Supplement to "De-biased lasso for stratified Cox models with application to the national kidney transplant data"

Lu Xia[a], Bin Nan[b], and Yi Li[c]

[a]Department of Biostatistics, University of Washington, Seattle, WA; [b]Department of Statistics, University of Californina, Irvine, CA; [c]Department of Biostatistics, University of Michigan, Ann Arbor, MI

For completeness of presentation, we first provide some useful technical lemmas and their proofs, and then give the proofs of the main theorems in this supplementary material. Since corollaries are direct results of corresponding theorems, their proofs are straightforward thus omitted.

1 Technical Lemmas

We first introduce some additional notation in counting processes. For the ith subject in the kth stratum, define the counting process $N_{ki}(t) = 1(Y_{ki} \le t, \delta_{ki} = 1)$. The corresponding intensity process $A_{ki}(t;\beta) = \int_0^t 1(Y_{ki} \ge s)\exp(X_{ki}^T\beta)d\Lambda_{0k}(s)$, where $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$ is the baseline cumulative hazard function for the kth stratum, $k = 1, \cdots, K$, $i = 1, \cdots, n_k$. Let $M_{ki}(t;\beta) = N_{ki}(t) - A_{ki}(t;\beta)$, then $M_{ki}(t;\beta^0)$ is a martingale with respect to the filtration $\mathcal{F}_{ki}(t) = \sigma\{N_{ki}(s), 1(Y_{ki} \ge s), X_{ki} : s \in (0,t]\}$.

Recall that the stratum-specific weighted covariate process $\widehat{\eta}_k(t;\beta) = \widehat{\mu}_{1k}(t;\beta)/\widehat{\mu}_{0k}(t;\beta)$, where $\widehat{\mu}_{rk}(t;\beta) = (1/n_k)\sum_{i=1}^{n_k} 1(Y_{ki} \ge t)X_{ki}^{\otimes r}\exp\{X_{ki}^T\beta\}$. Their population-level counterparts are $\mu_{rk}(t;\beta) = E[1(Y_{k1} \ge t)X_{k1}^{\otimes r}\exp\{X_{k1}^T\beta\}]$ and $\eta_{k0}(t;\beta) = \mu_{1k}(t;\beta)/\mu_{0k}(t;\beta)$, $r = 0, 1, 2$, $k = 1, \cdots, K$. It is easily seen that the process $\{X_{ki} - \widehat{\eta}_k(t;\beta^0)\}$ is predictable with respect to the filtration $\mathcal{F}(t) = \sigma\{N_{ki}(s), 1(Y_{ki} \ge s), X_{ki} : s \in (0,t], k = 1, \cdots, K,\ i = 1, \cdots, n_k\}$.

Lemma S.1. *Under Assumptions B.1–B.3, for $k = 1, \cdots, K$, we have*

$$\sup_{t\in[0,\tau]} |\widehat{\mu}_{0k}(t;\beta^0) - \mu_{0k}(t;\beta^0)| = \mathcal{O}_P(\sqrt{\log(p)/n_k}),$$
$$\sup_{t\in[0,\tau]} \|\widehat{\mu}_{1k}(t;\beta^0) - \mu_{1k}(t;\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_k}),$$
$$\sup_{t\in[0,\tau]} \|\widehat{\eta}_{k}(t;\beta^0) - \eta_{k0}(t;\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_k}).$$

Lemma S.1 is simply the result of Lemma A1 in Xia et al. (2021) applied to each of the K strata. We omit its proof here.

Lemma S.2. *Assume $p^2\log(p)/n_{min} \to 0$. Under Assumptions B.1–B.5, for any $c \in \mathbb{R}^p$ such that $\|c\|_2 = 1$ and $\|c\|_1 \le a_*$ with some absolute constant $a_* < \infty$, we have*

$$\frac{\sqrt{N}c^T\Theta_{\beta^0}\dot{\ell}(\beta^0)}{\sqrt{c^T\Theta_{\beta^0}c}} \xrightarrow{\mathcal{D}} N(0,1).$$

Proof of Lemma S.2. We rewrite

$$\begin{aligned}
\frac{-\sqrt{N}c^T\Theta_{\beta^0}\dot{\ell}(\beta^0)}{\sqrt{c^T\Theta_{\beta^0}c}} &= \frac{1}{\sqrt{N}}\sum_{k=1}^K\sum_{i=1}^{n_k}\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_{ki} - \frac{\widehat{\mu}_{1k}(Y_{ki};\beta^0)}{\widehat{\mu}_{0k}(Y_{ki};\beta^0)}\right\}\delta_{ki} \\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^K\sum_{i=1}^{n_k}\int_0^\tau\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_{ki} - \frac{\widehat{\mu}_{1k}(t;\beta^0)}{\widehat{\mu}_{0k}(t;\beta^0)}\right\}dN_{ki}(t) \\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^K\sum_{i=1}^{n_k}\int_0^\tau\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_{ki} - \frac{\widehat{\mu}_{1k}(t;\beta^0)}{\widehat{\mu}_{0k}(t;\beta^0)}\right\}dM_{ki}(t). \qquad \text{(S.1)}
\end{aligned}$$

Denote $U(t) = \frac{1}{\sqrt{N}}\sum_{k=1}^K\sum_{i=1}^{n_k}\int_0^t\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_{ki} - \frac{\widehat{\mu}_{1k}(s;\beta^0)}{\widehat{\mu}_{0k}(s;\beta^0)}\right\}dM_{ki}(s)$. Then the variation process for $U(t)$ is

$$\begin{aligned}
\langle U\rangle(t) &= \sum_{k=1}^K\sum_{i=1}^{n_k}\frac{1}{N}\int_0^t\frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0}c}\left\{X_{ki} - \widehat{\eta}_k(u;\beta^0)\right\}^{\otimes 2}1(Y_{ki}\ge u)e^{X_{ki}^T\beta^0}d\Lambda_{0k}(u)\Theta_{\beta^0}c \\
&= \frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0}c}\left[\sum_{k=1}^K\frac{n_k}{N}\int_0^t\left\{\widehat{\mu}_{2k}(u;\beta^0) - \frac{\widehat{\mu}_{1k}(u;\beta^0)\widehat{\mu}_{1k}^T(u;\beta^0)}{\widehat{\mu}_{0k}(u;\beta^0)}\right\}d\Lambda_{0k}(u)\right]\Theta_{\beta^0}c. \qquad \text{(S.2)}
\end{aligned}$$

By Assumption B.4, following the same lines in the proof of Lemma A2 in Xia et al. (2021), we have

$$
\begin{aligned}
&\frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0}c}\left[\int_0^t\left\{\widehat{\mu}_{2k}(u;\beta^0)-\frac{\widehat{\mu}_{1k}(u;\beta^0)\widehat{\mu}_{1k}^T(u;\beta^0)}{\widehat{\mu}_{0k}(u;\beta^0)}\right\}d\Lambda_{0k}(u)\right]\Theta_{\beta^0}c\\
=&\frac{c^T\Theta_{\beta^0}}{c^T\Theta_{\beta^0}c}\left[\int_0^t\left\{\mu_{2k}(u;\beta^0)-\frac{\mu_{1k}(u;\beta^0)\mu_{1k}^T(u;\beta^0)}{\mu_{0k}(u;\beta^0)}\right\}d\Lambda_{0k}(u)\right]\Theta_{\beta^0}c+o_P(1)\\
\to& v_k(t;c).
\end{aligned}
$$

Since $n_k/N \to r_k$, then $\langle U\rangle(t) \to_P \sum_{k=1}^K r_k v_k(t;c)$.

For any $\epsilon > 0$, define $G_{ki}(u) = \frac{1}{\sqrt{N}}\frac{c^T\Theta_{\beta^0}}{\sqrt{c^T\Theta_{\beta^0}c}}\left\{X_{ki}-\frac{\widehat{\mu}_{1k}(u;\beta^0)}{\widehat{\mu}_{0k}(u;\beta^0)}\right\}$ and the truncated process $U_\epsilon(t) = \sum_{k=1}^K\sum_{i=1}^{n_k}\int_0^t G_{ki}(u)1(|G_{ki}(u)|>\epsilon)dM_{ki}(u)$. The variation process of $U_\epsilon(t)$ is

$$
\langle U_\epsilon\rangle(t) = \sum_{k=1}^K\sum_{i=1}^{n_k}\int_0^t G_{ki}^2(u)1(|G_{ki}(u)|>\epsilon)dA_{ki}(u),
$$

where $dA_{ki}(u) = 1(Y_{ki}\geq u)e^{X_{ki}^T\beta^0}d\Lambda_{0k}(u)$. Since

$$
|\sqrt{N}G_{ki}(u)| \leq a_*\|\Theta_{\beta^0}\|_{1,1}2M\lambda_{\min}^{-1/2}(\Theta_{\beta^0}) = \mathcal{O}(\sqrt{p}),
$$

then $1(|G_{ki}(u)|>\epsilon) = 0$ almost surely as $p/N \to 0$. So $\langle U_\epsilon\rangle(t) \to_P 0$. By the martingale central limit theorem, we obtain the desirable result.

□

Lemma S.3. *Under Assumptions B.1–B.5, for* $\lambda \asymp \sqrt{\log(p)/n_{min}}$*, the lasso estimator* $\widehat{\beta}$ *satisfies*

$$
\|\widehat{\beta}-\beta^0\|_1 = \mathcal{O}_P(s_0\lambda),\ \frac{1}{N}\sum_{k=1}^K\sum_{i=1}^{n_k}|X_{ki}^T(\widehat{\beta}-\beta^0)|^2 = \mathcal{O}_P(s_0\lambda^2).
$$

Proof of Lemma S.3. This result is from the proof in Kong and Nan (2014), with minor modifications as follows. An intermediate replacement for the negative log-likelihood in the kth stratum

$$
\ell^{(k)}(\beta) = -\frac{1}{n_k}\sum_{i=1}^{n_k}\left[\beta^TX_{ki}-\log\left\{\frac{1}{n_k}\sum_{j=1}^{n_k}1(Y_{kj}\geq Y_{ki})\exp(\beta^TX_{kj})\right\}\right]\delta_{ki}
$$

can be defined as

$$\widetilde{\ell}^{(k)}(\beta) = -\frac{1}{n_k}\sum_{i=1}^{n_k}\left\{\beta^T X_{ki} - \log \mu_{0k}(Y_{ki};\beta)\right\}\delta_{ki},$$

which is a sum of n_k independent and identically distributed terms. The target parameter is

$$\bar{\beta} = \arg\min_{\beta} E\left\{\sum_{k=1}^{K}\frac{n_k}{N}\widetilde{\ell}^{(k)}(\beta)\right\}.$$

Then the excess risk for any given β is

$$\mathcal{E}(\beta) = E\left\{\sum_{k=1}^{K}\frac{n_k}{N}\widetilde{\ell}^{(k)}(\beta)\right\} - E\left\{\sum_{k=1}^{K}\frac{n_k}{N}\widetilde{\ell}^{(k)}(\bar{\beta})\right\}.$$

We refer remaining details to Kong and Nan (2014). □

Lemma S.4. *Under Assumptions B.1–B.5, assume* $\lambda \asymp \sqrt{\log(p)/n_{min}}$*, then it holds with probability going to one that* $\|\Theta_{\beta^0}\widehat{\Sigma} - I_p\|_\infty \le \gamma$*, with* $\gamma \asymp \|\Theta_{\beta^0}\|_{1,1}\{\max_{1\le k\le K}|n_k/N - r_k| + s_0\lambda\}$.

Lemma S.4 shows that when $\gamma \asymp \|\Theta_{\beta^0}\|_{1,1}\{\max_{1\le k\le K}|n_k/N - r_k| + s_0\lambda\}$, the jth row, $j = 1,\cdots,p$, of Θ_{β^0} will be feasible in the constraint of the corresponding quadratic programming problem with probability going to one.

Proof of Lemma S.4. We first derive the rate for $\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty$. Note that

$$\begin{aligned}
&\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty \\
&\quad \le \left\|\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\left[\{X_{ki} - \widehat{\eta}_k(t;\widehat{\beta})\}^{\otimes 2} - \{X_{ki} - \eta_{k0}(t;\beta^0)\}^{\otimes 2}\right]dN_{ki}(t)\right\|_\infty \\
&\qquad + \left\|\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\{X_{ki} - \eta_{k0}(t;\beta^0)\}^{\otimes 2}dN_{ki}(t) - \Sigma_{\beta^0}\right\|_\infty \\
&\quad \equiv a_{N1} + a_{N2}.
\end{aligned}$$

Due to the boundness Assumption B.1,

$$
\begin{aligned}
a_{N1} \leq & \left\| \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \{X_{ki} - \widehat{\eta}_k(t;\widehat{\beta})\}\{\eta_{k0}(t;\beta^0) - \widehat{\eta}_k(t;\widehat{\beta})\}^T dN_{ki}(t) \right\|_\infty \\
& + \left\| \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \{\eta_{k0}(t;\beta^0) - \widehat{\eta}_k(t;\widehat{\beta})\}\{X_{ki} - \eta_{k0}(t;\beta^0)\}^T dN_{ki}(t) \right\|_\infty \\
\leq & \frac{4M}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \|\eta_{k0}(t;\beta^0) - \widehat{\eta}_k(t;\widehat{\beta})\|_\infty dN_{ki}(t) \\
\leq & \frac{4M}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \|\eta_{k0}(t;\beta^0) - \widehat{\eta}_k(t;\beta^0)\|_\infty dN_{ki}(t) \\
& + \frac{4M}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \|\widehat{\eta}_k(t;\beta^0) - \widehat{\eta}_k(t;\widehat{\beta})\|_\infty dN_{ki}(t) \\
\leq & 4M\mathcal{O}_P(\sqrt{\log(p)/n_{min}}) + 4M\mathcal{O}_P(s_0\lambda) = \mathcal{O}_P(s_0\lambda),
\end{aligned}
$$

where the last inequality is a result of Lemma S.1 and the fact that $\sup_{t\in[0,\tau]} \|\widehat{\eta}_k(t;\beta^0) - \widehat{\eta}_k(t;\widehat{\beta})\|_\infty = \mathcal{O}_P(\|\widehat{\beta} - \beta^0\|_1) = \mathcal{O}_P(s_0\lambda)$ (see the proof of Lemma A4 in Xia et al. (2021)). Since $\Sigma_{\beta^0} = \sum_{k=1}^K r_k \Sigma_{\beta^0,k}$,

$$
\begin{aligned}
a_{N2} \leq & \left\| \sum_{k=1}^{K} \frac{n_k}{N} \left[\frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^\tau \{X_{ki} - \eta_{k0}(t;\beta^0)\}^{\otimes 2} dN_{ki}(t) - \Sigma_{\beta^0,k} \right] \right\|_\infty \\
& + \left\| \sum_{k=1}^{K} \left(\frac{n_k}{N} - r_k \right) \Sigma_{\beta^0,k} \right\|_\infty \\
\leq & \sum_{k=1}^{K} \frac{n_k}{N} \left\| \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^\tau \{X_{ki} - \eta_{k0}(t;\beta^0)\}^{\otimes 2} dN_{ki}(t) - \Sigma_{\beta^0,k} \right\|_\infty + \left\| \sum_{k=1}^{K} \left(\frac{n_k}{N} - r_k \right) \Sigma_{\beta^0,k} \right\|_\infty .
\end{aligned}
$$

The proof of Lemma A4 in Xia et al. (2021) shows that, for $k = 1, \cdots, K$,

$$
\left\| \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^\tau \{X_{ki} - \eta_{k0}(t;\beta^0)\}^{\otimes 2} dN_{ki}(t) - \Sigma_{\beta^0,k} \right\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_k})
$$

by Hoeffding's concentration inequality. So $a_{N2} = \mathcal{O}_P(\sqrt{\log(p)/n_{min}}) + \mathcal{O}(\max_k |n_k/N - r_k|)$. Then, combining the bounds on a_{N1} and a_{N2}, $\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty = \mathcal{O}_P(s_0\lambda + \max_k |n_k/N - r_k|)$.

Finally, it is easy to see that

$$\|\Theta_{\beta^0}\widehat{\Sigma} - I_p\|_\infty = \|\Theta_{\beta^0}(\widehat{\Sigma} - \Sigma_{\beta^0})\|_\infty \leq \|\Theta_{\beta^0}\|_{1,1}\|\widehat{\Sigma} - \Sigma_{\beta^0}\|_\infty,$$

and $\|\Theta_{\beta^0}\widehat{\Sigma} - I_p\|_\infty = \mathcal{O}_P(\|\Theta_{\beta^0}\|_{1,1}\{s_0\lambda + \max_k |n_k/N - r_k|\})$. □

Lemma S.5. *Under the assumptions in Lemma S.4, if we further assume for some constant* $\epsilon' \in (0,1)$*,* $\limsup_{n_{min}\to\infty} p\gamma \leq 1-\epsilon'$*, then we have* $\|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty = \mathcal{O}_P(\gamma\|\Theta_{\beta^0}\|_{1,1})$*.*

The proof of Lemma S.5 follows that of Lemma A5 in Xia et al. (2021), thus is omitted.

Lemma S.6. *Under Assumptions B.1–B.3 and B.5, for each* $t > 0$*,*

$$P(\|\dot{\ell}(\beta^0)\|_\infty > t) \leq 2Kpe^{-n_{min}t^2/(8M^2)}.$$

Proof of Lemma S.6. Since $\dot{\ell}(\beta^0) = \sum_{k=1}^K \frac{n_k}{N}\dot{\ell}^{(k)}(\beta^0)$, we have

$$\begin{aligned} P\left(\|\dot{\ell}(\beta^0)\|_\infty > t\right) &\leq P\left(\sum_{k=1}^K \frac{n_k}{N}\|\dot{\ell}^{(k)}(\beta^0)\|_\infty > t\right) \\ &\leq \sum_{k=1}^K P\left(\|\dot{\ell}^{(k)}(\beta^0)\|_\infty > t\right) \\ &\leq \sum_{k=1}^K 2pe^{-n_k t^2/(8M^2)}. \end{aligned}$$

Note that $\|X_{ki} - \widehat{\eta}_k(t;\beta^0)\|_\infty \leq 2M$ holds uniformly for all k and i. Then the last inequality is a direct result of Lemma 3.3(ii) in Huang et al. (2013) when applied to each of the K strata. □

2 Proofs of Main Theorems

Proof of Theorem 3.1. Let $\dot{\ell}_j(\beta)$ be the jth element of the derivative $\dot{\ell}(\beta)$. By the mean value theorem, there exists $\widetilde{\beta}^{(j)}$ between $\widehat{\beta}$ and β^0 such that $\dot{\ell}_j(\widehat{\beta}) - \dot{\ell}_j(\beta^0) = \frac{\partial \dot{\ell}_j(\beta)}{\partial \beta^T}\Big|_{\beta=\widetilde{\beta}^{(j)}}(\widehat{\beta} - \beta^0)$. Denote the $p \times p$ matrix $D = \left(\frac{\partial \dot{\ell}_1(\beta)}{\partial \beta}\Big|_{\beta=\widetilde{\beta}^{(1)}}, \cdots, \frac{\partial \dot{\ell}_p(\beta)}{\partial \beta}\Big|_{\beta=\widetilde{\beta}^{(p)}}\right)^T$. By the definition of the de-biased estimator $\widehat{b}$, we may decompose $c^T(\widehat{b}-\beta^0)$ as

$$
\begin{aligned}
c^T(\widehat{b}-\beta^0) &= -c^T\Theta_{\beta^0}\dot{\ell}(\beta^0) - c^T(\widehat{\Theta}-\Theta_{\beta^0})\dot{\ell}(\beta^0) \\
&\quad - c^T(\widehat{\Theta}\widehat{\Sigma}-I_p)(\widehat{\beta}-\beta^0) + c^T\widehat{\Theta}(\widehat{\Sigma}-D)(\widehat{\beta}-\beta^0) \\
&= -c^T\Theta_{\beta^0}\dot{\ell}(\beta^0) + (i) + (ii) + (iii),
\end{aligned}
$$

where $(i) = -c^T(\widehat{\Theta}-\Theta_{\beta^0})\dot{\ell}(\beta^0)$, $(ii) = -c^T(\widehat{\Theta}\widehat{\Sigma}-I_p)(\widehat{\beta}-\beta^0)$ and $(iii) = c^T\widehat{\Theta}(\widehat{\Sigma}-D)(\widehat{\beta}-\beta^0)$.

We first show $\sqrt{N}(i) = o_P(1)$ and $\sqrt{N}(ii) = o_P(1)$. By Lemma S.5 and Lemma S.6,

$$
\begin{aligned}
|\sqrt{N}(i)| &\le \sqrt{N}\|c\|_1 \cdot \|\widehat{\Theta}-\Theta_{\beta^0}\|_{\infty,\infty} \cdot \|\dot{\ell}(\beta^0)\|_\infty \\
&\le \sqrt{N}a_*\mathcal{O}_P(p\gamma\|\Theta_{\beta^0}\|_{1,1})\mathcal{O}_P(\sqrt{\log(p)/n_{min}}) \\
&= \mathcal{O}_P(\|\Theta_{\beta^0}\|_{1,1}p\gamma\sqrt{\log(p)}) \\
&= o_P(1),
\end{aligned}
$$

where the last equality is a direct result of the assumption that $\|\Theta_{\beta^0}\|_{1,1}^2\{\max_k |n_k/N - r_k| + s_0\lambda\}p\sqrt{\log(p)} \to 0$ when $\lambda \asymp \sqrt{\log(p)/n_{min}}$. By Lemma S.3,

$$
\begin{aligned}
|\sqrt{N}(ii)| &\le \sqrt{N}\|c\|_1\|(\widehat{\Theta}\widehat{\Sigma}-I_p)(\widehat{\beta}-\beta^0)\|_\infty \\
&\le \sqrt{N}a_*\|\widehat{\Theta}\widehat{\Sigma}-I_p\|_\infty\|\widehat{\beta}-\beta^0\|_1 \\
&\le \sqrt{N}a_*\gamma\|\widehat{\beta}-\beta^0\|_1 \\
&= \mathcal{O}_P(\sqrt{N}\gamma s_0\lambda) \\
&= o_P(1).
\end{aligned}
$$

We then show that $\sqrt{N}(iii) = o_P(1)$. Note that $\widehat{\Sigma} - D = (\widehat{\Sigma}-\Sigma_{\beta^0}) + (\Sigma_{\beta^0} - \ddot{\ell}(\beta^0)) + (\ddot{\ell}(\beta^0) - D)$. By the proof of Lemma S.4, we see that with $\lambda \asymp \sqrt{\log(p)/n_{min}}$, $\|\widehat{\Sigma}-\Sigma_{\beta^0}\|_\infty = \mathcal{O}_P(s_0\lambda + \max_k|n_k/N - r_k|)$. Based on the proof of Theorem 1 in Xia et al. (2021), for each stratum, $\|\ddot{\ell}^{(k)}(\beta^0) - D^{(k)}\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_k})$, where $D^{(k)} = \left(\frac{\partial\dot{\ell}_1^{(k)}(\beta)}{\partial\beta}\Big|_{\beta=\widetilde{\beta}^{(1)}}, \cdots, \frac{\partial\dot{\ell}_p^{(k)}(\beta)}{\partial\beta}\Big|_{\beta=\widetilde{\beta}^{(p)}}\right)^T$. Since the overall negative log partial likelihood $\ell(\beta) = \sum_{k=1}^K \frac{n_k}{N}\ell^{(k)}(\beta)$, and $D = \sum_{k=1}^K \frac{n_k}{N}D^{(k)}$, then $\|\ddot{\ell}(\beta^0) - D\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_{min}})$. Also, $\|\Sigma_{\beta^0,k} - \ddot{\ell}^{(k)}(\beta^0)\|_\infty = \mathcal{O}_P(\sqrt{\log(p)/n_k})$ by the proof of Theorem 1 in Xia et al. (2021).

Then

$$
\begin{aligned}
\|\Sigma_{\beta^0} - \ddot{\ell}(\beta^0)\|_\infty &\le \left\| \sum_{k=1}^K r_k \Sigma_{\beta^0,k} - \sum_{k=1}^K \frac{n_k}{N} \Sigma_{\beta^0,k} \right\|_\infty + \left\| \sum_{k=1}^K \frac{n_k}{N} \Sigma_{\beta^0,k} - \sum_{k=1}^K \frac{n_k}{N} \ddot{\ell}^{(k)}(\beta^0) \right\|_\infty \\
&\le K \max_k (|n_k/N - r_k| \|\Sigma_{\beta^0,k}\|_\infty) + K\mathcal{O}_P(\sqrt{\log(p)/n_{min}}) \\
&= \mathcal{O}_P(\max_k |n_k/N - r_k| + \sqrt{\log(p)/n_{min}}).
\end{aligned}
$$

Therefore, for $\lambda \asymp \sqrt{\log(p)/n_{min}}$, $\|\widehat{\Theta} - D\|_\infty = \mathcal{O}_P(s_0\lambda + \max_k |n_k/N - r_k|)$, and

$$
\begin{aligned}
|\sqrt{N}(iii)| &\le \sqrt{N}\|c\|_1 \|\widehat{\Theta}\|_{\infty,\infty} \|\widehat{\Sigma} - D\|_\infty \|\widehat{\beta} - \beta^0\|_1 \\
&\le \mathcal{O}_P\left(\sqrt{N}\|\Theta_{\beta^0}\|_{1,1}(s_0\lambda + \max_k |n_k/N - r_k|)\right) s_0\lambda \\
&\le \mathcal{O}_P\left(\sqrt{N/n_{min}}\|\Theta_{\beta^0}\|_{1,1}(s_0\lambda + \max_k |n_k/N - r_k|) p\sqrt{\log(p)}\right) \\
&= o_P(1).
\end{aligned}
$$

Finally, for the variance,

$$
\begin{aligned}
|c^T(\widehat{\Theta} - \Theta_{\beta^0})c| &\le \|c\|_1^2 \|\widehat{\Theta} - \Theta_{\beta^0}\|_\infty \\
&\le a_*^2 \mathcal{O}_P(\gamma \|\Theta_{\beta^0}\|_{1,1}) = o_P(1).
\end{aligned}
$$

By Slutsky's theorem and Lemma S.2, $\sqrt{n}c^T(\widehat{b} - \beta^0)/(c^T\widehat{\Theta}c)^{1/2} \xrightarrow{\mathcal{D}} N(0,1)$. □

Sketch proof of Theorem 3.4. Theorem 3.4 can be easily proved using Cramér-Wold device. For any $\omega \in \mathbb{R}^p$, since the dimension of ω is a fixed integer, we can invoke Theorem 3.1 by taking $c = J^T\omega$. Note that $\|J^T\omega\|_1 \le \|J^T\|_{1,1}\|\omega\|_1 = \|J\|_{\infty,\infty}\|\omega\|_1 = \mathcal{O}(1)$. □

References

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C.-H. (2013). Oracle inequalities for the lasso in the Cox model. *Annals of Statistics*, 41(3):1142–1165.

Kong, S. and Nan, B. (2014). Non-asymptotic oracle inequalities for the high-dimensional Cox regression via lasso. *Statistica Sinica*, 24(1):25–42.

Xia, L., Nan, B., and Li, Y. (2021). Statistical inference for Cox proportional hazards models with a diverging number of covariates. *arXiv preprint arXiv:2106.03244*.