Distribution

Mingwei Fei

December 17, 2022

1 Distribution Transformation

Let $X \sim Gamma(\alpha, 1)$ and $Y \sim Gamma(\beta, 1)$ where the paramaterization is such that α is the shape parameter. Then

$$\frac{X}{X+Y} \sim Beta(\alpha, \beta)$$

(i) How do we create the transformed variables $U = \frac{X}{X+Y}$, V, such that it is easy to get the distribution of U and V, and easy to integrate out V to get the margin distribution of U? This requires familiarity of Beta and Gamma distribution.

One proposal is $V = Y$, and the other is $V = X + Y$. Compare between the two and see which one is better.

Note: The way to choose new parameter is easier to get the original parameter by simple arithmetic calculation or just itself. If we are using the first set of new parameter, it would get the distribution very complicated, the Beta distribution has the form of $U, 1-U$, need to keep in mind of the achieving the product of two distribution form.

(ii) We could see that $V = X + Y$ is better as U and V are independent from each other.

$$f(X) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$$
$$f(Y) = \frac{1}{\Gamma(\beta)} y^{\beta-1} e^{-y}$$

Let

$$U = \frac{X}{X+Y}, \qquad V = X + Y$$

Then

$$X = UV, \qquad Y = V - UV$$

The Jacobian transformation matrix

$$J = \begin{pmatrix} \dfrac{\partial X}{\partial U} & \dfrac{\partial X}{\partial V} \\ \dfrac{\partial Y}{\partial U} & \dfrac{\partial Y}{\partial V} \end{pmatrix} = \begin{pmatrix} V & U \\ -V & 1-U \end{pmatrix}$$
$$|J| = V$$

X and Y are independent, so the joint distribution of (X, Y)

$$f(X,Y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}e^{-x}y^{\beta-1}e^{-y}$$
$$f(U,V) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} U^{\alpha-1}(1-U)^{\beta-1}V^{\alpha+\beta-1}e^{-V}$$

We don't always need to integrate out the other parameter, if we can write the distribution in the form that we can recognize, then will directly get the distribution.

$V \sim Gamma(\alpha+\beta, 1)$ and $U \sim Beta(\alpha, \beta)$.

1.1 Gamma distribution and Beta distribution

If $Y_1, ..., Y_{n+1}$ are i.i.d $Exp(\theta)$, then

$$Z_i = \frac{Y_1 + ... + Y_i}{Y_1 + .. + Y_{n+1}} \sim Beta(i, n-i+1)$$

Particularly, $(Z_1, ...Z_n)$ has the same distribution as the order statistics $(\xi_{n:1}, ..\xi_{n:n})$ of n Uniform(0,1) random variables.

1.1.1 Prove Beta distribution

The $Z_{n+1} = 1$, so we don't need to calculate this distribution. We can use the method in above problem to get the distribution of Z_i, and we need to get the joint distribution of $(Z_1, ...Z_n)$.

Let

$$U = Y_1 + ... + Y_i, \qquad V = Y_{i+1} + .. + Y_{n+1}$$

Then $U \sim Gamma(i, \theta), V \sim Gamma(n+1-i, \theta)$, Let

$$Z_i = U/(U+V), \qquad W = U+V$$

Consider the transformation $(U,V)^T \to (Z_i, W)^T$, note that the transform is one-to-one with the Jacobian

$$\left| \frac{\partial(U,V)}{\partial(Z_i, W)} \right| = |W|$$

For joint distribution of $(U,V)^T$,

$$\frac{1}{\Gamma(i)}\theta exp(-\theta u)(\theta u)^{i-1}I(u>0)\times\frac{1}{\Gamma(n+1-i)}\theta exp(-\theta v)(\theta v)^{n-i}I(v>0)$$

We obtain the joint density of (Z_i,W) as

$$\begin{aligned}&\frac{1}{\Gamma(i)}\theta exp(-\theta z_i w)(\theta z_i w)^{i-1}\\ &\times\frac{1}{\Gamma(n+1-i)}\theta exp(-\theta(1-z_i)w)(\theta(1-z_i)w)^{n-i}w\times I(0<z_i<1)I(w>0)\end{aligned}$$

Thus, the marginal density of $Z_i = X/(X+Y)$ is equal to

$$\begin{aligned}&(1-z_i)^{n-i}z_i^{i-1}I(0<z_i<1)\frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)}\int_w \theta exp(-\theta w)(\theta w)^n I(w>0)dw\\ &=\frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)}(1-z_i)^{n-i}z_i^{i-1}I(0<z_i<1)\end{aligned}$$

That is $Z_i \sim Beta(i,n+1-i)$

1.1.2 Prove Uniform distribution

We need to note that, we always construct a joint distribution as we are using the $Y_1,...,Y_{n+1}$ in joint distribution. So the transformed variables are also joint distribution.

Then we need to find the relationship between $Y_1,...,Y_{n+1}$ and $(Z_1,...Z_n)$. We can set the total sum as $S_{n+1} = Y_1+..+Y_{n+1}$.

$$\begin{aligned}Y_1 &= Z_1\times S_{n+1}\\ Y_2 &= Z_2\times S_{n+1}-Y_1=(Z_2-Z_1)\times S_{n+1}\\ Y_3 &= Z_3\times S_{n+1}-Y_1-Y_2=(Z_3-Z_2)\times S_{n+1}\\ &....\\ Y_n &= Z_n\times S_{n+1}-Y_1-...-Y_{n-1}=(Z_n-Z_{n-1})\times S_{n+1}\\ Y_{n+1} &= S_{n+1}-Y_1-...-Y_n=(1-Z_n)\times S_{n+1}\end{aligned}$$

Then we have the Jacobian transform distribution

$$J=\begin{pmatrix}\frac{\partial Y_1}{\partial Z_1} & \frac{\partial Y_1}{\partial Z_2}... & \frac{\partial Y_1}{\partial Z_n} & \frac{\partial Y_1}{\partial S_{n+1}}\\ .. & .. & .. & ..\\ \frac{\partial Y_n}{\partial Z_1} & \frac{\partial Y_n}{\partial Z_2}... & \frac{\partial Y_n}{\partial Z_n} & \frac{\partial Y_n}{\partial S_{n+1}}\\ \frac{\partial Y_{n+1}}{\partial Z_1} & \frac{\partial Y_{n+1}}{\partial Z_2} & ... & \frac{\partial Y_{n+1}}{\partial S_{n+1}}\end{pmatrix}=\begin{pmatrix}S_{n+1} & 0 & 0 & .. & .. & Z_1\\ -S_{n+1} & S_{n+1} & 0 & .. & .. & Z_2-Z_1\\ .. & .. & .. & & & \\ 0 & 0 & ..-S_{n+1} & S_{n+1} & 0 & Z_n-Z_{n-1}\\ 0 & 0 & .. & 0 & S_{n+1} & 1-Z_n\end{pmatrix}$$

$|J| = S_{n+1}^n$

(b) Let

$$W_1 = Y_1, W_2 = Y_1 + Y_2, ..., W_n = Y_1 + ... + Y_n, S = Y_1 + ... + Y_{n+1}.$$

Consider the transformation $(Y_1, ..., Y_{n+1})' \mapsto (W_1, ..., W_n, S)'$. Note that the transformation is one-to-one with the Jacobian

$$|det\left(\frac{\partial(Y_1, ..., Y_{n+1})}{\partial(W_1, ..., W_n, S)}\right)| = |det\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}| = 1$$

Figure 1: Alt method

The joint distribution of $Y_1, ..., Y_{n+1}$

$$\begin{aligned} f(Y_1, Y_2, ..Y_{n+1}) &= \prod_{i=1}^{n+1} \frac{1}{\theta} exp(-\frac{y_i}{\theta}) I(0 < y_i < \infty) \\ &= \frac{1}{\theta^{n+1}} exp(-\frac{\sum_{i=1}^{n+1} y_i}{\theta}) I(0 < y_i < \infty) \\ &= \frac{1}{\theta^{n+1}} exp(-\frac{S_{n+1}}{\theta}) I(0 < y_i < \infty) \end{aligned}$$

Then the transformed variables

$$f(Z_1, Z_2, ..Z_n, S_{n+1}) = \frac{1}{\theta^{n+1}} exp(-\frac{S_{n+1}}{\theta}) S_{n+1}^n I(0 < z_1 < z_2 < .. < z_n < 1) I(S_{n+1} > 0)$$

We obtain the joint distribution of $(Z_1, ...Z_n)$

$$\begin{aligned} f(Z_1, Z_2, ..Z_n) &= \int_s f(Z_1, Z_2, ..Z_n, S_{n+1}) dS \\ &= \int_s \frac{1}{\theta^{n+1}} exp(-\frac{S}{\theta}) S^n I(0 < z_1 < z_2 < .. < z_n < 1) I(S > 0) dS \\ &= n! I(0 < z_1 < z_2 < .. < z_n < 1), \qquad \text{Gamma integral} \end{aligned}$$

which is the joint density of order statistics of n uniform (0,1) random variables.

1.2 Lemma

Let $Q_i \sim \chi^2_{k_i}(\lambda_i)$ for $i = 1, \dots, n$, be independent. Then, $Q = \sum_{i=1}^n Q_i$ is a noncentral $\chi^2_k(\lambda)$, where $k = \sum_{i=1}^n k_i$ and $\lambda = \sum_{i=1}^n \lambda_i$.

Proof:

The distribution transformation use moment generating function.

From the joint density of $(Y_1, ..., Y_{n+1})'$,

$$\prod_{i=1}^{n+1} \frac{1}{\theta} \exp\{-\frac{1}{\theta} y_i\} I(0 < y_i < \infty) = (\frac{1}{\theta})^{n+1} \exp\{-\frac{\sum_{i=1}^{n+1} y_i}{\theta}\} I(0 < y_i < \infty)$$

we obtain the joint density of $(W_1, ..., W_n, S)$ as

$$(\frac{1}{\theta})^{n+1} \exp\{-\frac{s}{\theta}\} I(0 < w_1 < w_2 < ... < w_n < s < \infty),$$

Let

$$Z_1 = W_1/S, Z_2 = W_2/S, ..., Z_n = W_n/S, S.$$

Consider the transformation $(W_1, ..., W_n, S)' \mapsto (Z_1, ..., Z_n, S)'$. Note that the formation is one-to-one with the Jacobian

$$|det \left(\frac{\partial(W_1, ..., W_n, S)}{\partial(Z_1, ..., Z_n, S)} \right)| = |det \begin{pmatrix} s & 0 & \dots & 0 & z_1 \\ 0 & s & \dots & 0 & z_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & s & z_n \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}| = s^n$$

From the joint density of $(Z_1, ..., Z_n, S)'$,

$$= (\frac{1}{\theta})^{n+1} \exp\{-\frac{s}{\theta}\} s^n I(0 < z_1 < z_2 < ... < z_n < 1) I(s > 0)$$

Figure 2: Alt method 2

1.2.1 Moment Generating Function

We can get MGF from $E[x^2t]$

$$
\begin{aligned}
M_i(t) = E[x^2t] &= \frac{1}{\sqrt{2\pi}} \int exp(x^2t)exp\left(-\frac{(x-\mu)^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int exp\left((t-\frac{1}{2})x^2 + \mu x - \frac{\mu^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int exp\left(-\frac{1}{2}(1-2t)\{x^2 - \frac{2\mu x}{(1-2t)} + \frac{\mu^2}{(1-2t)^2}\} + \frac{\mu^2}{2(1-2t)} - \frac{\mu^2}{2}\right) dx \\
&= \frac{1}{\sqrt{(1-2t)}} \int \frac{(1-2t)}{\sqrt{2\pi}} exp\left(-\frac{(x-\frac{\mu}{1-2t})^2}{2(1-2t)^{-1}}\right) dx \left[exp\left(\frac{\mu^2 t}{1-2t}\right)\right] \\
&= \frac{1}{\sqrt{(1-2t)}} exp\left(\frac{\mu^2 t}{1-2t}\right), \qquad \lambda = \mu^2 \\
&= \frac{1}{\sqrt{(1-2t)}} exp\left(\frac{\lambda t}{1-2t}\right)
\end{aligned}
$$

Then the MGF for $Q_i \sim \chi^2_{k_i}(\lambda_i)$

$$
\begin{aligned}
M(t) = E[\sum_{i=1}^{k} x_i^2 t] &= \prod_{i=1}^{k} M_i(t) \\
&= \left(\frac{1}{\sqrt{(1-2t)}}\right)^k exp\left(\frac{\sum_{i=1}^{k} \lambda_i t}{1-2t}\right) \\
&= \left(\frac{1}{\sqrt{(1-2t)}}\right)^k exp\left(\frac{\lambda t}{1-2t}\right) \\
&= (1-2t)^{-k/2} exp\left(\frac{\lambda t}{1-2t}\right), \qquad \text{i.i.d}
\end{aligned}
$$

The general case of a linear combination of independent $\chi^2_{k_i}(\lambda_i)$

$$
Q = \sum_{i=1}^{k} a_i Q_i
$$

We also can prove using MGF.

1.2.2 Linear Combination of Chi-Square Distribution

The linear combination of chi-square distribution Y_j. Let us denote by $X \sim \Gamma(r, \lambda)$ the fact that the r.v. X has a Gamma distribution with shape parameter r and rate parameter λ

$$f_X(x) = \frac{\lambda^x}{\Gamma(r)} exp(-\lambda x) x^{r-1}, \qquad (r, \lambda > 0, x > 0)$$

Then we have, for $j = 1, ..p$,

$$Y_j \sim \Gamma(\frac{k_j}{2}, \frac{1}{2}) \to Z_j = w_j Y_j \sim \Gamma(\frac{k_j}{2}, \frac{1}{2w_j})$$

The MGF for linear combinations $Z_j = w_j Y_j$

$$\begin{aligned} M(t) &= E[exp(Y_j t)] = (1-2t)^{-k/2} exp\left(\frac{\lambda t}{1-2t}\right) \\ M_{Z_j}(t) &= E[exp(w_j Y_j t)] = E[exp(Y_j(w_j t))] \\ &= (1-2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1-2w_j t}\right) \end{aligned}$$

$$\begin{aligned} M_Y(t) &= E[exp(Yt)] = E[exp(t[w_1 Y_1 + w_2 Y_2 + w_3 Y_3 + ..w_n Y_n])] \\ &= E[exp(w_1 t Y_1)] E[exp(w_2 t Y_2)] ... E[exp(w_n t Y_n)] \\ &= M_{X_1}(w_1 t) M_{X_2}(w_2 t) M_{X_3}(w_3 t) .. M_{X_n}(w_n t) \\ &= \prod_{i=1}^{n} M_{X_i}(w_i t) \end{aligned}$$

The third equation comes from the properties of exponents, as wells as from the expectation of the product of functions of independent random variables.

I need to pay attention that, only under independent and identical situation, we can write

$$M_Y(t) = M_X(t)^n$$

Other than that, we can not further simplify that. So back to the non-central chi-square distribution, we have the MGF of Y

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(w_i t) \\ &= \prod_{i=1}^{n} (1-2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1-2w_j t}\right) \end{aligned}$$

Then we can see that the shape parameter is $\frac{1}{2w_i}$. If we want to have a non-central chi-square distribution for Y, then all w_j need to be the same.

1.3 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

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

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

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

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

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

The score function and observed fisher information is

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

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

$$\begin{aligned}
\psi &= log\frac{\mu_1}{\mu_2} \\
\lambda &= log\mu_2
\end{aligned}$$

(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

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

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

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