Web-based Supplementary Materials for "Outcome-Dependent Sampling with Interval-Censored Failure Time Data"

Qingning Zhou, Jianwen Cai and Haibo Zhou*
Department of Biostatistics, University of North Carolina at Chapel Hill
*email: zhou@bios.unc.edu

The Web-based Supplementary Materials include the Web Appendix referenced in Section 3.3 and Web Table 1 referenced in Section 4.

Web Appendix: Proofs of Theorems 1 and 2

In this appendix, we will sketch the proofs of Theorems 1 and 2 given in Section 3.3. Denote a single observation by $O = \{U, V, \Delta_1 = I(T \leq U), \Delta_2 = I(U < T \leq V), Z\}$, where U and V are two random examination times and Z is the p-dimensional vector of covariates. The following regularity conditions are needed for proving the theorems:

Condition (C1). There exists $\eta > 0$ such that $P(V - U \geq \eta) = 1$. The union of the supports of U and V is contained in the interval $[\sigma, \tau]$, where $0 < \sigma < \tau < +\infty$.

Condition (C2). The distribution of Z has a bounded support and is not concentrated on any proper subspace of R^p. Also $E\{var(Z|U)\}$ and $E\{var(Z|V)\}$ are positive definite.

Condition (C3). For $r = 1$ or 2, the function $\Lambda_0 \in \mathcal{M}$ is continuously differentiable up to order r in $[\sigma, \tau]$ with the first derivative being strictly positive, and satisfies $\alpha^{-1} < \Lambda_0(\sigma) < \Lambda_0(\tau) < \alpha$ for some positive constant α. Also $\xi_0' = (\beta_0', \pi_{10}, \pi_{20})$ is an interior point of $\mathcal{B}$, a compact subset of R^{p+2}. Here $\mathcal{B}$ and $\mathcal{M}$ are defined in Section 3.2.

Condition (C4). The conditional density $g(u, v|z)$ of (u, v) given z has bounded partial derivatives with respect to u and v, and the bounds of these partial derivatives do not depend on (u, v, z).

These conditions are commonly used in the studies of interval-censored data (Huang and Rossini, 1997; Zhang *et al.*, 2010). In the following, we will prove Theorems 1 and 2 under these conditions by employing the empirical process theory and some nonparametric methods or techniques. Note that under the proposed design, the samples I_0, I_1 and I_2 are independent and each sample consists of iid observations.

Before proving the theorems, we first establish two lemmas. Consider the class of functions $\mathcal{L}_n = \{l(\theta, O) : \theta \in \Theta_n\}$, where $l(\theta, O)$ is the log-likelihood function based on a single observation O given by

$$\begin{aligned} l(\theta, O) &= \Delta_1 \log(1 - S(U|Z)) + \Delta_2 \log(S(U|Z) - S(V|Z)) + (1 - \Delta_1 - \Delta_2) \log S(V|Z) \\ &\quad - \log \left\{ n_0 \left(1 + \sum_{k=1}^{2} \frac{n_k}{n_0 \pi_k} G_k(U, V, Z; \beta, \Lambda) \right) \right\} - \sum_{k=1}^{2} \frac{n_k}{n} \log \pi_k, \end{aligned}$$

with $S(t|z) = \exp\{-\Lambda(t) e^{\beta' z}\}$ being the survival function of T given $Z = z$ and G_k, $k = 1, 2$, given by $G_1(u, v, z; \beta, \Lambda) = I(u < a_1)(1 - S(u|z)) + I(v < a_1)(S(u|z) - S(v|z))$ and $G_2(u, v, z; \beta, \Lambda) = I(u > a_2)(S(u|z) - S(v|z))$. Let P_n denote the empirical measure. For any $\epsilon > 0$, define the covering number $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ as the smallest value of κ for which there exists $\{\theta^{(1)}, \ldots, \theta^{(\kappa)}\}$ such that

$$\min_{j \in \{1, \cdots, \kappa\}} \frac{1}{n} \sum_{i=1}^{n} \left| l(\theta, O_i) - l(\theta^{(j)}, O_i) \right| < \epsilon$$

for all $\theta \in \Theta_n$, where $\{O_1, \cdots, O_n\}$ represent the observed data under the ODS design and for $j = 1, \ldots, \kappa$, $\theta^{(j)} = (\xi^{(j)}, \Lambda^{(j)}) \in \Theta_n$. If no such κ exists, define $N(\epsilon, \mathcal{L}_n, L_1(P_n)) = \infty$.

Lemma 1 (Calculation of the covering number). Assume that Conditions (C1) - (C4) hold. Then the covering number of the class $\mathcal{L}_n = \{l(\theta, O) : \theta \in \Theta_n\}$ satisfies

$$N\Big(\epsilon, \mathcal{L}_n, L_1(P_n)\Big) \leq K M_n^{(m+1)} \epsilon^{-(p+m+3)}$$

for some constant K, where $m = o(n^\nu)$ with $\nu \in (0, 1)$ is the degree of Bernstein polynomials and $M_n = O(n^a)$ with $a > 0$ controls the size of the sieve space Θ_n.

Proof of Lemma 1

To investigate the covering number, first note that for any $\theta^1 = (\xi^1, \Lambda^1)$, $\theta^2 = (\xi^2, \Lambda^2) \in \Theta_n$, one can easily obtain that under Conditions (C1) - (C4),

$$|l(\theta^1, O) - l(\theta^2, O)| \leq K^* (\|\xi^1 - \xi^2\| + \|\Lambda^1 - \Lambda^2\|_\infty)$$

for some constant K^*, where $\|f\|_\infty = \sup_t |f(t)|$ for a function f.

Denote $\phi^j = (\phi_0^j, \cdots, \phi_m^j)'$ the Bernstein coefficients corresponding to Λ^j, $j = 1, 2$. Then it is easy to show that

$$\|\Lambda^1 - \Lambda^2\|_\infty = \sup_t \left| \sum_{k=0}^m \phi_k^1 B_k(t, m, \sigma, \tau) - \sum_{k=0}^m \phi_k^2 B_k(t, m, \sigma, \tau) \right|$$
$$\leq \max_{0 \leq k \leq m} |\phi_k^1 - \phi_k^2| \equiv \|\phi^1 - \phi^2\|_\infty.$$

Combining these results, we obtain

$$|l(\theta^1, O) - l(\theta^2, O)| \leq K^* \|\xi^1 - \xi^2\| + K^* \|\phi^1 - \phi^2\|_\infty .$$

It thus follows that for any $\theta \in \Theta_n$,

$$\frac{1}{n} \sum_{i=1}^n \left| l(\theta, O_i) - l(\theta^{(j)}, O_i) \right| \leq K^* \|\xi - \xi^{(j)}\| + K^* \|\phi - \phi^{(j)}\|_\infty.$$

By Lemma 2.5 of van de Geer (2000), one can show that $\{\xi \in R^{p+2}, \|\xi\| \leq M\}$ is covered by $(5M/(\epsilon/(2K^*)))^{p+2}$ balls with radius $\epsilon/(2K^*)$ and $\{\phi \in R^{m+1}, \sum_{0 \leq k \leq m} |\phi_k| \leq M_n\}$ is covered by $(5M_n/(\epsilon/(2K^*)))^{m+1}$ balls with radius $\epsilon/(2K^*)$. Therefore, the covering number of $\mathcal{L}_n$ satisfies

$$N\Big(\epsilon, \mathcal{L}_n, L_1(P_n)\Big) \leq \Big(\frac{10K^* M}{\epsilon}\Big)^{p+2} \cdot \Big(\frac{10K^* M_n}{\epsilon}\Big)^{m+1} \leq K M_n^{(m+1)} \epsilon^{-(p+m+3)} .$$

This completes the proof of Lemma 1.

Lemma 2 (Uniform convergence). Assume that Conditions (C1) - (C4) hold. Then we have

$$\sup_{\theta \in \Theta_n} \big| P_n l(\theta, O) - P l(\theta, O) \big| \to 0, \quad \text{almost surely},$$

where

$$P_n l(\theta, O) = \frac{1}{n} \sum_{i=1}^n l(\theta, O_i) = \frac{1}{n} \sum_{k=0}^2 \sum_{j=1}^{n_k} l(\theta, O_j) \quad \text{and} \quad P l(\theta, O) = \sum_{k=0}^2 \rho_k P^{(k)} l(\theta, O)$$

with $P^{(k)} l(\theta, O)$ being the expectation of $l(\theta, O)$ taken under the distribution $P^{(k)}$ that corresponds to the k-th stratum ($k = 0$ corresponds to the whole population).

Proof of Lemma 2

Note that $|l(\theta, O)|$ is bounded under Conditions (C1) - (C4). Without loss of generality, we assume $\sup_{\theta \in \Theta} |l(\theta, O)| \leq 1$. Then $P^{(k)}[l(\theta, O)]^2 \leq P^{(k)}(\sup_{\theta \in \Theta} |l(\theta, O)|)^2 \leq 1$. Let $\alpha_n =$

$n^{-1/2+\phi_1}(\log n)^{1/2}$ with $\nu/2 < \phi_1 < 1/2$. Then $\{\alpha_n\}$ is a nonincreasing sequence of positive numbers. Also for a given $\epsilon > 0$, let $\epsilon_n = \epsilon\alpha_n$. Then for sufficiently large n and any $\theta \in \Theta_n$, we have

$$\text{var}(P_n l(\theta, O))/(4\epsilon_n)^2 \leq \frac{(1/n^2)\sum_{k=0}^{2} n_k P^{(k)}[l(\theta, O)]^2}{16\epsilon^2\alpha_n^2} \leq \frac{1}{16\epsilon^2 n\alpha_n^2} = \frac{1}{16\epsilon^2 n^{2\phi_1}\log n} \leq \frac{1}{2}.$$

Let P_n^o denote the signed measure that places mass $\pm n^{-1}$ at each of the observations $\{O_1, \ldots, O_n\}$, with the random $\pm$ signs being decided independently of the O_i's. Then from Pollard (1984, p. 31) and $\text{var}(P_n l(\theta, O))/(4\epsilon_n)^2 \leq 1/2$, the following symmetrization inequality holds

$$P\Big(\sup_{\theta\in\Theta_n} |P_n l(\theta, O) - Pl(\theta, O)| > 8\epsilon_n\Big) \leq 4P\Big(\sup_{\theta\in\Theta_n} |P_n^o l(\theta, O)| > 2\epsilon_n\Big).$$

Let $\mathcal{O} = \{O_1, \cdots, O_n\}$. Given $\mathcal{O}$, choose $\theta^{(1)}, \ldots, \theta^{(\kappa)}$, where $\kappa = N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n))$, such that

$$\min_{j\in\{1,\ldots,\kappa\}} P_n\big|l(\theta, O) - l(\theta^{(j)}, O)\big| < \epsilon_n/2$$

for all $\theta \in \Theta_n$. For each $\theta \in \Theta_n$, write θ^* for the $\theta^{(j)}$ at which the minimum is achieved. Note that

$$\begin{aligned}
|P_n^o(l(\theta, O) - l(\theta^*, O))| &= \left|\frac{1}{n}\sum_{i=1}^{n} \pm(l(\theta, O_i) - l(\theta^*, O_i))\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} |l(\theta, O_i) - l(\theta^*, O_i)| = P_n|l(\theta, O) - l(\theta^*, O)|.
\end{aligned}$$

Then we have

$$\begin{aligned}
P\Big(\sup_{\theta\in\Theta_n} |P_n^o l(\theta, O)| > 2\epsilon_n|\mathcal{O}\Big) &\leq P\Big(\sup_{\theta\in\Theta_n} \big[|P_n^o l(\theta^*, O)| + P_n|l(\theta, O) - l(\theta^*, O)|\big] > 2\epsilon_n|\mathcal{O}\Big) \\
&\leq P\Big(\max_j |P_n^o l(\theta^{(j)}, O)| > 3\epsilon_n/2|\mathcal{O}\Big) \\
&\leq N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n)) \max_j P\Big(|P_n^o l(\theta^{(j)}, O)| > 3\epsilon_n/2|\mathcal{O}\Big).
\end{aligned}$$

From the definition of the covering number $N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n))$, for each $\theta^{(j)}$, there exists $\check{\theta}^{(j)} \in \Theta_n$ such that $P_n|l(\check{\theta}^{(j)}, O) - l(\theta^{(j)}, O)| < \epsilon_n/2$. Therefore, we obtain

$$\begin{aligned}
P\Big(|P_n^o l(\theta^{(j)}, O)| > 3\epsilon_n/2|\mathcal{O}\Big) &\leq P\Big(\big[P_n|l(\theta^{(j)}, O) - l(\check{\theta}^{(j)}, O)| + |P_n^o l(\check{\theta}^{(j)}, O)|\big] > 3\epsilon_n/2|\mathcal{O}\Big) \\
&\leq P\Big(|P_n^o l(\check{\theta}^{(j)}, O)| > \epsilon_n|\mathcal{O}\Big).
\end{aligned}$$

From Hoeffding's inequality (Pollard, 1984, Appendix B), we have

$$P\big(|P_n^o l(\check{\theta}^{(j)}, O)| > \epsilon_n | \mathcal{O}\big) = P\left(\left|\sum_{i=1}^n \pm l(\check{\theta}^{(j)}, O_i)\right| > n\epsilon_n | \mathcal{O}\right)$$
$$\leq 2\exp\left[-2(n\epsilon_n)^2 / \sum_{i=1}^n (2l(\check{\theta}^{(j)}, O_i))^2\right]$$
$$\leq 2\exp\big(-n\epsilon_n^2/2\big) \quad \text{(because } |l(\check{\theta}^{(j)}, O)| \leq 1\text{)}.$$

Combining the inequalities above together with Lemma 1, we obtain

$$P\big(\sup_{\theta\in\Theta_n} |P_n^o l(\theta, O)| > 2\epsilon_n | \mathcal{O}\big) \leq 2N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n)) \exp(-n\epsilon_n^2/2)$$
$$\leq 2KM_n^{(m+1)}(\epsilon_n/2)^{-(p+m+3)} \exp(-n\epsilon_n^2/2).$$

Note that the right-hand side does not depend on $\mathcal{O}$, then by taking expectations over $\mathcal{O}$, we have the following result

$$P\big(\sup_{\theta\in\Theta_n} |P_n^o l(\theta, O)| > 2\epsilon_n\big) \leq 2KM_n^{(m+1)}(\epsilon_n/2)^{-(p+m+3)} \exp(-n\epsilon_n^2/2).$$

Combining this result with the symmetrization inequality derived above and also noting that $M_n = O(n^a)$, $m = o(n^\nu)$ and $\phi_1 > \nu/2$, we obtain

$$P\big(\sup_{\theta\in\Theta_n} |P_n l(\theta, O) - Pl(\theta, O)| > 8\epsilon_n\big) \leq 4P\big(\sup_{\theta\in\Theta_n} |P_n^o l(\theta, O)| > 2\epsilon_n\big)$$
$$\leq 8KM_n^{(m+1)}(\epsilon_n/2)^{-(p+m+3)} \exp(-n\epsilon_n^2/2)$$
$$\leq 8K_1 \exp\left\{(m+1)a\log n - (p+m+3)[\log(\epsilon n^{-1/2+\phi_1}(\log n)^{1/2}) - \log 2] - n\epsilon^2 n^{-1+2\phi_1}\log n/2\right\}$$
$$\leq 8K_2 \exp\left\{(p+m+3)\big[(a+1/2-\phi_1)\log n - \log\log n/2 - \log\epsilon + \log 2\big] - \epsilon^2 n^{2\phi_1}\log n/2\right\}$$
$$\leq 8K_2 \exp\big(-K_3 n^{2\phi_1}\log n\big),$$

where K_1, K_2 and K_3 are constants. Hence $\sum_{n=1}^\infty P\big(\sup_{\theta\in\Theta_n} |P_n l(\theta, O) - Pl(\theta, O)| > 8\epsilon_n\big) < \infty$. By the Borel-Cantelli lemma, we have $\sup_{\theta\in\Theta_n} |P_n l(\theta, O) - Pl(\theta, O)| \to 0$ almost surely, which completes the proof of Lemma 2.

Now we are ready to prove Theorems 1 and 2.

Proof of Theorem 1

We first prove the strong consistency of $\hat{\theta}_n$. Let $l(\theta, O)$ denote the log-likelihood function based on a single observation O and consider the class of functions $\mathcal{L}_n = \{l(\theta, O) : \theta \in \Theta_n\}$.

Then based on Lemma 1, the covering number of $\mathcal{L}_n$ satisfies

$$N\Big(\epsilon, \mathcal{L}_n, L_1(P_n)\Big) \leq K M_n^{(m+1)} \epsilon^{-(p+m+3)} .$$

Furthermore, by Lemma 2, we have

$$\sup_{\theta \in \Theta_n} \Big| P_n l(\theta, O) - P l(\theta, O) \Big| \to 0 \quad \text{almost surely.} \tag{A.1}$$

Let $M(\theta, O) = -l(\theta, O)$, and define $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$ for $\epsilon > 0$ and

$$\zeta_{1n} = \sup_{\theta \in \Theta_n} |P_n M(\theta, O) - PM(\theta, O)|, \; \zeta_{2n} = P_n M(\theta_0, O) - PM(\theta_0, O).$$

Then one can show that

$$\inf_{K_\epsilon} PM(\theta, O) = \inf_{K_\epsilon} \Big\{PM(\theta, O) - P_n M(\theta, O) + P_n M(\theta, O)\Big\} \leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta, O). \tag{A.2}$$

If $\hat{\theta}_n \in K_\epsilon$, then we have

$$\inf_{K_\epsilon} P_n M(\theta, O) = P_n M(\hat{\theta}_n, O) \leq P_n M(\theta_0, O) = \zeta_{2n} + PM(\theta_0, O). \tag{A.3}$$

Define $\delta_\epsilon = \inf_{K_\epsilon} PM(\theta, O) - PM(\theta_0, O)$. Then under Condition (C2), using the same arguments as those in Zhang *et al.* (2010, p. 352), we can prove $\delta_\epsilon > 0$. It follows from (A.2) and (A.3) that

$$\inf_{K_\epsilon} PM(\theta, O) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0, O) = \zeta_n + PM(\theta_0, O)$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$, and hence $\zeta_n \geq \delta_\epsilon$. This gives $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$, and by (A.1) and the strong law of large numbers, we have both $\zeta_{1n} \to 0$ and $\zeta_{2n} \to 0$ almost surely. Therefore, $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$, which proves that $d(\hat{\theta}_n, \theta_0) \to 0$ almost surely.

Now we will show the convergence rate of $\hat{\theta}_n$ by using Theorem 3.4.1 of van der Vaart and Wellner (1996). Below we will use $\tilde{K}$ to denote a universal positive constant which may differ from place to place. First note from Theorem 1.6.2 of Lorentz (1986) that there exists a Bernstein polynomial Λ_{n0} such that $\|\Lambda_{n0} - \Lambda_0\|_\infty = O(m^{-r/2})$. Define $\theta_{n0} = (\xi_0, \Lambda_{n0})$. Then we have $d(\theta_{n0}, \theta_0) = O(n^{-r\nu/2})$. For any $\eta > 0$, define the class of functions $\mathcal{F}_\eta = \{l(\theta, O) - l(\theta_{n0}, O) : \theta \in \Theta_n, \, \eta/2 < d(\theta, \theta_{n0}) \leq \eta\}$. One can easily show that $P(l(\theta_0, O) - l(\theta_{n0}, O)) \leq \tilde{K} d(\theta_0, \theta_{n0}) \leq \tilde{K} n^{-r\nu/2}$. Also under Condition (C2), using the same arguments as those in Zhang *et al.* (2010, p. 352), we obtain $P(l(\theta_0, O) - l(\theta, O)) \geq \tilde{K} d^2(\theta_0, \theta)$. Thus, for large n, we have $P(l(\theta, O) - l(\theta_{n0}, O)) = P(l(\theta, O) - l(\theta_0, O)) + P(l(\theta_0, O) - l(\theta_{n0}, O)) \leq -\tilde{K}\eta^2 + \tilde{K} n^{-r\nu/2} = -\tilde{K}\eta^2$, for any $l(\theta, O) - l(\theta_{n0}, O) \in \mathcal{F}_\eta$.

Following the calculations in Shen and Wong (1994, p. 597), we can establish that for $0 < \varepsilon < \eta$, $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, L_2(P)) \leq \tilde{K} N \log(\eta/\varepsilon)$ with $N = m+1$. Moreover, some algebraic manipulations yield that $P(l(\theta, O) - l(\theta_{n0}, O))^2 \leq \tilde{K}\eta^2$ for any $l(\theta, O) - l(\theta_{n0}, O) \in \mathcal{F}_\eta$. Under Conditions (C1) - (C4), it is easy to see that $\mathcal{F}_\eta$ is uniformly bounded. Therefore, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain

$$E_P \| n^{1/2}(P_n - P) \|_{\mathcal{F}_\eta} \leq \tilde{K} J_{[]}(\eta, \mathcal{F}_\eta, L_2(P)) \left\{ 1 + \frac{J_{[]}(\eta, \mathcal{F}_\eta, L_2(P))}{\eta^2 n^{1/2}} \right\}$$

where $J_{[]}(\eta, \mathcal{F}_\eta, L_2(P)) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, L_2(P))\}^{1/2} d\varepsilon \leq \tilde{K} N^{1/2} \eta$. This yields $\phi_n(\eta) = N^{1/2}\eta + N/n^{1/2}$. It is easy to see that $\phi_n(\eta)/\eta$ is decreasing in η, and $r_n^2 \phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} \leq \tilde{K} n^{1/2}$, where $r_n = N^{-1/2} n^{1/2} = n^{(1-\nu)/2}$.

Finally note that $P_n(l(\hat{\theta}_n, O) - l(\theta_{n0}, O)) \geq 0$ and $d(\hat{\theta}_n, \theta_{n0}) \leq d(\hat{\theta}_n, \theta_0) + d(\theta_0, \theta_{n0}) \to 0$ in probability. Thus by applying Theorem 3.4.1 of van der Vaart and Wellner (1996), we have $n^{(1-\nu)/2} d(\hat{\theta}_n, \theta_{n0}) = O_P(1)$. This together with $d(\theta_{n0}, \theta_0) = O(n^{-r\nu/2})$ yields that $d(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu/2})$ and the proof is completed.

Proof of Theorem 2

To establish the asymptotic normality of $\hat{\xi}_n$, following the proof of Theorem 2 in Zhang *et al.* (2010), one can first obtain that

$$\sqrt{n}(\hat{\xi}_n - \xi_0) \;=\; \left\{ \sum_{k=0}^{2} \frac{n_k}{n} J_k(\xi_0) \right\}^{-1} \left\{ \sum_{k=0}^{2} \sqrt{\frac{n_k}{n}} \left(\frac{1}{\sqrt{n_k}} \sum_{i=1}^{n_k} h_k(\xi_0, \Lambda_0; O_i) \right) \right\} \;+\; o_p(1),$$

where $h_k(\xi, \Lambda; O_i)$ and $J_k(\xi)$ are the efficient score and information for ξ corresponding to the k-th stratum ($k = 0$ corresponds to the whole population), which can be derived similarly as in Zhang *et al.* (2010, p. 344) with our parameters (ξ, Λ) corresponding to theirs $(\theta, \exp(\phi))$. Note that $n_0/n \to \rho_0 > 0$ and $n_k/n \to \rho_k \geq 0$, $k = 1, 2$, as $n \to \infty$. Thus, we have

$$\sqrt{n}(\hat{\xi}_n - \xi_0) \to_d \; N(0, \Sigma),$$

where $\Sigma \;=\; \Gamma^{-1}\Psi\Gamma^{-1}$ with

$$\Gamma = \sum_{k=0}^{2} \rho_k \, J_k(\xi_0) \quad \text{and} \quad \Psi = \sum_{k=0}^{2} \rho_k \, var_k(h_k(\xi_0, \Lambda_0; O)).$$

Simulation results under sampling without replacement

We conducted a simulation study under the same setup as in Table 1 of the main paper by using sampling without replacement and the results are given in Web Table 1. The proposed method under sampling without replacement performs similarly as under independent Bernoulli sampling (see Table 1 of the main paper).

References

Huang, J. and Rossini, A. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, **92**(439), 960–967.

Lorentz, G. G. (1986). *Bernstein Polynomials*. New York: Chelsea Publishing Co.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics*, **22**(2), 580–615.

van de Geer, S. A. (2000). *Applications of Empirical Process Theory*. Cambridge Univ. Press.

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

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

Web Table 1: Simulation results for the estimation of β when $(n_0, n_1, n_2) = (470, 40, 40)$: the samples are selected by sampling without replacement

			$\beta = 0$					$\beta = \log 2$				
Pr(failure)	cutpoints		Bias	SSD	ESE	CP	RE	Bias	SSD	ESE	CP	RE
0.1	(20%, 80%)	$\hat{\beta}_{SRS_{n_0}}$	-0.000	0.145	0.144	0.94	0.90	-0.005	0.172	0.149	0.93	0.82
		$\hat{\beta}_{SRS_n}$	0.004	0.138	0.133	0.94	1.00	0.001	0.155	0.138	0.94	1.00
		$\hat{\beta}_{GCC}$	0.000	0.115	0.112	0.95	1.44	-0.001	0.138	0.131	0.94	1.26
		$\hat{\beta}_{IPW}$	-0.001	0.125	0.121	0.94	1.21	-0.004	0.146	0.139	0.94	1.13
		$\hat{\beta}_P$	-0.001	0.097	0.098	0.95	2.01	-0.002	0.106	0.122	0.93	2.14
	(10%, 90%)	$\hat{\beta}_{SRS_{n_0}}$	-0.003	0.143	0.144	0.96	0.96	-0.008	0.173	0.149	0.93	0.74
		$\hat{\beta}_{SRS_n}$	-0.005	0.140	0.133	0.94	1.00	0.006	0.149	0.138	0.94	1.00
		$\hat{\beta}_{GCC}$	0.001	0.110	0.113	0.96	1.62	-0.003	0.135	0.132	0.95	1.21
		$\hat{\beta}_{IPW}$	-0.002	0.130	0.130	0.95	1.17	-0.007	0.163	0.147	0.92	0.84
		$\hat{\beta}_P$	-0.002	0.102	0.100	0.95	1.89	-0.001	0.102	0.104	0.94	2.12
0.2	(20%, 80%)	$\hat{\beta}_{SRS_{n_0}}$	-0.000	0.108	0.104	0.95	0.79	0.007	0.112	0.108	0.95	0.88
		$\hat{\beta}_{SRS_n}$	0.001	0.096	0.095	0.95	1.00	0.009	0.105	0.099	0.94	1.00
		$\hat{\beta}_{GCC}$	-0.005	0.110	0.107	0.94	0.76	0.000	0.112	0.112	0.95	0.88
		$\hat{\beta}_{IPW}$	0.001	0.095	0.093	0.94	1.02	0.005	0.097	0.097	0.95	1.17
		$\hat{\beta}_P$	-0.001	0.086	0.083	0.94	1.24	-0.010	0.084	0.085	0.95	1.54
	(10%, 90%)	$\hat{\beta}_{SRS_{n_0}}$	0.001	0.106	0.103	0.96	0.88	0.006	0.112	0.108	0.95	0.88
		$\hat{\beta}_{SRS_n}$	0.002	0.099	0.096	0.94	1.00	0.009	0.105	0.099	0.94	1.00
		$\hat{\beta}_{GCC}$	-0.004	0.107	0.107	0.95	0.86	-0.000	0.112	0.112	0.95	0.87
		$\hat{\beta}_{IPW}$	0.002	0.098	0.097	0.95	1.03	0.007	0.102	0.102	0.95	1.06
		$\hat{\beta}_P$	0.000	0.087	0.084	0.95	1.32	-0.002	0.089	0.091	0.94	1.40
0.3	(20%, 80%)	$\hat{\beta}_{SRS_{n_0}}$	0.001	0.084	0.085	0.95	0.86	0.005	0.093	0.091	0.95	0.83
		$\hat{\beta}_{SRS_n}$	-0.004	0.078	0.078	0.95	1.00	0.003	0.084	0.084	0.95	1.00
		$\hat{\beta}_{GCC}$	-0.004	0.100	0.100	0.95	0.62	0.008	0.110	0.108	0.94	0.58
		$\hat{\beta}_{IPW}$	0.001	0.078	0.078	0.94	1.00	0.005	0.084	0.084	0.95	1.00
		$\hat{\beta}_P$	0.001	0.075	0.073	0.94	1.10	-0.018	0.079	0.076	0.93	1.13
	(10%, 90%)	$\hat{\beta}_{SRS_{n_0}}$	0.001	0.084	0.085	0.95	0.86	0.005	0.093	0.091	0.95	0.83
		$\hat{\beta}_{SRS_n}$	-0.004	0.078	0.078	0.95	1.00	0.003	0.084	0.084	0.95	1.00
		$\hat{\beta}_{GCC}$	-0.004	0.100	0.100	0.95	0.62	0.008	0.110	0.108	0.94	0.58
		$\hat{\beta}_{IPW}$	0.001	0.081	0.081	0.94	0.95	0.006	0.089	0.087	0.95	0.89
		$\hat{\beta}_P$	0.000	0.075	0.082	0.94	1.08	-0.003	0.080	0.079	0.95	1.09

$\hat{\beta}_{SRS_{n_0}}$, the sieve MLE based only on the SRS portion of the ODS sample; $\hat{\beta}_{SRS_n}$, the sieve MLE based on a SRS sample of the same size as the ODS sample; $\hat{\beta}_{GCC}$, the estimator based on the generalized case-cohort sample; $\hat{\beta}_{IPW}$, the inverse probability weighted estimator based on the ODS sample; $\hat{\beta}_P$, the proposed estimator based on the ODS sample.