2009 Theory II #1

1a) $\frac{\partial}{\partial(\pi,\lambda)}\left\{ l_n(\pi) + \lambda\left(1 - \sum\sum \pi_{ij}\right)\right\} \overset{set}{=} 0 \Rightarrow \begin{cases} \frac{n_{00}}{\hat{\pi}_{00}} = \lambda \\ \vdots \\ \frac{n_{11}}{\hat{\pi}_{11}} = \lambda \\ \sum\sum \hat{\pi}_{ij} = 1 \end{cases}$

$\Rightarrow \lambda = n \Rightarrow \hat{\pi}_{ij} = \frac{n_{ij}}{n}, \quad \forall (i,j)$

Since $\hat{\pi} = \bar{Y}_n$, by the CLT

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{L} N\left(0, \text{diag}(\pi) - \pi\pi'\right)$$

1b) Define $\bar{y}_1 = \frac{1}{n}(n_{10} + n_{11})$ and $\bar{y}_2 = \frac{1}{n}(n_{01} + n_{11})$. Then

$E\bar{Y}_1 = \frac{1}{n}(En_{10} + En_{11}) = \pi_{10} + \pi_{11} = \pi_{1\cdot}$

$E\bar{Y}_2 = \frac{1}{n}(En_{01} + En_{11}) = \pi_{01} + \pi_{11} = \pi_{\cdot 1}$

so that by CLT

$$\sqrt{n}\left(\begin{bmatrix}\bar{Y}_1 \\ \bar{Y}_2\end{bmatrix} - \begin{bmatrix}\pi_{1\cdot} \\ \pi_{\cdot 1}\end{bmatrix}\right) \xrightarrow{L} N\left(0, \begin{bmatrix} Var[Z_{10}+Z_{11}] & Cov[Z_{10}+Z_{11}, Z_{01}+Z_{11}] \\ Cov[Z_{10}+Z_{11}, Z_{01}+Z_{11}] & Var[Z_{01}+Z_{11}] \end{bmatrix}\right)$$

where $Z = (Z_{00}, Z_{11}, Z_{10}, Z_{11}) \sim mult(1, \pi)$. Then

$Var[Z_{10}+Z_{11}] = Var[Z_{10}] + 2Cov[Z_{10}Z_{11}] + Var[Z_{11}] = \pi_{10}(1-\pi_{10}) - 2\pi_{10}\pi_{11} + \pi_{11}(1-\pi_{11}) = \pi_{1\cdot}(1-\pi_{1\cdot})$

$Var[Z_{01}+Z_{11}] = \pi_{01}(1-\pi_{01}) - 2\pi_{01}\pi_{11} + \pi_{11}(1-\pi_{11}) = \pi_{\cdot 1}(1-\pi_{\cdot 1})$

$Cov[Z_{10}+Z_{11}, Z_{01}+Z_{11}] = \pi_{11}(1-\pi_{11}) - \pi_{01}\pi_{10} - \pi_{01}\pi_{11} - \pi_{10}\pi_{11} = \pi_{00}\pi_{01} - \pi_{01}\pi_{10}$

Now $\begin{cases} \pi_{10} = \operatorname{logit}^{-1}\alpha \\ \pi_{01} = \operatorname{logit}^{-1}(\alpha+\beta) \end{cases} \Rightarrow \begin{cases} \alpha = \operatorname{logit}\pi_{10} \\ \alpha+\beta = \operatorname{logit}\pi_{01} \end{cases}$

$$\Rightarrow \beta = \log\left(\frac{\pi_{01}}{1-\pi_{01}}\right) - \log\left(\frac{\pi_{10}}{1-\pi_{10}}\right) = \log\left(\frac{\pi_{01}(1-\pi_{10})}{(1-\pi_{01})\,\pi_{10}}\right)$$

Let $g(\pi_{10}, \pi_{01}) = \begin{bmatrix} \log\left(\frac{\pi_{10}}{1-\pi_{10}}\right) \\ \log\left(\frac{\pi_{01}(1-\pi_{10})}{(1-\pi_{01})\,\pi_{10}}\right) \end{bmatrix}$

Then

$$\dot{g}(\pi_{10}, \pi_{01}) = \begin{bmatrix} \frac{1}{\pi_{10}(1-\pi_{10})} & 0 \\ -\frac{1}{\pi_{10}(1-\pi_{10})} & \frac{1}{\pi_{01}(1-\pi_{01})} \end{bmatrix}$$

so that the (2,2) entry of

$$\dot{g}(\pi_{10}, \pi_{01})\,\Sigma\,[\dot{g}(\pi_{10}, \pi_{01})]' \equiv A\Sigma A' = A_{21}^2\Sigma_{11} + A_{22}^2\Sigma_{22} + 2A_{21}A_{22}\Sigma_{12}$$

$$= \left(\frac{1}{\pi_{10}(1-\pi_{10})}\right)^2 \pi_{10}(1+\pi_{10}) + \left(\frac{1}{\pi_{01}(1-\pi_{01})}\right)^2 \pi_{01}(1-\pi_{01}) - \frac{2(\pi_{00}\pi_{01} - \pi_{01}\pi_{10})}{\pi_{10}(1-\pi_{10})\,\pi_{01}(1-\pi_{01})}$$

$$= \frac{1}{\pi_{00}\pi_{10}} + \frac{1}{\pi_{00}\pi_{01}} - \frac{2(\pi_{00}\pi_{01} - \pi_{01}\pi_{10})}{\pi_{00}\pi_{10}\,\pi_{00}\pi_{01}}$$

1c) The model assumes that

$$P(Y_{i1}=1) = \frac{e^{\alpha_i}}{1+e^{\alpha_i}}, \quad P(Y_{i2}=1) = \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}$$

and that responses for different subjects and for the two observations on the same subject are independent (averaged over all the subjects, the responses are nonnegatively associated). We wish to make inference regarding β and consider α to be a nuisance param.

		Y_{i2}: 0	Y_{i2}: 1	
Y_{ij}	0	π_{00}	π_{01}	$\frac{1}{1+e^{\alpha_i}}$
	1	π_{10}	π_{11}	$\frac{e^{\alpha_i}}{1+e^{\alpha_i}}$
		$\frac{1}{1+e^{\alpha_i+\beta}}$	$\frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}$	

The joint mass function for $\{(y_{11}, y_{12}), \ldots, (y_{n1}, y_{n2})\}$ is given by

$$\prod_{i=1}^{n} \left(\frac{e^{\alpha_i}}{1+e^{\alpha_i}}\right)^{y_{i1}} \left(\frac{1}{1+e^{\alpha_i}}\right)^{1-y_{i1}} \left(\frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}\right)^{y_{i2}} \left(\frac{1}{1+e^{\alpha_i+\beta}}\right)^{1-y_{i2}}$$

$$= \prod_{i=1}^{n} e^{y_{i1}\alpha_i} (1+e^{\alpha_i})^{-1} e^{y_{i2}(\alpha_i+\beta)} (1+e^{\alpha_i+\beta})^{-1}$$

$$= \prod_{i=1}^{n} \exp\left\{ y_{i1}\alpha_i + y_{i2}(\alpha_i+\beta) - \log(1+e^{\alpha_i}) - \log(1+e^{\alpha_i+\beta}) \right\}$$

$$= \exp\left\{ \sum_{i=1}^{n} (y_{i1}+y_{i2})\alpha_i + \sum_{i=1}^{n} y_{i2}\beta + c(\alpha,\beta) \right\}$$

The parameter space is full rank ($\mathbb{R}^{n+1}$) so that $(y_{i1}+y_{i2})$ is sufficient for α_i, $i=1,\ldots,n$. Denote $S_i = Y_{i1}+Y_{i2}$, $i=1,\ldots,n$

Given $S_i = 0$ then $P(Y_{i1} = Y_{i2} = 0) = 1$, and given $S_i = 2$ then $P(Y_{i1} = Y_{i2} = 1) = 1$; the distribution depends on β only when $S_i = 1$. We have for $y_{i1} + y_{i2} = 1$

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid S_i = 1) = \frac{P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, S_i = 1)}{P(S_i = 1)}$$

$$= \frac{P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2})}{P(Y_{i1} = 0, Y_{i2} = 1) + P(Y_{i1} = 1, Y_{i2} = 0)}$$

$$= \frac{\left(\frac{e^{\alpha_i}}{1+e^{\alpha_i}}\right)^{y_{i1}} \left(\frac{1}{1+e^{\alpha_i}}\right)^{1-y_{i1}} \left(\frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}\right)^{y_{i2}} \left(\frac{1}{1+e^{\alpha_i+\beta}}\right)^{1-y_{i2}}}{\frac{1}{1+e^{\alpha_i}} \cdot \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}} + \frac{e^{\alpha_i}}{1+e^{\alpha_i}} \cdot \frac{1}{1+e^{\alpha_i+\beta}}}$$

$$= \frac{e^{y_{i2}\beta}}{1+e^{\beta}} = \begin{cases} \frac{e^{\beta}}{1+e^{\beta}}, & \{y_{i1} = 0, y_{i2} = 1\} \\ \frac{1}{1+e^{\beta}}, & \{y_{i1} = 1, y_{i2} = 0\} \end{cases}$$

Let $n_{12} = \sum_{i=1}^{n} I(y_{i1} = 1, y_{i2} = 0)$, $n^* = \sum_{i=1}^{n} I(s_i = 1)$

Then, conditional on $\{S_1, \dots, S_n\}$ the joint distribution of the matched pairs is given by

$$\prod_{\{i: s_i = 1\}} \left(\frac{1}{1+e^{\beta}}\right)^{y_{i1}} \left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{y_{i2}} = \frac{[e^{\beta}]^{n_{21}}}{[1+e^{\beta}]^{n^*}}$$

Then

$$\frac{\partial}{\partial\beta} \ell_n(\beta) = \frac{\partial}{\partial\beta}\left\{ n_{21}\beta - n^* \log(1+e^{\beta})\right\}$$

$$= n_{21} - \frac{n^* e^{\beta}}{1+e^{\beta}} \overset{\text{set}}{=} 0 \Rightarrow n_{21} + n_{21} e^{\hat\beta} = n^* e^{\hat\beta}$$

$$\Rightarrow n_{21} = (n^* - n_{21}) e^{\hat\beta} \Rightarrow \hat\beta = \log\left(\frac{n_{21}}{n^* - n_{21}}\right) = \log\left(\frac{n_{21}}{n_{12}}\right)$$

$$\frac{\partial^2}{\partial\beta^2} \ell_n(\beta) = -n^* \frac{e^{\beta}(1+e^{\beta}) - e^{2\beta}}{(1+e^{\beta})^2} = -n^* \frac{e^{\beta}}{1+e^{\beta}} \cdot \frac{1}{1+e^{\beta}}$$

Since n^* is fixed conditional on $(S_1, \ldots, S_n)$ we obtain

$$I_n(\beta) = \frac{n^* e^{\beta}}{(1+e^{\beta})^2}$$

and

$$I_n(\hat\beta) = \frac{\frac{n^* n_{21}}{n_{12}}}{\left(1+\frac{n_{21}}{n_{12}}\right)^2} = \frac{\frac{n^* n_{21}}{n_{12}}}{\left(\frac{n^*}{n_{12}}\right)^2} = \frac{n_{12} n_{21}}{n^*}$$

Then

$$\widehat{Var}[\hat\beta] = [I_n(\hat\beta)]^{-1} = \frac{n^*}{n_{12} n_{21}} = \frac{n_{12}+n_{21}}{n_{12} n_{21}} = \frac{1}{n_{12}} + \frac{1}{n_{21}}$$

and

$$\widehat{s.e.}[\hat\beta] = \sqrt{\frac{1}{n_{12}} + \frac{1}{n_{21}}}$$

1d)

$$l_n(\alpha,\beta) = \sum_{i=1}^{n}\left\{(y_{i1}+y_{i2})\alpha_i - \log(1+e^{\alpha_i}) + y_{i2}\beta - \log(1+e^{\alpha_i+\beta})\right\}$$

$$\frac{d}{d\alpha_i} l_n(\alpha,\beta) = y_{i1}+y_{i2} - \frac{e^{\alpha_i}}{1+e^{\alpha_i}} - \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}} \overset{\text{set}}{=} 0 \Rightarrow y_{i1}+y_{i2} = \frac{e^{\alpha_i}}{1+e^{\alpha_i}} + \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}}$$

$$\frac{d}{d\beta} l_n(\alpha,\beta) = \sum_{i=1}^{n} y_{i2} - \sum_{i=1}^{n}\frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}} \overset{\text{set}}{=} 0 \Rightarrow n_{\cdot 1} = \sum_{i=1}^{n}\frac{e^{\hat\alpha_i+\hat\beta}}{1+e^{\hat\alpha_i+\hat\beta}}$$

for i s.t.

$$\begin{cases} y_{i1}+y_{i2} = 0 \Rightarrow \hat\alpha_i = -\infty \\ y_{i1}+y_{i2} = 2 \Rightarrow \hat\alpha_i = \infty \\ y_{i1}+y_{i2} = 1 \Rightarrow \frac{1}{e^{\hat\alpha_i}} = \frac{e^{\hat\alpha_i+\hat\beta}}{1+e^{\hat\alpha_i+\hat\beta}}\cdot\frac{e^{-\hat\alpha_i-\hat\beta}}{e^{-\hat\alpha_i-\hat\beta}} = \frac{1}{1+e^{-\hat\alpha_i-\hat\beta}} \end{cases}$$

$$\Rightarrow \hat\alpha_i = -\hat\alpha_i - \hat\beta \Rightarrow \hat\beta = -2\hat\alpha_i$$

so that

$$n_{01} = \sum_{i=1}^{n}\frac{e^{\hat\alpha_i+\hat\beta}}{1+e^{\hat\alpha_i+\hat\beta}} = \sum_{\{i:\, S_i=0\}} 0 + \sum_{\{i:\, S_i=2\}} 1 + \sum_{\{i:\, S_i=1\}}\frac{e^{\hat\beta/2}}{1+e^{\hat\beta/2}} = n_{11} + (n_{01}+n_{10})\frac{e^{\hat\beta/2}}{1+e^{\hat\beta/2}}$$

$$\Rightarrow n_{01} + n_{01}e^{\hat\beta/2} = (n_{01}+n_{10})e^{\hat\beta/2} \Rightarrow n_{01} = n_{10}e^{\hat\beta/2}$$

$$\Rightarrow \hat\beta = 2\log\left(\frac{n_{01}}{n_{10}}\right) = 2\log\left(\frac{n_{01}/n}{n_{10}/n}\right) \xrightarrow{P} 2\log\left(\frac{\pi_{01}}{\pi_{10}}\right)$$

but we saw earlier that $\beta = \log\left(\frac{\pi_{\cdot 1}\pi_{0\cdot}}{\pi_{\cdot 0}\pi_{1\cdot}}\right) = \log\left(\frac{(\pi_{01}+\pi_{11})(\pi_{00}+\pi_{01})}{(\pi_{00}+\pi_{10})(\pi_{10}+\pi_{11})}\right)$

$= \log\left(\frac{\pi_{00}\pi_{01}+\pi_{00}\pi_{11}+\pi_{01}\pi_{11}+\pi_{01}^2}{\pi_{00}\pi_{10}+\pi_{00}\pi_{11}+\pi_{10}^2+\pi_{10}\pi_{11}}\right)$ so $\hat\beta$ is not consistent