Chi-square distribution

Mingwei Fei

December 6, 2022

1 Chi-square distribution

If $Z_1, ..., Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = Z_i^2 \sim \chi^2(k)$$
$$p(k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} exp(-\frac{x}{2})$$

The non-central chi-square distribution: Let $(X_1, X_2, \ldots, X_i, \ldots, X_k)$ be k independent, normally distributed random variables with means μ_i and unit variances. Then the random variable

$$Q = \sum_{i=1}^{k} X_i^2 \sim \chi^2(k, \lambda), \qquad \lambda = \sum_{i=1}^{k} \mu_i^2$$

The sample mean of n i.i.d. chi-squared variables of degree k is distributed according to a gamma distribution with shape α and scale θ parameters:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim Gamma(\alpha = nk/2, \theta = 2/n)$$

1.1 Normal distribution and variance

Suppose that $y_i \sim N(\mu_i, \sigma^2)$, and independent for $i = 1, ..n$, where μ_i and σ^2 are unknown.

(a) Suppose that $\mu_i = x_i^T \beta$, where x_i is a $p \times 1$ vector of covariates for the ith subject, $i = 1, ..n$. Suppose β is the vector parameter of interest and σ^2 is nuisance parameter. Derive the conditional likelihood of β and a closed form estimate of the conditional MLE of β, and compare the result to the MLE of β of full likelihood.

The matrix form and projection operator generally applies to normal distribution, so we will usually write in matrix form for multivariate normal distribution. Because the likelihood function could be considered as MVN, as we are estimating β, σ^2 using all the y_i simultaneously.

$$p(Y, \mu_i, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n exp\{\frac{(Y-\mu)^T(Y-\mu)}{2\sigma^2}\}$$

We can use $M = X(X^TX)^{-1}X^T$ to get the β

$$X\beta = MY$$
$$\hat{\beta} = (X^TX)^{-1}X^TY$$

Also we have

$$Z = \frac{Y-\mu}{\sigma I} \sim N(0, I)$$
$$\frac{Y^T(I-M)Y}{\sigma^2 I} \sim \chi^2_{n-p}$$

So we have distribution for σ^2

$$p(z) = \frac{1}{2^{k/2}\Gamma k/2} z^{k/2-1} exp(-\frac{z}{2})$$
$$x = \sigma^2 z, \qquad \frac{\partial z}{\partial x} = \sigma^{-2}$$
$$p(x) = \sigma^{-2} \frac{1}{2^{k/2}\Gamma(k/2)} \left(\frac{x}{\sigma^2}\right)^{k/2-1} exp(-\frac{x}{2\sigma^2})$$

The conditional log-likelihood of β

$$l_c = logp(y) - logp(x) = -nlog\sqrt{2\pi} - nlog\sigma I + \frac{-(Y-X\beta)^T(Y-X\beta)}{2\sigma^2}$$
$$- log\sigma I + (k/2-1)log\frac{x}{\sigma^2} - log2^{k/2}\Gamma(k/2) - \frac{x}{2\sigma^2}$$

Taking first derivative, we see that the distribution of σ^2 is free of β

$$\frac{\partial l_c}{\partial \beta} = -X^T(Y-X\beta) - (Y-X\beta)^TX = 0$$
$$\hat{\beta} = (X^TX)^{-1}X^TY$$

Thus the MLE of conditional distribution of β is equal to the MLE of β of the full likelihood.

Method 2: Use the sufficient statistics for σ^2, which is $(Y-\mu)^T(Y-\mu) = \sum_{i=1}^n (y_i - \mu_i)^2$. Fix β at β_0, and let $\mu_0 = X\beta_0$, then we have the non-central χ^2 distribution with $r = n-p$ degrees of freedom

$$p(S(\mu_0)) =$$

Then we also have log-likelihood

$$l_c = logp(y) - logp(S(\mu_0)) =$$

The conditional score statistic for β is

$$\begin{aligned}
U_\beta(\beta, \sigma^2|\beta_0) &= D^T \frac{\partial l_c(\mu, \sigma^2|\mu_0)}{\partial \mu}\bigg|_{\beta_0=\beta} = \frac{1}{\sigma^2}(Y - X\beta) \\
D &= X \\
U_\beta(\beta, \sigma^2|\beta_0) &= X^T(Y - X\beta) = 0 \\
\hat{\beta}_c &= (X^TX)^{-1}X^TY
\end{aligned}$$

(b) Suppose β is nuisance parameter, and σ^2 is the parameter of interest. Derive the conditional distribution and calculate conditional MLE.

If we need to assume value for σ, it is the S_n that includes σ, so need to separate out the situation from above question. Here I have to replace $\mu = x^T\beta$ as the derivative is about σ, not β. $S_n = \sum_{i=1}^n y_i x_i = X^TY$ is the sufficient statistics for $X\beta$, then

$$\begin{aligned}
E(S_n) &= X^TX\beta, \qquad Var(S_n) = \sigma^2 X^TX \\
S_n &\sim N(X^TX\beta, \sigma^2X^TX) \\
p(S_n) &= \frac{1}{\sqrt{2\pi}}\frac{n}{\sigma}exp\{\frac{-(S_n - X^TX\beta)^T(S_n - X^TX\beta)}{2\sigma^2}\} \\
l_c &= logp(y) - logp(S(\sigma_0^2)) \\
&= -\frac{n}{2}log(\sigma^2) - \frac{(Y - X\beta)^T(Y - X\beta)}{2\sigma^2} \\
&+ \frac{p}{2}log(\sigma^2) + \frac{(X^TY - X^TX\beta)^T(X^TX)^{-1}(X^TY - X^TX\beta)}{2\sigma^2}
\end{aligned}$$

Then take the first derivative of σ^2

$$\begin{aligned}
\frac{\partial^2 l_c}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(Y - X\beta)^T(Y - X\beta) \\
&+ \frac{p}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}(X^TY - X^TX\beta)^T(X^TX)^{-1}(X^TY - X^TX\beta) = 0
\end{aligned}$$

We could get the MLE of β from $p(S_n)$

$$\begin{aligned}\frac{\partial l_{S_n})}{\partial \beta} &= \frac{1}{2(\sigma^2)}2(X^TY - X^TX\beta) = 0\\ \hat{\beta} &= (X^TX)^{-1}X^Y\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial^2 l_c}{\partial \sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(Y-X\beta)^T(Y-X\beta) - \frac{1}{2(\sigma^2)^2}(X^TY - X^TX\beta)^T(X^TX)^{-1}(X^TY - X^TX\beta)\\ \sigma_c^2 &= \frac{(Y-X\beta)^T(Y-X\beta)}{n-p} = \frac{Y^T(I-M)Y}{n-p}\end{aligned}$$

The MLE of σ^2 from the full data likelihood is

$$\hat{\sigma}^2 = \frac{Y^T(I-M)Y}{n} \leq \hat{\sigma}_c^2$$

Assume that $y_1, , y_n$ are independent and y_i follows a Poisson distribution with mean $exp(\lambda+\psi x_i)$, where x_i is a covariate of interest. Suppose that λ is the nuisance parameter and ψ is the parameter of interest. The joint distribution of $(y_1, , y_n)$ is given by

$$exp\left(\sum_{i=1}^n y_i(\lambda + \psi x_i) - \sum_{i=1}^n exp(\lambda + \psi x_i) + c\right)$$

Thus, $S_n = \sum_{i=1}^n y_i$ is the sufficient and complete statistics for λ. Since S_n follows a poisson distribution with mean $\sum_{i=1}^n exp(\lambda + \psi x_i)$, the log-likelihood of conditional distribution of $\mathbf{Y}$) given $S_n = \sum_{i=1}^n y_i$ is given by

$$\begin{aligned}logp(\mathbf{Y};\xi) &= \sum_{i=1}^n y_i(\lambda + \psi x_i) - \sum_{i=1}^n exp(\lambda + \psi x_i) + c_1\\ logp(\mathbf{s};\xi) &= \sum_{i=1}^n y_i log\left(\sum_{i=1}^n exp(\lambda + \psi x_i)\right) - \sum_{i=1}^n exp(\lambda + \psi x_i) + c_2\\ log_c p(\psi) &= logp(\mathbf{Y};\xi) - logp(\mathbf{s};\xi)\\ &= \sum_{i=1}^n y_i(\psi x_i) - \sum_{i=1}^n y_i log\left(\sum_{i=1}^n exp(\psi x_i)\right)\end{aligned}$$

which is independent of λ.

1.1.1 Negative Binomial distribution - conditional probability free of nuisance parameters

Suppose that $y_1, , y_n$ are independently and identically distributed with density function

$$P(y) = \frac{\Gamma(\psi + y)}{\Gamma(y+1)\Gamma(\psi)} \frac{\lambda^y \psi^\psi}{(\lambda+\psi)^{y+\psi}}, y = 0, 1, ..$$

Find a conditional likelihood score function $U_\psi(\xi)$ for ψ.
Write the distribution in exponential family

$$P(y) = exp\left[log\left(\frac{\Gamma(\psi+y)}{\Gamma(y+1)\Gamma(\psi)}\right) + ylog\frac{\lambda}{\lambda+\psi} + \psi log\frac{\psi}{\lambda+\psi}\right]$$

In which,

$$\begin{aligned} \theta &= log\frac{\lambda}{\lambda+\psi} \\ b(\theta) &= -\psi log\frac{\psi}{\lambda+\psi} = -\psi log(1 - exp\theta) \end{aligned}$$

We can find the distribution from MGF or KGF function

$$\begin{aligned} &M_y(t) = exp\{\phi[b(\theta + t/\phi) - b(\theta)]\} \\ &K_y(t) = logM_y(t) = \phi[b(\theta + t/\phi) - b(\theta)], \qquad \phi = 1 \end{aligned}$$

Then

$$\begin{aligned} K_y(t) &= -\psi log\left(1 - exp(\theta + t)\right) + \psi log(1 - exp\theta) \\ &= log\left(\frac{1 - e(\theta)}{1 - e(\theta)e(t)}\right)^\psi \end{aligned}$$

Then

$$M_y(t) = \left(\frac{1 - e(\theta)}{1 - e(\theta)e(t)}\right)^\psi$$

which is the MGF for negative binomial distribution. Then we have

$$\begin{aligned} &\sum_{i=1}^n y_i \sim NB\left(n\psi, \frac{\lambda}{\lambda+\psi}\right) \\ P(S = &\sum_{i=1}^n y_i) = exp\left[log\left(\frac{\Gamma(n\psi + s)}{\Gamma(s+1)\Gamma(n\psi)}\right) + slog\frac{\lambda}{\lambda+\psi} + n\psi log\frac{\psi}{\lambda+\psi}\right] \end{aligned}$$

where s is a sufficient statistics for λ. Now

$$
\begin{aligned}
l_c(\psi) &= logP_y(y|\lambda,\psi) - logP_s(S) \\
&= \sum_{i=1}^{n} log\left(\frac{\Gamma(\psi+y_i)}{\Gamma(y_i+1)\Gamma(\psi)}\right) + \sum_{i=1}^{n} y_i log\frac{\lambda}{\lambda+\psi} + n\psi log\frac{\psi}{\lambda+\psi} \\
&- log\left(\frac{\Gamma(n\psi+s)}{\Gamma(s+1)\Gamma(n\psi)}\right) - slog\frac{\lambda}{\lambda+\psi} - n\psi log\frac{\psi}{\lambda+\psi} \\
&= \sum_{i=1}^{n} log\left(\frac{\Gamma(\psi+y_i)}{\Gamma(y_i+1)\Gamma(\psi)}\right) - log\left(\frac{\Gamma(n\psi+s)}{\Gamma(s+1)\Gamma(n\psi)}\right) \\
&= \sum_{i=1}^{n} log\left(\frac{\Gamma(\psi+y_i)}{\Gamma(y_i+1)\Gamma(\psi)}\right) - log\left(\frac{\Gamma(n\psi+\sum_{i=1}^{n} y_i)}{\Gamma(\sum_{i=1}^{n} y_i+1)\Gamma(n\psi)}\right)
\end{aligned}
$$

The score function

$$
\begin{aligned}
U_\psi(\xi) &= \partial_\psi\left[\sum_{i=1}^{n} log\left(\frac{\Gamma(\psi+y_i)}{\Gamma(y_i+1)\Gamma(\psi)}\right) - log\left(\frac{\Gamma(n\psi+\sum_{i=1}^{n} y_i)}{\Gamma(\sum_{i=1}^{n} y_i+1)\Gamma(n\psi)}\right)\right] \\
&= \partial_\psi\left[\sum_{i=1}^{n} log\Gamma(\psi+y_i) - log\Gamma(y_i+1) - log\Gamma(\psi) - log\Gamma(n\psi+\sum_{i=1}^{n} y_i) - log\Gamma(\sum_{i=1}^{n} y_i+1) - log\Gamma(n\psi)\right] \\
&= \frac{\Gamma'(\psi+y_i)}{\Gamma(\psi+y_i)} - \frac{n\Gamma'(\psi)}{\Gamma(\psi)} - \frac{n\Gamma'(n\psi+\sum_{i=1}^{n} y_i)}{\Gamma(n\psi+\sum_{i=1}^{n} y_i)} - \frac{n\Gamma'(n\psi)}{\Gamma(n\psi)}
\end{aligned}
$$

1.2 Sufficient statistics with distribution not free of interest parameter

The second scenario, the conditional distribution of $\mathbf{Y}$ given $\mathbf{s}_\lambda(\psi)$ is not well defined. Since $\mathbf{s}_\lambda(\psi)$ depends on ψ, it is difficult to calculate the conditional distribution of $\mathbf{Y}$ given $\mathbf{s}_\lambda(\psi)$. However, for a fixed ψ_0, we may use

$$
l_c(\xi,\psi_0) = logP(\mathbf{Y}|s_\lambda(\psi_0),\xi) = logP(\mathbf{Y}|\xi) - logP(s_\lambda(\psi_0),\xi) \tag{1}
$$

We can see that $P(\mathbf{Y}|\xi)$ is now conditional on ξ, because it is basically the same. And the conditional score statistics

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

It can be shown that

$$
\begin{aligned}
U_\psi(\xi) &= \frac{\partial logp(\mathbf{Y}|\xi)}{\partial\psi} - \frac{\partial logp(\mathbf{s};\xi)}{\partial\psi} \\
U_\psi(\xi) &= \partial_\psi logp(\mathbf{Y}|\xi) - E[\partial_\psi logp(\mathbf{Y}|\xi)|s_\lambda(\psi)]
\end{aligned}
$$

We can get conditional score statistics in an alternative way, which is $\frac{\partial log E[p(\mathbf{Y}|\xi,\mathbf{s})|\mathbf{s}]}{\partial \psi}$.
Proof

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

integral and expectation can switch, distribution integral with no ψ

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

2 Practice

2.1 Pair of variables

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 $S_n(\psi) = 0$.

2.2 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

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

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

2.3 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 $0 \leq \pi \leq 1, \theta > 0$ 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 $\psi \neq 0$.

2.4 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 $S = y_1 + y_2$.

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

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

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

2.5 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 $y_1;...y_n$

$$
\begin{aligned}
p(Y) &= \prod_{i=0}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)} \\
logp(Y) &= \sum_{i=0}^{n} y_i logp_i + (1-y_i)log(1-p_i) = \sum_{i=0}^{n} y_i log\frac{p_i}{1-p_i} + log(1-p_i) \\
logit(pi) &= log\frac{p_i}{1-p_i} = x_i'\beta, \qquad p_i = \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)} \\
logp(Y) &= \sum_{i=0}^{n} y_i x_i'\beta - log(1+exp(x_i'\beta)) \\
&= \sum_{i=0}^{n} y_i(x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + ..x_{ip}\beta_p) - log(1+exp(x_i'\beta))
\end{aligned}
$$

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 $S = (s_3, s_4, ..s_p)$

$$
\begin{aligned}
P(\beta_1,\beta_2|S) &= \frac{exp\left[\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2 + ..(y_i x_{ip})\beta_p - log(1+exp(x_i'\beta))\right]}{\sum_{t\in S} exp\left[(t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2 + ...(t_i x_{ip})\beta_p - log(1+exp(x_i^T\beta)\right]} \\
&= \frac{exp\left(\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2)\right)}{\sum_{t\in S} exp\left((t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2)\right)} \\
&= \frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}, \qquad S_j = \sum_{i=0}^{n}(y_i x_{ij}), S_j' = \sum_{i=0}^{n}(t_i x_{ij})
\end{aligned}
$$

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

$$
\begin{aligned}
l_c(\beta_1,\beta_2|S) &= logp(Y,\xi) - logp(s,\lambda,\psi_0) = logP(\beta_1,\beta_2|S) \\
l_c(\beta_1,\beta_2|S) &= log\frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} = S_1\beta_1 + S_2\beta_2 - log\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right) \\
\frac{\partial l_c}{\partial \beta_1} &= S_1 - \frac{\sum_{S'} S_1' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \\
\frac{\partial l_c}{\partial \beta_2} &= S_2 - \frac{\sum_{S'} S_2' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}
\end{aligned}
$$

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

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

(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 $SCn > \chi^2(1, \alpha)$.