THE L_q-NORM LEARNING FOR ULTRAHIGH-DIMENSIONAL SURVIVAL DATA : AN INTEGRATIVE FRAMEWORK

H. G. Hong[†], X. Chen[‡], J. Kang[*] and Y. Li[*]

[†] *Michigan State University, USA*

[‡] *Southwestern University of Finance and Economics, China,*

[*] *University of Michigan, USA*

Supplementary Material

S1. Underlying probability space

Let $(\Omega, \mathcal{F}, P)$ be the probability space that underlies all the random variables in the paper. Here, Ω is the sample space, $\mathcal{F}$ is the σ-algebra and P is the probability measure. Let $\omega \in \Omega$ denote a sample point. The ensuing proofs need the results of the strong convergence of Kaplan–Meier estimators and empirical quantiles for each covariate. We consider the individual subsets of Ω in which the convergence results hold. Specifically, Theorem 5.9 of Shao (1999) and Theorem 3.1 of Foldes and Rejto (1981) indicate that $\hat{S}(t)$, $\hat{S}(t \mid X_j)$, $\hat{Q}_{j(k)}$ and $\hat{Q}_{ju(k)}$ are strong consistent estimators of $S(t)$, $S(t \mid X_j)$, $Q_{j(k)}$ and $Q_{ju(k)}$, for $1 \leq j \leq p$. That is, there exists an Ω_j such that on Ω_j, $\sup_{0<t<\tau} |\hat{S}(t \mid X_j) - S(t \mid X_j)| = o(1)$, $\hat{Q}_{j(k)} - Q_{j(k)} = o(1)$ and $\hat{Q}_{ju(k)} - Q_{ju(k)} = o(1)$, where $\Omega_j \subset \Omega$ with $P(\Omega_j) = 1$ and $\tau = \inf\{t : P(T > t) = 0\}$. In addition, there exists an $\Omega_0 \subset \Omega$ where $P(\Omega_0) = 1$, such that on Ω_0, $\sup_{0<t<\tau} |\hat{S}(t) - S(t)| = o(1)$. Take $\Omega_* = \bigcap_{j=0}^{p} \Omega_j$. Then it follows that

$P(\Omega_*) = 1$ and $P(\Omega_*^c) = 0$, where Ω_*^c is the complement of Ω_*. All the events mentioned in the following proofs should implicitly be viewed as the intersections with Ω_*, which ensures that the aforementioned strong convergence results hold.

S2. Lemmas and proofs

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

Lemma 1. *Let $\tau = \inf\{t : P(T > t) = 0\}$. For a categorical covariate X_j with K_j categories, let $\hat{S}(t \mid X_j = k)$ be the Kaplan–Meier estimator of the conditional survival function within the category of $X_j = k, k = 1, \ldots, K_j$. There exist $d_0 > 0$, $d_1 > 0$, κ and v under Condition 2, for any $\epsilon > 0$ and n sufficiently large,*

$$P\left\{\max_{1\le k\le K_j}\sup_{t\in[0,\tau]}|\hat{S}(t \mid X_j = k) - S(t \mid X_j = k)| > \epsilon\right\} \le d_1 K \exp(-d_0\epsilon^2 n^{1-3\kappa}),$$

where $K = \max_{1\le j\le p} K_j$.

Proof. By the inequality in the last paragraph on page 1161 of Dabrowska (1989), there exist positive constants d_0 and d_1 not depending on ϵ, n and $S(t \mid X_j)$, such that

$$\begin{aligned} & P(\max_k \sup_{t\in[0,\tau]}|\hat{S}(t \mid X_j = k) - S(t \mid X_j = k)| > \epsilon) \\ \le\ & d_1 K_j \exp(-d_0\epsilon^2 \min_k n_k K_j^{-2}) \\ \le\ & d_1 K \exp(-d_0\epsilon^2 \min_k n_k K^{-2}) \end{aligned}$$

where n_k is the number of subjects within $X_j = k$. The result follows since $\min_k n_k \ge n/K = n^{1-\kappa}$ by Condition 2. □

Lemma 2. *Under the same constants and conditions for Lemma 1, for any $q \ge 1$, $\epsilon > 0$ and*

n sufficiently large,

$$P(|\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| > \epsilon) \leq 2d_1 K \exp\left(-\frac{d_0}{16G(q)^2}\epsilon^2 n^{1-3\kappa}\right),$$

where $G(q) = 2\{\min_k \int_0^\tau S^q(u \mid X_j = k) dS(u)/4\}^{(1/q)-1}$.

Proof. By the Minkowski inequality and the definition of τ,

$$\begin{aligned}
& |\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| \\
= & \left| \max_{k_1,k_2} \left\{ -\int_0^\infty |\hat{S}(u \mid X_j = k_1) - \hat{S}(u \mid X_j = k_2)|^q d\hat{S}(u) \right\}^{1/q} \right. \\
& \left. - \max_{k_1,k_2} \left\{ -\int_0^\infty |S(u \mid X_j = k_1) - S(u \mid X_j = k_2)|^q \, dS(u) \right\}^{1/q} \right| \\
\leq & \max_{k_1} \left| \left\{ \int_0^\infty |\hat{S}(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\infty |S(u \mid X_j = k_1)|^q \, dS(u) \right\}^{1/q} \right| \\
& + \max_{k_2} \left| \left\{ \int_0^\infty |\hat{S}(u \mid X_j = k_2)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\infty |S(u \mid X_j = k_2)|^q dS(u) \right\}^{1/q} \right| \\
= & \max_{k_1} \left| \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q \, dS(u) \right\}^{1/q} \right| \\
& + \max_{k_2} \left| \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_2)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_2)|^q dS(u) \right\}^{1/q} \right| \\
=: & I_{11} + I_{12}.
\end{aligned}$$

We next bound I_{11} and I_{12} separately. We first define

$$\psi(z) = \left\{ \int_0^\tau |z[\hat{S}(u \mid X_j = k_1) - S(u \mid X_j = k_1)] + S(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q}.$$

Here, $\psi(z)$ is continuous with respect to z on $z \in [0, 1]$ and

$$\psi(0) = \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} \text{ and } \psi(1) = \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q}.$$

We intent to apply the mean value theorem to bound $|\psi(1) - \psi(0)|$. Since $\psi(z)$ involves an absolute value, we need to compute the subgradient of $\psi(z)$, which is denoted by $\partial\psi(z)$.

Given $|\partial(|z|)| \leq 1$ and by the strong consistency of $\hat{S}(t \mid X_j)$ to $S(t \mid X_j)$ and $\hat{S}(t)$ to $S(t)$, for any $z \in [0,1]$, we have that

$$\begin{aligned}
|\partial\psi(z)| \;\leq\; & \left| \left[\int_0^\tau \left| |z\{\hat{S}(u \mid X_j = k_1) - S(u \mid X_j = k_1)\} + S(u \mid X_j = k_1)|^q d\hat{S}(u) \right]^{(1/q)-1} \right. \\
\times\; & \left[\int_0^\tau |z\{\hat{S}(u \mid X_j = k_1) - S(u \mid X_j = k_1)\} + S(u \mid X_j = k_1)|^{q-1} \right. \\
\times\; & \left. \left. \{\hat{S}(u \mid X_j = k_1) - S(u \mid X_j = k_1)\} \right| d\hat{S}(u) \right] \Bigg| \\
\leq\; & G(q) \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j = k_1) - S(t \mid X_j = k_1)|,
\end{aligned}$$

where $G(q) = 2\{\min_k \int_0^\tau S^q(u \mid X_j = k) dS(u)/4\}^{(1/q)-1}$.

Hence, by Rolle's mean value inequality theorem (Aussel et al., 1995),

$$|\psi(1) - \psi(0)| \leq G(q) \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j = k_1) - S(t \mid X_j = k_1)|.$$

Then, we have

$$\begin{aligned}
I_{11} \;=\; & \max_{k_1} \left| \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q \, dS(u) \right\}^{1/q} \right| \\
\leq\; & \max_{k_1} \left| \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q \, d\hat{S}(u) \right\}^{1/q} \right| \\
+\; & \max_{k_1} \left| \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_1)|^q \, dS(u) \right\}^{1/q} \right| \\
\leq\; & G(q) \times \max_{k_1} \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j = k_1) - S(t \mid X_j = k_1)| + \frac{\epsilon}{4},
\end{aligned}$$

when n is sufficiently large. Here, the last inequality involving $\epsilon/4$ stems from the uniform strong convergence of $\hat{S}(t)$ to $S(t)$ over $[0,\tau)$.

Similarly, we can obtain that

$$\begin{aligned}
I_{12} \;=\; & \max_{k_2} \left| \left\{ \int_0^\tau |\hat{S}(u \mid X_j = k_2)|^q d\hat{S}(u) \right\}^{1/q} - \left\{ \int_0^\tau |S(u \mid X_j = k_2)|^q \, dS(u) \right\}^{1/q} \right| \\
\leq\; & G(q) \times \max_{k_2} \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j = k_2) - S(t \mid X_j = k_2)| + \frac{\epsilon}{4}
\end{aligned}$$

for a sufficiently large n.

Therefore,

$$\begin{aligned}
P(|\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| > \epsilon) &\leq P(I_{11} > \epsilon/2) + P(I_{12} > \epsilon/2) \\
&\leq P\left\{ \max_{1 \leq k_1 \leq K_j} \sup_{t \in [0,\tau]} |\hat{S}(t \mid X_j = k_1) - S(t \mid X_j = k_1)| > \frac{\epsilon}{4G(q)} \right\} \\
&\quad + P\left\{ \max_{1 \leq k_2 \leq K_j} \sup_{t \in [0,\tau]} |\hat{S}(t \mid X_j = k_2) - S(t \mid X_j = k_2)| > \frac{\epsilon}{4G(q)} \right\} \\
&\leq 2d_1 K \exp\left(-\frac{d_0}{16G(q)^2} \epsilon^2 n^{1-3\kappa} \right).
\end{aligned}$$

□

Proof of Theorem 1. By Lemma 2, we have that

$$\begin{aligned}
P(\mathcal{M} \subset \widehat{\mathcal{M}}_1) &\geq P\left(|\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| \leq cn^{-v} \right) \\
&\geq P(\max_{1 \leq j \leq p} |\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| \leq cn^{-v}) \\
&\geq 1 - \sum_{j=1}^{p} P(|\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| > cn^{-v}) \\
&\geq 1 - \sum_{j=1}^{p} \left[2d_1 K \exp\left(-\frac{d_0 c^2}{16G(q)^2} n^{1-3\kappa-2v} \right) \right] \\
&= 1 - 2pd_1 c_0 n^{\kappa} \exp\left(-\frac{d_0 c^2}{16G(q)^2} n^{1-3\kappa-2v} \right) \\
&= 1 - c_2 p \exp(-c_1 n^{1-3\kappa-2v} + \kappa \log n),
\end{aligned}$$

where $c_2 = 2d_1 c_0$ and $c_1 = d_0 c^2 / 16G(q)^2$. □

Let $\hat{Q}_{j(k)}$ and $Q_{j(k)}$ be the empirical and theoretical $k/K_j \times 100$-th percentiles of X_j, for $k = 1, \ldots, K_j$. For notational simplicity, let $\hat{J}_k = [\hat{Q}_{j(k-1)}, \hat{Q}_{j(k)})$ and $J_k = [Q_{j(k-1)}, Q_{j(k)})$.

Lemma 3. *For a continuous covariate X_j, let $\hat{S}(t \mid X_j \in \hat{J}_k)$ be the Kaplan–Meier estimator of the conditional survival function within the subsamples of $X_j \in \hat{J}_k$. There exist constants*

$c_3 > 0$, $c_4 > 0$, κ and ρ under Condition 3, for sufficiently large n,

$$P\left\{\max_k \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j \in \hat{J}_k) - S(t \mid X_j \in J_k)| > \epsilon\right\} \leq d_3 K \exp(-d_2\epsilon^2 n^{1-3\kappa-2\rho}),$$

for any $1 \leq k \leq K_j$ and $K = \max_{1\leq j\leq p} K_j$.

Proof. By the strong consistency of $\hat{Q}_{j(k)}$ and the continuity of F_{X_j}, it follows that when n is sufficiently large

$$F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(\hat{Q}_{j(k-1)}) > 0.5\{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})\}$$

on Ω_* for $k = 1, \ldots, K_j$. Here, by convention, $\hat{Q}_{j(0)} = Q_{j(0)} = 0$ and $\hat{Q}_{j(K_j)} = Q_{j(K_j)} = \infty$.

Now for each $k = 1, \ldots, K_j$, by the mean value theorem,

$$\begin{aligned}
& |S(t \mid X_j \in \hat{J}_k) - S(t \mid X_j \in J_k)| \\
= & \left| \frac{P(T > t, X_j < \hat{Q}_{j(k)}) - P(T > t, X_j < \hat{Q}_{j(k-1)})}{F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(\hat{Q}_{j(k-1)})} \right. \\
& \left. - \frac{P(T > t, X_j < Q_{j(k)}) - P(T > t, X_j < Q_{j(k-1)})}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \right| \\
\leq & \left| \frac{P(T > t, X_j < \hat{Q}_{j(k)}) - P(T > t, X_j < \hat{Q}_{j(k-1)})}{F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(\hat{Q}_{j(k-1)})} \right. \\
& \left. - \frac{P(T > t, X_j < Q_{j(k)}) - P(T > t, X_j < Q_{j(k-1)})}{F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(\hat{Q}_{j(k-1)})} \right| \\
& + \left| \frac{P(T > t, X_j < Q_{j(k)}) - P(T > t, X_j < Q_{j(k-1)})}{F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(\hat{Q}_{j(k-1)})} \right. \\
& \left. - \frac{P(T > t, X_j < Q_{j(k)}) - P(T > t, X_j < Q_{j(k-1)})}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \right| \\
\leq & \frac{2|P(T > t, X_j < \hat{Q}_{j(k)}) - P(T > t, X_j < Q_{j(k)})|}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \\
& + \frac{2|P(T > t, X_j < \hat{Q}_{j(k-1)}) - P(T > t, X_j < Q_{j(k-1)})|}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \\
& + \frac{2|F_{X_j}(\hat{Q}_{j(k-1)}) - F_{X_j}(Q_{j(k-1)})|}{\{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})\}^2} + \frac{2|F_{X_j}(\hat{Q}_{j(k)}) - F_{X_j}(Q_{j(k)})|}{\{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})\}^2} \\
=: & \; I_{21} + I_{22} + I_{23} + I_{24}.
\end{aligned}$$

It is easy to show that $I_{21} = 0$ when $k = K_j$ as $\hat{Q}_{j(K_j)} = Q_{j(K_j)} = \infty$. Now for $k = 1, \ldots, K_j - 1$,

$$\begin{aligned} I_{21} &= \frac{2}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} |P(T > t, X_j < \hat{Q}_{j(k)}) - P(T > t, X_j < Q_{j(k)})| \\ &\leq \frac{2}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \left| \int_t^{\infty} f(s \mid Q^*_{j(k)}) f_{X_j}(Q^*_{j(k)}) ds \right| \max_k |\hat{Q}_{j(k)} - Q_{j(k)}| \\ &\leq \frac{2}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} f_{X_j}(Q^*_{j(k)}) \max_k |\hat{Q}_{j(k)} - Q_{j(k)}| \\ &\leq \frac{2}{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})} \max_x f_{X_j}(x) \max_k |\hat{Q}_{j(k)} - Q_{j(k)}|, \end{aligned}$$

where $Q^*_{j(k)}$ lies between $\hat{Q}_{j(k)}$ and $Q_{j(k)}$, for $k = 1, \ldots, K_j - 1$. By the strong consistency of $\hat{Q}_{j(k)}$, the continuity of f_{X_j} and Theorem 5.9 of Shao (1999), there exist positive constants b_{01} and b_{11} such that

$$\begin{aligned} P\left(I_{21} > \frac{\epsilon}{8}\right) &\leq P\left[\max_{1 \leq k \leq K_j - 1} |\hat{Q}_{j(k)} - Q_{j(k)}| > \frac{\epsilon\{F_{X_j}(Q_{j(k)}) - F_{X_j}(Q_{j(k-1)})\}}{16 \max_x f_{X_j}(x)}\right] \\ &\leq b_{11} K_j \exp(-b_{01} n \delta_\epsilon^2), \end{aligned}$$

where $\delta_\epsilon = \min_{1 \leq k \leq K_j - 1}\{F_{X_j}(Q_{j(k)} + \epsilon) - F(Q_{j(k)}), F(Q_{j(k)}) - F_{X_j}(Q_{j(k)} - \epsilon)\} \geq \min_{1 \leq k \leq K_j - 1} f(Q_{j(k)})\epsilon$. Hence, we have that $P(I_{21} > \epsilon/8) \leq b_{11} K \exp(-b_{01} c_3^2 n^{1-2\rho} \epsilon^2)$ by Condition 3.

Similarly, for $w = 2, 3, 4$, there exist constants b_{0w} and b_{1w} such that $P(I_{2w} > \epsilon/8) \leq$

$b_{1w}K\exp(-b_{0w}c_3^2n^{1-2\rho}\epsilon^2)$. Therefore,

$$\begin{aligned}
& P\{\max_k \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j \in \hat{J}_k) - S(t \mid X_j \in J_k)| > \epsilon\} \\
\leq\ & P\{\max_k \sup_{t\in[0,\tau]} |\hat{S}(t \mid X_j \in \hat{J}_k) - S(t \mid X_j \in \hat{J}_k)| > \epsilon/2\} \\
& +P\{\max_k \sup_{t\in[0,\tau]} |S(t \mid X_j \in \hat{J}_k) - S(t \mid X_j \in J_k)| > \epsilon/2\} \\
\leq\ & d_1K\exp\{-d_0(\epsilon/2)^2n^{1-3\kappa}\} + \sum_{w=1}^{4} P\left(I_{4w} > \frac{\epsilon}{8}\right) \\
\leq\ & d_1K\exp\left\{-\frac{d_0}{4}n^{1-3\kappa}\epsilon^2\right\} + \sum_{w=1}^{4} b_{1w}K\exp(-b_{0w}c_3^2n^{1-2\rho}\epsilon^2) \\
\leq\ & d_3K\exp(-d_2n^{1-3\kappa-2\rho}\epsilon^2),
\end{aligned}$$

where $d_3 = \max\{d_1, b_{11}, \ldots, b_{14}\}$ and $d_2 = \min\{d_0/4, b_{01}c_3^2, \ldots, b_{04}c_3^2\}$. □

Lemma 4. *If X_j is a continuous covariate, under Condition 3, there exist positive constants d_2' and d_3', for n sufficiently large,*

$$P(|\widehat{\Psi}_j^{(q)} - \Psi_j^{(q)}| > \epsilon) \leq d_3'K\exp(-d_2'\epsilon^2n^{1-3\kappa-2\rho}),$$

where $K = \max_{1\leq j\leq p} K_j$.

Proof. The proof of this lemma is similar to that of Lemma 2. By Lemma 3, the conclusion follows. □

Proof of Theorem 2. By Lemma 4, the proof of this theorem is similar to that of Theorem 1. □

For notational simplicity, we let $\hat{J}_{ur} = [\hat{Q}_{ju(r-1)}, \hat{Q}_{ju(r)})$ and $J_{ur} = [Q_{ju(r-1)}, Q_{ju(r)})$.

Lemma 5. *If X_j is a continuous covariate, there exist constants $\tilde{d}_0 > 0$, $\tilde{d}_1 > 0$, $\tilde{\kappa}$ and $\tilde{\rho}$ under Condition 5, for any $\epsilon > 0$ and n sufficiently large, we have that*

$$P(|\widetilde{\Psi}_j^{(q)} - \Psi_{jo}^{(q)}| > \epsilon) \leq \tilde{d}_1K\log n\exp(-\tilde{d}_0\epsilon^2n^{1-3\tilde{\kappa}-2\tilde{\rho}}/\log n),$$

where $K = \max_{1\le j\le p, 1\le u\le N} K_{ju}$.

Proof.

$$
\begin{aligned}
&|\widetilde{\Psi}_j^{(q)} - \Psi_{jo}^{(q)}| \le \sum_{u=1}^{N} |\widehat{\Psi}_{j,\Lambda_{ju}}^{(q)} - \Psi_{j,\Lambda_{juo}}^{(q)}| \\
&\le \sum_{u=1}^{N} \Bigg[\max_{k_1} \Bigg| \left\{ \int_0^{\tau} |\hat{S}(t \mid X_j \in \hat{J}_{uk_1})|^q d\hat{S}(t) \right\}^{1/q} - \left\{ \int_0^{\tau} |S(t \mid X_j \in J_{uk_1})|^q dS(t) \right\}^{1/q} \Bigg| \\
&+ \max_{k_2} \Bigg| \left\{ \int_0^{\tau} |\hat{S}(t \mid X_j \in \hat{J}_{uk_2})|^q d\hat{S}(t) \right\}^{1/q} - \left\{ \int_0^{\tau} |S(t \mid X_j \in J_{uk_2})|^q dS(t) \right\}^{1/q} \Bigg| \Bigg].
\end{aligned}
$$

The conclusion follows by using a proof similar to Lemma 2 and Lemma 4. □

Proof of Theorem 3. By Lemma 5, the proof is similar to that of Theorem 1. □

Proof of Theorem 4. Since q_l satisfies Condition 6, by Theorem 3, there exist constants $c_{2,l}$, $c_{3,l}$, κ_l, v_l and ρ_l such that

$$P\{\mathcal{M} \subset \widetilde{\mathcal{M}}^{(q_l)}\} \ge 1 - c_{3,l} p \log n \exp\{(-c_{2,l} n^{1-3\kappa_l - 2v_l - 2\rho_l}/\log n) + \kappa_l \log n\}.$$

Note that $\widetilde{\mathcal{M}}_h = \bigcup_{l=1}^{L} \widetilde{\mathcal{M}}^{(q_l)}$. Hence, we have $\widetilde{\mathcal{M}}^{(q_l)} \subset \widetilde{\mathcal{M}}_h$ and

$$
\begin{aligned}
P(\mathcal{M} \subset \widetilde{\mathcal{M}}_h) &\ge P(\mathcal{M} \subset \widetilde{\mathcal{M}}^{(q_l)}) \\
&\ge 1 - c_{3,l} p \log n \exp\{(-c_{2,l} n^{1-3\kappa_l - 2v_l - 2\rho_l}/\log n) + \kappa_l \log n\}.
\end{aligned}
$$

□

S3. Additional Simulation Results

We explored some dependent censoring situations, where the censoring times C_i depend on X. In the following, Example 3* is the same as Example 3, except that the censoring times C_i were generated from the following proportional hazards model

$$h_C(t \mid X) = c_0 \exp(\beta^{\mathrm{T}} X),$$

where $\beta = (0.3, 0.3, 0_{p-2}^{\mathrm{T}})^{\mathrm{T}}$ and c_0 was chosen to give approximately 20% and 40% of censoring proportions. Table S1 shows that the proposed method still provides good performance under the considered dependent censoring scenarios.

We next studied the performance of the proposed methods when the number of selected the top genes was 133, which was far more than 27 as reported in the main text. Table S2 reports the numbers of overlapping genes selected by different methods, showing that the variables selected by L_q-norm learning with different q did differ and the proposed method helped choose novel genes that were not identified by the existing methods.

We calculated and compared the c-statistics obtained by various methods. First, using the full dataset of 170 patients, we randomly generated 10 training/testing splits, with 133 in the training set and the rest in the testing set. In each training dataset, we fitted a random survival forests model based on the top 133 genes selected by each method. When fitting the random forests, a total of 100 trees were generated for each dataset. Then the fitted "forests" were applied to each testing dataset, for which a c-statistic was computed. Finally, for each method, the average of the c-statistics from all 10 testing datasets, along with its confidence interval, is listed in Table S3. The results showed that even with more selected genes, the c-statistics did not improve much across all the methods compared to the ones based on the top 27 genes.

References

Aussel, D., J.-N. Corvellec, and M. Lassonde (1995). Mean value property and subdifferential criteria for lower semicontinuous functions. *Transactions of the American Mathematical Society 347*(10), 4147–4161.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *The Annals of Statis-*

tics 17(3), 1157–1167.

Foldes, A. and L. Rejto (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *The Annals of Statistics 9*(1), 122–129.

Shao, J. (1999). *Mathematical Statistics*. New York: Springer.

Table S1: Performance of different variable screening methods with $(n,p) = (400, 1000)$ under Examples 3 and 3^*.

Example 3	20% CR			40% CR		
	MMS	TPR	PIT	MMS	TPR	PIT
L_1	2	1.00	1.00	2	1.00	1.00
L_2	2	1.00	1.00	2	1.00	1.00
L_5	2	1.00	1.00	2	1.00	1.00
L_{13}	2	1.00	1.00	2	1.00	0.99
L_{89}	2	1.00	1.00	2	0.99	0.99
L_∞	2	1.00	1.00	2	0.99	0.99
Hybrid	2	1.00	1.00	2	1.00	1.00
Example 3^*	20% CR			40% CR		
	MMS	TPR	PIT	MMS	TPR	PIT
L_1	2	1.00	1.00	2	0.99	0.99
L_2	2	1.00	1.00	2	0.99	0.99
L_5	2	1.00	1.00	2	0.99	0.98
L_{13}	2	1.00	1.00	2	0.99	0.99
L_{89}	2	1.00	1.00	2	0.99	0.98
L_∞	2	1.00	1.00	3	0.99	0.98
Hybrid	2	1.00	1.00	3	0.99	0.99

Table S2: The numbers of overlapping genes among the top 133 genes selected by various screening methods on the multiple myeloma training dataset.

	PSIS	CRIS	FAST	CS	QA	L_1	L_2	L_5	L_{13}	L_{89}	L_∞	Hybrid
PSIS	133	42	55	7	0	38	37	30	19	13	12	29
CRIS		133	16	5	1	30	31	27	17	13	17	26
FAST			133	9	1	15	14	11	4	2	1	10
CS				133	0	11	14	15	17	12	15	15
QA					133	0	0	0	0	0	2	0
L_1						133	122	88	53	36	33	82
L_2							133	94	57	37	34	84
L_5								133	94	72	69	109
L_{13}									133	108	97	99
L_{89}										133	115	84
L_∞											133	82
Hybrid												133

Table S3: Comparisons of the average c-statistics (95% CI) based on 10 random testing datasets of multiple myeloma.

PSIS	CRIS	FAST	CS	QA	Hybrid
0.61 (0.53, 0.70)	0.62 (0.49, 0.75)	0.56 (0.41, 0.72)	0.60 (0.43, 0.76)	0.56 (0.47, 0.65)	0.63 (0.56, 0.69)