BIOMETRICS x, 000–000 DOI: xxxxxxxxxxxxxxx
January 2014

Supplement to: Identification of Homogeneous and Heterogeneous Variables in Pooled Cohort Studies

Xin Cheng[1,*], Wenbin Lu[2,], and Mengling Liu[1,***]**

[1]Departments of Population Health and Environmental Medicine, New York University School of Medicine, New York, U.S.A

[2]Department of Statistics, North Carolina State University, Raleigh, North Carolina, U.S.A.

**email:* xc311@nyu.edu

***email:* lu@stat.ncsu.edu

****email:* mengling.liu@nyu.edu

In this supplement, we provide the proofs of the theorems.

Following the counting process notation, we define the counting process $N_{ki}(t) = I(T_{ki} \leqslant t, \delta_{ki} = 1)$, and the risk process $Y_{ki}(t) = I(T_{ki} \geqslant t)$. For simplicity, we assume that failure time T^*_{ki} takes values on a finite time interval $[0, \tau]$, and we still use Z_{ki} to denote the predictors corresponding to the transformed parameters θ_n. Then the log partial likelihood $\ell(\theta_n)$ could be expressed as

$$\ell(\theta_n) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^{\tau} \{\theta_n' Z_{ki} - \log(n_k S_k^{(0)}(\theta_n, t))\} dN_{ki}(t),$$

where $S_k^{(m)}(\theta_n, t) = n_k^{-1} \sum_{i=1}^{n_k} Y_{ki}(t) Z_{ki}^{\otimes m} \exp(\theta_n' Z_{ki})$, with $a^{\otimes m} = 1, a, aa', m = 1, 2, 3$ for a vector a.

Let $M_{ki}(t) = N_{ki}(t) - \int_0^t \lambda_{0k}(s) \exp(\theta_n' Z_{ki}) ds$ be the martingale for $N_{ki}(t)$. The regularity conditions are given as follows:

(A) $\int_0^{\tau} \lambda_{0k}(s) ds < \infty$ for $k = 1, \ldots, K$.

(B) There exists a neighborhood $\mathcal{B}$ of the true θ_n^* satisfying: (i) There exist a scalar, a vector, and a matrix $s_k^{(m)}(\theta, t)$ $(m = 0, 1, 2)$, such that $\sup_{t \in [0,\tau], \theta \in \mathcal{B}} \|S_k^{(m)}(\theta_n, t) - s_k^{(m)}(\theta_n, t)\| \to 0$ in probability. (ii) functions $s_k^{(m)}(\theta, t)$ are bounded, and $s_k^{(0)}(\theta, t)$ is bounded away from zero; $s_k^{(m)}(\cdot, t)$ are absolutely continuous for $\theta \in \mathcal{B}$, uniformly in $t \in [0, \tau]$. (iii) let $e_k(\theta_n, t) = s_k^{(1)}(\theta_n, t)/s_k^{(0)}(\theta_n, t), v_k(\theta_n, t) = s_k^{(2)}(\theta_n, t)/s_k^{(0)}(\theta_n, t) - (e_k(\theta_n, t))^{\otimes 2}$, and $I_k(\theta_n^*) = \int_0^{\tau} v_k(\theta_n^*, s) s_k^{(0)}(\theta_n^*, s) \lambda_{0k}(s) ds$ is positive definite with bounded eigenvalues, for $k = 1, \ldots, K$.

(C) For $k = 1, \ldots, K$, there exists a matrix $\Gamma_k = \Gamma_k(\theta_n^*)$ with bounded eigenvalues such that at true θ_n^*, $\|n_k^{-1} \sum_{i=1}^{n_k} Var(D_{ki}) - \Gamma_k\| \to 0$, where $D_{ki} = \int_0^{\tau} [Z_{ki} - e_k(\theta_n, t)] dM_{ki}(t)$.

(D) There exists a constant C such that $\sup_{k \in [1,K], i \in [1,n_k]} E(D_{kij} D_{kil})^2 < C$, where D_{kij}, D_{kil} are the j-th and l-th element of D_{ki}.

Conditions (A)-(D) are also required in Cai et al. (2005), which guarantee the local asymptotic quadratic property for the partial likelihood function and hence the asymptotic normality.

For simplicity, we denote $\lambda_l^{\mu} = \lambda_{1n} \omega_{0l}, \lambda_k^{\alpha} = \lambda_{2n} \omega_{1k}$, and define $a_n = \max\{\lambda_l^{\mu}, \lambda_k^{\alpha} : l \in \mathcal{A}_{1n}, k \in \mathcal{A}_{2n}\}$, and $b_n = \min\{\lambda_l^{\mu}, \lambda_k^{\alpha} : l \in \mathcal{A}_{1n}^c, k \in \mathcal{A}_{2n}^c\}$.

Proof. [of Theorem 1]

Let $\eta_n = \sqrt{q_n/n}$. We show that for any $\epsilon > 0$, there exists a large constant d, such that for any $\triangle u = (\triangle\mu_l', \triangle\alpha_k')'$,

$$P\{\inf_{\|\triangle u\|=d} Q_n(\theta_n^* + \eta_n\triangle u) > Q_n(\theta_n^*)\} > 1 - \epsilon. \tag{1}$$

$$\begin{aligned} Q_n(\theta_n^* + \eta_n\triangle u) - Q_n(\theta_n^*) \geqslant & -\ell(\theta_n^* + \eta_n\triangle u) + \ell(\theta_n^*) + \Big\{ \sum_{l\in\mathcal{A}_{1n}} \lambda_l^{\mu}(|\mu_l^* + \eta_n\triangle\mu_l| - |\mu_l^*|) \\ & + \sum_{k\in\mathcal{A}_{2n}} \lambda_k^{\alpha}(\|\alpha_k^* + \eta_n\triangle\alpha_k\| - \|\alpha_k^*\|)\Big\} \\ \triangleq & H_1 + H_2. \end{aligned}$$

With triangular inequality and Cauchy–Schwarz inequality,

$$\begin{aligned} H_2 \geqslant & -\sum_{l\in\mathcal{A}_{1n}} \lambda_l^{\mu}\eta_n|\triangle\mu_l| - \sum_{k\in\mathcal{A}_{2n}} \lambda_k^{\alpha}\eta_n\|\triangle\alpha_k\| \\ \geqslant & -\sum_{l\in\mathcal{A}_{1n}} a_n\eta_n|\triangle\mu_l| - \sum_{k\in\mathcal{A}_{2n}} a_n\eta_n\|\triangle\alpha_k\| \\ \geqslant & -a_n\eta_n\sqrt{q_n}d \geqslant -n\eta_n^2 d, \end{aligned}$$

the last step is due to the condition $\lambda_{1n}/\sqrt{n} \to 0$, $\lambda_{2n}/\sqrt{n} \to 0$, which implies $a_n/\sqrt{n} \to_p 0$, $a_n\sqrt{q_n} < \sqrt{n}\sqrt{q_n} < \sqrt{q_n/n}n = n\eta_n$.

With Taylor expansion and arguments in Cai et al. (2005),

$$\begin{aligned} H_1 = & -\nabla\ell(\theta_n^*)\eta_n\triangle u - \frac{1}{2}(\eta_n\triangle u)'\nabla^2\ell(\tilde{\theta}_n)(\eta_n\triangle u) \\ \triangleq & H_{11} + H_{12}, \end{aligned}$$

where $\tilde{\theta}_n$ lies between θ_n^* and $\theta_n^* + \eta_n\triangle u$.

$$|H_{11}| \leqslant \eta_n\|\triangle u\| \times \|\nabla\ell(\theta_n^*)\| = O_p(\eta_n\sqrt{nq_n})d = O_p(n\eta_n^2 d).$$

Using Chebyshev's inequality and the assumption $q_n^4/n \to 0$, $\|\frac{1}{n}\nabla^2\ell(\tilde{\theta}_n) + I(\theta_n^*)\| = o_p(1)$(*Caiet al.*, 2005),

$$\begin{aligned} H_{12} = & -\frac{1}{2}n\eta_n^2[\triangle u'\{\frac{1}{n}\nabla^2\ell(\tilde{\theta}_n) + I(\theta_n^*)\}\triangle u] + \frac{1}{2}n\eta_n^2\triangle u' I(\theta_n^*)\triangle u \\ = & \frac{1}{2}n\eta_n^2\triangle u' I(\theta_n^*)\triangle u - \frac{1}{2}n\eta_n^2 d^2 o_p(1). \end{aligned}$$

Therefore, combining H_{11}, H_{12} and H_2, we see H_{12} dominates the other two. So when $\|\triangle u\| = d$ is sufficiently large, $Q_n(\theta_n^* + \eta_n\triangle u) > Q_n(\theta_n^*)$. This completes the proof.

Proof. [of Theorem 2] We show $P(\hat{\theta}_{\mathcal{A}_n^c=0}) \to 1$. Without loss of generality, we assume the true value α_k^* of θ_n^* equals to for a certain k, and show in details $P(\hat{\alpha}_k = 0) \to 1$. Suppose $\hat{\alpha}_k \neq 0$, then Q_n becomes differentiable w.r.t α_k. Therefore,

$$0 = -\frac{\partial \ell}{\partial \alpha_k}(\hat{\theta}_n) + \lambda_k^{\alpha} \frac{\hat{\alpha}_k}{||\hat{\alpha}_k||}. \tag{2}$$

$$\begin{aligned} -\frac{\partial \ell}{\partial \alpha_k}(\hat{\theta}_n) &= -\frac{\partial \ell}{\partial \alpha_k}(\theta_n^*) - \sum_{j=1}^{q_n} \frac{\partial^2 \ell(\tilde{\theta}_n)}{\partial \alpha_k \partial \theta_j}(\hat{\theta}_j - \theta_j^*) \\ &\triangleq H_1 + H_2. \end{aligned}$$

We can easily see that $H_1 = O_p(\sqrt{nq_n})$, and for H_2,

$$\begin{aligned} H_2 = &- \sum_{j=1}^{q_n} \Big(\frac{\partial^2 \ell}{\partial \alpha_k \partial \theta_j} - E\big(\frac{\partial^2 \ell}{\partial \alpha_k \partial \theta_j}\big)\Big)(\hat{\theta}_j - \theta_j^*) - \sum_{j=1}^{q_n} E\big(\frac{\partial^2 \ell}{\partial \alpha_k \partial \theta_j}\big)(\hat{\theta}_j - \theta_j^*) \\ &\triangleq H_{21} + H_{22}. \end{aligned}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned} |H_{21}| &\leqslant \Big[\sum_{j=1}^{q_n} \{\frac{\partial^2 \ell}{\partial \alpha_k \partial \theta_j} - E(\frac{\partial^2 \ell}{\partial \alpha_k \partial \theta_j})\}^2\Big]^{1/2} ||\hat{\theta}_n - \theta_n^*|| \\ &= O_p(\sqrt{q_n n}) O_p(\sqrt{q_n/n}) = o_p(\sqrt{nq_n}), \end{aligned}$$

$$\begin{aligned} |H_{22}| &\leqslant n O_p(1) ||\hat{\theta}_n - \theta_n|| \\ &= n O_p(1) O_p(\sqrt{q_n/n}) = O_p(\sqrt{nq_n}), \end{aligned}$$

we get $H_2 = O_p(\sqrt{nq_n})$. The "$\leqslant$" in H_{22} is due to the finite eigenvalues of the information matrix.

Therefore, $H_1 + H_2 = O_p(\sqrt{nq_n})$. Since $\lambda_{1n}/q_n \to \infty$ and $\lambda_{2n}/q_n \to \infty$, $b_n/\sqrt{nq_n} \to \infty$, $\Big\| \lambda_k^{\alpha} \hat{\alpha}_k / ||\hat{\alpha}_k|| \Big\| \geqslant b_n = \sqrt{nq_n}(b_n/\sqrt{nq_n}) > \sqrt{nq_n} O_p(1)$. That implies the "=" in (2) cannot be satisfied. Proof is completed.

Proof. [of Theorem 3]

We first show $I_{\mathcal{A}_n}(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}) = \frac{1}{n}\nabla_{\mathcal{A}_n} \ell(\theta^*) + o_p(n^{-1/2})$. Then for any $m \times s_n$ matrix B_n, Lindeberg–Feller central limit theorem gives

$$\sqrt{n} B_n I_{\mathcal{A}_n}(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}) = \sqrt{n} B_n I_{\mathcal{A}_n}^{-1/2} \{\frac{1}{n} \nabla_{\mathcal{A}_n} \ell(\theta_n^*)\} \to_d N(0, G).$$

Since $0 = -\nabla_{\mathcal{A}_n}\ell(\hat{\theta}_n) + D(\hat{\theta}_n)$, $D(\hat{\theta}) = (\lambda_l^{\mu}\mathrm{sgn}(\hat{\mu}_l), \lambda_k^{\alpha}\hat{\alpha}_k/\|\hat{\alpha}_k\|)_{l\in\mathcal{A}_{1n},k\in\mathcal{A}_{2n}}$, $\|D(\hat{\theta}_n)\|^2 \leqslant s_n a_n^2$, and λ_{1n}, λ_{2n} satisfy the conditions in Theorem 3, then $a_n^2 = o_p(n/q_n)$, $D(\hat{\theta}_n) = \sqrt{s_n o_p(n/q_n)} = o_p(\sqrt{n})$.

By Taylor expansion,

$$-\nabla_{\mathcal{A}_n}\ell(\hat{\theta}_n) = -\nabla_{\mathcal{A}_n}\ell(\theta_n^*) - \nabla^2_{\mathcal{A}_n}\ell(\tilde{\theta}_n)(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}),$$

$$I_{\mathcal{A}_n}(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}) = -\frac{1}{n}\nabla^2_{\mathcal{A}_n}\ell(\tilde{\theta}_n)(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}) + \left\{I_{\mathcal{A}_n} + \frac{1}{n}\nabla^2_{\mathcal{A}_n}\ell(\tilde{\theta}_n)\right\}(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n})$$
$$\triangleq H_1 + H_2.$$

By Cauchy–Schwarz inequality, we can easily see that $H_2 = o_p(1/\sqrt{n})$. Therefore, $I_{\mathcal{A}_n}(\hat{\theta}_{\mathcal{A}_n} - \theta^*_{\mathcal{A}_n}) = \frac{1}{n}\nabla_{\mathcal{A}_n}\ell(\theta^*) + o_p(n^{-1/2})$.

Now we justify the conditions for Lindeberg–Feller central limit theorem. Let $G_{ki} = \frac{1}{\sqrt{n}}B_n I_{\mathcal{A}_n}^{-1/2}D_{ki.}$, where $D_{ki.}$ corresponds to the nonzero elements in D_{ki}. Since

$$\sum_{i=1}^{n_k} E\left[\|G_{ki}\|^2 I\{\|G_{ki}\| \geqslant \epsilon\}\right] \leqslant \left[\sum_{i=1}^{n_k} E\|G_{ki}\|^4\right]^{1/2}\left[\sum_{i=1}^{n_k} E(I\{\|G_{ki}\| \geqslant \epsilon\})\right]^{1/2}$$
$$\leqslant \sqrt{\frac{1}{n^2}E\left\|\sum_{i=1}^{n_k} B_n I_{\mathcal{A}_n}^{-1/2}D_{ki.}\right\|^4}\sqrt{\frac{\sum_{i=1}^{n_k} E\|G_{ki}\|^2}{\epsilon^2}}$$
$$\leqslant \sqrt{\frac{1}{n^2}\lambda^2_{max}(B_n'B_n)\lambda^2_{min}(B_n'B_n)O_p(s_n^2)} \times O_p(1)$$
$$= o_p(1),$$

then $\sum_{k=1}^{K}\sum_{i=1}^{n_k} E\left[\|G_{ki}\|^2 I\{\|G_{ki}\| \geqslant \epsilon\}\right] = o_p(1)$. By central limit theorem, we prove the asymptotic normality.

References

Cai, J., Fan, J., Li, R., and Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika* **92,** 303–316.