# STUDENT SOLUTION MANUAL

## 2012 THEORY SECTION, PART II

BYRON[1]

March 8, 2015

## NOTATION

| Symbol | Meaning |
|---|---|
| $\mathcal{V}(Y)$ | *Variance of* $Y$ |
| $\mathcal{V}(\mathbf{Y})$ | *Covariance Matrix of* $\mathbf{Y}$ |
| $\mathcal{V}(X, Y)$ | *Covariance of* $X$ *and* $Y$ |
| $\partial_x(y)$ | $\partial y / \partial x$ |
| $\partial_{x,z}(y)$ | $\partial^2 y / \partial x \partial z$ |
| $\partial_{x^2,z}(y)$ | $\partial^3 y / \partial x \partial x \partial z$ |
| $\mathcal{S}(\theta)$ | *Score Function of* $\theta$ |
| $\mathcal{H}(\theta)$ | *Hessian Matrix of* $\theta$ |
| $\mathcal{I}(\theta)$ | *Fisher Information Matrix of* $\theta$ |

[1]If you have comments or suggestions, you can reach me at *byron.jaeger@gmail.com*

# 1 PROBLEM # 1

*After a certain surgical procedure, some patients develop a wound infection. Typically, the infection is treated and cleared. However, some patients develop another wound infection. The first infection is called the "primary infection", while the second is called a "secondary infection". An investigator is interested in the question whether the risk of a secondary infection in those who have had a primary infection is the same as the risk of a primary infection.*
*Data are collected on a random sample of n patients. Assume that the* $n$ *responses are independent and identically distributed. For the* $i^{th}$ *patient,* $1 \leqslant i \leqslant n$, *let*

$$Y_{i1} = \begin{cases} 1 & \textit{if subject i developed a primary infection} \\ 0 & \textit{otherwise} \end{cases}$$

*and*

$$Y_{i2} = \begin{cases} 1 & \textit{if subject i developed a secondary infection} \\ 0 & \textit{otherwise} \end{cases}$$

*Define*

$$\alpha = \mathbb{P}\left(Y_{i1} = 1\right) \;\; \textit{and} \;\; \beta = \mathbb{P}\left(Y_{i2} = 1 \mid Y_{i1} = 1\right)$$

*Both* $\alpha$ *and* $\beta$ *take values in* $(0, 1)$. *Suppose there are*

- $X_1$ *patients with* $Y_{i1} = 1, Y_{i2} = 1$
- $X_2$ *patients with* $Y_{i1} = 1, Y_{i2} = 0$
- $X_3$ *patients with* $Y_{i1} = 0, Y_{i2} = 0$

*Also, note that*

- $X_1 + X_2 + X_3 = n$.
- *By definition, a secondary infection can occur only in patients who have had a primary infection.*

---

Part A

*Does the distribution of the data have the form of the exponential family? Give details*

---

**Solution:**

From the description of the data, we can write

$$\mathbf{X} \overset{\text{set}}{=} (X_1, X_2, X_3) \sim \text{Multinomial}\left(n, \mathbf{p}\right), \text{ where } \mathbf{p} = \left(\, \alpha\beta, \, \alpha(1-\beta), \, 1-\alpha \,\right)$$

Therefore, the log-likelihood is given by

$$\ell(\alpha, \beta) = \log\binom{n}{x_1, x_2, x_3} + x_1 \log(\alpha\beta) + x_2 \log(\alpha(1-\beta)) + x_3 \log(1-\alpha) \tag{1}$$

$$= c(\boldsymbol{x}) + x_1 \log(\alpha\beta) + x_2 \log(\alpha(1-\beta)) + (n - x_1 - x_2) \log(1-\alpha) \tag{2}$$

$$= c(\boldsymbol{x}) + x_1 \log\left(\frac{\alpha\beta}{1-\alpha}\right) + x_2 \log\left(\frac{\alpha \cdot (1-\beta)}{1-\alpha}\right) - n \log\left(\frac{1}{1-\alpha}\right) \tag{3}$$

$$= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \cdot \boldsymbol{\theta} - b(\boldsymbol{\theta}) + c(\boldsymbol{x}) \tag{4}$$

where (see below for details)

- $\boldsymbol{\theta}^T = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix} = \begin{pmatrix} \log\left(\frac{\alpha \cdot \beta}{1-\alpha}\right) & \log\left(\frac{\alpha \cdot (1-\beta)}{1-\alpha}\right) \end{pmatrix}$
- $b(\boldsymbol{\theta}) = n \cdot \log\left(1 + e^{\theta_1} + e^{\theta_2}\right)$
- $c(\boldsymbol{x}) = \log\binom{n}{x_1, x_2, x_3}$

Therefore, the distribution of the data fit the form of the exponential family. We will need the explicit form of the inverse link functions for later. First, we solve for $\beta$.

$$\frac{\alpha \cdot \beta}{1-\alpha} = e^{\theta_1} \Longrightarrow \beta = \left(\frac{1-\alpha}{\alpha}\right) e^{\theta_1}$$

$$\frac{\alpha \cdot (1-\beta)}{1-\alpha} = e^{\theta_2} \Longrightarrow \frac{1-\alpha}{\alpha} = (1-\beta)e^{-\theta_2}$$

Thus,

$$\beta = (1-\beta)e^{\theta_1 - \theta_2} \Longrightarrow \beta = \frac{e^{\theta_1 - \theta_2}}{1 + e^{\theta_1 - \theta_2}}$$

Now we can find $\alpha$

$$\frac{1-\alpha}{\alpha} = \frac{e^{-\theta_2}}{1 + e^{\theta_1 - \theta_2}} \Longrightarrow \alpha = \frac{1 + e^{\theta_1 - \theta_2}}{1 + e^{-\theta_2} + e^{\theta_1 - \theta_2}}$$

Note that

$$\frac{1}{1-\alpha} = \left(\frac{e^{-\theta_2}}{1 + e^{-\theta_2} + e^{\theta_1 - \theta_2}}\right)^{-1} = 1 + e^{\theta_1} + e^{\theta_2}$$

---

Part B

*Derive the maximum-likelihood estimators of $\alpha$ and $\beta$.*

---

**Solution:**

The properties of multinomial distributions give us

$$\alpha = \mathbb{P}(Y_{i1} = 1) \Rightarrow \widehat{\alpha} = \frac{X_1 + X_2}{n}$$

and

$$\beta = \mathbb{P}(Y_{i2} = 1 \mid Y_{i1} = 1) = \frac{X_1}{X_1 + X_2}$$

We can verify this with the log likelihood in 3.

$$\frac{\partial(\ell(\alpha, \beta))}{\partial(\beta)} = x_1 \cdot \frac{1}{\beta} - x_2 \cdot \frac{1}{1-\beta} \overset{\text{set}}{=} 0 \Rightarrow \widehat{\beta} = \frac{x_1}{x_1 + x_2}$$

$$\frac{\partial(\ell(\alpha, \beta))}{\partial(\alpha)} = \frac{x_1 + x_2 - n\alpha}{\alpha \cdot (1-\alpha)} \overset{\text{set}}{=} 0 \Rightarrow \widehat{\alpha} = \frac{x_1 + x_2}{n}$$

---

Part C

*Derive the asymptotic covariance matrix of the estimators derived above.*

---

**Solution:**

the score function is

$$\mathcal{S}(\alpha, \beta) = \begin{pmatrix} x_1 \cdot \frac{1}{\beta} - x_2 \cdot \frac{1}{1-\beta} \\ \frac{x_1 + x_2 - n\alpha}{\alpha \cdot (1-\alpha)} \end{pmatrix} \tag{5}$$

A few derivatives later, the hessian matrix is given by

$$\mathcal{H}(\alpha, \beta) = \begin{pmatrix} -\frac{x_2}{(1-\beta)^2} - \frac{x_1}{\beta^2} & 0 \\ 0 & \frac{(2\alpha - 1) \cdot (x_1 + x_2 - n \cdot \alpha)}{\alpha^2 \cdot (1-\alpha^2)} - \frac{n}{\alpha \cdot (1-\alpha)} \end{pmatrix}$$

Recall The expectations of $X_1$ and $X_2$ are given by

- $\mathbb{E}[X_1] = n \cdot \mathbb{P}(Y_{i1} = 1, Y_{i2} = 1) = n \cdot \alpha\beta$
- $\mathbb{E}[X_2] = n \cdot \mathbb{P}(Y_{i1} = 1, Y_{i2} = 0) = n\alpha \cdot (1-\beta)$

So the fisher information matrix is

$$\mathcal{I}(\alpha, \beta) = -\mathbb{E}\left[\mathcal{H}(\alpha, \beta)\right] = \begin{pmatrix} \dfrac{n \cdot \alpha}{\beta \cdot (1-\beta)} & 0 \\ 0 & \dfrac{n}{\alpha \cdot (1-\alpha)} \end{pmatrix} \tag{6}$$

Now by MLE theory,

$$\sqrt{n} \cdot \left[\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right] \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \cdot (1-\alpha) & 0 \\ 0 & \dfrac{\beta \cdot (1-\beta)}{\alpha} \end{pmatrix}\right)$$

And the asymptotic covariance matrix for $(\widehat{\alpha}, \widehat{\beta})^T$ is the covariance of the multivariate normal.

---

Part D

*Does there exist a UMP test for testing*

$$H_0 : \beta = 0.5 \ \textit{versus} \ H_1 : \beta > 0.5?$$

*If so, then please find it. If not, then explain why such a test does not exist.*

---

**Solution:**

No. We can only derive UMPU tests for the two parameter exponential family.

---

Part E

*Derive the likelihood-ratio test statistic for testing*

$$H_0 : \alpha - \beta = 0 \ \textit{versus} \ H_1 : \alpha - \beta \neq 0.$$

---

**Solution:**

The likelihood ratio test is based on

$$\Lambda = \frac{\sup_{\alpha=\beta} \mathcal{L}(\alpha, \beta)}{\sup_{\alpha\neq\beta} \mathcal{L}(\alpha, \beta)}$$

Consider 3 when $\alpha = \beta \stackrel{\text{set}}{=} \alpha_0$:

$$\ell(\alpha_0) \propto x_1 \log\left(\frac{\alpha_0^2}{1-\alpha_0}\right) + x_2 \log(\alpha_0) - n \log\left(\frac{1}{1-\alpha_0}\right)$$

Taking the first derivative and maximizing this log-likelihood gives

$$\dot{\ell}(\alpha_0) = \frac{2x_1}{\alpha_0} + \frac{x_1}{1-\alpha_0} + \frac{x_2}{\alpha_0} - \frac{n}{1-\alpha_0} \stackrel{\text{set}}{=} 0 \Longrightarrow \widehat{\alpha}_0 = \frac{2x_1 + x_2}{2x_1 + 2x_2 + x_3}$$

So the likelihood ratio statistic is

$$\Lambda = \frac{\left(\widehat{\alpha}_0^2\right)^{x_1} \left(\widehat{\alpha}_0 \cdot (1-\widehat{\alpha}_0)\right)^{x_2} \left(1-\widehat{\alpha}_0\right)^{x_3}}{\left(\widehat{\alpha} \cdot \widehat{\beta}\right)^{x_1} \left(\widehat{\alpha} \cdot (1-\widehat{\beta})\right)^{x_2} \left(1-\widehat{\alpha}\right)^{x_3}}$$

Then the likelihood ratio test rejects $H_0$ when $-2\log(\Lambda) > \chi^2_{1,1-\tilde{\alpha}}$, where $\chi^2_{1,1-\tilde{\alpha}}$ is the $(1-\tilde{\alpha})$ quantile of a chi-square distribution with 1 degree of freedom, and $\tilde{\alpha}$ is the desired level of the test.

---

## Part F

*Derive the score test for the hypotheses in part (e).*

---

**Solution:**

The score test is given by

$$SC_n = \left[\dot{\ell}(\alpha, \beta)\right]^T \cdot \left[\mathfrak{I}(\alpha, \beta)\right]^{-1} \cdot \left[\dot{\ell}(\alpha, \beta)\right] \Big|_{\alpha=\beta=\widehat{\alpha}_0} \xrightarrow{\mathcal{D}} \chi_1^2$$

where

- $\dot{\ell}(\alpha, \beta)$ is the score function given by equation 5.
- $\mathfrak{I}(\alpha, \beta)$ is the fisher information matrix given by equation 6

We will reject $H_0$ when $SC_n > \chi^2_{1,1-\tilde{\alpha}}$

---

## Part G

*Derive the Wald test statistic for the hypotheses in part (e).*

---

**Solution:**

The wald test is given by

$$W_n = \left[\mathbf{C}\widehat{\theta} - 0\right]^T \left[\frac{\mathbf{C} \cdot \mathfrak{I}(\theta)^{-1} \mathbf{C}^T}{n}\right]^{-1}_{\theta=\widehat{\theta}} \left[\mathbf{C}\widehat{\theta} - 0\right] \xrightarrow{\mathcal{D}} \chi_1^2$$

where

- $\widehat{\theta} = \begin{pmatrix} \widehat{\alpha} & \widehat{\beta} \end{pmatrix}$
- $\mathbf{C} = \begin{pmatrix} 1 & -1 \end{pmatrix}$
- $\mathfrak{I}(\theta)$ is the fisher information matrix from equation 6.

We will reject $H_0$ when $W_n > \chi^2_{1,1-\tilde{\alpha}}$

---

Part H

*Now, suppose we are interested in inference about $\beta$ only, while considering $\alpha$ as a nuisance parameter. Derive a conditional likelihood for $\beta$ which does not depend on $\alpha$. Compute the maximum likelihood estimator for $\beta$ and compare with the estimator for $\beta$ in part (b). Is the result intuitive?*

---

**Solution:**

First we need a complete sufficient statistic for $\alpha$. We can write the log-likelihood as

$$\ell(\alpha,\beta) \propto x_1 \log\left(\frac{\alpha\beta}{1-\alpha}\right) + x_2 \log\left(\frac{\alpha\cdot(1-\beta)}{1-\alpha}\right) - n\log\left(\frac{1}{1-\alpha}\right)$$

$$= (x_1+x_2)\log(\alpha) + x_1\log(\beta) - (x_1+x_2)\log(1-\alpha) + x_2\log(1-\beta) + n\log(1-\alpha)$$

$$= (x_1+x_2)\cdot\log\left(\frac{\alpha}{1-\alpha}\right) + x_1\cdot\log\left(\frac{\beta}{1-\beta}\right) + (x_1+x_2)\log(1-\beta) + n\log(1-\alpha)$$

So we want the conditional likelihood of $X_1 \mid (X_1+X_2=t)$. Note that $X_1+X_2 \sim \text{Binomial}\ (n,\ \alpha)$

$$\mathbb{P}(X_1 = x_1 \mid X_1+X_2=t) = \frac{\mathbb{P}(X_1=x_1, X_2=t-x_1)}{\mathbb{P}(X_1+X_2=t)}$$

$$= \frac{\binom{n}{x_1,\, t-x_1,\, n-t}(\alpha\cdot\beta)^{x_1}(\alpha\cdot(1-\beta))^{t-x_1}(1-\alpha)^{n-t}}{\binom{n}{t}\alpha^t\cdot(1-\alpha)^{n-t}}$$

$$= \frac{t!}{x_1!\,(t-x_1)!}(\beta)^{x_1}(1-\beta)^{t-x_1}$$

$$\sim \text{Binomial}\ (t,\beta)$$

$$\Rightarrow \widehat{\beta}_{\text{cond}} = \frac{x_1}{t} = \frac{x_1}{x_1+x_2}$$

This is the same estimator from part (b). The result is intuitive because $\beta$ itself is a conditional probability, which means inference from the joint and corresponding conditional likelihoods should be equivalent.

---

# 2 PROBLEM # 3

*Consider independent observations* $y_1, \ldots, y_n$*, where* $y_i = (y_{i1}, y_{i2})^T$ *is a bivariate binary random vector such that* $y_{ij}$ *takes values 0 and 1 for* $j = 1, 2$*. Suppose that* $y_i \sim \mathcal{QE}(\theta, \lambda)$ *where* $\mathcal{QE}(\theta, \lambda)$ *is a bivariate binary distribution of quadratic exponential form*

$$p(y_i \mid \theta, \lambda) = \Delta(\theta, \lambda)^{-1} exp \left\{ y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1}, y_{i2}) \right\}$$

*where* $\Delta(\theta, \lambda)$ *is a normalizing constant and* $C(y_{i1}, y_{i2})$ *if a shape function independent of* $\theta = (\theta_1, \theta_2)^T$ *and* $\lambda$

---

## 2.1 Part A

*Derive both the marginal distribution of* $y_{i1}$ *and the conditional distribution of* $y_{i2}$ *given* $y_{i1}$*. Specify a sufficient and necessary condition such that* $y_{i1}$ *and* $y_{i2}$ *are independent.*

---

**Solution:**

the marginal distribution is given by

$$\begin{aligned} p(y_{i1}) &= \sum_{k=0}^{1} p(y_{i1}, y_{i2} = k) \\ &= \Delta(\theta, \lambda)^{-1} \Big( \exp\left\{ y_{i1}\theta_1 - C(y_{i1}, 0) \right\} + \exp\left\{ y_{i1}\theta_1 + \theta_2 + y_{i1}\lambda - C(y_{i1}, 1) \right\} \Big) \end{aligned}$$

and the conditional distribution is,

$$\begin{aligned} p(y_{i2} \mid y_{i1} = t) &= \frac{p(y_{i2}, y_{i1} = t)}{p(y_{i1} = t)} \\ &= \frac{\exp\left\{ t \cdot \theta_1 + y_{i2} \cdot \theta_2 + y_{i2} t\lambda - C(t, y_{i2}) \right\}}{\exp\left\{ t \cdot \theta_1 - C(t, 0) \right\} + \exp\left\{ t \cdot \theta_1 + \theta_2 + y_{i1}\lambda - C(y_{i1}, 1) \right\}} \\ &= \frac{\exp\left\{ y_{i2} \cdot \theta_2 + y_{i2} t\lambda - C(t, y_{i2}) \right\}}{\exp\left\{ -C(t, 0) \right\} + \exp\left\{ \theta_2 + y_{i1}\lambda - C(y_{i1}, 1) \right\}} \end{aligned}$$

We know that $y_{i1} \perp\!\!\!\perp y_{i2} \iff p(y_{i1}, y_{i2}) = p(y_{i1})p(y_{i2})$. We'll need to use the marginal distribution of $y_{i2}$, which is

$$\begin{aligned} p(y_{i2}) &= \sum_{k=0}^{1} p(y_{i1} = k, y_{i2}) \\ &= \Delta(\theta, \lambda)^{-1} \Big( \exp\left\{ y_{i2}\theta_2 - C(0, y_{i2}) \right\} + \exp\left\{ \theta_1 + y_{i2}\theta_2 + y_{i2}\lambda - C(1, y_{i2}) \right\} \Big) \end{aligned}$$

Ignoring the normalizing constant , we have

$$
\begin{aligned}
y_{i1} \perp\!\!\!\perp y_{i2} &\iff p(y_{i1}, y_{i2}) = p(y_{i1})p(y_{i2}) \\
&\iff \exp\{y_{i1}y_{i2}\lambda - C(y_{i1}, y_{i2})\} \\
&\quad = \Big[\exp\{-C(y_{i1}, 0)\} + \exp\{\theta_2 + y_{i1}\lambda - C(y_{i1}, 1)\}\Big] \\
&\quad \cdot \Big[\exp\{-C(0, y_{i2})\} + \exp\{\theta_1 + y_{i2}\lambda - C(1, y_{i2})\}\Big] \\
&\iff \lambda = 0 \text{ and } -C(y_{i1}, y_{i2}) = h_1(y_{i1} \mid \theta) + h_2(y_{i2} \mid \theta)
\end{aligned}
$$

where

- $h_1(y_{i1} \mid \theta) = \log\left(\exp\{-C(y_{i1}, 0)\} + \exp\{\theta_2 - C(y_{i1}, 1)\}\right)$
- $h_2(y_{i2} \mid \theta) = \log\left(\exp\{-C(0, y_{i2})\} + \exp\{\theta_1 - C(1, y_{i2})\}\right)$

---

## 2.2 Part B

*Calculate*

- *The marginal mean of* $y_i$*, denoted by* $\mu = (\mu_1, \mu_2)' = \mathbb{E}[y_i]$
- *The marginal product moment of* $y_{i1}y_{i2}$*, denoted by* $\eta_{12} = \mathbb{E}[y_{i1}y_{i2}]$
- *The marginal product centered moment of* $(y_{i1} - \mu_1)(y_{i2} - \mu_2)$*, denoted by*

$$\sigma_{12} = \mathbb{E}[(y_{i1} - \mu_1)(y_{i2} - \mu_2)]$$

---

**Solution:**

$$
\begin{aligned}
\mu_1 &= \mathbb{E}[y_{i1}] = \mathbb{P}(y_{i1} = 1) \\
&= \Delta(\theta, \lambda)^{-1} \exp\{\theta_1\} \Big[\exp\{-C(1, 0)\} + \exp\{\theta_2 + \lambda - C(1, 1)\}\Big]
\end{aligned}
$$

$$
\begin{aligned}
\mu_2 &= \mathbb{E}[y_{i2}] = \mathbb{P}(y_{i2} = 1) \\
&= \Delta(\theta, \lambda)^{-1} \exp\{\theta_2\} \Big[\exp\{-C(0, 1)\} + \exp\{\theta_1 + \lambda - C(1, 1)\}\Big]
\end{aligned}
$$

$$\eta_{12} = \mathbb{E}\left[y_{i1} \cdot y_{i2}\right] = \mathbb{P}\left(y_{i1} = 1, y_{i2} = 1\right)$$

$$= \Delta(\theta, \lambda)^{-1} \exp\left\{\theta_1 + \theta_2 + \lambda - C(1,1)\right\}$$

$$\sigma_{12} = \mathbb{E}\left[(y_{i1} - \mu_1)(y_{i2} - \mu_2)\right] = \mathbb{E}\left[y_{i1} \cdot y_{i2}\right] - \mathbb{E}\left[y_{i1}\right] \cdot \mathbb{E}\left[y_{i2}\right]$$

$$= \eta_{12} - \mu_1 \cdot \mu_2$$

$$= \frac{e^{\theta_1 + \theta_2 + \lambda - C(1,1)}}{\Delta(\theta, \lambda)} - \frac{\left[e^{\theta_1 - C(1,0)} + e^{\theta_1 + \theta_2 + \lambda - C(1,1)}\right] \cdot \left[e^{\theta_2 - C(0,1)} + e^{\theta_1 + \theta_2 + \lambda - C(1,1)}\right]}{\Delta(\theta, \lambda)^2}$$

---

## 2.3 Part C

*Calculate the Jacobian of the transformation from the canonical parameters $\theta$ and $\lambda$ to the marginal parameters $\mu$ and $\eta_{12}$, denoted by $\mathbf{V} = \partial(\theta, \lambda)/\partial(\mu, \eta_{12})$. Use $\mathbf{V}^{-1}$ to characterize the covariance matrix of $(y_i', \; y_{i1} \cdot y_{i2})'$ and specify a sufficient and necessary condition such that this transformation is one to one*

---

**Solution:**

This solution will get very detailed. It is important to write out the big ideas for a question like this one before we get into the messy parts. Here's the plan :

1. We need to derive the inverse function $h(\theta, \lambda) = (\mu, \eta_{12})$.
2. Next, we find $\mathbf{V}^{-1} = \dfrac{\partial\left(\mu, \eta_{12}\right)}{\partial\left(\theta, \lambda\right)}$. This is a straightforward procedure, but the calculation is tedious.
3. Next, note that the likelihood function for $\theta, \lambda$ can be expressed as

$$p(y_i \mid \theta, \lambda) = \exp\left\{ \begin{pmatrix} y_{i1} & y_{i2} & y_{i1} \cdot y_{i2} \end{pmatrix} \cdot \begin{pmatrix} \theta_1 \\ \theta_2 \\ \lambda \end{pmatrix} - \log\left(\Delta(\theta, \lambda)\right) - C(y_{i1}, y_{i2}) \right\}$$

Familiar theory on the exponential family tells us that the covariance of the sufficient statistic $\mathbf{y} = \begin{pmatrix} y_{i1} & y_{i2} & y_{i1} \cdot y_{i2} \end{pmatrix}$ is found by calculating

$$\ddot{b}(\theta, \lambda) = \frac{\partial^2 \log\left(\Delta\left(\theta, \lambda\right)\right)}{\partial\left(\theta, \lambda\right) \partial\left(\theta, \lambda\right)'}$$

Addressing the points in order, we can first write

$$\sum_{y_{i1}=0}^{1}\sum_{y_{i2}=0}^{1} p(y_i \mid \theta, \lambda) = 1 \Rightarrow \Delta(\theta, \lambda) = \pi_{00} + \pi_{10} + \pi_{01} + \pi_{11}$$

where

- $\pi_{00} = \exp\{-C(0,0)\}$
- $\pi_{01} = \exp\{\theta_2 - C(0,1)\}$
- $\pi_{10} = \exp\{\theta_1 - C(1,0)\}$
- $\pi_{11} = \exp\{\theta_1 + \theta_2 + \lambda - C(1,1)\}$

This notation will help us write the inverse link functions in a more concise way. For example, recall from part (b) that

$$\mu_1 = \frac{\exp\{\theta_1 - C(1,0)\} + \exp\{\theta_1 + \theta_2 + \lambda - C(1,1)\}}{\Delta(\theta, \lambda)} = \frac{\pi_{10} + \pi_{11}}{\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11}}$$

Similarly,

$$\mu_2 = \frac{\pi_{01} + \pi_{11}}{\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11}}, \quad \text{and} \quad \eta_{12} = \frac{\pi_{11}}{\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11}}$$

This means

$$\mathbf{V}^{-1} = \frac{\partial(\mu_1, \mu_2, \eta_{12})}{\partial(\theta, \lambda)} = \frac{\partial(\mu_1, \mu_2, \eta_{12})}{\partial(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})} \frac{\partial(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})}{\partial(\theta, \lambda)}$$

This is the tedious part. Taking lots of derivatives gives $\dfrac{\partial(\mu_1, \mu_2, \eta_{12})}{\partial(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})} =$

$$\frac{1}{\Delta(\theta, \lambda)^2} \begin{pmatrix} -(\pi_{11} + \pi_{10}) & -(\pi_{11} + \pi_{10}) & \pi_{00} + \pi_{01} & \pi_{00} + \pi_{01} \\ -(\pi_{11} + \pi_{01}) & \pi_{00} + \pi_{10} & -(\pi_{11} + \pi_{01}) & \pi_{00} + \pi_{10} \\ -\pi_{11} & -\pi_{11} & -\pi_{11} & \pi_{00} + \pi_{01} + \pi_{10} \end{pmatrix}$$

And referring to the definitions given above,

$$\frac{\partial(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})}{\partial(\theta_1, \theta_2, \lambda)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \pi_{01} & 0 \\ \pi_{10} & 0 & 0 \\ \pi_{11} & \pi_{11} & \pi_{11} \end{pmatrix}$$

Taking the desired product, we have

$$\mathbf{V}^{-1} = \frac{1}{\Delta(\theta, \lambda)^2} \begin{pmatrix} (\pi_{00} + \pi_{01})(\pi_{10} + \pi_{11}) & \pi_{00}\pi_{11} - \pi_{01}\pi_{10} & \pi_{11}(\pi_{01} + \pi_{00}) \\ \pi_{00}\pi_{11} - \pi_{01}\pi_{10} & (\pi_{10} + \pi_{00})(\pi_{01} + \pi_{11}) & \pi_{11}(\pi_{10} + \pi_{00}) \\ \pi_{11}(\pi_{00} + \pi_{01}) & \pi_{11}(\pi_{00} + \pi_{10}) & \pi_{11}(\pi_{00} + \pi_{01} + \pi_{10}) \end{pmatrix}$$

Which, applying the inverse functions, is equivalent to

$$\begin{pmatrix} \mu_1(1-\mu_1) & \eta_{12}-\mu_1\mu_2 & \eta_{12}(1-\mu_1) \\ \eta_{12}-\mu_1\mu_2 & (1-\mu_2)\mu_2 & \eta_{12}(1-\mu_2) \\ \eta_{12}(1-\mu_1) & \eta_{12}(1-\mu_2) & \eta_{12}(1-\eta_{12}) \end{pmatrix}$$

These are straightforward except for one. Here are the details for that one:

$$\begin{aligned}
\frac{\pi_{00}\pi_{11}-\pi_{01}\pi_{10}}{\Delta(\theta,\lambda)^2} &= \frac{\pi_{00}\pi_{11}+\pi_{01}\pi_{11}-\pi_{01}\pi_{10}-\pi_{01}\pi_{11}}{\Delta(\theta,\lambda)^2} \\
&= \frac{\pi_{11}(\pi_{00}+\pi_{01})-\pi_{01}(\pi_{11}+\pi_{10})}{\Delta(\theta,\lambda)^2} \\
&= \frac{\pi_{11}(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11})-\pi_{01}(\pi_{11}+\pi_{10})-\pi_{11}(\pi_{11}+\pi_{01})}{\Delta(\theta,\lambda)^2} \\
&= \frac{\pi_{11}}{\Delta(\theta,\lambda)} - \left(\frac{\pi_{10}+\pi_{11}}{\Delta(\theta,\lambda)}\right)\cdot\left(\frac{\pi_{01}+\pi_{11}}{\Delta(\theta,\lambda)}\right) \\
&= \eta_{12}-\mu_1\mu_2
\end{aligned}$$

Now we can find the covariance matrix of $(y_i', y_{i1}\cdot y_{i2})$ using the properties of the exponential family. Taking the first derivative,

$$\frac{\partial \log(\Delta(\theta,\lambda))}{\partial(\theta,\lambda)} = \begin{pmatrix} \frac{\pi_{11}+\pi_{10}}{\Delta(\theta,\lambda)} & \frac{\pi_{11}+\pi_{01}}{\Delta(\theta,\lambda)} & \frac{\pi_{11}}{\Delta(\theta,\lambda)} \end{pmatrix} = \begin{pmatrix} \mu_1 & \mu_2 & \eta_{12} \end{pmatrix}$$

And some more:

$$\frac{\partial(\mu_1)}{\partial(\theta_1)} = \frac{\Delta(\theta,\lambda)(\pi_{10}+\pi_{11})-(\pi_{10}+\pi_{11})^2}{\Delta(\theta,\lambda)^2} = \frac{(\pi_{10}+\pi_{11})(\pi_{01}+\pi_{00})}{\Delta(\theta,\lambda)^2} = \mu_1(1-\mu_1)$$

$$\frac{\partial(\mu_1)}{\partial(\theta_2)} = \frac{\Delta(\theta,\lambda)(\pi_{11})-(\pi_{10}+\pi_{11})(\pi_{11}+\pi_{01})}{\Delta(\theta,\lambda)^2} = \eta_{12}-\mu_1\mu_2$$

$$\frac{\partial(\mu_1)}{\partial(\lambda)} = \frac{\Delta(\theta,\lambda)(\pi_{11})-(\pi_{10}+\pi_{11})\pi_{11}}{\Delta(\theta,\lambda)^2} = \frac{\pi_{11}}{\Delta(\theta,\lambda)}\cdot\frac{\pi_{00}+\pi_{01}}{\Delta(\theta,\lambda)} = \eta_{12}(1-\mu_1)$$

$$\frac{\partial(\mu_2)}{\partial(\theta_1)} = \frac{\Delta(\theta,\lambda)(\pi_{11})-(\pi_{01}+\pi_{11})(\pi_{10}+\pi_{11})}{\Delta(\theta,\lambda)^2} = \eta_{12}-\mu_1\mu_2$$

$$\frac{\partial(\mu_2)}{\partial(\theta_2)} = \frac{\Delta(\theta,\lambda)(\pi_{01}+\pi_{11})-(\pi_{01}+\pi_{11})(\pi_{01}+\pi_{11})}{\Delta(\theta,\lambda)^2} = \mu_2(1-\mu_2)$$

$$\frac{\partial(\mu_2)}{\partial(\lambda)} = \frac{\Delta(\theta,\lambda)(\pi_{11})-(\pi_{01}+\pi_{11})\pi_{11}}{\Delta(\theta,\lambda)^2} = \frac{\pi_{11}}{\Delta(\theta,\lambda)}\cdot\frac{\pi_{00}+\pi_{10}}{\Delta(\theta,\lambda)} = \eta_{12}(1-\mu_2)$$

$$\frac{\partial(\eta_{12})}{\partial(\theta_1)} = \frac{\Delta(\theta,\lambda)(\pi_{11}) - \pi_{11}(\pi_{11}+\pi_{10})}{\Delta(\theta,\lambda)^2} = \eta_{12}(1-\mu_1)$$

$$\frac{\partial(\eta_{12})}{\partial(\theta_2)} = \frac{\Delta(\theta,\lambda)(\pi_{11}) - \pi_{11}(\pi_{11}+\pi_{01})}{\Delta(\theta,\lambda)^2} = \eta_{12}(1-\mu_2)$$

$$\frac{\partial(\eta_{12})}{\partial(\lambda)} = \frac{\Delta(\theta,\lambda)(\pi_{11}) - \pi_{11}\pi_{11}}{\Delta(\theta,\lambda)^2} = \frac{\pi_{11}}{\Delta(\theta,\lambda)} - \left(\frac{\pi_{11}}{\Delta(\theta,\lambda)}\right)^2 = \eta_{12}(1-\eta_{12})$$

Now we can put all the derivatives together to see that $\mathcal{V}\left(y_i', y_{i1}\cdot y_{i2}\right) = \mathbf{V}^{-1}$, the inverse of the Jacobian transformation from canonical to marginal parameters. This means the inverse Jacobian matrix has the same properties as a covariance matrix; namely, it is positive semi-definite. Therefore, the transformation is one to one iff $\mathbf{V}^{-1}$ is non-singular.

---

## 2.4 Part D

*Suppose that we also observe a $p\times 1$ column vector $x_i$ for each $i$ and that conditionally on $x_i$, $y_i \sim \mathcal{QE}(\theta_i, \lambda_i)$, where $\theta_i = (\theta_{i1}, \theta_{i2})$ and $\lambda_i$ may depend on $x_i$, for $i = 1, \ldots, n$. Consider the model*

$$\mathbb{E}\left[y_i \mid x_i\right] = \mu_i = (\mu_{i1}, \mu_{i2})' = \mu(x_i, \beta)$$

*and*

$$\mathbb{E}\left[(y_{i1}-\mu_{i1})(y_{i2}-\mu_{i2})\right] = \sigma_{i12} = \sigma_{12}(x_i, \beta, \alpha)$$

*where $\beta$ is an unknown $p\times 1$ regression parameter and $\alpha$ is an unknown scalar parameter. Derive the likelihood score equations for $(\alpha, \beta)'$ and simplify them using the result obtained in part (c). Please clarify whether such estimating equations explicitly involve $C(y_{i1}, y_{i2})$*

---

**Solution:**

It is helpful to remember the flow of GLMs here. The canonical parameter, $\theta_i$, is a function of the mean $\mu_i$. The mean vector, $\mu_i$, is a function of the linear predictor, which in turn is determined by $x_i, \beta$. Furthermore, we saw in part (b) that $\sigma_{i12}$ is a function of $\mu_i$ and $\mathbb{E}\left[y_{i1}y_{i2}\right] = \eta_{i12} = \eta_{12}(x_i, \beta, \lambda)$. Thus, to model the mean and covariance of the two binary outcomes, we write $\mathbf{Y}_i = (y_{i1}, y_{i2}, y_{i1}\cdot y_{i2})$ and $\boldsymbol{\mu}_i^* = (\mu_{i1}, \mu_{i2}, \eta_{i12})'$. In part (c), we showed that

$$\mathcal{V}(\mathbf{Y}_i) = \mathbf{V}^{-1} = \begin{pmatrix} \mu_{i1}(1-\mu_{i1}) & \eta_{12}-\mu_{i1}\mu_{i2} & \eta_{12}(1-\mu_{i1}) \\ \eta_{12}-\mu_{i1}\mu_{i2} & (1-\mu_{i2})\mu_{i2} & \eta_{12}(1-\mu_{i2}) \\ \eta_{12}(1-\mu_{i1}) & \eta_{12}(1-\mu_{i2}) & \eta_{12}(1-\eta_{12}) \end{pmatrix}$$

Now recall that the score function for exponential families takes the form

$$\mathcal{S}(\alpha,\beta) = \sum_{i=1}^{n} (\mathbf{Y}_i - \boldsymbol{\mu}_i^*)\, \mathbf{V}_i^{-1} \frac{\partial(\mu_i^*)}{\partial(\alpha,\beta)}$$

where

$$\frac{\partial\left(\mu_i^*\right)}{\partial\left(\alpha,\beta\right)} = \begin{pmatrix} \frac{\partial\left(\mu_{i1}\right)}{\partial\left(\beta\right)} & \frac{\partial\left(\mu_{i1}\right)}{\partial\left(\alpha\right)} \\ \frac{\partial\left(\mu_{i2}\right)}{\partial\left(\beta\right)} & \frac{\partial\left(\mu_{i2}\right)}{\partial\left(\alpha\right)} \\ \frac{\partial\left(\eta_{i12}\right)}{\partial\left(\beta\right)} & \frac{\partial\left(\eta_{i12}\right)}{\partial\left(\alpha\right)} \end{pmatrix} = \begin{pmatrix} \frac{\partial\left(\mu_{i1}\right)}{\partial\left(\beta\right)} & 0 \\ \frac{\partial\left(\mu_{i2}\right)}{\partial\left(\beta\right)} & 0 \\ \frac{\partial\left(\eta_{i12}\right)}{\partial\left(\beta\right)} & \frac{\partial\left(\eta_{i12}\right)}{\partial\left(\alpha\right)} \end{pmatrix}$$

Note that $\mu_i$ is determined by $x_i, \beta$ and $\sigma_{i12} = \eta_{i12} + \mu_{i1}\mu_{i2}$. Thus the score function does not depend on $C(\cdot,\cdot)$

---

## 2.5 Part E

*Consider the generalized estimating equations given by*

$$\sum_{i=1}^{n} \frac{\partial\left(\mu_i,\sigma_{i12}\right)}{\partial\left(\alpha,\beta\right)} \cdot \frac{\partial\ell(y_i \mid \theta_i,\lambda_i)}{\partial(\theta,\lambda_i)} = 0$$

*Compare the estimate of $(\alpha,\beta')'$ in part (d) with that in part (e) in terms of the statistical efficiency. To do so, provide an explicit comparison of the asymptotic variances of these estimators.*

---

**Solution:**

$$\sum_{i=1}^{n} \frac{\partial\left(\mu_i,\sigma_{i12}\right)}{\partial\left(\alpha,\beta\right)} \cdot \frac{\partial\ell(y_i \mid \theta_i,\lambda_i)}{\partial(\theta,\lambda_i)} = \sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)$$

---

## 2.6 Part F

*Will the results in parts (a)-(e) be changed if $y_{i1}$ and $y_{i2}$ are continuous variables instead of binary variables? Please explain. If so, then derive the corresponding results and compare with those obtained above.*

---

**Solution:**

a. If $y_{i1}$ and $y_{i2}$ are continuous,

$$p(y_{i1} \mid \theta, \lambda) = \int p(y_{i1}, y_{i2} \mid \theta, \lambda) dy_{i2}$$

and

$$p(y_{i2} \mid y_{i1}) = \frac{p(y_{i1}, y_{i2})}{p(y_{i1})}$$

and the independence condition is still given as

$$y_{i1} \perp\!\!\!\perp y_{i2} \iff p(y_{i1}, y_{i2}) = p(y_{i1})p(y_{i2})$$

b Now there are no simple forms for $\mu_1$, $\mu_2$ and $\eta_{12}$

c We now have

$$\Delta(\theta, \lambda) = \int \exp\left\{y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1}, y_{i2})\right\} dy_{i1} dy_{i2}$$

So

$$\frac{\partial}{\partial\theta_1} \log\left(\Delta(\theta, \lambda)\right) = \frac{1}{\Delta(\theta, \lambda)} \int y_{i1} \cdot e^{y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1}, y_{i2})} dy_{i1} dy_{i2} = \mu_{i1}$$

Similarly, $\partial_{\theta_2}\left(\log\left(\Delta(\theta, \lambda)\right)\right) = \mu_{i2}$ and $\partial_{\lambda}\left(\log\left(\Delta(\theta, \lambda)\right)\right) = \eta_{i12}$. So the results in part c should not be affected.

d These results are based on exponential families, so they should not change either.

---