2009W Qualifying Exam Section 2

February 21, 2019

1 Question 1

1.a 1.a

We can write the likelihood function as

$$L(\pi) = \prod_{i=1}^{2}\prod_{j=1}^{2} \pi_{ij}^{n_{ij}}$$

So the log-likelihood function is given by

$$\ell(\pi) = \sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\log\pi_{ij}$$

Since $\sum_i \sum_j \pi_{ij} = 1$ we have a constrained optimization problem. Let the objective function be

$$Q(\pi, \lambda) = \sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\log\pi_{ij} + \lambda\left(1 - \sum_{i=1}^{2}\sum_{j=1}^{2}\pi_{ij}\right)$$

We have for $i = 1, 2$, $j = 1, 2$,

$$\begin{aligned} \frac{\partial Q}{\partial \pi_{ij}} &= \frac{n_{ij}}{\pi_{ij}} - \lambda \overset{\text{SET}}{=} 0 \\ &\implies \hat{\pi}_{ij} = \frac{n_{ij}}{\lambda} \end{aligned} \tag{1.1}$$

Plugging in the constraint, we have

$$1 = \sum_{i=1}^{2}\sum_{j=1}^{2}\hat{\pi}_{ij} = \frac{1}{\lambda}\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij} = \frac{n}{\lambda} \implies \lambda = n \tag{1.2}$$

Plugging (1.2) into (1.1), the maximum likelihood estimate of π_{ij}, $i = 1, 2$, $j = 1, 2$ is given by

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

Note that we can write

$$n_{ij} = \sum_{k=1}^{n} \mathbf{1}_{\{Y_{k1}=i, Y_{k2}=j\}}$$

So that $\hat{\pi}_{ij}$ is simply an i.i.d. average of the indicator functions. Let $U_{ij} \equiv \mathbf{1}_{\{Y_{11}=i, Y_{12}=j\}}$. Then we have

$$\mathbb{E}\hat{\pi}_{ij} = \mathbb{E}U_{ij} = \Pr(Y_{11} = i, Y_{12} = j) = \pi_{ij}$$

$$\begin{aligned}
\mathrm{Cov}(U_{ij}, U_{kl}) &= \mathbb{E}U_{ij}U_{kl} - (\mathbb{E}U_{ij})(\mathbb{E}U_{kl}) \\
&= \mathbb{E}U_{ij}U_{kl} - \pi_{ij}\pi_{kl} \\
&= \begin{cases} -\pi_{ij}\pi_{kl} & \text{if } i \neq k \text{ or } j \neq l \\ \mathrm{Var}(U_{ij}) = \pi_{ij} - \pi_{ij}^2 & \text{if } i = k, j = l \end{cases}
\end{aligned}$$

since the U_{ij} are indicators. It is clear then that we can write $\Sigma = \mathrm{diag}(\pi) - \pi\pi^T$

Since the MLEs are just an iid average, by the multivariate central limit theorem we have

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{\mathrm{d}} N(0, \Sigma)$$

1.b 1.b

We have

$$
\begin{aligned}
&\pi_{1+} = \pi_1 1 + \pi_{10} = \text{expit}(\alpha) \\
\implies &\alpha = \text{logit}(\pi_{11} + \pi_{10})
\end{aligned}
$$

$$
\begin{aligned}
&\pi_{+1} = \pi_{11} + \pi_{01} = \text{expit}(\alpha + \beta) \\
\implies &\beta = \text{logit}(\pi_{11} + \pi_{01}) - \alpha
\end{aligned}
$$

By the invariance property of MLEs,

$$\hat{\alpha}_M = \text{logit}(\hat{\pi}_{11} + \hat{\pi}_{10})$$

$$\hat{\beta}_M = \text{logit}(\hat{\pi}_{11} + \hat{\pi}_{01}) - \hat{\alpha}_M$$

1.c 1.c

The likelihood function is

$$\begin{aligned}
L_i(\alpha, \beta) &= L(\alpha, \beta | y_{i1}, y_{i2}) \\
&= \left(\frac{e^{\alpha_i}}{1 + e^{\alpha_i}}\right)^{y_{i1}} \left(\frac{1}{1 + e^{\alpha_i}}\right)^{1-y_{i1}} \left(\frac{e^{\alpha_i+\beta}}{1 + e^{\alpha_i+\beta}}\right)^{y_{i2}} \left(\frac{1}{1 + e^{\alpha_i+\beta}}\right)^{1-y_{i2}} \\
&= (e^{\alpha_i})^{y_{i1}} \left(e^{\alpha_i+\beta}\right)^{y_{i2}} \left[(1 + e^{\alpha_i+\beta})(1 + e^{\alpha_i})\right]^{-1} \\
&= (e^{\beta})^{y_{i2}} (e^{\alpha_i})^{y_{i1}+y_{i2}} \left[(1 + e^{\alpha_i+\beta})(1 + e^{\alpha_i})\right]^{-1}
\end{aligned}$$

Thus, since e^x is a bijective function, we have that $s_i \equiv y_{i1}+y_{i2}$ is sufficient for α_i, $i = 1, \dots, n$. We want to find the conditional distribution given s_i.

Note that $P(y_{i1} = 0, y_{i2} = 0 | s_i = 0) = 1 = P(y_{i1} = 1, y_{i2} = 1 | s_i = 2)$

We have

$$P(s_{i1} = 1) = P(y_{i1} = 1, y_{i2} = 0) + P(y_{i1} = 0, y_{i2} = 0)$$

$$= \frac{e^{\alpha_i}}{1 + e^{\alpha_i}} \frac{1}{1 + e^{\alpha_i+\beta}} + \frac{1}{1 + e^{\alpha_i}} \frac{e^{\alpha_i+\beta}}{1 + e^{\alpha_i+\beta}}$$

$$= \frac{e^{\alpha_i}(1 + e^{\beta})}{(1 + e^{\alpha_i})(1 + e^{\alpha_i+\beta})}$$

Moreover,

$$P(y_{i2} = 1, s_i = 1) = P(y_{i2} = 1, y_{i1} + y_{i2} = 1) = P(y_{i1} = 0, y_{i1} = 1)$$

$$= \frac{1}{1 + e^{\alpha_i}} \frac{e^{\alpha_i+\beta}}{1 + e^{\alpha_i+\beta}}$$

$$= \frac{e^{\alpha_i} e^{\beta}}{(1 + e^{\alpha_i})(1 + e^{\alpha_i+\beta})}$$

Thus, we have

$$P(y_{i2} = 1 | s_i = 1) = \frac{P(y_{i2} = 1, s_i = 1)}{P(s_i = 1)}$$

$$\begin{aligned}
&= \frac{e^{\alpha_i} e^{\beta}}{e^{\alpha_i}(1 + e^{\beta})} \\
&= \frac{e^{\beta}}{1 + e^{\beta}}
\end{aligned}$$

Let $L_c(\beta)$ represent the conditional likelihood. Then when $s_i = 1$ the conditional likelihood is given by

$$\begin{aligned} L_c(\beta) &= \prod_{\{i:s_i=1\}} \left(\frac{e^\beta}{1+e^\beta}\right)^{y_{i2}} \left(\frac{1}{1+e^\beta}\right)^{1-y_{i2}} \\ &= \left(e^\beta\right)^{n_{01}} \left(1+e^\beta\right)^{n_*} \end{aligned}$$

Thus,

$$\ell_c = n_{01}\beta - n_* \log\left(1+e^\beta\right)$$

$$\begin{aligned} \frac{\mathrm{d}\ell_c}{\mathrm{d}\beta} &= n_{01} - n_* \frac{e^\beta}{1+e^\beta} \overset{\text{SET}}{=} 0 \\ \implies \frac{e^\beta}{1+e^\beta} &= \frac{n_{01}}{n_*} \end{aligned}$$

$$\begin{aligned} \implies \hat{\beta}_C &= \log \frac{n_{01}/n_*}{1 - n_{01}/n_*} \\ &= \log \frac{n_{01}/n_*}{(n_* - n_{01})/n_*} \\ &= \log \frac{n_{01}}{n_* - n_{01}} = \log \frac{n_{01}}{n_{10}} \end{aligned}$$

Note since $\hat{\pi}_{ij} = n_{ij}/n$, we can write

$$\hat{\beta}_C = \log \frac{n_{01}}{n_{10}} = \log \frac{n\hat{\pi}_{01}}{n\hat{\pi}_{10}} = \log \frac{\hat{\pi}_{01}}{\hat{\pi}_{10}} = \log \hat{\pi}_{01} - \log \hat{\pi}_{10}$$

Let $\pi_* = (\pi_{01}\pi_{10})^T$ and $\hat{\pi}_*$ be the estimate of π_*. By part a and the continuous mapping theorem, we have

$$\sqrt{n}(\hat{\pi}_* - \pi_*) \xrightarrow{\mathrm{d}} N(0, \Sigma_*)$$

where $\Sigma_* = \begin{pmatrix} \pi_{01}(1-\pi_{01}) & -\pi_{01}\pi_{10} \\ -\pi_{01}\pi_{10} & \pi_{10}(1-\pi_{10}) \end{pmatrix}$

Let $g : \mathbb{R}^2 \to \mathbb{R}^1$ be given by

$$g(\pi_{01}, \pi_{10}) = \log \pi_{01} - \log \pi_{10}$$

Then

$$\begin{aligned} \ell_1 &\equiv \frac{\partial g}{\partial \pi_{01}} = \frac{1}{\pi_{01}} \\ \ell_2 &\equiv \frac{\partial g}{\partial \pi_{10}} = -\frac{1}{\pi_{10}} \end{aligned}$$

Let $\ell = (\ell_1, \ell_2)^T$. Since $\pi_{ij} \neq 0$, by the Delta Method we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\text{d}} N(0, \tau^2)$$

where

$$\begin{aligned}
\tau^2 &= \ell^T \Sigma_* \ell \\
&= (1/\pi_{01})^2 \pi_{01}(1 - \pi_{01}) + (1/\pi_{10})^2 \pi_{10}(1 - \pi_{10}) + 2(1/\pi_{01})(-1/\pi_{10})(-\pi_{01}\pi_{10}) \\
&= \frac{1 - \pi_{01}}{\pi_{01}} + \frac{1 - \pi_{10}}{\pi_{10}} + 2 \\
&= \frac{1}{\pi_{01}} - 1 + \frac{1}{\pi_{10}} - 1 + 2 \\
&= \frac{1}{\pi_{01}} + \frac{1}{\pi_{10}}
\end{aligned}$$

2 Problem 2

2.a

2.a.1

We can express X as

$$X = \begin{pmatrix} \mathbf{1}_{n_1} & \boldsymbol{z}_1 & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \boldsymbol{z}_2 \end{pmatrix}$$

where $\boldsymbol{z}_i = (z_{ij}) = x_{ij} - \bar{x}_i, \quad i = 1, 2$ and $\boldsymbol{z}_i \in R^{n_i}$

Since ϵ_{ij} are i.i.d. $N(0, \sigma^2)$, we have that

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$$

2.a.2

By the Gauss Markov Theorem, the BLUE is given by $\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}$. This is valid since it is clear that X is full rank since its columns are clearly linearly independent since x_{ij} are not all equal, and hence $\boldsymbol{z}_i \neq \mathbf{0}$ and not proportional to $\mathbf{1}$.

2.b

Let $\boldsymbol{a} = (0, 1, 0, -1)^T$. Then $\boldsymbol{a}^T \beta = \gamma_1 - \gamma_2$. Note that $a^T \beta$ is estimable if and only if $a \in C(X^T)$. We have

$$X^T = \begin{pmatrix} \mathbf{1}_{n_1}^T & \mathbf{0}_{n_2}^T \\ \\ \boldsymbol{z}_1^T & \mathbf{0}_{n_2}^T \\ \\ \mathbf{0}_{n_1}^T & \mathbf{1}_{n_2}^T \\ \\ \mathbf{0}_{n_1}^T & \boldsymbol{z}_2^T \end{pmatrix}$$

Let $P = X^T$ and write $P = (p_{11}, \ldots p_{1n_1}, p_{21}, \ldots p_{2n_2})$ If $\boldsymbol{a}^T \beta \in C(X^T)$, then we have

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} = \sum_{i=1}^{n_1} c_{1i} \boldsymbol{p}_{1i} + \sum_{j=1}^{n_2} c_{2j} \boldsymbol{p}_{1j}$$

Since the first element in a is 0, we must have $c_{1i} = 0$ for $i = 1, \ldots, n_1$ since the second n_2 columns of P are 0. Similarly, we must have $c_{2j} = 0$ for $j = 1, \ldots, n_2$ since the third element in a is 0 and the first n_1 columns of P are 0. Clearly, this does not make the vector a, so $a^T \beta$ is not estimable.