Supplementary Materials for
"Semiparametric Accelerated Intensity Models for Correlated Recurrent and Terminal Events"

By Sangbum Choi, Xuelin Huang, Hyunsu Ju, and Jing Ning

Korea University, The University of Texas MD Anderson Cancer Center, and The University of Texas Medical Branch

A Asymptotic Results

This section outlines the asymptotic properties for the proposed estimators. The proofs of consistency and asymptotic normality of the NPMLEs essentially follow the steps in Zeng and Lin (2007, 2009). With abuse of notation, we denote the entire set of parameters by $\vartheta = \{\alpha, \beta, \theta, r(\cdot), \lambda(\cdot)\}$ with $r(\cdot)$ and $\lambda(\cdot)$ being the unknown baseline intensity functions. Likewise, let $\hat{\vartheta}_n = \{\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n, \hat{r}_n(\cdot), \hat{\lambda}_n(\cdot)\}$ and $\vartheta_0 = \{\alpha_0, \beta_0, \theta_0, r_0(\cdot), \lambda_0(\cdot)\}$ denote the NPMLE and the true value of ϑ.

We impose the following regularity conditions:

(C1) The parameter value $(\alpha_0, \beta_0, \theta_0)$ belongs to the interior of a known compact set Θ in $\mathcal{R}^d$. The covariate matrix $\mathbf{z} \in \mathcal{R}^p$ is bounded and has a full rank.

(C2) The true rate function of $r_0(t)$ and the true hazard function of $\lambda_0(t)$ are both positive, at least twice continuously differentiable and have bounded variations over $t \in [0, \tau]$.

(C3) With probability one, there exists a $\kappa_0 > 0$ such that $P(C \geq \tau|\mathbf{z}) > \kappa_0$. With probability one, $E[N^{R*}(\tau)] < \infty$ and $E[N^{D*}(\tau)] < \infty$.

(C4) The kernel function $K(\cdot)$ is thrice continuously differentiable and the rth derivative $K^{(r)}(\cdot), r = 0, 1, 2, 3$ has bounded variation in $(-\infty, \infty)$.

(C5) The information matrix $\mathcal{I}_0$ is finite and positive definite.

Theorem A.1. *Suppose that conditions (C1)–(C4) hold and that as* $n \to \infty$, $na_n^2 \to \infty$, $na_n^4 \to 0$, $nb_n^2 \to \infty$, *and* $nb_n^4 \to 0$. *Then,* $\hat{\vartheta}$ *is strongly consistent for* ϑ_0 *as* $n \to \infty$.

Theorem A.2. *Suppose that conditions (C1)–(C5) hold and that as $n \to \infty$, $na_n^2 \to \infty$, $na_n^4 \to 0$, $nb_n^2 \to \infty$, and $nb_n^4 \to 0$. As $n \to \infty$, $\sqrt{n}(\hat{\alpha} - \alpha_0)$, $\sqrt{n}(\hat{\beta} - \beta_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0)$ converge in distribution to respective mean-0 normal random vectors.*

A.1 Proof of Theorem 1

Define the observed log-likelihood function by

$$l_n(\vartheta) = \prod_{i=1}^{n} \log \int_{\nu_i} \left[\prod_t \{\nu_i e^{\alpha' \mathbf{z}_i} r(te^{\alpha' \mathbf{z}_i})\}^{dN_i^R(t)} \exp\{-\nu_i R(x_i e^{\alpha' \mathbf{z}_i})\} \right] \\ \times \left[\{\nu_i e^{\beta' \mathbf{z}_i} \lambda(x_i e^{\beta' \mathbf{z}_i})\}^{\delta_i} \exp\{-\nu_i \Lambda(x_i e^{\beta' \mathbf{z}_i})\} \right] f_\theta(\nu_i) d\nu_i,$$

and let $l_0(\vartheta)$ denote the expected version of $l_n(\vartheta)$. To establish the consistency of the NPMLE, it suffices to show that $\sup_{\vartheta \in \Theta} |l_n(\hat{\vartheta}_n) - l_0(\vartheta)| \xrightarrow{a.s.} 0$ as $n \to \infty$ and that the maximization is uniquely determined at $\vartheta = \vartheta_0$. Toward this end, let us define smoothed estimators for $R(t)$ and $\Lambda(t)$, respectively, as

$$\tilde{R}_n(t) = \int_{-\infty}^{t} \left[\frac{(na_n)^{-1} \sum_{i=1}^{n} \int_0^\infty K\{(u-s)/a_n\} dN_i^R(ue^{-\alpha_0' \mathbf{z}_i})}{n^{-1} \sum_{i=1}^{n} \tilde{\nu}_i \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha_0)} - s)/a_n} K(u) du} \right] ds,$$

and

$$\tilde{\Lambda}_n(t) = \int_{-\infty}^{t} \left[\frac{(nb_n)^{-1} \sum_{i=1}^{n} \int_0^\infty K\{(u-s)/b_n\} dN_i^D(ue^{-\beta_0' \mathbf{z}_i})}{n^{-1} \sum_{i=1}^{n} \tilde{\nu}_i \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\beta_0)} - s)/b_n} K(u) du} \right] ds,$$

where $\tilde{\nu}_i = E(\nu_i | \mathbf{O})$ for $i = 1, ..., n$. By Lemma 2.4 of Schuster (1969) and Theorem 2.4.3 of van der Vaart and Wellner (1996), it can be shown that as $n \to \infty$,

$$\sup_{t \in [0,\tau]} \left| \frac{1}{na_n} \sum_{i=1}^{n} \int_0^\infty K\{(u-t)/a_n\} dN_i^R(ue^{-\alpha_0' \mathbf{z}_i}) - \frac{E\{dN_i^R(te^{-\alpha_0' \mathbf{z}_i})\}}{dt} \right| \xrightarrow{a.s.} 0,$$

$$\sup_{t \in [0,\tau]} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{\nu}_i \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha_0)} - t)/a_n} K(u) du - E\{\tilde{\nu}_i I(e^{\tilde{\varepsilon}_i(\alpha_0)} \geq t)\} \right| \xrightarrow{a.s.} 0.$$

In addition, we have $E\{dN_i^R(te^{-\alpha_0' \mathbf{z}_i})\}/dt = r_0(t) E\{\tilde{\nu}_i I(e^{\tilde{\varepsilon}_i(\alpha_0)} \geq t)\}$. Therefore,

$$\sup_{t \in [0,\tau]} \left| \frac{(na_n)^{-1} \sum_{i=1}^{n} \int_0^\infty K\{(u-t)/a_n\} dN_i^R(ue^{-\alpha_0' \mathbf{z}_i})}{n^{-1} \sum_{i=1}^{n} \tilde{\nu}_i \int_{-\infty}^{(e^{\tilde{\varepsilon}_i(\alpha_0)} - t)/b_n} K(u) du} - r_0(t) \right| \xrightarrow{a.s.} 0,$$

which implies that $\tilde{R}_n(t) \to R_0(t)$ almost surely in $t \in [0, \tau]$. The pointwise consistency can be strengthened to the uniform consistency due to the monotonicity and boundedness of $\tilde{R}_n(t)$ and $R_0(t)$, i.e., $\sup_{t\in[0,\tau]} |\tilde{R}_n(t) - R_0(t)| \to 0$ a.s. Similarly, one can show the uniform consistency of $\tilde{\Lambda}_n(t)$ to $\Lambda_0(t)$ on $[0, \tau]$.

It follows from Helly's selection theorem that there exists a convergent subsequence such that $\hat{\vartheta}_{n_k} \to \vartheta_*$. Clearly, $n_k{}^{-1}\{l_{n_k}(\hat{\vartheta}_{n_k}) - l_0(\tilde{\vartheta}_0)\} \geq 0$, where $\tilde{\vartheta}_0 = (\alpha_0, \beta_0, \theta_0, \tilde{R}_{n_k}, \tilde{\Lambda}_{n_k})$. It can be checked that the Kullback-Leibler information between the density indexed by ϑ_* and the true density is negative. These implies that

$$\begin{aligned}
&\log \int_{\nu_i} \left[\prod_t \{\nu_i e^{\alpha_*' \mathbf{z}_i} r_*(t e^{\alpha_*' \mathbf{z}_i})\}^{dN_i^R(t)} \exp\{-\nu_i r_*(x_i e^{\alpha_*' \mathbf{z}_i})\} \right] \\
&\qquad \times \left[\{\nu_i e^{\beta_*' \mathbf{z}_i} \lambda_*(x_i e^{\beta_*' \mathbf{z}_i})\}^{\delta_i} \exp\{-\nu_i \Lambda_*(x_i e^{\beta_*' \mathbf{z}_i})\} \right] f_{\theta_*}(\nu_i) d\nu_i \\
=& \log \int_{\nu_i} \left[\prod_t \{\nu_i e^{\alpha_0' \mathbf{z}_i} r_0(t e^{\alpha_0' \mathbf{z}_i})\}^{dN_i^R(t)} \exp\{-\nu_i r_0(x_i e^{\alpha_0' \mathbf{z}_i})\} \right] \\
&\qquad \times \left[\{\nu_i e^{\beta_0' \mathbf{z}_i} \lambda_0(x_i e^{\beta_0' \mathbf{z}_i})\}^{\delta_i} \exp\{-\nu_i \Lambda_0(x_i e^{\beta_0' \mathbf{z}_i})\} \right] f_{\theta_0}(\nu_i) d\nu_i,
\end{aligned}$$

Some manipulations with the above equations by following the lines of Zeng and Lin (2009) show that $\alpha_* = \alpha_0$, $\beta_* = \beta_0$, $\theta_* = \theta_0$, and $R_*(t) = R_0(t)$ and $\Lambda_*(t) = \Lambda_0(t)$ for $t \in [0, \tau]$ with probability one. Therefore, $\hat{\vartheta}_{n_k}$ should converge to ϑ_0. Helly's theorem completes the proof for consistency of $\hat{\vartheta}_n$ to ϑ_0. Furthermore, the point-wise consistency can be strengthened to uniform convergence on $[0, \tau]$ by applying the Glivenko-Cantelli theorem.

A.2 Proof of Theorem 2

Let $\|.\|_{l^\infty[0,\tau]}$ denote the supremum norm in $[0, \tau]$, $BV[0, \tau]$ the space of bounded variation functions on $[0, \tau]$, $\|\mu\|_{BV[0,\tau]}$ the total variation of $\mu(t)$ in $[0, \tau]$, and $\mathcal{H} = \{\mu(t) : \|\mu\|_{BV[0,\tau]} \leq 1\}$. Let $\mathcal{P}_n$ denote the empirical measure and $\mathcal{P}_0$ the probability measure. For ease of presentation, we further define $\phi = (\alpha', \beta', \theta)'$ and $A(\cdot) = \{R(\cdot), \Lambda(\cdot)\}$, which, respectively, represent finite- and infinite-dimensional parameters. Likewise, we let $\hat{\vartheta}_n = (\hat{\phi}_n, \hat{A}_n)$ and $\vartheta_0 = (\phi_0, A_0)$ denote the NPMLE and the true parameter. Note that $\hat{A}(t)$ can be regarded as a bounded linear functional in $l^\infty(\mathcal{H})$, and $\{\hat{\phi}_n - \phi_0, \hat{A}_n(\cdot) - A_0(\cdot)\}$ is a random element in $\mathcal{R}^d \times l^\infty(\mathcal{H})$, where $l^\infty(\mathcal{H})$ is the space of bounded real-valued functions on $\mathcal{H}$ under the supremum norm. We claim that $n^{1/2}\{\hat{\phi}_n - \phi_0, \hat{A}_n(\cdot) - A_0(\cdot)\}$ converges weakly to a zero-mean Gaussian process in the metric space $\mathcal{R}^d \times l^\infty(\mathcal{H})$. This can be accomplished by checking four conditions in Theorem 3.3.1 of van der Vaart and Wellner (1996).

Let $h = (h_1, h_2)$, where h_1 is a vector of the same dimension with ϕ, say d, and h_2 is a bounded function, i.e., $h = (h_1, h_2) \in \mathcal{R}^d \times BV[0,\tau]$. Consider $\phi_s = \phi + sh_1$ and $A_s(t) = \int_0^t \{1 + sh_2(u)\} dA(u)$. Let $U_n(\phi, A)(h_1, h_2) = dl_n(\phi_s, A_s)/ds|_{s=0} = U_{1n}(\phi, A)(h_1) + U_{2n}(\phi, A)(h_2)$, where U_{1n} and U_{2n} are the directional derivatives with respect to ϕ and A, respectively. Also, define the limit of $U_n(\phi, A)(h_1, h_2)$ by $U_0(\phi, A)(h_1, h_2) = U_{10}(h_1) + U_{20}(h_2)$, i.e., $U_0(\phi, A)(h) = \mathcal{P}_0\{U_n(\phi, A)(h)\}$. Clearly, $U_n(\hat{\phi}_n, \hat{A}_n)(h_1, h_2) = U_0(\phi_0, A_0)(h_1, h_2) = 0$. It can be checked that $U_0(\phi, A)$ is Fréchet differentiable and let $\dot{U}(\phi_0, A_0)(\phi - \phi_0, A - A_0)(h_1, h_2)$ denote the Fréchet derivative of $U_0(\phi, A)$ at (ϕ_0, A_0).

Note that score operators U_{1n} and U_{2n} can be written as $U_{1n}(\phi_0, A_0)(h_1) = \mathcal{P}_n\{h_1' U_{1n}^o(\phi_0, A_0)\} = n^{-1}\sum_{i=1}^n h_1' U_{1i}^o(\phi_0, A_0)$, and $U_{2n}(\phi_0, A_0)(h_2) = \mathcal{P}_n\{U_{2n}^o(\phi_0, A_0)(h_2)\} = n^{-1}\sum_{i=1}^n U_{2i}^o(\phi_0, A_0)$ for some U_{1i}^o and U_{2i}^o. Consider

$$\mathcal{A}_1(\phi_0, A_0) = \{h_1' U_{1i}^o(\phi_0, A_0); \|h_1\| < \infty\},$$
$$\mathcal{A}_2(\phi_0, A_0) = \{U_{2i}^o(\phi_0, A_0)(h_2); h_2 \in BV[0,\tau]\}.$$

Under conditions (C1)–(C6), it can be checked that U_{1i}^o is sufficiently smooth and bounded functions, thus $\mathcal{A}_1$ is a Donsker class. Similarly, $\mathcal{A}_2$ is also Donsker. Therefore, we see that $n^{1/2}\{U_n(\phi_0, A_0)(h) - U_0(\phi_0, A_0)(h)\}$ converges weakly to a Gaussian process $\mathcal{G}$ in the metric space $\mathcal{R}^d \times l^\infty(\mathcal{H})$. Furthermore, $\|\phi - \phi_0\| + \sup_{t\in[0,\tau]} |A(t) - A_0(t)| = o_p(1)$, we can show that $\mathcal{A}_1(\phi, A)$ and $\mathcal{A}_2(\phi, A)$ are both Donsker and that $n^{1/2}\{(U_n - U_0)(\hat{\phi}_n, \hat{A}_n)(h) - (U_n - U_0)(\phi_0, A_0)(h)\} = o_p(1)$.

Finally, we need to show that the information operator $\dot{U}(\phi_0, A_0)(h)$ is continuously convertible on its range. To this end, it is key to show that $\dot{U}_0(\phi_0, A_0)(h) = 0$ almost surely implies $h \equiv 0$. By the definition of $\dot{U}_0(\phi, A)(h)$, $\dot{U}_0(\phi_0, A_0)(h) = 0$ implies $E\{h_1' U_{10}^o(\phi_0, A_0) + U_{20}^o(\phi_0, A_0)(h_2)\}^2 = 0$. Under conditions (C1)–(C6), we have that $h_1' U_{10}^o(\phi_0, A_0) + U_{20}^o(\phi_0, A_0)(h_2) = 0$ almost surely. By using similar techniques used for the consistency of the estimators, it is easily seen that $h_1 = 0$ and $h_2(\cdot) = 0$ everywhere, thus $h \equiv 0$. This implies that the information operator is a one-to-one map and thus $\dot{U}(\phi_0, A_0)(h)$ or equivalently $-\mathcal{I}_0$ is invertible. Therefore, $n^{1/2}\{(\hat{\phi}_n, \hat{A}_n) - (\phi_0, A_0)\}$ converges weakly to the Gaussian process $\mathcal{I}_0^{-1}\mathcal{G}$. The estimation and consistency of the variance-covariance function can be shown along the lines of Parner (1998) and Zeng and Lin (2007). In addition, we can show that the asymptotic variance of $n^{1/2}\{(\hat{\phi}_n, \hat{A}_n) - (\phi_0, A_0)\}$ achieves the semiparametric efficiency bound $\mathcal{I}_0^{-1}$, using the theory of Bickel et al. (1993).

A.3 Comparison of Frailty Distributions

One reviewer suggested to compare two underlying frailty distributions. Corresponding to the results in simulation #2 (Table 3.2), we displayed two distributions in the following four cases: (a) gamma distribution with $\theta = 0.8$ is a true model and log-normal (LN) distribution is a working

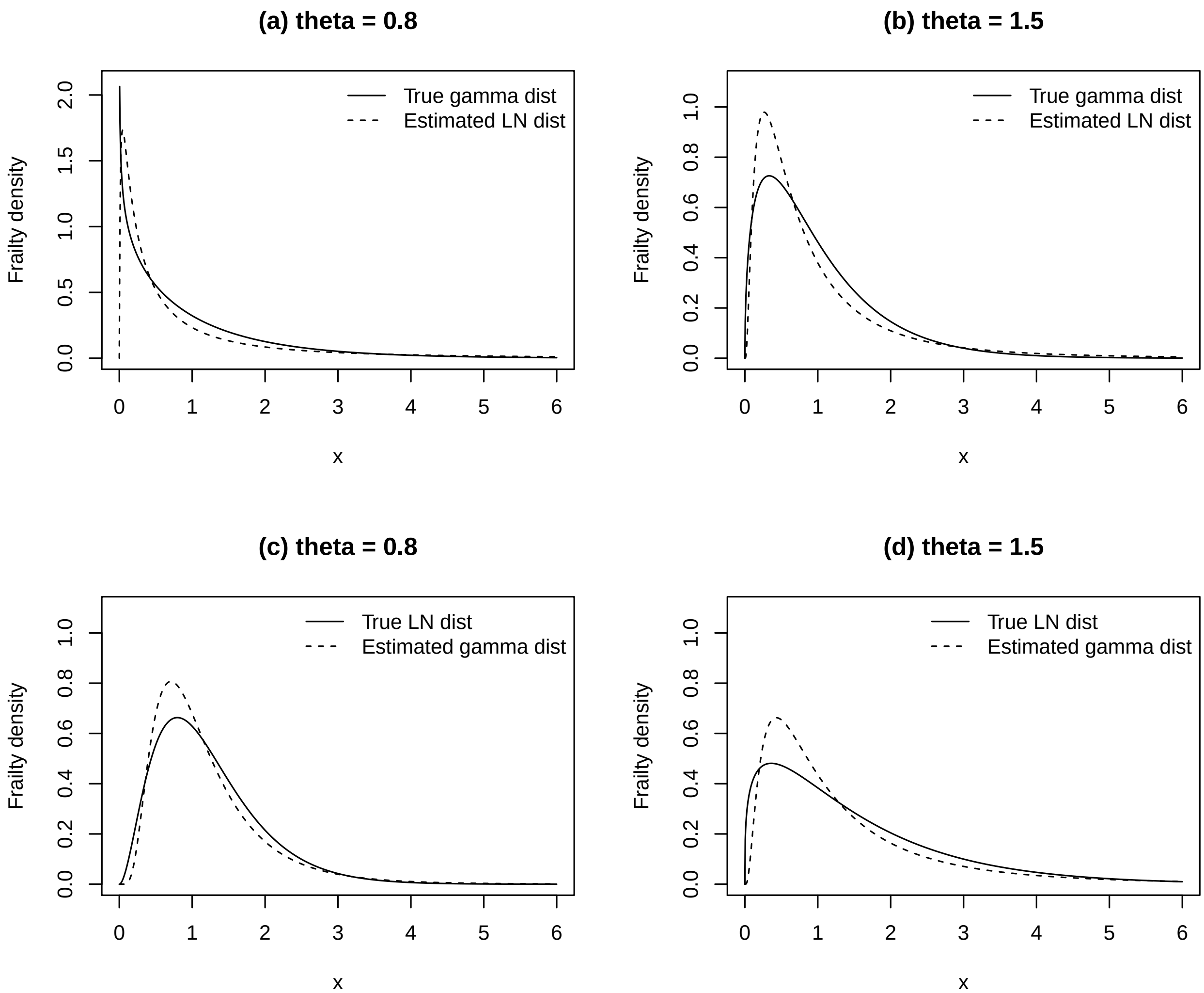

Figure 1: Comparison of gamma and log-normal (LN) frailty distributions under model misspecification.

model; (b) gamma distribution with $\theta = 1.5$ is a true model and LN distribution is a working model; (c) LN distribution with $\theta = \exp(\sigma^2) - 1 = 0.8$ is a true model and gamma distribution is a working model; and (d) LN distribution with $\theta = \exp(\sigma^2) - 1 = 1.5$ is a true model and gamma distribution is a working model. Although there exists some discrepancy between underlying and estimated distributions, two density functions appear to be fairly similar.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

van der Vaart, A. W. and Wellner, J. A. (1996). *Empirical Processes and Weak Convergence.* Springer, New York.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric models with censored data (with discussion). *J. Royal Statist. Soc. B* **69**, 507–564.

Zeng, D. and Lin, D. Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65**, 746–752.