Electronic Journal of Statistics
Vol. 8 (2014) 1345–1379
ISSN: 1935-7524
DOI: 10.1214/14-EJS927

Bootstrapping a change-point Cox model for survival data

Gongjun Xu

School of Statistics, University of Minnesota
e-mail: xuxxx360@umn.edu

Bodhisattva Sen and Zhiliang Ying

Department of Statistics, Columbia University
e-mail: bodhi@stat.columbia.edu; zying@stat.columbia.edu

Abstract: This paper investigates the (in)-consistency of various bootstrap methods for making inference on a change-point in time in the Cox model with right censored survival data. A criterion is established for the consistency of any bootstrap method. It is shown that the usual nonparametric bootstrap is inconsistent for the maximum partial likelihood estimation of the change-point. A new model-based bootstrap approach is proposed and its consistency established. Simulation studies are carried out to assess the performance of various bootstrap schemes.

AMS 2000 subject classifications: Primary 62N02, 62G09.
Keywords and phrases: Change-point in time, (in)-consistency of bootstrap, m-out-of-n bootstrap, non-standard asymptotics, smoothed bootstrap.

Received February 2014.

Contents

1. Introduction

The proportional hazards model of Cox (1972) specifies that the hazard function of survival time for a subject with possibly time-dependent covariate vector Z is

$$\lambda(t|Z(s), s \leq t) = \exp\left(\beta_0' Z(t)\right) \lambda_0(t), \tag{1.1}$$

where β_0 is a p-dimensional vector of regression parameters and λ_0 an unknown baseline hazard function. Inference on the regression parameter β_0 is usually based on the partial likelihood (Cox, 1975). The theoretical properties of the maximum partial likelihood estimator (MPLE) of β_0 have been studied extensively in the literature; see Andersen and Gill (1982), Fleming and Harrington (1991), and Kalbfleisch and Prentice (2002).

It is sometimes plausible to postulate that the regression coefficient changes its value at a certain time, resulting in a change-point extension of the Cox model. For clinical trial data, Meinert (1986) and Zucker and Lakatos (1990) argue that the treatment effect may manifest only after a period of time. To model such lag effect, a two-phase Cox model with a change-point in time is usually considered and the hazard function is written as

$$\lambda(t|Z(s), s \leq t) = \exp\left(\alpha_0' Z(t) 1_{t \leq \zeta_0} + \beta_0' Z(t) 1_{t > \zeta_0}\right) \lambda_0(t), \tag{1.2}$$

where a second regression parameter vector α_0 is added to model (1.1) and ζ_0 is the change-point parameter. It is clear that estimation of the change-point ζ_0 is an important step in the model based inference. For the identifiability of model (1.2), we assume throughout that $\alpha_0 \neq \beta_0$ since otherwise this model reduces to (1.1) and ζ_0 is not identifiable.

Model (1.2) has been extensively studied in the literature. Liang, Self and Liu (1990) considered the problem of testing the null hypothesis of no change-point effect based on a maximal score statistic. Luo, Turnbull and Clark (1997) focused on testing $H_0 : \zeta = \zeta_0$ versus $H_1 : \zeta \neq \zeta_0$ for a pre-specified ζ_0 and derived the asymptotic distribution of the partial likelihood ratio test statistic under H_0. For estimation of the change-point parameter ζ_0, Luo (1996) and Pons (2002) showed that the MPLE of ζ_0 is n^{-1} consistent while that of the regression parameter vector is $n^{-1/2}$ consistent. This is largely due to the fact that the partial likelihood function is not differentiable with respect to the change-point parameter and therefore the usual Taylor expansion is not applicable. This "non-standard" asymptotic behavior of the MPLE of ζ_0 is typical in change-point regression problems; see Kosorok and Song (2007), Lan, Banerjee and Michailidis (2009), and Seijo and Sen (2011a) for examples of different change-point models; see also Delsol and Van Keilegom (2011) for a general approach to non-standard semiparametric M-estimation problems.

Although the asymptotic distribution of the MPLE of ζ_0 has been derived in the literature (Luo, 1996; Pons, 2002), it cannot be directly used for making inference for ζ_0 due to the presence of nuisance parameters in the limiting distribution. Bootstrap methods bypass the difficulty of estimating the nuisance parameters and are generally reliable in standard $n^{-1/2}$ convergence problems; see Efron and Tibshirani (1993) and Davison and Hinkley (1997). Various bootstrap procedures have been applied to the standard Cox model (1.1) (Davison and Hinkley, 1997). However, when these bootstrap procedures are applied to the change-point model (1.2), they yield invalid confidence intervals (CIs) for ζ_0 (see Section 4). The failure of the usual bootstrap methods in non-standard problems has been documented in the literature; see Abrevaya and Huang (2005) and Sen, Banerjee and Woodroofe (2010) for situations giving rise to $n^{-1/3}$ asymptotics; also see Bose and Chatterjee (2001) and Cheng and Huang (2010) for M-estimation problems. The performance of different bootstrap methods for the Cox model (1.2) has not been investigated in the literature. This problem differs from the problems considered by the above authors in that compound Poisson processes, as opposed to Gaussian processes, form the backbone of the asymptotic distributions of the estimators. In addition, as in many survival analysis problems, statistical inference for the change-point parameter is further complicated by censoring.

In this paper, we propose a general framework for studying various bootstrap methods for model (1.2). We present strong theoretical and empirical evidence to suggest the inconsistency of the most commonly used bootstrap methods, including sampling directly from the empirical distribution (ED) and sampling fixing the covariates (Burr, 1994). We show that the bootstrap estimates constructed by these methods are the smallest maximizers of certain stochastic processes that, conditional on the data, do not have any weak limit. This strongly suggests not only the inconsistency but also the non-existence of any weak limit for the corresponding bootstrap estimates, although a complete proof of this non-existence is difficult and still an open problem. The main difficulty arises from the non-linearity of the smallest argmax functional. Moreover, to get consistent bootstrap procedures, we propose a smoothed bootstrap method that, fixing the covariates, draws samples from a smooth approximation of the distribution of the survival time and from an estimate of the distribution of the censoring time. A key step in this new approach is the smooth approximation to the distribution of the survival time, which makes the bootstrap scheme successfully mimic the local behavior of the true distribution function at the location ζ_0. As a result, the proposed approach yields asymptotically valid CIs for ζ_0. Furthermore, the asymptotic theory is also validated through simulation studies with reasonable sample sizes.

We would like to emphasize that although there has been some work in the literature studying the performance of bootstrap methods for change-point regression models (see, e.g., Kosorok, 2008; Seijo and Sen, 2011a), none of the authors consider the problem of change-point in time, which has wide applications in survival analysis. Also, our proof techniques differ considerably from the above mentioned papers and our conclusions can be extended to other semiparametric models with a change-point in time as well as to multiple change-point problems.

The rest of this paper is organized as follows. In Section 2 we describe the model setup and introduce the different bootstrap schemes. We state a series of convergence results in in Section 3. In Section 4 we study the inconsistency of the standard bootstrap methods, including sampling from the ED, and we prove the consistency of the smoothed and the m-out-of-n bootstrap procedures. We compare the finite sample performance of the different bootstrap methods through a simulation study in Section 5. Proofs of the main results are presented in Section 6. Proofs of several lemmas are provided in Appendix.

2. Model setup and bootstrap schemes

We use T to denote survival time and C censoring time. Throughout, $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Let $\tilde{T} = T \wedge C$ and $\delta = 1_{T \leq C}$ indicating failure (1) or censoring (0). Furthermore, there is a p-dimensional covariate process $Z(t)$, cáglád (left-continuous with right-hand limits), which may include an individual's treatment assignment and certain relevant characteristics. In this paper we focus on the external time-dependent covariate and assume that Z is observed over the study interval $[0, \tau]$, $\tau < \infty$. An external time-dependent covariate means that its value path is not directly generated by the individual under the study. Examples include the age of an individual and the air pollution level for asthma study; see Chapter 6.3.1 in Kalbfleisch and Prentice (2002) for more discussion. Furthermore, we assume that Z is of bounded total variation on $[0, \tau]$ and the covariance matrix $Var(Z(t))$ is strictly positive definite for any $t \in [0, \tau]$.

Given covariate Z, the survival time T is assumed to be conditionally independent of the censoring time C. The hazard rate function of T follows the change-point Cox model (1.2) with α_0 and β_0 belonging to bounded convex sets Θ_α and Θ_β in $\mathbb{R}^p$, respectively. We assume that the baseline hazard function $\lambda_0(\cdot)$ is bounded on $[0, \tau]$ with $\inf_{t \in [0,\tau]} \lambda_0(t) > 0$ and the conditional distribution of censoring time $G(\cdot|Z)$ satisfies $\sup_{z \in \mathcal{V}} G(\tau|z) < 1$, where $\mathcal{V}$ is the set of all possible paths of Z. To ensure the identifiability of ζ_0, we further assume that $\lambda_0(\cdot)$ and $G(\cdot|z)$, for any $z \in \mathcal{V}$, are continuous at ζ_0.

The observed data $(\tilde{T}_i, \delta_i, Z_i), i = 1, \ldots, n$, consist of n i.i.d. realizations of $(\tilde{T}, \delta, Z)$. The Cox partial likelihood (Cox, 1975) is

$$L_n(\alpha, \beta, \zeta) = \prod_{1 \leq i \leq n, \delta_i = 1} \frac{e^{\alpha' Z_i(T_i) 1_{T_i \leq \zeta} + \beta' Z_i(T_i) 1_{T_i > \zeta}}}{\sum_{1 \leq j \leq n, \tilde{T}_j \geq T_i} e^{\alpha' Z_j(T_i) 1_{T_i \leq \zeta} + \beta' Z_j(T_i) 1_{T_i > \zeta}}}. \tag{2.1}$$

Let $l_n(\alpha, \beta, \zeta) = \log L_n(\alpha, \beta, \zeta)$, which is continuous in α and β but cádlág in ζ. In fact, it is a step function in ζ and hence could have multiple maximizers. To avoid ambiguity, we say that $(\tilde{\alpha}_n', \tilde{\beta}_n', \tilde{\zeta}_n)' \in \Theta := \Theta_\alpha \times \Theta_\beta \times [0, \tau]$ is a maximizer if

$$l_n(\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\zeta}_n-) \vee l_n(\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\zeta}_n) = \sup_{(\alpha', \beta', \zeta)' \in \Theta} l_n(\alpha, \beta, \zeta).$$

Since, for each ζ, $l_n(\alpha, \beta, \zeta)$ as a function of α and β has a unique maximizer, we can choose as our MPLE the maximizer with the smallest value of ζ. In other words, our estimator $(\hat{\alpha}_n', \hat{\beta}_n', \hat{\zeta}_n)' \in \Theta$ will be the only maximizer such that if $(\tilde{\alpha}_n', \tilde{\beta}_n', \tilde{\zeta}_n)' \in \Theta$ is any other maximizer, then $\hat{\zeta}_n < \tilde{\zeta}_n$. In this case, we say that $\hat{\theta}_n := (\hat{\alpha}_n', \hat{\beta}_n', \hat{\zeta}_n)'$ is the smallest argmax of l_n and write it as

$$\hat{\theta}_n := (\hat{\alpha}_n', \hat{\beta}_n', \hat{\zeta}_n)' := \text{sargmax}_{(\alpha', \beta', \zeta)' \in \Theta} l_n(\alpha, \beta, \zeta). \tag{2.2}$$

Asymptotic properties of the MPLE $\hat{\theta}_n$ have been studied in Pons (2002). The change point estimator $\hat{\zeta}_n$ is n^{-1} consistent and converges to a minimizer of a two-sided compound Poisson process that depends on the distributions of T, C and Z (see Corollary 11). As discussed in the Introduction, it is not practical to directly use the limiting distribution of $\hat{\zeta}_n$ for constructing CIs for ζ_0. Thus it is desirable to develop bootstrap approaches.

2.1. Bootstrap procedures

We start with a brief review of bootstrap procedures. Consider a sample $\mathbf{X}_n = \{X_1, \ldots, X_n\} \overset{\text{iid}}{\sim} F_X$. Suppose that we are interested in estimating the distribution function F_{R_n} of a random variable $R_n(\mathbf{X}_n, F_X)$. A bootstrap procedure generates $\mathbf{X}_n^* = \{X_1^*, \ldots, X_{m_n}^*\} \overset{\text{iid}}{\sim} \hat{F}_{X,n}$ given $\mathbf{X}_n$, where $\hat{F}_{X,n}$ is an estimator of F_X from $\mathbf{X}_n$ and m_n is a constant depending on n, and then estimates F_{R_n} by $F_{R_n}^*$, the conditional distribution function of $R_n(\mathbf{X}_n^*, \hat{F}_{X,n})$ given $\mathbf{X}_n$. Let d denote any metric metrizing weak convergence of distributions. We say that $F_{R_n}^*$ is *weakly consistent* if $d(F_{R_n}, F_{R_n}^*) \to 0$ in probability. If F_{R_n} has a weak limit F_R, then weak consistency requires $F_{R_n}^*$ to converge weakly to F_R, in probability.

In the current context, we are interested in the distribution of $n(\hat{\zeta}_n - \zeta_0)$. Then for a consistent bootstrap procedure, the conditional distribution of $m_n(\hat{\zeta}_n^* - \hat{\zeta}_n)$ given the data must provide a good approximation to the distribution function of $n(\hat{\zeta}_n - \zeta_0)$, where $\hat{\zeta}_n^*$ is the estimator of ζ_0 obtained from the bootstrap sample. In the following we introduce several bootstrap methods commonly used in the literature for the model (1.2). We start with the classical bootstrap based on the ED.

Method 1 (Classical bootstrap)**.** Draw a random sample $\{(\tilde{T}_{n,i}^*, \delta_{n,i}^*, Z_{n,i}^*) : i = 1, \ldots, n\}$ from the ED of the data $\{(\tilde{T}_i, \delta_i, Z_i) : i = 1, \ldots, n\}$.

An alternative to the usual nonparametric bootstrap method (method 1) considered in non-regular problems is the m-out-of-n bootstrap; see, e.g., Bickel, Götze and van Zwet (1997).

Method 2 (m-out-of-n bootstrap)**.** Choose an increasing sequence $\{m_n\}_{n=1}^{\infty}$ such that $m_n = o(n)$ and $m_n \to \infty$. Draw a random sample $\{(\tilde{T}_{n,i}^*, \delta_{n,i}^*, Z_{n,i}^*) : i = 1, \ldots, m_n\}$ from the ED of the data $\{(\tilde{T}_i, \delta_i, Z_i) : i = 1, \ldots, n\}$.

Two widely used bootstrap procedures, involving sampling while fixing the covariates, for the Cox model are given in methods 3 and 4 below; see Burr (1994). These methods are model-based and need estimators of the conditional distributions of T and C given Z.

Method 3 (Bootstrap fixing the covariates)**.**

1. Fit the Cox regression model and construct an estimator of the conditional distribution of T given Z as

$$\hat{F}_n^b(t|Z) = 1 - \exp\Big(- \int_0^t e^{\hat{\alpha}_n' Z(s) 1_{s \le \hat{\zeta}_n} + \hat{\beta}_n' Z(s) 1_{s > \hat{\zeta}_n}} d\hat{\Lambda}_{n,0}^b(s)\Big), \tag{2.3}$$

where $\hat{\Lambda}_{n,0}^b(s)$ is the Breslow estimator of the cumulative baseline hazard function Λ_0, i.e.,

$$\hat{\Lambda}_{n,0}^b(t) = \int_0^t \Big(\sum_{j=1}^n Y_j(s) e^{\hat{\alpha}_n' Z_j(s) 1_{s \le \hat{\zeta}_n} + \hat{\beta}_n' Z_j(s) 1_{s > \hat{\zeta}_n}} \Big)^{-1} d\Big(\sum_{i=1}^n N_i(s)\Big)$$

with $Y_i(t) = 1_{\tilde{T}_i \ge t}$ and $N_i(t) = 1_{\tilde{T}_i \le t, \delta_i = 1}$. In addition, we construct a conditional distribution estimator $\hat{G}_n(\cdot|Z)$ of $G(\cdot|Z)$; see Section 4.1.2 for more discussion on estimating $G(\cdot|Z)$.

2. For given $Z_1, \ldots, Z_n$, generate i.i.d. replicates $\{T_{n,i}^*, C_{n,i}^* : i = 1, \ldots, n\}$ from the conditional distribution estimators $\{\hat{F}_n^b(\cdot|Z_i), \hat{G}_n(\cdot|Z_i) : i = 1, \ldots, n\}$, respectively. Then we obtain a bootstrap sample $\{(\tilde{T}_{n,i}^*, \delta_{n,i}^*, Z_i) : i = 1, \ldots, n\}$, where $\tilde{T}_{n,i}^* = T_{n,i}^* \wedge C_{n,i}^*$ and $\delta_{n,i}^* = 1_{T_{n,i}^* \le C_{n,i}^*}$.

Method 4 (Bootstrap fixing covariates and censoring)**.**

1. Same as Step 1 in method 3.

2. For given $Z_1, \ldots, Z_n$, generate $T_{n,i}^*$ from $\hat{F}_n^b(\cdot|Z_i)$. If $\delta_i = 0$, let $C_{n,i}^* = C_i$; otherwise, generate $C_{n,i}^*$ from $\hat{G}_n(\cdot|Z_i)$ conditioning on $C_{n,i}^* > T_i$.

Methods 1, 3 and 4 are the most widely used bootstrap methods for the Cox regression model. In the following sections, we demonstrate through theoretical derivation and simulation the inconsistency of these methods for constructing CIs for ζ_0. To get a consistent estimate of the distribution of $n(\hat{\zeta}_n - \zeta_0)$, we propose the following smoothed bootstrap procedures.

Method 5 (Smoothed bootstrap fixing the covariates)**.**

1. Choose an appropriate nonparametric smoothing procedure (e.g., kernel estimation method in Wells (1994)) to build an estimator $\hat{\lambda}_{n,0}$ of λ_0. The associated estimator of $F(t|Z)$ is

$$\hat{F}_n^s(t|Z) = 1 - \exp\Big(- \int_0^t e^{\hat{\alpha}_n' Z(s) 1_{s \le \hat{\zeta}_n} + \hat{\beta}_n' Z(s) 1_{s > \hat{\zeta}_n}} \hat{\lambda}_{n,0}(s) ds\Big). \tag{2.4}$$

2. For given $Z_1, \ldots, Z_n$, generate i.i.d. replicates $\{T_{n,i}^*, C_{n,i}^* : i = 1, \ldots, n\}$ from the conditional distribution estimators $\{\hat{F}_n^s(\cdot|Z_i), \hat{G}_n(\cdot|Z_i) : i = 1, \ldots, n\}$, respectively. Then we obtain a bootstrap sample $\{\tilde{T}_{n,i}^*, \delta_{n,i}^*, Z_i : i = 1, \ldots, n\}$.

Method 6 (Smoothed bootstrap fixing the covariates and censoring).

1. Same as Step 1 in method 5.
2. For given $Z_1, \ldots, Z_n$, generate $T^*_{n,i}$ from $\hat{F}^s_n(\cdot|Z_i)$. If $\delta_i = 0$, let $C^*_{n,i} = C_i$; otherwise, generate $C^*_{n,i}$ from $\hat{G}_n(\cdot|Z_i)$ conditioning on $C^*_{n,i} > T_i$.

We will use a general convergence result established in Section 3 to prove that the smoothed bootstrap procedures (methods 5 and 6) and the m-out-of-n procedure (method 2) are consistent. We will also illustrate through a simulation study that the smoothed bootstrap methods outperform the m-out-of-n method.

3. A general convergence result

In this section we prove a general convergence theorem for triangular arrays of random variables in the non-regular Cox proportional hazard model with a change-point in time. This theorem will be applied to show the consistency of the bootstrap procedures introduced in the previous section.

We first introduce some notation. Let $\mathbb{P}$ be a distribution satisfying the change-point Cox model (1.2) for some parameter $\theta_0 := (\alpha_0', \beta_0', \zeta_0)' \in \Theta := \Theta_\alpha \times \Theta_\beta \times [0, \tau]$. Consider a triangular array of independent random samples $\{(\tilde{T}_{n,i}, \delta_{n,i}, Z_{n,i}) : i = 1, \ldots, m_n\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, where $\tilde{T}_{n,i} = T_{n,i} \wedge C_{n,i}$, $\delta_{n,i} = 1_{T_{n,i} \le C_{n,i}}$, and $m_n \to \infty$ as $n \to \infty$. We use $\mathbf{E}$ to denote the expectation operator with respect to $\mathbf{P}$. Furthermore, we assume that $\{(\tilde{T}_{n,i}, \delta_{n,i}, Z_{n,i}) : i = 1, \ldots, m_n\}$ jointly follows a distribution $\mathbb{Q}_n$, and for each i, the distribution of $(\tilde{T}_{n,i}, \delta_{n,i}, Z_{n,i})$ is $\mathbb{Q}_{n,i}$.

As in Section 2, we assume that under $\mathbb{Q}_n$, the covariate process $Z(t)$ is cáglád and has bounded total variation on $[0, \tau]$. We write $Z^{\otimes 0} = 1$, $Z^{\otimes 1} = Z$, and $Z^{\otimes 2} = ZZ'$. For the ith subject, let $Y_{n,i}(t) = 1_{\tilde{T}_{n,i} \ge t}$ and $N_{n,i}(t) = 1_{\tilde{T}_{n,i} \le t, \delta_{n,i}=1}$. For $\gamma \in \mathbb{R}^p$ and $k = 0, 1$ and 2, let

$$\begin{aligned}
S_{n,k}(t;\gamma) &= \frac{1}{m_n} \sum_{i=1}^{m_n} Y_{n,i}(t) Z^{\otimes k}_{n,i}(t) \exp(\gamma' Z_{n,i}(t)),\\
s_{n,k}(t;\gamma) &= \mathbb{Q}_n\Big(\frac{1}{m_n} \sum_{i=1}^{m_n} Y_{n,i}(t) Z^{\otimes k}_{n,i}(t) \exp(\gamma' Z_{n,i}(t))\Big),\\
s_k(t;\gamma) &= \mathbb{P}\left(Y(t) Z^{\otimes k}(t) \exp(\gamma' Z)\right),\\
A_{n,k}(t) &= \mathbb{Q}_n\Big(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^t Z^{\otimes k}_{n,i}(s) dN_{n,i}(s)\Big),\\
A_k(t) &= \mathbb{P}\Big(\int_0^t Z^{\otimes k}(s) dN(s)\Big) = \int_0^t s_k(s; \alpha_0 1_{s \le \zeta_0} + \beta_0 1_{s > \zeta_0}) \lambda_0(s) ds,
\end{aligned}$$

where we use $\mathbb{Q}_n(\cdot)$ and $\mathbb{P}(\cdot)$ to denote the expectation operators under the distributions $\mathbb{Q}_n$ and $\mathbb{P}$, respectively. We write

$$\bar{Z}_n(t;\gamma) = \frac{S_{n,1}(t;\gamma)}{S_{n,0}(t;\gamma)}, \quad \bar{z}_n(t;\gamma) = \frac{s_{n,1}(t;\gamma)}{s_{n,0}(t;\gamma)}, \quad \bar{z}(t;\gamma) = \frac{s_1(t;\gamma)}{s_0(t;\gamma)}.$$

Further we denote the ratio between $S_{n,0}(t;\gamma_1)$ and $S_{n,0}(t;\gamma_2)$ by

$$R_n(t;\gamma_1,\gamma_2) = \frac{S_{n,0}(t;\gamma_1)}{S_{n,0}(t;\gamma_2)}.$$

Similarly we write

$$r_n(t;\gamma_1,\gamma_2) = \frac{s_{n,0}(t;\gamma_1)}{s_{n,0}(t;\gamma_2)}, \quad r(t;\gamma_1,\gamma_2) = \frac{s_0(t;\gamma_1)}{s_0(t;\gamma_2)}.$$

Using the above notation, for $\theta = (\alpha', \beta', \zeta)'$, the log partial likelihood function of $\{(\tilde{T}_{n,i}, \delta_{n,i}, Z_{n,i}) : i = 1, \dots, m_n\}$ takes the form

$$l_n^*(\theta) = \sum_{i=1}^{m_n} \int_0^\tau \Big((\alpha 1_{s\le\zeta} + \beta 1_{s>\zeta})' Z_{n,i} - \log S_{n,0}(s; \alpha 1_{s\le\zeta} + \beta 1_{s>\zeta}) \Big) dN_{n,i}(s).$$

Denote the MPLE of $l_n^*(\theta)$ by $\theta_n^* = (\alpha_n^{*\prime}, \beta_n^{*\prime}, \zeta_n^*)'$, i.e.,

$$\theta_n^* := \operatorname{sargmax}_{\theta\in\Theta} l_n^*(\theta).$$

Let $\theta_n = (\alpha_n', \beta_n', \zeta_n)'$ be given by

$$\theta_n := \operatorname{sargmax}_{\theta\in\Theta} \mathbb{Q}_n \Bigg(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau \big((\alpha 1_{s\le\zeta} + \beta 1_{s>\zeta})' Z_{n,i} - \log s_{n,0}(s; \alpha 1_{s\le\zeta} + \beta 1_{s>\zeta}) \big) dN_{n,i}(s) \Bigg).$$

The existence of θ_n is guaranteed as the above objective function is concave in α and β for every fixed ζ and bounded and cádlág as a function of ζ. The concavity of the objective function in α and β, for fixed ζ, follows from the concavity of the function

$$q(\gamma) := \mathbb{Q}_n \bigg(\int_{\zeta_1}^{\zeta_2} \big(\gamma' Z_{n,1} - \log s_{n,0}(s;\gamma) \big) dN_{n,1}(s) \bigg),$$

where $0 \le \zeta_1 < \zeta_2 \le \tau$; i.e., $q(c\gamma_1 + (1-c)\gamma_2) \ge cq(\gamma_1) + (1-c)q(\gamma_2)$ for any $c \in (0,1)$ and $\gamma_1, \gamma_2 \in \mathbb{R}^p$.

When $\mathbb{Q}_n$ is the ED of a sample generated from model (1.2), θ_n becomes the usual MPLE $\hat{\theta}_n$ of $l_n(\theta)$ as defined in (2.2).

In the following, we derive sufficient conditions on the distribution $\mathbb{Q}_n$ that guarantees the weak convergence of $(\sqrt{m_n}(\alpha_n^* - \alpha_n)', \sqrt{m_n}(\beta_n^* - \beta_n)', m_n(\zeta_n^* - \zeta_n))'$.

3.1. *Consistency and the rate of convergence*

We first show the consistency of the MPLE θ_n^* of l_n^*, whose proof is given in Section 6. We need the following assumption.

A1 For $k = 0, 1$ and 2, as $n \to \infty$,

$$\sup_{t\in[0,\tau],\gamma\in\Theta_\alpha\cup\Theta_\beta} |s_{n,k}(t;\gamma) - s_k(t;\gamma)| \to 0 \text{ and } \sup_{t\in[0,\tau]} |A_{n,k}(t) - A_k(t)| \to 0,$$

where $|\cdot|$ denotes the L_1 norm.

Condition A1 indicates that $\mathbb{Q}_n$ approaches the distribution satisfying the Cox model (1.2) in the sense that the difference between expectations of $S_{n,k}$ ($A_{n,k}$) under distributions $\mathbb{Q}_n$ and $\mathbb{P}$ goes to 0 as $n \to \infty$. When $\mathbb{Q}_n$ is the ED of a sample from model (1.2), the uniform law of large numbers implies A1; see Section 4.1.1 for more details.

Theorem 7. *Under condition A1,* $\theta_n \to \theta_0$ *and* $\theta_n^* \xrightarrow{\mathbf{P}} \theta_0$.

We consider the rate of convergence of θ_n^* and show that the estimators of the "regular" parameters, α_n^* and β_n^*, converge at a rate of ${m_n}^{-1/2}$ while the change-point ζ_n^* converges at rate m_n^{-1}. To guarantee the right rate of convergence, we need the following condition.

A2 There exist positive constants ρ_1 and ρ_2 such that, for any sequence $\{h_n\}$ satisfying $h_n \to \infty$ and $h_n/m_n \to 0$ as $n \to \infty$, the following holds:

$$\lim_{n\to\infty} \sup_{|h|>\frac{h_n}{m_n}} \frac{1}{h}\bigg| \int_{\zeta_n}^{\zeta_n+h} dA_{n,1}(s) - \int_{\zeta_n}^{\zeta_n+h} \bar{z}_n(s;\alpha_n 1_{s\le\zeta_n} + \beta_n 1_{s>\zeta_n}) dA_{n,0}(s)\bigg| = 0,$$

and

$$\rho_1 < \lim_{n\to\infty} \inf_{|h|>\frac{h_n}{m_n}} \frac{1}{h} |A_{n,0}(\zeta_n + h) - A_{n,0}(\zeta_n)| \le \lim_{n\to\infty} \sup_{|h|>\frac{h_n}{m_n}} \frac{1}{h} |A_{n,0}(\zeta_n + h) - A_{n,0}(\zeta_n)| < \rho_2.$$

Note that condition A2 holds if under $\mathbb{Q}_n$ the survival time T has uniformly bounded baseline hazard rate function $\lambda_{n,0}$ in some neighborhood of ζ_n. In this case, $A_{n,0}$ has right derivative $s_{n,0}(\zeta_n;\beta_n)\lambda_{n,0}(\zeta_n)$ at ζ_n^+ and left derivative $s_{n,0}(\zeta_n;\alpha_n)\lambda_{n,0}(\zeta_n)$ at ζ_n^-, which implies A2.

Theorem 8. *Under conditions A1 and A2,*

$$\big|\big(\sqrt{m_n}(\alpha_n^* - \alpha_n)', \sqrt{m_n}(\beta_n^* - \beta_n)', m_n(\zeta_n^* - \zeta_n)\big)'\big| = O_{\mathbf{P}}(1).$$

3.2. Asymptotic distribution

To compute the asymptotic distribution of θ_n^*, we need the following assumption.

A3 For any $t \in \mathbb{R}$, $h_1 < h_2$ and $0 \notin (h_1, h_2)$,

$$\mathbb{Q}_n\bigg(\sum_{k=1}^{m_n}\int_{\zeta_n+h_1/m_n}^{\zeta_n+h_2/m_n} e^{\imath t(\alpha_n-\beta_n)'Z_{n,k}(s)}dN_{n,k}(s)\bigg) \\ \to s_0\left(\zeta_0;\gamma_0+\imath t(\alpha_0-\beta_0)\right)\lambda_0(\zeta_0)(h_2-h_1),$$

where $\imath$ is the imaginary unit and $\gamma_0 = \alpha_0 1_{h_2\leq 0} + \beta_0 1_{0\leq h_1}$.

Condition A3 holds if under $\mathbb{Q}_n$ the survival time T has uniformly bounded baseline hazard rate function $\lambda_{n,0}$ converging uniformly to λ_0 in some neighborhood of ζ_0. This is satisfied by the smoothed bootstrap methods introduced in Section 2.1 and therefore guarantees their consistency; see Section 4.2.1 for more details.

We write $X_n(\theta) = {m_n}^{-1}(l_n^*(\theta) - l_n^*(\theta_n))$. Note that θ_n^* is also the maximizer of X_n. For $h := (h'_\alpha, h'_\beta, h_\zeta)' \in \mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}$, consider the multiparameter process

$$U_n^*(h) := m_n X_n\Big(\alpha_n + \frac{h_\alpha}{\sqrt{m_n}}, \beta_n + \frac{h_\beta}{\sqrt{m_n}}, \zeta_n + \frac{h_\zeta}{m_n}\Big)$$

and observe that

$$(\sqrt{m_n}(\alpha_n^*-\alpha_n)', \sqrt{m_n}(\beta_n^*-\beta_n)', m_n(\zeta_n^*-\zeta_n))' = \text{sargmax}_{h\in\mathbb{R}^{2p+1}} U_n^*(h). \quad (3.1)$$

We proceed to describe the limit law of the process U_n^*. Let Γ^- and Γ^+ be two homogenous Poisson processes with intensities

$$\gamma^- = s_0(\zeta_0;\alpha_0)\lambda_0(\zeta_0) \text{ and } \gamma^+ = s_0(\zeta_0;\beta_0)\lambda_0(\zeta_0),$$

respectively. Define two sequences of i.i.d. random variables $\mathbf{v}^- = (v_i^-)_{i=1}^\infty$ and $\mathbf{v}^+ = (v_i^+)_{i=1}^\infty$ such that v_i^- and v_i^+ follow distributions:

$$\mathbf{P}(v_i^- \leq z) = s_0(\zeta_0;\alpha_0)^{-1}\mathbf{E}\left[1_{Z(\zeta_0)\leq z}Y(\zeta_0)e^{\alpha_0' Z(\zeta_0)}\right]$$

and

$$\mathbf{P}(v_i^+ \leq z) = s_0(\zeta_0;\beta_0)^{-1}\mathbf{E}\left[1_{Z(\zeta_0)\leq z}Y(\zeta_0)e^{\beta_0' Z(\zeta_0)}\right].$$

Additionally, take two Gaussian $\mathbb{R}^p$-valued random vectors

$$U_1 \sim N\bigg(0, \int_0^{\zeta_0} Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\bigg),$$
$$U_6 \sim N\bigg(0, \int_{\zeta_0}^{\tau} Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\bigg),$$

where for $s\in\mathbb{R}$ and $\gamma\in\mathbb{R}^p$, $Q(s;\gamma)$ is defined as

$$Q(s;\gamma) = \frac{s_2(s;\gamma)}{s_0(s;\gamma)} - \bar{z}(s;\gamma)^{\otimes 2}. \quad (3.2)$$

Suppose that Γ^-,Γ^+, $\mathbf{v}^-, \mathbf{v}^+$, U_1 and U_6 are all independent. For $h_\zeta \in \mathbb{R}$, define the vector-valued process (U_2, U_3, U_4, U_5) as

$$\begin{aligned} U_2(h_\zeta) &:= 1_{h_\zeta<0}\Big(\Gamma^-(-h_\zeta)\log r(\zeta_0;\alpha_0,\beta_0) + \sum_{1\le i\le \Gamma^-(-h_\zeta)}(\beta_0-\alpha_0)v_i^-\Big), \\ U_3(h_\zeta) &:= 1_{h_\zeta<0}\Gamma^-(-h_\zeta), \\ U_4(h_\zeta) &:= 1_{h_\zeta>0}\Big(\Gamma^+(h_\zeta)\log r(\zeta_0;\beta_0,\alpha_0) + \sum_{1\le i\le \Gamma^+(h_\zeta)}(\alpha_0-\beta_0)v_i^+\Big), \\ U_5(h_\zeta) &:= 1_{h_\zeta>0}\Gamma^+(h_\zeta). \end{aligned}$$

Furthermore, define processes $J(h_\zeta) := U_3(h_\zeta) + U_5(h_\zeta)$ and

$$\begin{aligned} U(h_\alpha, h_\beta, h_\zeta) &:= h_\alpha' U_1 - \frac{1}{2}h_\alpha'\Big(\int_0^{\zeta_0} Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\Big)h_\alpha + U_2(h_\zeta) \\ &\quad + h_\beta' U_6 - \frac{1}{2}h_\beta'\Big(\int_{\zeta_0}^{\tau} Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\Big)h_\beta + U_4(h_\zeta). \end{aligned}$$

Observe that J is the sequence of jumps of U. Our goal is to show that the asymptotic distribution of the MPLE is exactly that of the smallest argmax of U. Before doing this, we state the following result about the smallest argmax of U.

Lemma 9. *Let $\phi = (\phi_\alpha', \phi_\beta', \phi_\zeta)' = sargmax_{h\in\mathbb{R}^{2p+1}}U(h)$ with ϕ_α, ϕ_β and ϕ_ζ corresponding to the first p, the second p and the last component of ϕ, respectively. Then ϕ is well-defined. Moreover, ϕ_α, ϕ_β and ϕ_ζ are mutually independent and*

$$\phi_\alpha \sim N\Big(0, \Big(\int_0^{\zeta_0} Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\Big)^{-1}\Big), \tag{3.3}$$

$$\phi_\beta \sim N\Big(0, \Big(\int_{\zeta_0}^{\tau} Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\Big)^{-1}\Big). \tag{3.4}$$

We are now in a position to give our main result. To state the result, we need to introduce some further notation. For any given compact set $K \subset \mathbb{R}^d, d \in \mathbb{N}$, we define the space $\mathcal{D}_K$ as the Skorohod space of functions $f : K \to \mathbb{R}$ having "quadrant limits" and continuous from above; see Neuhaus (1971) and Seijo and Sen (2011a) for more information about this space. Further, we take $\mathcal{D}_K$ as a metric space endowed with the Skorohod metric, which ensures the existence of conditional probability distributions for its random elements; see Neuhaus (1971) and Theorem 10.2.2 of Dudley (2002).

Theorem 10. *Under conditions A1–A3, for a compact rectangle $\Theta \subset \mathbb{R}^{2p+1}$, U_n^* converges weakly in the Skorohod topology to U in $\mathcal{D}_\Theta$. Moreover,*

$$\begin{pmatrix} \sqrt{m_n}(\alpha_n^* - \alpha_n) \\ \sqrt{m_n}(\beta_n^* - \beta_n) \\ m_n(\zeta_n^* - \zeta_n) \end{pmatrix} \rightsquigarrow sargmax_{h\in\mathbb{R}^{2p+1}}U(h),$$

where $\rightsquigarrow$ denotes weak convergence. Specifically, $\sqrt{m_n}(\alpha_n^* - \alpha_n)$, $\sqrt{m_n}(\beta_n^* - \beta_n)$, and $m_n(\zeta_n^* - \zeta_n)$ are asymptotically mutually independent, and marginally

$$\sqrt{m_n}(\alpha_n^* - \alpha_n) \quad \rightsquigarrow \quad N\left(0, \left(\int_0^{\zeta_0} Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\right)^{-1}\right),$$
$$\sqrt{m_n}(\beta_n^* - \beta_n) \quad \rightsquigarrow \quad N\left(0, \left(\int_{\zeta_0}^{\tau} Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\right)^{-1}\right),$$

and

$$m_n(\zeta_n^* - \zeta_n) \rightsquigarrow \textit{sargmax}_{h\in\mathbb{R}^{2p+1}}\{U_2(h_\zeta) + U_4(h_\zeta)\}.$$

We consider the MPLE $\hat{\theta}_n$ of $l_n(\theta)$ as defined in (2.2). In this case, we can take $m_n = n$, $\mathbb{Q}_{n,i} = \mathbb{P}$ and $\theta_n = \theta_0$. Then conditions A1–A3 automatically hold and we immediately obtain the following corollary from Theorem 10; see also Pons (2002).

Corollary 11. *Under the model setup in Section 2, for the MPLE $\hat{\theta}_n = (\hat{\alpha}_n', \hat{\beta}_n', \hat{\zeta}_n)'$,*

$$\begin{pmatrix} \sqrt{n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \\ n(\hat{\zeta}_n - \zeta_0) \end{pmatrix} \rightsquigarrow \textit{sargmax}_{h\in\mathbb{R}^{2p+1}} U(h).$$

4. Large sample properties of the bootstrap procedures

In this section we use the results from the previous section to prove the (in)-consistency of different bootstrap methods introduced in Section 2.1. In Section 4.1, we argue that the classical bootstrap method (method 1) and the methods based on fixing the covariates (methods 3 and 4) are inconsistent. In Section 4.2, we prove the consistency of the smoothed bootstrap (methods 5 and 6) and the m-out-of-n bootstrap (method 2).

Recall the notation and definitions in the beginning of Section 2. In particular, note that we have i.i.d. random vectors $(\tilde{T}_i, \delta_i, Z_i)_{i=1}^{\infty}$ from (1.2). Let $\mathcal{X}$ be the σ-algebra generated by the sequence $(\tilde{T}_i, \delta_i, Z_i)_{i=1}^{\infty}$. For a metric space $(\mathtt{X}, d)$, consider $\mathtt{X}$-valued random elements $(V_n)_{n=1}^{\infty}$ and V defined on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$. We say that V_n converges conditionally in probability to V, in probability, if for any given $\epsilon > 0$

$$\mathbf{P}(d(V_n, V) > \epsilon \mid \mathcal{X}) \xrightarrow{\mathbf{P}} 0,$$

and we write $V_n \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathcal{X}}} V$.

4.1. Inconsistent bootstrap methods

4.1.1. Classical bootstrap

Consider the classical bootstrap method 1 introduced in Section 2.1. We set $m_n = n$ and $\mathbb{Q}_{n,j}, j = 1, \ldots, n$, to be the ED of the data $\{(\tilde{T}_i, \delta_i, Z_i) : i =$

$1, \dots, n\}$. This implies that for $k = 0, 1, 2$,

$$\begin{aligned} s_{n,k}(t;\gamma) &= \mathbb{Q}_n\left(\frac{1}{m_n}\sum_{i=1}^{m_n} Y_{n,i}(t)Z_{n,i}^{\otimes k}(t)\exp(\gamma' Z_{n,i}(t))\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)Z_i^{\otimes k}(t)\exp(\gamma' Z_i(t)), \qquad (4.1) \\ A_{n,k}(t) &= \mathbb{Q}_n\left(\frac{1}{m_n}\sum_{i=1}^{m_n}\int_0^t Z_{n,i}^{\otimes k}(s)dN_{n,i}(s)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t Z_i^{\otimes k}(s)dN_i(s). \qquad (4.2) \end{aligned}$$

Therefore, $\theta_n = \hat{\theta}_n$ and condition A1 holds. Apply Theorem 7 and we have that the bootstrap estimator θ_n^* converges conditionally in probability to the true value θ_0, in probability.

Proposition 12. *For method 1,* $\theta_n^* \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathcal{X}}} \theta_0$.

As for the weak convergence, we show in Lemma 13 that condition A3 does not hold. Hence, Theorem 10 is not applicable in this case.

Lemma 13. *For method 1, there is $h_0 > 0$ such that for any $h > h_0$, the sequences*

$$\left\{\sum_{i=1}^{n}\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+\frac{h}{n}} dN_i(s)\right\}_{n=1}^{\infty} \quad \textit{and} \quad \left\{\sum_{i=1}^{n}\int_{\hat{\zeta}_n-\frac{h}{n}}^{\hat{\zeta}_n} dN_i(s)\right\}_{n=1}^{\infty} \qquad (4.3)$$

do not converge in probability. Furthermore,

$$\left\{\sum_{i=1}^{n}\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+h/n} \phi_i(s)dN_i(s)\right\}_{n=1}^{\infty} \quad \textit{and} \quad \left\{\sum_{i=1}^{n}\int_{\hat{\zeta}_n-\frac{h}{n}}^{\hat{\zeta}_n} \phi_i(s)dN_i(s)\right\}_{n=1}^{\infty}, \qquad (4.4)$$

where $\phi_i(s) := e^{\imath t((\hat{\alpha}_n-\hat{\beta}_n)' Z_i(s) - \log r_n(s;\hat{\alpha}_n,\hat{\beta}_n))} - 1$, *do not converge in probability.*

The following theorem shows that, conditional on the data, $(U_n^*)_{n=1}^{\infty}$ does not have any weak limit in probability. Consider the Skorohod space $\mathcal{D}_\Theta$ with compact set $\Theta \subset \mathbb{R}^{2p+1}$. We say that $(U_n^*)_{n=1}^{\infty}$ has no weak limit in probability in $\mathcal{D}_\Theta$ if there is no probability measure μ defined on $\mathcal{D}_\Theta$ such that $\rho(\mu_n, \mu) \xrightarrow{\mathbf{P}} 0$, where μ_n is the conditional distribution of U_n^* given $\mathcal{X}$, and ρ is a metric metrizing weak convergence on $\mathcal{D}_\Theta$.

Theorem 14. *There is a compact set $\Theta \in \mathbb{R}^{2p+1}$ such that, conditional on the data, U_n^* does not have a weak limit in probability in $\mathcal{D}_\Theta$.*

Proof of Theorem 14. It suffices to show that there is some $h > 0$ such that, conditional on the data, $U_n^*(\mathbf{0}, \mathbf{0}, h)$ does not have a weak limit in probability.

In this case,

$$U_n^*(\mathbf{0},\mathbf{0},h) = \sum_{i=1}^n \int_{\hat\zeta_n}^{\hat\zeta_n+h/n} \big((\hat\alpha_n - \hat\beta_n)' Z_{n,i}(s) - \log R_n(s;\hat\alpha_n,\hat\beta_n)\big) dN_{n,i}(s).$$

Consider the conditional characteristic function of $U_n^*(\mathbf{0},\mathbf{0},h)$ given $\mathcal{X}$. A similar argument as in the proof of Lemma 21 implies that

$$\mathbf{E}[e^{\imath t U_n^*(\mathbf{0},\mathbf{0},h)}|\mathcal{X}] = (1+o_{\mathbf{P}}(1))\exp\Big(\sum_{i=1}^n \int_{\hat\zeta_n}^{\hat\zeta_n+h/n} \phi_i(s) dN_i(s)\Big),$$

where $\phi_i(s)$ is defined as in Lemma 13. Then Lemma 13 implies the desired conclusion. □

The result that U_n^* does not have any weak limit in probability makes the existence of a weak limit for $n(\zeta_n^* - \hat\zeta_n)$ very unlikely; see (3.1). But a complete proof of the non-existence may be complicated due to the non-linearity of the smallest argmax functional. For this reason, theoretically we do not pursue this problem any further, and we will use simulation results to illustrate the inconsistency in Section 5.

4.1.2. Bootstrap fixing the covariates

For methods 3 and 4 in Section 2.1, we consider that $m_n = n$, $\mathbb{Q}_{n,i}(Z_{n,i} = Z_i) = 1$, and the cumulative hazard function of T takes the form $\Lambda_{n,0} = \hat\Lambda_{n,0}^b$, where $\hat\Lambda_{n,0}^b$ is the Breslow estimator as defined in method 3. Therefore, for $k = 0,1,2$,

$$s_{n,k}(t;\gamma) = \frac{1}{n}\sum_{i=1}^n \mathbb{Q}_{n,i}(Y_{n,i}(t)|Z_i) Z_i^{\otimes k}(t)\exp(\gamma' Z_i(t))$$

and

$$A_{n,k}(t) = \int_0^t s_{n,k}(s;\hat\alpha_n 1_{s\le\hat\zeta_n} + \hat\beta_n 1_{s>\hat\zeta_n}) d\hat\Lambda_{n,0}^b(s).$$

Thus $\theta_n = \hat\theta_n$.

A uniformly consistent estimator of G is usually needed in bootstrap methods, where sampling is done fixing the covariates, for the Cox model. We assume that

$$\sup_{t\in[0,\tau],z\in\mathcal{V}} |\hat G_n(t|z) - G(t|z)| \xrightarrow{\mathbf{P}} 0, \tag{4.5}$$

where $\mathcal{V}$ is the set of all possible sample paths of covariate Z. Note that $\hat G_n$ can be taken as the Kaplan-Meier estimator when C_i's are i.i.d. or Z_i's are time-independent and categorical; see also Beran (1981) for a class of nonparametric estimates of the conditional distribution. For more general time-dependent covariates Z, it is hard, if not impossible, to obtain a consistent estimator of G without further model assumption. In the literature, a common approach is to assume that the censoring time follows the Cox model (1.1), in which case a consistent estimator of G can be constructed based on the usual Breslow estimator (Cox and Oakes, 1984). In this paper we assume that (4.5) holds and do not go into the problem of estimating G any further.

Under the setup in Section 2, it is known that $\sup_{t\in[0,\tau]}|\Lambda_{n,0}(t)-\Lambda_0(t)| \xrightarrow{\mathbf{P}} 0$ (Andersen *et al.*, 1993). Together with (4.5), this implies condition A1. Apply Theorem 7 and we have the following convergence result.

Proposition 15. *For methods 3 and 4, if (4.5) holds, then* $\theta_n^* \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathcal{X}}} \theta_0$.

As with method 1, we argue that methods 3 and 4 are also inconsistent. We start with the following lemma.

Lemma 16. *For methods 3 and 4, there is* $h_0 > 0$ *such that for any* $h > h_0$, *the sequences*

$$\left\{n\left(\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n+\frac{h}{n}\right)-\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n\right)\right)\right\}_{n=1}^{\infty} \quad \text{and} \quad \left\{n\left(\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n\right)-\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n-\frac{h}{n}\right)\right)\right\}_{n=1}^{\infty}$$

do not converge in probability.

Proof of Lemma 16. We only need to show that the first sequence does not converge in probability. For $h > 0$,

$$\begin{aligned}
& n\left(\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n+\frac{h}{n}\right)-\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n\right)\right) \\
= \; & n\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+\frac{h}{n}} \Big(\sum_{j=1}^n Y_j(s)e^{\hat{\alpha}_n' Z_j(s)1_{s\le\hat{\zeta}_n}+\hat{\beta}_n' Z_j(s)1_{s>\hat{\zeta}_n}}\Big)^{-1} d\Big(\sum_{i=1}^n N_i(s)\Big) \\
= \; & (1+o_{\mathbf{P}}(1))\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+\frac{h}{n}} s_0(s;\beta_0)^{-1} d\Big(\sum_{i=1}^n N_i(s)\Big) \\
= \; & (1+o_{\mathbf{P}}(1)) s_0(\zeta_0;\beta_0)^{-1}\sum_{i=1}^n \int_{\hat{\zeta}_n}^{\hat{\zeta}_n+\frac{h}{n}} dN_i(s).
\end{aligned}$$

Thus, it suffices to show that $\sum_{i=1}^n \int_{\hat{\zeta}_n}^{\hat{\zeta}_n+\frac{h}{n}} dN_i(s)$ does not converge in probability. Apply Lemma 13 and we have the desired conclusion. □

Based on Lemma 16, we further show that, conditional on the data, the sequence $\{U_n^*\}_{n=1}^{\infty}$ does not have a weak limit in probability.

Theorem 17. *There is a compact set* $\Theta \in \mathbb{R}^{2p+1}$ *such that, conditional on the data,* U_n^* *does not have a weak limit in probability in* $\mathcal{D}_\Theta$.

Proof of Theorem 17. For $h > 0$, consider the conditional characteristic function of $U_n^*(\mathbf{0},\mathbf{0},h)$ given $\mathcal{X}$. A similar argument as in the proof of Lemma 21 implies that

$$\mathbf{E}[e^{\imath t U_n^*(\mathbf{0},\mathbf{0},h)}|\mathcal{X}] = (1+o_{\mathbf{P}}(1))\exp\Big\{n\Big(\hat{\Lambda}^b_{n,0}\Big(\hat{\zeta}_n+\frac{h}{n}\Big)-\hat{\Lambda}^b_{n,0}\left(\hat{\zeta}_n\right)\Big) \times\Big(e^{-\log r(\zeta_0;\alpha_0,\beta_0)}\; s_0(\zeta_0;\imath t(\alpha_0-\beta_0)+\beta_0)-s_0(\zeta_0;\beta_0)\Big)\Big\}.$$

Hence, Lemma 16 implies the desired conclusion. □

4.2. Consistent bootstrap methods

In this section we show that the smoothed bootstrap (methods 5 and 6) and the m-out-of-n bootstrap (method 2) are consistent for constructing CIs for ζ_0.

The results from Section 3 can be directly applied to derive sufficient conditions on the distribution from which the bootstrap samples are generated. Let $\hat{\mathbb{Q}}_n$ be a distribution constructed from the data $\{(\tilde{T}_i,\delta_i,Z_i): i=1,\ldots,n\}$. If conditions A1–A3 hold with $\mathbb{Q}_n=\hat{\mathbb{Q}}_n$, then the weak convergence of the bootstrap estimate follows from Theorem 10 applied conditionally given the data.

4.2.1. Smoothed bootstrap

Consider methods 5 and 6. To prove the consistence, thanks to Theorem 10, we only need to show conditions A1–A3 hold conditionally on the data with $m_n=n$ and $\mathbb{Q}_n$ the distribution of the bootstrap sample. Recall that $\hat{\lambda}_{n,0}(\cdot)$ and $\hat{G}(\cdot|Z)$ are the estimated smoothed baseline hazard rate function of T and the conditional distribution of C given Z, respectively. In addition to (4.5), we need the following convergence result:

$$\sup_{t\in[0,\tau]}|\hat{\lambda}_{n,0}(t)-\lambda_0(t)|\xrightarrow{\mathbf{P}}0. \tag{4.6}$$

Note that (4.6) is fulfilled if $\hat{\lambda}_{n,0}$ is the usual kernel estimator (Wells, 1994).

Similarly as in Section 4.1.2, we have $\theta_n=\hat{\theta}_n$, $s_{n,k}(t;\gamma)\xrightarrow{\mathbf{P}}s_k(t;\gamma)$, and $A_{n,k}(t)\xrightarrow{\mathbf{P}}A_k(t)$. In addition,

$$\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+h}dA_{n,1}(s)-\int_{\hat{\zeta}_n}^{\hat{\zeta}_n+h}\bar{z}_n(s;\hat{\alpha}_n 1_{s\le\hat{\zeta}_n}+\hat{\beta}_n 1_{s>\hat{\zeta}_n})dA_{n,0}(s)=0,$$

and for any $t\in\mathbb{R}$, $h_1<h_2$ and $0\notin(h_1,h_2)$,

$$\begin{aligned}&\mathbb{Q}_n\Big(\sum_{i=1}^n\int_{\hat{\zeta}_n+h_1/n}^{\hat{\zeta}_n+h_2/n}e^{\imath t(\hat{\alpha}_n-\hat{\beta}_n)'Z_{n,k}(s)}dN_{n,i}(s)\Big)\\ =\;& n\int_{\hat{\zeta}_n+h_1/n}^{\hat{\zeta}_n+h_2/n}s_0\big(s;\gamma_n+\imath t(\hat{\alpha}_n-\hat{\beta}_n)\big)\hat{\lambda}_0(s)ds\end{aligned}$$

$$\xrightarrow{\mathbf{P}} \quad s_0\big(\zeta_0;\gamma_0+\imath t(\alpha_0-\beta_0)\big)\lambda_0(\zeta_0)(h_2-h_1),$$

where $\gamma_n=\alpha_n 1_{h_2\le 0}+\beta_n 1_{0\le h_1}$ and $\gamma_0=\alpha_0 1_{h_2\le 0}+\beta_0 1_{0\le h_1}$. Therefore, A1–A3 hold and Theorems 7 and 10 give the weak consistency result.

Proposition 18. *For methods 5 and 6, if* (4.5) *and* (4.6) *hold, then* $\theta_n^* \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathcal{X}}} \theta_0$ *and conditional on the data,*

$$\begin{pmatrix} \sqrt{n}(\alpha_n^*-\hat\alpha_n) \\ \sqrt{n}(\beta_n^*-\hat\beta_n) \\ n(\zeta_n^*-\hat\zeta_n) \end{pmatrix} \rightsquigarrow \mathit{sargmax}_{h\in\mathbb{R}^{2p+1}} U(h). \tag{4.7}$$

4.2.2. m-out-of-n bootstrap

Consider the m-out-of-n bootstrap (method 2). We will again use the consistency results established in Section 3. We set $m_n\to\infty$ and $m_n/n\to 0$ as $n\to\infty$. Similar to the classical bootstrap, $\mathbb{Q}_{n,j}$, $j=1,\dots,n$, is the ED of the data $\{(\tilde T_i,\delta_i,Z_i): i=1,\dots,n\}$, and (4.1) and (4.2) hold. Therefore $\theta_n=\hat\theta_n$ and condition A1 holds. Consider the first equation in A2 and we have

$$\begin{aligned}
&\sup_{|h|>h_n/m_n}\frac{1}{h}\bigg|\int_{\hat\zeta_n}^{\hat\zeta_n+h} dA_{n,1}(s)-\int_{\hat\zeta_n}^{\hat\zeta_n+h}\bar z_n(s;\hat\alpha_n 1_{s\le\hat\zeta_n}+\hat\beta_n 1_{s>\hat\zeta_n})dA_{n,0}(s)\bigg| \\
=&\sup_{|h|>h_n/m_n}\frac{1}{nh}\bigg|\sum_{i=1}^n\int_{\hat\zeta_n}^{\hat\zeta_n+h}\big[Z_i-\bar z_n(s;\hat\alpha_n 1_{s\le\hat\zeta_n}+\hat\beta_n 1_{s>\hat\zeta_n})\big]dN_i(s)\bigg| \\
\xrightarrow{\mathbf{P}}&\ 0.
\end{aligned}$$

As to condition A3, for any $t\in\mathbb{R}$, $h_1<h_2$ and $0\notin(h_1,h_2)$,

$$\begin{aligned}
&\mathbb{Q}_n\bigg(\sum_{i=1}^{m_n}\int_{\hat\zeta_n+h_1/m_n}^{\hat\zeta_n+h_2/m_n} e^{\imath t(\hat\alpha_n-\hat\beta_n)'Z_{n,i}(s)}dN_{n,i}(s)\bigg) \\
=&\ \frac{m_n}{n}\sum_{i=1}^n\int_{\hat\zeta_n+h_1/m_n}^{\hat\zeta_n+h_2/m_n} e^{\imath t(\hat\alpha_n-\hat\beta_n)'Z_i(s)}dN_i(s) \\
\xrightarrow{\mathbf{P}}&\ s_0\left(\zeta_0;\gamma_0+\imath t(\alpha_0-\beta_0)\right)\lambda_0(\zeta_0)(h_2-h_1),
\end{aligned}$$

where $\gamma_0=\alpha_0 1_{h_2\le 0}+\beta_0 1_{0\le h_1}$. Therefore, A1–A3 hold and we have the following proposition.

Proposition 19. *For the m-out-of-n bootstrap method, if* $m_n\to\infty$ *and* $m_n/n\to 0$ *as* $n\to\infty$, *then* $\theta_n^* \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathcal{X}}} \theta_0$ *and conditional on the data*

$$\begin{pmatrix} \sqrt{m_n}(\alpha_n^*-\hat\alpha_n) \\ \sqrt{m_n}(\beta_n^*-\hat\beta_n) \\ m_n(\zeta_n^*-\hat\zeta_n) \end{pmatrix} \rightsquigarrow \mathit{sargmax}_{h\in\mathbb{R}^{2p+1}} U(h).$$

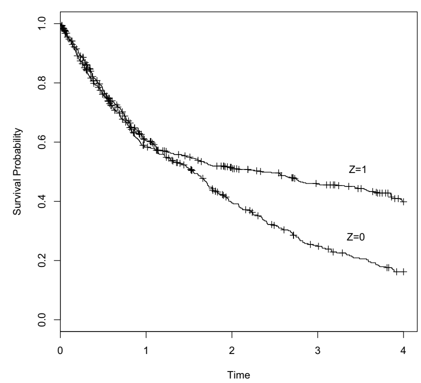

FIG 1. *Kaplan-Meier curves of a simulated sample.*

Remark 20. As pointed out in the Introduction, our results of the different bootstrap methods can be extended to multiple change-point problems. In particular, for the Cox model with multiple change-points, if the number of change points is known or can be estimated consistently, we may conclude from a similar analysis that bootstrapping from the ED and bootstrapping fixing the covariates will not be consistent, while the smoothed bootstrap can be easily adapted to produce valid CIs. However, when the number of change-points are unknown, the estimation of the change-points is more complicated. Binary segmentation procedures and its variants (see Fryzlewicz (2014) and the references therein) can be used to detect the number of change points.

5. Simulation studies

In this section we compare the finite sample performance of the different bootstrap schemes introduced in Section 2.1. We consider a single covariate Z which has a Bernoulli distribution with parameter 0.5. That is, a subject is equally likely to be assigned to the control group ($Z = 0$) and the treatment group ($Z = 1$). The model parameter values are set at $\alpha_0 = 0, \beta_0 = -1.5$, and $\zeta_0 = 1$. The baseline hazard rate is assumed constant and taken as $\lambda_0(t) = 0.5$. Note that, at $\zeta_0 = 1$, the cumulative mortality for the control group is $1-\exp(-0.5) =$ 39%. The censoring times are chosen to be independent and follow an exponential distribution with rate parameter 0.1 and truncated at $\tau = 4$. This results in a censoring rate of about 36%. Figure 1 gives the Kaplan-Meier curves of a simulated sample of size $n = 1000$, which clearly shows the lag feature around the change-point time $\zeta_0 = 1$.

We consider 1000 random samples of sample sizes $n = 100, 300, 500$, and 1000. For each simulated sample and for each bootstrap method, 1000 bootstrap replicates are generated to approximate the bootstrap distribution. The conditional censoring distribution estimator $\hat{G}(\cdot|Z)$ is taken as the Kaplan-Meier estimators for each group ($Z = 0$ and $Z = 1$). For the smoothed bootstrap, we

TABLE 1
The estimated coverage rates and average lengths of nominal 95% CIs for ζ_0

n		M5	M6	M1	M3	M4	$n^{4/5}$	$n^{9/10}$	$n^{14/15}$
100	Coverage	0.91	0.90	0.78	0.81	0.80	0.87	0.84	0.78
	Length	0.85	0.86	0.64	0.69	0.70	0.74	0.71	0.67
300	Coverage	0.93	0.94	0.88	0.87	0.86	0.97	0.93	0.91
	Length	0.64	0.62	0.56	0.54	0.55	0.80	0.69	0.64
500	Coverage	0.96	0.94	0.89	0.89	0.85	0.98	0.96	0.94
	Length	0.46	0.45	0.48	0.46	0.49	0.77	0.62	0.58
1000	Coverage	0.95	0.96	0.89	0.88	0.86	0.99	0.97	0.94
	Length	0.24	0.25	0.30	0.29	0.28	0.69	0.48	0.42

use a kernel density estimator based on the Gaussian kernel and choose the so-called "normal-reference rule" (Scott, 1992). For the m-out-of-n bootstrap, we try three different choices of m_n: $n^{4/5}$, $n^{9/10}$ and $n^{14/15}$. To reduce the computation complexity, we restrict $\zeta \in [0.5, 1.5]$ when calculating the MPLE of ζ_0.

Table 1 provides the simulation results of coverage proportions and average lengths of nominal 95% CIs for ζ_0 that are estimated using different bootstrap methods. The first two columns ("M5" and "M6") give the results of smoothed bootstrap methods 5 and 6, the third column ("M1") corresponds to the classical bootstrap method 1, the columns ("M3" and "M4") correspond to bootstrap methods 3 and 4, and the last three columns correspond to the m-out-of-n bootstrap with different choices of m_n.

We can see from Table 1 that the smoothed bootstrap outperforms all the others in terms of coverage rate. The results of methods 4 and 6 are similar to those of methods 3 and 5, respectively. The m-out-of-n bootstrap also performs reasonably well for large sample sizes, but the average length is bigger than that of the smoothed bootstrap. This may be due to the fact that the m-out-of-n bootstrap method converges at rate m_n^{-1} instead of n^{-1}. Table 1 also shows that the commonly used bootstrap methods 1, 3 and 4 provide under-coverage, which indicates their inconsistency.

To further illustrate the performance of different bootstrap methods, we compare the histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$, obtained from 1000 random samples of sample size 1000, and its bootstrap estimates from a single sample. All bootstrap estimates are based on 1000 bootstrap replicates. It is clearly shown in Figure 2 that the smoothed bootstrap (top right panel) provides the best approximation to the actual distribution obtained from 1000 random samples (top left panel).

In Table 2, we present the estimated coverage rates and average lengths for different β_0 values and fixed $\alpha_0 = 0$ using the smoothed bootstrap method 5. We can see that sharper differences between α_0 and β_0 give better coverage rates and shorter CIs.

6. Proof of Theorems

This section contains proofs of Theorems 7, 8 and 10.

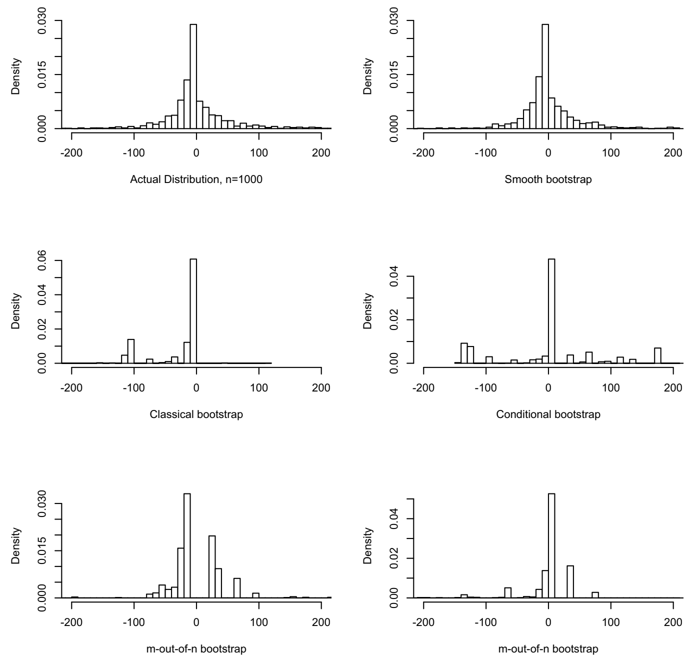

Fig 2. *Histograms of the distribution of* $n(\hat{\zeta}_n - \zeta_0)$ *and its bootstrap estimates. The top left panel shows the distribution of* $n(\hat{\zeta}_n - \zeta_0)$ *obtained from 1000 random samples; the top right panel shows the distribution of* $n(\hat{\zeta}_n^* - \hat{\zeta}_n)$ *for the smoothed bootstrap (method 5), the middle left panel shows the classical bootstrap (method 1), the middle right panel shows the bootstrap method fixing the covariates (method 3). The bottom left panel shows the distribution of* $m_n(\hat{\zeta}_n^* - \hat{\zeta}_n)$ *with* $m_n = n^{9/10}$*, and the bottom right for* $m_n = n^{14/15}$.

Table 2
The estimated coverage rates and average lengths of nominal 95% CIs for different β_0*'s. The sample size is 500*

β_0	-0.5	-1	-1.5	-2
Coverage	0.84	0.90	0.96	0.94
Length	0.90	0.68	0.46	0.31

6.1. Proof of Theorem 7

We first show that $\theta_n \to \theta_0$. Write $\Theta = \Theta_\alpha \times \Theta_\beta \times [0, \tau]$. By the definition, $\theta_n = (\alpha_n', \beta_n', \zeta_n)'$ is the smallest maximizer of

$$X_0(\theta) := X_0(\alpha, \beta, \zeta) := \int_0^\tau \left((\alpha - \alpha_0)1_{s \le \zeta} + (\beta - \beta_0)1_{s > \zeta}\right)' dA_{n,1}(s) \\ - \int_0^\tau \log \frac{s_{n,0}(s; \alpha 1_{s \le \zeta} + \beta 1_{s > \zeta})}{s_{n,0}(s; \alpha_0 1_{s \le \zeta_0} + \beta_0 1_{s > \zeta_0})} dA_{n,0}(s).$$

Thus, $X_0(\theta_n) \geq 0$. By condition A1, we have

$$|X_0(\theta_n) - X(\theta_n)| \leq \sup_{\theta \in \Theta} |X_0(\theta) - X(\theta)| \to 0,$$

where

$$\begin{aligned} X(\theta) = & \int_0^\tau ((\alpha - \alpha_0)' s_1(s; \alpha_0) - s_0(s; \alpha_0) \log r(s; \alpha, \alpha_0)) 1_{s \leq \zeta \wedge \zeta_0} d\Lambda_0(s) \\ & + \int_0^\tau ((\alpha - \beta_0)' s_1(s; \beta_0) - s_0(s; \beta_0) \log r(s; \alpha, \beta_0)) 1_{\zeta_0 < s \leq \zeta} d\Lambda_0(s) \\ & + \int_0^\tau ((\beta - \alpha_0)' s_1(s; \alpha_0) - s_0(s; \alpha_0) \log r(s; \beta, \alpha_0)) 1_{\zeta < s \leq \zeta_0} d\Lambda_0(s) \\ & + \int_0^\tau ((\beta - \beta_0)' s_1(s; \beta_0) - s_0(s; \beta_0) \log r(s; \beta, \beta_0)) 1_{s > \zeta \vee \zeta_0} d\Lambda_0(s). \end{aligned} \quad (6.1)$$

Then by the continuous mapping theorem, the conclusion follows from the fact that θ_0 is the unique maximizer of $X(\theta)$ and that $X(\theta_0) = 0$.

We now show the consistency of $\hat{\theta}_n^*$. For $\gamma_n \in \{\alpha_n, \beta_n\}$ and $\gamma \in \Theta_\alpha \cup \Theta_\beta$, let

$$w_n(t; \gamma) := \int_0^t (\gamma - \gamma_n)' dM_{n,1}(s) - \int_0^t \log r_n(s; \gamma, \gamma_n) dM_{n,0}(s),$$

where

$$M_{n,k}(t) = \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^t Z_{n,i}^{\otimes k}(s) dN_{n,i}(s) - A_{n,k}(t), \quad \text{for } k = 0, 1. \quad (6.2)$$

For any $\epsilon_1 > 0$, we have

$$\begin{aligned} \mathbb{Q}_n\Big(\sup_{t \in [0,\tau]} |w_n(t; \gamma)| \geq 2\epsilon_1 \Big) \leq & \mathbb{Q}_n\Big(\sup_{t \in [0,\tau]} \Big| \int_0^t (\gamma - \gamma_n)' dM_{n,1}(s) \Big|^2 \geq \epsilon_1^2 \Big) \\ & + \mathbb{Q}_n\Big(\sup_{t \in [0,\tau]} \Big| \int_0^t \log r_n(s; \gamma, \gamma_n) dM_{n,0}(s) \Big|^2 \geq \epsilon_1^2 \Big). \end{aligned}$$

Then by Lenglart's inequality for cádlág processes (Jacod and Shiryaev, 2002, p. 35), we have that for $\epsilon_2 > 0$, there exists a constant $B > 0$ such that

$$\begin{aligned} & \mathbb{Q}_n\Big(\sup_{t \in [0,\tau]} |w_n(t; \gamma)| \geq 2\epsilon_1 \Big) \\ \leq & \; 2\Big(\frac{\epsilon_2}{\epsilon_1} + \frac{B}{\epsilon_1^2 m_n} \Big) + \mathbb{Q}_n\Big(\frac{1}{m_n^2} \sum_{i=1}^{m_n} \int_0^t \big((\gamma - \gamma_n)' Z_{n,i}(s) \big)^2 dN_{n,i}(s) > \epsilon_2 \Big) \\ & + \mathbb{Q}_n\Big(\frac{1}{m_n^2} \sum_{i=1}^{m_n} \int_0^t \big(\log r_n(s; \gamma, \gamma_n) \big)^2 dN_{n,i}(s) > \epsilon_2 \Big) \\ \leq & \; 2\Big(\frac{\epsilon_2}{\epsilon_1^2} + \frac{B}{\epsilon_1^2 m_n} \Big) + \frac{2B}{\epsilon_2 m_n^2}, \end{aligned} \quad (6.3)$$

where the last inequality follows from Chebyshev inequality. Since ϵ_1 and ϵ_2 are arbitrary, it follows that

$$\sup_{t\in[0,\tau]} |w_n(t;\gamma)| \xrightarrow{\mathbf{P}} 0, \quad \text{for } \gamma \in \Theta_\alpha \cup \Theta_\beta. \tag{6.4}$$

On the other hand, by Lemma 23 in the Appendix, we have

$$\sup_{t\in[0,\tau]} \left| \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^t (\log R_n(s;\gamma,\gamma_n) - \log r_n(s;\gamma,\gamma_n))\, dN_{n,i}(s) \right| \to 0. \tag{6.5}$$

Thus, (6.4) and (6.5) imply that for $\theta = (\alpha', \beta', \zeta)'$ with $\alpha \in \Theta_\alpha$ and $\beta \in \Theta_\beta$,

$$\sup_{\zeta\in[0,\tau]} |X_n(\theta) - X_n^*(\theta)| \xrightarrow{\mathbf{P}} 0,$$

where $X_n(\theta) = {m_n}^{-1}(l_n^*(\theta) - l_n^*(\theta_n))$ and

$$\begin{aligned} X_n^*(\theta) = {} & \mathbb{Q}_n\left(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau ((\alpha-\alpha_n)' Z_{n,i}(s) - \log r_n(s;\alpha,\alpha_n))\, 1_{s\le\zeta\wedge\zeta_n} dN_{n,i}(s) \right) \\ & + \mathbb{Q}_n\left(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau ((\alpha-\beta_n)' Z_{n,i}(s) - \log r_n(s;\alpha,\beta_n))\, 1_{\zeta_n< s\le\zeta} dN_{n,i}(s) \right) \\ & + \mathbb{Q}_n\left(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau ((\beta-\alpha_n)' Z_{n,i}(s) - \log r_n(s;\beta,\alpha_n))\, 1_{\zeta< s\le\zeta_n} dN_{n,i}(s) \right) \\ & + \mathbb{Q}_n\left(\frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau ((\beta-\beta_n)' Z_{n,i}(s) - \log r_n(s;\beta,\beta_n))\, 1_{s>\zeta\vee\zeta_n} dN_{n,i}(s) \right). \end{aligned}$$

A similar argument as in the proof of Theorem II.1 in Andersen and Gill (1982) implies that the convergence of $|X_n(\theta) - X_n^*(\theta)|$ is uniform in $\theta \in \Theta$; cf. Lemma 1 in Hjort and Pollard (2011). Then by the result that $\theta_n \to \theta_0$ and condition A1, we have that

$$\sup_{\theta\in\Theta} |X_n(\theta) - X(\theta)| \xrightarrow{\mathbf{P}} 0,$$

where $X(\theta)$ is defined as in (6.1). Apply Corollary 3.2.3 (ii) in van der Vaart and Wellner (1996) and we obtain the desired convergence result.

6.2. Proof of Theorem 8

For $\alpha = (\alpha_1, \dots, \alpha_p)'$ and $\beta = (\beta_1, \dots, \beta_p)'$, let $Conv(\alpha, \beta) := \{\gamma = (\gamma_1, \dots, \gamma_p)' : \beta_i \wedge \alpha_i \le \gamma_i \le \beta_i \vee \alpha_i, i = 1, \dots, p\}$. By the definition of MPLE, we have $0 \le X_n(\theta_n^*) - X_n(\theta_n)$. Take Taylor's expansion of $X_n(\theta_n^*) - X_n(\theta_n)$ with respect to α_n and β_n and we obtain that there exist $\tilde{\alpha}_n \in Conv(\alpha_n^*, \alpha_n)$ and $\tilde{\beta}_n \in Conv(\beta_n^*, \beta_n)$ such that

$$0 \le \sqrt{m_n}|\alpha_n^* - \alpha_n| I_1 + \sqrt{m_n}|\beta_n^* - \beta_n| I_2$$

$$- \frac{m_n}{2}|\alpha_n^* - \alpha_n|^2 \Big| \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau Q_n(s;\tilde{\alpha}_n) 1_{s \le \zeta_n} dN_{n,i}(s) \Big|$$
$$- \frac{m_n}{2}|\beta_n^* - \beta_n|^2 \Big| \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau Q_n(s;\tilde{\beta}_n) 1_{s > \zeta_n} dN_{n,i}(s) \Big|$$
$$+ m_n|\zeta_n - \zeta_n^*| a_n + m_n |\zeta_n - \zeta_n^*| b_n,$$

where

$$I_1 = \Big| \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \int_0^\tau \left(Z_{n,i}(s) - \bar{Z}_n(s;\alpha_n) \right) 1_{s \le \zeta_n} dN_{n,i}(s) \Big|,$$
$$I_2 = \Big| \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \int_0^\tau \left(Z_{n,i}(s) - \bar{Z}_n(s;\beta_n) \right) 1_{s > \zeta_n} dN_{n,i}(s) \Big|,$$

$$a_n = \frac{1}{m_n|\zeta_n - \zeta_n^*|} \sum_{i=1}^{m_n} \int_0^\tau ((\alpha_n^* - \beta_n^*)' Z_{n,i}(s) - \log R_n(s;\alpha_n^*,\beta_n^*))\, 1_{\zeta_n < s \le \zeta_n^*} dN_{n,i}(s),$$
$$b_n = \frac{1}{m_n|\zeta_n - \zeta_n^*|} \sum_{i=1}^{m_n} \int_0^\tau ((\beta_n^* - \alpha_n^*)' Z_{n,i}(s) - \log R_n(s;\beta_n^*,\alpha_n^*))\, 1_{\zeta_n^* < s \le \zeta_n} dN_{n,i}(s),$$

and Q_n is defined as

$$Q_n(s;\alpha) := \frac{S_{n,2}(s;\alpha)}{S_{n,0}(s;\alpha)} - \bar{Z}_n(s;\alpha)^{\otimes 2}.$$

Lemma 23 implies that $Q_n(s;\alpha)$ converges uniformly to $Q(s;\alpha)$, where $Q(s;\alpha) = s_2(s;\alpha)/s_0(s;\alpha) - \bar{z}(s;\alpha)^{\otimes 2}$ is as defined in (3.2). Let $\sigma_0(A)$ denote the smallest eigenvalue of a matrix A. Since σ_0 is a continuous function on $\mathbb{R}^{p\times p}$, we have $\sigma_0(Q_n)$ converges to $\sigma_0(Q)$ and therefore $\sigma_0(Q_n(s;\tilde{\alpha}_n))$ and $\sigma_0(Q_n(s;\tilde{\beta}_n))$ are positive for all large n. Then by the positive definiteness of Q_n, we obtain

$$0 \le \sqrt{m_n}|\alpha_n^* - \alpha_n| I_1 + \sqrt{m_n}|\beta_n^* - \beta_n| I_2 - \frac{m_n}{2}|\alpha_n^* - \alpha_n|^2 I_3 - \frac{m_n}{2}|\beta_n^* - \beta_n|^2 I_4$$
$$+ m_n|\zeta_n - \zeta_n^*| a_n + m_n|\zeta_n - \zeta_n^*| b_n, \tag{6.6}$$

where

$$I_3 = \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau \sigma_0(Q_n(s;\tilde{\alpha}_n)) 1_{s \le \zeta_n} dN_{n,i}(s),$$
$$I_4 = \frac{1}{m_n} \sum_{i=1}^{m_n} \int_0^\tau \sigma_0(Q_n(s;\tilde{\beta}_n)) 1_{s > \zeta_n} dN_{n,i}(s).$$

We consider the quantities in (6.6) one by one. For I_1, we have

$$I_1 \quad \le \quad \Big| \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \int_0^\tau \left(Z_{n,i}(s) - \bar{z}_n(s;\alpha_n) \right) 1_{s \le \zeta_n} dN_{n,i}(s) \Big|$$

$$+\left|\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\int_0^\tau \left(\bar{z}_n(s;\alpha_n)-\bar{Z}_n(s;\alpha_n)\right)1_{s\le\zeta_n}dN_{n,i}(s)\right|. \qquad (6.7)$$

By the definition of α_n and ζ_n, the first term in (6.7) equals

$$\sqrt{m_n}\left|\int_0^{\zeta_n} dM_{n,1}(s)-\int_0^{\zeta_n}\bar{z}_n(s;\alpha_n)dM_{n,0}(s)\right|,$$

where $M_{n,k}, k=0,1$, is defined as in (6.2). Then similarly as in the derivation of (6.3), Lenglart's inequality implies that the above quantity is $O_{\mathbf{P}}(1)$. Consider the second term in (6.7). From the proof of Lemma 23, we know that $\{Y_{n,i}(t)Z_{n,i}^{\otimes k}e^{\gamma' Z_{n,i}(t)}\}$ is manageable. Then by the boundedness property of Z and inequality (7.10) in page 38 of Pollard (1990), there exists constant $B>0$ such that

$$\mathbf{E}\Big[\sup_{s\in[0,\tau]}\left|\bar{z}_n(s;\alpha_n)-\bar{Z}_n(s;\alpha_n)\right|\Big]\le\frac{B}{\sqrt{m_n}},$$

which implies that the second term in (6.7) is also $O_{\mathbf{P}}(1)$. Therefore,

$$I_1=O_{\mathbf{P}}(1).$$

Similarly, we obtain that $I_2=O_{\mathbf{P}}(1)$.

Consider I_3 and we have that

$$\begin{aligned}&\left|I_3-\int_0^{\zeta_n}\sigma_0(Q(s;\tilde{\alpha}_n))dA_{n,0}(s)\right|\\ \le\ &\left|\frac{1}{m_n}\sum_{i=1}^{m_n}\int_0^{\zeta_n}\left(\sigma_0(Q_n(s;\tilde{\alpha}_n))-\sigma_0(Q(s;\tilde{\alpha}_n))\right)dN_{n,i}(s)\right|\\ &+\sup_{\zeta\in[0,\tau]}\left|\int_0^{\zeta}\sigma_0(Q(s;\tilde{\alpha}_n))dM_{n,0}(s)\right|.\end{aligned}$$

The first term in the right hand side of the above display converges to 0 due to the convergence of $\sigma_0(Q_n)$. The second term also converges to 0 in probability by Lenglart's inequality. Thus, together with condition A1 and the convergence of θ_n to θ_0, we have

$$I_3=(1+o_{\mathbf{P}}(1))\int_0^{\zeta_0}\sigma_0(Q(s;\alpha_0))s_0(s;\alpha_0)d\Lambda_0(s).$$

Similarly, $I_4=(1+o_{\mathbf{P}}(1))\int_{\zeta_0}^{\tau}\sigma_0(Q(s;\beta_0))s_0(s;\beta_0)d\Lambda_0(s)$.

Consider a_n in (6.6) and we have that

$$\left|a_n-\frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\Big(\int_{\zeta_n}^{\zeta_n^*}(\alpha_n^*-\beta_n^*)'dA_{n,1}(s)-\int_{\zeta_n}^{\zeta_n^*}\log r_n(s;\alpha_n^*,\beta_n^*)dA_{n,0}(s)\Big)\right| \qquad (6.8)$$

$$\begin{aligned}
&\le \frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\Big|\int_{\zeta_n}^{\zeta_n^*}(\alpha_n^*-\beta_n^*)'dM_{n,1}(s)\Big| \\
&\quad+\frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\Big|\int_{\zeta_n}^{\zeta_n^*}\log r_n(s;\alpha_n^*,\beta_n^*)dM_{n,0}(s)\Big| \\
&\quad+\frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\sup_{s\in[\zeta_n,\zeta_n^*]}\big|\log R_n(s;\alpha_n^*,\beta_n^*)-\log r_n(s;\alpha_n^*,\beta_n^*)\big|\Big|\int_{\zeta_n}^{\zeta_n^*}dM_{n,0}(s)\Big| \\
&\quad+\frac{1_{\zeta_n<\zeta_n^*}}{\zeta_n^*-\zeta_n}\sup_{s\in[\zeta_n,\zeta_n^*]}\big|\log R_n(s;\alpha_n^*,\beta_n^*)-\log r_n(s;\alpha_n^*,\beta_n^*)\big|\int_{\zeta_n}^{\zeta_n^*}dA_{n,0}(s) \\
&=: a_{n,1}+a_{n,2}+a_{n,3}+a_{n,4}.
\end{aligned}$$

For $a_{n,k}$, $k=1,2,3$, by Lenglart's inequality and condition A2, we have that for any positive ϵ and $\epsilon_{n,j}, j\in\mathbb{Z}^+$, there exists a constant $B>0$ such that

$$\begin{aligned}
\mathbb{Q}_n\Big(\sup_{m_n|\zeta_n^*-\zeta_n|>h_n}a_{n,k}>\epsilon\Big) &\le \sum_{j=1}^{\infty}\mathbb{Q}_n\Big(\sup_{2^{j-1}h_n\le m_n|\zeta_n^*-\zeta_n|<2^jh_n}a_{n,k}^2>\epsilon^2\Big) \\
&\le \sum_{j=1}^{\infty}\frac{\epsilon_{n,j}}{\epsilon^2}+\frac{B\rho_2}{\epsilon_{n,j}2^jh_n},
\end{aligned}$$

where ρ_2 is defined as in condition A2. Since $\epsilon_{n,j}$ are arbitrary, it follows that $\sup_{m_n|\zeta_n^*-\zeta_n|>h_n} a_{n,k}\xrightarrow{\mathbf{P}}0$, $k=1,2,3$. In addition, for $a_{n,4}$, we have

$$\sup_{m_n|\zeta_n^*-\zeta_n|>h_n}a_{n,4}=o_{\mathbf{P}}(1)\sup_{m_n|\zeta_n^*-\zeta_n|>h_n}\frac{1_{\zeta_n<\zeta_n^*}}{\zeta_n^*-\zeta_n}\int_{\zeta_n}^{\zeta_n^*}dA_{n,0}(s)\xrightarrow{\mathbf{P}}0.$$

Thus (6.8) $\xrightarrow{\mathbf{P}}0$. From condition A2 and Theorem 7, we have

$$\begin{aligned}
&\sup_{m_n|\zeta_n^*-\zeta_n|>h_n}\frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\Big(\int_{\zeta_n}^{\zeta_n^*}(\alpha_n^*-\beta_n^*)'dA_{n,1}(s) \\
&\qquad-\int_{\zeta_n}^{\zeta_n^*}\log r_n(s;\alpha_n^*,\beta_n^*)dA_{n,0}(s)\Big) \\
&\quad=\sup_{m_n|\zeta_n^*-\zeta_n|>h_n}\frac{1_{\zeta_n<\zeta_n^*}}{|\zeta_n-\zeta_n^*|}\int_{\zeta_n}^{\zeta_n^*}\left((\alpha_0-\beta_0)'\bar{z}_n(s;\beta_0)\right. \\
&\qquad\left.-\log r_n(s;\alpha_0,\beta_0)\right)dA_{n,0}(s)+o_{\mathbf{P}}(1).
\end{aligned}$$

Since $(\alpha_0-\beta_0)'\bar{z}_n(s;\beta_0)-\log r_n(s;\alpha_0,\beta_0)<0$ and it is continuous in a neighborhood of ζ_0, there exists a constant $\kappa_0<0$ such that for any sequence $h_n\to\infty$ and $h_n/m_n\to 0$,

$$0<-\kappa_0\rho_1\le\inf_{m_n|\zeta_n-\zeta_n^*|>h_n}\{-a_n\}\le\sup_{m_n|\zeta_n-\zeta_n^*|>h_n}\{-a_n\}\le-\rho_2\kappa_0$$

holds with probability tending to 1 as $n \to \infty$. Similarly, we have

$$0 < -2\kappa_0\rho_1 \le \inf_{m_n|\zeta_n-\zeta_n^*|>h_n}\{-a_n-b_n\} \le \sup_{m_n|\zeta_n-\zeta_n^*|>h_n}\{-a_n-b_n\} \le -2\rho_2\kappa_0$$

holds with probability tending to 1 as $n \to \infty$.

Combining the above derivations for (6.6), we have that

$$\begin{aligned} -m_n|\zeta_n-\zeta_n^*|(a_n+b_n) &\le \sqrt{m_n}|\alpha_n^*-\alpha_n|O_{\mathbf{P}}(1)+\sqrt{m_n}|\beta_n^*-\beta_n|O_{\mathbf{P}}(1) \\ &\quad - O_{\mathbf{P}(1)}m_n(|\alpha_n^*-\alpha_n|^2+|\beta_n^*-\beta_n|^2) \\ &= O_{\mathbf{P}}(1). \end{aligned} \tag{6.9}$$

Thus $m_n(\zeta_n-\zeta_n^*)=O_{\mathbf{P}}(1)$. As a consequence, (6.9) implies that

$$\begin{aligned} &\sqrt{m_n}|\alpha_n^*-\alpha_n|O_{\mathbf{P}}(1)+\sqrt{m_n}|\beta_n^*-\beta_n|O_{\mathbf{P}}(1) \\ &\quad - O_{\mathbf{P}(1)}m_n(|\alpha_n^*-\alpha_n|^2+|\beta_n^*-\beta_n|^2) = O_{\mathbf{P}}(1). \end{aligned}$$

This gives that $|\sqrt{m_n}(\alpha_n^*-\alpha_n)|=O_{\mathbf{P}}(1)$ and $|\sqrt{m_n}(\beta_n^*-\beta_n)|=O_{\mathbf{P}}(1)$.

6.3. Proof of Theorem 10

Let

$$\begin{aligned} U_{n,1} &= \frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\int_0^{\zeta_n}(Z_{n,i}(s)-\bar{z}_n(s;\alpha_n))\,dN_{n,i}(s), \\ U_{n,2}(h_\zeta) &= \sum_{i=1}^{m_n}\int_{\zeta_n+h_\zeta/m_n}^{\zeta_n}((\beta_n-\alpha_n)'Z_{n,i}(s)-\log r_n(s;\beta_n,\alpha_n))\,1_{h_\zeta<0}dN_{n,i}(s), \\ U_{n,3}(h_\zeta) &= \sum_{i=1}^{m_n}\int_{\zeta_n+h_\zeta/m_n}^{\zeta_n}1_{h_\zeta<0}dN_{n,i}(s), \\ U_{n,4}(h_\zeta) &= \sum_{i=1}^{m_n}\int_{\zeta_n}^{\zeta_n+h_\zeta/m_n}((\alpha_n-\beta_n)'Z_{n,i}(s)-\log r_n(s;\alpha_n,\beta_n))\,1_{h_\zeta>0}dN_{n,i}(s), \\ U_{n,5}(h_\zeta) &= \sum_{i=1}^{m_n}\int_{\zeta_n}^{\zeta_n+h_\zeta/m_n}1_{h_\zeta>0}dN_{n,i}(s), \\ U_{n,6} &= \frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\int_{\zeta_n}^{\tau}(Z_{n,i}(s)-\bar{z}_n(s;\beta_n))\,dN_{n,i}(s). \end{aligned}$$

We define processes $J_n(h_\zeta):=U_{n,3}(h_\zeta)+U_{n,5}(h_\zeta)$ and

$$\begin{aligned} U_n(h) := {} & h_\alpha'U_{n,1}-\frac{1}{2}h_\alpha'\Big(\int_0^{\zeta_0}Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\Big)h_\alpha+U_{n,2}(h_\zeta) \\ & + h_\beta'U_{n,6}-\frac{1}{2}h_\beta'\Big(\int_{\zeta_0}^{\tau}Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\Big)h_\beta+U_{n,4}(h_\zeta). \end{aligned}$$

The limit law of U_n and J_n can be deduced from that of $U_{n,i}$ and is given as follows.

Lemma 21. *Let $K \subset \mathbb{R}$ be a compact interval and $\Theta = \tilde{\Theta} \times K \subset \mathbb{R}^{2p+1}$ a compact set. Then, under conditions A1–A3, (U_n, J_n) converges weakly in the Skorohod topology to (U, J) in $\mathcal{D}_\Theta \times \mathcal{D}_K$.*

Next we show that processes U_n and U_n^* have the same asymptotic distribution.

Lemma 22. *Let Θ be a compact set in $\mathbb{R}^{2p+1}$. Then under conditions A1 and A2,*

$$\sup_{h\in\Theta} |U_n(h) - U_n^*(h)| \xrightarrow{\mathbf{P}} 0.$$

Thus, (U_n^*, J_n^*) also converges weakly in the Skorohod topology to (U, J) in $\mathcal{D}_\Theta \times \mathcal{D}_K$. Then by Theorem 3.1 in Seijo and Sen (2011b), we have the desired conclusion.

Acknowledgements

We thank the Editor and a referee for their feedback and helpful comments. B. Sen's research is supported by a (CAREER) grant from the NSF (DMS-1150435). Z. Ying's research is partially supported by grants from the NIH (R37GM047845) and the NSF (DMS-1308566).

Appendix

This appendix contains proofs of Lemmas 9, 13, 21, 22 and 23 (stated below).

Proof of Lemma 9. From the definition of U it is easily seen that

$$\begin{aligned}
\phi_\alpha &= \left(\int_0^{\zeta_0} Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\right)^{-1} U_1, \\
\phi_\beta &= \left(\int_{\zeta_0}^{\tau} Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\right)^{-1} U_6, \\
\phi_\zeta &= \text{sargmax}_{h\in\mathbb{R}^{2p+1}}\{U_2(h_\zeta) + U_4(h_\zeta)\}.
\end{aligned}$$

Due to the independence of U_1, U_2, U_4 and U_6, ϕ_α, ϕ_β, and ϕ_ζ are independent. In addition, (3.3) and (3.5) hold.

We now show the existence of ϕ_ζ. It suffices to show that $U_2(h_\zeta) + U_4(h_\zeta) \to -\infty$ as $|h_\zeta| \to \infty$. For $h_\zeta > 0$,

$$\begin{aligned}
U_4(h_\zeta) = &-\Gamma^+(h_\zeta)\log r(\zeta_0;\alpha_0,\beta_0) + \sum_{1\le i\le\Gamma^+(h_\zeta)} (\alpha_0-\beta_0)v_i^+ \\
= &\sum_{1\le i\le\Gamma^+(h_\zeta)} \left\{(\alpha_0-\beta_0)v_i^+ - \mathbf{E}\left[(\alpha_0-\beta_0)v_i^+\right]\right\} \\
&+ \mathbf{E}\left[(\alpha_0-\beta_0)v_i^+ - \log r(\zeta_0;\alpha_0,\beta_0)\right]\Gamma^+(h_\zeta).
\end{aligned}$$

Since $\Gamma^+(h_\zeta) \to \infty$ as $h_\zeta \to \infty$ and

$$\mathbf{E}\left[(\alpha_0 - \beta_0)v_i^+ - \log r(\zeta_0; \alpha_0, \beta_0)\right] = (\alpha_0 - \beta_0)' z(\zeta_0; \beta_0) - \log r(\zeta_0; \alpha_0, \beta_0) < 0,$$

$U_4(h_\zeta) \to -\infty$ as $h_\zeta \to \infty$. A similar argument gives $U_2(h_\zeta) \to -\infty$ as $h_\zeta \to -\infty$, which completes the proof. □

Proof of Lemma 13. For (4.3), we only need to show that the first sequence does not converge in probability. Take $\epsilon < 1/4$. From Theorem 8, there exists a constant $B_\epsilon > 0$ such that $P(n|\hat{\zeta}_n - \zeta_0| \le B_\epsilon) > 1 - \epsilon$ for all large n. Choose $h > 2B_\epsilon$ and let

$$\hat{E}_n = \sum_{i=1}^n \int_{\hat{\zeta}_n}^{\hat{\zeta}_n + \frac{h}{n}} dN_i(s), \quad E_{n,1} = \sum_{i=1}^n \int_{\zeta_0 + \frac{B_\epsilon}{n}}^{\zeta_0 + \frac{h - B_\epsilon}{n}} dN_i(s),$$
$$E_{n,2} = \sum_{i=1}^n \int_{\zeta_0 - \frac{B_\epsilon}{n}}^{\zeta_0 + \frac{h + B_\epsilon}{n}} dN_i(s).$$

Then,

$$P(E_{n,1} \le \hat{E}_n \le E_{n,2}) \ge P(n|\hat{\zeta}_n - \zeta_0| \le B_\epsilon) > 1 - \epsilon. \tag{A.10}$$

We know that for any $h_1 < h_2$,

$$\sum_{i=1}^n \int_{\zeta_0 + \frac{h_1}{n}}^{\zeta_0 + \frac{h_2}{n}} dN_i(s) \rightsquigarrow \text{Poisson}\big(\lambda_0(\zeta_0)[(h_2 - h_1 \vee 0)s_0(\zeta_0; \beta_0) - (h_1 \wedge 0)s_0(\zeta_0; \alpha_0)]\big).$$

Therefore, $E_{n,1} \rightsquigarrow \text{Poisson}(\lambda_0(\zeta_0)(h - 2B_\epsilon)s_0(\zeta_0; \beta_0))$ and $E_{n,2} \rightsquigarrow \text{Poisson}(\lambda_0(\zeta_0)(h + B_\epsilon)s_0(\zeta_0; \beta_0) + \lambda_0(\zeta_0)B_\epsilon s_0(\zeta_0; \alpha_0))$. Then by Lemma A.4 in Seijo and Sen (2011a), there is a constant h_0 such that when $h > h_0$, we can find two numbers $N_{1,h} < N_{2,h} \in \mathbb{N}$ satisfying

$$\liminf_{n\to\infty} \mathbf{P}(E_{n,1} > N_{2,h}) > 2\epsilon \text{ and } \liminf_{n\to\infty} \mathbf{P}(E_{n,2} < N_{1,h}) > 2\epsilon.$$

Combining with (A.10), we have

$$\mathbf{P}(\hat{E}_n \ge E_{n,1} > N_{2,h},\ i.o.) > \epsilon \text{ and } \mathbf{P}(\hat{E}_n \le E_{n,2} < N_{1,h},\ i.o.) > \epsilon.$$

Then by the Hewitt-Savage 0-1 law, the permutation invariant events $\{\hat{E}_n > N_{2,h},\ i.o.\}$ and $\{\hat{E}_n < N_{1,h},\ i.o.\}$ occur with probability 1, which implies that $\hat{E}_n$ does not have an almost sure limit. A similar argument applies for any increasing sequence of natural numbers $\{n_k\}_{k=1}^\infty$ and gives that $\hat{E}_n$ does not converge in probability.

For (4.4), consider the real part of ϕ_i, $Re(\phi_i)$, and define

$$\hat{E}_n^\phi = \sum_{i=1}^n \int_{\hat{\zeta}_n}^{\hat{\zeta}_n + \frac{h}{n}} Re(\phi_i(s)) dN_i(s), \quad E_{n,1}^\phi = \sum_{i=1}^n \int_{\zeta_0 + \frac{B_\epsilon}{n}}^{\zeta_0 + \frac{h - B_\epsilon}{n}} Re(\phi_i(s)) dN_i(s),$$

$$E^{\phi}_{n,2} = \sum_{i=1}^{n} \int_{\zeta_0 - \frac{B_\epsilon}{n}}^{\zeta_0 + \frac{h+B_\epsilon}{n}} Re(\phi_i(s)) dN_i(s).$$

It is sufficient to show that $\hat{E}^{\phi}_n$ does not converge. Since for any $h_1 < h_2$, $\sum_{i=1}^{n} \int_{\zeta_0+\frac{h_1}{n}}^{\zeta_0+\frac{h_2}{n}} Re(\phi_i(s)) dN_i(s)$ converges to a compound Poisson distribution, then a similar argument as above gives the desired conclusion. □

Proof of Lemma 21. It is sufficient to show the weak convergence in probability of $(U_{n,1}, \ldots, U_{n,6})$ to $(U_1, \ldots, U_6)$. We first prove the convergence of its finite dimensional joint characteristic function. Consider real numbers $h_{-N} < \cdots < h_{-1} < 0 = h_0 < h_1 < \cdots < h_N$ and the linear combination

$$\begin{aligned} W_n = {} & \mu U_{n,1} + v U_{n,6} \\ & + \sum_{-N \le j \le -1} \{ q_j (U_{n,2}(h_j) - U_{n,2}(h_{j+1})) + p_j (U_{n,3}(h_j) - U_{n,3}(h_{j+1})) \} \\ & + \sum_{1 \le j \le N} \{ q_j (U_{n,4}(h_j) - U_{n,4}(h_{j-1})) + p_j (U_{n,5}(h_j) - U_{n,5}(h_{j+1})) \}, \end{aligned}$$

where $p_j, q_j \in \mathbb{R}, j = -N, \ldots, N$, and $\mu, v \in \mathbb{R}^{1 \times p}$. For simplicity, we write $h_{j,n} = h_j / m_n$.

The characteristic function of W_n is $\mathbf{E}[e^{\imath t W_n}]$ and can be expressed as

$$\begin{aligned} & \mathbf{E}\Bigg[\exp\Bigg(\imath t \mu \frac{1}{\sqrt{m_n}} \sum_{k=1}^{m_n} \int_0^{\zeta_n} \big(Z_{n,k}(s) - \bar{z}_n(s;\alpha_n) \big) \, dN_{n,k}(s) \\ & + \imath t v \int_{\zeta_n}^{\tau} \big(Z_{n,k}(s) - \bar{z}_n(s;\beta_n) \big) \, dN_{n,k}(s) \\ & + \imath t \sum_{j=-N}^{-1} \sum_{k=1}^{m_n} \int_{\zeta_n + h_{j,n}}^{\zeta_n + h_{j+1,n}} \big(q_j (\beta_n - \alpha_n)' Z_{n,k}(s) - q_j \log r_n(s;\beta_n,\alpha_n) + p_j \big) \, dN_{n,k}(s) \\ & + \imath t \sum_{j=1}^{N} \sum_{k=1}^{m_n} \int_{\zeta_n + h_{j-1,n}}^{\zeta_n + h_{j,n}} \big(q_j (\alpha_n - \beta_n)' Z_{n,k}(s) - q_j \log r_n(s;\alpha_n,\beta_n) + p_j \big) \, dN_{n,k}(s) \Bigg) \Bigg]. \end{aligned}$$

By the independence of the observations $\{(\tilde{T}_{n,k}, \delta_{n,k}, Z_{n,k}) : k = 1, \ldots, m_n\}$, $\mathbf{E}[e^{\imath t W_n}]$ can be further written as

$$\begin{aligned} & \prod_{k=1}^{m_n} \mathbb{Q}_n \Bigg\{ 1 + \int_0^{\tau} \Big[e^{\imath t \frac{1}{\sqrt{m_n}} \mu \left(Z_{n,k}(s) - \bar{z}_n(s;\alpha_n) \right) 1_{s<\zeta_n}} - 1 \Big] dN_{n,k}(s) \\ & + \int_0^{\tau} \Big[e^{\imath t v \frac{1}{\sqrt{m_n}} \left(Z_{n,k}(s) - \bar{z}_n(s;\beta_n) \right) 1_{s>\zeta_n}} - 1 \Big] dN_{n,k}(s) \\ & + \sum_{-N \le j \le -1} \int_{\zeta_n + h_{j,n}}^{\zeta_n + h_{j+1,n}} \Big[e^{\imath t \left(q_j (\beta_n - \alpha_n)' Z_{n,k}(s) - q_j \log r_n(s;\beta_n,\alpha_n) + p_j \right)} - 1 \Big] dN_{n,k}(s) \\ & + \sum_{1 \le j \le N} \int_{\zeta_n + h_{j-1,n}}^{\zeta_n + h_{j,n}} \Big[e^{\imath t \left(q_j (\alpha_n - \beta_n)' Z_{n,k}(s) - q_j \log r_n(s;\alpha_n,\beta_n) + p_j \right)} - 1 \Big] dN_{n,k}(s) \Bigg\}. \end{aligned}$$

For the first two integrals, take Taylor's expansions of the exponential functions and we have that $\mathbf{E}[e^{\imath t W_n}]$ equals

$$\prod_{k=1}^{m_n}\Bigg\{1+\mathbb{Q}_n\left(\int_0^\tau \frac{1}{\sqrt{m_n}}\imath t\mu\Big(Z_{n,k}(s)-\bar{z}_n(s;\alpha_n)\Big)1_{s<\zeta_n}dN_{n,k}(s)\right)$$
$$+\mathbb{Q}_n\left(\int_0^\tau \frac{1}{\sqrt{m_n}}\imath tv\Big(Z_{n,k}(s)-\bar{z}_n(s;\beta_n)\Big)1_{s>\zeta_n}dN_{n,k}(s)\right)$$
$$-\frac{1}{2m_n}\mathbb{Q}_n\left(\int_0^\tau t^2\mu\Big(Z_{n,k}(s)-\bar{z}_n(s;\alpha_n)\Big)^{\otimes 2}\mu'\ 1_{s<\zeta_n}dN_{n,k}(s)\right)$$
$$-\frac{1}{2m_n}\mathbb{Q}_n\left(\int_0^\tau t^2v\Big(Z_{n,k}(s)-\bar{z}_n(s;\beta_n)\Big)^{\otimes 2}v'\ 1_{s>\zeta_n}dN_{n,k}(s)\right)+o(m_n^{-1})$$
$$+\sum_{-N\le j\le -1}\mathbb{Q}_n\left(\int_{\zeta_n+h_{j,n}}^{\zeta_n+h_{j+1,n}}\Big[e^{\imath t\left(q_j(\beta_n-\alpha_n)'Z_{n,k}(s)-q_j\log r_n(s;\beta_n,\alpha_n)+p_j\right)}-1\Big]dN_{n,k}(s)\right)$$
$$+\sum_{1\le j\le N}\mathbb{Q}_n\left(\int_{\zeta_n+h_{j-1,n}}^{\zeta_n+h_{j,n}}\Big[e^{\imath t\left(q_j(\alpha_n-\beta_n)'Z_{n,k}(s)-q_j\log r_n(s;\alpha_n,\beta_n)+p_j\right)}-1\Big]dN_{n,k}(s)\right)\Bigg\}.$$

By condition A2, $\mathbb{Q}_n(\sum_{i=1}^{m_n}\int_0^{\zeta_n}(Z_{n,k}(s)-\bar{z}_n(s;\alpha_n))dN_{n,k}(s))$ and $\mathbb{Q}_n(\sum_{i=1}^{m_n}\int_{\zeta_n}^{\tau}(Z_{n,k}(s)-\bar{z}_n(s;\beta_n))dN_{n,k}(s))$ converge to 0. Thus $\mathbf{E}[e^{\imath t W_n}]$ equals

$$(1+o(1))\exp\Bigg\{-\frac{1}{2}\sum_{k=1}^{m_n}\mathbb{Q}_n\left(\int_0^{\zeta_n}t^2\mu\big(Z_{n,k}(s)-\bar{z}_n(s;\alpha_n)\big)^{\otimes 2}\mu'\ dN_{n,k}(s)\right)$$
$$-\frac{1}{2}\sum_{k=1}^{m_n}\mathbb{Q}_n\left(\int_{\zeta_n}^{\tau}t^2v\big(Z_{n,k}(s)-\bar{z}_n(s;\beta_n)\big)^{\otimes 2}v\ dN_{n,k}(s)\right)\Bigg\}$$
$$\times\exp\Bigg\{\sum_{j=-N}^{-1}\mathbb{Q}_n\Bigg(\sum_{k=1}^{m_n}\int_{\zeta_n+h_{j,n}}^{\zeta_n+h_{j+1,n}}\Big[e^{\imath t\left(q_j(\beta_n-\alpha_n)'Z_{n,k}(s)-q_j\log r_n(s;\beta_n,\alpha_n)+p_j\right)}$$
$$-1\Big]dN_{n,k}(s)\Bigg)\Bigg\}$$
$$\times\exp\Bigg\{\sum_{j=1}^{N}\mathbb{Q}_n\Bigg(\sum_{k=1}^{m_n}\int_{\zeta_n+h_{j-1,n}}^{\zeta_n+h_{j,n}}\Big[e^{\imath t\left(q_j(\alpha_n-\beta_n)'Z_{n,k}(s)-q_j\log r_n(s;\alpha_n,\beta_n)+p_j\right)}$$
$$-1\Big]dN_{n,k}(s)\Bigg)\Bigg\}.$$

It is easily seen that the first exponential component in the above display converges to $\mathbf{E}[e^{\imath t\mu U_1+\imath tvU_6}]$. Therefore, Lemma 23 together with condition A3 implies that

$$\mathbf{E}[e^{\imath t W_n}]=(1+o(1))\mathbf{E}[e^{\imath t\mu U_1+\imath tvU_6}]$$
$$\times\exp\Bigg\{\lambda_0(\zeta_0)\sum_{-N\le j\le -1}(h_{j+1}-h_j)$$
$$\times\Big[e^{-q_j\log r(\zeta_0;\beta_0,\alpha_0)+p_j}\ s_0(\zeta_0;\imath tq_j(\beta_0-\alpha_0)+\alpha_0)-s_0(\zeta_0;\alpha_0)\Big]\Bigg\}$$

$$\times \exp\Bigg\{\lambda_0(\zeta_0) \sum_{1\le j\le N} (h_j - h_{j-1}) \\ \times \Big[e^{-q_j \log r(\zeta_0;\alpha_0,\beta_0)+p_j}\ s_0(\zeta_0; itq_j(\alpha_0-\beta_0)+\beta_0) - s_0(\zeta_0;\beta_0)\Big]\Bigg\}.$$

For $(U_1,\dots,U_6)$ as defined in Section 3.2, define the linear combination

$$W := \mu U_1 + vU_6 + \sum_{-N\le j\le -1} \big\{q_j(U_{n,2}(h_j) - U_2(h_{j+1})) + p_j(U_3(h_j) - U_{n,3}(h_{j+1}))\big\} \\ + \sum_{1\le j\le N} \big\{q_j(U_{n,4}(h_j) - U_4(h_{j-1})) + p_j(U_5(h_j) - U_{n,5}(h_{j+1}))\big\}.$$

By the definition of $(U_1,\dots,U_6)$, we know that the characteristic function of W has the same form as the limit of $\mathbf{E}[e^{\imath tW_n}]$. Thus, we have the weak convergence of the finite dimensional distributions.

To further prove the weak convergence of $(U_1,\dots,U_6)$, we use Theorem 15.6 in Billingsley (1968). It's sufficient to show for each $U_{n,i}, i = 2,3,4,5$, there exists a nondecreasing, continuous function F such that for any $h_1 < h < h_2$,

$$\mathbf{E}|U_{n,i}(h_1) - U_{n,i}(h)||U_{n,i}(h_2) - U_{n,i}(h)| \le (F(h_2) - F(h_1))^2. \tag{A.11}$$

Consider $U_{n.2}$. For $h_1 < h < h_2 < 0$,

$$\begin{aligned} & \mathbf{E}|U_{n,2}(h_1) - U_{n,2}(h)||U_{n,2}(h_2) - U_{n,2}(h)| \\ \le\ & \mathbf{E}\sum_{i=1}^{m_n} \int_{\zeta_n+h_1/m_n}^{\zeta_n+h/m_n} |(\beta_n-\alpha_n)'Z_{n,i}(s) - \log r_n(s;\beta_n,\alpha_n)|\, dN_{n,i}(s) \\ & \times \sum_{i=1}^{m_n} \int_{\zeta_n+h/m_n}^{\zeta_n+h_2/m_n} |(\beta_n-\alpha_n)'Z_{n,i}(s) - \log r_n(s;\beta_n,\alpha_n)|\, dN_{n,i}(s) \\ \le\ & \sup_{\substack{1\le i\le m_n \\ s\in[\zeta_n+\frac{h_1}{m_n},\zeta_n+\frac{h_2}{m_n}]}} |(\beta_n-\alpha_n)'Z_{n,i}(s) - \log r_n(s;\beta_n,\alpha_n)| \\ & \times m_n^2 \mathbf{E}\Big[\int_{\zeta_n+h_1/m_n}^{\zeta_n+h/m_n} dN_{n,i}(s)\Big]\mathbf{E}\Big[\int_{\zeta_n+h/m_n}^{\zeta_n+h_2/m_n} dN_{n,i}(s)\Big] \\ \le\ & B|h_2-h_1|^2, \end{aligned}$$

where $B > 0$ is some constant. Thus, (A.11) holds for $U_{n,2}$. Similar arguments give that (A.11) is satisfied for $U_{n,i}, i = 3,4,5$. Then our conclusion follows from Theorem 15.6 in Billingsley (1968). □

Proof of Lemma 22. For notational simplicity, we write

$$h_{\alpha,n} = \frac{h_\alpha}{\sqrt{m_n}}, h_{\beta,n} = \frac{h_\beta}{\sqrt{m_n}}, h_{\zeta,n} = \frac{h_\zeta}{m_n}.$$

We start by writing U_n^* as follows:

$$U_n^*(h) := u_{n,1}(h) + u_{n,2}(h) + u_{n,3}(h) + u_{n,4}(h)$$

where

$$u_{n,1}(h) = \sum_{i=1}^{m_n} \int_0^{\tau} \big(h'_{\alpha,n} Z_{n,i}(s) - \log R_n(s;\alpha_n + h_{\alpha,n},\alpha_n)\big) 1_{s \le \zeta_n \wedge (\zeta_n + h_{\zeta,n})} dN_{n,i}(s),$$

$$u_{n,2}(h) = \sum_{i=1}^{m_n} \int_0^{\tau} \big((\alpha_n - \beta_n + h_{\alpha,n})' Z_{n,i}(s) \\ - \log R_n(s;\alpha_n + h_{\alpha,n},\beta_n)\big) 1_{\zeta_n < s \le \zeta_n + h_{\zeta,n}} dN_{n,i}(s),$$

$$u_{n,3}(h) = \sum_{i=1}^{m_n} \int_0^{\tau} \big((\beta_n - \alpha_n + h_{\beta,n})' Z_{n,i}(s) \\ - \log R_n(s;\beta_n + h_{\beta,n},\alpha_n)\big) 1_{\zeta_n + h_{\zeta,n} < s \le \zeta_n} dN_{n,i}(s),$$

$$u_{n,4}(h) = \sum_{i=1}^{m_n} \int_0^{\tau} \big(h'_{\beta,n} Z_{n,i}(s) - \log R_n(s;\beta_n + h_{\beta,n},\beta_n)\big) 1_{s > \zeta_n \vee (\zeta_n + h_{\zeta,n})} dN_{n,i}(s).$$

For $h = (h'_\alpha, h'_\beta, h_\zeta)' \in \Theta$, consider the difference between $u_{n,1}$ and the first two terms in U_n:

$$\begin{aligned} &\Big|u_{n,1}(h) - h'_\alpha U_{n,1} + \frac{1}{2} h'_\alpha \Big(\int_0^{\zeta_0} Q(s;\alpha_0) s_0(s;\alpha_0)\lambda_0(s) ds\Big) h_\alpha\Big| \\ \le& \Big|\sum_{i=1}^{m_n} \int_0^{\tau} \big(h'_{\alpha,n} Z_{n,i}(s) - \log R_n(s;\alpha_n + h_{\alpha,n},\alpha_n)\big) 1_{\zeta_n \wedge (\zeta_n + h_{\zeta,n}) < s \le \zeta_n} dN_{n,i}(s)\Big| \\ &+ \Big|\sum_{i=1}^{m_n} \int_0^{\zeta_n} \big(h'_{\alpha,n} Z_{n,i}(s) - \log R_n(s;\alpha_n + h_{\alpha,n},\alpha_n)\big) dN_{n,i}(s) - h'_\alpha U_{n,1} \\ &\quad + \frac{1}{2} h'_\alpha \Big(\int_0^{\zeta_0} Q(s;\alpha_0) s_0(s;\alpha_0)\lambda_0(s) ds\Big) h_\alpha\Big|. \end{aligned} \tag{A.12}$$

It is easily seen that

$$\begin{aligned} &\sup_{h\in\Theta}\Big|\sum_{i=1}^{m_n} \int_0^{\tau} \big(h'_{\alpha,n} Z_{n,i}(s) - \log R_n(s;\alpha_n + h_{\alpha,n},\alpha_n)\big)\, 1_{\zeta_n \wedge (\zeta_n + h_{\zeta,n}) < s \le \zeta_n} dN_{n,i}(s)\Big| \\ \le& \frac{B_1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \int_0^{\tau} 1_{\zeta_n - \frac{B_2}{m_n} < s \le \zeta_n} dN_{n,i}(s) \xrightarrow{\mathbf{P}} 0, \end{aligned} \tag{A.13}$$

where B_1 and B_2 are some constants. On the other hand, by Taylor's expansion, we have that

$$\begin{aligned} &\sum_{i=1}^{m_n} \int_0^{\zeta_n} \big(h'_{\alpha,n} Z_{n,i}(s) - \log R_n(s;\alpha_n + h_{\alpha,n},\alpha_n)\big)\, dN_{n,i}(s) \\ =\;& h'_\alpha U_{n,1} - \frac{1}{2} h'_\alpha \Big(\frac{1}{n}\sum_{i=1}^{m_n} \int_0^{\zeta_n} Q_n(s;\alpha_n) dN_i(s)\Big) h_\alpha + o(1). \end{aligned}$$

Then by condition A1 and the uniform convergence of Q_n, the second term of (A.12) converges to 0 uniformly in probability. Thus,

$$\sup_{h\in\Theta}\Big|u_{n,1}(h)-h'_\alpha U_{n,1}+\frac{1}{2}h'_\alpha\Big(\int_0^{\zeta_0}Q(s;\alpha_0)s_0(s;\alpha_0)\lambda_0(s)ds\Big)h_\alpha\Big|\xrightarrow{\mathbf{P}}0.$$

A similar argument gives that $|u_{n,2}-U_{n,4}|\xrightarrow{\mathbf{P}}0$, $|u_{n,3}-U_{n,2}|\xrightarrow{\mathbf{P}}0$, and

$$\sup_{h\in\Theta}\Big|u_{n,4}(h)-h'_\beta U_{n,6}-\frac{1}{2}h'_\beta\Big(\int_{\zeta_0}^{\tau}Q(s;\beta_0)s_0(s;\beta_0)\lambda_0(s)ds\Big)h_\beta\Big|\xrightarrow{\mathbf{P}}0.$$

This completes our proof. □

Lemma 23. *Under condition A1, for* $k=1,2,$ *and* 3*, as* $n\to\infty$*,*

$$\sup_{\substack{t\in[0,\tau]\\ \gamma\in\Theta_\alpha\cup\Theta_\beta}}|S_{n,k}(t;\gamma)-s_{n,k}(t;\gamma)|\to 0 \text{ and } \sup_{\substack{t\in[0,\tau]\\ \gamma\in\Theta_\alpha\cup\Theta_\beta}}|S_{n,k}(t;\gamma)-s_k(t;\gamma)|\to 0.$$

Proof of Lemma 23. We only need to show the first convergence result. The second result then follows from condition A1. We start with the case $k=0$. In view of Theorem 8.3 of Pollard (1990), it suffices to show that $\{Y_{n,i}(t)e^{\gamma' Z_{n,i}(t)}\}$ is manageable. Since the total variation of $Z_{n,i}$ is bounded, we can write $e^{\gamma' Z_{n,i}(t)}=e^{\gamma' Z^+_{n,i}(t)-\gamma' Z^-_{n,i}(t)}$, where component-wise, $Z^+_{n,i}(t)$ and $Z^-_{n,i}(t)$ are non-negative, non-increasing, and bounded by some constant B. By Lemma A.2 in Bilias, Gu and Ying (1997), we have the manageability of $\{Y_{n,i}(t)\}$, $\{Z^+_{n,i}(t)\}$ and $\{Z^-_{n,i}(t)\}$. Then, (5.2) in Pollard (1990) implies $\{\gamma' Z_{n,i}(t)\}$ is manageable, and further $\{Y_{n,i}(t)e^{\gamma' Z_{n,i}(t)}\}$ is manageable. Thus,

$$\sup_{t\in[0,\tau],\gamma\in\Theta_\alpha\cup\Theta_\beta}|S_{n,0}(t;\gamma)-s_{n,0}(t;\gamma)|\to 0.$$

A similar argument yields the result for $k=1$ and 2. □

References

Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, **73**, 1175–1204. MR2149245

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, **10**, 1100–1120. MR0673646

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Unpublished technical report, University of California, Berkeley.*

Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica*, **7**, 1–31. MR1441142

Bilias, Y., Gu, M. and Ying, Z. (1997). A general asymptotic theory for Cox model with staggered entry. *Ann. Statist.*, **25**, 662–682. MR1439318

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York. MR0233396

Bose, A. and Chatterjee, S. (2001). Generalised bootstrap in non-regular M-estimation problems. *Statist. Probab. Lett.*, **55**, 319–328. MR1867535

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *J. Amer. Statist. Assoc.*, **89**, 1290–1302. MR1310223

Cheng, G. and Huang, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.*, **38**, 2884–2915. MR2722459

Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc. B*, **34**, 187–220. MR0341758

Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276. MR0400509

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London. MR0751780

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge. MR1478673

Delsol, L. and Van Keilegom, I. (2011). Semiparametric M-estimation with non-smooth criterion functions. *Technical Report*.

Dudley, R. M. (2002). *Real Analysis and Probability*, vol. 74. Cambridge University Press, Cambridge. MR1932358

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York. MR1270903

Fleming, T. R. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons Inc., New York. MR1100924

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* (to appear).

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. arXiv:1107.3806.

Jacod, J. and Shiryaev, A. (2002). *Limit Theorems for Stochastic Processes*. Springer, New York. MR0959133

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York. MR1924807

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York. URL `http://dx.doi.org/10.1007/978-0-387-74978-5`. MR2724368

Kosorok, M. R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.*, **35**, 957–989. MR2341694

Lan, Y., Banerjee, M. and Michailidis, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.*, **37**, 1752–1791. MR2533471

Liang, K.-Y., Self, S. and Liu, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics*, **46**, 783–793.

Luo, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *J. Statist. Plann. Inference*, **53**, 33–61. MR1405736

Luo, X., Turnbull, B. and Clark, L. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika*, **84**, 555–565. MR1603981

Meinert, C. (1986). *Clinical Trials: Design, Conduct, and Analysis*. Oxford University Press.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, **42**, 1285–1295. MR0293706

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, Hayward, CA. MR1089429

Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time. *Statistics*, **36**, 101–124. MR1910255

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York. MR1191168

Seijo, E. and Sen, B. (2011a). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.*, **39**, 1580–1607. MR2850213

Seijo, E. and Sen, B. (2011b). A continuous mapping theorem for the smallest argmax functional. *Electron. J. Stat.*, **5**, 421–439. MR2802050

Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: The grenander estimator. *Ann. Statist.*, **38**, 1953–1977. MR2676880

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

Wells, M. T. (1994). Nonparametric kernel estimation in counting processes with explanatory variables. *Biometrika*, **81**, 795–801. MR1326428

Zucker, D. and Lakatos, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika*, **77**, 853–864. MR1086695