Cure Model with Current Status Data

Shuangge Ma

Yale University

Appendix

Proof of Lemma 2.

The parameter set of (α, β) is compact by assumption A2. The parameter set for Λ is compact relative to the weak topology. Theorem 5.14 of van der Vaart (1998) shows that the distance between $(\hat{\alpha}, \hat{\beta}, \hat{\Lambda})$ and the set of maximizers of the Kullback-Leibler distance converges to zero. Lemma 2 follows from the identifiability assumption A4.

Proof of Lemma 3.

Definition (Bracketing number). Let $(\mathbb{F}, ||\cdot||)$ be a subset of a normed space of real function f on some set. Given two functions f_1 and f_2, the bracket $[f_1, f_2]$ is the set of all functions f with $f_1 \leq f \leq f_2$. An ϵ bracket is a bracket $[f_1, f_2]$ with $||f_1 - f_2|| \leq \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathbb{F}, ||\cdot||)$ is the minimum number of ϵ brackets needed to cover $\mathbb{F}$. The entropy with bracketing is the logarithm of the bracketing number.

Lemma 25.84 of van der Vaart (1998) shows that there exists a constant K_3 such that for every $\epsilon > 0$, $\log N_{[]}(\epsilon, \{\Lambda\}, L_2(P)) \leq K_3(\frac{1}{\epsilon})$, if assumption A3 is satisfied. Since the log-likelihood function l_1 is Hellinger differentiable and considering the compactness assumptions A2 and A3, we have $\log N_{[]}(\epsilon, \{l_1(\alpha, \beta, \Lambda)\}, L_2(P)) \leq K_4(\frac{1}{\epsilon})$, for a constant K_4.

Apply Theorem 3.2.5 of van der Vaart and Wellner (1996). Considering the consistency result in Lemma 2, we have

$$\mathrm{P}^* \sup_{d((\alpha,\beta,\Lambda),(\alpha_0,\beta_0,\Lambda_0))<\eta} |\sqrt{n}(\mathrm{P}_n - \mathrm{P})(l_1(\alpha, \beta, \Lambda) - l_1(\alpha_0, \beta_0, \Lambda_0))|$$
$$= O_p(1)\eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}}K_5\right), \tag{7.1}$$

for a constant K_5, where P^* is the outer expectation. So conditions of Theorem 3.2.1 of van der Vaart and Wellner (1996) are satisfied. Equation(7.1) and assumption A4 imply

$$d((\hat{\alpha}, \hat{\beta}, \hat{\Lambda}), (\alpha_0, \beta_0, \Lambda_0)) = O_p(n^{-1/3}).$$

Proof of Lemma 5.

Lemma 5 can be proved using Theorem 3.4 of Huang (1996). A slightly different version is presented as Theorem 1 in Ma and Kosorok (2005b). We refer to those papers for details.

Since $\mathrm{P}_n l_1(\alpha, \beta, \Lambda)$ is maximized at $(\hat{\alpha}, \hat{\beta}, \hat{\Lambda})$, we have

$$\mathrm{P}_n \dot{l}_{1\alpha}(\hat{\alpha}, \hat{\beta}, \hat{\Lambda}) = 0,\ \mathrm{P}_n \dot{l}_{1\beta}(\hat{\alpha}, \hat{\beta}, \hat{\Lambda}) = 0,\ \text{and } \mathrm{P}_n \tilde{l}_{1\Lambda}(\hat{\alpha}, \hat{\beta}, \hat{\Lambda})a = 0,$$

for any $a \in \mathbb{A}$. We also have

1. (Consistency and convergence rate). $||\hat{\alpha} - \alpha_0|| = O_p(n^{-1/3})$; $||\hat{\beta} - \beta_0|| = O_p(n^{-1/3})$ and $||\hat{\Lambda} - \Lambda_0||_2 = O_p(n^{-1/3})$ from Lemma 3.

2. (Positive information) The Fisher Information matrix is positive definite and component wise bounded from assumption A5.

3. (Stochastic equicontinuity). For any $\delta_n \to 0$ and constant $K_6 > 0$, within the neighborhood $\{||\alpha - \alpha_0|| < \delta_n, ||\beta - \beta_0|| < \delta_n, ||\Lambda - \Lambda_0||_2 < K_6 n^{-1/3}\}$,

$$\begin{aligned}
&\sup \sqrt{n}|(\mathrm{P}_n - \mathrm{P})(\dot{l}_{1\alpha}(\alpha,\beta,\Lambda) - \dot{l}_{1\alpha}(\alpha_0,\beta_0,\Lambda_0))| = o_p(1),\\
&\sup \sqrt{n}|(\mathrm{P}_n - \mathrm{P})(\dot{l}_{1\beta}(\alpha,\beta,\Lambda) - \dot{l}_{1\beta}(\alpha_0,\beta_0,\Lambda_0))| = o_p(1),\\
&\sup \sqrt{n}\left|(\mathrm{P}_n - \mathrm{P})\left(\tilde{l}_{1\Lambda}(\alpha,\beta,\Lambda)\frac{\mathrm{P}(\dot{l}_{1\alpha}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)}\right.\right.\\
&\qquad\qquad\left.\left.-\tilde{l}_{1\Lambda}(\alpha_0,\beta_0,\Lambda_0)\frac{\mathrm{P}(\dot{l}_{1\alpha}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)}\right)\right| = o_p(1),\\
&\sup \sqrt{n}\left|(\mathrm{P}_n - \mathrm{P})\left(\tilde{l}_{1\Lambda}(\alpha,\beta,\Lambda)\frac{\mathrm{P}(\dot{l}_{1\beta}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)}\right.\right.\\
&\qquad\qquad\left.\left.-\tilde{l}_{1\Lambda}(\alpha_0,\beta_0,\Lambda_0)\frac{\mathrm{P}(\dot{l}_{1\beta}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)}\right)\right| = o_p(1).
\end{aligned}$$

The above equations can be proved by applying Theorem 3.2.5 of van der Vaart and Wellner (1996) and the entropy result.

4. (Smoothness of the model). For (α,β,Λ) within the neighborhood $\{||\alpha - \alpha_0|| < \delta_n, ||\beta - \beta_0|| < \delta_n, ||\Lambda - \Lambda_0||_2 < K_6 n^{-1/3}\}$, the expectations of $\dot{l}_{1\alpha}, \dot{l}_{1\beta}$ and $\tilde{l}_{1\Lambda}$ are Hellinger differentiable.

Conditions in Theorem 3.4 of Huang (1996) are satisfied and hence Lemma 5 follows.

Proof of Lemma 7.

van de Geer (2000) shows that for the class

$$\tilde{\mathbb{H}} = \{h : [0,1] \to [0,1] \int (h^{(s_0)}(x))^2 dx < 1\},$$

we have $\log N_{[]}(\epsilon, \tilde{\mathbb{H}}, L_2(P)) \le K_7 \epsilon^{-1/s_0}$, for a fixed constant K_7, $s_0 \ge 1$ and all ϵ. This result, combined with the entropy calculation for $\{\Lambda\}$, gives that

$$\log N_{[]}(\epsilon, \{l_2(\alpha,\beta,h,\Lambda)\}, L_2(P)) \le K_8 \epsilon^{-1}, \tag{7.2}$$

for a fixed constant K_8.

From the definition of the PMLE, we have

$$\mathrm{P}_n l_2(\hat{\alpha},\hat{\beta},\hat{h},\hat{\Lambda}) - \lambda_n^2 J^2(\hat{h}) \ge \mathrm{P}_n l_2(\alpha_0,\beta_0,h_0,\Lambda_0) - \lambda_n^2 J^2(h_0), \tag{7.3}$$

which can also be written as

$$\lambda_n^2 J^2(\hat{h}) + \mathrm{P}[l_2(\alpha_0, \beta_0, h_0, \Lambda_0) - l_2(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \\ \leq \lambda_n^2 J^2(h_0) + (\mathrm{P}_n - \mathrm{P})[l_2(\alpha_0, \beta_0, h_0, \Lambda_0) - l_2(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \tag{7.4}$$

Apply the entropy result in (7.2). We have

$$(\mathrm{P}_n - \mathrm{P})[l_2(\alpha_0, \beta_0, h_0, \Lambda_0) - l_2(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] = (1 + J(h_0) + J(\hat{h}))o_p(n^{-1/2}). \tag{7.5}$$

Combine inequalities (7.4) and (7.5). Simple calculations show that $\lambda_n J(\hat{h}) = o_p(1)$.

Equation (7.4) and assumption B3 hence yield

$$K_2 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \leq o_p(1) + (1 + J(h_0) + J(\hat{h}))o_p(n^{-1/2}).$$

We can then conclude that the PMLE is consistent.

To prove the rate of convergence, we use the following result.

(Theorem in van de Geer 2000, Page 79). Consider a uniformly bounded class of functions Γ, with $\sup_{\gamma \in \Gamma} |\gamma - \gamma_0|_\infty < \infty$ with a fixed $\gamma_0 \in \Gamma$, and $\log N_{[]}(\epsilon, \Gamma, P) \leq K_9 \epsilon^{-b}$ for all $\epsilon > 0$, where $b \in (0, 2)$ and K_9 is a fixed constant. Then for $\delta_n = n^{-1/(2+b)}$,

$$\sup_{\gamma \in \Gamma} \frac{|(\mathrm{P}_n - \mathrm{P})(\gamma - \gamma_0)|}{||\gamma - \gamma_0||_2^{1-b/2} \vee \sqrt{n}\delta_n^2} = O_p(n^{-1/2}), \tag{7.6}$$

where $x \vee y = max(x, y)$. Considering the compactness assumptions A2, A3 and B1, uniqueness assumption B2 and the smoothness of the objective function, we have

$$\begin{aligned} K_2 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) &\leq \mathrm{P}[l_2(\alpha_0, \beta_0, h_0, \Lambda_0) - l_2(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \\ &\leq K_{10} d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)), \end{aligned} \tag{7.7}$$

for a fixed constant K_{10}.

Combining (7.4) with (7.6) for $b = 1$ and (7.7), we have

$$\begin{aligned} &\lambda_n^2 J^2(\hat{h}) + K_2 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \\ &\quad \leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h}))O_p(n^{-1/2}) \\ &\qquad \times (d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}). \end{aligned} \tag{7.8}$$

We thus conclude from (7.8) that

$$\begin{aligned} &\lambda_n^2 J^2(\hat{h}) \leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h}))O_p(n^{-1/2}) \\ &\qquad \times (d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}), \\ &K_2 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h})) \\ &\quad \times O_p(n^{-1/2})(d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}). \end{aligned}$$

Simple calculations give that

$$J(\hat{h}) = O_p(1) \text{ and } d((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) = O_p(n^{-1/3}).$$

Proof of Lemma 9.

The proof of Lemma 9 follows from arguments similar to those in proof of Lemma 5. Note that in the proof of Lemma 5, we only need $(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})$ to nearly maximize the empirical likelihood function, i.e.,

$$\mathrm{P}_n l_2(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) \geq max \mathrm{P}_n l_2(\alpha, \beta, h, \Lambda) - o_p(n^{-1/2}).$$

Note that assumption B3 assumes $\lambda_n = O_p(n^{-1/3})$ and Lemma 5 proves that $J(\hat{h}) = O_p(1)$. So the above nearly maximization requirement is satisfied and Lemma 9 can be proved.