Quasi-likelihood

Mingwei Fei

January 30, 2023

1 Quasi function

Suppose that $y_1, ..., y_n$ are independent where $E(y_i) = \mu_i, Var(y_i) = \mu_i^2$ and assume that $\mu_i = exp(x_i^T \beta)$. Derive the quasi log-likelihood and quasi score function of β.

The quasi log-likelihood function

$$\begin{aligned}
I_q(\mu) &= \sum_{i=1}^{n} lq(\mu_i) \\
l_q(\mu_i) &= \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 V_i(t)} dt \\
Var(y_i) &= \mu_i^2, \qquad \sigma^2 = 1, V_i(\mu_i) = \mu_i^2 \\
V_i(t) &= \mu_i^2, \qquad \frac{y_i - t}{\sigma^2 V_i(t)} = \frac{y_i - t}{t^2}
\end{aligned}$$

By integrating of t

$$\begin{aligned}
lq(\mu_i) &= \int_{y_i}^{\mu_i} \frac{y_i - t}{t^2} dt \\
&= -\frac{y_i}{\mu_i} + 1 - log\frac{\mu_i}{y_i} = 1 - \frac{y_i}{\mu_i} + log\frac{y_i}{\mu_i}
\end{aligned}$$

Then the quasi log-likelihood

$$\begin{aligned}
I_q(\mu) &= \sum_{i=1}^{n} 1 - \frac{y_i}{\mu_i} + log\frac{y_i}{\mu_i} \\
I_q(\beta) &= \sum_{i=1}^{n} 1 - \frac{y_i}{exp(x_i^T \beta)} + log\frac{y_i}{exp(x_i^T \beta)} \\
&= \sum_{i=1}^{n} 1 - \frac{y_i}{exp(x_i^T \beta)} + log y_i - (x_i^T \beta)
\end{aligned}$$

The quasi score function

$$Q(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{y_i - exp(x_i^T \beta)}{exp(2x_i^T \beta)}$$
$$\frac{\partial \mu_i}{\partial \beta} = exp(x_i^T \beta) x_{ij}$$
$$V(\beta) = diag\{exp(2x_i^T \beta)\}$$
$$Q(\beta) = \sum_{i=1}^{n} \frac{y_i - exp(x_i^T \beta)}{exp(x_i^T \beta)} x_i = D^T V(\beta)^{-1}(Y - \mu)$$

The D_i could be considered as the product of the diagonal matrix of $diag\{exp(x_i^T \beta)\}$ and x_i

$$\mu_i = exp(X_i^T \beta) = exp(x_{i1}\beta_1 + x_{i2}\beta_2 + .. + x_{ip}\beta_p)$$
$$\frac{\partial \mu_i}{\partial \beta_j} = exp(x_i^T \beta) x_{ij}$$
$$D_i(\beta) = \frac{\partial \mu_i}{\partial \beta} = \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta_1} \\ \frac{\partial \mu_i}{\partial \beta_2} \\ .. \\ \frac{\partial \mu_i}{\partial \beta_p} \end{pmatrix}$$
$$= exp(x_i^T \beta) \begin{pmatrix} x_{i1} \\ x_{i2} \\ .. \\ x_{ip} \end{pmatrix} = \begin{pmatrix} exp(x_i^T \beta) x_{i1} \\ exp(x_i^T \beta) x_{i2} \\ .. \\ exp(x_i^T \beta) x_{ip} \end{pmatrix}$$
$$D(\beta) = \begin{pmatrix} exp(x_1^T \beta) & 0 & 0... \\ 0 & exp(x_2^T \beta) & 0.. \\ . & .. & .. \\ .. & .. & exp(x_n^T \beta) \end{pmatrix} \begin{pmatrix} x_1^T \\ x_2^T \\ .. \\ x_n^T \end{pmatrix}$$
$$= \begin{pmatrix} exp(x_1^T \beta) & 0 & 0... \\ 0 & exp(x_2^T \beta) & 0.. \\ . & .. & .. \\ .. & .. & exp(x_n^T \beta) \end{pmatrix} \begin{pmatrix} x_{11} & x_{12}... & x_{1p} \\ x_{21} & x_{22}... & x_{2p} \\ .. & .. & .. \\ x_{n1} & x_{n2}... & x_{np} \end{pmatrix}$$

Then the formula

$$D^T V(\beta)^{-1} e = \begin{pmatrix} x_{11} & x_{21}... & x_{n1} \\ x_{12} & x_{22}... & x_{n2} \\ .. & .. & .. \\ x_{1p} & x_{2p}... & x_{np} \end{pmatrix} \begin{pmatrix} exp(x_1^T\beta) & 0 & 0... \\ 0 & exp(x_2^T\beta) & 0.. \\ . & .. & .. \\ .. & .. & exp(x_n^T\beta) \end{pmatrix}$$

$$\begin{pmatrix} exp(2x_1^T\beta)^{-1} & 0 & 0... \\ 0 & exp(2x_2^T\beta)^{-1} & 0.. \\ . & .. & .. \\ .. & .. & exp(2x_n^T\beta)^{-1} \end{pmatrix} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_1 \\ .. \\ y_n - \mu_1 \end{pmatrix}$$

$$= \begin{pmatrix} x_{11} & x_{21}... & x_{n1} \\ x_{12} & x_{22}... & x_{n2} \\ .. & .. & .. \\ x_{1p} & x_{2p}... & x_{np} \end{pmatrix} \begin{pmatrix} exp(-x_1^T\beta) & 0 & 0... \\ 0 & exp(-x_2^T\beta) & 0.. \\ . & .. & .. \\ .. & .. & exp(-x_n^T\beta) \end{pmatrix} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_1 \\ .. \\ y_n - \mu_1 \end{pmatrix}$$

$$U(\beta_j) = \sum_{i=1}^{n} x_{ij} exp(-x_i^T\beta)(y_i - exp(x_i^T\beta))$$

$$U(\beta) = \sum_{i=1}^{n} x_i exp(-x_i^T\beta)(y_i - exp(x_i^T\beta)) = \sum_{i=1}^{n} \frac{x_i y_i}{exp(x_i^T\beta)} - x_i$$

1.1 Asymptotic Distribution of β

The sandwich theorem could be used to derive the covariance of β. By Taylor expansion, we have

$$\hat{\beta} - \beta_* = -\left(\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}\right)^{-1} \frac{\partial ln(\beta_*}{\partial\beta}(1 + O(n))$$

$$\sqrt{n}(\hat{\beta} - \beta_*) = -\left(\frac{1}{n}\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}\right)^{-1} \frac{1}{\sqrt{n}}\frac{\partial ln(\beta_*}{\partial\beta}$$

β_* is the solution to the score function

$$E[\frac{\partial ln(\beta_*)}{\partial\beta}] = 0$$

While $\hat{\beta}$ is the solution to the score function

$$\frac{\partial ln(\hat{\beta})}{\partial\beta} = 0$$

By WLLN,

$$\left(\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}\right) = E[\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}] = D^T V(\beta)^{-1} D$$

So we have the covariance of β

$$
\begin{aligned}
Cov\left(\sqrt{n}(\hat{\beta})\right) &= \left(\frac{1}{n}\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}\right)^{-1}\frac{1}{n}Cov(\frac{\partial ln(\hat{\beta})}{\partial\beta})[\left(\frac{1}{n}\frac{\partial^2 ln(\beta_*)}{\partial\beta\,\partial\beta}\right)^{-1}]^T \\
Cov(\frac{\partial ln(\hat{\beta})}{\partial\beta}) &= D^T V(\beta)^{-1}V(\beta)V(\beta)^{-1}D \\
&= D^T V(\beta)^{-1}D
\end{aligned}
$$

By Sandwich theorem,

$$
\begin{aligned}
Cov\left(\sqrt{n}(\hat{\beta})\right) &= \left(\frac{1}{n}D^T V(\beta)^{-1}D\right)^{-1}\left(\frac{1}{n}D^T V(\beta)^{-1}D\right)\left(\frac{1}{n}D^T V(\beta)^{-1}D\right)^{-1} \\
&= \{D^T V(\beta)^{-1}D\}^{-1}
\end{aligned}
$$

By the above formula,

$$Cov(\hat{\beta}) = \{D^T V(\beta)^{-1}D\}^{-1} = (X^T X)^{-1}$$

Thus the asymptotic distribution

$$\sqrt{n}(\hat{\beta} - \beta_*) = N(0, n(X^T X)^{-1})$$

1.1.1 Godambe Efficiency

Suppose that X_i, Y_i are independent random variables with an exponential distribution, with $E(X_i) = 1/(\psi\lambda_i)$ and $E(Y_i) = 1/\lambda_i$, for $i = 1, 2, ..n$. The parameters of interest is ψ, the λ_i is being unknown nuisance parameters.

(a) Write log-likelihood function $ln(\psi, \lambda_1, \lambda_2, ..\lambda_n)$ based on $(X_i, Y_i), i = 1, ..n$. Derive the score function (only depends on ψ) that the maximum likelihood estimator for ψ based on ln, and denote the score equation by $S_n(\psi) = 0$.

1.2 Exercise

Consider pairs of independent random variables $(y_{i1}, y_{i2}), i = 1, , n$ such that both y_{i1} and y_{i2} follow a $N(\mu_i, \psi)$ distribution. Let ψ be the parameter of interest and the μ_i are nuisance parameters.

(a) Show that the maximum likelihood estimate of ψ is inconsistent.
The joint density of y_{i1}, y_{i2}

$$
\begin{aligned}
P(y_{i1}, y_{i2}) &= \frac{1}{2\pi\psi}exp\left(-\frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}\right) \\
P(y_1, y_2) &= \prod_{i=1}^{n}\frac{1}{(2\pi\psi)^n}exp\left(-\sum_{i=1}^{n}\frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}\right)
\end{aligned}
$$

The log-likelihood function

$$ln(y_1, y_2) = -nlog(2\pi) - nlog\psi - \sum_{i=1}^{n} \frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}$$

Obtain MLE of μ_i, ψ

$$\begin{aligned}
\partial_{\mu_i} ln &= -1/(2\psi) \sum_{i=1}^{n} -2(y_{i1} - \mu_i + y_{i2} - \mu_i) = 0, \qquad \hat{\mu_i} \\
\mu_i &= 1/2(y_{i1} + y_{i2}) \\
\partial_{\psi} ln &= -n/\psi + \frac{\sum_{i=1}^{n}[(y_{i1} - \mu_1)^2 + (y_{i2} - \mu_2)^2]}{2\psi^2} = 0 \\
\hat{\psi} &= 1/2n \left(\sum_{i=1}^{n}[(y_{i1} - \mu_1)^2 + (y_{i2} - \mu_2)^2] \right) \\
&= \frac{1}{4n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2
\end{aligned}$$

As $E(y_{i1} - y_{i2}) = 0, Var(y_{i1} - y_{i2}) = 2\psi$

$$Var(y_{i1} - y_{i2}) = E(y_{i1} - y_{i2})^2 - [E(y_{i1} - y_{i2})]^2 = 2\psi, \qquad E(y_{i1} - y_{i2})^2 = 2\psi$$

By WLLN,

$$\hat{\psi} = \frac{1}{4n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2 \xrightarrow{n \to \infty} 1/4E(y_{i1} - y_{i2})^2 = \psi/2 \neq \psi$$

So MLE of ψ is not consistent.

(b) Construct a consistent estimate for ψ based on the available information.
From part(a), we can construct $\tilde{\psi} = 2\hat{\psi} = \frac{1}{2n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2$. By WLLN, the

$$\tilde{\psi} = \frac{1}{2n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2 \xrightarrow[n \to \infty]{p} = \psi$$

(c) Assume that y_{i1} and y_{i2} follow a $N(\mu_i, \psi_i)$ distribution for $i = 1, , n$, where $\mu_i = \beta_0 + \beta_1(x_i - \bar{x})$ and $\psi_i = exp(\alpha_0 + \alpha_1(x_i - \bar{x}))$, in which x_i is a covariate of interest and $\bar{x}$ is the mean of the x_is. Derive the score test statistic for testing homogeneous variance.
The hypothesis are

$$\begin{aligned}
H_0 &: \alpha_1 = 0 \\
H_1 &: \alpha_1 \neq 0
\end{aligned}$$

The log-likelihood function

$$
\begin{aligned}
\xi &= (\beta_0, \beta_1, \alpha_0, \alpha_1)^T \\
ln(y_1, y_2, \mu_i, \psi_i) &= -nlog(2\pi) - \sum_{i=1}^{n} log\psi_i - \sum_{i=1}^{n} \frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi_i} \\
ln(y_1, y_2, \xi) &= -nlog(2\pi) - \sum_{i=1}^{n} (\alpha_0 + \alpha_1(x_i - \bar{x})) \\
&\quad - \sum_{i=1}^{n} \frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))^2}{2exp(\alpha_0 + \alpha_1(x_i - \bar{x}))}, \qquad \sum x_i - \bar{x} = 0 \\
&= -nlog(2\pi) - n\alpha_0 - 1/2 \sum_{i=1}^{n} \frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 + \beta_1(x_i - \bar{x}))^2}{exp(\alpha_0 + \alpha_1(x_i - \bar{x}))}
\end{aligned}
$$

We will get the score function and Fisher information for ξ

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \alpha_0} &= -n + 1/2 \sum_{i=1}^{n} \frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))^2}{exp(\alpha_0 + \alpha_1(x_i - \bar{x}))} \\
&= -n + 1/2 \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2] \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_0^2} &= -1/2 \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \alpha_1} &= 1/2 \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_1^2} &= -1/2 \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x})^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \beta_0} &= \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))] \\
\frac{\partial^2 ln(\xi)}{\partial \beta_0^2} &= -2 \sum_{i=1}^{n} \psi_i^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \beta_1} &= \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \beta_1^2} &= -2 \sum_{i=1}^{n} \psi_i^{-1} (x_i - \bar{x})^2
\end{aligned}
$$

Other derivatives

$$\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \alpha_1} = -1/2 \sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x})$$

$$\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_0} = -\sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x})$$

$$\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_1} = -\sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x})$$

$$\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_0} = -\sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i) + (y_{i2} - \mu_i)](x_i - \bar{x})$$

$$\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_1} = -\sum_{i=1}^{n} \psi_i^{-1} [(y_{i1} - \mu_i) + (y_{i2} - \mu_i)](x_i - \bar{x})^2$$

$$\frac{\partial^2 ln(\xi)}{\partial \beta_0 \beta_1} = -2 \sum_{i=1}^{n} \psi_i^{-1} (x_i - \bar{x})$$

Taking expectation as $I(\xi) = -E(\partial^2 \xi)$

$$E(y_{i1} - \mu_i)^2 = \psi_i, \qquad E(y_{i1}) = E(y_{i2}) = \mu_i, \qquad \sum_{i=1}^{n} x_i - n\bar{x} = 0$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0^2}] = -1/2 \sum_{i=1}^{n} \psi_i^{-1} [E(y_{i1} - \mu_i)^2 + E(y_{i2} - \mu_i)^2] = -n$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1^2}] = -\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \beta_0^2}] = -2 \sum_{i=1}^{n} \psi_i^{-1}$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \beta_1^2}] = -2 \sum_{i=1}^{n} \psi_i^{-1} (x_i - \bar{x})^2$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \alpha_1}] = -1/2 \sum_{i=1}^{n} \psi_i^{-1} [E(y_{i1} - \mu_i)^2 + E(y_{i2} - \mu_i)^2] E(x_i - \bar{x}) = 0$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_0}] = 0, \qquad E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_1}] = 0$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_0}] = 0, \qquad E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_1}] = 0$$

$$E[\frac{\partial^2 ln(\xi)}{\partial \beta_0 \beta_1}] = -2 \sum_{i=1}^{n} \psi_i^{-1} (x_i - \bar{x})$$

Then

$$I(\xi) = -E(\partial^2 \xi) = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^n \psi_i^{-1} & 2\sum_{i=1}^n \psi_i^{-1}(x_i - \bar{x}) \\ 0 & 0 & 2\sum_{i=1}^n \psi_i^{-1}(x_i - \bar{x}) & 2\sum_{i=1}^n \psi_i^{-1}(x_i - \bar{x})^2 \end{bmatrix}$$

Under null hypothesis, we have score test statistics follows a chi-square distribution

$$\frac{\partial ln}{\partial \tilde{\xi}}^T I(\tilde{\xi})^{-1} \frac{\partial ln}{\partial \tilde{\xi}} \sim \chi^2(1)$$

So we have $\tilde{\psi} = exp(\tilde{\alpha_0})$, then $\tilde{\alpha_0} = ln(\tilde{\psi})$.
From part (a) which ψ is constant, we have $\psi = \frac{1}{4n}\sum_{i=1}^n (y_{i1} - y_{i2})^2$ and then,

$$\hat{\mu_i} = 1/2(y_{i1} + y_{i2})$$

$$\hat{\psi} = \frac{1}{4n}\sum_{i=1}^n (y_{i1} - y_{i2})^2$$

then the score function under $\tilde{\xi}$

$$\dot{l}(\xi) = \begin{bmatrix} \partial_{\alpha_0} l(\xi) & = -n + 1/2\sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2] = 0 \\ \partial_{\alpha_1} l(\xi) & = 1/2\sum_{i=1}^n \tilde{\psi}^{-1} 1/2(y_{i1} - y_{i2})^2(x_i - \bar{x}) = \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1} - y_{i2})^2(x_i - \bar{x}) \\ \partial_{\beta_0} l(\xi) & = \sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))] = 0 \\ \partial_{\beta_1} l(\xi) & = \sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))](x_i - \bar{x}) = 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1} - y_{i2})^2(x_i - \bar{x}) \\ 0 \\ 0 \end{bmatrix}$$

Under null hypothesis, $2\sum_{i=1}^n \psi_i^{-1}(x_i - \bar{x}) = 0$, then

$$I_n(\tilde{\xi}) = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1} & 0 \\ 0 & 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1}(x_i - \bar{x})^2 \end{bmatrix}$$

The score test statistics

$$SCn = \frac{\partial ln}{\partial \tilde{\xi}}^T I_n(\tilde{\xi})^{-1} \frac{\partial ln}{\partial \tilde{\xi}} = (0, \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1} - y_{i2})^2(x_i - \bar{x}), 0, 0)$$

$$\begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1} & 0 \\ 0 & 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1}(x_i - \bar{x})^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1} - y_{i2})^2(x_i - \bar{x}) \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{\left[\frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1} - y_{i2})^2(x_i - \bar{x})\right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

With $\tilde{\psi} = \frac{1}{4n}\sum_{i=1}^n (y_{i1} - y_{i2})^2$, we have

$$SCn = \frac{\left[n^2 \sum_{i=1}^n (y_{i1} - y_{i2})^2 (x_i - \bar{x})\right]^2}{[\sum_{i=1}^n (y_{i1} - y_{i2})^2]^2 \sum_{i=1}^n (x_i - \bar{x})^2} \sim \chi^2(1)$$

We will reject the H_0 if $SCn > \chi^2(1, 1-\alpha)$.

1.3 e

Suppose that the vector $Y = (Y_0; Y_1; Y_2)^T$ follows a multinomial distribution with total count m and probability vector $(\gamma_0; \gamma_1; \gamma_2)^T$ with

$$\gamma_j = \binom{2}{j} \pi^j (1-\pi)^{2-j} \theta^{-j(2-j)} / f(\pi, \theta), \qquad j = 0, 1, 2$$

where

$$f(\pi, \theta) = \sum_{k=0}^{2} \binom{2}{k} \pi^k (1-\pi)^{2-k} \theta^{-k(2-k)}$$

and $0 \le \pi \le 1, \theta > 0$ are parameters. Furthermore, define $\lambda = log\frac{\pi}{1-\pi}$ and $\psi = log\theta$.

(a) Derive a sufficient statistic for λ assuming $\psi = \psi_0$ is known. Derive a conditional likelihood for ψ.
Write the joint distribution of Y

$$\begin{aligned} P(Y) &= \binom{m}{y_0, y_1, y_2} \gamma_1^{y_1} \gamma_2^{y_2} \gamma_0^{y_0} \\ &= exp\left[log\binom{m}{y_0, y_1, y_2} + y_0 log\gamma_0 + y_1 log\gamma_1 + y_2 log\gamma_2\right] \end{aligned}$$

$$\begin{aligned} \gamma_0 &= \binom{2}{0} \pi^0 (1-\pi)^2 \theta^0 / f(\pi, \theta) = (1-\pi)^2 / f(\pi, \theta) \\ \gamma_1 &= \binom{2}{1} \pi^1 (1-\pi)^1 \theta^{-1} / f(\pi, \theta) = 2\pi(1-\pi)\theta^{-1} / f(\pi, \theta) \\ \gamma_2 &= \binom{2}{2} \pi^2 (1-\pi)^0 \theta^0 / f(\pi, \theta) = \pi^2 / f(\pi, \theta) \end{aligned}$$

$$
\begin{aligned}
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + y_0[2log(1-\pi) - logf(\pi,\theta)] \\
&+ y_1[log2\pi(1-\pi) - log\theta - logf(\pi,\theta)] + y_2[2log\pi - logf(\pi,\theta)] \\
f(\pi,\theta) &= \binom{2}{0}\pi^0(1-\pi)^2\theta^0 + \binom{2}{1}\pi^1(1-\pi)^1\theta^{-1} + \binom{2}{2}\pi^2(1-\pi)^0\theta^0 \\
logf(\pi,\theta) &= 2log(1-\pi) + log2\pi(1-\pi) - log\theta + 2log\pi \\
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + (2y_0 + y_1)log(1-\pi) \\
&- (y_0 + y_1 + y_2)logf(\pi,\theta) + (y_1 + 2y_2)log\pi + y_1 log2 - y_1 log\theta \\
m &= y_0 + y_1 + y_2, \qquad y_1 = m - y_0 - y_2 \\
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + (m + y_0 - y_2)log(1-\pi) - mlogf(\pi,\theta) \\
&+ (m - y_0 + y_2)log\pi + y_1 log2 - y_1 log\theta \\
&= log\binom{m}{y_0, y_1, y_2} + mlog\left[\frac{e^\lambda}{1+e^\lambda}\frac{1}{1+e^\lambda}\frac{(1+e^\lambda)^2}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right] \\
&- (y_0 - y_2)\lambda + y_1 log2 - y_1\psi
\end{aligned}
$$

If assume $\psi = \psi_0$ is known, then a sufficient statistics is $m, y_0 - y_2$.

$$
logP(Y) = log\binom{m}{y_0, y_1, y_2} + mlog\left[\frac{e^\lambda}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right] - (y_0 - y_2)\lambda + y_1 log2 - y_1\psi
$$

Let $y_2 - y_0 = t$,

$$
\begin{aligned}
P(t) &= \sum_t \binom{m}{y_0, y_1, y_2}\left[\frac{e^\lambda}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1}exp(-\psi y_1) \\
P(y_1|t) &= \frac{P(t,Y)}{P(t)} = \frac{\binom{m}{y_0,y_1,y_2}\left[\frac{e^\lambda}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1}exp(-\psi y_1)}{\sum_t \binom{m}{y_0,y_1,y_2}\left[\frac{e^\lambda}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1}exp(-\psi y_1)} \\
&= \frac{\frac{1}{y_0!y_1!y_2!}2^{y_1}exp(-\psi y_1)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')}
\end{aligned}
$$

The conditional distribution for ψ

$$
P(y_1, \psi|t) = \frac{\frac{1}{y_0!y_1!y_2!}2^{y_1}exp(-\psi y_1)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')}
$$

(b) The data $y_0 = 3; y_1 = 0; y_2 = 2$ were observed. Based on the conditional likelihood of Part (a), compute the exact one-sided p-value for testing $H0 : \theta = 1$ against

$H_0 : \theta > 1$ with λ unspecified.
The null hypothesis could be written as

$$H_0 : \psi = 0 \qquad vs. \qquad H_1 : \psi \neq 0$$

From $y_0 = 3; y_1 = 0; y_2 = 2$, we have $t = y_2 - y_0 = -1, m = 5$. There are possible 3 combinations that t=-1 as below

y_1	y_2	y_0	t	case
0	2	3	-1	1
2	1	2	-1	2
4	0	1	-1	3

So under H_0, the conditional probability for y_1 in the above 3 cases are

$$\begin{aligned} denominator &= \frac{1}{0!2!3!}2^0 exp(-\psi 0) + \frac{1}{1!2!2!}2^2 exp(-\psi 2) + \frac{1}{0!4!1!}2^4 exp(-\psi 4) \\ &= 2/3exp(-4\psi) + exp(-2\psi) + 1/12 = 21/12 \end{aligned}$$

$$P(y_1 = 0, \psi | t = -1) = \frac{\frac{1}{0!2!3!}2^0 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 1/21$$

$$P(y_1 = 2, \psi | t = -1) = \frac{\frac{1}{1!2!2!}2^2 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 12/21$$

$$P(y_1 = 4, \psi | t = -1) = \frac{\frac{1}{0!4!1!}2^4 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 8/21$$

We will reject H_0 if $P(y_1|t = -1) < 0.05$. Under the current sample, one sided test p-value for $P(y_1 = 0|t = -1) = 1/21 = 0.0476$, that $\psi \neq 0$.

1.4 b

Consider the following

(a) For an arbitrary model, consider the conditional score statistic

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi}$$

Show that the conditional score statistic for any model can be written as

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

The conditional score statistic is the derivative of the conditional distribution

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi}$$

$$p(\mathbf{Y}|\xi) = p(\mathbf{Y}|s_\lambda(\psi_0), \xi)p(s_\lambda(\psi_0)|\xi), \qquad p(\mathbf{Y}|s_\lambda(\psi_0), \xi) = \frac{p(\mathbf{Y}|\xi)}{p(s_\lambda(\psi_0)|\xi)}$$

$$l_c(\xi, \psi_0) = logp(\mathbf{Y}|s_\lambda(\psi_0), \xi) = logp(\mathbf{Y}|\xi) - logp(s_\lambda(\psi_0)|\xi)$$

Then we need to prove

$$U_\psi(\xi) = \frac{\partial l_c(\xi,\psi_0)}{\partial\psi}|_{\psi_0=\psi} = \partial_\psi logp(\mathbf{Y}|\xi) - \partial_\psi logp(s_\lambda(\psi_0)|\xi)$$

$$\partial_\psi logp(s_\lambda(\psi_0)|\xi) = E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

We can write

$$logp(\mathbf{Y}|\xi) = logp(\mathbf{Y}|s_\lambda(\psi_0),\xi) + logp(s_\lambda(\psi_0)|\xi)$$

$$E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) = E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0),\xi)|s_\lambda]\right) + E\left(\partial_\psi[logp(s_\lambda(\psi_0),\xi)|s_\lambda]\right)$$

in which, the integral and expectation can switch, then we have

$$E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0),\xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|s_\lambda(\psi_0),\xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|\xi)]\right) = 0$$

So,

$$E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) = \partial_\psi logp(s_\lambda(\psi_0),\xi)$$

Then we show

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

(b) Suppose that $y_1;..y_n$ are independent and y_i follows a Poisson distribution with mean $exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2})$, where $(x_{i1};x_{i2})$ are covariates, $\lambda = (\lambda_0;\lambda_1)$ is the nuisance parameter vector and ψ is the parameter of interest. Derive the conditional likelihood of ψ and show that this conditional likelihood is free of λ.
The joint distribution of $(y_1,,y_n)$ is given by

$$P(Y|\lambda,\psi) = exp\left(\sum_{i=1}^n y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)$$

Thus, $S_0 = \sum_{i=1}^n y_i$ is the sufficient and complete statistics for λ_0, and $S_1 = \sum_{i=1}^n y_i x_{i1}$ is the sufficient and complete statistics for λ_1.
The conditional distribution of ψ given S_0, S_1 is given by

$$\begin{aligned}
p(\mathbf{Y},\psi|S=(S_0,S_1)) &= \frac{exp\left(\sum_{i=1}^n y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)}{\sum_{y'\in S} exp\left(\sum_{i=1}^n y_i'(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy\right.} \\
&= \frac{exp\left(S_1\lambda_0 + S_2\lambda_1 + S_3\psi\right) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!)}{\sum_{y'\in S} exp\left(S_1'\lambda_0 + S_2'\lambda_1 + S_3'\psi\right) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i'!)} \\
&= \frac{exp\left(S_3\psi - logy_i!\right)}{\sum_{y'\in S} exp\left(S_3'\psi - logy_i'!\right)}, \qquad S_3 = \sum_{i=1}^n y_i x_{i2}, S_3' = \sum_{i=1}^n y_i' x_{i2}
\end{aligned}$$

which is independent of λ.

(c) Derive the conditional score statistic for part (b) and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ based on $U_\psi(\xi)$.
The log likelihood of the conditional distribution is

$$l_c(\psi) = S_3\psi - logy_i! - log\left[\sum_{y'\in S} exp\left(S_3'\psi - logy_i'!\right)\right], \qquad S_3 = \sum_{i=1}^{n} y_i x_{i2}, S_3' = \sum_{i=1}^{n} y_i' x_{i2}$$

The score function and observed fisher information is

$$\begin{aligned}
U_\psi(\xi) &= \frac{\partial l_c(\xi,\psi_0)}{\partial\psi}|_{\psi_0=\psi} \\
&= \psi - \frac{\sum_{y'\in S} S_3' exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y'\in S} exp\left(S_3'\psi - logy_i'!\right)} \\
\frac{\partial^2 l_c(\xi,\psi_0)}{\partial\psi^2} &= \left[\frac{\sum_{y'\in S} S_3' exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y'\in S} exp\left(S_3'\psi - logy_i'!\right)}\right]^2 - \frac{\sum_{y'\in S} S_3'^2 exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y'\in S} exp\left(S_3'\psi - logy_i'!\right)}
\end{aligned}$$

The newton-Raphson algorithm

$$\psi^{k+1} = \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial\psi^2}\right]^{-1} U_\psi(\psi^k)$$

where $\frac{\partial^2 l_c(\psi^k)}{\partial\psi^2}, U_\psi(\psi^k)$ are from above equations.

(d) Now suppose that we only have two random variables $y_1 \sim Poisson(\mu_1)$ and $y_2 \sim Poisson(\mu_2)$, where y_1 and y_2 are independent. We are interested in making inferences on the ratio $\psi = \mu_1/\mu_2$. Let $\xi = (\psi, \lambda)$, where λ represents the nuisance parameter.

(i) Show that the log-likelihood function of ξ can be written as

$$l(\xi) = (y_1 + y_2)\lambda + y_1 log(\psi) - exp(\lambda)(1 + \psi)$$

where λ is a function of μ_2. Explicitly state what λ is.
Write the joint distribution of y_1, y_2

$$\begin{aligned}
P(y_1, y_2) &= \frac{\mu_1^{y_1} e^{-\mu_1}}{y_1!} \frac{\mu_2^{y_2} e^{-\mu_2}}{y_2!} \\
logP(y_1, y_2) &= y_1 log\mu_1 - \mu_1 + y_2 \log\mu_2 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + y_1 log\mu_2 + y_2 log\mu_2 - \mu_1 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + (y_1 + y_2) log\mu_2 - \mu_2(\mu_1/\mu_2 + 1) - logy_1! - logy_2!
\end{aligned}$$

where

$$\psi = log\frac{\mu_1}{\mu_2}$$
$$\lambda = log\mu_2$$

(ii) Derive the conditional likelihood of ψ and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ .
From part (a), we see $y_1 + y_2$ is the sufficient statistics for λ, while $y_1 + y_2 \sim Poission(\mu_1 + \mu_2)$ then we have conditional distribution of ψ condition on $S = y_1 + y_2$.

$$\begin{aligned}
Y(\psi|S = y_1 + y_2, \lambda) &= \frac{exp\left[y_1\psi + (y_1 + y_2)\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[(y_1 + y_2)log(\mu_1 + \mu_2) - (\mu_1 + \mu_2) - log(y_1 + y_2)!\right]} \\
&= \frac{exp\left[y_1\psi + S\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[S(\lambda + log(\psi + 1)) - exp(\lambda)(\psi + 1) - logS!\right]} \\
&= \frac{exp\left[y_1\psi - logy_1! - logy_2!\right]}{exp\left[(y_1 + S - y_1)log(\psi + 1)) - logS!\right]} \\
&= \binom{S}{y_1}\left(\frac{\psi}{1+\psi}\right)^{y_1}\left(\frac{1}{1+\psi}\right)^{S-y_1}
\end{aligned}$$

The conditional distribution is a binomial, $B(S, \psi/(1 + \psi))$.
The score function and observed fisher information

$$\begin{aligned}
logY(\psi|S, \lambda) &= y_1log\psi - Slog(1 + \psi) + log\binom{S}{y_1} \\
\partial_\psi logY(\psi|S, \lambda) &= \frac{y_1}{\psi} - \frac{S}{1+\psi} = 0, \qquad \hat{\psi} = y_1/(S - y_1) \\
\partial^2_\psi logY(\psi|S, \lambda) &= -\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}
\end{aligned}$$

The $CMLE = \hat{\psi} = y_1/(S - y_1)$. And the newton-Raphson equation

$$\begin{aligned}
\psi^{k+1} &= \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial\psi^2}\right]^{-1} U_\psi(\psi^k) \\
&= \psi^k - \left[-\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}\right]^{-1}\left[\frac{y_1}{\psi} - \frac{S}{1+\psi}\right]|_{\psi=\psi^k} \\
&= \psi^k + \frac{y_1/\psi^k - S/(1 + \psi^k)}{y_1/{\psi^k}^2 - S/(1 + \psi^k)^2}
\end{aligned}$$

1.5 a

Suppose that $y_1; ...y_n$ are independent Bernoulli random variables, where $y_i \sim Bernoulli(\pi)$, and we consider a logistic regression so that $logit(\pi) = x_i'\beta$, where $\beta = (\beta_1; ...\beta_p)$. Our interest is inference on $(\beta_1; \beta_2)$, with all other parameters being treated as nuisance.

(a) Derive the conditional likelihood of $(\beta_1; \beta_2)$ and express it in the simplest possible form.
The joint distribution of $y_1; ...y_n$

$$p(Y) = \prod_{i=0}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)}$$

$$logp(Y) = \sum_{i=0}^{n} y_i logp_i + (1-y_i)log(1-p_i) = \sum_{i=0}^{n} y_i log\frac{p_i}{1-p_i} + log(1-p_i)$$

$$logit(pi) = log\frac{p_i}{1-p_i} = x_i'\beta, \qquad p_i = \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}$$

$$logp(Y) = \sum_{i=0}^{n} y_i x_i'\beta - log(1+exp(x_i'\beta))$$

$$= \sum_{i=0}^{n} y_i(x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + ..x_{ip}\beta_p) - log(1+exp(x_i'\beta))$$

We can see that $\sum_{i=0}^{n} x_{i1}y_i$ is a sufficient and complete statistics for β_1. When only $(\beta_1; \beta_2)$ are the interest, and all other parameters being treated as nuisance. Then $s_j = \sum_{i=0}^{n} y_i x_{ij}$ is sufficient statistics for β_j. Let $S = (s_3, s_4, ..s_p)$

$$P(\beta_1, \beta_2|S) = \frac{exp\left[\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2 + ..(y_i x_{ip})\beta_p - log(1+exp(x_i'\beta))\right]}{\sum_{t\in S} exp\left[(t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2 + ...(t_i x_{ip})\beta_p - log(1+exp(x_i^T\beta)\right]}$$

$$= \frac{exp\left(\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2)\right)}{\sum_{t\in S} exp\left((t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2)\right)}$$

$$= \frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}, \qquad S_j = \sum_{i=0}^{n}(y_i x_{ij}), S_j' = \sum_{i=0}^{n}(t_i x_{ij})$$

(b) Derive the score equations for $(\beta_1; \beta_2)$ based on the conditional likelihood derived in part (a).
The log conditional distribution is

$$l_c(\beta_1, \beta_2|S) = logp(Y, \xi) - logp(s, \lambda, \psi_0) = logP(\beta_1, \beta_2|S)$$

$$l_c(\beta_1, \beta_2|S) = log\frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} = S_1\beta_1 + S_2\beta_2 - log\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)$$

$$\frac{\partial l_c}{\partial \beta_1} = S_1 - \frac{\sum_{S'} S_1' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}$$

$$\frac{\partial l_c}{\partial \beta_2} = S_2 - \frac{\sum_{S'} S_2' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}$$

The score equations are setting the score function to 0

$$SCn = 0 = \begin{bmatrix} S_1 - \frac{\sum_{S'} S_1' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \\ S_2 - \frac{\sum_{S'} S_2' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(c) Derive the asymptotic covariance matrix of the conditional maximum likelihood estimates of $(\beta_1; \beta_2)$.
The Fisher information of $(\beta_1; \beta_2)$

$$\frac{\partial^2 l_c}{\partial \beta_1^2} = \left[\frac{\sum_T T_1 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}\right]^2 - \frac{\sum_T T_1^2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}$$

$$\frac{\partial^2 l_c}{\partial \beta_2^2} = \left[\frac{\sum_T T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}\right]^2 - \frac{\sum_T T_2^2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}$$

$$\frac{\partial^2 l_c}{\partial \beta_1 \beta_2} = \frac{\left[\sum_T T_1 exp\left(T_1\beta_1 + T_2\beta_2\right)\right]\left[\sum_T T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)\right]}{\left[\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)\right]^2} - \frac{\sum_T T_1 T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}$$

Thus the asymptotic covariance matrix $Cov(\beta_1, \beta_2)$ is

$$\begin{aligned} Cov(\beta_1, \beta_2) &= I(\beta_1, \beta_2)^{-1} \\ I(\beta_1, \beta_2) &= -E\left[\frac{\partial^2 l_c}{\partial \beta^2}\right] = -\lim_{n\to\infty} \frac{I_n(\beta)}{n} \\ I_n(\beta) &= -\begin{bmatrix} \frac{\partial^2 l_c}{\partial \beta_1^2} & \frac{\partial^2 l_c}{\partial \beta_1 \beta_2} \\ \frac{\partial^2 l_c}{\partial \beta_1 \beta_2} & \frac{\partial^2 l_c}{\partial \beta_2^2} \end{bmatrix} \end{aligned}$$

(d) Derive the conditional score test for testing $H_0 : \beta_1 = \beta_2 = 0$.

$$SCn = \frac{\partial l_c}{\partial \tilde{\beta}}^T I_n(\tilde{\beta})^{-1} \frac{\partial l_c}{\partial \tilde{\beta}} \sim \chi^2(1)$$

SCn is estimated under $H_0, \beta_1 = \beta_2 = 0$. The SCn quadratic form is rank 1, so the degrees of freedom is 1.
We will reject H_0 if $SCn > \chi^2(1, \alpha)$.