BIOMETRICS 000, ??–10 DOI: 000
000 0000

Web-based Supplementary Materials for

Integrated Powered Density: Screening Ultrahigh Dimensional Covariates with Survival Outcomes

by Hyokyoung G. Hong, Xuerong Chen, David C. Christiani, and Yi Li

This paper has been submitted for consideration for publication in *Biometrics*

1. Proof of the main results

We present several useful lemmas before proving the theoretical results in the main text.

LEMMA 1: *For a categorical covariate X_j with R_j categories, let $\hat{S}_{T|X_j}(t|r)$ be the Kaplan-Meier estimator of conditional survival function within the subsample $X_j = r, r = 1, \ldots, R_j$. Under conditions (C1) and (C5), we have*

$$P(\max_{1\leqslant r\leqslant R_j} \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)| > \epsilon) \leqslant d_3 R \exp(-d_4\epsilon^2\theta_1^{25} n^{1-3\kappa}),$$

where d_3 and d_4 are positive constants, $R = \max_{1\leqslant j\leqslant p} R_j$.

Proof. By the inequality in the last paragraph on page 1161 of Dabrowska (1989), we have

$$\begin{aligned} & P(\max_{r} \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)| > \epsilon) \\ \leqslant\ & d_3 R_j \exp(-d_4\epsilon^2\theta_1^{25} \min_r n_r R_j^{-2}) \\ \leqslant\ & d_3 R \exp(-d_4\epsilon^2\theta_1^{25} \min_r n_r R^{-2}) \end{aligned}$$

where n_r is the subsample size of $X_j = r$. By condition (C6), we have $\min_r n_r \geqslant n/R = n^{1-\kappa}$. $\square$

LEMMA 2: *Under (C1)-(C5), for a categorical covariate X_j with R_j categories, we have*

$$P(\max_{1\leqslant r\leqslant R_j} \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|r) - f_{T|X_j}(t|r)| > \epsilon) \leqslant d_3 R \exp\left(-\frac{1}{4} d_4\epsilon^2\theta_1^{25} n^{1-3\kappa} h_n^2\right),$$

where $R = \max_{1\leqslant j\leqslant p} R_j$.

Proof. Note that

$$
\begin{aligned}
& \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|r) - f_{T|X_j}(t|r)| \\
\leqslant\ & \sup_{t\in[0,\tau]} \left| -\int K_{h_n}(t-s)d\hat{S}_{T|X_j}(s|r) + \int K_{h_n}(t-s)dS_{T|X_j}(s|r)\right| \\
& + \sup_{t\in[0,\tau]} \left| -\int K_{h_n}(t-s)dS_{T|X_j}(s|r) - f_{T|X_j}(t|r)\right| \\
\leqslant\ & \sup_{t\in[0,\tau]} \left| -\int K_{h_n}(t-s)d[\hat{S}_{T|X_j}(s|r) - S_{T|X_j}(s|r)]\right| \\
& + \sup_{t\in[0,\tau]} \left| -\int K_{h_n}(t-s)dS_{T|X_j}(s|r) - f_{T|X_j}(t|r)\right| \\
=:\ & I_1 + I_2.
\end{aligned}
$$

Assume that there exists a constant C_0 such that $|K| \leqslant C_0$. Integration by parts yields that

$$
\begin{aligned}
I_1 &= \left| -[\hat{S}_{T|X_j}(s|r) - S_{T|X_j}(s|r)]K_{h_n}(t-s)|_0^\tau + \int [\hat{S}_{T|X_j}(s|r) - S_{T|X_j}(s|r)]dK_{h_n}(t-s)\right| \\
&\leqslant C_0 h_n^{-1} \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)| + V_K h_n^{-1} \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)| \\
&\leqslant (C_0 + V_K) h_n^{-1} \max_r \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)|.
\end{aligned}
$$

For I_2, we have

$$
\begin{aligned}
I_2 &= \sup_{t\in[0,\tau]} \left| \int K_{h_n}(s-t) f_{T|X_j}(s|r)ds - f_{T|X_j}(t|r)\right| \\
&= \sup_{t\in[0,\tau]} \left| \int K(u) f_{T|X_j}(t+uh_n|r)du - f_{T|X_j}(t|r)\right| = O(h_n^2).
\end{aligned}
$$

Note that $P(I_2 > \epsilon/2) = 0$. Therefore, by Lemma 1, we have

$$
\begin{aligned}
& P(\max_r \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|r) - f_{T|X_j}(t|r) > \epsilon| \\
\leqslant\ & P(I_1 > \frac{\epsilon}{2}) + P(I_2 > \frac{\epsilon}{2}) \\
\leqslant\ & P(\sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|r) - S_{T|X_j}(t|r)| > \frac{\epsilon h_n}{2}) \\
\leqslant\ & d_3 R \exp\left(-\frac{1}{4} d_4 \epsilon^2 \theta_1^{25} n^{1-3\kappa} h_n^2 \right).
\end{aligned}
$$

□

LEMMA 3: *Under (C1)-(C5), for a categorical covariate X_j with R_j categories, i.e., $X_j =$*

r for $1 \leqslant r \leqslant R_j$, we have

$$P(|\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| > \epsilon) \leqslant d_6 R \exp(-d_5 \epsilon^2 n^{1-3\kappa} h_n^2),$$

where d_5 and d_6 are positive constants.

Proof. Note that

$$\begin{aligned}
& |\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| \\
= & \left| \max_{r_1, r_2} \sup_{t \in [0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_1) ds - \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_2) ds \right| \right. \\
& \left. - \max_{r_1, r_2} \sup_{t \in [0,\tau]} \left| \int_0^t f^{\gamma}_{T|X_j}(s|X_j = r_1) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j = r_2) ds \right| \right| \\
\leqslant & \max_{r_1} \sup_{t \in [0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_1) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j = r_1) ds \right| \\
& + \max_{r_2} \sup_{t \in [0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_2) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j = r_2) ds \right| \\
=: & I_{31} + I_{32}.
\end{aligned}$$

By Lemma 2 and the mean value theorem,

$$\begin{aligned}
& \hat{f}^{\gamma}_{T|X_j}(t|X_j = r_1) - f^{\gamma}_{T|X_j}(t|X_j = r_1) \\
= & \{f_{T|X_j}(t|X_j = r_1) + [\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)]\}^{\gamma} - f^{\gamma}_{T|X_j}(t|X_j = r_1) \\
= & \gamma\{f_{T|X_j}(t|X_j = r_1) + \zeta^*[\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)]\}^{\gamma-1} \\
& \times[\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)] \\
=: & \gamma\psi(\zeta^*)[\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)],
\end{aligned}$$

where ζ^* is a constant between 0 and 1. For $\gamma > 1$, we have

$$\begin{aligned}
|\psi(\zeta^*)| & = |\{f_{T|X_j}(t|X_j = r_1) + \zeta^*[\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)]\}^{\gamma-1}| \\
& \leqslant [3 f_{T|X_j}(t|X_j = r_1)]^{\gamma-1} \\
& \leqslant 3^{\gamma-1}[\sup_{t \in [0,\tau]} f_{T|X_j}(t|X_j = r_1)]^{\gamma-1},
\end{aligned}$$

and for $\gamma < 1$, we have

$$\begin{aligned}
|\psi(\zeta^*)| &= |\{f_{T|X_j}(t|X_j = r_1) + \zeta^*[\hat{f}_{T|X_j}(t|X_j = r_1) - f_{T|X_j}(t|X_j = r_1)]\}^{\gamma-1}| \\
&\leqslant \left[\frac{1}{2} f_{T|X_j}(t|X_j = r_1)\right]^{\gamma-1} \\
&\leqslant \left(\frac{1}{2}\right)^{\gamma-1} [\inf_{s\in[0,\tau]} f_{T|X_j}(t|X_j = r_1)]^{\gamma-1}.
\end{aligned}$$

Let

$$G_1(\gamma) = \begin{cases} 3^{\gamma-1}[\sup_{t\in[0,\tau]} f_{T|X_j}(t|X_j = r_1)]^{\gamma-1}, \text{ if } \gamma > 1, \\ 1, \text{ if } \gamma = 1, \\ (\frac{1}{2})^{\gamma-1}[\inf_{t\in[0,\tau]} f_{T|X_j}(t|X_j = r_1)]^{\gamma-1}, \text{ if } \gamma < 1. \end{cases}$$

Then we have

$$\begin{aligned}
I_{31} &= \max_{r_1} \sup_{t\in[0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_1) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j = r_1) ds \right| \\
&\leqslant \max_{r_1} \sup_{t\in[0,\tau]} \int_0^t \left| \hat{f}^{\gamma}_{T|X_j}(s|X_j = r_1) - f^{\gamma}_{T|X_j}(s|X_j = r_1) \right| ds \\
&\leqslant |\gamma| G_1(\gamma) \tau \max_{r} \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|r) - f_{T|X_j}(t|r)|.
\end{aligned}$$

Similarly,

$$I_{32} \leqslant |\gamma| G_2(\gamma) \tau \max_{r} \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|r) - f_{T|X_j}(t|r)|,$$

where

$$G_2(\gamma) = \begin{cases} 3^{\gamma-1}[\sup_{t\in[0,\tau]} f_{T|X_j}(t|X_j = r_2)]^{\gamma-1}, \text{ if } \gamma > 1, \\ 1, \text{ if } \gamma = 1, \\ (\frac{1}{2})^{\gamma-1}[\inf_{t\in[0,\tau]} f_{T|X_j}(t|X_j = r_2)]^{\gamma-1}, \text{ if } \gamma < 1. \end{cases}$$

The result follows from Lemma 2. □

Proof of Theorem 1. By Lemma 3, we have

$$\begin{aligned}
P(\mathcal{M} \subset \widehat{\mathcal{M}}_1) &\geqslant P\bigg(|\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| \leqslant cn^{-v}\bigg) \\
&\geqslant P(\max_{1\leqslant j\leqslant p} |\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| \leqslant cn^{-v}) \\
&\geqslant 1 - \sum_{j=1}^{p} P(|\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| > cn^{-v}) \\
&\geqslant 1 - \sum_{j=1}^{p} \bigg[d_6 R \exp\bigg(-\frac{1}{4} d_5 c^2 n^{1-3\kappa-2v} h_n^2\bigg)\bigg] \\
&= 1 - O(pn^{\kappa}) \exp\bigg(-\frac{1}{4} d_5 c^2 n^{1-3\kappa-2v} h_n^2\bigg) \\
&= 1 - O(p \exp\{-b_0 n^{1-3\kappa-2v} h_n^2 + \kappa \log n\}),
\end{aligned}$$

where b_0 is a positive constant. □

Proof of Corollary 1. Under the assumption $\sum_{j=1}^{p} \mathcal{I}_j^{(\gamma)} = O(\zeta)$, it is easy to obtain that the cardinality of $\{j : \mathcal{I}_j^{(\gamma)} \geqslant cn^{-v}\}$ is no greater than $O(n^{\zeta+v})$. Hence, on the set

$$\Omega_n = \{\sup_{1\leqslant j\leqslant p} |\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| \leqslant cn^{-v}\},$$

we have

$$\{j : \widehat{\mathcal{I}}_j^{(\gamma)} \geqslant 2cn^{-v}\} \leqslant \{j : \mathcal{I}_j^{(\gamma)} \geqslant cn^{-v}\} = O(n^{\zeta+v}).$$

By Lemma 3, we have

$$P(\sup_{1\leqslant j\leqslant p} |\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| > cn^{-v}) \leqslant O(R) \exp(-d_5 \epsilon^2 n^{1-3\kappa-2v}).$$

□

Let $q_{j(r)}$ be the r/R_j theoretical quantile of X_j, for $r = 1, \cdots, R_j$. For notational simplicity, let $\hat{J}_r = [\hat{q}_{j(r-1)}, \hat{q}_{j(r)})$ and $J_r = [q_{j(r-1)}, q_{j(r)})$ in the following statements.

Lemma 4: *For continuous covariate X_j, let $\hat{S}_{T|X_j}(t|X_j \in \hat{J}_r)$ be the Kaplan-Meier estimator of the conditional survival function within the subsample $X_j \in \hat{J}_r$, and assume conditions (C1),(C5) and (C6) hold. Then,*

$$P(\max_{r} \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|X_j \in \hat{J}_r) - S_{T|X_j}(t|X_j \in J_r)| > \epsilon) \leqslant d_7 R \exp(-d_8 \epsilon^2 n^{1-3\kappa-2\rho}),$$

for any $1 \leqslant r \leqslant R_j$*, and* $R = \max_{1\leqslant j\leqslant p} R_j$*, where* d_7 *and* d_8 *are positive constants.*

Proof. By consistency of $\hat{q}_{j(r)}$, it is easy to obtain that,

$$F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(\hat{q}_{j(r-1)}) > 0.5[F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})].$$

By the mean value theorem,

$$\begin{aligned}
& |S_{T|X_j}(t|X_j \in \hat{J}_r) - S_{T|X_j}(t|X_j \in J_r)| \\
= & \left|\frac{P(T>t, X_j<\hat{q}_{j(r)}) - P(T>t, X_j<\hat{q}_{j(r-1)})}{F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(\hat{q}_{j(r-1)})}\right. \\
& \left.-\frac{P(T>t, X_j<q_{j(r)}) - P(T>t, X_j<q_{j(r-1)})}{F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})}\right| \\
\leqslant & \left|\frac{P(T>t, X_j<\hat{q}_{j(r)}) - P(T>t, X_j<\hat{q}_{j(r-1)})}{F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(\hat{q}_{j(r-1)})}\right. \\
& \left.-\frac{P(T>t, X_j<q_{j(r)}) - P(T>t, X_j<q_{j(r-1)})}{F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(\hat{q}_{j(r-1)})}\right| \\
& +\left|\frac{P(T>t, X_j<q_{j(r)}) - P(T>t, X_j<q_{j(r-1)})}{F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(\hat{q}_{j(r-1)})}\right. \\
& \left.-\frac{P(T>t, X_j<q_{j(r)}) - P(T>t, X_j<q_{j(r-1)})}{F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})}\right| \\
\leqslant & \frac{2}{F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})}\Big[|P(T>t, X_j<\hat{q}_{j(r)}) - P(T>t, X_j<q_{j(r)})| \\
& +|P(T>t, X_j<\hat{q}_{j(r-1)}) - P(T>t, X_j<q_{j(r-1)})|\Big] \\
& +\frac{2}{[F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})]^2}[|F_{X_j}(\hat{q}_{j(r-1)}) - F_{X_j}(q_{j(r-1)})| + |F_{X_j}(\hat{q}_{j(r)}) - F_{X_j}(q_{j(r)})|] \\
=: & I_{41} + I_{42} + I_{43} + I_{44}.
\end{aligned}$$

For I_{41}, we have

$$\begin{aligned}
I_{41} &= \frac{2}{F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})}|P(T>t, X_j<\hat{q}_{j(r)}) - P(T>t, X_j<q_{j(r)})| \\
&\leqslant \frac{2}{F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})}\left|\int_t^\infty f_{T|X_j}(s|q^*_{j(r)}) f_{X_j}(q^*_{j(r)}) ds\right| \max_r |\hat{q}_{j(r)} - q_{j(r)}|,
\end{aligned}$$

where $q^*_{j(r)}$ lies between $\hat{q}_{j(r)}$ and $q_{j(r)}$. Hence,

$$\begin{aligned}
& P\left(I_{41} > \frac{\epsilon}{8}\right) \\
\leqslant\ & P\left(\max_r |\hat{q}_{j(r)} - q_{j(r)}| > \frac{\epsilon[F_{X_j}(q_{j(r)}) - F_{X_j}(q_{j(r-1)})]}{16|\int_t^\infty f_{T|X_j}(s|q^*_{j(r)}) f_{X_j}(q^*_{j(r)}) ds|}\right) \\
\leqslant\ & b_2 R_j \exp(-b_1 n^{1-2\rho}\epsilon^2) \\
\leqslant\ & b_2 R \exp(-b_1 n^{1-2\rho}\epsilon^2),
\end{aligned}$$

where b_1 and b_2 are positive constants, and the second inequality is obtained by Lemma A.2 from Ni and Fang (2016). Similarly, we can have $P(I_{4k} > \epsilon/8) \leqslant b_{2k} R \exp(-b_k n^{1-2\rho}\epsilon^2)$, for $k = 2, 3, 4$ and where b_k and b_{2k} are positive constants. Therefore, we have

$$\begin{aligned}
& P(\max_r \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|X_j \in \hat{J}_r) - S_{T|X_j}(t|X_j \in J_r)| > \epsilon) \\
\leqslant\ & P(\max_r \sup_{t\in[0,\tau]} |\hat{S}_{T|X_j}(t|X_j \in \hat{J}_r) - S_{T|X_j}(t|X_j \in \hat{J}_r)| > \epsilon/2) \\
& + P(\max_r \sup_{t\in[0,\tau]} |S_{T|X_j}(t|X_j \in \hat{J}_r) - S_{T|X_j}(t|X_j \in J_r)| > \epsilon/2) \\
\leqslant\ & d_3 R \exp(-d_4(\epsilon/2)^2 \theta_2^{25} n^{1-3\kappa}) + \sum_{k=1}^{4} P\left(I_{4k} > \frac{\epsilon}{8}\right) \\
\leqslant\ & d_7 R \exp(d_8 \epsilon^2 n^{1-3\kappa-2\rho}).
\end{aligned}$$

□

Lemma 5: *Under (C1)-(C4) and (C6), for a continuous covariate X_j, we have*

$$P(\max_r \sup_{t\in[0,\tau]} |\hat{f}_{T|X_j}(t|X_j \in \hat{J}_r) - f_{T|X_j}(t|X_j \in J_r)| > \epsilon) \leqslant d_9 \exp(-d_{10}\epsilon^2 n^{1-3\kappa-2\rho-2\mu}),$$

where d_9, d_{10} are positive constants.

Proof. The proof of this lemma is similar to that of Lemma 2, and is omitted. □

Lemma 6: *Under (C1)-(C4) and (C6), for a continuous covariate X_j, we have*

$$P(|\widehat{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_j^{(\gamma)}| > \epsilon) \leqslant d_{11} R \exp(-d_{12}\epsilon^2 n^{1-3\kappa-2\rho-2\mu}),$$

where d_{11}, d_{12} are positive constants, and $R = \max_{1\leqslant j\leqslant p} R_j$.

Proof. The proof of this lemma is similar to that of Lemma 3. By Lemmas 4 and 5, it is easy to obtain the conclusion. □

Proof of Theorem 2. By Lemma 6, the proof of this theorem is similar to that of Theorem 1, and hence is omitted. □

Proof of Corollary 2. The proof of it is similar to that of Corollary 1, and we omit it here. □

For simplicity, let $\hat{J}_{ur} = [\hat{q}_{ju(r-1)}, \hat{q}_{ju(r)})$, and $J_{ur} = [q_{ju(r-1)}, q_{ju(r)})$.

LEMMA 7: *Under (C1)-(C4) and (C6), for a continuous covariate X_j, we have*

$$P(|\widetilde{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_{jo}^{(\gamma)}| > \epsilon) \leqslant d_{13} N R \exp(-d_{14}\epsilon^2 n^{1-3\kappa-2\rho-2\mu}),$$

where d_{13}, d_{14} are positive constants, and $R = \max_{1\leqslant j\leqslant p, 1\leqslant u\leqslant N} R_{ju}$.

Proof. Note that

$$\begin{aligned} & |\widetilde{\mathcal{I}}_j^{(\gamma)} - \mathcal{I}_{jo}^{(\gamma)}| \\ \leqslant & \sum_{u=1}^{N} |\widehat{\mathcal{I}}_{j,\Lambda_{ju}}^{(\gamma)} - \mathcal{I}_{j,\Lambda_{juo}}^{(\gamma)}| \\ \leqslant & \sum_{u=1}^{N} \left[\max_{r_1} \sup_{t\in[0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j \in \hat{J}_{ur_1}) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j \in J_{ur_1}) ds \right| \right. \\ & \left. + \max_{r_2} \sup_{t\in[0,\tau]} \left| \int_0^t \hat{f}^{\gamma}_{T|X_j}(s|X_j \in \hat{J}_{ur_2}) ds - \int_0^t f^{\gamma}_{T|X_j}(s|X_j \in J_{ur_2}) ds \right| \right]. \end{aligned}$$

By Lemma 6, similar to the proof of Lemma 3, it is easy to obtain the conclusion. □

Proof of Theorem 3. By Lemma 7, the proof is similar to that of Theorem 1, and hence is omitted. □

Proof of Corollary 3. The proof is similar to that of Corollary 1, and is omitted. □

2. On the Choice of bandwidth h_n

From Theorem 2.2 of Lo et al. (1989), we can obtain that

$$\begin{aligned} \mathrm{E}[\hat{f}_T(t)] &= f(t) + \frac{f''(t)h_n^2}{2} \int s^2 K(s) ds + o(h_n) + o((nh_n)^{-1/2}), \\ Var[\hat{f}_T(t)] &= \frac{1}{nh_n} \frac{f(t)}{P(Y_i > t)} \int K^2(s) ds + o((nh_n)^{-1}). \end{aligned}$$

Obviously there is a trade-off: when h_n increases, the bias becomes larger, while the variance become smaller; when h_n decreases, the bias becomes smaller, while the variance become larger. An optimal h_n could be selected by minimizing the mean squared error (MSE) of $\hat{f}(t)$, which strikes a balance between bias and variance:

$$\text{MSE} = \left[\frac{f''(t)h_n^2}{2}\int s^2K(s)ds\right]^2 + \frac{1}{nh_n}\frac{f(t)}{P(Y_i > t)}\int K^2(s)ds + o((nh_n)^{-1}) + o(h_n^4).$$

It follows that the minimal of MSE could be achieved when $h_n = O(n^{-1/5})$. That is, the optimal bandwidth is in the order $O(n^{-1/5})$.

To explore how the bandwidth can impact the results with various γ, we present in Figure S1 the boxplots of the MMS for IPOD in Example 1 with $(n, p) = (500, 1000)$, $\gamma = 0.1, 0.5, 0.8, 1, 1.2, 1.5, 2.0, 2.5, 3.0$, and $h_n = h_0 n^{-1/5}$ with $h_0 = 0.4,\ 2,\ 5,\ 10$, respectively. Figure S1 shows a U-shaped relationship between γ and MMS. The impact of the bandwidth appeared negligible unless the bandwidth was too narrow or too wide. In addition, if a γ was too distant from 1, it did not help detect differences in distributions and produced less meaningful results. On the other hand, using γ from 0.7 to 1.5 might help IPOD detect early or late differences.

[Supplemental Material, Figure 1 about here.]

3. Additional Numerical Results

Example 5. *The survival time was generated from a Cox model,* $\lambda(t|\mathbf{X}) = 0.2\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ *where the covariates* X_j *were from a multivariate normal distribution and* $\boldsymbol{\beta} = (\mathbf{0.3}_5^{\mathrm{T}}, \mathbf{0}_{p-5}^{\mathrm{T}})^{\mathrm{T}}$. *For the true covariance, we considered an exchangeable correlation structure with an equal correlation of 0.5. The censoring times* C_i *were independently generated from a uniform distribution* $U[0, c]$*, with* c *chosen to give approximately 20% and 50% of censoring proportions.*

Example 5*. *The setup was the same as in Example 5 except that the censoring times* C_i *were* *covariate-dependent and generated from* $\lambda_C(t|\mathbf{X}) = c\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$*, where* $\boldsymbol{\beta} = (\mathbf{0.3}_2^{\mathrm{T}}, \mathbf{0}_{p-2}^{\mathrm{T}})^{\mathrm{T}}$*, and* c *was chosen to give approximately 20% and 50% of censoring proportions.*

[Supplemental Material, Table 1 about here.]

Table S1 indicates that when the censoring time depended on covariates (Example 5*), the results were not impacted, suggesting the validity of the results under dependent censoring.

[Supplemental Material, Table 2 about here.]

Table S2 reports the average computing time under Example 1 by various screening methods. It shows that the IPOD procedure is on par with the competing methods, but more computationally efficient than SII and CRIS, the nonparametric competitors.

References

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics* **17,** 1157–1167.

Ni, L. and Fang, F. (2016). Entropy-based model-free feature screening for ultrahigh-dimensional multiclass classification. *Journal of Nonparametric Statistics* **28,** 515–530.

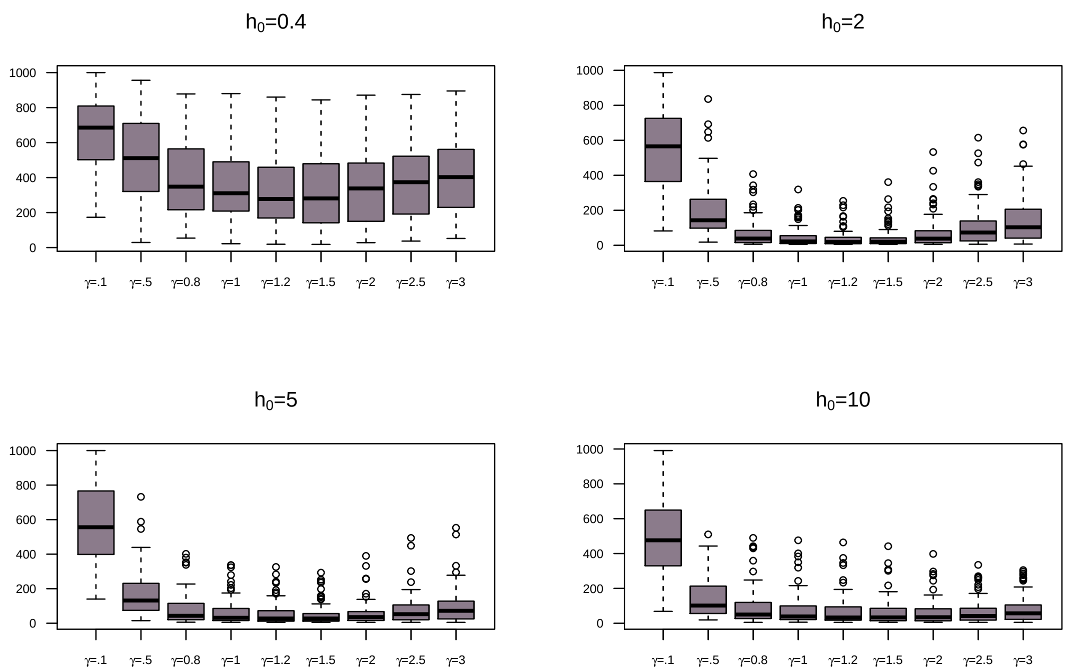

Figure S1: The boxplots of MMS obtained from IPOD with various γ's and bandwidths under Example 1.

Table S1: Comparisons of competing methods with $(n, p) = (500, 1000)$ in terms of MMS (with interquartile range in parentheses), TPR, and PIT

Method	**MMS**	**TPR**	**PIT**	**MMS**	**TPR**	**PIT**
Example 5	CR=20%			CR=50%		
IPOD ($\gamma = .8$)	46 (73)	0.93	0.71	89 (153)	0.86	0.47
IPOD ($\gamma = 1$)	29 (51)	0.96	0.80	66 (116)	0.90	0.56
IPOD ($\gamma = 1.2$)	23 (42)	0.97	0.86	49 (83)	0.92	0.66
PSIS	6 (5)	1.00	0.99	14 (28)	0.98	0.90
CRIS	7 (6)	1.00	0.98	30 (70)	0.94	0.74
CS	5 (1)	1.00	1.00	8 (10)	0.99	0.96
SII	13 (21)	0.99	0.94	20 (31)	0.98	0.90
Example 5*	CR=20%			CR=50%		
IPOD ($\gamma = 0.8$)	46 (63)	0.94	0.70	100 (162)	0.85	0.44
IPOD ($\gamma = 1$)	32 (45)	0.96	0.81	70(124)	0.89	0.54
IPOD ($\gamma = 1.2$)	23 (47)	0.97	0.85	58 (98)	0.91	0.63
PSIS	6 (7)	1.00	0.98	15 (27)	0.98	0.88
CRIS	7 (9)	1.00	0.98	30 (62)	0.95	0.78
CS	5 (1)	1.00	1.00	7 (9)	0.99	0.97
SII	24 (69)	0.95	0.77	273 (330)	0.70	0.15

Table S2: Average runtime (seconds) of different screening methods in Example 1 on a CPU with 2.9 GHz Intel Core i5 and 8GB of memory

	PSIS	CS	CRIS	SII	IPOD
$(n, p) = (500, 1000)$	3.59	3.17	127.55	356.92	5.60
$(n, p) = (300, 10000)$	29.21	28.74	458.28	1259.82	40.01