Supplement to "Refining Kaplan-Meier Estimation with the Generalized Pareto Model for Survival Analysis"

Yi He
University of Amsterdam

Liang Peng
Georgia State University

Dabao Zhang
University of California, Irvine

Zifeng Zhao
University of Notre Dame

October 24, 2024

This supplement contains three appendices. Appendix A presents some general limit theories that will be used in the proofs of Theorem 2.1 in Appendix B and Theorem 2.2 in Appendix C.

A General Limit Theory

In this section, we generalize the universal law of large numbers and universal central limit theorem in He et al. (2022) for adaptive inference with a general threshold statistic under censoring. We invoke the notion of stable function therein: a real-valued function ϕ on $D \subset \mathbb{R}_+$ is called stable if it is bounded in a neighborhood of 1, and there exists a finite collection of k functions h_j and f_j such that

$$\phi(x) - \phi(y) = \sum_{j=1}^{k} h_j(y) f_j\left(\frac{x}{y}\right), \quad x, y \in D, \tag{1}$$

where f_j is continuous at 1 and $f_j(1) = 0$ for $j = 1, \ldots, k$. It is easy to verify that the constant function, logarithm function, and power function are all stable. Furthermore, for any two stable functions ϕ and ψ on a common domain D, $\phi + \psi$, $\phi - \psi$, and $\phi \cdot \psi$ are all stable. This implies that, for example, any polynomial function is stable, and $\phi(x) = \log^m(x)$ is stable for all integer $m \geq 1$.

Proposition 1 (Universal LLN under Censoring)**.** *Let $\{(U_i, \delta_i) : i = 1, \ldots, n\}$ be independent copies of some bi-variate random vector $(U, \delta) \in (0,1) \times \{0,1\}$, where U is not necessarily a uniform variable and may be dependent of δ. Suppose that we generate i.i.d. random weights $\{\xi_1, \ldots, \xi_n\}$ from a distribution with mean one and finite first absolute moment, possibly degenerate with $\xi_1 = \ldots = \xi_n = 1$, independent of $\{(U_i, \delta_i) : 1 \leq i \leq n\}$. Consider an arbitrary measurable statistic $\alpha_n \in (0,1)$ such that $\alpha_n \xrightarrow{p} \bar{\alpha} \in (0,1)$, and suppose that the marginal distribution of U is Lipschitz continuous in a neighborhood of $\bar{\alpha}$. For any stable function ϕ on $(0,1]$ satisfying the decomposition* (1) *and*

$$\frac{1}{\bar{\alpha}} \mathbb{E}\left[\delta \mathbb{1}[U < \bar{\alpha}] |h_j(U < \bar{\alpha})|\right] < \infty, \quad \textit{for all } j = 1, \ldots, k,$$

we have that

$$\begin{aligned} \frac{1}{n\alpha_n} \sum_{i=1}^n \xi_i \delta_i \mathbb{1}[U_i < \alpha_n] \phi\left(\frac{U_i}{\alpha_n}\right) =& \frac{1}{n\bar{\alpha}} \sum_{i=1}^n \xi_i \delta_i \mathbb{1}[U_i < \bar{\alpha}] \phi\left(\frac{U_i}{\bar{\alpha}}\right) + o_p(1) \\ =& \frac{1}{\bar{\alpha}} \mathbb{E}\left[\delta \mathbb{1}[U < \bar{\alpha}] \phi\left(\frac{U}{\bar{\alpha}}\right)\right] + o_p(1). \end{aligned}$$

if provided the existence of the limit $\frac{1}{\bar{\alpha}} \mathbb{E}\left[\delta \mathbb{1}[U < \bar{\alpha}] \phi\left(\frac{U}{\bar{\alpha}}\right)\right]$. Note that the results remain true if we replace δ with $1 - \delta$ everywhere.

Proof. We only prove the first equation, as the second one follows directly from the law of large numbers. It suffices to show that

$$\begin{aligned}
&\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}\xi_i\delta_i\mathbb{1}[U_i<\alpha_n]\phi\left(\frac{U_i}{\alpha_n}\right)-\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}\xi_i\delta_i\mathbb{1}[U_i<\bar{\alpha}]\phi\left(\frac{U_i}{\bar{\alpha}}\right)\\
=&\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}\xi_i\delta_i\mathbb{1}[U_i<\alpha_n]\left(\phi\left(\frac{U_i}{\alpha_n}\right)-\phi\left(\frac{U_i}{\bar{\alpha}}\right)\right)\\
&+\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}\xi_i\delta_i\left(\mathbb{1}[U_i<\alpha_n]-\mathbb{1}[U_i<\bar{\alpha}]\right)\phi\left(\frac{U_i}{\bar{\alpha}}\right)=:T_1+T_2=o_p(1).
\end{aligned}$$

First, we show that $T_1=o_p(1)$. Recalling the decomposition (1) for stable function ϕ and using triangle inequality, there exists functions h_j and f_j such that

$$|T_1|\leq\sum_{j=1}^{k}\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}|\xi_i|\delta_i\mathbb{1}(U_i<\bar{\alpha})\left|h_j\left(U_i/\bar{\alpha}\right)\right|\left|f_j(\bar{\alpha}/\alpha_n)\right|$$

Observe that for every j, $f_j(\bar{\alpha}/\alpha_n)=f_j(1)+o_p(1)=o_p(1)$, and by assumption

$$\mathbb{E}\left\{\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}|\xi_i|\mathbb{1}(U_i<\bar{\alpha})\left|h_j\left(U_i/\bar{\alpha}\right)\right|\right\}=\mathbb{E}|\xi_i|\cdot\frac{1}{\bar{\alpha}}\mathbb{E}\left[\delta_i\mathbb{1}[U_i<\bar{\alpha}]|h_j(U_i<\bar{\alpha})|\right]=O(1).$$

Because k is finite, together with Markov inequality,

$$T_1=\sum_{j=1}^{k}O_p(1)\cdot o_p(1)=o_p(1).$$

Next, we show that $T_2=o_p(1)$. We only need to consider those U_i's lying between α_n and $\bar{\alpha}$, i.e. those U_i/α_n lying between one and $\bar{\alpha}/\alpha_n$. For any small $\varepsilon>0$ and large $M>0$, with probability tending to 1, $|\alpha_n/\bar{\alpha}-1|<\varepsilon^2/M$ and then

$$\begin{aligned}
|T_2|\leq&\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}|\xi_i\delta_i|\mathbb{1}(|U_i-\bar{\alpha}|\leq\bar{\alpha}\varepsilon^2/M)\cdot\sup_{|x-1|\leq\varepsilon^2/M}|\phi(x)|\\
\leq&\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}|\xi_i|\mathbb{1}(|U_i-\bar{\alpha}|\leq\bar{\alpha}\varepsilon^2/M)\cdot\sup_{|x-1|\leq\varepsilon^2/M}|\phi(x)|.
\end{aligned}$$

But by assumption, for some Lipschitz constant K,

$$\begin{aligned}
&\frac{1}{\bar{\alpha}}\mathbb{E}\left\{|\xi_i|\mathbb{1}(|U_i-\bar{\alpha}|\leq\bar{\alpha}\varepsilon^2/M)\right\}\\
=&\frac{\mathbb{E}|\xi_i|}{\bar{\alpha}}\left(\mathbb{P}(U_i\leq\bar{\alpha}+\bar{\alpha}\varepsilon^2/M)-\mathbb{P}(U_i\leq\bar{\alpha}-\bar{\alpha}\varepsilon^2/M)\right)\leq 2\mathbb{E}|\xi_i|K/M\varepsilon^2.
\end{aligned}$$

Combining with Markov inequality yields that, for all large n

$$
\begin{aligned}
\mathbb{P}\left(|T_2|>\varepsilon\right) \leq & \mathbb{P}\left(|T_2|>\varepsilon \mid |\alpha_n/\bar{\alpha}-1|<\varepsilon^2/M\right)+\mathbb{P}\left(|\alpha_n/\bar{\alpha}-1|>\varepsilon^2/M\right) \\
\leq & 2\mathbb{P}\left(\frac{1}{n\bar{\alpha}}\sum_{i=1}^{n}|\xi_i|\mathbb{1}(|U_i-\bar{\alpha}|\leq\bar{\alpha}\varepsilon^2/M)\cdot\sup_{|x-1|\leq\varepsilon^2/M}|\phi(x)|>\varepsilon\right)+\varepsilon/2 \\
\leq & 2\frac{1}{\varepsilon}\frac{1}{\bar{\alpha}}\mathbb{E}\left\{|\xi_i|\mathbb{1}(|U_i-\bar{\alpha}|\leq\bar{\alpha}\varepsilon^2/M)\right\}\cdot\sup_{|x-1|\leq\varepsilon}|\phi(x)|+\varepsilon/2 \\
\leq & 2\frac{1}{\varepsilon}\frac{2\mathbb{E}|\xi_i|K}{M}\varepsilon^2\cdot\sup_{|x-1|\leq\varepsilon}|\phi(x)|+\varepsilon/2<\varepsilon,
\end{aligned}
$$

by taking a sufficiently large M not depending on ε. It follows that $T_2=o_p(1)$ as $\varepsilon>0$ can be arbitrarily small. □

Before showing the universal CLT under censoring, we first introduce some important notations. With probability 1, $(X_i,\delta_i)=(Q(1-U_i),\delta_i)$ where U_i's are i.i.d. uniform variables on $[0,1]$ and Q is the generalized quantile function of X_i. Define $V_i=\widetilde{\rho}(U_i)$, where $\widetilde{\rho}(x)=\mathbb{P}(U_i<x,\delta_i=1)$ is a continuous improper distribution function with total mass $p_1:=\widetilde{\rho}(\infty)=\mathbb{P}(\delta_i=1)>0$. Consider the empirical processes

$$
\begin{aligned}
\bar{\mathbb{U}}_n(x):=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}[U_i<x]-x),\ 0\leq x\leq 1; \\
\bar{\mathbb{V}}_n(x):=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}[V_i<x,\delta_i=1]-x),\ 0\leq x\leq p_1.
\end{aligned}
$$

Similarly, define the random weighted empirical processes

$$
\begin{aligned}
\widehat{\mathbb{U}}_n(x):=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathbb{1}[U_i<x]-x),\ 0\leq x\leq 1; \\
\widehat{\mathbb{V}}_n(x):=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathbb{1}[V_i<x,\delta_i=1]-x),\ 0\leq x\leq p_1.
\end{aligned}
$$

The following lemma follows from the Chibisov-O'Reilly theorems.

Lemma A.1 (Chibisov-O'Reilly Theorem)**.** *Let $\xi_1,\xi_2,\ldots$ be nondegenerate random weights with mean one taken from a sequence of i.i.d. random variables with subexponential distri-*

bution. For any $\eta \in (0, 1/2)$, under Skorokhod construction,

$$\sup_{x\in[0,1]} \frac{|\bar{\mathbb{U}}_n(x) - \bar{B}_1(x)|}{x^\eta} \xrightarrow{p} 0, \quad \sup_{x\in[0,p_1]} \frac{|\bar{\mathbb{V}}_n(x) - \bar{B}_2(x)|}{x^\eta} \xrightarrow{p} 0,$$

and similarly

$$\sup_{x\in[0,1]} \frac{|\widehat{\mathbb{U}}_n(x) - \widehat{B}_1(x)|}{x^\eta} \xrightarrow{p} 0, \quad \sup_{x\in[0,p_1]} \frac{|\widehat{\mathbb{V}}_n(x) - \widehat{B}_2(x)|}{x^\eta} \xrightarrow{p} 0,$$

where $(\bar{B}_1, \bar{B}_2)$ and $(\widehat{B}_1, \widehat{B}_2)$ are independent copies of (B_1, B_2), a bivariate Gaussian process whose margins B_1, B_2 are Brownian bridges and the cross-covariance structure is given by

$$cov(B_1(s), B_2(t)) = \min\{\widetilde{\rho}(s), t\} - st.$$

Proof. Following the proof of Lemma D.2 in He et al. (2022), it suffices to prove the lemma substantiating $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$ with the approximate processes $(\underline{\widehat{\mathbb{U}}}_n, \underline{\widehat{\mathbb{V}}}_n)$ given by

$$\underline{\widehat{\mathbb{U}}}_n(x) := \frac{1}{\sqrt{\sum_{i=1}^n \widetilde{\xi}_i^2}} \sum_{i=1}^n \widetilde{\xi}_i(\mathbb{1}[U_i < x] - x), \ 0 \le x \le 1;$$

$$\underline{\widehat{\mathbb{V}}}_n(x) := \frac{1}{\sqrt{\sum_{i=1}^n \widetilde{\xi}_i^2}} \sum_{i=1}^n \widetilde{\xi}_i(\mathbb{1}[V_i < x, \delta_i = 1] - x), \ 0 \le x \le p_1,$$

where $\widetilde{\xi}_i = \xi_i - \frac{1}{n}\sum_{i=1}^n \xi_i$. This is because they have shown that

$$\sup_{x\in[0,1]} \frac{\left|\widehat{\mathbb{U}}_n(x) - \widehat{B}_1(x)\right|}{t^\eta} = (1 + o_p(1)) \sup_{x\in[0,1]} \frac{\left|\underline{\widehat{\mathbb{U}}}_n(x) - \widehat{B}_1(t)\right|}{t^\eta} + o_p(1),$$

and similar arguments hold for $\widehat{V}_n$.

Consider the combined weighted process $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n, \underline{\widehat{\mathbb{U}}}_n, \underline{\widehat{\mathbb{V}}}_n)$. By the Lebesgue dominated theorem, it suffices to consider a conditional statement given the sample path of random weights $\xi_i = \xi_i(\omega)$, or equivalently $\widetilde{\xi}_i = \widetilde{\xi}_i(\omega)$, for any $\omega \in \Omega$ from a set Ω with probability measure 1. This is because the weights are independent of the observations, and the joint limiting distribution does not depend on the weights. In particular, for sub-exponential weights ξ_i, we can easily choose such a set via the Borel–Cantelli lemma satisfying the asymptotic negligibility conditions required for the following:

(i) Finite-dimensional convergence holds by applying the Lindeberg central limit theorem with the boundedness of the indicator function.

(ii) The marginal tightness of $\widehat{\underline{\mathbb{U}}}_n$ and $\widehat{\underline{\mathbb{V}}}_n$ follows from Theorem 1, Chapter 3, in Shorack and Wellner (1986).

Moreover, the marginal tightness of $\mathbb{U}_n$ and $\mathbb{V}_n$ are available from Shorack and Wellner (1982), and these processes do not depend on the random weights. Hence, one can conclude that, conditional on the sample path of random weights from a set with probability 1,

$$(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n, \widehat{\underline{\mathbb{U}}}_n, \widehat{\underline{\mathbb{V}}}_n) \xrightarrow{w} \left(\bar{B}_1, \bar{B}_2, \widehat{B}_1, \widehat{B}_2\right)$$

in the product of generalized Skorohod space on $[0,1] \times [0,p_1] \times [0,1] \times [0,p_1]$ of left-continuous functions, where '$\xrightarrow{w}$' denotes weak convergence. Therefore, under Skorohod construction,

$$(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n, \widehat{\underline{\mathbb{U}}}_n, \widehat{\underline{\mathbb{V}}}_n) \xrightarrow{a.s.} \left(\bar{B}_1, \bar{B}_2, \widehat{B}_1, \widehat{B}_2\right)$$

where the processes on the left are equal to the original ones only in distribution. Note that we can extend the probability space to include the random weights. The rest follows by applying the Chibisov-O'Reilly Theorem to each marginal process in this probability space (see Chibisov, 1964, O'Reilly, 1974 and Shorack and Wellner, 1982). □

Now, we present our universal CLT under censoring.

Proposition 2 (Universal CLT under Censoring)**.** *Suppose that the threshold statistic u_n satisfies Assumption 2.2, where we define adaptive exceeding probability $\alpha_n = S_0(u_n)$ and its limit $\bar{\alpha} \in (0,1)$. Consider a stable function ϕ defined in* (1) *on an open domain $D \subset (0,\infty)$ including $(0,1]$ such that* *(i) ϕ and its base functions h_j, $1 \le j \le k$, all have Lebesgue integrable derivatives on every closed sub-interval of D; note that each sub-interval is bounded away from zero.*

(ii) For some $\eta \in (0, 1/2)$, $\lim_{t \downarrow 0} t^\eta |\phi(t)| = 0$, $\int_0^1 t^\eta |\phi'(t)| dt < \infty$ and $\int_0^1 t^\eta |h_j'(t)| dt < \infty$ for all $1 \le j \le k$.

(iii) The derivatives ϕ' and h_j' are continuous in a neighborhood of one.

Let $(\mathbb{U}_n, \mathbb{V}_n)$ denote one of the multivariate processes, either $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$ or $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$. Correspondingly, define the limiting processes (B_1, B_2) to be $(\bar{B}_1, \bar{B}_2)$ or $(\widehat{B}_1, \widehat{B}_2)$, respectively from Lemma A.1. Under the same probability space of Lemma A.1,

$$\int_0^1 \phi(t) d\mathbb{W}_{n,i}(\alpha_n t) = \phi(1) W_i(\bar{\alpha}) - \int_0^1 W_i(\bar{\alpha} t) d\phi(t) + o_p(1), \quad i = 1, 2.$$

where $\mathbb{W}_{n,1}(t) = \mathbb{U}_n(S(Q_0(1-t)))$, $\mathbb{W}_{n,2}(t) = \mathbb{V}_n(\widetilde{S}(Q_0(1-t)))$, $W_1(t) = B_1(S(Q_0(1-t)))$, and $W_2(t) = B_2(\widetilde{S}(Q_0(1-t)))$.

Proof. First, we show that $\lim_{t \downarrow 0} \phi(t) \mathbb{W}_{n,i}(\alpha_n t) = \lim_{t \downarrow 0} t^\eta \phi(t) \cdot \lim_{t \downarrow 0} t^{-\eta} \mathbb{W}_{n,i}(\alpha_n t) = o_p(1)$. We only need to show $\lim_{t \downarrow 0} t^{-\eta} \mathbb{W}_{n,i}(\alpha_n t) = O_p(1)$. By Lemma A.1

$$\lim_{t \downarrow 0} t^{-\eta} \mathbb{W}_{n,i}(\alpha_n t) = \lim_{t \downarrow 0} t^{-\eta} W_i(\alpha_n t) + o_p(1).$$

Consider the case for $i = 1$. By O'Reilly (1974) theorem, we also have that, with probability 1, uniformly for $t \in (0, 1]$

$$\begin{aligned} t^{-\eta} |W_1(\alpha_n t)| &= \left| \frac{B_1(S(Q_0(1 - \alpha_n t)))}{(S(Q_0(1 - \alpha_n t)))^\eta} \right| \left(\frac{S(Q_0(1 - \alpha_n t))}{t} \right)^\eta \\ &\le \sup_{x \in (0,1)} \left| \frac{B_1(x)}{x^\eta} \right| \cdot \left(\frac{\alpha_n t}{t} \right)^\eta \le \sup_{x \in (0,1)} \left| \frac{B_1(x)}{x^\eta} \right| = O_p(1), \end{aligned}$$

where we use the fact that $0 < \widetilde{S}(Q_0(1 - \bar{\alpha} t)) \le S(Q_0(1 - \bar{\alpha} t)) \le \bar{\alpha} t$ all $t \in (0, 1)$ for the first inequality. The case for $i = 2$ is similar and omitted.

The integration by parts formula then gives that

$$
\begin{aligned}
\int_0^1 \phi(t) d\mathbb{W}_{n,i}(\alpha_n t) =& \phi(1)\mathbb{W}_{n,i}(\alpha_n) - o_p(1) - \int_0^1 \mathbb{W}_{n,i}(\alpha_n t) d\phi(t) \\
=& \phi(1)\mathbb{W}_{n,i}(\alpha_n) - o_p(1) - \int_0^{\alpha_n} \mathbb{W}_{n,i}(t) d\phi(t/\alpha_n). \qquad (2)
\end{aligned}
$$

Using the uniform convergence in Lemma A.1 and stochastic continuity of Brownian bridge (and the continuity assumption of S and $\widetilde{S}$ around the neighborhood of $\bar{u} = Q_0(1-\bar{\alpha}t)$),

$$
\mathbb{W}_{n,i}(\alpha_n) = W_i(\alpha_n) + o_p(1) = W_i(\bar{\alpha}) + o_p(1). \qquad (3)
$$

Furthermore, using expansion (1) for stable function ϕ,

$$
\phi(t/\alpha_n) - \phi(t/\bar{\alpha}) = \sum_{j=1}^{k} h_j(t/\bar{\alpha}) f_j(\alpha_n/\bar{\alpha}).
$$

It follows that

$$
\begin{aligned}
\int_0^{\alpha_n} \mathbb{W}_{n,i}(t) d\phi(t/\alpha_n) =& \int_0^{\alpha_n} \mathbb{W}_{n,i}(t) d\phi(t/\bar{\alpha}) \\
&+ \sum_{j=1}^{k} f_j(\alpha_n/\bar{\alpha}) \int_0^{\alpha_n} \mathbb{W}_{n,i}(t) dh_j(t/\bar{\alpha}).
\end{aligned}
$$

The leading term (i.e., first term above) equals to

$$
\begin{aligned}
\int_0^{\alpha_n/\bar{\alpha}} \mathbb{W}_{n,i}(\bar{\alpha}t) d\phi(t) =& \int_0^1 \mathbb{W}_{n,i}(\bar{\alpha}t) d\phi(t) + \int_1^{\alpha_n/\bar{\alpha}} \mathbb{W}_{n,i}(\bar{\alpha}t) d\phi(t) \\
=& \int_0^1 \mathbb{W}_{n,i}(\bar{\alpha}t) d\phi(t) + o_p(1),
\end{aligned}
$$

where we use the stochastic continuity of $\mathbb{W}_{n,i}(\bar{\alpha}t)$ from Lemma A.1 and the boundedness of ϕ' around one in the last step. Similarly, the remainder terms

$$
\begin{aligned}
\int_0^{\alpha_n} \mathbb{W}_{n,i}(t) dh_j(t/\bar{\alpha}) =& \int_0^1 \mathbb{W}_{n,i}(\bar{\alpha}t) dh_j(t) + \int_1^{\alpha_n/\bar{\alpha}} \mathbb{W}_{n,i}(\bar{\alpha}t) dh_j(t) \\
=& \int_0^1 \mathbb{W}_{n,i}(\bar{\alpha}t) dh_j(t) + o_p(1) \\
=& \int_0^1 \left(t^{-\eta}\mathbb{W}_{n,i}(\bar{\alpha}t)\right) t^{\eta} h_j'(t) dt + o_p(1) = O_p(1),
\end{aligned}
$$

where in the last step we invoke from above that $\sup_{t\in(0,1]}|t^{-\eta}\mathbb{W}_{n,i}(\bar{\alpha}t)| = O_p(1)$ and the assumption that $\int_0^1 t^\eta |h_j'(t)|dt < \infty$. But $f_j(\alpha_n/\bar{\alpha}) = f_j(1) + o_p(1) = o_p(1)$ by continuous mapping theorem for every j. Collecting the asymptotic approximations above gives that

$$\int_0^{\alpha_n} \mathbb{W}_{n,i}(t)d\phi(t/\alpha_n) = \int_0^1 \mathbb{W}_{n,i}(\bar{\alpha}t)d\phi(t) + o_p(1).$$

Recall that $0 < \widetilde{S}(Q_0(1-\bar{\alpha}t)) \le S(Q_0(1-\bar{\alpha}t)) \le \bar{\alpha}t$ all $t \in (0,1)$. Applying Lemma A.1, we can replace $\mathbb{W}_{n,i}$ with their W_i in the last line. Specifically, when $i=1$,

$$\begin{aligned}
\int_0^1 \mathbb{W}_{n,1}(\bar{\alpha}t)h_j'(t)dt &= \int_0^1 \frac{\mathbb{U}_n(S(Q_0(1-\bar{\alpha}t)))}{S^\eta(Q_0(1-\bar{\alpha}t))}\left(\frac{S(Q_0(1-\bar{\alpha}t))}{\bar{\alpha}t}\right)^\eta (\bar{\alpha}t)^\eta h_j'(t)dt. \\
&= \int_0^1 \frac{B_1(S(Q_0(1-\bar{\alpha}t)))}{S^\eta(Q_0(1-\bar{\alpha}t))}\left(\frac{S(Q_0(1-\bar{\alpha}t))}{\bar{\alpha}t}\right)^\eta (\bar{\alpha}t)^\eta h_j'(t)dt + o_p(1) \\
&= \int_0^1 W_1(\bar{\alpha}t)h_j'(t)dt + o_p(1),
\end{aligned}$$

where the second step follows from Lemma A.1 and the fact that

$$\int_0^1 \left(\frac{S(Q_0(1-\bar{\alpha}t))}{\bar{\alpha}t}\right)^\eta (\bar{\alpha}t)^\eta |h_j'(t)|dt \le \int_0^1 (\bar{\alpha}t)^\eta |h_j'(t)|dt = \bar{\alpha}^\eta \int_0^1 t^\eta |h_j'(t)|dt < \infty.$$

The case for $i=2$ is completely analogous, and we omit the details. Therefore, we have the final approximation that

$$\int_0^{\alpha_n} \mathbb{W}_{n,i}(t)d\phi(t/\alpha_n) = \int_0^1 W_i(\bar{\alpha}t)d\phi(t) + o_p(1). \tag{4}$$

Combining equations (2)–(4) completes the proof. □

B Proof of Theorem 2.1

Let $\boldsymbol{\theta} = (\gamma, \log\sigma)^\top$ and $\Theta_n^\varepsilon = \{\boldsymbol{\theta} \in (-1/2,\infty)\times\mathbb{R} : \left\|\boldsymbol{\theta} - \boldsymbol{\theta}_0^{(n)}\right\| < n^{-1/2+\varepsilon}\}$, where $\boldsymbol{\theta}_0^{(n)} = (\gamma_0, \log\sigma_{\alpha_n})^\top$ denotes the adaptive true value. For i-th observation, denote its exceedance likelihood by $l_i(\boldsymbol{\theta}|u_n) = \mathbb{1}[X_i > u_n]\ell_i(\boldsymbol{\theta}|u_n)$ and define the sample score vector for $\boldsymbol{\theta}$ by

$$g_i(\boldsymbol{\theta}|u_n) = \nabla_{\boldsymbol{\theta}} l_i(\boldsymbol{\theta}|u_n) = \mathbb{1}[X_i > u_n]\nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}|u_n),$$

where, for $X_i > u_n$,

$$\nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}|u_n) =: \delta_i s(\boldsymbol{\theta}|X_i - u_n) + (1-\delta_i) w(\boldsymbol{\theta}|X_i - u_n).$$

The function $s(\boldsymbol{\theta}|x) = \nabla_{\boldsymbol{\theta}} \log(-G'(x|\gamma,\sigma))$ is the lifetime score function for generalized Pareto distributions with respect to $\boldsymbol{\theta}$, and the function $w(\boldsymbol{\theta}|x) = \nabla_{\boldsymbol{\theta}} \log(G(x|\gamma,\sigma))$ is the censoring score function. For presentation convenience, whenever needed, all the limit elements are defined on the same probability space via the Skorohod construction in Lemma A.1. They are only equal in distribution to the original elements, and the joint convergence in probability in this probability space implies the joint weak convergence in the original space.

B.1 Adaptive Maximum Likelihood Estimation

We establish the following fundamental results for our maximum likelihood estimation adaptive to a universal threshold statistic u_n. Let '$\xrightarrow{p}$' denote convergence in probability.

(a) With probability tending to 1, the log-likelihood function is well-defined, that is,

$$\sum_{i=1}^{n} l_i(\boldsymbol{\theta}|u_n) > -\infty$$

uniformly in the parameter space $\mathrm{cl}(\Theta_n^{\varepsilon})$ with $\varepsilon \in (0, \min\{\gamma_0 + 1/2, 1/2\})$, where $\mathrm{cl}(\cdot)$ denotes the set closure.

(b) Under the Skorohod construction of Lemma A.1 for $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$,

$$\frac{1}{\sqrt{n\alpha_n}} \sum_{i=1}^{n} g_i(\boldsymbol{\theta}_0^{(n)}|u_n) \xrightarrow{p} \Upsilon$$

where Υ is defined the same way in Theorem 2.1 through the Brownian bridges (B_1, B_2).

Note that here the random elements are only equal to the original ones in distributions.

(c) $\sup_{\boldsymbol{\theta}\in \mathrm{cl}(\Theta_n^\varepsilon)} \left\| \frac{1}{n\alpha_n} \sum_{i=1}^n \nabla g_i(\boldsymbol{\theta}|u_n) + \mathcal{I}(\bar{\alpha}) \right\| \xrightarrow{p} 0$, where the unconditional Fisher information matrix $\mathcal{I}(\bar{\alpha})$ is defined in (7) in the main document.

Claim (a) is straightforward for the case $\gamma_0 \geq 0$ where $\left\{1+\gamma\frac{X_i-u_n}{\sigma} : X_i - u_n\right\}$ is bounded below by a strictly positive number, say, $1/2$ with probability tending to one. When $\gamma_0 < 0$, using the fact that $X_{n:n} \geq T_{n:n}$, we can show that

$$1+\gamma\cdot\frac{X_{n:n}-u_n}{\sigma} \geq 1+\gamma\cdot\frac{T_{n:n}-u_n}{\sigma} > 0$$

uniformly for Θ_n^ε with probability tending to 1, where the proof of the last step is available in Section E.2 of He et al. (2022).

Next, we prove claim (b). Recall the score functions of $\ell(\boldsymbol{\theta}|x,\delta)$ are given by

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}|x,\delta) =& \delta s(\boldsymbol{\theta}|x) + (1-\delta)w(\boldsymbol{\theta}|x)\\ =& -\delta(w(\boldsymbol{\theta}) - s(\boldsymbol{\theta}|x)) + w(\boldsymbol{\theta}|x) =: -\delta h(\boldsymbol{\theta}|x) + w(\boldsymbol{\theta}|x),\end{aligned}$$

where $h(\boldsymbol{\theta}|x) = \nabla_{\boldsymbol{\theta}}\log\lambda(\boldsymbol{\theta}|x)$. The next lemma follows from the integration by parts formula.

Lemma B.1. *Write in short that $h_i(x) = h_i(\gamma_0, \log\sigma_{\alpha_n}|x)$ where $\alpha_n = S_0(u_n)$. The adaptive score equations are correct, that is, with probability 1*

$$-\int_{u_n}^{\infty} h(x-u_n)d\widetilde{S}(x) + \int_{u_n}^{\infty} w(x-u_n)dS(x) = \mathbf{0}, \tag{5}$$

where $\widetilde{S}(x) = \mathbb{P}(X > x, \delta = 1)$ is a possibly improper censored survival function.

Proof. Observe that $\widetilde{S}(x) = \mathbb{P}(X > x, \delta = 1) = -\int_x^\infty (1-H)dS_0$ and $S(x) = S_0(x)(1 - H(x))$. Substituting $S_0(x) = \alpha_n G_0(x-u_n)$ with $G_0(x) := G(x|\gamma_0, \sigma_{\alpha_n})$ for $x > u_n > u_0$

yields that

$$S(x) = \alpha_n G_0(x - u_n)(1 - H(x))$$

and

$$\widetilde{S}'(x) = \alpha_n (1 - H(x)) G_0'(x - u_n) = -S(x)\lambda_0(x - u_n),$$

where $\lambda_0(x) := \lambda(\boldsymbol{\theta}_0^{(n)}|x)$. It follows that

$$\begin{aligned}\int_{u_n}^{\infty} h(x - u_n) d\widetilde{S}(x) &= \int_{u_n}^{\infty} \nabla_{\boldsymbol{\theta}} \lambda(\boldsymbol{\theta}_0^{(n)}|x) S(x) dx \\ &= -\int_{u_n}^{\infty} \frac{\partial}{\partial x} w(\boldsymbol{\theta}_0^{(n)}|x - u_n) S(x) dx = \int_{u_n}^{\infty} w(\boldsymbol{\theta}_0^{(n)}|x - u_n) dS(x)\end{aligned}$$

where the last step follows from the integration by parts formula. □

We have the following formulas by direct calculations. The first part for lifetime score functions is due to Fact 2 in He et al. (2022), and the other part for censoring score functions follows from similar calculations.

Fact 1. *For all $\alpha, \alpha t \in (0, \alpha_0)$,*

$$-\nabla_{\boldsymbol{\theta}} s(\gamma_0, \log \sigma_\alpha | Q_0(1 - \alpha t) - Q_0(1 - \alpha)) = \begin{bmatrix} \phi_{1,1,\gamma_0}(t) & \phi_{1,2,\gamma_0}(t) \\ \phi_{1,2,\gamma_0}(t) & \phi_{2,2,\gamma_0}(t) \end{bmatrix} =: \Phi_{\gamma_0}(t)$$

and

$$-\nabla_{\boldsymbol{\theta}} w(\gamma_0, \log \sigma_\alpha | Q_0(1 - \alpha t) - Q_0(1 - \alpha)) = \begin{bmatrix} \psi_{1,1,\gamma_0}(t) & \psi_{1,2,\gamma_0}(t) \\ \psi_{1,2,\gamma_0}(t) & \psi_{2,2,\gamma_0}(t) \end{bmatrix} := \Psi_{\gamma_0}(t),$$

where

$$\phi_{1,1,\gamma_0}(t) = \frac{2}{\gamma_0^3} \left(-\gamma_0 \log t - \frac{3 + \gamma_0}{2} + (\gamma_0 + 2) t^{\gamma_0} - \frac{1 + \gamma_0}{2} t^{2\gamma_0} \right),$$

$$\phi_{1,2,\gamma_0}(t) = \frac{1}{\gamma_0^2} \left(1 - (2 + \gamma_0) t^{\gamma_0} + (1 + \gamma_0) t^{2\gamma_0}\right), \quad \phi_{2,2,\gamma_0}(t) = \frac{1 + \gamma_0}{\gamma_0} \left(t^{\gamma_0} - t^{2\gamma_0}\right),$$

$$\psi_{1,1,\gamma_0}(t) = \frac{2}{\gamma_0^3} \left(-\gamma_0 \log t - \frac{3}{2} + 2 t^{\gamma_0} - \frac{1}{2} t^{2\gamma_0} \right),$$

$$\psi_{1,2,\gamma_0}(t) = \frac{1}{\gamma_0^2} (1 - t^{\gamma_0})^2, \quad \psi_{2,2,\gamma_0}(t) = \frac{1}{\gamma_0} \left(t^{\gamma_0} - t^{2\gamma_0}\right)$$

being well defined for $\gamma_0 = 0$ by continuity as

$$\phi_{1,1,0}(t) = -\frac{2}{3}(\log t)^3 - (\log t)^2,\ \phi_{1,2,0}(t) = (\log t)^2 + \log t,\ \phi_{2,2,0}(t) = -\log t,$$
$$\psi_{1,1,0}(t) = -\frac{2}{3}(\log t)^3,\ \psi_{1,2,0}(t) = (\log t)^2,\ \psi_{2,2,0}(t) = -\log t.$$

Observe that

$$\widetilde{S}(x) = \mathbb{P}(X_i > x, \delta_i = 1) = \mathbb{P}(U_i < S(x), \delta_i = 1) = \mathbb{P}(V_i < \widetilde{\rho}(S(x)), \delta_i = 1) = \widetilde{\rho}(S(x)).$$

With probability 1, using Lemma B.1 we can rewrite that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} g_i(\gamma_0, \log\sigma_{\bar{\alpha}_n}|X_i - u_n, \delta) \\
&= \int_{u_n}^{\infty} h(x-u_n)d\bar{\mathbb{V}}_n(\widetilde{S}(x)) - \int_{u_n}^{\infty} w(x-u_n)d\bar{\mathbb{U}}_n(S(x)) \\
&= -\int_0^1 h(Q_0(1-\alpha_n t) - Q_0(1-\alpha_n))d\bar{\mathbb{V}}_n(\widetilde{S}(Q_0(1-\alpha_n t))) \\
&\quad + \int_0^1 w(Q_0(1-\alpha_n t) - Q_0(1-\alpha_n))d\bar{\mathbb{U}}_n(S(Q_0(1-\alpha_n t))) =: J_1 + J_2.
\end{aligned}$$

Recall from Proposition 2 that $\mathbb{W}_{n,1}(\alpha_n t) = \bar{\mathbb{U}}_n(S(Q_0(1-\alpha_n t)))$ and $\mathbb{W}_{n,2}(\alpha_n t) = \bar{\mathbb{V}}_n(\widetilde{S}(Q_0(1-\alpha_n t)))$. Also recall the expansion of the high quantile function Q_0 for $0 < \alpha_n t \le \alpha_n < \alpha_0$ given by

$$Q_0(1-\alpha_n t) - Q_0(1-\alpha_n) = \begin{cases} \frac{\sigma_{\alpha_n}}{\gamma_0}\left(t^{-\gamma_0} - 1\right) & \gamma_0 \neq 0, \\ \sigma_{\alpha_n}\log(1/t) & \gamma_0 = 0, \end{cases}$$

where $\alpha_0 = S_0(u_0)$. Therefore, using Fact 1, we can rewrite that

$$J_1 = -\int_0^1 h(Q_0(1-\bar{\alpha}_n t) - Q_0(1-\bar{\alpha}_n))d\mathbb{W}_{n,2}(\alpha_n t) = -\int_0^1 \phi_{1,\gamma_0}(t)d\mathbb{W}_{n,2}(\alpha_n t)$$

and

$$J_2 = \int_0^1 w(Q_0(1-\bar{\alpha}_n t) - Q_0(1-\bar{\alpha}_n))d\mathbb{W}_{n,1}(\alpha_n t) = \int_0^1 \phi_{2,\gamma_0}(t)d\mathbb{W}_{n,1}(\alpha_n t),$$

where

$$\phi_{1,\gamma_0}(t) = \begin{bmatrix} \gamma_0^{-1}(1-t^{\gamma_0}) \\ \\ t^{\gamma_0} \end{bmatrix} \quad \text{and} \quad \phi_{2,\gamma_0}(t) = \begin{bmatrix} \gamma_0^{-2}(-\gamma_0 \log t + t^{\gamma_0} - 1) \\ \\ \gamma_0^{-1}(1-t^{\gamma_0}) \end{bmatrix}.$$

When $\gamma_0 = 0$, we interpret ϕ_{1,γ_0} and ϕ_{1,γ_0} as their continuous extension given by $\phi_{1,0}(t) = [-\log t, 1]^T$ and $\phi_{2,0}(t) = [\frac{1}{2}(\log t)^2, \ -\log t]^T$, respectively. The rest follows from Proposition 2. Note that we rewrite the limit using $\psi_{i,\gamma_0}(t) = \phi'_{i,\gamma_0}(t)$ for $i = 1, 2$.

It remains to prove claim (c). The following formula is due to Gertsbakh (1995).

Fact 2. *For any fixed $u > u_0$ with $\alpha = S_0(u) > 0$,*

$$-\frac{1}{\alpha}\mathbb{E}[\nabla g_i(\gamma, \log\sigma|u)|C_i - u = z > 0] = \mathcal{I}(\gamma, \log\sigma|z).$$

Moreover, we have the following formulas for the Hessian matrix. The first part for lifetime score functions $s(\gamma, \log\sigma|x) = (s_1(\gamma, \log\sigma|x), s_2(\gamma, \log\sigma|x))^\top$ is due to Fact 1 in He et al. (2022), and the other part for censoring score $w(\gamma, \log\sigma|x) = (w_1(\gamma, \log\sigma|x), w_2(\gamma, \log\sigma|x))^\top$ can be obtained similarly by straightforward calculations.

Fact 3. *Given any threshold statistic u, the negative Hessian matrix is given by*

$$\begin{aligned} \frac{1}{n\alpha}\sum_{i=1}^n \nabla g_i(\gamma, \log\sigma|u) =& \frac{1}{n\alpha}\sum_{i=1}^n \delta_i \nabla s(\gamma, \log\sigma|T_i - u) \\ &+ \frac{1}{n\alpha}\sum_{i=1}^n (1-\delta_i)\nabla w(\gamma, \log\sigma|C_i - u), \end{aligned}$$

where

$$\begin{aligned} -\frac{\partial s_1(\gamma, \log\sigma|x)}{\partial\gamma} =& \frac{2}{\gamma^3}\left(\log\left(1+\gamma\frac{x}{\sigma}\right) - \frac{\gamma\cdot x/\sigma}{1+\gamma\frac{x}{\sigma}} - \frac{\gamma^2+\gamma^3}{2}\frac{x^2/\sigma^2}{\left(1+\gamma\frac{x}{\sigma}\right)^2}\right), \\ -\frac{\partial s_1(\gamma, \log\sigma|x)}{\partial\log\sigma} =& -\frac{\partial s_2(x;\gamma, \log\sigma, u)}{\partial\gamma} = \frac{x^2/\sigma^2 - x/\sigma}{\left(1+\gamma\frac{x}{\sigma}\right)^2}, \\ -\frac{\partial s_2(\gamma, \log\sigma|x)}{\partial\log\sigma} =& (1+\gamma)\frac{x/\sigma}{\left(1+\gamma\cdot\frac{x}{\sigma}\right)^2}, \end{aligned}$$

$$
\begin{aligned}
-\frac{\partial w_1(\gamma,\log\sigma|x)}{\partial\gamma} =& \frac{2}{\gamma^3}\left(\log\left(1+\gamma\frac{x}{\sigma}\right)-\frac{\gamma\cdot x/\sigma}{1+\gamma\frac{x}{\sigma}}-\frac{\gamma^2}{2}\frac{x^2/\sigma^2}{\left(1+\gamma\frac{x}{\sigma}\right)^2}\right),\\
-\frac{\partial w_1(\gamma,\log\sigma|x)}{\partial\log\sigma} =& -\frac{\partial w_2(x;\gamma,\log\sigma,u)}{\partial\gamma}=\frac{x^2/\sigma^2}{\left(1+\gamma\frac{x}{\sigma}\right)^2},\\
-\frac{\partial w_2(\gamma,\log\sigma|x)}{\partial\log\sigma} =& \frac{x/\sigma}{\left(1+\gamma\cdot\frac{x}{\sigma}\right)^2}.
\end{aligned}
$$

Note that the above derivatives are well defined for $\gamma=0$ by continuity.

To control the Hessian matrix in the entire local parameter space Θ_n^ε, we use the following lemma. Its proof is the same as that of Lemma E.3 in He et al. (2022) and omitted.

Lemma B.2. *Uniformly for $(\gamma,\log\sigma)^\top\in\Theta_n^\varepsilon$ with $\varepsilon\in\left(0,\min\{\gamma_0+\frac{1}{2},\frac{1}{2}\}\right)$,*

$$
\|\nabla_{\boldsymbol{\theta}}s(\gamma,\log\sigma|Q_0(1-\alpha_n t)-u_n)-\nabla_{\boldsymbol{\theta}}s(\gamma_0,\log\sigma_{\alpha_n}|Q_0(1-\alpha_n t)-u_n)\|=o_p(1)\cdot\phi(t)
$$

and

$$
\|\nabla_{\boldsymbol{\theta}}w(\gamma,\log\sigma|Q_0(1-\alpha_n t)-u_n)-\nabla_{\boldsymbol{\theta}}w(\gamma_0,\log\sigma_{\alpha_n}|Q_0(1-\alpha_n t)-u_n)\|=o_p(1)\cdot\phi(t),
$$

where $\phi(t)=\sum_{i=1}^3(-\log t)^i+t^{\gamma_0}+t^{2\gamma_0}$, $t\in(0,1)$ and the $o_p(1)$-terms are uniform for $t\in\left(\frac{1}{2n\bar{\alpha}},1\right)$.

With a slight abuse of notation, define $U_i=S_0(T_i)$ and $V_i=S_0(C_i)$, where S_0 denotes the (uncensored) survival distribution of T. Combining Facts 1 and 3, we have

$$
\begin{aligned}
&-\frac{1}{n\alpha_n}\sum_{i=1}^n\nabla g_i(\boldsymbol{\theta}_0^{(n)})\\
=&\frac{1}{n\alpha_n}\sum_{i=1}^n\delta_i\mathbb{1}[U_i<\alpha_n]\Phi_{\gamma_0}(U_i/\alpha_n)+\frac{1}{n\alpha_n}\sum_{i=1}^n(1-\delta_i)\mathbb{1}[V_i<\alpha_n]\Psi_{\gamma_0}(V_i/\alpha_n).
\end{aligned}
$$

Observe that

$$
\mathbb{P}(V_i\le x,\delta_i=1)\le\mathbb{P}(U_i\le x,\delta_i=1)\le\mathbb{P}(U_i\le x)\le x.
$$

Together with integration by parts formula, we have, for $a \in (-1, \infty)$ and $\alpha \in (0, 1)$

$$
\begin{aligned}
0 \leq & \int_0^\alpha x^a d\mathbb{P}(V_i \leq x, \delta_i = 1) \\
\leq & \alpha^{a+1} + \{|a| + 1\} \int_0^\alpha \mathbb{P}(V_i \leq x, \delta_i = 1) x^{a-1} dx \\
\leq & \alpha^{a+1} + \{|a| + 1\} \int_0^\alpha x^a dx = \alpha^{a+1} + \{|a| + 1\} \frac{\alpha^{a+1}}{a+1} < \infty. \qquad (6)
\end{aligned}
$$

Similarly, we can show that for $1 \leq a \leq 3$ and $\alpha \in (0, 1)$

$$
0 \leq \int_0^\alpha (-\log x)^a d\mathbb{P}(V_i \leq x, \delta_i = 1) \leq (-\log \alpha)^a \alpha + (|a| + 1) \int_0^\alpha (-\log x)^{a-1} dx < \infty. \quad (7)
$$

Applying Proposition 1 entry-by-entry twice using (6), (7), and the assumption $2\gamma_0 > -1$ to $\{(U_i, \delta_i) : 1 \leq i \leq n\}$ and $\{(V_i, 1 - \delta_i) : 1 \leq i \leq n\}$, respectively, we have

$$
\begin{aligned}
-\frac{1}{n\alpha_n} \sum_{i=1}^n \nabla g_i(\boldsymbol{\theta}_0^{(n)}) =& -\frac{1}{n\bar{\alpha}} \sum_{i=1}^n \nabla g_i(\boldsymbol{\theta}_0) + o_p(1) \\
& \xrightarrow{p} -\frac{1}{\bar{\alpha}} \mathbb{E}[\nabla g_i(\boldsymbol{\theta}_0)] = \mathcal{I}(\bar{\alpha}),
\end{aligned}
$$

where $\boldsymbol{\theta}_0 = (\gamma_0, \log \sigma_{\bar{\alpha}})^\top$ and the last equation follows from Fact 2 and the law of iterated expectations. Furthermore, because U_i's are i.i.d. uniform variables, we have

$$
\min_{1 \leq i \leq n} S_0(X_i) = \min_{1 \leq i \leq n} \max\{U_i, V_i\} \geq \min_{1 \leq i \leq n} U_i \geq \frac{1}{2n}
$$

with probability tending to 1. Applying Lemma B.2 and the stable function $\phi(t) > 0$ therein, we have

$$
\begin{aligned}
& \sup_{\boldsymbol{\theta} \in \Theta_n^\varepsilon} \left\| \frac{1}{n\alpha_n} \sum_{i=1}^n \nabla g_i(\boldsymbol{\theta}) - \frac{1}{n\alpha_n} \sum_{i=1}^n \nabla g_i(\boldsymbol{\theta}_0^{(n)}) \right\| \\
&= o_p(1) \cdot \frac{1}{n\alpha_n} \sum_{i=1}^n \delta_i \mathbb{1}[U_i < \alpha_n] \phi\left(\frac{U_i}{\alpha_n}\right) + o_p(1) \cdot \frac{1}{n\alpha_n} \sum_{i=1}^n (1 - \delta_i) \mathbb{1}[V_i < \alpha_n] \phi\left(\frac{V_i}{\alpha_n}\right) \\
&\xrightarrow{p} 0 \cdot \frac{1}{\bar{\alpha}} \mathbb{E}\left[\delta_i \mathbb{1}[U_i < \bar{\alpha}] \phi\left(\frac{U_i}{\bar{\alpha}}\right)\right] + 0 \cdot \frac{1}{\bar{\alpha}} \mathbb{E}\left[(1 - \delta_i) \mathbb{1}[V_i < \bar{\alpha}] \phi\left(\frac{V_i}{\bar{\alpha}}\right)\right] = 0,
\end{aligned}
$$

where we apply Proposition 1 twice to $\{(U_i, \delta_i) : 1 \leq i \leq n\}$ and $\{(V_i, 1 - \delta_i) : 1 \leq i \leq n\}$, respectively, in the last line by invoking (6) and (7) again.

B.2 Existence of MLE and Joint Convergence

We first prove the existence of the maximum likelihood estimator, that is, part (1) of Theorem 2.1. It follows from claim (a) in the last section that, with probability tending to 1, the likelihood function $\sum_{i=1}^{n} l_i(\boldsymbol{\theta}|u_n)$ is well-defined and continuous in $\mathrm{cl}(\bar{\Theta}_n^{\varepsilon})$. Under such an event, applying the Weierstrass theorem yields the existence of a maximum point $\widehat{\boldsymbol{\theta}}_n \in \mathrm{cl}(\bar{\Theta}_n^{\varepsilon})$. However, uniformly for all boundary points $\boldsymbol{\theta} = \boldsymbol{\theta}_0^{(n)} + n^{-1/2+\varepsilon}\boldsymbol{w} \in \mathrm{cl}(\bar{\Theta}_n^{\varepsilon}) \setminus \bar{\Theta}_n^{\varepsilon}$ with $\|\boldsymbol{w}\| = 1$, it follows from Taylor expansion in conjunction with claims (b) and (c) that

$$
\begin{aligned}
&\frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\boldsymbol{\theta}|u_n) \\
=&\frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\boldsymbol{\theta}_0^{(n)}|u_n) + n^{-1/2+\varepsilon}\frac{1}{n\alpha_n}\sum_{i=1}^{n} g_i(\boldsymbol{\theta}_0^{(n)}|u_n)\boldsymbol{w} - n^{-1+2\varepsilon}\frac{1}{2}\boldsymbol{w}^T\left(\mathcal{I}(\bar{\alpha}) + o_p(1)\right)\boldsymbol{w} \\
=&\frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\boldsymbol{\theta}_0^{(n)}|u_n) + o_p((n\alpha_n)^{-1+2\varepsilon}) - n^{-1+2\varepsilon}\frac{1}{2}\boldsymbol{w}^T\mathcal{I}(\bar{\alpha})\boldsymbol{w}.
\end{aligned}
$$

As $\mathcal{I}(\bar{\alpha})$ is positive definite, with probability tending to 1, for these boundary points, we have

$$
\frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\boldsymbol{\theta}|u_n) < \frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\boldsymbol{\theta}_0^{(n)}|u_n) \le \frac{1}{n\alpha_n}\sum_{i=1}^{n} l_i(\widehat{\boldsymbol{\theta}}_n|u_n),
$$

implying that the estimator $\widehat{\boldsymbol{\theta}}_n$ is in the interior of $\mathrm{cl}(\Theta_n^{\varepsilon})$.

Next, we prove part (2) of Theorem 2.1. Using Taylor expansion, for some $\widehat{\boldsymbol{\theta}}^* \in \Theta_n^{\varepsilon}$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_n^{(0)}$

$$
\sqrt{n\alpha_n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_n^{(0)}\right) = \left(\frac{1}{n\alpha_n}\sum_{i=1}^{n}\nabla g_i(\widehat{\boldsymbol{\theta}}^*|u_n)\right)^{-1}\frac{1}{\sqrt{n\alpha_n}}\sum_{i=1}^{n} g_i\left(\boldsymbol{\theta}_0^{(n)}|u_n\right).
$$

Using claims (b) and (c) in the last section and replacing α_n with limit $\bar{\alpha}$ yields that, under the same Skorohod construction of Lemma A.1,

$$
\sqrt{n\bar{\alpha}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_n^{(0)}\right) = \sqrt{n\bar{\alpha}}\left(\widehat{\gamma} - \gamma_0, \log\widehat{\sigma} - \log\sigma_{\alpha_n}\right)^{\top} \xrightarrow{p} \mathcal{I}(\bar{\alpha})^{-1}\Upsilon;
$$

That is, using the delta method,

$$
\sqrt{n\bar{\alpha}}\left(\widehat{\gamma} - \gamma_0, \frac{\widehat{\sigma}}{\sigma_{\alpha_n}} - 1\right)^{\top} \xrightarrow{p} \mathcal{I}(\bar{\alpha})^{-1}\Upsilon. \tag{8}
$$

It remains to derive the limit of the product-limit process under the same Skorohod construction. Define the empirical process of cumulative hazard and the product-limit process by

$$\bar{\mathbb{L}}_n(x) = \sqrt{n}\left(\widehat{\Lambda}(x) - \Lambda_0(x)\right) \text{ and } \quad \bar{\mathbb{S}}_n(x) = \sqrt{n}\left(\widehat{S}_0(x) - S_0(x)\right)$$

The following lemma first appears in Breslow and Crowley (1974), Theorems 4 and 5. See Burke, Csörgő, and Horváth (1981), Theorem 4.2, for comments on the proofs in the aforementioned paper and corrections in the context of strong approximation. Recall the construction $(X_i, \delta_i) = (Q(1-U_i), \delta_i)$ and $V_i = \widetilde{\rho}(U_i)$ from Appendix A. The key idea is to exploit the relation

$$-\log S_0(t) = \Lambda_0(t) = \int_{S(t)}^{1} x^{-1} d\widetilde{\rho}(x),$$

and to approximate $-\log \widehat{S}_0(t)$ by the empirical cumulative hazard rate $\widehat{\Lambda}(t) = \sum_{X_i \le t} \frac{\delta_i}{\sum_j \mathbb{1}[X_j > X_i]}$ which can be rewritten as

$$\widehat{\Lambda}(t) = \int_{S(t)}^{1} [\rho_n(x)]^{-1} d\widetilde{\rho}_n(x),$$

Here, $(\rho_n(x), \widetilde{\rho}_n(x))$ are the empirical version of $(x, \widetilde{\rho}(x))$ given by

$$\begin{aligned}\rho_n(x) :=& n^{-1}\sum_{i=1}^{n} \mathbb{1}\left[U_i < x\right] \\ \widetilde{\rho}_n(x) :=& n^{-1}\sum_{i=1}^{n} \mathbb{1}\left[U_i < x, \delta_i = 1\right].\end{aligned}$$

Observe that $\bar{\mathbb{U}}_n = \sqrt{n}(\rho_n(x) - x)$ and $\bar{\mathbb{V}}_n(\widetilde{\rho}(x)) = \sqrt{n}(\widetilde{\rho}_n(x) - \widetilde{\rho}(x))$. Our proof is completely analogous to that of its bootstrap version in the next section, namely Lemma C.2 below, and therefore omitted. Note that the everywhere-continuity assumption of the censoring time C is not necessary as in Horváth and Yandell (1987); see also Horváth (1980).

Lemma B.3. *Under the Skorohod construction of Lemma A.1 for* $(\mathbb{U}_n, \mathbb{V}_n) = (\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$*,*

$$\sup_{0 \le x \le \tau} \left|\bar{\mathbb{L}}_n(x) + Z(S(x))\right| \xrightarrow{p} 0 \textit{ and } \sup_{0 \le x \le \tau} \left|\bar{\mathbb{S}}_n(x) - S_0(x) Z(S(x))\right| \xrightarrow{p} 0,$$

where $\tau \geq 0$ is any finite time point with $S(\tau) > 0$ and Z is the Gaussian process defined in Theorem 2.1 via the limit (B_1, B_2) of $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$.

Finally, combining (8) and Lemma B.3 completes the proof of Part (b) of Theorem 2.1.

C Proof of Theorem 2.2

C.1 Weak Convergence of Random Weighted Product-Limit Process

For presentation convenience, we omit the superscript and just write the random weights $\xi = (\xi_1, \ldots, \xi_n)^\top$ in short of $\xi^{(b)} = (\xi_1^{(b)}, \ldots, \xi_n^{(b)})^\top$. Define the random weighted estimator of the cumulative hazard rate by

$$\widehat{\Lambda}(t;\xi) := \sum_{X_i \leq t} \frac{\delta_i \xi_i}{\sum_{X_j > X_i} \xi_j}.$$

Note that the denominator $\sum_{X_i > t} \xi_i \geq \xi_{n,n} > 0$ when $X_{n,n} > t$, which occurs with probability tending to 1 uniformly for $t \in [0, \tau]$. The following lemma bounds the difference between the estimators $-\log \widehat{S}(t;\xi)$ and $\widehat{\Lambda}(t;\xi)$ of the cumulative hazard rate stochastically.

Lemma C.1. *Let τ be any finite value such that $S(\tau) > 0$. For i.i.d. random weights ξ_i's from a subexponential distribution with mean one (possibly degenerate),*

$$\sup_{0 \leq t \leq \tau} \left| -\log \widehat{S}_0(t;\xi) - \widehat{\Lambda}(t;\xi) \right| = O_p\left((\log n)^2 n^{-1} \right).$$

Proof. Let $N(t) = \sum_{i=1}^n \mathbb{1}[X_i > t]$ and $\Xi(i) = \sum_{j > i} \xi_{j,n}$. Using the condition $S(\tau) > 0$, with probability tending to 1, $\tau < X_{n,n}$ and thus $\Xi(i) > 0$ uniformly for i such that $X_{i:n} \leq \tau$. First, define an approximation of $\widehat{\Lambda}(t;\xi)$ given by

$$\widetilde{\Lambda}(t;\xi) = \sum_{X_{i:n} \leq t} \frac{\delta_{i,n} \xi_{i,n}}{\Xi(i) + \xi_{i,n}}$$

and we shall show that

$$\sup_{0\le t\le \tau}\left|-\log \widehat{S}_0(t;\xi)-\widetilde{\Lambda}(t;\xi)\right| = O_p\left((\log n)^2\, n^{-1}\right). \tag{9}$$

Note the following elementary inequality:

$$0 < -\log(1-(z+1)^{-1})-(z+1)^{-1} < (z(z+1))^{-1} < z^{-2}, \quad z>0.$$

Substituting $z=\Xi(i)/\xi_{i,n}$ gives that

$$\begin{aligned}
0 &< -\log \Pi_{X_{i:n}\le t}\left(1-\frac{\xi_{i,n}}{\Xi(i)+\xi_{i,n}}\right)^{\delta_{i,n}} - \sum_{X_{i:n}\le t}\frac{\delta_{i,n}\xi_{i,n}}{\Xi(i)+\xi_{i,n}} \\
&= \sum_{X_i\le t}\delta_i\left\{-\log\left(1-\frac{\xi_i}{\Xi(i)+\xi_{i,n}}\right)-\frac{\xi_i}{\Xi(i)+\xi_{i,n}}\right\} \\
&< \sum_{X_i\le t}\delta_i\left(\frac{\Xi(i)}{\xi_i}\right)^{-2} \le \left(\max_{1\le i\le n}\xi_i\right)^2\sum_{i=1}^{n-N(t)}(\Xi(i))^{-2} \le \left(\max_{1\le i\le n}\xi_i\right)^2\sum_{i=1}^{n-N(\tau)}(\Xi(i))^{-2}.
\end{aligned} \tag{10}$$

Next, we use the well-known bound of the sample maximum for sub-exponential distribution:

$$\max_{1\le i\le n}\xi_i = O_p(\log n), \tag{11}$$

which can be shown using the Bonferroni method and the sub-exponential tail probability bound. Moreover, it is elementary to show that $N(\tau)/(nS(\tau)) \xrightarrow{p} 1$ using the Hoeffding's inequality, and therefore with probability tending to one

$$\sum_{j=1}^{n-N(\tau)}(\Xi(j))^{-2} \le \sum_{j=1}^{n-\frac{1}{2}nS(\tau)}(\Xi(j))^{-2}.$$

But by Bernstein's inequality, there exists some absolute constant $M>0$ and sub-exponential norm $K>0$ such that for small $t>0$,

$$\begin{aligned}
\sum_{j=1}^{n-\frac{1}{2}nS(\tau)}\mathbb{P}\left(\left|\frac{\Xi(j)}{n-j}-1\right|>t\right) &\le \sum_{j=1}^{n-\frac{1}{2}NS(\tau)}2(n-j)\exp\left(-M\frac{(n-j)t^2}{K^2}\right) \\
&\le \sum_{j=\frac{1}{2}nS(\tau)}^{n-1}2j\exp\left(-M\frac{jt^2}{K^2}\right)\to 0
\end{aligned}$$

as the lower bound $\frac{1}{2}nS(\tau) \to \infty$. It follows from the Bonferroni method that

$$\sum_{j=1}^{n-N(\tau)} (\Xi(j))^{-2} = O_p\left(\sum_{j=1}^{n-\frac{1}{2}nS(\tau)} (n-j)^2\right) = O_p(1/(nS(\tau))) = O_p(n^{-1}).$$

Combining this with (10) and (11) yields (9). Next, observe that

$$\begin{aligned}
0 \leq & \widehat{\Lambda}(t;\xi) - \widetilde{\Lambda}(t;\xi) \\
= & \sum_{X_{i:n} \leq t} \frac{\delta_{i,n}\xi_{i,n}(\sum_{j \geq i} \xi_{j,n} - \sum_{X_{j:n} > X_{i:n}} \xi_{j,n})}{\left(\sum_{X_j > X_{i:n}} \xi_j\right)\left(\sum_{j \geq i} \xi_{j,n}\right)} \\
\leq & \left(\sum_{X_j > \tau} \xi_j\right)^{-2} \sum_{X_i \leq \tau} \left(\delta_i \xi_i \sum_{X_j = X_i} \xi_j\right) \\
\leq & \left(\sum_i \xi_i \mathbb{1}[X_i > \tau]\right)^{-2} \sum_i \sum_j \left(\delta_i \xi_i \xi_j \mathbb{1}[X_j = T_i]\right).
\end{aligned}$$

The first term $\sum_i \xi_i \mathbb{1}[X_i > \tau] \geq \xi_{n,n} \mathbb{1}[X_{n:n} > \tau] > 0$ is positive with probability tending to one. Furthermore, because $\mathbb{P}(T_j = T_i) = 0$ by the continuity of S_0,

$$\mathbb{P}(X_j = T_i, \delta_i = 1) = \mathbb{P}(C_j = T_i, \delta_i = 1) = 0,$$

where we also use the independence between C_j and (T_i, δ_i) in the last step. This implies that $\sum_i \sum_j (\delta_i \xi_i \xi_j \mathbb{1}[X_j = T_i]) = 0$ with probability one. It follows that $\widehat{\Lambda}(t;\xi) = \widetilde{\Lambda}(t;\xi)$ uniformly for $0 \leq t \leq \tau$ with probability tending to one. Combining with (9) completes the proof. $\square$

Define the random weighted bootstrapped empirical processes of cumulative hazard rate and survival function by

$$\widehat{\mathbb{L}}_n(x) = \sqrt{n}\left(\widehat{\Lambda}(x;\xi) - \widehat{\Lambda}(x)\right), \quad \widehat{\mathbb{S}}_n(x) = \sqrt{n}\left(\widehat{S}_0(x;\xi) - \widehat{S}_0(x)\right).$$

The following lemma is a random weighted bootstrap analogy of Lemma B.3.

Lemma C.2. *Under the same Skorohod construction of Lemma A.1 but for* $(\mathbb{U}_n, \mathbb{V}_n) = (\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$,

$$\sup_{0 \le t \le \tau} \left| \widehat{\mathbb{L}}_n(t) + Z(S(t)) \right| \xrightarrow{p} 0, \quad \sup_{0 \le t \le \tau} \left| \widehat{\mathbb{S}}_n(t) - S_0(x) Z(S(t)) \right| \xrightarrow{p} 0$$

for any finite τ *with* $S(\tau) > 0$, *where* Z *is a Gaussian process defined as in Theorem 2.1 in terms of the limit* (B_1, B_2) *of* $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$.

Proof. By Lemma C.1,

$$\sup_{0 \le t \le \tau} \left| \widehat{S}_0(x; \xi) - \exp(-\widehat{\Lambda}(t; \xi)) \right| = O_p((\log n)^2 n^{-1}) = o_p(n^{-1/2}) \text{ and}$$
$$\sup_{0 \le t \le \tau} \left| \widehat{S}_0(x) - \exp(-\widehat{\Lambda}(t)) \right| = O_p((\log n)^2 n^{-1}) = o_p(n^{-1/2}),$$

where the second equation is for the special case $\xi_1 = \ldots = \xi_n = 1$. It follows that

$$\sup_{0 \le t \le \tau} \left| \widehat{\mathbb{S}}_n(t) - \sqrt{n} \left(\exp(-\widehat{\Lambda}(t; \xi)) - \exp(-\widehat{\Lambda}(t)) \right) \right| \xrightarrow{p} 0.$$

This means that we only need to prove the convergence of $\widehat{\mathbb{L}}_n$, as the convergence of $\widehat{\mathbb{S}}_n$ follows by the delta method. Recall from Appendix A that, with probability one, $(X_i, \delta_i) = (Q(1 - U_i), \delta_i)$ where U_i's are i.i.d. uniform variables on $[0, 1]$ and Q is the generalized quantile function of X_i. Moreover, invoke the definition of $V_i = \widetilde{\rho}(U_i)$, where $\widetilde{\rho}(x) = \mathbb{P}(U_i < x, \delta_i = 1)$ is a continuous improper distribution function with total mass $p_1 := \widetilde{\rho}(\infty) = \mathbb{P}(\delta_i = 1) > 0$. Using the same trick in Breslow and Crowley (1974) and integrating by substitution, we can rewrite that

$$\widehat{\Lambda}(t; \xi) = \int_{S(t)}^{1} [\rho_n(x; \xi)]^{-1} d\widetilde{\rho}_n(x; \xi) \text{ and } \widehat{\Lambda}(t) = \int_{S(t)}^{1} [\rho_n(x)]^{-1} d\widetilde{\rho}_n(x),$$

with probability one, where

$$\rho_n(x; \xi) := n^{-1} \sum_{i=1}^{n} \xi_i \mathbb{1}\left[U_i < x\right] \text{ and } \widetilde{\rho}_n(x; \xi) := n^{-1} \sum_{i=1}^{n} \xi_i \mathbb{1}\left[U_i < x, \delta_i = 1\right].$$

The functions $\rho_n(\cdot)$ and $\widetilde{\rho}_n(\cdot)$ are defined similarly as in section B.2 but with equal weights. Observe that $\widehat{\mathbb{U}}_n(x) = \sqrt{n}(\rho_n(x;\xi) - \rho_n(x))$ and $\widehat{\mathbb{V}}_n(\widetilde{\rho}(x)) = \sqrt{n}(\widetilde{\rho}_n(x;\xi) - \widetilde{\rho}_n(x))$. Following the proof of Lemma 6.1 in Horváth and Yandell (1987) and integrating by substitution, we decompose that

$$\widehat{\mathbb{L}}_n(t) = -\mathbb{Z}_n(S(t)) + A^{(1)}(S(t)) + A^{(2)}(S(t)) + A^{(3)}(S(t)),$$

with

$$\mathbb{Z}_n(s) = \int_s^1 \widehat{\mathbb{U}}_n(x)x^{-2}d\widetilde{\rho}(x) - \int_s^1 x^{-1}d\widehat{\mathbb{V}}_n(\widetilde{\rho}(x)),$$

$$\begin{aligned}
A^{(1)}(s) &= \int_s^1 \mathbb{U}_n(x)\left(x^{-2} - [\rho_n(x;\xi)\rho_n(x)]^{-1}\right) d\widetilde{\rho}_n(x;\xi),\\
A^{(2)}(s) &= \int_s^1 \left([\rho_n(x)]^{-1} - x^{-1}\right) d\widehat{\mathbb{V}}_n(\widetilde{\rho}(x)),\\
A^{(3)}(s) &= \int_s^1 \mathbb{U}_n(x)x^{-2}d(\widetilde{\rho}(x) - \widetilde{\rho}_n(x;\xi)).
\end{aligned}$$

The leading term (i.e., the first term above) gives the desired limit by using Lemma A.1 with $(\mathbb{U}_n, \mathbb{V}_n) = (\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$ and under the Skorohod construction therein,

$$\sup_{S(\tau)\le s\le 1} |\mathbb{Z}_n(s) - Z(s)| \xrightarrow{p} 0.$$

It remains to show that $\sup_{S(\tau)\le s\le 1} |A^{(i)}(s)| = o_p(1)$ for all $i = 1, 2, 3$. By the uniform consistency of $\rho_n(x;\xi)$ and $\rho_n(x)$ on $[S(\tau), 1]$ (implies by the functional CLT) and the stochastic boundedness of $\mathbb{U}_n(x)$,

$$\begin{aligned}
\sup_{S(\tau)\le s\le 1} |A^{(1)}(s)| \le & \sup_{S(\tau)\le x\le 1} |\mathbb{U}_n(x)| \cdot \sup_{S(\tau)\le x\le 1} |x^{-2} - [\rho_n(x;\xi)\rho_n(x)]^{-1}| \cdot \widetilde{\rho}_n(1;\xi)\\
=& O_p(1)\cdot o_p(1)\cdot O_p(1) = o_p(1).
\end{aligned}$$

Let $\widehat{\mathbb{W}}_n(x) = \widehat{\mathbb{V}}_n(\widetilde{\rho}(x)) - \widehat{\mathbb{V}}_n(\widetilde{\rho}(1))$, $x \in [0, 1]$, such that $\widehat{\mathbb{W}}_n(1) = 0$ by construction. By Lemma A.1,

$$\sup_{S(\tau)\le x\le 1} \left|\widehat{\mathbb{W}}_n(x) - W_2(x)\right| \xrightarrow{p} 0 \text{ and } W_2(x) = B_2(\widetilde{\rho}(x)) - B_2(\widetilde{\rho}(1)). \tag{12}$$

We can rewrite that

$$A^{(2)}(s) = \int_s^1 \left([\rho_n(x)]^{-1} - x^{-1}\right) d\widehat{\mathbb{W}}_n(x) = \int_s^1 [\rho_n(x)]^{-1} d\widehat{\mathbb{W}}_n(x) - \int_s^1 x^{-1} d\widehat{\mathbb{W}}_n(x).$$

Integrating by parts and using (12), uniformly for all $s \geq S(\tau) > 0$, we have

$$\int_s^1 x^{-1} d\widehat{\mathbb{W}}_n(x) = -s^{-1}\widehat{\mathbb{W}}_n(s) + \int_s^1 \widehat{\mathbb{W}}_n(x) x^{-2} dx = -s^{-1} W_2(s) - \int_s^1 W_2(x) dx^{-1} + o_p(1).$$

On the other hand,

$$\int_s^1 [\rho_n(x)]^{-1} d\widehat{\mathbb{W}}_n(x) = -[\rho_n(s)]^{-1}\widehat{\mathbb{W}}_n(s) - \int_s^1 \widehat{\mathbb{W}}_n(x) d[\rho_n(x)]^{-1}.$$

Again, the first term above converges in probability to $-s^{-1}W_2(s)$ uniformly for all $s \geq S(\tau) > 0$. It remains to verify the limit of the second term. Let $U_{1,n} \geq \ldots \geq U_{n,n}$ be the order statistics of $U_1, \ldots, U_n$. For every s, define $N(s) = \sum_{i=1}^n \mathbb{1}[U_i \geq s]$. Like above, we have that

$$\begin{aligned}
&\int_s^1 \widehat{\mathbb{W}}_n(x) d[\rho_n(x)]^{-1} \\
=&\sum_{i=1}^{N(s)} \widehat{\mathbb{W}}_n(U_{i,n}) \left([\rho_n(U_{i,n})]^{-1} - [\rho_n(U_{i+1,n})]^{-1}\right) \\
=&\sum_{i=1}^{N(s)} W_2(U_{i,n}) \left(U_{i,n}^{-1} - U_{i+1,n}^{-1}\right) + o_p(1) =: R_n(s) + o_p(1)
\end{aligned}$$

uniformly for $s \geq S(\tau)$. Note that the first term $R_n(s)$ is an approximation of $R(s) := \int_s^1 W_2(x) dx^{-1}$. Take any $p_\tau \in (1 - S(\tau), 1)$. Note that $N(s)/n \leq N(S(\tau))/n \xrightarrow{p} 1 - S(\tau)$, and thus $N(s) \leq np_\tau$ uniformly for $s \geq S(\tau)$ with probability tending to one. Moreover, by the uniform convergence of uniform variables (see, e.g., Theorem 0 in Wellner, 1978),

$$\sup_{1 \leq i \leq np_\tau} |U_{i,n} - (1 - i/n)| \xrightarrow{p} 0.$$

Using triangle inequality, uniformly for $s \in [S(\tau), 1]$

$$\begin{aligned}
\sup_{1 \leq i \leq N(s)} |U_{i+1,n} - U_{i,n}| \leq& \sup_{1 \leq i \leq np_\tau - 1} |U_{i+1,n} - U_{i,n}| \\
\leq& \frac{1}{n} + 2 \sup_{1 \leq i \leq np_\tau} |U_{i,n} - (1 - i/n)| = o_p(1),
\end{aligned}$$

where the first inequality holds with probability tending to one. As the sample path of W_2 is uniformly continuous on $[S(\tau), 1]$, this implies that

$$
\begin{aligned}
R_n(s) =& \sum_{i=1}^{N(s)} W_2(1-i/n)\left((1-i/n)^{-1} - (1-(i+1)/n)^{-1}\right) + o_p(1) \\
=& \frac{1}{n}\sum_{i=1}^{N(s)} W_2(1-i/n)(1-i/n)^{-2} + o_p(1)
\end{aligned}
$$

uniformly for $s \in [S(\tau), 1]$. Now,

$$
\begin{aligned}
\sup_{S(\tau)\le s\le 1} |R_n(s) - R(s)| \le& \sum_{i=1}^{np_\tau} \int_{1-i/n}^{1-(i-1)/n} \left|W_2(x)x^{-2} - W_2(1-i/n)(1-i/n)^{-2}\right| dx \\
\le& \frac{1}{n}\sum_{i=1}^{np_\tau} \sup_{1-i/n\le x,y\le 1-(i-1)/n} \left|W_2(x)x^{-2} - W_2(y)y^{-2}\right| \\
\le& p_\tau \sup_{|x-y|<1/n, x,y\ge 1-p_\tau} \left|W_2(x)x^{-2} - W_2(y)y^{-2}\right| = o_p(1),
\end{aligned}
$$

where the last step follows from the stochastic continuity of W_2 (due to that of the Brownian bridge B_2). This completes the proof of $\sup_{S(\tau)\le s\le 1} |A^{(2)}(s)| = o_p(1)$. Finally, decompose that

$$
A^{(3)}(s) = -\int_s^1 (\rho_n(x)/x^2 - x^{-1}) d\widehat{\mathbb{V}}_n(\widetilde{\rho}(x)) - \int_s^1 (\rho_n(x)/x^2 - x^{-1}) d\mathbb{V}_n(\widetilde{\rho}(x)).
$$

Following similar arguments for $A^{(2)}(s)$, we can show that each term above is $o_p(1)$ uniformly for $s \in [S(\tau), 1]$. This completes the proof. $\square$

C.2 Existence of Boostrap MLE and Joint Convergence of Boostrapped Elements

Like in Appendix B, we have the following results:

(a) With probability tending to one, the log-likelihood function is well-defined, that is,

$$
\sum_{i=1}^{n} \xi_i^{(b)} l_i(\boldsymbol{\theta}|u_n) > -\infty
$$

uniformly in the parameter space $\mathrm{cl}(\Theta^{\varepsilon}_{n,b})$ with $\varepsilon \in (0, \min\{\gamma_0 + 1/2, 1/2\})$, where $\mathrm{cl}(\cdot)$ denotes the set closure.

(b) Under the probability space of Lemma A.1,

$$\frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} \left(\xi_i^{(b)} - 1\right) g_i(\boldsymbol{\theta}_0^{(n,b)} | u_n) \xrightarrow{p} \widehat{\Upsilon},$$

where $\widehat{\Upsilon}$ is defined the same way as Υ in the proof of Theorem 2.1 in terms of the limiting Brownian bridges (B_1, B_2) of $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$ instead of $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$. Note that here all random elements are only equal to the original ones in distributions.

(c) $\sup_{\boldsymbol{\theta} \in \mathrm{cl}(\Theta^{\varepsilon}_{n,b})} \left\| \frac{1}{n\alpha_n} \sum_{i=1}^{n} \xi_i^{(b)} \nabla g_i(\boldsymbol{\theta}|u_n) + \mathcal{I}(\bar{\alpha}) \right\| \xrightarrow{p} 0$, where the unconditional Fisher information matrix $\mathcal{I}(\bar{\alpha})$ is defined in equation (7) in the main document.

The proofs of these claims are completely analogous to that in Appendix B: Claim (a) follows because the weights $\xi_i^{(b)} > 0$ does not change the finiteness of the likelihood function; The proof of claim (b) is completely analogous, except replacing $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$ with $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$ everywhere; The proof of claim (c) is also the same by applying Proposition 1 with the random weights. We do not repeat the details.

For part (1) of Theorem 2.2, following the proof of part (1) of Theorem 2.1 in Appendix B, it remains to check that

$$\frac{1}{\sqrt{n\alpha_n}} \sum_{i=1}^{n} \xi_i^{(b)} g_i \left(\boldsymbol{\theta}_0^{(n,b)} | u_n^{(b)}\right) = O_p(1),$$

by combining claim (b) here with claim (b) in Subsection B by taking $u_n = u_n^{(b)}$ therein.

Next, we prove part (2) of the theorem. Invoke the probability space from Lemma A.1. Let $\sqrt{n\bar{\alpha}}(\widehat{\theta} - \theta_0) \xrightarrow{p} Z$, where the limiting variable Z depends only on $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$ and its distribution is given in Theorem 2.1 or Corollaries 2.1–2.2. Following Csörgő and Mason (1989), we only need to show that

$$\sqrt{n\bar{\alpha}}(\widehat{\theta}^{(b)} - \widehat{\theta}) \xrightarrow{p} \widetilde{Z} \sim Z, \tag{13}$$

where $\widetilde{Z}$ is a random variable only based on $(\widehat{\mathbb{U}}_n, \widehat{\mathbb{V}}_n)$ that is independent of $(\bar{\mathbb{U}}_n, \bar{\mathbb{V}}_n)$.

By part (1) of Theorem 2.1, in the same neighborhood $\Theta^{\varepsilon}_{n,b}$ for the bootstrap threshold $u_n^{(b)}$, with probability tending to 1 there exists a unweighted MLE $\widetilde{\boldsymbol{\theta}}^{(b)} = (\widetilde{\gamma}^{(b)}, \log \widetilde{\sigma}^{(b)})^\top$ solving

$$\sum_{i=1}^{n} g_i(\widetilde{\boldsymbol{\theta}}^{(b)} | u_n^{(b)}) = 0.$$

Using claims (b) and (c) above and Taylor expansion,

$$\begin{aligned}\frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} \xi_i^{(b)} g_i(\boldsymbol{\theta}_0^{(n,b)} | u_n^{(b)}) =& \frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} \xi_i^{(b)} g_i(\boldsymbol{\theta}_0^{(n,b)} | u_n^{(b)}) - \frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} \xi_i^{(b)} g_i(\widehat{\boldsymbol{\theta}}^{(n,b)} | u_n^{(b)}) \\ =& - (\mathcal{I}(\bar{\alpha}) + o_p(1)) \sqrt{n\alpha_n^{(b)}} (\boldsymbol{\theta}_0^{(n,b)} - \widehat{\boldsymbol{\theta}}^{(n,b)}) \\ =& - \mathcal{I}(\bar{\alpha}) \sqrt{n\alpha_n^{(b)}} (\boldsymbol{\theta}_0^{(n,b)} - \widehat{\boldsymbol{\theta}}^{(n,b)}) + o_p(1),\end{aligned}$$

where in the last step, we recall that the beginning term in the first line is $O_p(1)$ from the proof of part (1), and so is $\sqrt{n\alpha_n^{(b)}}(\boldsymbol{\theta}_0^{(n,b)} - \widehat{\boldsymbol{\theta}}^{(n,b)})$ by the penultimate step. Note that we use $\widehat{\boldsymbol{\theta}}^{(n,b)}$ to denote the random weighted maximum likelihood estimator for $\boldsymbol{\theta}_0^{(n,b)}$. Similarly, but using claims (b) and (c) from Subsection B,

$$\begin{aligned}\frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} g_i(\boldsymbol{\theta}_0^{(n,b)} | u_n^{(b)}) =& \frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} g_i(\boldsymbol{\theta}_0^{(n,b)} | u_n^{(b)}) - \frac{1}{\sqrt{n\alpha_n^{(b)}}} \sum_{i=1}^{n} g_i(\widetilde{\boldsymbol{\theta}}^{(b)} | u_n^{(b)}) \\ =& - (\mathcal{I}(\bar{\alpha}) + o_p(1)) \sqrt{n\alpha_n^{(b)}} (\boldsymbol{\theta}_0^{(n,b)} - \widetilde{\boldsymbol{\theta}}^{(b)}) \\ =& - \mathcal{I}(\bar{\alpha}) \sqrt{n\alpha_n^{(b)}} (\boldsymbol{\theta}_0^{(n,b)} - \widetilde{\boldsymbol{\theta}}^{(b)}) + o_p(1).\end{aligned}$$

Subtracting these expansions and inverting the information matrix $\mathcal{I}(\bar{\alpha})$,

$$\sqrt{n\bar{\alpha}}(\widehat{\boldsymbol{\theta}}^{(n,b)} - \widetilde{\boldsymbol{\theta}}^{(b)}) \xrightarrow{p} \mathcal{I}^{-1}(\bar{\alpha})\widehat{\Upsilon}. \tag{14}$$

Note that $\widehat{\boldsymbol{\theta}}^{(n,b)}$ is centered around $\widetilde{\boldsymbol{\theta}}^{(b)}$ rather than $\widehat{\boldsymbol{\theta}}$ (the bootstraped estimator is centered around the empirical estimator rather than the population value) in the previous equations. Introduce the intermediate estimator $\widetilde{\theta}^{(b)}$ to be the estimator of θ_0 using the threshold $u_n^{(b)}$

instead of u_n. Applying the delta method with (14) and using Lemma C.2, we can show that

$$\sqrt{n\bar{\alpha}}(\widehat{\theta}^{(b)} - \widetilde{\theta}^{(b)}) \xrightarrow{p} \widetilde{Z} \sim Z. \tag{15}$$

Next, it is essential to observe that both $\widetilde{\theta}^{(b)}$ and $\widehat{\theta}$ have the same probabilistic limit centering around a common population value θ_0 under the Skorohod construction in the proof of Theorem 2.1. That is, under such Skorohod construction,

$$\sqrt{n\bar{\alpha}}(\widetilde{\theta}^{(b)} - \theta_0) = \sqrt{n\bar{\alpha}}(\widehat{\theta} - \theta_0) + o_p(1).$$

Canceling the common terms yields that

$$\sqrt{n\bar{\alpha}}(\widetilde{\theta}^{(b)} - \widehat{\theta}) = o_p(1).$$

Now combining this with (15) yields (13).

References

Breslow, N., and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," *The Annals of Statistics*, 2(3), 437–453.

Burke, M. D., Csörgő, S., and Horváth, L. (1981), "Strong Approximations of Some Biometric Estimates under Random Censorship," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 56, 87–112.

Chibisov, D. (1964), "Some Theorems on the Limiting Behaviour of Empirical Distribution Functions," *Selected Translations in Mathematical Statistics and Probability*, 6, 147–156.

Csörgő, S., and Mason, D. M. (1989), "Bootstrapping Empirical Functions," *The Annals of Statistics*, 17(4), 1447–1471.

Gertsbakh, I. (1995), "On the Fisher Information in Type-I Censored and Quantal Response Data," *Statistics & Probability Letters*, 23, 297–306.

He, Y., Peng, L., Zhang, D., and Zhao, Z. (2022), "Risk Analysis via Generalized Pareto Distributions," *Journal of Business & Economic Statistics*, 40(2), 852–867.

Horváth, L. (1980), "Dropping Continuity and Independence Assumptions in Random Censorship Models," *Studia Scientiarum Mathematicarum Hungarica*, 15, 381–389.

Horváth, L., and Yandell, B. S. (1987), "Convergence Rates for the Bootstrapped Product-Limit Process," *The Annals of Statistics*, 15(3), 1155–1173.

O'Reilly, N. E. (1974), "On the Weak Convergence of Empirical Processes in Sup-Norm Metrics," *The Annals of Probability*, 2(4), 642–651.

Shorack, G. R., and Wellner, J. A. (1982), "Limit Theorems and Inequalities for the Uniform Empirical Process Indexed by Intervals," *The Annals of Probability*, 10(3), 639–652.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, New York: Wiley.

Wellner, J. A. (1978), "Limit Theorems for the Ratio of the Empirical Distribution Function to the True Distribution Function," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 45, 73–88.