde{\lambda}_n(x) \leqslant z$ if and only if $U_n^L(z) \geqslant x$, with probability one, which gives that

$$n^{1/3}\left[\tilde{\lambda}_n(x_0 + n^{-1/3}x) - \theta_0\right] \leqslant z \Leftrightarrow n^{1/3}\left[U_n^L(\theta_0 + n^{-1/3}z) - x_0\right] \geqslant x,$$

with probability one. Moreover,

$$n^{1/3}\left[U_n^L(\theta_0 + n^{-1/3}z) - x_0\right] = \underset{x\in I_n^L(x_0)}{\text{argmin}} \left\{\mathbb{Z}_n(x) - S_n(x)z\right\},$$

where $I_n^L(x_0) = [-n^{1/3}(x_0 - T_{(1)}), n^{1/3}(T_{(m+1)} - x_0)]$. Denote by

$$Z_n(z,x) = \mathbb{Z}_n(x) - S_n(x)z.$$

As for the unconstrained estimator, we aim to apply Theorem 2.7 in Kim and Pollard (1990). As Theorem 2.7 in Kim and Pollard (1990) applies to the argmax of processes on the whole real line, we extend the above process in the following manner

$$Z_n^-(z,x) = \begin{cases} Z_n(z, -n^{1/3}(x_0 - T_{(1)})) & x < -n^{1/3}(x_0 - T_{(1)}), \\ Z_n(z,x) & -n^{1/3}(x_0 - T_{(1)}) \leqslant x \leqslant n^{1/3}(T_{(m+1)} - x_0), \\ Z_n(z, n^{1/3}(T_{(m+1)} - x_0)) + 1 & x > n^{1/3}(T_{(m+1)} - x_0). \end{cases}$$

Then, $Z_n^-(z,x) \in B_{loc}(\mathbb{R})$ and

$$n^{1/3}\left[U_n^L(\theta_0 + n^{-1/3}z) - x_0\right] = \underset{x\in\mathbb{R}}{\text{argmin}} \left\{Z_n^-(z,x)\right\} = \underset{x\in\mathbb{R}}{\text{argmax}} \left\{-Z_n^-(z,x)\right\}.$$

Since $\lambda_0(x_0) = \theta_0 > 0$ and λ_0 is continuously differentiable in a neighborhood of x_0, it follows by a Taylor expansion and by Lemma 2.5 in Devroye (1981) that $n^{1/3}(T_{(m+1)} - x_0) = \mathcal{O}_p(n^{-1}\log n)$. Therefore, by virtue of Lemma 8 and Lemma 9 in Lopuhaä and Nane (2013), the process $x \mapsto -Z_n^-(z,x)$ converges weakly to $Z^-(x) \in \mathbb{C}_{max}(\mathbb{R})$, for any fixed z, where

$$Z_n^-(z,x) = \begin{cases} -X_{a,b}(x) + zx & x \leqslant 0, \\ 1 & x > 0, \end{cases}$$

for a and b defined in (4.3). Hence, the first condition of Theorem 2.7 in Kim and Pollard (1990) is verified. The second condition follows directly from Lemma 11 in Lopuhaä and Nane (2013), while the third condition is trivially fulfilled. Thus, for any z fixed,

$$n^{1/3}\left[U_n^L(\theta_0+n^{-1/3}z)-x_0\right]\xrightarrow{d}U^-(z),$$

where $U^-(z)=\sup\{t\leqslant 0: X_{a,b}(t)-zt \text{ is minimal}\}$. Concluding, for $x<0$,

$$\begin{aligned}P\left(\widetilde{Y}_n(x)\leqslant z\right)&=P\left(n^{1/3}\left[\tilde{\lambda}_n(x_0+n^{-1/3}x)-\theta_0\right]\leqslant z\right)\\&=P\left(n^{1/3}\left[U_n^L(\theta_0+n^{-1/3}z)-x_0\right]\geqslant x\right)\\&\to P\left(U^-(z)\geqslant x\right).\end{aligned}$$

The switching relationship for the limiting process gives that $U^-(z)\geqslant x\Leftrightarrow D_L(X_{a,b})(x)\leqslant z$, with probability one, where $D_L(X_{a,b})(x)$ has been defined as the left-hand slope of the GCM of $X_{a,b}$, at a point $x<0$. Hence, for $x<0$, $\widetilde{Y}_n(x)\xrightarrow{d}D_L(X_{a,b})(x)$.

Completely analogous, $\widetilde{Y}_n(x)\xrightarrow{d}D_R(X_{a,b})(x)$, for $x>0$. By continuous mapping theorem and by (S4.2), it can be concluded that for fixed $x\in[-k,k]$,

$$Y_n(x)\xrightarrow{d}g_{a,b}^0(x),$$

where $g_{a,b}^0$ has been defined in (3.12).

Our next objective is to apply Theorem 6.1 in Huang and Wellner (1995). The first condition of Theorem 6.1 is trivially fulfilled. The second condition follows by Lemma 11 in Lopuhaä and Nane (2013), while the third condition follows by the definition of the inverse processes. Hence, for fixed x,

$$P\left(X_n(x)\leqslant z, Y_n(x)\leqslant z\right)\to P\left(g_{a,b}(x)\leqslant z, g_{a,b}^0(x)\leqslant z\right),$$

for a and b defined in (4.3). The arguments for one dimensional marginal convergence can be extended to the finite dimensional convergence, as in the proof of Theorem 3.6.2 in Banerjee (2000), by making use of Lemma 3.6.10 in Banerjee (2000). Hence, we can conclude that the finite dimensional marginals of the process (X_n, Y_n) converge to the finite dimensional marginals of the process $(g_{a,b}, g_{a,b}^0)$. This completes the proof. □

S4 Nonincreasing baseline hazard

The characterization of the unconstrained and the constrained NPMLE estimators of a nonincreasing baseline hazard function follows analogously to the characterization of the nondecreasing estimators. The unconstrained NPMLE $\hat{\lambda}_n(x;\beta)$ is obtained by maximizing the (pseudo) likelihood function in (3.1) over all $\lambda_0(T_{(1)}) \geqslant \ldots \geqslant \lambda(T_{(n)}) \geqslant 0$. Lopuhaä and Nane (2013) showed that the likelihood is maximized by a nonincreasing step function that is constant on $(T_{(i-1)}, T_{(i)}]$, for $i = 1, \ldots, n$ and where $T_{(0)} = 0$. The (pseudo) loglikelihood in (3.1) becomes then

$$\sum_{i=1}^{n} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \tag{S4.1}$$

The lemmas below provide the characterization of the unconstrained estimator $\hat{\lambda}_n(x;\beta)$ and the constrained estimator $\hat{\lambda}_n^0(x;\beta)$. Their proofs follow by arguments similar to those in the proofs of Lemma 1 and Lemma 2, as well as the necessary and sufficient conditions that uniquely characterize these estimators..

LEMMA 1. *Let $T_{(1)} < \ldots < T_{(n)}$ be the ordered follow-up times and consider a fixed $\beta \in \mathbb{R}^p$.*

(i) Let W_n be defined in (3.3) *and let*

$$\bar{V}_n(x) = \int \delta\{u \leqslant x\} \, \mathrm{d}P_n(u, \delta, z). \tag{S4.2}$$

Then, the NPMLE $\hat{\lambda}_n(x;\beta)$ of a nonincreasing baseline hazard function λ_0 is given by

$$\hat{\lambda}_n(x;\beta) = \begin{cases} \hat{\lambda}_i & T_{(i-1)} < x \leqslant T_{(i)}, \text{ for } i = 1, \ldots, n, \\ 0 & x > T_{(n)}, \end{cases}$$

for $i = 1, \ldots, n$, with $T_{(0)} = 0$ and where $\hat{\lambda}_i$ is the left derivative of the least concave majorant (LCM) at the point P_i of the cumulative sum diagram (CSD) consisting of the points

$$P_j = \Big(W_n(\beta, T_{(j)}), \bar{V}_n(T_{(j)}) \Big), \tag{S4.3}$$

for $j = 1, \ldots, n$ and $P_0 = (0, 0)$.

(ii) *Let $B_1,\dots,B_k$ be blocks of indices such that $\hat\lambda_n(x;\beta)$ is constant on each block and $B_1\cup\dots\cup B_k=\{1,\dots,n\}$. Denote by $v_{nj}(\beta)$, the value of the estimator on block B_j. Then*

$$v_{nj}(\beta)=\frac{\sum_{i\in B_j}\Delta_{(i)}}{\sum_{i\in B_j}\left[T_{(i)}-T_{(i-1)}\right]\sum_{l=i}^{n}\mathrm{e}^{\beta' Z_{(l)}}}.$$

In fact, for $x\geqslant T_{(n)}$, $\hat\lambda_n(x;\beta)$ can take any value smaller than $\hat\lambda_n$, the left derivative of the LCM at the point P_n of the CSD. As before, we propose $\hat\lambda_n(x)=\hat\lambda_n(x;\hat\beta_n)$ as the estimator of λ_0 and $\hat v_{nj}=v_{nj}(\hat\beta_n)$, where $\hat\beta_n$ denotes the maximum partial likelihood estimator of β_0. Fenchel conditions as in (S1.3) and (S1.4) can be derived analogously.

The NPMLE estimator $\hat\lambda_n^0$ maximizes the (pseudo) loglikelihood function in (S4.1) over the set $\lambda_0(T_{(1)})\geqslant\dots\geqslant\lambda_0(T_{(m)})\geqslant\theta_0\geqslant\lambda_0(T_{(m+1)})\geqslant\dots\geqslant\lambda_0(T_{(n)})\geqslant 0$. It can be argued that the constrained estimator has to be a nonincreasing step function that is constant on $(T_{(i-1)},T_{(i)}]$, for $i=1,\dots,n$, is θ_0 on the interval $(T_{(m)},x_0]$, and is zero for $x\geqslant T_{(n)}$. Hence, the (pseudo) loglikelihood function becomes

$$\begin{aligned}&\sum_{i=1}^{m}\left\{\Delta_{(i)}\log\lambda_0(T_{(i)})-\lambda_0(T_{(i)})\left[T_{(i)}-T_{(i-1)}\right]\sum_{l=i}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\\&\quad+\Delta_{(m+1)}\log\lambda_0(T_{(m+1)})\\&\quad-\left\{\theta_0\left[x_0-T_{(m)}\right]-\lambda_0(T_{(m+1)})\left[T_{(m+1)}-x_0\right]\right\}\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\\&\quad+\sum_{i=m+2}^{n}\left\{\Delta_{(i)}\log\lambda_0(T_{(i)})-\lambda_0(T_{(i)})\left[T_{(i)}-T_{(i-1)}\right]\sum_{l=i}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}.\end{aligned}$$

The characterization of the constrained NPMLE $\hat\lambda_n^0$ is provided with the next lemma.

LEMMA 2. *Let $x_0\in(0,\tau_H)$ fixed, such that $T_{(m)}<x_0<T_{(m+1)}$, for a given $1\leqslant m\leqslant n-1$. Consider a fixed $\beta\in\mathbb{R}^p$.*

(i) *For $i=1,\dots,m$, let $\hat\lambda_i^L$ to be the left derivative of the LCM at the point P_i^L of the CSD consisting of the points $P_j^L=P_j$, for $j=1,\dots,m$, with P_j defined in* (S4.3)*, and $P_0^L=(0,0)$. Moreover, for $i=m+1,\dots,n$, let $\hat\lambda_i^R$*

be the left derivative of the LCM at the point P_i^R of the CSD consisting of the points $P_j^R = P_j$, for $j = m, \ldots, n$, with P_j defined in (S4.3)*. Then, the NPMLE $\hat{\lambda}_n^0(x;\beta)$ of a nonincreasing baseline hazard function λ_0, under the null hypothesis $H_0 : \lambda_0 = \theta_0$, is given by*

$$\hat{\lambda}_n^0(x;\beta) = \begin{cases} \hat{\lambda}_i^0 & T_{(i-1)} < x \leqslant T_{(i)}, \text{ for } i \in \{1,\ldots n\}\backslash\{m+1\}, \\ \theta_0 & T_{(m)} < x \leqslant x_0, \\ \hat{\lambda}_{m+1}^0 & x_0 < x \leqslant T_{(m+1)}, \\ 0 & x > T_{(n)}, \end{cases} \tag{S4.4}$$

where $T_{(0)} = 0$ and where $\hat{\lambda}_i^0 = \max(\hat{\lambda}_i^L, \theta_0)$, for $i = 1, \ldots, m$, and $\hat{\lambda}_i^0 = \min(\hat{\lambda}_i^R, \theta_0)$, for $i = m+1, \ldots, n$.

(ii) For $k \geqslant 1$, let $B_1^0, \ldots, B_k^0$ be blocks of indices such that $\hat{\lambda}_n^0(x;\beta)$ is constant on each block and $B_1^0 \cup \ldots \cup B_k^0 = \{1, \ldots, n\}$. There is one block, say B_r^0, on which $\hat{\lambda}_n^0(x;\beta)$ is θ_0, and one block, say B_p^0, that contains $m+1$. On all other blocks B_j^0, denote by $v_{nj}^0(\beta)$ the value of $\hat{\lambda}_n^0(x;\beta)$ on block B_j^0. Then,

$$v_{nj}^0(\beta) = \frac{\sum_{i \in B_j^0} \Delta_{(i)}}{\sum_{i \in B_j^0} \left[T_{(j)} - T_{(j-1)}\right] \sum_{l=j}^n e^{\beta' Z_{(l)}}}.$$

On the block B_p^0, that contains $m+1$,

$$v_{np}^0(\beta) = \frac{\sum_{i \in B_p^0} \Delta_{(i)}}{\sum_{i \in B_p^0 \backslash \{m+1\}} \left[T_{(i)} - T_{(i-1)}\right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}} + \left[T_{(m+1)} - x_0\right] \sum_{l=m+1}^n e^{\beta' Z_{(l)}}}.$$

We propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x;\hat{\beta}_n)$ as the constrained estimator of a nonincreasing baseline hazard function λ_0, as well as $\hat{v}_{nj}^0 = v_{nj}^0(\hat{\beta}_n)$ on blocks of indices where the estimator is constant. The Fenchel conditions corresponding to (S2.3)-(S2.6) can be derived in the same manner as for the constrained estimator in the nondecreasing case.

Let slolcm(f, I) be the left-hand slope of the LCM of the restriction of the real-valued function f to the interval I. Denote by slolcm$(f) =$ slolcm$(f, \mathbb{R})$. For

$a, b > 0$, let $\bar{X}_{a,b}(t) = a\mathbb{W}(t) - bt^2$, where $\mathbb{W}$ is a standard two-sided Brownian motion originating from zero. Denote by $L_{a,b}$ the LCM of $\bar{X}_{a,b}$ and let

$$l_{a,b}(t) = \text{slolcm}(\bar{X}_{a,b})(t), \tag{S4.5}$$

be the left-hand slope of $L_{a,b}$, at point t. Additionally, set

$$\text{slolcm}^0(f) = \max\left(\text{slolcm}(f, (-\infty, 0]), 0\right) 1_{(-\infty,0]} + \min\left(\text{slolcm}(f, (0, \infty)), 0\right) 1_{(0,\infty)}.$$

For $t \leqslant 0$, construct the LCM of $\bar{X}_{a,b}$, that will be denoted by $L^L_{a,b}$ and take its left-hand slope at point t, denoted by $D_L(\bar{X}_{a,b})(t)$. When the slopes fall behind zero, replace them by zero. In the same manner, for $t > 0$, denote the LCM of $\bar{X}_{a,b}$ by $L^R_{a,b}$ and its slope at point t by $D_R(\bar{X}_{a,b})(t)$. Replace the slopes by zero when they exceed zero. This slope process will be denoted by $l^0_{a,b}$, which is thus given by

$$l^0_{a,b}(t) = \begin{cases} \max\left(D_L(\bar{X}_{a,b})(t), 0\right) & t < 0, \\ 0 & t = 0, \\ \min\left(D_R(\bar{X}_{a,b})(t), 0\right) & t > 0. \end{cases} \tag{S4.6}$$

Observe that $l^0_{a,b}(t) = \text{slolcm}^0(\bar{X}_{a,b})(t)$.

By making use of results in Lopuhaä and Nane (2013), a completely similar result holds in the nonincreasing setting.

LEMMA 3. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is nonincreasing on $[0, \infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) < 0$. Moreover, assume that the functions $x \to \Phi(\beta_0, x)$ and $H^{uc}(x)$, defined in* (4.1) *and above* (4.1)*, are continuously differentiable in a neighborhood of x_0.*
Then, for a and b defined in (4.3)*, (X_n, Y_n) converge jointly to $\left(l_{a,b}, l^0_{a,b}\right)$ in $\mathcal{L} \times \mathcal{L}$, where the processes $l_{a,b}$ and $l^0_{a,b}$ have been defined in* (S4.5) *and* (S4.6)*.*

The asymptotic distribution of the likelihood ratio statistic in the nonincreasing baseline hazard setting can be derived completely analogous.

THEOREM 1. *Suppose (A1) and (A2) hold and let $x_0 \in (0, \tau_H)$. Assume that λ_0 is nonincresing on $[0, \infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) < 0$. Moreover, assume that $H^{uc}(x)$ and*

$x \to \Phi(\beta_0, x)$, defined in (4.1) and above (4.1), are continuously differentiable in a neighborhood of x_0. Let $2\log\xi_n(\theta_0)$ be the likelihood ratio statistic for testing $H_0 : \lambda_0(x_0) = \theta_0$, as defined in (3.2). Then,

$$2\log\xi_n(\theta_0) \xrightarrow{d} \mathbb{D}.$$

Proof. Following the same reasoning as in the proof of Theorem 1 and by Lemma 8, it can be deduced that

$$2\log\xi_n(\theta_0) \xrightarrow{d} \frac{1}{a^2}\int \left[(l_{a,b}(x))^2 - \left(l^0_{a,b}(x)\right)^2\right]\left\{x \in \bar{D}_{a,b}\right\}\,\mathrm{d}x,$$

where $\bar{D}_{a,b}$ is the set on which $l_{a,b}$ and $l^0_{a,b}$ differ. By continuous mapping theorem, it suffices to show that, for t fixed, $l_{a,b}(\bar{X}_{a,b})(t)$ has the same distribution as $g_{a,b}(X_{a,b})(t)$ and $l^0_{a,b}(\bar{X}_{a,b})(t)$ has the same distribution as $g^0_{a,b}(X_{a,b})(t)$. It is noteworthy that

$$\mathrm{slolcm}(\bar{X}_{a,b})(t) = -\mathrm{slogcm}(-\bar{X}_{a,b})(t).$$

Thus, by Brownian motion properties and continuous mapping theorem,

$$\begin{aligned} P\left(l_{a,b}(t) \leqslant z\right) &= P\left(-\mathrm{slogmc}(-a\mathbb{W}(t) + t^2) \leqslant z\right) \\ &= P\left(-\mathrm{slogmc}(a\mathbb{W}(t) + t^2) \leqslant z\right) = P\left(-g_{a,b}(t) \leqslant z\right). \end{aligned}$$

Concluding, $l_{a,b}(\bar{X}_{a,b})(t) \stackrel{d}{=} -g_{a,b}(X_{a,b})(t)$, and a similar reasoning can be applied to show that $l^0_{a,b}(\bar{X}_{a,b})(t) \stackrel{d}{=} -g^0_{a,b}(X_{a,b})(t)$. The proof is then immediate, by continuous mapping theorem. □

References

Banerjee, M. (2000). Likelihood Ratio Inference in Regular and Nonregular Problems. *PhD Dissertation* University of Washington.

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699-1731.

Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings *Ann. Prob.* **9**, 860-867.

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* **2**, 539-555.

Groeneboom, P. (1998). Special Topics Course 593C: Nonparametric Estimation for Inverse Problems: Algorithms and Asymptotics. *Technical Report 344, Department of Statistics, University of Washington.*

Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22**, 3-33.

Huang, Y. and Zhang, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22**, 1256-1274.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.

Lopuhaä, H. P. and Nane, G. F. (2013). Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model. *Scand. J. Statist.* **40**(3), 619-646.