BIOMETRICS 000, 000–000 DOI: 000
000 0000

Supporting Information for "Autologistic Network Model on Binary Data for Disease Progression Study"

Yei Eun Shin[1,*], Huiyan Sang[2], Dawei Liu[3], Toby A. Ferguson[3] and Peter X. K. Song[4]

[1]Biostatistics Branch, Division of Cancer Epidemiology and Genetics,

National Cancer Institute, Rockville, Maryland, U.S.A.

[2]Department of Statistics, Texas A&M University, College Station, Texas, U.S.A.

[3]Biogen, Cambridge, Massachusetts, U.S.A.

[4]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A.

**email:* yei-eun.shin@nih.gov

This paper has been submitted for consideration for publication in *Biometrics*

S1. Interpretation of η-parameters in Section 2

We illustrate a simple example to promote better understanding on the model parameters, especially for autocovariates, denoted by η. Consider two locations s_1 and s_2, and assume that the effect of s_2 on s_1 is of interest when s_1 is normal at $t-1$. The probability of s_1 being diseased at t, $p_m(s_1,t)$, follows the proposed model (1) given the status of s_2 and other locations at $t-1$ and t. Model parameters η_{012} and η_{112} will characterize the effect of s_2 on s_1 as follows. If s_2 were both previously and currently normal (Case 1), $\text{logit}\{p_m(s_1,t)\}$ would decrease as much as $\eta_{012}(0-\kappa_m)$. If s_2 were previously normal but currently diseased (Case 2), $\text{logit}\{p_m(s_1,t)\}$ would increase as much as $\eta_{012}(1-\kappa_m)$. These two cases imply that strongly linked locations with high value of η_{012} are more likely to stay healthy (or be diseased) simultaneously. On the other hand, s_1 would always be ill-affected if s_2 were diseased at the previous time (Case 3); $\text{logit}\{p_m(s_1,t)\}$ will increase as much as $\eta_{112}(1-\kappa_m)$. There are no other cases for s_2 such as being previously diseased but currently normal because of absorbing feature. Likewise, η_{0jk} and η_{1jk} characterize the effects of s_k on s_j for $j \neq k$. See Table S1 below.

[Table 1 about here.]

S2. Derivation of Joint Distribution in Section 3

We build a negpotenial function Q following Besag (1974) and Kaiser and Cressie (2000) to derive a valid joint distribution from conditionals. For a fixed subject m and time t whose previous state is zero (i.e. $Y_m(s_j,t-1)=0$ and so $Y_m(s_j,t) \in \mathcal{P}^0_{mt}$), the conditional density of a response $Y_m(s_j,t)$ at y is

$$f_j\{y|\boldsymbol{X}_m, Y_m(s_k,t-1), Y_m(s_k,t) \text{ for } \forall k \neq j; \boldsymbol{\theta}\} = p_m(s_j,t)^y\{1-p_m(s_j,t)\}^{1-y}.$$

From the model specification in (1), we have its log-conditional density as

$$\log f_j\{Y_m(s_j,t)|\boldsymbol{X}_m, Y_m(s_k,t-1), Y_m(s_k,t) \text{ for } \forall k \neq j; \boldsymbol{\theta}\}$$
$$= Y_m(s_j,t)\Big[\boldsymbol{X}_m^{\mathrm{T}}\boldsymbol{\beta} + \sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}\{Y_m(s_k,t)-\kappa_m\} + \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}\{Y_m(s_k,t)-\kappa_m\}\Big]$$
$$-\log\left(1+\exp\Big[\boldsymbol{X}_m^{\mathrm{T}}\boldsymbol{\beta} + \sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}\{Y_m(s_k,t)-\kappa_m\} + \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}\{Y_m(s_k,t)-\kappa_m\}\Big]\right)$$

for all $Y_m(s_j,t)$ in the active set $\mathcal{P}_{mt}^0$. As the above conditionals indicate only pairwise dependencies, the negpotential function of all responses in the active set has only the first and second order of cliques, so it has the following permutation invariance form,

$$Q(\boldsymbol{Y}_{mt}|\boldsymbol{\theta}) = \sum_{j:\ Y_m(s_j,t)\in\mathcal{A}_{mt}} H_j\{Y_m(s_j,t)\} + \sum_{\substack{j:\,Y_m(s_j,t)\in\mathcal{A}_{mt}\\ k:\,j<k\leq N_s}} H_{j,k}\{Y_m(s_j,t), Y_m(s_k,t)\}.$$

To derive $H_j\{Y_m(s_j,t)\}$ and $H_{j,k}\{Y_m(s_j,t), Y_m(s_k,t)\}$, we follow Besag (1974) and define

$$H_j\{Y_m(s_j,t)\} = \log\frac{f_j\{Y_m(s_j,t)|Y_m^*(s_{-j},t)\}}{f_j\{Y_m^*(s_j,t)|Y_m^*(s_{-j},t)\}},$$
$$H_{j,k}\{Y_m(s_j,t), Y_m(s_k,t)\} = \log\frac{f_j\{Y_m(s_j,t)|Y_m(s_k,t), Y_m^*(s_{-j,-k},t)\} f_j\{Y_m^*(s_j,t)|Y_m^*(s_{-j},t)\}}{f_j\{Y_m^*(s_j,t)|Y_m(s_k,t), Y_m^*(s_{-j,-k},t)\} f_j\{Y_m(s_j,t)|Y_m^*(s_{-j},t)\}}$$

Choosing $Y_m^*(s_j,t)=0$ for each j in active set $\mathcal{P}_{mt}^0$, we obtain

$$H_j\{Y_m(s_j,t)\} = Y_m(s_j,t)\Big\{\boldsymbol{X}_m^{\mathrm{T}}\boldsymbol{\beta} - \sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}\kappa_m - \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}\kappa_m\Big\},$$

$$H_{j,k}\{Y_m(s_j,t), Y_m(s_k,t)\} = \sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}Y_m(s_j,t)Y_m(s_k,t) + \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}Y_m(s_j,t)Y_m(s_k,t).$$

The negpotential function then takes the form in (2) as

$$Q(\boldsymbol{Y}_{mt}|\boldsymbol{\theta}) = \sum_{j\in\mathcal{P}_{mt}^0} Y_m(s_j,t)\Big\{\boldsymbol{X}_m^{\mathrm{T}}\boldsymbol{\beta} - \sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}\kappa_m - \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}\kappa_m\Big\}$$
$$+\frac{1}{2}\sum_{j\in\mathcal{P}_{mt}^0}\Big\{\sum_{k\in\mathcal{P}_{mt}^0\setminus\{j\}}\eta_{0jk}Y_m(s_j,t)Y_m(s_k,t) + \sum_{k\in\mathcal{P}_{mt}^1\setminus\{j\}}\eta_{1jk}Y_m(s_j,t)Y_m(s_k,t)\Big\}.$$

and finally, the joint distribution of $\boldsymbol{Y}_{mt}$ in the support set $\mathcal{S}_{mt}$ given a complete set of conditional distributions, denoted by f, can be specified up to a normalizing constant by Theorem 3 in Kaiser and Cressie (2000).

S3. Proof of Theorem 1 in Section 4

Proof. We first introduce notation to simplify mathematical expressions. For a function $\rho : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, write $\mathcal{P}_n\rho = \sum_{i=1}^{n} \rho_i/n$, and $\mathcal{P}\rho = E(\mathcal{P}_n\rho)$. Also define a function $\rho(\alpha, y) = y\alpha - \log\{1 + \exp(\alpha)\}$. For the binary logistic regression model, we can rewrite the pseudo loglikelihood function as $\ell_c(\boldsymbol{\theta}) = \mathcal{P}_n\rho(\boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}, Y_i)$.

When condition (C2) holds, the second derivative of $\rho(\alpha, y)$ with respect to α is $\ddot{\rho}(\alpha, y) = \text{logit}^{-1}(\alpha)\{1 - \text{logit}^{-1}(\alpha)\}$, which is positive and bounded away from zero. It indicates that $\rho(\alpha, y)$ behaves quadratically near $\alpha^* = \boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}^*$ and hence the quadratic margin condition holds (see, e.g., Section 6.4 of Bühlmann and Van De Geer (2011)), i.e., $\mathcal{P}_n\{\rho(\boldsymbol{\mathcal{X}}_i\widehat{\boldsymbol{\theta}}_\lambda, Y_i) - \rho(\boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}^*, Y_i)\} \geq c\|\boldsymbol{\mathcal{X}}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)\|_2^2/n$ for some constant c.

Furthermore, the restricted eigenvalue condition (C1) implies that the compatibility condition required in Theorem 6.4 in Bühlmann and Van De Geer (2011) holds. Combining these two conditions together, the oracle inequality of the LASSO estimator can be established as $c\|\boldsymbol{\mathcal{X}}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)\|_2^2/n + \lambda\|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|_1 \leq \mathcal{P}_n\{\rho(\boldsymbol{\mathcal{X}}_i\widehat{\boldsymbol{\theta}}_\lambda, Y_i) - \rho(\boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}^*, Y_i)\} + \lambda\|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|_1 = \mathcal{O}(s_0\lambda^2)$, which provides asymptotic bounds for both the prediction error and the ℓ_1 error, i.e., $\|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|_1 = \mathcal{O}_{\mathcal{P}}(s_0\lambda)$, $\|\boldsymbol{\mathcal{X}}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)\|_2^2/n = \mathcal{O}(s_0\lambda^2)$. Under condition (C3), $\widehat{\boldsymbol{\theta}}_\lambda$ is a consistent estimator of $\boldsymbol{\theta}$.

Notice that $\widehat{\boldsymbol{H}} = \frac{1}{n}\boldsymbol{\mathcal{X}}^T\text{diag}\big[\widehat{\pi}_1(\widehat{\boldsymbol{\theta}}_\lambda)\{1 - \widehat{\pi}_1(\widehat{\boldsymbol{\theta}}_\lambda)\}, \ldots, \widehat{\pi}_n(\widehat{\boldsymbol{\theta}}_\lambda)\{1 - \widehat{\pi}_n(\widehat{\boldsymbol{\theta}}_\lambda\}\big]\boldsymbol{\mathcal{X}}$. By conditions (C1) and (C2), $\Lambda_{\min}\{\widehat{\boldsymbol{H}}\} = \min_{\|u\|_2=1} u^T(\frac{1}{n}\boldsymbol{\mathcal{X}}^T\text{diag}\big[\widehat{\pi}_1(\widehat{\boldsymbol{\theta}})\{1 - \widehat{\pi}_1(\widehat{\boldsymbol{\theta}})\}, \ldots, \widehat{\pi}_n(\widehat{\boldsymbol{\theta}})\{1 - \widehat{\pi}_n(\widehat{\boldsymbol{\theta}}\}\big]\boldsymbol{\mathcal{X}})u = \mathcal{O}\{\min_{\|u\|_2=1} u^T(\frac{1}{n}\boldsymbol{\mathcal{X}}^T\boldsymbol{\mathcal{X}})u\} = \mathcal{O}(\Lambda_{\min}\{\boldsymbol{\mathcal{X}}^T\boldsymbol{\mathcal{X}}/n)\}$. Similarly, $\Lambda_{\max}\{\widehat{\boldsymbol{H}}\} = \mathcal{O}\{\Lambda_{\max}(\boldsymbol{\mathcal{X}}^T\boldsymbol{\mathcal{X}}/n)\}$, which indicates $\widehat{\boldsymbol{H}}$ is strictly positive definite. Consider the inverse matrix of $\widehat{\boldsymbol{H}}$, and define it as $\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{H}} = I$. Recall $\Lambda_{\max}(\boldsymbol{\Theta}) = 1/\Lambda_{\min}(\widehat{\boldsymbol{H}})$, and $\Lambda_{\min}(\boldsymbol{\Theta}) = 1/\Lambda_{\max}(\boldsymbol{H})$, which suggests that $\widehat{\boldsymbol{\Theta}}$ is also strictly positive definite with bounded eigenvalues and hence $\|\widehat{\boldsymbol{\Theta}}\|_2 = \Lambda_{\max}(\widehat{\boldsymbol{\Theta}}) = \mathcal{O}(1)$. Combing this fact with $S_n(\widehat{\boldsymbol{\theta}}_\lambda) = \lambda\widehat{\kappa}(\widehat{\boldsymbol{\theta}}_\lambda) \to 0$ and $\widehat{\boldsymbol{\theta}}_\lambda \to \boldsymbol{\theta}^*$, we prove the consistency of $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_\lambda + \widehat{\boldsymbol{\Theta}}S_n(\widehat{\boldsymbol{\theta}}_\lambda)$, i.e. $\widetilde{\boldsymbol{\theta}} \to \boldsymbol{\theta^*}$ as $n \to \infty$.

We next show the asymptotic normality of the biased-corrected estimator. When condition (C2) holds, the third derivative of $\rho(\alpha, y)$ with respect to α exists and its absolute value is bounded by 1, which ensures that the second derivative of $\rho(\alpha, y)$ with respect to α is locally Lipschitz with a universal constant.

From the Taylor expansion of $\dot\rho(\boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}, y)$ and the Lipschitz conditions on $\ddot\rho(\boldsymbol{\mathcal{X}}_i\boldsymbol{\theta}, y)$ for $\forall\theta \in N_\delta(\boldsymbol{\theta}^*)$, we have $\dot\rho(\boldsymbol{\mathcal{X}}_i^T\widehat{\boldsymbol{\theta}}_\lambda, Y_i) = \dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i) + \ddot\rho(\boldsymbol{\mathcal{X}}_i^T\widehat{\boldsymbol{\theta}}_\lambda, Y_i)\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*) + \mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)$. Therefore,

$$
\begin{aligned}
&\widehat{\boldsymbol{\theta}}_\lambda + \widehat{\boldsymbol{\Theta}}S_n(\widehat{\boldsymbol{\theta}}_\lambda) - \boldsymbol{\theta}^*\\
=&\widehat{\boldsymbol{\theta}}_\lambda + \widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\widehat{\boldsymbol{\theta}}_\lambda, Y_i)\boldsymbol{\mathcal{X}}_i\} - \boldsymbol{\theta}^*\\
=&\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^* + \widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{X}}_i\ddot\rho(\boldsymbol{\mathcal{X}}_i^T\widehat{\boldsymbol{\theta}}_\lambda, Y_i)\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*) + \boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)\}\\
=&\widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)\} + [\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^* + \widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\boldsymbol{\mathcal{X}}_i\ddot\rho(\boldsymbol{\mathcal{X}}_i^T\widehat{\boldsymbol{\theta}}_\lambda, Y_i)\boldsymbol{\mathcal{X}}_i^T\}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)]\\
=&\widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)\} + [\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^* + \widehat{\boldsymbol{\Theta}}\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)]\\
=&\widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)\} + [\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^* + \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{H}}(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)]\\
=&\underbrace{\boldsymbol{\Theta}^*\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i\}}_{\boldsymbol{T}_1} + \underbrace{\widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*)|^2)\}}_{\boldsymbol{T}_2} + \underbrace{(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*)\mathcal{P}_n\{\dot\rho(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i\}}_{\boldsymbol{T}_3}
\end{aligned}
\tag{A1}
$$

When conditions (C1)–(C3) hold, by Hölder's inequality, the second term in (A1) is

$$\begin{aligned}\|\boldsymbol{T}_2\|_\infty &= \|\widehat{\boldsymbol{\Theta}}\mathcal{P}_n\{\boldsymbol{\mathcal{X}}_i\mathcal{O}(|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda-\boldsymbol{\theta}^*)|^2)\}\|_\infty \le \mathcal{O}(\mathcal{P}_n\{\|\widehat{\boldsymbol{\Theta}}\boldsymbol{\mathcal{X}}_i\|_\infty|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda-\boldsymbol{\theta}^*)|^2\}) \\ &\le \mathcal{O}(\mathcal{P}_n\{\|\widehat{\boldsymbol{\Theta}}\|_1\|\boldsymbol{\mathcal{X}}_i\|_\infty|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda-\boldsymbol{\theta}^*)|^2\}) \\ &\le \mathcal{O}(\mathcal{P}_n\{\sqrt{p}\|\widehat{\boldsymbol{\Theta}}\|_2\|\boldsymbol{\mathcal{X}}_i\|_\infty|\boldsymbol{\mathcal{X}}_i^T(\widehat{\boldsymbol{\theta}}_\lambda-\boldsymbol{\theta}^*)|^2\}) \\ &= \mathcal{O}(\sqrt{p}\|\boldsymbol{\mathcal{X}}^T(\widehat{\boldsymbol{\theta}}_\lambda-\boldsymbol{\theta}^*)\|_2^2/n) = \mathcal{O}(\sqrt{p}s^*\lambda^2) = o_p(1/\sqrt{n})\end{aligned}$$

Notice that we consider the case with $p < n$, and hence $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_1 = \mathcal{O}(1/\sqrt{n})$. By Hölder's inequality, the third term in (A1) is as follows

$$\begin{aligned}\|\boldsymbol{T}_3\|_\infty &\le \|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_1\|\mathcal{P}_n\{\dot{\rho}(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*, Y_i)\boldsymbol{\mathcal{X}}_i\}\|_\infty = \|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_1\|\{Y_i - \text{logit}^{-1}(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*)\}\boldsymbol{\mathcal{X}}_i\|_\infty \\ &= \|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^*\|_1\|\mathcal{P}_n\{Y_i - \text{logit}^{-1}(\boldsymbol{\mathcal{X}}_i^T\boldsymbol{\theta}^*)\}\|_1\|\mathcal{P}_n\boldsymbol{\mathcal{X}}_i\|_\infty \le o_p(1/\sqrt{n})\end{aligned}$$

We now consider $n^{1/2}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) = n^{1/2}\boldsymbol{A}\boldsymbol{T}_1 + n^{1/2}\boldsymbol{A}(\boldsymbol{T}_2+\boldsymbol{T}_3)$. For any $\boldsymbol{A} \in \mathcal{A}_r$ with fixed r, we have $\|n^{1/2}\boldsymbol{A}(\boldsymbol{T}_2+\boldsymbol{T}_3)\|_\infty \le \|\boldsymbol{A}\|_1\|\sqrt{n}(\boldsymbol{T}_2+\boldsymbol{T}_3)\|_\infty = o_r(1)$

Also recall the Fisher information for the logistic regression is $\boldsymbol{J}^* = \text{var}\{\frac{1}{n}\boldsymbol{\mathcal{X}}^T(\boldsymbol{\mathcal{Y}}-\boldsymbol{\pi}^*)\}$, and the Hessian information is $\boldsymbol{H}^* = \frac{1}{n}\boldsymbol{\mathcal{X}}^T\text{diag}\big[\pi_1^*(1-\pi_1^*),\ldots,\pi_n^*(1-\pi_n^*)\big]\boldsymbol{\mathcal{X}}$, where $\pi_i^* = \text{logit}^{-1}\{\boldsymbol{\mathcal{X}}_i(\kappa_i)^{\mathrm{T}}\boldsymbol{\theta}^*\}$. From conditions (C1) and (C2), both $\boldsymbol{J}^*$ and $\{\boldsymbol{H}^*\}^{-1}$ exist. When $\boldsymbol{A}\boldsymbol{A}^T$ is positive definite with bounded eigenvalues, $\boldsymbol{\Sigma}^{-1/2}$ exists.

From the central limit theorem and the theory of unbiased estimating equation theory (see, e.g., Chapter 3 of Song (2007)), we have $n^{1/2}\boldsymbol{A}\boldsymbol{T}_1 \xrightarrow{d} \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Theta}^*\boldsymbol{J}^*\boldsymbol{\Theta}^*\boldsymbol{A}^T)$.

Finally, we prove that

$$n^{1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{I}_r)$$

References

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Mark S Kaiser and Noel Cressie. The construction of multivariate distributions from Markov random fields. *Journal of Multivariate Analysis*, 73(2):199–220, 2000.

Peter X-K Song. *Correlated data analysis: modeling, analytics, and applications*. Springer Science & Business Media, 2007.

Table S1: Illustrating η-parameters that describe the effect of s_2 on the probability of s_1 being diseased, $p_m(s_1, t)$, which depends on the status of s_2 at previous and current times.

	$Y_m(s_2, t-1)$	$Y_m(s_2, t)$	change in logit$\{p_m(s_1, t)\}$
Case 1	0	0	$\eta_{012}(0 - \kappa_m)$
Case 2	0	1	$\eta_{012}(1 - \kappa_m)$
Case 3	1	1	$\eta_{112}(1 - \kappa_m)$