Computational Statistics

Assignment 2

Hanbin Liu 11912410

2.15

(a)

The pmf is given by

$$f_Y(y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, 1, ..., n$$

Then

$$\begin{aligned}
L(\theta|Y_{\text{obs}}) &= f_Y(y) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \\
l(\theta|Y_{\text{obs}}) &= \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta) \\
\nabla l(\theta|Y_{\text{obs}}) &= \frac{y}{\theta} - \frac{n-y}{1-\theta} \\
I(\theta|Y_{\text{obs}}) &= -\nabla^2 l(\theta|Y_{\text{obs}}) = \frac{y}{\theta^2} + \frac{n-y}{(1-\theta)^2} \\
J(\theta) &= E\{I(\theta|Y_{\text{obs}})\} = \frac{E(y)}{\theta^2} + \frac{E(n-y)}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}
\end{aligned}$$

(b)

The pmf is given by

$$f_Y(y) = e^{-\theta}\frac{\theta^y}{y!}, \quad y = 0, 1, ..., \infty$$

Then

$$\begin{aligned}
L(\theta|Y_{\text{obs}}) &= f_Y(y) = e^{-\theta}\frac{\theta^y}{y!} \\
l(\theta|Y_{\text{obs}}) &= y\log\theta - \log y! - \theta \\
\nabla l(\theta|Y_{\text{obs}}) &= \frac{y}{\theta} - 1 \\
I(\theta|Y_{\text{obs}}) &= -\nabla^2 l(\theta|Y_{\text{obs}}) = \frac{y}{\theta^2} \\
J(\theta) &= E\{I(\theta|Y_{\text{obs}})\} = \frac{E(y)}{\theta^2} = \frac{1}{\theta}
\end{aligned}$$

(c)

The pmf is given by

$$f_Y(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, \quad y \geq 0$$

Then

$$\begin{aligned}
L(\theta|Y_{\text{obs}}) &= f_Y(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}} \\
l(\theta|Y_{\text{obs}}) &= -\log\theta - \frac{y}{\theta} \\
\nabla l(\theta|Y_{\text{obs}}) &= \frac{y}{\theta^2} - \frac{1}{\theta} \\
I(\theta|Y_{\text{obs}}) &= -\nabla^2 l(\theta|Y_{\text{obs}}) = \frac{2y}{\theta^3} - \frac{1}{\theta^2} \\
J(\theta) &= E\{I(\theta|Y_{\text{obs}})\} = \frac{E(2y)}{\theta^3} - \frac{1}{\theta^2} = \frac{2\theta}{\theta^3} - \frac{1}{\theta^2} = \frac{1}{\theta^2}
\end{aligned}$$

(d)

The pmf is given by

$$f_Y(y) = \binom{N}{y_1, ..., y_n} \prod_{i=1}^{n} \theta_i^{y_i}, \quad y_i \geq 0, \quad \sum_{i=1}^{n} y_i = N$$

Then

$$\begin{aligned}
L(\theta|Y_{\text{obs}}) &= f_Y(y) = \binom{N}{y_1, ..., y_n} \prod_{i=1}^{n} \theta_i^{y_i} \\
l(\theta|Y_{\text{obs}}) &= \log \binom{N}{y_1, ..., y_n} + \sum_{i=1}^{n} y_i \log\theta_i
\end{aligned}$$

Since $\theta_n = 1 - \theta_1 - ... - \theta_{n-1}$ and $E(y_i) = N\theta_i$, it then follows that

$$\begin{aligned}
\nabla l(\theta|Y_{\text{obs}}) &= \begin{pmatrix} \frac{\partial l(\theta|Y_{\text{obs}})}{\partial\theta_1} \\ \vdots \\ \frac{\partial l(\theta|Y_{\text{obs}})}{\partial\theta_{n-1}} \end{pmatrix} = \begin{pmatrix} \frac{y_1}{\theta_1} - \frac{y_n}{\theta_n} \\ \vdots \\ \frac{y_{n-1}}{\theta_{n-1}} - \frac{y_n}{\theta_n} \end{pmatrix} \\
I(\theta|Y_{\text{obs}}) = -\nabla^2 l(\theta|Y_{\text{obs}}) &= \begin{pmatrix} -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_1^2} & -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_1\partial\theta_2} & \cdots & -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_1\partial\theta_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_{n-1}\partial\theta_1} & -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_{n-1}\partial\theta_2} & \cdots & -\frac{\partial^2 l(\theta|Y_{\text{obs}})}{\partial\theta_{n-1}^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{y_1}{\theta_1^2} + \frac{y_n}{\theta_n^2} & \frac{y_n}{\theta_n^2} & \cdots & \frac{y_n}{\theta_n^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{y_n}{\theta_n^2} & \frac{y_n}{\theta_n^2} & \cdots & \frac{y_{n-1}}{\theta_{n-1}^2} + \frac{y_n}{\theta_n^2} \end{pmatrix} \\
&= \text{diag}(\frac{y_1}{\theta_1^2}, ..., \frac{y_{n-1}}{\theta_{n-1}^2}) + \frac{y_n}{\theta_n^2}\mathbf{J} \\
J(\theta) = E\{I(\theta|Y_{\text{obs}})\} &= \text{diag}(\frac{E(y_1)}{\theta_1^2}, ..., \frac{E(y_{n-1})}{\theta_{n-1}^2}) + \frac{E(y_n)}{\theta_n^2}\mathbf{J} \\
&= \text{diag}(\frac{N}{\theta_1}, ..., \frac{N}{\theta_{n-1}}) + \frac{N}{\theta_n}\mathbf{J},
\end{aligned}$$

where

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{(n-1)\times(n-1)}$$

□

2.16

The likelihood function is

$$\begin{aligned}L(\theta,\sigma^2|x) &= \prod_{i=1}^n (2\pi\sigma^2/w_i)^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2/w_i}(x_i-\mu_i(\theta))^2\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\Big(\prod_{i=1}^n w_i^{\frac{1}{2}}\Big)\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2\}\end{aligned}$$

Then

$$\begin{aligned}
l(\theta,\sigma^2|x) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \frac{1}{2}\sum_{i=1}^n \log w_i - \frac{1}{2\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2 \\
\nabla l(\theta,\sigma^2|x) &= \begin{pmatrix} \frac{\partial l}{\partial\theta_1} \\ \vdots \\ \frac{\partial l}{\partial\theta_n} \\ \frac{\partial l}{\partial\sigma^2}\end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))\frac{\partial\mu_i(\theta)}{\partial\theta_1} \\ \vdots \\ \frac{1}{\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))\frac{\partial\mu_i(\theta)}{\partial\theta_n} \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2\end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))\nabla\mu_i(\theta) \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2\end{pmatrix} \\
I(\theta,\sigma^2|x) &= -\nabla^2 l(\theta,\sigma^2|x) \\
&= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n w_i[\nabla\mu_i(\theta)(\nabla\mu_i(\theta))^\top - (x_i-\mu_i(\theta))\nabla^2\mu_i(\theta)] & \frac{1}{\sigma^4}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))\nabla\mu_i(\theta) \\ \frac{1}{\sigma^4}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))(\nabla\mu_i(\theta))^\top & -\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2\end{pmatrix} \\
J(\theta,\sigma^2) &= E(I(\theta,\sigma^2|x)) = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n w_i[\nabla\mu_i(\theta)(\nabla\mu_i(\theta))^\top & \mathbf{0} \\ \mathbf{0}^\top & \frac{n}{2\sigma^4}\end{pmatrix}
\end{aligned}$$

Sovling $\nabla l(\theta,\sigma^2|x)=0$, yields that $\sigma^2 = \frac{1}{n}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))^2$. Thus, we do not need to use Fisher scoring algorithm to find the MLE of σ^2. If we find the MLEs of θ, then the MLE of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n w_i(x_i-\mu_i(\hat{\theta}))^2$$

We use the Fisher scoring algorithm to obtain the MLEs of θ. Note that $J(\theta,\sigma^2)$ is a block matrix, if we let

$$\begin{aligned}a(\theta) &= \frac{1}{\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta))\nabla\mu_i(\theta) \\ \mathbf{A}(\theta) &= \frac{1}{\sigma^2}\sum_{i=1}^n w_i[\nabla\mu_i(\theta)(\nabla\mu_i(\theta))^\top\end{aligned}$$

then

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} + \mathbf{A}^{-1}(\theta^{(t)})a(\theta^{(t)}) \\ &= \theta^{(t)} + \Big[\frac{1}{\sigma^2}\sum_{i=1}^n w_i\nabla\mu_i(\theta^{(t)})(\nabla\mu_i(\theta^{(t)}))^\top\Big]^{-1}\frac{1}{\sigma^2}\sum_{i=1}^n w_i(x_i-\mu_i(\theta^{(t)}))\nabla\mu_i(\theta^{(t)}) \\ &= \theta^{(t)} + \Big[\sum_{i=1}^n w_i\nabla\mu_i(\theta^{(t)})(\nabla\mu_i(\theta^{(t)}))^\top\Big]^{-1}\sum_{i=1}^n w_i(x_i-\mu_i(\theta^{(t)}))\nabla\mu_i(\theta^{(t)})\end{aligned}$$

□

2.17

(a)

The likelihood and log-likehood function are given by

$$L(\beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$$
$$l(\beta) = \sum_{i=1}^{n} y_i \log p_i + (1-y_i)\log(1-p_i)$$

Since $p_i = \Phi(x_{(i)}^\top \beta)$, we have

$$\frac{\partial p_i}{\partial \beta} = \phi(x_{(i)}^\top \beta) x_{(i)},$$

where $\phi(\cdot)$ is the pdf of $N(0,1)$ and $\phi'(x) = -x\phi(x)$. Thus,

$$\nabla l(\beta) = \sum_{i=1}^{n} (\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}) \phi(x_{(i)}^\top \beta) x_{(i)}$$
$$I(\beta) = -\nabla^2 l(\beta) = \sum_{i=1}^{n} \left[\frac{y_i}{p_i^2} + \frac{1-y_i}{(1-p_i)^2}\right] (\phi(x_{(i)}^\top \beta))^2 x_{(i)} x_{(i)}^\top + \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right) x_{(i)}^\top \beta \phi(x_{(i)}^\top \beta) x_{(i)} x_{(i)}^\top$$

(b)

The Newton–Raphson algorithm is

$$\beta^{(t+1)} = \beta^{(t)} + I^{-1}(\beta^{(t)}) \nabla l(\beta^{(t)})$$

The expected information matrix is given by

$$J(\beta) = E(I(\beta)) = \sum_{i=1}^{n} \left[\frac{1}{p_i(1-p_i)}\right] (\phi(x_{(i)}^\top \beta))^2 x_{(i)} x_{(i)}^\top$$

Therefore, the estimated asymptotic covariance matrix of $\hat{\beta}$ is

$$\hat{\text{Cov}}(\hat{\beta}) = J^{-1}(\hat{\beta})$$

or

$$\hat{\text{Cov}}(\hat{\beta}) = I^{-1}(\hat{\beta})$$

□

2.19

To split the term $(\theta_1+\theta_2), (\theta_3+\theta_4), (\theta_1+\theta_3), (\theta_2+\theta_4)$ we introduce latent variables such that

$$n_{12} = Z_1 + Z_2 \quad n_{34} = Z_3 + Z_4$$
$$n_{13} = W_1 + W_3 \quad n_{24} = W_2 + W_4$$

then the complete-data likelihood is given by

$$L(\theta | Y_{\text{obs}}, Z) \propto \prod_{i=1}^{4} \theta_i^{n_i + Z_i + W_i}$$

where $Z = (Z_1, Z_3, W_1, W_2)^\top$ is the latent vector. The MLEs of θ based on the complete data are given by

$$\hat{\theta}_i = \frac{n_i + Z_i + W_i}{N}, \ i = 1, 2, 3, 4 \tag{1}$$

where $N = n_1 + n_2 + n_3 + n_4 + n_{12} + n_{34} + n_{13} + n_{24}$. The conditional predictive distribution is given by

$$\begin{aligned} f(Z|Y_{\text{obs}}, \theta) = \text{Binomial}(Z_1|n_{12}, \frac{\theta_1}{\theta_1 + \theta_2}) \times \text{Binomial}(Z_3|n_{34}, \frac{\theta_3}{\theta_3 + \theta_4}) \\ \times \text{Binomial}(W_1|n_{13}, \frac{\theta_1}{\theta_1 + \theta_3}) \times \text{Binomial}(W_2|n_{24}, \frac{\theta_2}{\theta_2 + \theta_4}) \end{aligned}$$

E-step:

$$\begin{aligned} E(Z_1|Y_{\text{obs}}, \theta^{(t)}) &= \frac{n_{12}\theta_1^{(t)}}{\theta_1^{(t)} + \theta_2^{(t)}}, \ E(Z_3|Y_{\text{obs}}, \theta^{(t)}) = \frac{n_{34}\theta_3^{(t)}}{\theta_3^{(t)} + \theta_4^{(t)}} \\ E(W_1|Y_{\text{obs}}, \theta^{(t)}) &= \frac{n_{13}\theta_1^{(t)}}{\theta_1^{(t)} + \theta_3^{(t)}}, \ E(W_2|Y_{\text{obs}}, \theta^{(t)}) = \frac{n_{24}\theta_2^{(t)}}{\theta_2^{(t)} + \theta_4^{(t)}} \end{aligned}$$

M-step: By replacing Z_i, W_i with $E(Z_i|Y_{\text{obs}}, \theta^{(t)}), E(W_i|Y_{\text{obs}}, \theta^{(t)})$ in (1), we have

$$\begin{aligned} \theta_1^{(t+1)} &= \frac{n_1 + n_{12}\frac{\theta_1^{(t)}}{\theta_1^{(t)}+\theta_2^{(t)}} + n_{13}\frac{\theta_1^{(t)}}{\theta_1^{(t)}+\theta_3^{(t)}}}{N} \\ \theta_2^{(t+1)} &= \frac{n_2 + n_{12}\frac{\theta_2^{(t)}}{\theta_1^{(t)}+\theta_2^{(t)}} + n_{24}\frac{\theta_2^{(t)}}{\theta_2^{(t)}+\theta_4^{(t)}}}{N} \\ \theta_3^{(t+1)} &= \frac{n_3 + n_{34}\frac{\theta_3^{(t)}}{\theta_3^{(t)}+\theta_4^{(t)}} + n_{13}\frac{\theta_3^{(t)}}{\theta_1^{(t)}+\theta_3^{(t)}}}{N} \\ \theta_4^{(t+1)} &= 1 - \theta_1^{(t+1)} - \theta_2^{(t+1)} - \theta_3^{(t+1)} \end{aligned}$$

□

2.20

(a)

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $\Delta = \frac{1}{n}\sum_{i=1}^n x_i^2$. Then

$$\begin{cases} \bar{x} = E(X) = E(Y) + E(U) = \theta + \lambda, \\ \Delta = E(X^2) = E(Y^2) + E(U^2) + 2E(Y)E(U) = \theta + \lambda^2 + \lambda + 2\theta\lambda. \end{cases}$$

Solving the equation yields that

$$\hat{\theta}^M = \sqrt{(\bar{x})^2 + \bar{x} - \Delta}, \quad \hat{\lambda}^M = \bar{x} - \sqrt{(\bar{x})^2 + \bar{x} - \Delta}.$$

(b)

If $x = 0$, then

$$\Pr(X = 0) = \Pr(Y = 0, U = 0) = \Pr(U = 0)\Pr(Y = 0) = (1 - \theta)e^{-\lambda}.$$

If $x \geq 1$, then

$$\begin{aligned}
\Pr(X = x) &= \Pr(Y + U = x) = \Pr(Y = 1, U = x - 1) + \Pr(Y = 0, U = x) \\
&= \Pr(Y = 1)\Pr(U = x - 1) + \Pr(Y = 0)\Pr(U = x) \\
&= \theta e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} + (1 - \theta)e^{-\lambda}\frac{\lambda^x}{x!}.
\end{aligned}$$

(c)

If $x = 0$, then

$$\Pr(Y = y \mid X = x) = \Pr(Y = 0 \mid X = 0) = 1.$$

That is, the conditional distribution of Y given $X = 0$ is Denegerate(0).
If $x \geq 1$, then

$$\begin{aligned}
\Pr(Y = 1 \mid X = x) &= \frac{\Pr(Y = 1, X = x)}{\Pr(X = x)} = \frac{\Pr(Y = 1)\Pr(U = x - 1)}{\Pr(X = x)} \\
&= \frac{\theta e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!}}{\theta e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} + (1 - \theta)e^{-\lambda}\frac{\lambda^x}{x!}} \\
&= \frac{\theta x}{\theta x + \lambda - \theta\lambda},
\end{aligned}$$

and

$$\begin{aligned}
\Pr(Y = 0 \mid X = x) &= \frac{\Pr(Y = 0, X = x)}{\Pr(X = x)} = \frac{\Pr(Y = 0)\Pr(U = x)}{\Pr(X = x)} \\
&= \frac{(1 - \theta)e^{-\lambda}\frac{\lambda^x}{x!}}{\theta e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} + (1 - \theta)e^{-\lambda}\frac{\lambda^x}{x!}} \\
&= \frac{\lambda - \theta\lambda}{\theta x + \lambda - \theta\lambda}.
\end{aligned}$$

That is,

$$Y|(X = x) \sim \text{Berboulli}\Big(\frac{\theta x}{\theta x + \lambda - \theta\lambda}\Big).$$

(d)

The observed data are $Y_{\text{obs}} = \{x_i, i = 1, ..., n\}$. The complete data are $Y_{\text{com}} = \{(Y_i, U_i), i = 1, ..., n\}$. The complete-data likelihood is given by

$$L(\theta, \lambda|Y_{\text{com}}) = \prod_{i=1}^{n} \theta^{y_i}(1 - \theta)^{1-y_i}e^{-\lambda}\frac{\lambda^{u_i}}{u_i!}$$

so that the MLEs of (θ, λ) based on the complete data are

$$\hat{\theta} = \bar{y}, \ \hat{\lambda} = \bar{u} = \bar{x} - \bar{y} \tag{2}$$

The conditional predictive distribution has been obtained in (c).
E-step:

$$E(Y_i|x_i, \theta, \lambda) = \begin{cases} 0, & x_i = 0 \\ \frac{\theta x_i}{\theta x_i + \lambda - \theta\lambda}, & x_i \geq 1 \end{cases}$$

Since $\frac{\theta x_i}{\theta x_i+\lambda-\theta\lambda} = 0$ for $x_i = 0$, we can write

$$E(Y_i|x_i, \theta, \lambda) = \frac{\theta x_i}{\theta x_i + \lambda - \theta\lambda}$$

M-step: Replacing y_i with $E(Y_i|x_i, \theta^{(t)}, \lambda^{(t)})$ in (2), we have

$$\begin{aligned}
\theta^{(t+1)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\theta^{(t)} x_i}{\theta^{(t)} x_i + \lambda^{(t)} - \theta^{(t)}\lambda^{(t)}} \\
\lambda^{(t+1)} &= \bar{x} - \theta^{(t+1)}
\end{aligned}$$

□

2.21

(a)

$Y \sim \text{ZIP}(\phi, \lambda)$, then the pmf of Y is given by

$$\Pr(Y = y) = \begin{cases} \phi + (1-\phi)e^{-\lambda}, & y = 0 \\ (1-\phi)e^{-\lambda}\frac{\lambda^y}{y!}, & y \geq 1. \end{cases}$$

If $y = 0$, then

$$\Pr(Z = 1|Y = 0) = \frac{\Pr(Z = 1, Y = 0)}{\Pr(Y = 0)} = \frac{\Pr(Z = 1)\Pr(X = 0)}{\phi + (1-\phi)e^{-\lambda}} = \frac{(1-\phi)e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}},$$

and

$$\Pr(Z = 0|Y = 0) = \frac{\Pr(Z = 0, Y = 0)}{\Pr(Y = 0)} = \frac{\Pr(Z = 0)}{\phi + (1-\phi)e^{-\lambda}} = \frac{\phi}{\phi + (1-\phi)e^{-\lambda}}.$$

That is,

$$Z|(Y = 0) \sim \text{Bernoulli}\Big(\frac{(1-\phi)e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}}\Big).$$

If $y \geq 1$, then

$$\Pr(Z = 1|Y = y) = \frac{\Pr(Z = 1, Y = y)}{\Pr(Y = y)} = \frac{\Pr(Z = 1, X = y)}{\Pr(Y = y)} = \frac{(1-\phi)e^{-\lambda}\frac{\lambda^y}{y!}}{(1-\phi)e^{-\lambda}\frac{\lambda^y}{y!}} = 1.$$

That is,

$$Z|(Y = y) \sim \text{Denegerate}(1).$$

Therefore,

$$Z|(Y = y) \sim \begin{cases} \text{Bernoulli}\Big(\frac{(1-\phi)e^{-\lambda}}{\phi+(1-\phi)e^{-\lambda}}\Big), & y = 0, \\ \text{Denegerate}(1), & y \geq 1. \end{cases}$$

(b)

If $y = 0$, then

$$\Pr(X = x|Y = 0) = \frac{\Pr(X = x, Y = 0)}{\Pr(Y = 0)} = \begin{cases} \frac{e^{-\lambda}}{\phi+(1-\phi)e^{-\lambda}}, & x = 0, \\ \frac{\phi e^{-\lambda}\frac{\lambda^x}{x!}}{\phi+(1-\phi)e^{-\lambda}}, & x \geq 1. \end{cases}$$

If $y \geq 1$, then

$$\Pr(X = y|Y = y) = \frac{\Pr(X = y, Y = y)}{\Pr(Y = y)} = \frac{\Pr(X = y, Z = 1)}{\Pr(Y = y)} = \frac{(1-\phi)e^{-\lambda}\frac{\lambda^y}{y!}}{(1-\phi)e^{-\lambda}\frac{\lambda^y}{y!}} = 1.$$

Therefore,

$$X|(Y = y) \sim \begin{cases} \text{ZIP}(\phi_0, \lambda), & y = 0, \\ \text{Denegerate}(y), & y \geq 1, \end{cases}$$

where $\phi_0 = \frac{(1-\phi)e^{-\lambda}}{\phi+(1-\phi)e^{-\lambda}}$.

(c)

$$E(Y) = E(ZX) = E(Z)E(X) = (1-\phi)\lambda.$$

(d)

The complete-data likelihood function is given by

$$L(\phi, \lambda|Y_{\text{com}}) = e^{-n\lambda}\prod_{i=1}^{n}(1-\phi)^{z_i}\phi^{1-z_i}\frac{\lambda^{x_i}}{x_i!}$$

M-Step Thus, the complete-data MLEs are

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} \quad \hat{\phi} = \frac{n - \sum_{i=1}^n z_i}{n}$$

E-Step

$$E(z_i|y_i, \phi, \lambda) = p_0 \cdot I(y_i = 0) + I(y_i \geq 1)$$
$$E(x_i|y_i, \phi, \lambda) = (1-p_0)\lambda \cdot I(y_i = 0) + y_i \cdot I(y_i \geq 1)$$

where $p_0 = \frac{(1-\phi)e^{-\lambda}}{\phi+(1-\phi)e^{-\lambda}}$. Then

$$\sum_{i=1}^n E(z_i|y_i, \phi, \lambda) = p_0 m + (n-m) \quad \sum_{i=1}^n E(x_i|y_i, \phi, \lambda) = (1-p_0)\lambda m + N$$

where $m = \#$ of $y_i = 0$ and $N = \sum_{i=1}^n y_i$.
Therefore, the iteration for ϕ is

$$\phi^{(t+1)} = \frac{n - \sum_{i=1}^n E(z_i|y_i, \phi^{(t)}, \lambda^{(t)})}{n} = (1-p_0^{(t)}) \cdot \frac{m}{n} = \frac{\phi^{(t)}}{\phi^{(t)} + (1-\phi^{(t)})e^{-\lambda^{(t)}}} \cdot \frac{m}{n}.$$

Note that $\lambda = \frac{\sum_{i=1}^n E(x_i|y_i,\phi,\lambda)}{n} = \frac{(1-p_0)\lambda m + N}{n}$ yields that $\lambda = \frac{N/n}{1-(1-p_0)\cdot\frac{m}{n}} = \frac{N/n}{1-\phi}$. The iteration for λ is

$$\lambda^{(t+1)} = \frac{N/n}{1-\phi^{(t+1)}}$$

Iteration

$$\phi^{(t+1)} = \frac{\phi^{(t)}}{\phi^{(t)} + (1-\phi^{(t)})e^{-\lambda^{(t)}}} \cdot \frac{m}{n}$$
$$\lambda^{(t+1)} = \frac{N/n}{1-\phi^{(t+1)}}$$

□

2.22

(a)

The pmf of ZX is given by

$$\Pr(ZX = k) = \begin{cases} \Pr(ZX = 0) = \Pr(Z = 0) = e^{-\lambda}, & k = 0, \\ \Pr(ZX = k) = \Pr(Z = 1, X = k) = (1 - e^{-\lambda})c \cdot \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\frac{\lambda^k}{k!}, & k \geq 1. \end{cases}$$

Thus, the distribution of ZX is Poisson(λ), i.e.,$Y \overset{d}{=} ZX$.

(b)

(b1)

Introduce the latent variables $Z_i, i = 1, ..., n$, then $Y_{\text{com}} = \{y_i, i = 1, ..., n\} = \{(z_i, x_i), i = 1, ..., n\}$. The complete-data likelihood function is given by

$$L(\lambda|Y_{\text{com}}) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}$$

so that the MLE of λ based on the complete data is

$$\hat{\lambda} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i z_i \tag{3}$$

E-step: Since Z_i and X_i are independent, $E(Z_i|Y_{\text{obs}}, \lambda) = E(Z_i|X_i, \lambda) = E(Z_i|\lambda) = 1 - e^{-\lambda}$.
M-step: Replacing z_i with $E(Z_i|Y_{\text{obs}}, \lambda)$ in (3), we have

$$\lambda^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} x_i(1 - e^{-\lambda^{(t)}}) = \bar{x}(1 - e^{-\lambda^{(t)}})$$

(b2)

The observed-data likelihood is

$$L(\lambda|Y_{\text{obs}}) = \prod_{i=1}^{n} c\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}$$

so that the observed-data log-likehood is

$$\begin{aligned} l(\lambda|Y_{\text{obs}}) &= n\log ce^{-\lambda} + \sum_{i=1}^{n} x_i \log\lambda - \log x_i! \\ &= n(\log e^{-\lambda} - \log(1 - e^{-\lambda})) + \sum_{i=1}^{n} x_i \log\lambda - \log x_i! \\ &= n\{\bar{x}\log\lambda - \lambda + g(\lambda)\} - \sum_{i=1}^{n} \log x_i!, \end{aligned}$$

where $g(\lambda) = -\log(1 - e^{-\lambda})$.

(b3)

It suffices to show that $g(\lambda),\ \lambda > 0$ is a convex function. For all $\lambda > 0$,

$$g'(\lambda) = \frac{-e^{-\lambda}}{1 - e^{-\lambda}}$$
$$g''(\lambda) = \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2} > 0$$

thus $g(\lambda)$ is a convex function, by Excercise 2.5(c), we have

$$g(\lambda) \geq g(\lambda_0) + (\lambda - \lambda_0) g'(\lambda_0), \forall \lambda, \lambda_0 > 0$$

(b4)

Let

$$Q(\lambda|\lambda^{(t)}) = n\Big[\bar{x}\log\lambda - \lambda + g(\lambda^{(t)}) + (\lambda - \lambda^{(t)})g'(\lambda^{(t)})\Big] - \sum_{i=1}^{n} \log x_i!$$

Then by (b3), we have

$$\begin{aligned} Q(\lambda|\lambda^{(t)}) &= n\Big[\bar{x}\log\lambda - \lambda + g(\lambda^{(t)}) + (\lambda - \lambda^{(t)})g'(\lambda^{(t)})\Big] - \sum_{i=1}^{n} \log x_i! \\ &\leq n\Big[\bar{x}\log\lambda - \lambda + g(\lambda)\Big] - \sum_{i=1}^{n} \log x_i! \\ &= l(\lambda|Y_{\text{obs}}) \end{aligned}$$

and $Q(\lambda^{(t)}|\lambda^{(t)}) = l(\lambda^{(t)}|Y_{\text{obs}})$. Thus, by MM algorithm, we have

$$\lambda^{(t+1)} = \underset{\lambda}{\text{argmax}}\ Q(\lambda|\lambda^{(t)})$$

Solving $\frac{dQ(\lambda|\lambda^{(t)})}{d\lambda} = 0$ yields that

$$\lambda^{(t+1)} = \bar{x}(1 - e^{-\lambda^{(t)}})$$

□

2.23

The pdf of $\mathbf{x}_i$ is

$$f_{\mathbf{x}_i}(x) = \frac{\prod_{j=1}^{n} x_{ij}^{a_j - 1}}{B_n(a)},$$

where $B_n(a) = (\prod_{j=1}^{n} \Gamma(a_j))/\Gamma(a^+)$ and $a^+ := \sum_{j=1}^{n} a_j$. Then the likelihood function of $\mathbf{x}_1, ..., \mathbf{x}_m$ is given by

$$L(a) = \prod_{i=1}^{m} f_{\mathbf{x}_i}(x) = \frac{\prod_{i=1}^{m}\prod_{j=1}^{n} x_{ij}^{a_j - 1}}{B_n^m(a)}$$

so that the log-likelihood is

$$\begin{aligned} l(a) &= -m \log B_n(a) + \sum_{i=1}^{m}\sum_{j=1}^{n}(a_j - 1)\log x_{ij} \\ &= m\Big[\log\Gamma(a^+) - \sum_{j=1}^{n}\log\Gamma(a_j)\Big] + \sum_{i=1}^{m}\sum_{j=1}^{n}(a_j - 1)\log x_{ij} \end{aligned}$$

Then

$$\nabla l(a) = \begin{pmatrix} \frac{\partial l(a)}{\partial a_1} \\ \frac{\partial l(a)}{\partial a_2} \\ \vdots \\ \frac{\partial l(a)}{\partial a_n} \end{pmatrix} = \begin{pmatrix} m\big[\psi(a^+) - \psi(a_1)\big] + \sum_{i=1}^m \log x_{i1} \\ m\big[\psi(a^+) - \psi(a_2)\big] + \sum_{i=1}^m \log x_{i2} \\ \vdots \\ m\big[\psi(a^+) - \psi(a_n)\big] + \sum_{i=1}^m \log x_{in} \end{pmatrix}$$

$$\begin{aligned} I(a) = -\nabla^2 l(a) &= \begin{pmatrix} -m\big[\psi'(a^+) - \psi'(a_1)\big] & \dots & -m\psi'(a^+) \\ -m\psi'(a^+) & \dots & -m\psi'(a^+) \\ \vdots & \ddots & \vdots \\ -m\psi'(a^+) & \dots & -m\big[\psi'(a^+) - \psi'(a_n)\big] \end{pmatrix} \\ &= m \cdot \mathrm{diag}\Big(\psi'(a_1), \dots, \psi'(a_n)\Big) - m \cdot \psi'(a^+)\mathbf{J} \end{aligned}$$

where $\psi(x)$ is the digamma function, $\psi'(x)$ is the trigamma function and

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{n \times n}$$

Therefore, the Newton–Raphson algorithm is

$$a^{(t+1)} = a^{(t)} + I^{-1}(a^{(t)})\nabla l(a^{(t)})$$

$\square$

2.24

(a)

Since

$$\mathrm{BBinomial}(x|n, \alpha, \beta) = \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)},$$

for $x = 0, 1, \dots, n$, it then follows that

$$L(\alpha, \beta) = \prod_{i=1}^m \Big[\binom{n_i}{x_i} \frac{B(x_i+\alpha, n_i-x_i+\beta)}{B(\alpha, \beta)}\Big].$$

Then the log-likelihood function is given by

$$
\begin{aligned}
l(\alpha,\beta) &= \sum_{i=1}^{m}\Big[\log\binom{n_i}{x_i} + \log B(x_i+\alpha, n_i-x_i+\beta) + \log B(\alpha,\beta)\Big] \\
&= c + \sum_{i=1}^{m}\Big[\log B(x_i+\alpha, n_i-x_i+\beta) - \log B(\alpha,\beta)\Big] \\
&= c + \sum_{i=1}^{m}\Big[\log\frac{(x_i+\alpha-1)!(n_i-x_i+\beta-1)!}{(n_i+\alpha+\beta-1)!} - \log\frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}\Big] \\
&= c + \sum_{i=1}^{m}\Big[\log\Big\{\frac{(x_i+\alpha-1)!(n_i-x_i+\beta-1)!}{(n_i+\alpha+\beta-1)!}\Big/\frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}\Big\}\Big] \\
&= c + \sum_{i=1}^{m}\Big[\log\frac{(x_i+\alpha-1)!}{(\alpha-1)!} + \log\frac{(n_i-x_i+\beta-1)!}{(\beta-1)!} - \log\frac{(n_i+\alpha+\beta-1)!}{(\alpha+\beta-1)!}\Big] \\
&= c + \sum_{i=1}^{m}\Big[\sum_{j=0}^{x_i-1}\log(\alpha+j) + \sum_{j=0}^{n_i-x_i-1}\log(\beta+j) - \sum_{j=0}^{n_i-1}\log(\alpha+\beta+j)\Big] \\
&= c + \sum_{i=1}^{m}\left\{\sum_{j=0}^{x_i-1}\log(\alpha+j) + \sum_{j=0}^{n_i-x_i-1}\log(\beta+j)\right\} - \sum_{i=1}^{m}\sum_{j=0}^{n_i-1}\log(\alpha+\beta+j),
\end{aligned}
$$

where c is a constant free from (α,β).

(b)

Since $\log(\cdot)$ is concave, $-\log(\cdot)$ is convex. Then

$$
\begin{aligned}
-\log(\alpha+j) &= -\log\left(\frac{\alpha^{(t)}}{\alpha^{(t)}+j}\cdot\frac{\alpha^{(t)}+j}{\alpha^{(t)}}\cdot\alpha + \frac{j}{\alpha^{(t)}+j}\cdot\frac{\alpha^{(t)}+j}{j}\cdot\right) \\
&\le -\frac{\alpha^{(t)}}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{\alpha^{(t)}}\alpha\right) - \frac{j}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{j}j\right),
\end{aligned}
$$

which is equivalent to

$$
\log(\alpha+j) \ge \frac{\alpha^{(t)}}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{\alpha^{(t)}}\alpha\right) + \frac{j}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{j}j\right)
$$

(c)

The support superplane inequality for $-\log(\cdot)$ is

$$
-\log(x) \ge -\log(x_0) - \frac{1}{x_0}(x-x_0)
$$

By replacing x and x_0 by $\alpha+\beta+j$ and $\alpha^{(t)}+\beta^{(t)}+j$ respectively, we have

$$
-\log(\alpha+\beta+j) \ge -\log(\alpha^{(t)}+\beta^{(t)}+j) - \frac{\alpha+\beta-\alpha^{(t)}-\beta^{(t)}}{\alpha^{(t)}+\beta^{(t)}+j}
$$

(d)

Define

$$Q(\alpha,\beta|\alpha^{(t)},\beta^{(t)}) = c + \sum_{i=1}^{m}\left\{\sum_{j=0}^{x_i-1} Q_1(\alpha|\alpha^{(t)}) + \sum_{j=0}^{n_i-x_i-1} Q_2(\beta|\beta^{(t)})\right\} + \sum_{i=1}^{m}\sum_{j=0}^{n_i-1} Q_3(\alpha,\beta|\alpha^{(t)},\beta^{(t)})$$

where

$$\begin{aligned}
Q_1(\alpha|\alpha^{(t)}) &= \frac{\alpha^{(t)}}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{\alpha^{(t)}}\alpha\right) + \frac{j}{\alpha^{(t)}+j}\log\left(\frac{\alpha^{(t)}+j}{j}j\right) \\
Q_1(\beta|\beta^{(t)}) &= \frac{\beta^{(t)}}{\beta^{(t)}+j}\log\left(\frac{\beta^{(t)}+j}{\beta^{(t)}}\beta\right) + \frac{j}{\beta^{(t)}+j}\log\left(\frac{\beta^{(t)}+j}{j}j\right) \\
Q_3(\alpha,\beta|\alpha^{(t)},\beta^{(t)}) &= -\log(\alpha^{(t)}+\beta^{(t)}+j) - \frac{\alpha+\beta-\alpha^{(t)}-\beta^{(t)}}{\alpha^{(t)}+\beta^{(t)}+j}
\end{aligned}$$

By (b) and (c), we have

$$\begin{aligned}
Q(\alpha,\beta|\alpha^{(t)},\beta^{(t)}) &= c + \sum_{i=1}^{m}\left\{\sum_{j=0}^{x_i-1} Q_1(\alpha|\alpha^{(t)}) + \sum_{j=0}^{n_i-x_i-1} Q_2(\beta|\beta^{(t)})\right\} + \sum_{i=1}^{m}\sum_{j=0}^{n_i-1} Q_3(\alpha,\beta|\alpha^{(t)},\beta^{(t)}) \\
&\le c + \sum_{i=1}^{m}\left\{\sum_{j=0}^{x_i-1}\log(\alpha+j) + \sum_{j=0}^{n_i-x_i-1}\log(\beta+j)\right\} - \sum_{i=1}^{m}\sum_{j=0}^{n_i-1}\log(\alpha+\beta+j) \\
&= l(\alpha,\beta)
\end{aligned}$$

and $Q(\alpha^{(t)},\beta^{(t)}|\alpha^{(t)},\beta^{(t)}) = l(\alpha^{(t)},\beta^{(t)})$. By MM algorithm, we have

$$(\alpha^{(t+1)},\beta^{(t+1)}) = \underset{\alpha,\beta}{\operatorname{argmax}}\ Q(\alpha,\beta|\alpha^{(t)},\beta^{(t)})$$

Solving

$$\begin{cases} \frac{\partial Q(\alpha,\beta|\alpha^{(t)},\beta^{(t)})}{\partial\alpha} = 0 \\ \frac{\partial Q(\alpha,\beta|\alpha^{(t)},\beta^{(t)})}{\partial\beta} = 0 \end{cases}$$

yields that

$$\begin{aligned}
\alpha^{(t+1)} &= \frac{\sum_{i=1}^{m}\sum_{j=0}^{x_i-1}\frac{\alpha^{(t)}}{\alpha^{(t)}+j}}{\sum_{i=1}^{m}\sum_{j=0}^{n_i-1}\frac{1}{\alpha^{(t)}+\beta^{(t)}+j}} \\
\beta^{(t+1)} &= \frac{\sum_{i=1}^{m}\sum_{j=0}^{n_i-x_i-1}\frac{\beta^{(t)}}{\beta^{(t)}+j}}{\sum_{i=1}^{m}\sum_{j=0}^{n_i-1}\frac{1}{\alpha^{(t)}+\beta^{(t)}+j}}
\end{aligned}$$

$\square$

2.28

Question Let $\{Y_i\}_{i=1}^n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i)$ and $\lambda_i = \Phi(x_{(i)}^\top \beta)$, where $\Phi(\cdot)$ is the cdf of $N(0,1)$, $x_{(i)}$ is a known vector of covariates for subject i, and $\beta_{(p+1)\times 1}$ is an unknown vector of parameters. Use the Newton–Raphson algorithm to find the MLEs $\hat{\beta}$ of β.
Solution The likelihood and log-likehood function are given by

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^n e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!} \\
l(\beta) &= -\sum_{i=1}^n \lambda_i + \sum_{i=1}^n y_i \log \lambda_i - \log y_i!
\end{aligned}
$$

Since $\lambda_i = \Phi(x_{(i)}^\top \beta)$, we have

$$
\frac{\partial \lambda_i}{\partial \beta} = \phi(x_{(i)}^\top \beta) x_{(i)},
$$

where $\phi(\cdot)$ is the pdf of $N(0,1)$ and $\phi'(x) = -x\phi(x)$. Thus,

$$
\begin{aligned}
\nabla l(\beta) &= -\sum_{i=1}^n \frac{\partial \lambda_i}{\partial \beta} + \sum_{i=1}^n \frac{y_i}{\lambda_i} \frac{\partial \lambda_i}{\partial \beta} \\
&= -\sum_{i=1}^n \phi(x_{(i)}^\top \beta) x_{(i)} + \sum_{i=1}^n \frac{y_i}{\lambda_i} \phi(x_{(i)}^\top \beta) x_{(i)} \\
I(\beta) = -\nabla^2 l(\beta) &= \sum_{i=1}^n \left[\frac{y_i}{\lambda_i^2} \phi^2(x_{(i)}^\top \beta) x_{(i)} x_{(i)}^\top + \frac{y_i - \lambda_i}{\lambda_i} \phi(x_{(i)}^\top \beta) x_{(i)}^\top \beta x_{(i)} x_{(i)}^\top \right]
\end{aligned}
$$

Therefore, the Newton-Raphson algorithm is

$$
\beta^{(t+1)} = \beta^{(t)} + I^{-1}(\beta^{(t)}) \nabla l(\beta^{(t)})
$$

□