Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Appendix

Proof of Theorem 1

Let β_0 be the true parameter value generating the observed data, $\dot{\ell}_n(\beta)=\partial\ell_n(\beta)/\partial\beta$ and $\ddot{\ell}_n(\beta)=\partial^2\ell_n(\beta)/\left[\partial\beta\partial\beta'\right]$. Since $\hat{\beta}$ is the MLE, $\dot{\ell}_n(\hat{\beta})=0$, so we have $-\dot{\ell}_n(\beta_0)=\dot{\ell}_n(\hat{\beta})-\dot{\ell}_n(\beta_0)=\ddot{\ell}_n(\beta_n)(\hat{\beta}-\beta_0)$,

Where β_n lies between $\hat{\beta}$ and β_0. Since $\hat{\beta}\to\beta_0$ (a.s.), we have $\beta_n\to\beta_0$ (a.s.). We get

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)=\sqrt{n}\left(-n^{-1}\ddot{\ell}_n(\beta_n)\right)^{-1}n^{-1}\dot{\ell}_n(\beta_0)$$

Note that $-n^{-1}\ddot{\ell}_n(\beta_n)\xrightarrow{P}I(\beta_0)$ and that $\dot{\ell}_n(\beta_0)=\sum_{i=1}^n v_i$

$v_i=\left[\partial f(y_i-x_i\beta_0)/\partial\beta\right]/f(y_i-x_i\beta_0)$. The v_i's are iid with $E(v_i)=0$ and $Var(v_i)=E(v_iv_i^T)=I(\beta_0)$ consequently,

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)\xrightarrow{D}N\left(0,I^{-1}(\beta_0)\right) \quad \text{(A.1)}$$

To simplify notation, assume the columns with partial deletions are $(1,\ldots,k)$. Denote $(n_1,\ldots,n_k)$ be the numbers of x_{ij}'s deleted from columns $(1,\ldots,k)$ of X_n, so $(n_1,\ldots,n_k)/n=(\gamma_1,\ldots,\gamma_k)$ denote $n_0=n-(n_1+\cdots+n_k)$ and $\gamma_0=n_0/n$. Let $\ell_n^-(\beta_0)$ be the likelihood with proportions $(\gamma_1,\ldots,\gamma_k)$ be deleted from columns $(1,\ldots,k)$ in X_n, and denote $\dot{\ell}_n^-(\beta)$ and $\ddot{\ell}_n^-(\beta)$ be the partial derivatives accordingly. Write

$$\ell_n^-(\beta)=\ell_{0,n_0}(\beta)+\sum_{j=1}^k\ell_{-j,n_j}(\beta),$$

Where $\ell_{0,n_0}(\beta)$ is the part of log-likelihood for the n_0 data without covariate deletion, and $\ell_{-j,n_j}(\beta)$ is that for all the n_j data with the j^{th} covariate deleted. Denote $\ddot{\ell}_n^-(\beta)=\ddot{\ell}_{0,n_0}(beta)+\sum_{j=1}^k\ddot{\ell}_{-j,n_j}(\beta)$ accordingly.

For the log-likelihood without data deletion we make the similar decomposition as $\ell_n(\beta)=\ell_{0,n_0}(\beta)+\sum_{j=1}^k\ell_{j,n_j}(\beta)$, accordingly.

Where $\ell_{j,n_j}(\beta)$ is the part of log-likelihood using data from the same individuals as those in $\ell_{-j,n_j}(\beta)$ but without covariate deletion, and denote $\ddot{\ell}_n(\beta)=\ddot{\ell}_{0,n_0}(\beta)+\sum_{j=1}^k\ddot{\ell}_{j,n_j}(\beta)$ accordingly. Note that the same term $\ell_{0,n_0}(\beta)$ appears in both the decompositions of $\ell_n(\beta)$ and $\ell_n^-(\beta)$, the same term $\ddot{\ell}_{0,n_0}(\beta)$ appears in both the decompositions of $\ddot{\ell}_n(\beta)$ and $\ddot{\ell}_n^-(\beta)$ and that for $j=1,\ldots,k$, $\ell_{j,n_j}(\beta)$ is different from $\ell_{-j,n_j}(\beta)$ in that there is no deletion in $\ell_{j,n_j}(\beta)$ although both partial log-likelihoods use data from the same set.

We have

$$\ell_n(\hat{\beta})=\ell_n(\beta_0)-\dot{\ell}_n(\hat{\beta})(\beta_0-\hat{\beta})-\frac{1}{2}(\hat{\beta}-\beta_0)'\ddot{\ell}_n(\beta_n)(\hat{\beta}-\beta_0)$$

$$=\ell_n(\beta_0)-\frac{1}{2}\left((\hat{\beta}-\beta_0)'\ddot{\ell}_{0,n_0}(\beta_n)(\hat{\beta}-\beta_0)+\sum_{j=1}^k(\hat{\beta}-\beta_0)'\ddot{\ell}_{j,n_j}(\beta_n)(\hat{\beta}-\beta_0)\right).$$

Also, let $Z=(Z_1,\ldots,Z_d)$ with Z_j's iid $N(0,1)$, and note $-n_j^{-1}\ddot{\ell}_{j,n_j}(\beta_n)\xrightarrow{P}I(\beta_0)$ for $(j=0,\ldots,k)$, so we have

$$-(\hat{\beta}-\beta_0)'\ddot{\ell}_{j,n_j}(\beta_n)(\hat{\beta}-\beta_0)=-\gamma_j\sqrt{n}(\hat{\beta}-\beta_0)'n_j^{-1}\ddot{\ell}_{j,n_j}(\beta_n)\sqrt{n}(\hat{\beta}-\beta_0)\xrightarrow{D}\gamma_jZ'Z.$$

So we get

$$\ell_n(\hat{\beta})=\ell_n(\beta_0)+\frac{1}{2}\sum_{j=0}^k\gamma_jZ'Z+o_p. \quad (1)$$

Similarly to (A.1) we have, with β_n^- lies between $\hat{\beta}^-$ and β_0,

$$\sqrt{n}\left(\hat{\beta}^{-}-\beta_0\right)=\sqrt{n}\left(-n^{-1}\ddot{\ell}_n^{-}\left(\hat{\beta}^{-}\right)\right)^{-1}n^{-1}\dot{\ell}_n^{-}\left(\beta_0\right).$$

Consequently,

$$\sqrt{n}\left(\hat{\beta}^{-}-\beta_0\right)\xrightarrow{D}N\left(0,I_{-}^{-1}\left(\beta_0\right)\right), \quad \text{(A.2)}$$

Where $I_{-}(\beta_0)=E_{H_0}\left(v_i^{-}v_i^{-'}\right)$ and v_i^{-} is v_i with the j^{th} covariate being removed with probability $\gamma_j\,(j=1,\ldots,k)$.

Also,

$$\ell_n^{-}\left(\hat{\beta}^{-}\right)=\ell_n^{-}\left(\beta_0\right)-\frac{1}{2}\left(\left(\hat{\beta}^{-}-\beta_0\right)'\ddot{\ell}_{0,n_0}\left(\beta_n^{-}\right)\left(\hat{\beta}^{-}-\beta_0\right)+\sum_{j=1}^{k}\left(\hat{\beta}^{-}-\beta_0\right)'\ddot{\ell}_{-j,n_j}\left(\beta_n^{-}\right)\left(\hat{\beta}^{-}-\beta_0\right)\right)$$

and we have

$$-\left(\hat{\beta}^{-}-\beta_0\right)'\ddot{\ell}_{0,n_0}\left(\beta_n^{-}\right)\left(\hat{\beta}^{-}-\beta_0\right)=-\gamma_0\sqrt{n}\left(\hat{\beta}^{-}-\beta_0\right)'n_0^{-1}\ddot{\ell}_{0,n_0}\left(\beta_n^{-}\right)\sqrt{n}\left(\hat{\beta}^{-}-\beta_0\right)\xrightarrow{D}\gamma_0 Z'Z.$$

Note that $\ddot{\ell}_{-j,n_j}\left(\beta_n^{-}\right)$ is a $d\times d$ matrix with the j^{th} row and j^{th} column be zeros, so

$$\left(\hat{\beta}^{-}-\beta_0\right)'\ddot{\ell}_{-j,n_j}\left(\beta_n^{-}\right)\left(\hat{\beta}^{-}-\beta_0\right)=\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)'\ddot{\tilde{\ell}}_{-j,n_j}\left(\beta_n^{-}\right)\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)$$

Where $\hat{\beta}_{-j}$ is the $(d-1)$ dimensional vector with the j^{th} element removed from $\hat{\beta}$ $\hat{\beta}^{-}$, is the $(d-1)$ dimensional vector with the j^{th} element removed from β_0, and $\ddot{\tilde{\ell}}_{-j,n_j}(\beta_n^{-})$ is the $(d-1)\times(d-1)$ matrix with the j^{th} column and j^{th} row removed from $\ddot{\ell}_{-j,n_j}\left(\beta_n\right)$.

Since under H_0 $\ell_n(\beta_0)=\ell_n^{-}(\beta_0)$ now we have, under

$$2\left[\ell_n\left(\hat{\beta}\right)-\ell_n^{-}\left(\hat{\beta}^{-}\right)\right]=\sum_{j=0}^{k}\gamma_j\sqrt{n}\left(\hat{\beta}-\beta_0\right)'\left(-n_j^{-1}\ddot{\ell}_{j,n_j}\left(\beta_n\right)\right)\sqrt{n}\left(\hat{\beta}-\beta_0\right)$$

$$-\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)^{T}\left(-n_0^{-1}\ddot{\tilde{\ell}}_{-j,n_j}\left(\beta_n^{-}\right)\right)\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)$$

$$=\sum_{j=1}^{k}\gamma_j(\sqrt{n}\left(\hat{\beta}-\beta_0\right)'\left(-n_j^{-1}\ddot{\ell}_{j,n_j}\left(\beta_n\right)\right)\sqrt{n}\left(\hat{\beta}-\beta_0\right)$$

$$-\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)^{T}\left(-n_0^{-1}\ddot{\tilde{\ell}}_{-j,n_j}\left(\beta_n^{-}\right)\right)\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)+o_p(1),$$

note that the first term in the above bracket is asymptotically a χ^2 random variable with d-degrees of freedom, while the second term is asymptotically a χ^2 random variable with $(d-1)$ degrees of freedom. As in the proof of Wilks' Theorem (or some more recent proofs, such as in Stat701, 2002), for each j we have

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)^{T}\left(-n_j^{-1}\ddot{\ell}_{j,n_j}\left(\beta_n\right)\right)\sqrt{n}\left(\hat{\beta}-\beta_0\right)$$

$$-\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)'\left(-n_j^{-1}\ddot{\tilde{\ell}}_{-j,n_j}\left(\beta_n^{-}\right)\right)\sqrt{n}\left(\hat{\beta}_{-j}-\beta_{0,-j}\right)\xrightarrow{D}\chi_j^2,$$

Where χ_j^2 is a chi-squared distribution with 1-degree of freedom, and the χ_j^2's are independent for different j, hence we get

$$2\left[\ell_n\left(\hat{\beta}\right)-\ell_n^{-}\left(\hat{\beta}^{-}\right)\right]\xrightarrow{D}\sum_{j=1}^{k}\gamma_j\chi_j^2.$$

Proof of Theorem 2

i) Is from standard argument for the consistency of MLE.

ii) After deleting the irrelevant covariates, the model is $y_i = x_i^{-}\beta + \epsilon_i, \quad \epsilon_i \sim f(\cdot),$

where the x_i^{-}'s are i.i.d. x^{-}, and $x^{-} = x_r^{-}$ with probability $\gamma_r\,(r = 0,1,\ldots,k)$ where $x_{\cdot\cdot}^{-}$ is an i.i.d. copy of the $x_{i,r}^{-}$'s, whose components with index in C_{jr} in particular C_{j0} is the index set for those covariates without partial deletion. The log-likelihood is

$\ell_n^{-}(\beta) = \sum_{i=1}^{n} \log f\left(y_i - x_i^{-}\beta\right)$ By the standard result on regression parameter estimation (eg. Proposition 4.3.1 D and Example 4.3.1 in [17]), the efficient score for β based on $\ell^{-}(\beta)$ is $\dot{\ell}(\beta) = \dfrac{\partial \log f\left(y - x^{-}\beta\right)}{\partial \beta} = \dfrac{\dot{f}\left(y - x^{-}\beta\right)}{f\left(y - x^{-}\beta\right)}\left(x^{-} - \mu^{-}\right),$

Where $\mu_j^{-} = E\left(x^{-}\right)$. Under the common assumption that $\epsilon = y - x^{-}\beta_0$ is independent of x^{-} (or just conditioning on x^{-}), it follows that

$\sqrt{n}\left(\hat{\beta}^{-} - \beta_0\right) \xrightarrow{D} N(0, \Omega)$

Where,

$$\Omega = E_{\beta_0}\left[\dot{\ell}(\beta_0)\dot{\ell}'(\beta_0)\right] = E\left[\left(x^{-} - \mu^{-}\right)\left(x^{-} - \mu^{-}\right)'\right]\int \frac{\dot{f}^2(\epsilon)}{f(\epsilon)}\,d\epsilon.$$