A Supplemental Appendix Section

Weighted NPMLE for the subdistribution of a competing risk

by Anna Bellach, Michael R Kosorok, Ludger Rüschendorf and Jason P Fine

A.1 Model conditions

M1) The cumulative baseline $A_0(t)$ is a strictly increasing and continuously differentiable function and β_0 lies in the interior of a compact set $\mathcal{C}$.

M2) The vector of covariates $Z(t)$ is P-almost surely of bounded variation on the observed interval $[0, \tau]$.

M3) The endpoint of the study τ is chosen in a way that P-almost surely there exists a constant $\delta > 0$ such that $P(C \geq \tau | Z) > \delta$ and $P(X \geq \tau | Z) > \delta$.

M4) $\mathcal{G}$ is a thrice continuously differentiable and strictly increasing function with $\mathcal{G}(0) = 0$, $\mathcal{G}'(0) > 0$ and $\mathcal{G}(\infty) = \infty$. In addition to that one of the following conditions is required:

a) $\mathcal{G}''(x) \leq 0$ for $x > 0$ or

b) $\mathcal{G}''(x) \geq 0$ for $x > 0$. In addition to that for any $a \in (0, \infty)$ and for any sequence $(x_n) \subset \mathbb{R}$ with $x_n \to \infty$ as $n \to \infty$,

$$(*) \quad \lim_{n\to\infty} \mathbb{P}(X \leq \tau, \Delta\varepsilon = 1) \left[\frac{\log(ax_n)}{\mathcal{G}(a^{-1}x_n)} + \frac{\log \mathcal{G}'(ax_n)}{\mathcal{G}(a^{-1}x_n)} \right] < P(X \geq \tau).$$

M5) *Identifiability condition*: If $h_1 \in \mathbb{R}^d$ and $h_2 \in \mathcal{D}[0, \tau]$ exist such that $h_1^T Z(t) + h_2(t) = 0$ P-almost surely, then $h_1 = 0$ and $h_2(t) = 0$ $\forall t \in [0, \tau]$.

M6) For any $h_1 \in \mathbb{R}^d$ and for any $h_2 \in \mathcal{D}[0, \tau]$ exists a subset $\mathcal{S} \subset [0, \tau]$ of nonzero Lebesgue measure such that $\forall t \in \mathcal{S}$

$$h_1^T Z(t) + h_2(t) \neq -\frac{\mathcal{G}''\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u)\right)}{\mathcal{G}'\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u)\right)} \int_0^t e^{\beta^T Z(u)} \big(h_1^T Z(u) + h_2(u)\big) dA_0(u).$$

For the consistency of the weighted NPMLE, model conditions M1)-M5) and twice continuous differentiability of $\mathcal{G}$ are sufficient. Model conditions M1)-M2),M5)-M6) and thrice continuous differentiability of $\mathcal{G}$ are sufficient to obtain weak convergence. For the proportional hazards (Fine-Gray) model conditions M5) and M6) are equivalent. Condition M4a) is satisfied for example for the class of logarithmic transformation models and for the Box-Cox transformation models with $\rho \in [0,1]$. M4b) holds for the Box-Cox transformation models with $\rho > 1$. It is sufficient for $(*)$ to prove that $\log(ax_n)/\mathcal{G}(a^{-1}x_n) \to 0$ and $\log \mathcal{G}'(ax_n)/\mathcal{G}(a^{-1}x_n) \to 0$ for $x_n \to \infty$.

A.2 Consistency

We show that $\widehat{A}_n^0(\tau)$ is bounded P-almost surely, thereby ascertaining the existence of the weighted NPMLE. Then $\widehat{A}_n^0(t)$ is bounded on $[0,\tau]$ uniformly P-almost surely. From Helly's selection theorem it is then obtained that $\exists n_k \subset \mathbb{N}: \widehat{A}_{n_k}^0 \xrightarrow{*} A^*$, for a limit A^*, with $*$ denoting outer almost sure convergence. With a Kullback-Leibler argument it is ascertained that every subsequence $(\widehat{\beta}_{n_k}, \widehat{A}_{n_k}^0)$ converges to the true parameter (β_0, A_0). $\widehat{A}_n(t)$ is a sequence of monotone increasing functions, and the limit $A_0(t)$ is continuous. From this $\|\widehat{A}_n - A_0\|_{\ell^\infty[0,\tau]} \to 0\,[P]$ and $|\widehat{\beta}_n - \beta_0| \to 0\,[P]$ where $\ell^\infty[0,\tau]$ denotes the space of bounded functions on $[0,\tau]$.

Existence of $(\widehat{\beta}_n, \widehat{A}_n^0)$ and boundedness of $\widehat{A}_n^0(\tau)$ under model condition M4a). We define $\widetilde{A}_n^0(t) := n^{-1}\sum_{i=1}^n N_i(t)$. By Jensen's inequality it is ascertained that

$$n^{-1}\left[\ell(\widehat{A}_n^0, \widehat{\beta}_n) - \ell(\widetilde{A}_n^0, \beta_0)\right]$$

$$\leq \left[\mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\} \frac{\log \mathcal{G}\big(e^{-M}\widehat{A}_n^0(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}_n^0(\tau)\big)} - \mathbb{P}_n \mathbb{1}\{X \geq \tau\}\right] \mathcal{G}\big(e^{-M}\widehat{A}_n^0(\tau)\big) + O_p(1).$$

Existence of $(\widehat{\beta}_n, \widehat{A}^0_n)$ and boundedness of $\widehat{A}^0_n(\tau)$ under model condition M4b).

$$n^{-1}\big[\ell(\widehat{A}^0_n, \widehat{\beta}_n) - \ell(\widehat{A}^0_n, \beta_0)\big]$$

$$\leq \left\{\mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\}\left[\frac{\log \mathcal{G}'\big(e^M \widehat{A}^0_n(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}^0_n(\tau)\big)} + \frac{\log\big(e^M \widehat{A}^0_n(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}^0_n(\tau)\big)}\right] - \mathbb{P}_n \mathbb{1}\{X \geq \tau\}\right\}$$
$$\times \mathcal{G}\big(e^{-M}\widehat{A}^0_n(\tau)\big) + O_p(1).$$

The conclusion in both cases is that if $\widehat{A}^0_n(\tau)$ became infinitely large, the right hand side would go to $-\infty$, which contradicts the definition of $(\widehat{A}_n, \widehat{\beta}_n)$ as a maximum likelihood estimator.

Competing risks setting with administrative censoring. From differentiating the discretized log-likelihood with respect to jump sizes we obtain

$$\widehat{A}^0_n(t) = \sum_{i:\, X_i \leq t,\, \Delta_i \varepsilon_i = 1} \Big[n \cdot \Phi^a_n(X_i, \widehat{A}^0_n, \widehat{\beta}_n)\Big]^{-1} = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s) dN_i(s)}{|\Phi^a_n(s, \widehat{A}^0_n, \widehat{\beta}_n)|},$$

with $\Phi^a_n(s, \widehat{A}^0_n, \widehat{\beta}_n)$ as provided in our technical report. Substituting estimated parameters by the true model parameters we obtain

$$\widetilde{A}^a_n(t) = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s) dN_i(s)}{|\widetilde{\Phi}^a_n(s, A_0, \beta_0)|},$$

and by Doob decomposition of the counting process $N_i(t)$

$$\widetilde{\Phi}^a_n(s, \beta_0, A_0) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s) Y_i(s) e^{\beta_0^T Z_i(s)} \mathcal{G}'\left(\int_0^s e^{\beta_0^T Z_i(u)} dA_0(u)\right) + o_p(1).$$

By the Glivenko-Cantelli theorem $\widetilde{\Phi}^a_n(s, \beta_0, A_0)$ converges uniformly to $E\big[\eta(s, \beta_0, A_0)\big]$ with

$$\eta(s, \beta_0, A_0) \equiv \mathbb{1}(C \geq s) Y(s) e^{\beta_0^T Z(s)} \mathcal{G}'\left(\textstyle\int_0^s Y(u) e^{\beta_0^T Z(u)} dA_0(u)\right)$$

and another application of Doob decomposition implies

$$\widetilde{A}^a_n(t) = A_0(t) + n^{-1}\sum_{i=1}^n \int_0^t \frac{\mathbb{1}(C_i \geq s) dM_i(s)}{E\big[\eta(s, \beta_0, A_0)\big] + o_p(1)} + o_p(1).$$

By the Glivenko-Cantelli theorem $\widetilde{A}_n^a(t) \to A_0(t)$ for $n \to \infty$ uniformly almost surely in t. Further the Glivenko-Cantelli theorem implies that $\Phi_n^a(s, \widehat{A}_n^0, \widehat{\beta}_n)$ converges uniformly to a continuously differentiable function $\Phi_*(s, A_*, \beta_*)$ and the limit is bounded away from zero. $\widehat{A}_n^0(t)$ is absolutely continuous with respect $\widetilde{A}_n^a(t)$ and $\widehat{A}_n^0(t)/\widetilde{A}_n^a(t)$ converges to $\alpha_*(t) = E\big[\eta(s, \beta_0, A_0)\big]/\Phi_*(s, A_*, \beta_*)$. Therefore,

$$\widehat{A}_n^0(t) \to A_*(t) = \int_0^t \frac{E\big[\eta(s, \beta_0, A_0)\big]}{\Phi_*(s, A_*, \beta_*)} dA_0(s).$$

With a Kullback-Leibler argument it is then obtained that the limit $A_*(t)$ is P-almost surely the true baseline hazard $A_0(t)$. For the setting with administrative censoring we define the Kullback-Leibler distance

$$\mathcal{K}_a(\theta_0, \theta_*) = E_{F_{\theta_0}, G}\left[\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta \mathbb{1}(\varepsilon=1)} \left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{1-\Delta} \left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{\Delta \mathbb{1}(\varepsilon \neq 1)}\right\}\right],$$

thereby denoting $\theta_0 = (\beta_0, A_0)$ and $\theta_* = (\beta_*, A_*)$ and with $f_\theta(t) = \alpha_\theta(t) S_\theta(t)$ being the subdensity for the event of interest. As derived in Section B.3 this Kullback-Leibler distance is nonnegative. On the other hand, as $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{A}_n^0(t))$ maximizes the log-likelihood function it is obtained that $\mathcal{K}_a(\theta_0, \theta_*) \leq 0$ and thus $\mathcal{K}_a(\theta_0, \theta_*) = 0$. It is then ascertained in Section B.3 of our technical report that the Kullback-Leibler distance takes the value zero if and only if $\beta_* = \beta_0$ and $A_* = A_0$. From this we conclude that $\widehat{\beta}_n \to \beta_0$ P-almost surely and $\widehat{A}_{n_k}^0 \to A_0$ uniformly P-almost surely.

Competing risks setting with independent right censoring

From maximizing the discretized likelihood function we obtain

$$\widehat{A}_n^0(t) = \sum_{i:\, X_i \leq t,\, \Delta_i \varepsilon_i = 1} \left[n \cdot \Phi_n(X_i, \widehat{A}_n^0, \widehat{\beta}_n)\right]^{-1} = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s) dN_i(s)}{|\Phi_n(s, \widehat{A}_n^0, \widehat{\beta}_n)|},$$

with $\Phi_n(s, \widehat{A}_n^0, \widehat{\beta}_n)$ as provided in our technical report. Substituting estimated parameters by the true model parameters we define

$$\widetilde{A}_n^0(t) = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s)dN_i(s)}{|\widetilde{\Phi}_n(s, A_0, \beta_0)|}.$$

With an application of empirical process theory it is ascertained that $\widetilde{\Phi}_n(s, \beta_0, A_0) = \widetilde{\Phi}_n^a(s, \beta_0, A_0) + o_p(1)$ and thus $\widetilde{A}_n(t) \to A_0(t)$. By the Glivenko-Cantelli theorem $\Phi_n(s, \widehat{\beta}_n, \widehat{A}_n)$ converges uniformly to a continuously differentiable function $\Phi^*(s, A^*, \beta^*)$ with $\Phi^*(s, A^*, \beta^*) > 0$ for $s \in [0, \tau]$ and

$$\widehat{A}_n^0(t) \to A_*(t) = \int_0^t \frac{E\big[\eta(s, \beta_0, A_0)\big]}{\Phi^*\big(s, \beta^*, A^*\big)} dA_0(s).$$

$A^*(t)$ is absolutely continuous with respect to the Lebesgue measure and the Radon–Nikodym derivative takes the form $\alpha^*(t) = E\big[\eta(t, \beta_0, A_0)\big]/\Phi^*\big(t, \beta^*, A^*\big)\alpha_0(t)$.

We define the Kullback-Leibler distance corresponding to the weighted log-likelihood function as

$$\mathcal{K}_{w^*}(\theta_0, \theta_*) = E_{F_{\theta_0},G}\left[\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{1-\Delta}\right.\right.$$
$$\left.\left.\times\left(\frac{\exp\big(-\int_0^\tau \widetilde{w}(t)dA_{\theta_0}(t)\big)}{\exp\big(-\int_0^\tau \widetilde{w}(t)dA_{\theta_*}(t)}\right)^{\Delta\mathbb{1}(\varepsilon\neq 1)}\right\}\right].$$

From the asymptotic equivalence of $\mathcal{K}_a(\theta_0, \theta_*)$ and $\mathcal{K}_{w^*}(\theta_0, \theta_*)$ it can be concluded that $\beta_* = \beta_0$ and $A_* = A_0$.

A.3 Weak convergence

Let $\mathcal{H}$ denote the space of elements $h = (h_1, h_2)$ with $h_1 \in \mathbb{R}^d$ and $h_2 \in \mathcal{D}[0, \tau]$. A norm on $\mathcal{H}$ is then defined by $\|h\|_{\mathcal{H}} = \|h_1\| + \|h_2\|_v$, with $\|\cdot\|$ denoting the Euclidean norm and $\|\cdot\|_v$ denoting the total variation norm. For $p < \infty$ we define $\mathcal{H}_p = \{h \in \mathcal{H} : \|h\|_{\mathcal{H}} \leq p\}$. The parameter space is denoted by $\Theta = \{\theta = (\beta, A_0)$, with $\beta \in \mathbb{R}^d$ and A_0 being a monotone increasing element of $\mathcal{D}[0, \tau]\}$. For $h \in \mathcal{H}_p$ we define $\theta(h) = h_1^T\beta + \int_0^\tau h_2(u)dA_0(u)$, so that $\Theta \subset \ell^\infty(\mathcal{H}_p)$. One-dimensional submodels of the form $t \to \theta_t = \theta + t\big(h_1, \int_0^{(\cdot)} h_2(u)dA_0(u)\big)$ are considered with $h \in \mathcal{H}_p$ to define the empirical score operator

$$\Psi_n(\theta, h, w) = \frac{\partial}{\partial t}\,\ell_n(\theta_t, h, w)\bigg|_{t=0} = \mathbb{P}_n\psi(\theta, h, w) = \mathbb{P}_n\big(\psi_1(\theta, h, w) + \psi_2(\theta, h, w)\big),$$

for a measurable function $\psi(\theta, h, w)$, where $\Psi^w_{n1}(\theta, h, w) = \mathbb{P}_n\psi_1(\theta, h, w)$ is the component related to the derivative with regard to β and $\Psi^w_{n2}(\theta, h, w) = \mathbb{P}_n\psi_2(\theta, h, w)$ is the component related to the derivative along the submodel for A_0. The limiting version Ψ is defined by replacing the empirical measure $\mathbb{P}_n$ by the probability measure $\mathcal{P}$.

Weak convergence is ascertained by a new lemma for weighted $\mathcal{Z}-$estimators, that is based on Theorem 3.3.1. of van der Vaart and Wellner (1996):

Lemma 1. *Let the parameter set Θ be a subset of a Banach space. Let $\widetilde{w}(t)$ be a bounded deterministic weight function and let $\widehat{w}_n(t)$ be a sequence of bounded random weight functions with values in $\mathbb{R}_+$. Let Ψ_n and Ψ be a linear random map and a linear deterministic map, respectively from $\Theta \times \mathbb{R}_+$ into a Banach space such that*

$$a) \quad \sqrt{n}(\Psi_n - \Psi)(\widehat{\theta}_n, \widehat{w}_n) - \sqrt{n}(\Psi_n - \Psi)(\theta_0, \widehat{w}_n) = o_p^*\big(1 + \sqrt{n}\|\widehat{\theta}_n - \theta_0\|\big),$$

such that $\sqrt{n}(\Psi_n - \Psi)(\theta_0, \widetilde{w})$ converges to a tight limit $\mathcal{Z}_1$ and $\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widehat{w}_n) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big)$ converges to a tight limit $\mathcal{Z}_2$, and the sequences jointly converge to $(\mathcal{Z}_1, \mathcal{Z}_2)$.

Let $(\theta, w) \to \Psi(\theta, w)$ be Fréchet-differentiable at $(\theta_0, \widetilde{w})$ with a continuously invertible derivative $\dot{\Psi}^{\widetilde{w}}_{\theta_0}$. If $\Psi(\theta_0, \widetilde{w}) = 0$ and $\widehat{\theta}_n$ satisfies $\Psi_n(\widehat{\theta}_n, \widehat{w}_n) = o_p^*(n^{-1/2})$, if $\Psi_n(\widehat{\theta}_n, \widetilde{w}) = \Psi_n(\widehat{\theta}_n, \widehat{w}_n) + o_p(1)$ and if $\Psi_n(\widehat{\theta}_n, \widetilde{w})$ converges in outer probability to $\Psi(\theta_0, \widetilde{w})$, then

$$\sqrt{n}\,\dot{\Psi}^{\widetilde{w}}_{\theta_0}(\widehat{\theta}_n - \theta_0) = -\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widetilde{w}) + (\Psi_n - \Psi)(\theta_0, \widehat{w}) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big) + o_p^*(1) \tag{1}$$

and $\sqrt{n}(\widehat{\theta}_n - \theta_0) \rightsquigarrow -\big[\dot{\Psi}^{\widetilde{w}}_{\theta_0}\big]^{-1}(\mathcal{Z}_1 + \mathcal{Z}_2)$, with $\rightsquigarrow$ denoting weak convergence.

As our model is based on iid observations, for condition $a)$ in the above Lemma 1 it is sufficient to verify the two conditions of Lemma 5 in our technical report. By the Donsker theorem $\sqrt{n}(\Psi_n - \Psi)(\theta_0, \widetilde{w})$ converges in distribution to the tight random element $\mathcal{Z}_1$. Also by Donsker theorem $\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widehat{w}_n) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big)$ converges in distribution to a tight random element $\mathcal{Z}_2$. Joint convergence follows from the asymptotic linearity of the two components marginally, combined with the fact that the composition of two Donsker classes is also Donsker. Per definition $\Psi_n(\widehat{\theta}_n, \widehat{w}_n) = 0$. As argued in Parner (1998) from the Kullback-Leibler information being positive and by interchanging expectation and differentiation we obtain $\Psi(\theta_0, \widetilde{w}) = 0$. Arguments for the continuous invertibility of $\dot{\Psi}_0$ and $\dot{\Psi}_0^a$ under model conditions $M6), M7)$ are provided in our technical report.

From Lemma 1 we obtain weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and the covariances for the limiting process $-\dot{\Psi}_0^{-1}(\mathcal{Z}_1 + \mathcal{Z}_2)$, with $\mathcal{Z}_1(h) = \lim_{n\to\infty} \mathbb{P}_n \, \psi(\widehat{\beta}_n, \widehat{A}_n, h, \widetilde{w})$ and $\mathcal{Z}_2(h) = \lim_{n\to\infty} \mathbb{P}_n \big[(\psi(\widehat{\beta}_n, \widehat{A}_n, h, \widehat{w}_n) - \psi(\widehat{\beta}_n, \widehat{A}_n, h, \widetilde{w})\big]$, are given by

$$\mathcal{P}\left\{\Big[(\mathcal{Z}_1 + \mathcal{Z}_2)\big(\dot{\Psi}_0^{-1}(g)\big)\Big]\,\Big[(\mathcal{Z}_1 + \mathcal{Z}_2)\big(\dot{\Psi}_0^{-1}(h)\big)\Big]\right\}$$

for $g, h \in \mathcal{H}$ (Kosorok, 2008, page 302).

A.4 Variance estimation

The middle part of the sandwich variance estimator, as proposed in section **??**, is obtained from the iid decomposition of the score with $\eta_{i,n} = \big(\eta_i^1, \ldots, \eta_i^{d+k(n)}\big)$ and $\psi_{i,n} = \big(\psi_i^1, \ldots, \psi_i^{d+k(n)}\big)$ defined as

$$\eta_{i,n}^{\ell} = \left[Z_{i\ell}(X_i) + \sum_{\substack{j: X_j \le X_i \\ \Delta_j \varepsilon_j = 1}} Z_{i\ell}(X_j) e^{\beta^T Z_i(X_j)} A_n\{X_j\} \frac{\mathcal{G}''\left(\int_0^{X_i} e^{\beta_0^T Z_i(u)} dA_n(u)\right)}{\mathcal{G}'\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u)\right)} \right] \mathbb{1}(\Delta_i \varepsilon_i = 1)$$

$$
\begin{aligned}
&- \sum_{\substack{j:X_j \le X_i \wedge \tau \\ \Delta_j \varepsilon_j = 1}} Z_{i\ell}(X_j) e^{\beta^T Z_i(X_j)} A_n\{X_j\}\, \mathcal{G}'\left(\int_0^{X_i \wedge \tau} e^{\beta^T Z_i(u)} dA_n(u) \right) \\
&- \left[\int_{X_i}^{\tau} \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)} Z_{i\ell}(t) \mathcal{G}'\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u) \right) dA_n(t) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2) \\
&- \left[\int_{X_i}^{\tau} \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)} \left(\int_0^t Z_{i\ell}(u) e^{\beta^T Z_i(u)} dA_n(u) \right) \mathcal{G}''\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u) \right) dA_n(t) \right] \\
&\qquad \times \mathbb{1}(\Delta_i \varepsilon_i = 2)
\end{aligned}
$$

for $\ell \in \{1, \ldots, d\}$ and

$$
\begin{aligned}
\eta_{i,n}^{\ell} = &\Bigg[\mathbb{1}(X_\ell = X_i) \big[A_n\{X_\ell\}\big]^{-1} + \mathbb{1}\big(X_\ell \le (X_i \wedge \tau)\big) e^{\beta^T Z_i(X_\ell)} \\
&\qquad \times \mathcal{G}''\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u) \right) \Big/ \mathcal{G}'\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u) \right) \Bigg] \mathbb{1}(\Delta_i \varepsilon_i = 1) \\
&- \mathbb{1}\big(X_\ell \le (X_i \wedge \tau)\big) e^{\beta^T Z_i(X_\ell)} \mathcal{G}'\left(\int_0^{X_\ell} e^{\beta^T Z_i(u)} dA_n(u) \right) \\
&- \mathbb{1}(X_i \le X_\ell) \left[\widetilde{w}_i^*(X_\ell) e^{\beta^T Z_i(X_\ell)} \mathcal{G}'\left(\int_0^{X_\ell} e^{\beta^T Z_i(u)} dA_n(u) \right) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2) \\
&- \left[e^{\beta^T Z_i(X_\ell)} \int_{X_i \wedge X_\ell}^{\tau} \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)}\, \mathcal{G}''\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u) \right) dA_n(t) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2)
\end{aligned}
$$

for $\ell \in \{d+1, \ldots, d+k(n)\}$. To calculate $\psi_{i,n}^{\ell}$ we apply

$$
\frac{\widehat{G}_c(t)}{\widehat{G}_c(X_j)} - \frac{G_c(t)}{G_c(X_j)} = -\frac{G_c(t)}{G_c(X_j)} \sum_{i=1}^{n} \int_{X_j}^{t} \frac{1}{\sum_{k=1}^{n} \mathbb{1}(X_k \ge u)} \times dM_i^c(u) + o_p(n^{-1/2}),
$$

where $M_i^c(t) = \mathbb{1}(X_i \le t,\, \Delta_i = 0) - \int_0^t \mathbb{1}(X_i \ge u) dA^c(u)$ is the martingale associated with the censoring process and $A^c(t)$ is the cumulative hazard of the censoring distribution. From this we obtain the representation

$$
\psi_{i,n}^{\ell} = \int_0^{\infty} q_n^{\ell}(u) \{\pi_n(u)\}^{-1} dM_i^c(u), \text{ with}
$$

$$
\begin{aligned}
q_n^\ell(u) &= \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j=2}\left[\int_0^\tau \widetilde{w}_j^*(t)\mathbb{1}(X_j \le u \le t)e^{\beta^T Z_j(t)} Z_{j\ell}(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z_j(s)}dA_n(s)\right)dA_n(t)\right] \\
&\quad + \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j=2}\left[\int_0^\tau \widetilde{w}_j^*(t)\mathbb{1}(X_j \le u \le t)e^{\beta^T Z_j(t)}\left(\int_0^t Z_{j\ell}(s)e^{\beta^T Z_j(s)}dA_n(s)\right)\right. \\
&\qquad \left.\times\mathcal{G}''\left(\int_0^t e^{\beta^T Z_j(s)}dA_n(s)\right)dA_n(t)\right] \text{ for } \ell\in\{1,\dots,d\},
\end{aligned}
$$

$$
q_n^\ell(u) = \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j=2}\left[\widetilde{w}_j^*(X_\ell)\mathbb{1}(X_j \le u \le X_\ell)e^{\beta^T Z_j(X_\ell)}\,\mathcal{G}'\big(\int_0^{X_\ell} e^{\beta^T Z_j(s)}dA_n(s)\big)\right]
$$

$$
+\frac{1}{n}\sum_{j:\Delta_j\varepsilon_j=2}\left[e^{\beta^T Z_j(X_\ell)}\int_0^\tau \mathbb{1}(X_\ell\le t)\mathbb{1}(X_j \le u \le t)\widetilde{w}_j^*(t)e^{\beta^T Z_j(t)}\mathcal{G}''\big(\int_0^t e^{\beta^T Z_j(s)}dA_n(s)\big)dA_n(t)\right]
$$

for $\ell \in \{d+1,\dots,d+k(n)\}$ and $\pi_n(u) = n^{-1}\sum_{j=1}^n \mathbb{1}(X_j \ge u)$.