Scandinavian Journal of Statistics (2015), **42**(2): 414–426
DOI: 10.1111/sjos.12113

A Cox-Aalen model for interval-censored data

AUDREY BORUVKA
Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: ajboruvka@uwaterloo.ca

RICHARD J COOK
Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

Summary

The Cox-Aalen model, obtained by replacing the baseline hazard function in the well-known Cox model with a covariate-dependent Aalen model, allows for both fixed and dynamic covariate effects. In this paper, we examine maximum likelihood estimation for a Cox-Aalen model based on interval-censored failure times with fixed covariates. The resulting estimator globally converges to the truth slower than the parametric rate, but its finite-dimensional component is asymptotically efficient. Numerical studies show that estimation via a constrained Newton method performs well in terms of both finite sample properties and processing time for moderate-to-large samples with few covariates. We conclude with an application of the proposed methods to assess risk factors for disease progression in psoriatic arthritis.

Keywords: maximum likelihood estimation; profile likelihood; quadratic programming; semiparametric model; survival data

1 Introduction

Most approaches to regression modelling of time-to-event data account for the possibility where some events are right-censored, but other forms of censoring are routinely encountered in practice. Current status data are obtained when each subject is assessed for the occurrence of an event at one random inspection time. Periodic assessment at a fixed number k of times results in case k interval-censored data. Often, the number of inspections is randomly-distributed. Mixed case interval censoring arises when both the number and timing of inspections are random (Sun, 2006, Section 1.3).

The Cox (1972) proportional hazards model has been adapted for use with interval-censored data in various ways. Many impose additional structure through a discrete (Finkelstein, 1986) or smooth (e.g. Cai and Betensky, 2003) cumulative baseline hazard function. A semiparametric maximum likelihood approach may be preferable because it avoids the need to further specify the Cox model. Early work in this area dealt with case 1 and 2 interval-censoring (e.g. Huang, 1996, Huang and Wellner, 1997). More recently, Kim (2003) considered partially case 2 interval-censored data with the remaining observations subject only to right censoring, Zeng et al. (2006) constructed a semiparametric model with fixed additive effects under case 2 interval censoring, and Wen (2012) devised a proportional hazards model accounting for both mixed case interval censoring and covariate error. Much of this development was made possible through Murphy and van der Vaart's (2000) profile likelihood theory. The semiparametric framework has since seen further extensions. Zhang et al. (2010), for example, devised a spline-based sieve approach, and Wellner and Zhang (2007) derived an M-theorem under model misspecification. These contributions were applied to estimation for the Cox model under case 2 interval censoring and the proportional mean model from panel count data, respectively.

This paper considers semiparametric maximum likelihood estimation of a Cox-Aalen model in which the event time T arises from a cumulative hazard function of the form

$$W'\Lambda(t)\exp(Z'\theta), \tag{1}$$

where $W = (1, W_2, \ldots, W_{d_w})'$ and $Z = (Z_1, \ldots, Z_{d_z})'$ are fixed covariates, θ is a regression coefficient quantifying the multiplicative effect of Z and $\Lambda = (\Lambda_1, \ldots, \Lambda_{d_w})'$ is a vector of cumulative regression functions tracking the additive effect of W. The inner product $W'\Lambda$ is a baseline cumulative hazard function with respect to Z, but Λ is otherwise unspecified. With the first component of W fixed at 1, the remaining entries of W are typically rescaled so that Λ_1 can be interpreted as a reference level of risk and $\Lambda_2, \ldots, \Lambda_{d_w}$ account for time-varying departures. When the $W_2, \ldots, W_{d_w}$ represent levels in a set of factors, (1) reduces to the stratified Cox model (Kalbfleisch and Prentice, 2002, Section 4.4). To our knowledge no methods for estimating either of these Cox model variants from interval-censored data have previously been developed.

The general Cox-Aalen model permitting time-dependent covariates and recurrent events was developed by Scheike and Zhang (2002) as an extension of the Cox (1972) and Aalen (1980) regression models. Its approximate maximum likelihood estimator from independently right-censored data converges weakly to a mean-zero Gaussian process at the parametric rate $\sqrt{n}$. Martinussen and Scheike (2006, Section 7.1.2) describe hypothesis tests about the functional form, but these are based on event times subject only to right-censoring and thus cannot be applied to interval-censored data. In practice one might decide, presumably from knowledge about the underlying process, that the proportional hazards assumption is unrealistic for certain covariates. Under this setting our estimator offers a novel approach to account for both departures from the Cox model and mixed case interval censoring.

2 Notation and basic assumptions

Herein, we consider interval-censored data arising from K random inspections occurring at the random times $Y_K = (Y_{K,1}, \ldots, Y_{K,K})$ on the observation period $[0, \tau]$, $\tau < \infty$. These assessments give $\Delta_K = (\Delta_{K,1}, \ldots, \Delta_{K,K})$, where $\Delta_{K,j} = 1_{(Y_{K,j-1}, Y_{K,j}]}(T)$ $(j = 1, \ldots, K+1)$, $Y_{K,0} \equiv 0$ and $Y_{K,K+1} \equiv \infty$. Consider the following basic assumptions.

C1. Let Θ be a compact subset of R^{d_z} and H the set of cumulative regression functions $\{\Lambda\}$ with $\Lambda(0) \equiv 0$, $\Lambda_1(\infty) \equiv \infty$, and $0 < W'\Lambda(\sigma-) < W'\Lambda(\tau) < M$, almost surely, for some fixed $0 < \sigma < \tau$ and $0 < M < \infty$. The true parameter (θ_0, Λ_0) belongs to $\Theta \times H$ with θ_0 an interior point of Θ.

C2. The conditional distribution of (Δ_K, Y_K, K) given $(T = t, W, Z)$ is invariant with respect to all t compatible with (Δ_K, Y_K, K); that is, all $t \in (Y_{K,j-1}, Y_{K,j}]$ such that $\Delta_{K,j} = 1$ $(j = 1, \ldots, K+1)$. This distribution is specified by some parameter distinct from (θ, Λ).

C3. The product measure of the marginals for $W_2, \ldots, W_p$ is absolutely continuous with respect to the distribution function of W, F_W.

C4. The probability $\mathrm{pr}(T > \tau \mid W)$ is almost surely bounded away from zero.

C5. There are known w_0, w_1 in the support of F_W, $\mathrm{supp}(F_W)$, such that $w_0 \leq W \leq w_1$, almost surely.

Remark 1. Condition C2 implies that T is coarsened at random (Gill et al., 1997, p. 274). This can be motivated with the requirement that (Y_K, K) be conditionally independent of T given (W, Z).

Remark 2. A cumulative regression function Λ ensures $W'\Lambda$ is almost surely nondecreasing. Conditions C3 and C4 imply Λ is valid if $w'\Lambda$ is nondecreasing for every $w \in \mathrm{supp}(F_W)$. From C5 this holds provided that $\mathbf{w}\Lambda$ has nondecreasing entries, where $\mathbf{w}$ is a matrix whose rows run through combinations of values from $(0, w_0, w_1)$.

Our goal is to estimate (θ_0, Λ_0) from n independent observations $X_i = (\Delta^i_{K_i}, Y^i_{K_i}, K_i, W_i, Z_i)$ $(i = 1, \ldots, n)$ assuming conditions C1–C5.

3 Maximum likelihood estimation

Under C2, the density of $X = x = (\delta_k, y_k, k, w, z)$ is

$$p_{\theta,\Lambda}(x) = \prod_{j=1}^{k} \big[\exp\{-w'\Lambda(y_{k,j-1})e^{z'\theta}\} - \exp\{-w'\Lambda(y_{k,j})e^{z'\theta}\}\big]^{\delta_{k,j}}, \tag{2}$$

with respect to a dominating measure ν determined by the distribution of (K, Y_K, W, Z). The corresponding log-likelihood is

$$\begin{aligned}\ell_n(\theta,\Lambda) &= \sum_{i=1}^{n} \log p_{\theta,\Lambda}(X_i) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{K_i} \log\big[\exp\{-W_i'\Lambda(Y^i_{K_i,j-1})e^{Z_i'\theta}\} - \exp\{-W_i'\Lambda(Y^i_{K_i,j})e^{Z_i'\theta}\}\big]^{\Delta^i_{K_i,j}}.\end{aligned}$$

However only some inspections are relevant to the likelihood function.

Definition 1. Let $Y_{(1)}, \ldots, Y_{(m)}$ be the order statistics of the set of all $Y^i_{K_i,j}$ for which $\Delta^i_{K_i,j} + \Delta^i_{K_i,j+1} = 1$ $(j = 1, \ldots, K_i,\ i = 1, \ldots, n)$. Also let $(W_{(l)}, \Delta_{(l)})$ denote some $(W_i, \Delta^i_{K_i,j})$ corresponding to $Y^i_{K_i,j} = Y_{(l)}$ $(l = 1, \ldots, m)$.

Suppose that the smallest relevant inspection time $Y_{(1)}$ corresponds to a right-censored observation; that is, $\Delta_{(1)} = 0$. Then any Λ maximizing the likelihood should satisfy $\Lambda(Y_{(1)}) = 0$. Now assume that the largest relevant time $Y_{(m)}$ is left-censored; that is, $\Delta_{(m)} = 1$. Then $W'_{(m)}\Lambda(Y_{(m)}) = \infty$ almost surely; or, in other words, $\Lambda_1(Y_{(m)}) = \infty$. These two cases make no contribution to the likelihood, so without loss of generality assume that $\Delta_{(1)} = 1$ and $\Delta_{(m)} = 0$. This implies that the bounds imposed on H in C1 incur no loss in generality. The maximum likelihood estimator $(\hat\theta_n, \hat\Lambda_n)$ is characterized by

$$\ell_n(\hat\theta_n, \hat\Lambda_n) = \max_{\theta\in\Theta,\Lambda\in H} \ell_n(\theta,\Lambda).$$

With $\Delta_{(1)} = 1$, the log-likelihood is concave in (θ, Λ). So the semiparametric maximum likelihood estimator (SPMLE) concentrates the survivor distribution on a subset of $Y_{(1)}, \ldots, Y_{(m)}$. This subset is unknown a priori, but a maximal subset can be found by adapting Turnbull (1976, Lemmas 1 and 2).

Definition 2. From the censoring intervals, $(L_i, R_i] = (Y^i_{K_i,j-1}, Y^i_{K_i,j}]$ having $\Delta^i_{K_i,j} = 1$ ($i = 1, \ldots, n$, $j = 1, \ldots, K_i + 1$), define the maximal intersections (Wong and Yu, 1999)

$$\mathcal{I} = \{(s_1, t_1], \ldots, (s_d, t_d]\},$$

whose left and right endpoints are selected respectively from $\{L_1, \ldots, L_n\}$ and $\{R_1, \ldots, R_n\}$ such that $(s_j, t_j] \cap (L_i, R_i]$ is either $(s_j, t_j]$ or $\emptyset$, for every $j = 1, \ldots, d$ and $i = 1, \ldots, n$.

Proposition 1. *$W'\hat{\Lambda}_n$ is almost surely constant outside $\mathcal{I}$. Moreover for fixed $\hat{\Lambda}_n$ on the boundary of $\mathcal{I}$, the likelihood is invariant to the behaviour of $\hat{\Lambda}_n$ on the interior of $\mathcal{I}$.*

This result follows from a straightforward adaptation of the proof for Alioum and Commenges (1996, Lemmas 1 and 2). Without loss of generality we take the SPMLE $(\hat{\theta}_n, \hat{\Lambda}_n)$ as the discrete maximizer of $\ell_n(\theta, \Lambda)$ concentrating mass on the right endpoints $\{t_1, \ldots, t_d\}$ of $\mathcal{I}$.

4 Computation

For $\theta \in \Theta$ and $\Lambda \in H$ discrete on $t_1, \ldots, t_d$, define $\lambda_j = \Lambda(t_j)$ ($j = 1, \ldots, d$), $\lambda = (\lambda_1', \ldots, \lambda_d')'$, $\phi = (\theta', \lambda')'$ and $\ell_n(\phi) \equiv \ell_n(\theta, \Lambda)$. Under conditions C3–C5 the requirement that $W'\Lambda$ be nondecreasing is met by $0 \le \mathbf{w}\Lambda(t_j) \le \mathbf{w}\Lambda(t_k)$, $j < k$, with $\mathbf{w}$ as defined in Remark 2. This can in turn be written in the form $A\lambda \ge 0$, where A is the block diagonal matrix

$$A = \begin{bmatrix} \mathbf{w} & 0 & 0 & 0 & \cdots & 0 \\ -\mathbf{w} & \mathbf{w} & 0 & 0 & \cdots & 0 \\ & & \cdots & & & \\ 0 & 0 & \cdots & 0 & -\mathbf{w} & \mathbf{w} \end{bmatrix}.$$

One could compute ϕ by alternating between iterative algorithms for θ and λ, but such a strategy can be slow to converge. A reduction in computing time might be obtained by jointly updating the estimates in a single iteration, as demonstrated in Pan's (1999) extension of the iterative convex minorant algorithm (Jongbloed, 1998) to the Cox model. A more recent example can be found in Cheng et al. (2011), where the iterative convex minorant is recast into quadratic programming. We propose a similar approach under a Lagrangian framework general enough to accommodate the Cox-Aalen model. The algorithm is summarized as follows.

Step 1 (Initial value). Set $r = 0$, $\theta^{(0)} = 0$ and $\lambda_j^{(0)} = (t_j, 0'_{d_w-1})'$.

Step 2 (Candidate step). Evaluate $\eta^{(r)} = (\eta_\theta^{(r)}, \eta_\lambda^{(r)})$, where

$$\eta_\theta^{(r)} = -\nabla_\theta^2 \ell_n(\phi^{(r)})^{-1} \nabla_\theta \ell_n(\phi^{(r)}),$$

is the Newton-Raphson offset, and

$$\begin{aligned} \eta_\lambda^{(r)} &= \underset{\eta_\lambda : A(\lambda^{(r)} + \eta_\lambda) \ge 0}{\arg\max} \nabla_\lambda \ell_n(\phi^{(r)})' \eta_\lambda + \frac{1}{2} \eta_\lambda' \nabla_\lambda^2 \ell_n(\phi^{(r)}) \eta_\lambda \\ &= \underset{\lambda : A\lambda \ge 0}{\arg\min} \frac{1}{2} \lambda' \nabla_\lambda^2 \ell_n(\phi^{(r)})' \lambda + \{\nabla_\lambda \ell_n(\phi^{(r)}) - \nabla_\lambda^2 \ell_n(\phi^{(r)})' \lambda^{(r)}\}' \lambda - \lambda^{(r)}. \end{aligned}$$

is the maximizer of a quadratic approximation to the increment in the log-likelihood function.

Step 3 (Line search). To avoid overshooting the maximum, set $\phi^{(r+1)} = \phi^{(r)} + \eta^{(r)}/2^j$, where j is the smallest nonnegative integer satisfying

$$\ell_n(\phi^{(r)}) - \ell_n(\phi^{(r)} + \eta^{(r)}/2^j) \leq \alpha \nabla_\phi \ell_n(\phi^{(r)})' \eta^{(r)}/2^j,$$

for some fixed $0 < \alpha < 1/2$.

Step 4 (Stopping rule). If $\|\phi^{(r+1)} - \phi^{(r)}\|_\infty \leq \varepsilon$ for some small $\varepsilon > 0$, then stop. Otherwise, increment r and return to Step 2.

Condition C1 and properties of the log-likelihood function satisfy Dümbgen et al.'s (2006, Section 3.1 and 3.2) requirements for convergence of $\phi^{(r)}$ to the semiparametric maximum likelihood estimator as $r \to \infty$. Following from Fenchel's duality theorem, the SPMLE can be directly characterized through the stopping rule $|\nabla_\theta \ell_n(\phi^{(r)})' \phi^{(r)}| \leq \varepsilon$ (Groeneboom, 1996, Lemma 2.1; Jongbloed, 1998). Unlike the supremum norm in Step 4 this inner product, when equal to zero, characterizes the SPMLE.

Computation time is largely determined by the size of (d_w, d_z, d), processing power and the software used to carry out quadratic programming in Step 2. We achieved a relatively fast estimation routine by implementing the algorithm in C and drawing from IBM's (2012) ILOG CPLEX Callable Library.

5 Asymptotic properties

Under some additional conditions $(\hat{\theta}_n, \hat{\Lambda}_n)$ is globally $n^{1/3}$-consistent, but $\hat{\theta}_n$ is asymptotically efficient at (θ_0, Λ_0). The limiting distribution of $\hat{\Lambda}_n$ remains an open problem. The current section describes the details of these results. Proofs can be found in this paper's supporting information.

Consistency requires that the semiparametric model is identifiable, which is easily ensured by conditions C5 and C6–C8 below.

C6. The support of Z, $\mathrm{supp}(F_Z)$, is a bounded subset of R^{d_z}.

C7. There is an integer $1 \leq k_0 < \infty$, such that $K < k_0$, almost surely.

C8. For any $a \in R$, $b \in R^{d_w}$ and $c \in R^{d_z}$ such that $b, c \neq 0$, both $\mathrm{pr}(W'b \neq a)$ and $\mathrm{pr}(Z'c \neq a)$ are bounded away from zero.

Identifiability is clearly limited to the support of the inspection times, so we consider convergence in measure. Let $p_k(w, z) = \mathrm{pr}(K = k \mid W = w, Z = z)$. Adapting van der Vaart and Wellner (2000, p. xiv) and Wellner and Zhang (2007, p. 2110), define for any $B \in \mathcal{B}[0, \tau]$ and $C \in \mathcal{B}(R^{d_w + d_z})$

$$\mu(B \times C) = \int_C \sum_{k=1}^{\infty} p_k(w, z) \sum_{j=1}^{k} \mathrm{pr}(Y_{k,j} \in B \mid W = w, Z = z) dF_{W,Z}(w, z),$$

$\mu_y(B) = \mu(B \times R^{d_w + d_z})$.

C9. $\mu_y \times F_W \times F_Z \ll \mu$.

Theorem 1. *Under the aforementioned conditions $\hat{\theta}_n \to \theta_0$, almost surely, and $\hat{\Lambda}_n \to \Lambda_0$, μ_y-almost everywhere.*

The overall rate of convergence for $(\hat{\theta}_n, \hat{\Lambda}_n)$ is derived with one additional assumption.

C10. Put $\mu_z = \mu/\mu\{[0,\tau]\times\mathrm{supp}(F_W)\}$. There exists some $0 < c < 1$ such that $a'\,\mathrm{var}(Z \mid Y, W)a \le ca'E(ZZ' \mid Y, W)a$, μ_z-almost–everywhere, for all $a \in R^{d_z}$.

Remark 3. Following Wellner and Zhang (2007, Remark 3.4), C10 can be justified as follows. From C8 and the Markov inequality, $E(ZZ')$ is positive definite. Assume that $\mathrm{var}_{\mu_z}(Z \mid Y, W)$ and $E(ZZ' \mid Y, W)$ are also positive definite. Condition C10 is then satisfied with c no larger than the ratio of the smallest eigenvalue of $\mathrm{var}_{\mu_z}(Z \mid Y, W)$ to the largest eigenvalue of $E(ZZ' \mid Y, W)$, provided that this ratio is bounded away from zero uniformly in (Y, W).

Theorem 2. *Under the aforementioned conditions,* $\|\hat\theta_n - \theta_0\| + \|\hat\Lambda_n - \Lambda_0\|_{\mu_y,2} = O_P(n^{-1/3})$, *where* $\|\hat\Lambda_n - \Lambda_0\|_{\mu_y,2} = \sum_{j=1}^{d_w}(\int |\hat\Lambda_{n,j} - \Lambda_{0,j}|^2 d\mu_y)^{1/2}$ *is the* $L_2(\mu_y)$ *distance between* $\hat\Lambda_n$ *and* Λ_0.

The limiting distribution of $\hat\theta_n$ is obtained by application of Murphy and van der Vaart's (2000) profile likelihood theory. This considers an asymptotic expansion of the profile log-likelihood function $\ell_n^p(\theta) = \sup_{\Lambda\in H}\ell_n(\theta,\Lambda)$ at $\hat\theta_n$.

C11. The true cumulative regression function Λ_0 has a bounded continuous derivative λ_0 such that $\mathbf{w}\lambda_0 > 0$ on $[\sigma,\tau]$.

C12. There is $y_0 > 0$ such that $\mathrm{pr}(Y_{K,j} - Y_{K,j-1} \ge y_0) = 1$ $(j = 1,\dots,K)$.

C13. The conditional density functions $f_{Y_{k,j}|W,Z}(u \mid w, z)$ and $f_{Y_{k,j},Y_{k,j+1}|W,Z}(u, v \mid w, z)$ exist for every $k = 1, 2, \dots$ and $j = 1,\dots,k-1$. Moreover the conditional expectations $E_{K|W,Z}\{\sum_{j=1}^{K} f_{Y_{K,j}|W,Z}(u \mid w, z)\}$ and $E_{K|W,Z}\{\sum_{j=1}^{K-1} f_{Y_{K,j},Y_{K,j+1}|W,Z}(u, v \mid w, z)\}$ have partial derivatives with respect to u and v that are bounded uniformly in $(w, z) \in \mathrm{supp}(F_{W,Z})$.

Theorem 3. *Under the aforementioned conditions, the sequence* $\sqrt{n}(\hat\theta_n - \theta_0)$ *is asymptotically normal with mean zero and variance equal to the inverse of efficient information matrix* $\tilde I_0$. *Moreover, for any* $v_n \to v \in R^{d_z}$ *and* $h_n \to 0$ *in probability as* $n \to \infty$ *such that* $(\sqrt{n}h_n)^{-1} = O_p(1)$,

$$-2\frac{\ell_n^p(\hat\theta_n + h_n v_n) - \ell_n^p(\hat\theta_n)}{nh_n^2} \to v'\tilde I_0 v, \tag{3}$$

in probability as $n \to \infty$.

Remark 4. Conditions C11–C13 greatly simplify the proof of Theorem 3 but have practical implications. A consequence of C12 is that the event times must be strictly left-censored, interval-censored or right-censored. The availability of some exact times would only improve the rate of convergence. So although Theorem 3 may imply asymptotic efficiency in the case of partially interval-censored data, our proof does not formally address it. Condition C13 precludes consideration of any discretely-distributed inspection process. However methods for grouped data (e.g. Lawless, 2003, Section 7.3) are better suited in this setting.

6 Variance estimation

The limit in (3) gives an approximation to the entries of $\tilde I_0$. The resulting matrix can be inverted to obtain a variance estimator for $\hat\theta_n$. Let $e_1,\dots,e_{d_z}$ denote the unit vectors in R^{d_z}. Expanding $(e_i + e_j)'\tilde I_0(e_i + e_j)$, the (i,j)th entry of $\tilde I_0$ is consistently estimated by

$$\frac{1}{nh_{ii}^2}\{\ell_n^p(\hat\theta_n + h_{ii}e_i) - \ell_n^p(\hat\theta_n)\} + \frac{1}{nh_{jj}^2}\{\ell_n^p(\hat\theta_n + h_{jj}e_j) - \ell_n^p(\hat\theta_n)\} - \frac{1}{nh_{ij}^2}[\ell_n^p\{\hat\theta_n + h_{ij}(e_i + e_j)\} - \ell_n^p(\hat\theta_n)] \quad (i,j = 1,\dots,d_z). \tag{4}$$

Because θ and Λ are variation independent, $\ell_n^p(\hat{\theta}_n) = \ell_n(\hat{\theta}_n, \hat{\Lambda}_n)$. So this variance estimator calls for maximizing $d_z(d_z+1)/2$ profile likelihood functions. Such a task is carried out by fixing $\theta^{(r)}$ in the parameter estimation algorithm of Section 4 and revising the stopping rule to convergence in $\ell_n(\phi^{(r)})$.

The tuning parameters $h_{ij} = h_{ji}$ $(i, j = 1, \ldots, d_z)$ must converge to zero no faster than $\sqrt{n}$. In practice, some fixed value proportional to $1/\sqrt{n}$ may be chosen empirically. Borrowing methods from numerical derivatives, we propose a data-driven approach that reduces the choice to specifying typical and large values for θ.

For a continuously differentiable function $f : R \to R$, consider the numerical derivative based on the first-order finite-difference approximation $f'(x) \approx \{f(x+h) - f(x)\}/h$. Typically, one chooses $h \sim \sqrt{\epsilon}\,\mathrm{curv}(x)$, where ϵ is the error in evaluating f and $\mathrm{curv} = (f/f'')^{1/2}$ is the curvature scale of f. This choice of h minimizes both the truncation error $h^3 f''$ and the round-off error $\epsilon|f/h|$ in the approximation to f'. Often little is known about f'', so $h \sim \sqrt{\epsilon}x$. To handle x close to zero, this choice is revised to $h \sim \sqrt{\epsilon}\,\mathrm{sign}(x)\max(|x|, \mathrm{typ}\,x)$, where $\mathrm{typ}\,x$ is the typical magnitude of x (e.g. Press et al., 2007, Section 5.7). The approximation given by (3) is essentially a second-order finite-difference approximation to the curvature in the profile log-likelihood at $\hat{\theta}_n$. The corresponding curvature scale $\mathrm{curv}_{ij}(\hat{\theta}_n)$ is the cube root of $-2\ell_n^p(\hat{\theta}_n)/(e_i \vee e_j)'\nabla_\theta^3 \ell_n^p(\hat{\theta}_n)$, where $e_i \vee e_j$ is the element-wise maximum of e_i and e_j. Using the size of θ to replace extreme values in the curvature scale, a straightforward extension of this selection strategy gives

$$h_{ij} = n^{-1/2}\,\mathrm{sign}\{\mathrm{curv}_{ij}(\hat{\theta}_n)\} \times \max[\min\{|\,\mathrm{curv}_{ij}(\hat{\theta}_n)|, \sup\theta\}, |\hat{\theta}_{n,i}|, |\hat{\theta}_{n,j}|, \mathrm{typ}\,\theta], \tag{5}$$

for $i, j = 1, \ldots, d_z$.

7 SIMULATION STUDY

We assessed the frequency properties of the semiparametric maximum estimator under an inspection scheme that roughly followed a predetermined schedule. Event times were generated from the Weibull-type cumulative hazard function $(t^{3/2}W_1 + t^{2/3}W_2)\exp(\theta_1 Z_1 + \theta_2 Z_2)$, where $W_1 = 1$, W_2 is uniform on $(0, 1)$, Z_1 is standard normal and Z_2 is uniform on $\{0, 1\}$. Over the observation period $(0, 2)$, a total of k scheduled visits were evenly spaced. The actual inspection times were generated from k independent normal distribution functions having mean equal to one of the schedule times, standard deviation $1/\{2(k+1)\}$ and truncation points at zero, the midpoints between scheduled visit times and 2. This ensured that the support of the inspections times covered $(0, 2)$, with most inspections occurring close to their scheduled target. To reflect skipped visits, each inspection after the first one was missed with probability $p(W, Z) = \mathrm{expit}(\beta_0 + \beta_1 Z_2)$. The values examined for k and (β_0, β_1) were set according to one of three scenarios.

Scenario 1. Independent censoring with $k = 8$, $\beta_0 = \log(1/9)$ and $\beta_1 = 0$.

Scenario 2. Independent censoring with $k = 4$, $\beta_0 = \log(1/9)$ and $\beta_1 = 0$.

Scenario 3. Conditionally independent censoring with $k = 8$, $e^{\beta_0} = 1/4$ and $e^{\beta_1} = 4/9$.

Under $\exp(\beta) = (1/9, 1)'$, the probability of missing the jth $(j = 2, \ldots, k)$ inspection is $1/10$. In Scenario 3 the probability remained the same if $Z_2 = 1$. Those with $Z_2 = 0$ were twice as likely to miss an inspection. The rates of left-, interval-, and right-censoring under Scenarios 1 and 3 were roughly 17, 62 and 16%, respectively. Under Scenario 2 the corresponding rates were 26, 46 and 27%.

For each scenario the SPMLE was fit to 1000 Monte Carlo samples of size $n = 100, 200$ or 500. For the tuning parameters, we set $\alpha = 1/3$ in the line search of Step 3, $\varepsilon = 10^{-7}$ in the stopping

Table 1: Simulation study results for the regression coefficients under Scenario 1

		$n = 100$				$n = 200$				$n = 500$			
		Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP
θ_1	SPMLE	5.64	1.57	1.60	0.96	3.10	1.03	1.04	0.94	1.34	0.58	0.62	0.96
	Mid	0.45	1.39	1.35	0.94	-0.55	0.95	0.92	0.94	-1.18	0.54	0.57	0.95
	End	-0.66	1.38	1.35	0.94	-1.73	0.94	0.92	0.94	-2.31	0.54	0.57	0.94
	Latent	1.79	1.35	1.33	0.94	0.98	0.95	0.91	0.94	0.41	0.54	0.56	0.96
θ_2	SPMLE	-3.48	2.57	2.57	0.95	-3.00	1.75	1.74	0.95	-0.97	1.06	1.07	0.95
	Mid	1.24	2.41	2.41	0.95	0.45	1.67	1.66	0.94	1.22	1.02	1.04	0.95
	End	3.37	2.40	2.43	0.95	2.03	1.68	1.67	0.94	2.38	1.03	1.04	0.95
	Latent	0.07	2.36	2.32	0.95	-0.76	1.63	1.60	0.94	-0.17	1.01	1.00	0.95

Bias, average of estimates minus the truth; SD, standard deviation of estimates; ASE, average of standard error estimates; CP, proportion of 95% confidence intervals that contained the truth; SPMLE, semiparametric maximum likelihood estimator; Mid, End and Latent: semiparametric estimators from midpoint-imputed, right-endpoint –imputed and latent right-censored data.

Table 2: Simulation study results for the regression coefficients under Scenario 2

		$n = 100$				$n = 200$				$n = 500$			
		Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP
θ_1	SPMLE	8.07	1.81	1.86	0.96	4.48	1.14	1.19	0.96	2.31	0.63	0.70	0.96
	Mid	-2.40	1.41	1.37	0.93	-3.47	0.97	0.93	0.92	-3.75	0.54	0.58	0.90
	End	-5.89	1.44	1.37	0.90	-7.34	1.00	0.92	0.83	-7.53	0.57	0.57	0.72
θ_2	SPMLE	-5.60	2.88	2.87	0.96	-4.41	1.94	1.91	0.95	-2.06	1.17	1.17	0.94
	Mid	3.70	2.50	2.51	0.94	2.94	1.72	1.73	0.95	3.50	1.08	1.08	0.94
	End	8.68	2.59	2.56	0.92	7.64	1.76	1.74	0.92	7.51	1.08	1.08	0.89

See notes for Table 1.

rule of Step 4, and $\operatorname{typ}\theta = 1$ and $\sup\theta = 10$ in (5). This ensured convergence within a reasonable number of iterations in all scenarios and sample sizes. For comparison, estimates were also obtained from the corresponding midpoint-imputed, right-endpoint–imputed and the latent right-censored data using Martinussen and Scheike's (2006) timereg package for R (R Core Team, 2013).

The simulation results for $\hat{\theta}_n$ reported in Tables 1 and 2 and Table S1 in the supporting information. The finite sample behaviour is compatible with the asymptotic properties outlined in Section 5; across all scenarios bias becomes negligible with larger sample size, the empirical standard deviation is reasonably approximated by the average of the standard error estimates, and the empirical coverage rates of the 95% confidence intervals are close to the nominal level. Imputing to the midpoint of the censoring interval generally achieved better results than imputing to the right-endpoint. Midpoint imputation also outperformed the maximum likelihood estimator in smaller samples with more frequent inspections. However performance of both imputation-based estimators degraded with increasing sample size and decreasing frequency of inspection.

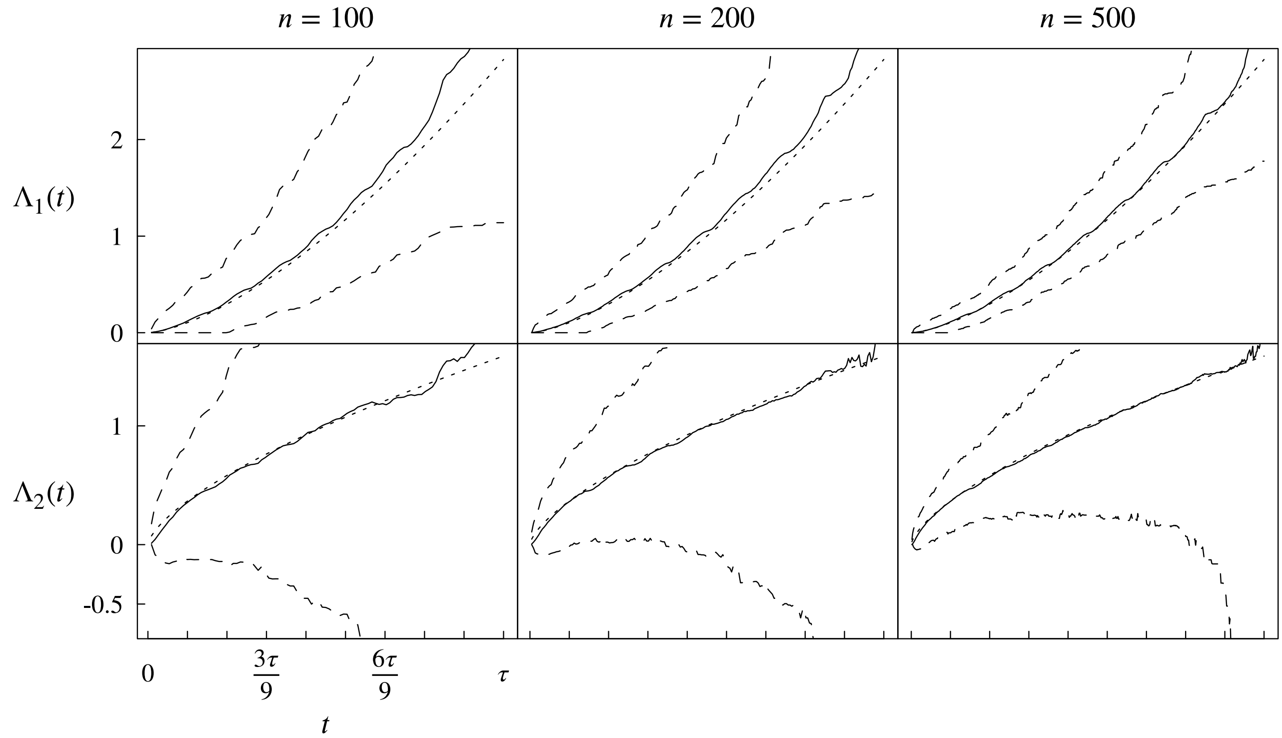

Figure 1: The true parameter value (dotted) displayed with the pointwise means (solid) and 2.5th percentiles (dashed) of $\hat{\Lambda}_n$ under Scenario 1

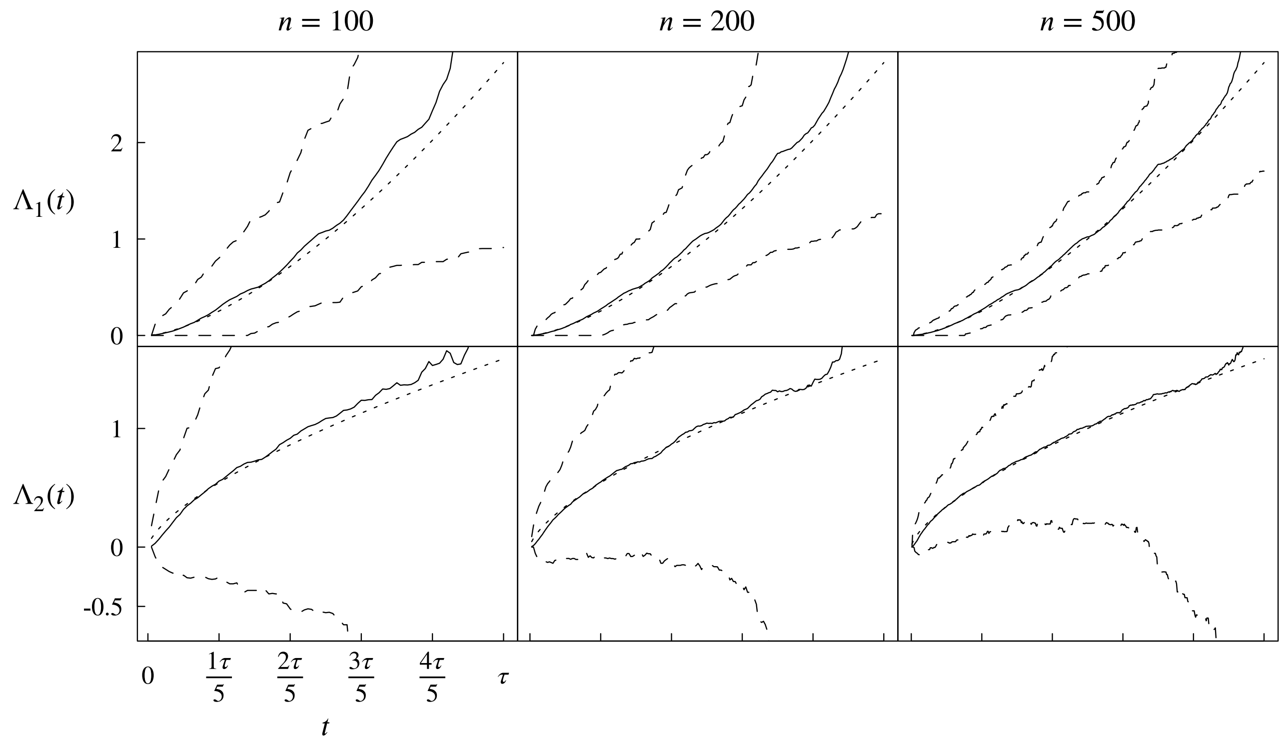

Figure 2: The true parameter value (dotted) displayed with the pointwise means (solid) and 2.5th percentiles (dashed) of $\hat{\Lambda}_n$ under Scenario 2

Pointwise empirical means and 2.5th percentiles for $\hat{\Lambda}_n$ are depicted in Figures 1 and 2 and Figure S1 in the supporting information. Empirical bias and variability in $\hat{\Lambda}_n$ decrease with increasing sample size and number of scheduled inspections. Bias also appeared to decrease closer to the scheduled visit times, where inspections were relatively frequent. Estimates of the cumulative coefficient Λ_2 vary considerably more that those for the baseline regression function Λ_1.

The constrained Newton algorithm described in Section 4 does not scale particularly well in larger samples (Table 3); estimation under $n = 500$ was over fifty times slower than with $n = 100$. This rate of increase is sharper when inspections are more frequent. The number of iterations is fairly stable with n, suggesting that the computational burden lies in quadratic programming. Because the number of constraints is exponential in d_w, the processing time may become unreasonably long under very large samples with numerous covariates. Such scaling issues are typical of methods to compute nonparametric and semiparametric estimators. However, our approach appears to perform relatively well compared to Pan's (1999) extension of the iterative convex minorant to the Cox model (Zhang et al., 2010, Table 1). On average, the maximum norm reached ε faster than the gradient-based norm. Either stopping rule appears adequate, but the maximum norm is preferable in terms of processing time.

Table 3: Monte Carlo sample average of processing time, iterations and gradient-based stopping norm

	$k = 4$			$k = 8$		
	$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
CPU time (minutes)	0.13	0.72	6.60	0.22	1.32	14.02
Number of iterations	210	201	177	356	360	372
$\nabla_\phi \ell_n(\hat{\phi})'\hat{\phi} \times 10^5$ at convergence	0.79	0.88	1.05	6.20	2.75	4.61

CPU time, processing time in minutes for both parameter and variance estimation on an Opteron 6200 processor core rated at 3GHz; Iterations, number of iterations to algorithm convergence.

8 Application

A severe form of joint destruction known as arthritis mutilans is estimated to arise in 2 to 16% of patients with psoriatic arthritis (Gladman et al., 2005). Prognostic studies suggest that genetic factors play a role in progressive joint damage. Using data collected from a Toronto-based psoriatic arthritis clinic, Gladman and Farewell (1995) established an association between certain human leukocyte antigen genes and various stages of progression. Chandran et al. (2011) revisited this problem with an updated sample of 610 patients, including data on killer-cell immunoglobulin-like receptor genes. They identified a number of potential genetic factors for the development in arthritis mutilans, characterized by the presence of at least five severely damaged joints. A total of 49 patients (8%) had arthritis mutilans by their first biannual radiographic survey, yielding left-censored times. An additional 49 patients developed arthritis mutilans over course of follow-up, leaving the majority (512 or 84%) of the sample right-censored.

Chandran et al. (2011) addressed interval-censoring through the use of a parametric Weibull hazard model. This gave hazard ratios of various genetic markers, adjusted for both sex and age at diagnosis of psoriatic arthritis. Juvenile psoriatic arthritis, defined by onset before the age of 16 years, is thought to be different from psoriatic arthritis arising in adulthood in terms both immunogenetics and disease course (Hamilton et al., 1990). We thus fit a Cox-Aalen model to the same data considered by Chandran et al. (2011), with the baseline hazard function stratified by age at diagnosis.

For initial covariate selection, we carried out Sun's (1996) nonparametric log-rank test of differences in survival curves using the R interval package (Fay and Shaw, 2010). Genetic markers having

Table 4: Covariate effects on the risk of arthritis mutilans, adjusted for age at diagnosis of psoriatic arthritis

Factor	$\hat{\theta}_n$	ese($\hat{\theta}_n$)	p-value	95% CI for e^{θ}
Female	-0.08	0.21	0.71	0.61 – 1.40
HLA-A11 present	-0.87	0.44	0.05	0.18 – 1.00
HLA-A29 present	-1.47	0.69	0.03	0.06 – 0.89
HLA-B27 present	0.59	0.23	0.01	1.15 – 2.82
HLA-DQB1*02 present	0.52	0.21	0.01	1.11 – 2.56
KIR3DS1 present, HLA-Bw4*80I absent	0.54	0.22	0.01	1.12 – 2.63

HLA-A, B, C, DQB1: human leukocyte antigen gene classes; KIR, killer-cell immunoglobulin-like receptor; ese, estimated standard error; CI, confidence interval.

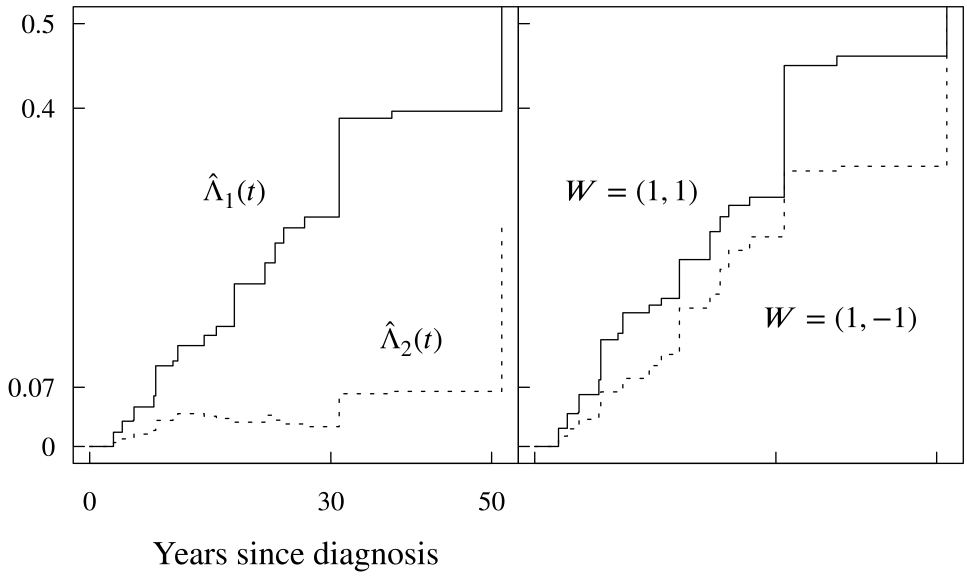

Figure 3: Left panel: Cumulative baseline hazard estimate for individuals diagnosed at the sample average of 36 years (solid) and cumulative coefficient for standardized age at diagnosis (dotted). Right panel: Cumulative baseline hazard estimate for individuals with age at diagnosis one standard deviation older (solid) and younger (dotted) than 36 years

a p-value less than 10% were added to a regression model containing sex and age at diagnosis. The set of markers was then reduced further by backward elimination until all remaining p-values were no greater than 5%. This selection procedure ultimately identified the same set of genetic risk factors (Table 4) as those found by Chandran et al. (2011).

Although the effect of age at diagnosis on the risk of arthritis mutilans is not of primary scientific interest, we depict the estimate for Λ in Figure 3 to fully illustrate the model fit. Patients diagnosed in older age tended to develop arthritis mutilans sooner. This difference appears to vary over time, but further investigation is needed to determine if it is significantly different from zero.

Additional work is needed to assess the requirements of the SPMLE, particularly C2; the assumption that the time to arthritis mutilans is coarsened at random. Although this goes beyond the simple demonstration we aimed for here, any departures from the condition are likely limited as assessments of arthritis mutilans were scheduled at 6-month intervals, staggered according to study entry. Some observations are right-censored at the last visit prior to death, but this form of potentially dependent censoring arises in less than 5% of the sample.

9 Discussion

In this paper, we derived the SPMLE for the Cox-Aalen model with fixed covariates from interval-censored data and showed that it performs well under moderate to large samples and relatively few covariates. Although the estimator relying on midpoint-imputation attained smaller empirical bias in some simulation scenarios, this edge in finite sample performance may not hold under other models for the event time and observation scheme. Moreover, our simulation study showed that estimators based on systematic imputation generally do not achieve smaller bias with increasing sample size—a property needed to reasonably carry out inference on the regression coefficient.

Derivation of the limiting distribution of $\hat{\Lambda}_n$ or subsampling-based pointwise confidence intervals for Λ would permit inference about additive effects and the survivor distribution. Under right-censored data Scheike and Zhang (2003) propose a test statistic to infer whether or not Λ_j, $j > 1$, is time-varying. An analogous test with interval-censored data would be useful in assessing departures from the Cox model. These are some potential areas for further development.

Acknowledgements

This article is based on an excerpt from Audrey Boruvka's Ph.D. thesis. The authors thank Dr. Dafna Gladman for permission to use data from the University of Toronto Psoriatic Arthritis Program. Funding support was provided by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN 155849) and a grant from the Canadian Institutes of Health Research (FRN 13887). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Section S1 Proofs for Theorems 1–3

Table S1 Simulation results for $\hat{\theta}_n$ under Scenario 3

Figure S1 Simulation results for $\hat{\Lambda}_n$ under Scenario 3

Section S3 Notes on obtaining our software implementation of Section 4

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In Witold, K., Kozek, A., and Rosinski, J., editors, *Mathematical Statistics and Probability Theory: Proceedings, Sixth International Conference, Wisa (Poland) 1978*, pages 1–25. Springer, New York.

Alioum, A. and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52(2):512–524.

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59(3):570–579.

Chandran, V., Thavaneswaran, A., Pellett, F. J., and Gladman, D. D. (2011). The association between human leukocyte antigen and killer-cell immunoglobulin-like receptor gene variants and the development of arthritis mutilans in patients with psoriatic arthritis. *Arthritis Rheum.*, 63(S10):S307. American College of Rheumatology Meeting abstract.

Cheng, G., Zhang, Y., and Lu, L. (2011). Efficient algorithms for computing the non and semi-parametric maximum likelihood estimates with panel count data. *J. Nonparametr. Stat.*, 23(2):567–579.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34(2):187–220.

Dümbgen, L., Freitag-Wolf, S., and Jongbloed, G. (2006). Estimating a unimodal distribution from interval-censored data. *J. Amer. Statist. Assoc.*, 101(475):1094–1106.

Fay, M. P. and Shaw, P. A. (2010). Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *J. Stat. Software*, 36(2):1–34.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42(4):845–854.

Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In Lin, D. Y. and Fleming, T. R., editors, *Proceedings of the First Seattle Symposium: Survival Analysis*, pages 255–294, New York. Springer.

Gladman, D. D., Antoni, C., Mease, P., Clegg, D. O., and Nash, P. (2005). Psoriatic arthritis: Epidemiology, clinical features, course, and outcome. *Ann. Rheum. Dis.*, 64(suppl 2):ii14–ii17.

Gladman, D. D. and Farewell, V. T. (1995). The role of HLA antigens as indicators of disease progression in psoriatic arthritis. *Arthritis Rheum.*, 38(6):845–850.

Groeneboom, P. (1996). Lectures on inverse problems. In Bernard, P., editor, *Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXIV – 1994*, pages 67–164. Springer-Verlag, Berlin.

Hamilton, M. L., Gladman, D. D., Shore, A., Laxer, R. M., and Silverman, E. D. (1990). Juvenile psoriatic arthritis and HLA antigens. *Ann. Rheum. Dis.*, 49(9):694–697.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.*, 24(2):540–568.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: A review of recent progress. In Lin, D. Y. and Fleming, T. R., editors, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pages 123–169, New York. Springer.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Stat.*, 7(3):310–321.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd edition.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. Roy. Statist. Soc. Ser. B*, 65(2):489–502.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, New York, 2nd edition.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.

IBM (2012). *ILOG CPLEX Optimization Studio*. Version 12.5.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.*, 95(450):449–465.

Pan, W. (1999). Extending the iterative convex minorant algorithm to the Cox model for interval-censored data. *J. Comput. Graph. Stat.*, 8(1):109–120.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, 3rd edition.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Scheike, T. H. and Zhang, M.-J. (2002). An additive-multiplicative Cox-Aalen regression model. *Scand. J. Stat.*, 29(1):75–88.

Scheike, T. H. and Zhang, M.-J. (2003). Extensions and applications of the cox-aalen survival model. *Biometrics*, 59(4):1036–1045.

Sun, J. (1996). A non-parametric test for interval-censored failure time data with application to AIDS studies. *Stat. Med.*, 15(13):1387–1395.

Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B*, 38(3):290–295.

van der Vaart, A. and Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In Giné, E., Mason, D. M., and Wellner, J. A., editors, *High Dimensional Probability II*, pages 115–133. Birkhäuser, Boston.

Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.*, 35(5):2106–2142.

Wen, C.-C. (2012). Cox regression for mixed case interval-censored data with covariate errors. *Lifetime Data Anal.*, 18(3):321–338.

Wong, G. Y. C. and Yu, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multivariate Anal.*, 69(2):155–166.

Zeng, D., Cai, J., and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statist. Sinica*, 16(1):287–302.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Stat.*, 37(2):338–354.

Published in final edited form as:
Scandinavian Journal of Statistics (2015), **42**(2): 414–426
DOI: 10.1111/sjos.12113

A Cox-Aalen model for interval-censored data: Supporting information

AUDREY BORUVKA
Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: ajboruvka@uwaterloo.ca

RICHARD J COOK
Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

S1 Proofs

Asymptotic properties stated in Section 5 are derived here by application of empirical process theory. This exercise largely amounts to adapting results from Huang (1996), Huang and Wellner (1997), Murphy and van der Vaart (1997), van der Vaart and Wellner (2000) and Wellner and Zhang (2007).

Theorem 1 follows from van der Vaart (1998, Theorem 5.7), an approach routinely used to establish consistency of M-estimators. Instead of $\log p_{\theta,\Lambda}$, we use the technically more convenient criterion $m_{\theta,\Lambda} = \log\{(p_{\theta,\Lambda} + p_0)/2\}$, where the subscript 0 is shorthand for θ_0, Λ_0.

Let $S_{\theta,\Lambda}(y \mid w, z) = \exp\{-w'\Lambda(y)e^{z'\theta}\}$ denote the survivor function with $S_{\theta,\Lambda}(0 \mid w, z) \equiv 1$, $S_{\theta,\Lambda}(\infty \mid w, z) \equiv 0$ and $F_{\theta,\Lambda} = 1 - S_{\theta,\Lambda}$. Then $p_{\theta,\Lambda} = p_0$ almost surely implies

$$0 = \int \sum_{k=1}^{\infty} p_k(w, z) \sum_{j=1}^{k+1} \int |(F_{\theta,\Lambda} - F_0)(y_{k,j} \mid w, z) - (F_{\theta,\Lambda} - F_0)(y_{k,j-1} \mid w, z)\}| dF_{Y_K|K,W,Z}(y_k \mid k, w, z) dF_{W,Z}(w, z).$$

Adapting van der Vaart and Wellner (2000, Lemma 4), C7 ensures that the inner summation has lower bound

$$\max_{1\le j\le k} \int |(F_{\theta,\Lambda} - F_0)(y_{k,j} \mid w, z)| dF_{Y_{K,j}|K}(y_{k,j} \mid k, w, z) \geq \frac{1}{k}\sum_{j=1}^{k} \int |(S_{\theta,\Lambda} - S_0)(y_{k,j} \mid w, z)| dF_{Y_{K,j}|K,W,Z}(y_{k,j} \mid k, w, z).$$

Thus $0 = \int |p_{\theta,\Lambda} - p_0| d\nu \geq \int |S_{\theta,\Lambda} - S_0| d\tilde{\mu}$, where $\tilde{\mu}$ is the measure obtained by scaling p_k by $1/k$ in μ. From C5–C7, μ and $\tilde{\mu}$ are both finite. Since $\mu \ll \tilde{\mu}$ the dominated convergence theorem gives $\int |S_{\theta,\Lambda} - S_0| d\mu = 0$. From C1, $W'\Lambda_0$ is μ-almost everywhere bounded away from zero. Therefore $e^{z'(\theta_0-\theta)} = w'\Lambda(y)/w'\Lambda_0(y)$, μ-a.e., and $Z'(\theta_0-\theta)$ is then degenerate given $Y \sim \mu_y$. Under conditions C8 and C9 this implies that $\theta = \theta_0$ and hence $w'\{\Lambda(y) - \Lambda_0(y)\} = 0$, μ-a.e. Appealing to C8 and

C9 again yields $\Lambda = \Lambda_0$, μ_y-a.e. This establishes identifiability. Since $\log a \le 2(\sqrt{a} - 1)$ for every $a \ge 0$,

$$P(m_{\theta,\Lambda} - m_0) \le 2 \int \left\{ \frac{(p_{\theta,\Lambda} + p_0)p_0}{2} \right\}^{1/2} d\nu - 2 \le - \int (p_{\theta,\Lambda}^{1/2} - p_0^{1/2})^2 d\nu. \tag{S1}$$

The true measure p_0 is identifiable, so this upper bound is zero only if $\theta = \theta_0$ and $\Lambda = \Lambda_0$ μ_y-almost–everywhere. Since the logarithm is concave and $(\hat\theta_n, \hat\Lambda_n)$ is the unique maximizer of the log-likelihood, $\mathbb{P}_n(m_{\hat\theta_n,\hat\Lambda_n} - m_0) = \mathbb{P}_n \log\{(p_{\hat\theta_n,\hat\Lambda_n} + p_0)/(2p_0)\} \ge \mathbb{P}_n\{\log(p_{\hat\theta_n,\hat\Lambda_n}) - \log(p_0)\}/2 \ge 0$. Thus

$$\begin{aligned} P(m_0 - m_{\hat\theta_n,\hat\Lambda_n}) &\le \mathbb{P}_n\, m_{\hat\theta_n,\hat\Lambda_n} - P m_{\hat\theta_n,\hat\Lambda_n} + o(1) \\ &\le \sup_{(\theta,\Lambda)\in\Theta\times H} |\,\mathbb{P}_n\, m_{\theta,\Lambda} - P m_{\theta,\Lambda}| + o(1). \end{aligned} \tag{S2}$$

Any $\Lambda \in H$ can be written as $\Lambda = \Lambda^+ - \Lambda^-$, where both Λ^+ and Λ^- are bounded and monotone on $[0, \tau]$. From C1, C5 and C6, $(\theta, \Lambda) \to S_{\theta,\Lambda}$ is uniformly bounded and Lipschitz for every x. From C7 and van der Vaart and Wellner (1996, Theorem 2.7.5), $\mathcal{M} = \{m_{\theta,\Lambda} : \theta \in \Theta, \Lambda \in H\}$ is P-Glivenko–Cantelli with bracketing number

$$N_{[]}\{\varepsilon, \mathcal{M}, L_2(P)\} \lesssim (\operatorname{diam}\Theta/\varepsilon)^{d_z k_0} \times \exp(2 d_w k_0/\varepsilon). \tag{S3}$$

Thus the upper bound in (S2) or, equivalently, the probability of the event $\{P m_{\hat\theta_n,\hat\Lambda_n} < P m_0\}$ almost-surely tends to zero. From (S1), $\{\hat\theta_n \ne \theta_0\}$ and $\{\hat\Lambda_n \ne \Lambda_0 \text{ on } \operatorname{supp}(\mu_y)\}$ are subsets of this event, so their probabilities must also almost-surely converge to zero.

Proving Theorem 2 amounts to verifying the requirements of van der Vaart and Wellner (1996, Theorem 3.2.5), a well-known method for establishing an M-estimator's rate of convergence.

Since p_0 is bounded away from zero, $p_{\theta,\Lambda}$ is bounded above by one, $\mu \ll \tilde\mu$ with $\mu, \tilde\mu < \infty$ and $|p - q|^2 \le |p - q|$ for every $p, q \in [0, 1]$, the proof of identifiability above gives $\int (p_{\theta,\Lambda}^{1/2} - p_0^{1/2})^2 d\nu \gtrsim \int (p_{\theta,\Lambda} - p_0)^2 d\nu \ge \int |S_{\theta,\Lambda} - S_0| d\mu$. Let $\theta_t = t\theta + (1-t)\theta_0$ and $\Lambda_t = t\Lambda + (1-t)\Lambda_0$. From the mean value theorem there is $t \in (0, 1)$ depending on (y, w, z) such that

$$(S_{\theta,\Lambda} - S_0)(y \mid w, z) = S_{\theta_t,\Lambda_t}(y \mid w, z) e^{z'\theta_t}\{w'(\Lambda - \Lambda_0)(y) + (\theta - \theta_0)' z w' \Lambda_t(y)\}.$$

For $(Y, W, Z) \sim \mu$, define $g_0(Z) = 1 + t(\theta - \theta_0)'Z$, $g_1(Y, W) = W'(\Lambda - \Lambda_0)(Y)$ and $g_2(Y, W, Z) = (\theta - \theta_0)' Z W' \Lambda_0(Y)$. So $(S_{\theta,\Lambda} - S_0)(Y \mid W, Z)$ is equal to $g_0(Z)g_1(Y, W) + g_2(Y, W, Z)$ up to the factor $S_{\theta_t,\Lambda_t}(y \mid W, Z)e^{Z'\theta_t}$, which is bounded away from zero under C1 and C6. Adapting Wellner and Zhang (2007, pp. 2126–2127), the Cauchy-Schwarz inequality and C10 give

$$\begin{aligned} \{E_\mu(g_1 g_2)\}^2 &\le E_\mu(g_1^2) E_\mu\big([W'\Lambda_0(Y) E_\mu\{(\theta - \theta_0)'Z \mid Y, W\}]^2\big) \\ &\le (1 - c) E_\mu(g_1^2) E_\mu(g_2^2). \end{aligned}$$

Since c is bounded away from zero and $g_0(z)$ is uniformly close to one for θ near θ_0, Murphy and van der Vaart (1997, Lemma A.6) gives $\int (S_{\theta,\Lambda} - S_0)^2 d\mu \gtrsim \mu g_2^2 + \mu g_1^2 \gtrsim \|\theta - \theta_0\| + \|\Lambda - \Lambda_0\|_{\mu_y}^2$, where the last inequality up to a constant holds under C1 and C5. Thus $P(m_{\theta,\Lambda} - m_0) \lesssim -\|\theta - \theta_0\|^2 - \|\Lambda - \Lambda_0\|_{\mu_y,2}^2$. Let $\mathcal{M}_\delta = \{m_{\theta,\Lambda} - m_0 : \|\theta - \theta_0\| + \|\Lambda - \Lambda_0\|_{\mu_y,2} < \delta_n\}$. From (S3), this class has bracketing integral $J_{[]}\{\delta, \mathcal{M}_\delta, L_2(P)\} \lesssim \delta^{-1/2}$. By van der Vaart and Wellner (1996, Lemma 3.4.2), $E^*\|\,\mathbb{G}_n\,\|_{\mathcal{M}_\delta} \lesssim \sqrt{\delta}\{1 + \delta^{-1}(\delta n)^{-1/2}\}$.

Theorem 3 is derived by checking Murphy and van der Vaart's (2000, Theorem 1) conditions for semiparametric efficiency. Existence of the efficient score for θ at (θ_0, Λ_0) is shown by adapting the proof of Huang and Wellner (1997, Theorem 4.1). Throughout Z is assumed scalar ($d_z = 1$). The case where $d_z > 1$ follows by applying the same arguments to each entry of θ.

Let $R_{\theta,\Lambda}(u,v \mid w,z) = S_{\theta,\Lambda}(v \mid w,z)/\{S_{\theta,\Lambda}(u \mid w,z) - S_{\theta,\Lambda}(v \mid w,z)\}$ and $Q_{\theta,\Lambda} = 1 - R_{\theta,\Lambda}$. The score for θ is $\dot{\ell}_{\theta,\Lambda} = \sum_{j=1}^{d_w} \dot{\ell}_{\theta,\Lambda}^{j}$, where

$$\begin{aligned}\dot{\ell}_{\theta,\Lambda}^{j}(x) = ze^{z\theta}\Big[&\delta_{k,1} w_j \Lambda_j(y_{k,1}) R_{\theta,\Lambda}(0, y_{k,1} \mid w,z) - \delta_{k,k+1} w_j \Lambda_j(y_{k,k}) \\ &+ \sum_{j=2}^{k} \delta_{k,j}\big\{ w_j \Lambda_j(y_{k,j}) R_{\theta,\Lambda}(y_{k,j-1}, y_{k,j} \mid w,z) \\ &\qquad - w_j \Lambda_j(y_{k,j-1}) Q_{\theta,\Lambda}(y_{k,j-1}, y_{k,j} \mid w,z)\big\}\Big].\end{aligned}$$

Perturbing each entry in Λ generates a tangent set with respect to the product space $\{\Lambda_1 \times \cdots \times \Lambda_{d_w}\}$ of which H is a subset. Consider a one-dimensional submodel $s \mapsto \Lambda_{s,1} \times \cdots \times \Lambda_{s,d_w}$ with direction $h = (h_1, \ldots, h_{d_w})$ satisfying $h_j = \partial/\partial s_{|s=0} \Lambda_{s,j}$ $(j = 1, \ldots, d_w)$. For now assume that h is chosen so that $\Lambda_s \in H$. Then a score function for Λ is $L_{\theta,\Lambda} h = \sum_{j=1}^{d_w} L_{\theta,\Lambda} h_j$, where

$$\begin{aligned}L_{\theta,\Lambda} h_j(x) = e^{z\theta}\Big[&\delta_{k,1} w_j h_j(y_{k,1}) R_{\theta,\Lambda}(0, y_{k,1} \mid w,z) - \delta_{k,k+1} w_j h_j(y_{k,k}) \\ &+ \sum_{l=2}^{k} \delta_{k,l}\big\{ w_j h_j(y_{k,l}) R_{\theta,\Lambda}(y_{k,l-1}, y_{k,l} \mid w,z) \\ &\qquad - w_j h_j(y_{k,l-1}) Q_{\theta,\Lambda}(y_{k,l-1}, y_{k,l} \mid w,z)\big\}\Big].\end{aligned}$$

Considering the event time T as the unobserved variable in an information loss model (e.g. van der Vaart, 1998, Section 25.5.2), the adjoint $L_{\theta,\Lambda}^{*}$ of the score operator $L_{\theta,\Lambda}$ is the conditional expectation under (θ, Λ) given $\{T = t\}$. Moreover if $L_{\theta,\Lambda}^{*} \dot{\ell}_{\theta,\Lambda} = L_{\theta,\Lambda}^{*} L_{\theta,\Lambda} h$ then h is a least favourable direction. By C2,

$$\begin{aligned}L_{\theta,\Lambda}^{*} \dot{\ell}_{\theta,\Lambda}(t) &= \sum_{j=1}^{d_w} E_{W,Z}[E\{\dot{\ell}_{\theta,\Lambda}^{j}(X) \mid T = t, W, Z\}] \\ &= \sum_{j=1}^{d_w} E_{W,Z}[E\{1_{(L,R]}(t) \dot{\ell}_{\theta,\Lambda}^{j}(X) \mid W, Z\}] \\ &= \sum_{j=1}^{d_w} E_{W,Z}\Big[\sum_{k=1}^{\infty} p_k(W,Z) E\{1_{(L,R]}(t) \dot{\ell}_{\theta,\Lambda}^{j}(X) \mid K = k, W, Z\}\Big].\end{aligned}$$

Owing to the similar score structure, $L_{\theta,\Lambda}^{*} L_{\theta,\Lambda} h(t)$ is the right side of this expression with $\dot{\ell}_{\theta,\Lambda}^{j}$ replaced by $L_{\theta,\Lambda} h_j$. For $u < v$ and $j = 1, \ldots, d_w$, put

$$A_j(u,v) = R_{\theta,\Lambda}(u,v \mid W,Z) W_j e^{Z\theta} \sum_{k=1}^{\infty} p_k(W,Z) \sum_{l=1}^{k} f_{Y_{K,l-1}, Y_{K,l} \mid W,Z}(u,v \mid W,Z)$$

and similarly define $B_j(u,v)$ with $Q_{\theta,\Lambda}$ in place of $R_{\theta,\Lambda}$. Let a_j, b_j, c_j and d_j denote the respective expectations of A_j, B_j, ZA_j and ZB_j relative to (W,Z). From C2 and C12,

$$\begin{aligned}q_j(t) \equiv L_{\theta,\Lambda}^{*} \dot{\ell}_{\theta,\Lambda}^{j}(t) = &\int_t^{\tau} \Lambda_j(u) c_j(0,u) du - \int_{\sigma}^{t} \Lambda_j(u) d_j(u,\infty) du \\ &+ \int_{u=\sigma}^{t} \int_{v=t}^{\tau} \{\Lambda_j(v) c_j(u,v) - \Lambda_j(u) d_j(u,v)\} 1(v - u \geq y_0) dv du\end{aligned}$$

and $r_j(t) \equiv L^*_{\theta,\Lambda} L_{\theta,\Lambda} h_j(t)$ has similar form, obtained by replacing Λ_j, c_j and d_j with h_j, a_j and b_j, respectively ($j = 1, \ldots, d_w$). Let $h^j_{\theta,\Lambda}$ denote the h_j for which $q'_j = r'_j$ ($j = 1, \ldots, d_w$). Then $h^j_{\theta,\Lambda}$ is the solution to the Fredholm integral equation

$$h^j_{\theta,\Lambda}(t) = g_j(t) + \int K_j(s,t) h^j_{\theta,\Lambda}(s) ds, \tag{S4}$$

where $g_j(t) = -q_j(t)/s_j(t)$, $K_j(u,t) = \{a_j(t,u)1(u - t \geq y_0) - b_j(u,t)1(t - u \geq y_0)\}/s_j(t)$ and

$$s_j(t) = a_j(0,t) + b_j(t,\infty) + \int_\sigma^t a_j(u,t)1(t-u \geq y_0)du + \int_t^\tau b_j(t,v)1(v - t \geq y_0)dv.$$

At the true parameter (θ_0, Λ_0), $g_j = g_{0,j}$ and $K_j = K_{0,j}$ are bounded by C1, C5, C6 and C12. From Fredholm's first theorem (e.g. Kanwal, 1997, p. 48), (S4) at the truth has the μ_y-almost–everywhere unique solution $h^j_0(t) = g_{0,j}(t) + \int \Gamma_{0,j}(u,t) g_{0,j}(u) du$, where $\Gamma_{0,j}$ is completely determined by $K_{0,j}$ and is identically zero only if $g_{0,j} = 0$. Thus $\tilde{\ell}_0 = \dot{\ell}_0 - L_0 h_0$ is the efficient score for θ at (θ_0, Λ_0) and the efficient information matrix $\tilde{I}_0 = P_0 \tilde{\ell}_0 \tilde{\ell}'_0$ is positive definite. We now identify a submodel that is indexed by h_0 and satisfies the structural requirements of Murphy and van der Vaart (2000, Theorem 1). Extending the arguments of Huang (1996, pp. 563–564) and van der Vaart (1998, p. 411), consider

$$\Lambda_s(\theta, \Lambda) = \Lambda + (\theta - s)\varphi(\Lambda)(h_0 \circ \Lambda_{0,1}^{-1} \circ \Lambda_1), \tag{S5}$$

where $\Lambda_{0,1}$ and Λ_1 are the first components of Λ_0 and Λ, respectively, and φ is a smooth approximation to $1_{(0,M)}(\mathbf{w}y)$ ensuring that $0 < \mathbf{w}\Lambda_s(\theta,\Lambda) < M$ on $[\sigma, \tau]$ and $\partial/\partial s_{|s=0} \Lambda_s(\theta, \Lambda_0) = h_0$. In particular $\varphi(\Lambda) = 1$ on $[\Lambda_0(\sigma), \Lambda_0(\tau)]$, $\Lambda \mapsto \varphi(\Lambda)$ is Lipschitz and, for every $\Lambda \in H$, $0 \leq \mathbf{w}\Lambda\varphi(\Lambda) \lesssim \mathbf{w}\Lambda \wedge (M - \mathbf{w}\Lambda)$ with the last inequality satisfied up to a constant depending only on (θ_0, Λ_0). From C1, φ exists. From C11, $\Lambda_{0,1}$ is strictly increasing and continuous, so its inverse is well-defined. With C13, $h_0 \circ \Lambda_{0,1}^{-1}$ is bounded and Lipschitz. Since the composition $h_0 \circ \Lambda_{0,1}^{-1} \circ \Lambda_1$ has the same jump discontinuities as Λ_1 we have, for every $u \leq v$ and s sufficiently close to θ,

$$\mathbf{w}\{\Lambda_s(\theta,\Lambda)(u) - \Lambda_s(\theta,\Lambda)(v)\} \leq \mathbf{w}\{\Lambda(u) - \Lambda(v)\}(1 - |\theta - s|c_0),$$

where c_0 is the Lipschitz constant of $\Lambda \mapsto \varphi(\Lambda) h_0 \circ \Lambda_{0,1}^{-1}(\Lambda)$. Thus (S5) defines an approximately least favourable submodel such that $\Lambda_\theta(\theta, \Lambda) = \Lambda$ and the map $s \mapsto \log p_{s,\Lambda_s(\theta,\Lambda)}(x) \equiv \ell(s,\theta,\Lambda)(x)$ is twice continuously differentiable with $\dot{\ell}(\theta_0,\theta_0,\Lambda_0) = \tilde{\ell}_0$. For any consistent estimator $\tilde{\theta}_n$, the profile maximizer $\arg\max_{\Lambda \in H} \ell_n(\tilde{\theta}_n, \Lambda)$ tends to Λ_0 in probability due to Theorem **??** and the fact that θ and Λ are variation independent. For fixed x each term on the right-hand side of

$$\begin{aligned} P_0 \dot{\ell}(\theta_0,\theta_0,\Lambda) = P_0 \Big[\frac{p_0 - p_{\theta_0,\Lambda}}{p_0} \{\dot{\ell}(\theta_0,\theta_0,\Lambda) - \dot{\ell}(\theta_0,\theta_0,\Lambda_0)\} \Big] \\ - P_0 \dot{\ell}(\theta_0,\theta_0,\Lambda) \Big\{ \frac{p_{\theta_0,\Lambda} - p_0}{p_0} - L_0(\Lambda - \Lambda_0) \Big\} \end{aligned}$$

depends on Λ only through one or both of $\Lambda(y_{k,j-1})$ and $\Lambda(y_{k,j})$ with $\delta_{k,j} = 1$. Without loss of generality suppose that $1 < j < k$. Following Murphy and van der Vaart (2000, p. 460), ordinary Taylor expansions at $(\Lambda(y_{k,j-1}), \Lambda(y_{k,j}))$ yield the inequalities

$$\begin{aligned} |p_{\theta_0,\Lambda} - p_0|(x) &\lesssim |\Lambda - \Lambda_0|(y_{k,j-1}) + |\Lambda - \Lambda_0|(y_{k,j}), \\ |\dot{\ell}(\theta_0,\theta_0,\Lambda) - \dot{\ell}(\theta_0,\theta_0,\Lambda_0)|(x) &\lesssim |\Lambda - \Lambda_0|(y_{k,j-1}) + |\Lambda - \Lambda_0|(y_{k,j}), \\ |p_{\theta_0,\Lambda} - p_0 - L_0(\Lambda - \Lambda_0)p_0|(x) &\lesssim |\Lambda - \Lambda_0|^2(y_{k,j-1}) + |\Lambda - \Lambda_0|^2(y_{k,j}), \end{aligned}$$

since the first and second derivatives with respect to $\Lambda(y_{k,j-1})$ and $\Lambda(y_{k,j})$ are uniformly bounded under C1, C5 and C6. Thus $P_0 \dot{\ell}(\theta_0,\theta_0,\Lambda) \lesssim \|\Lambda - \Lambda_0\|^2_{\mu_y,2}$. By Theorem 2, the right side of this inequality up to a constant is $O_P(n^{-2/3})$, which is more than enough to establish the no-bias condition (Murphy and van der Vaart, 2000, equation 11). For the same x, $\dot{\ell}(s, \theta, \Lambda)(x)$ and $\ddot{\ell}(s, \theta, \Lambda)(x)$ are Lipschitz in z, $e^{z\theta}$, $w'\Lambda(y_{k,j-1})$ and $w'\Lambda(y_{k,j})$. From C1 and C7, van der Vaart and Wellner (1996, Theorem 2.7.5) and arguments similar to those preceding display (S3), $\Theta \times H$ is P_0-Donsker. Thus $x \to \ddot{\ell}(s, \theta, \Lambda)(x)$ and $x \to \dot{\ell}(s, \theta, \Lambda)(x)$ form P_0-Glivenko–Cantelli and P_0-Donsker classes, respectively, for (θ, Λ) running through $\Theta \times H$.

S2 Additional simulation study results

Table S1: Simulation study results for the regression coefficients under Scenario 3

		$n = 100$				$n = 200$				$n = 500$			
		Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP	Bias $\times 10^2$	SD $\times 10$	ASE $\times 10$	CP
θ_1	SPMLE	5.75	1.59	1.61	0.95	3.17	1.04	1.05	0.95	1.32	0.58	0.62	0.96
	Mid	0.19	1.39	1.35	0.94	-0.80	0.96	0.92	0.94	-1.48	0.54	0.57	0.95
	End	-1.22	1.38	1.35	0.94	-2.37	0.96	0.92	0.93	-3.07	0.54	0.56	0.92
θ_2	SPMLE	-3.78	2.59	2.58	0.95	-3.07	1.75	1.75	0.94	-0.99	1.07	1.08	0.95
	Mid	1.91	2.42	2.41	0.94	1.19	1.66	1.67	0.94	1.99	1.03	1.04	0.95
	End	6.77	2.40	2.43	0.94	5.50	1.69	1.68	0.93	5.85	1.04	1.04	0.90

Bias, average of estimates minus the truth; SD, standard deviation of estimates; ASE, average of standard error estimates; CP, proportion of 95% confidence intervals that contained the truth; MLE, semiparametric maximum likelihood estimator; Mid, End and Latent: semiparametric estimators from midpoint-imputed, right-endpoint–imputed and latent right-censored data.

S3 Software

The C routine mentioned in Section 4 is available as part of the coxinterval R package, which is currently maintained at a GitHub repository by the same name:

`https://github.com/aboruvka/coxinterval`

System requirements and installation instructions are available on the repository's home page.

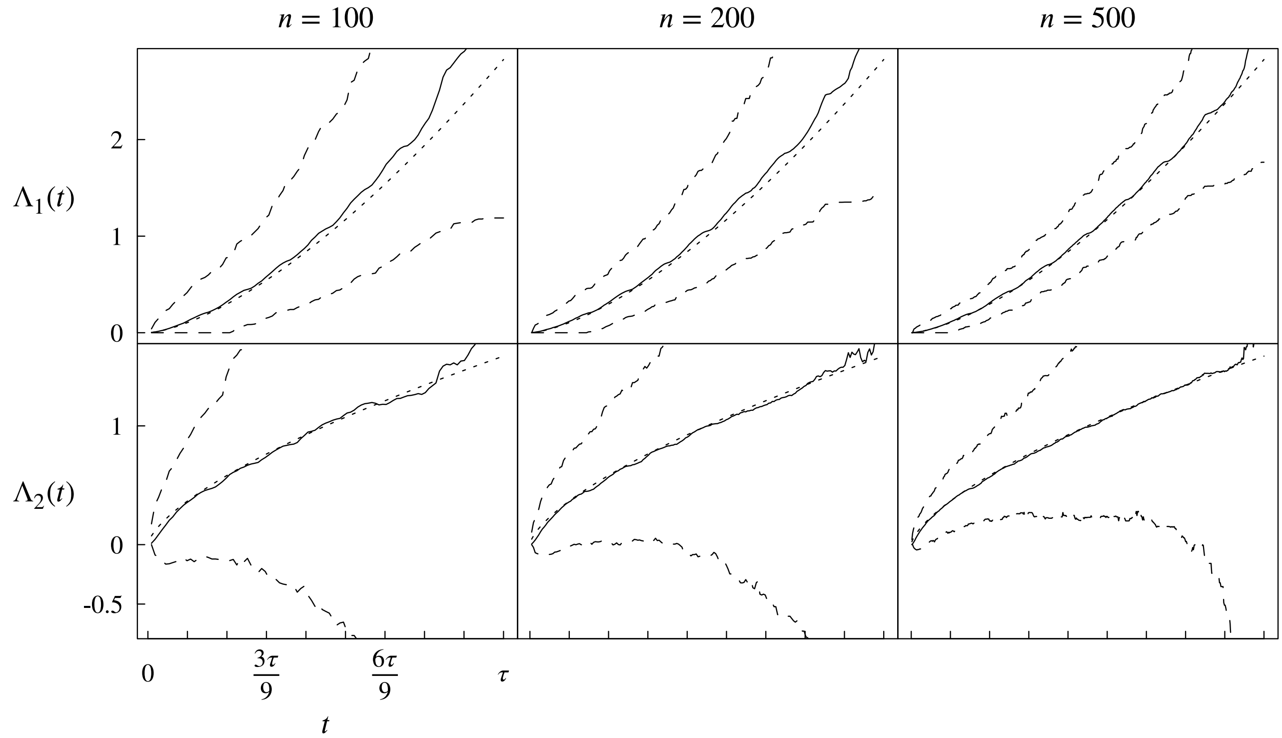

Figure S1: The true parameter value (dotted) displayed with the pointwise means (solid) and 2.5th percentiles (dashed) of $\hat{\Lambda}_n$ under Scenario 3

REFERENCES

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.*, 24(2):540–568.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: A review of recent progress. In Lin, D. Y. and Fleming, T. R., editors, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pages 123–169, New York. Springer.

Kanwal, R. P. (1997). *Linear Integral Equations*. Birkhäuser, Boston, 2nd edition.

Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.*, 25(4):1471–1509.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.*, 95(450):449–465.

van der Vaart, A. and Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In Giné, E., Mason, D. M., and Wellner, J. A., editors, *High Dimensional Probability II*, pages 115–133. Birkhäuser, Boston.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.*, 35(5):2106–2142.