MAT7035: Computational Statistics

Suggested Solutions to Assignment 2

2.15 Solution: (a) Let $Y \sim \text{Binomial}(n, \theta)$, the pmf of Y is

$$f(y;\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, 1, \ldots, n,$$

so that the log-likelihood function of θ is given by

$$\ell(\theta|y) = \log\left\{\binom{n}{y}\right\} + y\log(\theta) + (n-y)\log(1-\theta).$$

The score, the observed information, and the expected information are given by

$$\begin{aligned}
\ell'(\theta|y) &= \frac{y}{\theta} - \frac{n-y}{1-\theta} = \frac{y-n\theta}{\theta(1-\theta)}, \\
-\ell''(\theta|y) &= \frac{y}{\theta^2} + \frac{n-y}{(1-\theta)^2} = \frac{y(1-2\theta)+n\theta^2}{\theta^2(1-\theta)^2}, \\
J(\theta) &= E\{-\ell''(\theta|y)\} = \frac{E(Y)(1-2\theta)+n\theta^2}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.
\end{aligned}$$

(b) Let $Y \sim \text{Poisson}(\theta)$, the pmf of Y is

$$f(y;\theta) = \frac{\theta^y}{y!}\,\mathrm{e}^{-\theta}, \quad y = 0, 1, \ldots, \infty,$$

so that the log-likelihood function of θ is given by

$$\ell(\theta|y) = -\log(y!) + y\log(\theta) - \theta.$$

The score, the observed information, and the expected information are

$$\begin{aligned} \ell'(\theta|y) &= \frac{y}{\theta} - 1, \quad -\ell''(\theta|y) = \frac{y}{\theta^2}, \\ J(\theta) &= E\{-\ell''(\theta|y)\} = \frac{E(Y)}{\theta^2} = \frac{1}{\theta}. \end{aligned}$$

(c) Let $Y \sim \text{Exponential}(1/\theta)$, the pdf of Y is

$$f(y;\theta) = \frac{1}{\theta}\,\mathrm{e}^{-y/\theta}, \quad y \geqslant 0,$$

so that the log-likelihood function of θ is given by

$$\ell(\theta|y) = -\log(\theta) - y/\theta.$$

Since $E(Y) = \theta$, the score, the observed information, and the expected information are given by

$$\begin{aligned} \ell'(\theta|y) &= -\frac{1}{\theta} + \frac{y}{\theta^2}, \quad -\ell''(\theta|y) = -\frac{1}{\theta^2} + \frac{2y}{\theta^3}, \\ J(\theta) &= E\{-\ell''(\theta|y)\} = -\frac{1}{\theta^2} + \frac{2E(Y)}{\theta^3} = \frac{1}{\theta^2}. \end{aligned}$$

(d) Let $\mathbf{y} = (Y_1, \ldots, Y_n)^\top \sim \text{Multinomial}_n(N, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\top \in \mathbb{T}_n$. The pmf of $\mathbf{y}$ is

$$f(\boldsymbol{y};\boldsymbol{\theta}) = \binom{N}{y_1, \ldots, y_n} \prod_{i=1}^n \theta_i^{y_i}, \quad y_i \geqslant 0, \quad \sum_{i=1}^n y_i = N,$$

so that the log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}) = c + \sum_{i=1}^n y_i \log(\theta_i).$$

Note that $\theta_n = 1 - \theta_1 - \cdots - \theta_{n-1}$ and[1]

$$E(Y_i) = N\theta_i, \quad i = 1, \ldots, n,$$

[1]cf. Appendix A.1.7

the score vector, the observed information matrix, and the expected information matrix are given by

$$\nabla \ell(\boldsymbol{\theta}|\boldsymbol{y}) = \begin{pmatrix} \dfrac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_1} \\ \vdots \\ \dfrac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_{n-1}} \end{pmatrix} = \begin{pmatrix} \dfrac{y_1}{\theta_1} - \dfrac{y_n}{\theta_n} \\ \vdots \\ \dfrac{y_{n-1}}{\theta_{n-1}} - \dfrac{y_n}{\theta_n} \end{pmatrix},$$

$$\begin{aligned} -\nabla^2 \ell(\boldsymbol{\theta}|\boldsymbol{y}) &= \begin{pmatrix} -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_1^2} & -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_1 \partial \theta_2} & \cdots & -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_1 \partial \theta_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_{n-1} \partial \theta_1} & -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_{n-1} \partial \theta_2} & \cdots & -\dfrac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_{n-1}^2} \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{y_1}{\theta_1^2} + \dfrac{y_n}{\theta_n^2} & \dfrac{y_n}{\theta_n^2} & \cdots & \dfrac{y_n}{\theta_n^2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{y_n}{\theta_n^2} & \dfrac{y_n}{\theta_n^2} & \cdots & \dfrac{y_{n-1}}{\theta_{n-1}^2} + \dfrac{y_n}{\theta_n^2} \end{pmatrix} \\ &= \operatorname{diag}\left(\frac{y_1}{\theta_1^2}, \ldots, \frac{y_{n-1}}{\theta_{n-1}^2} \right) + \frac{y_n}{\theta_n^2} \mathbf{1}_n \mathbf{1}_n^\top, \end{aligned}$$

$$\begin{aligned} \boldsymbol{J}(\boldsymbol{\theta}) &= E\{-\nabla^2 \ell(\boldsymbol{\theta}|\boldsymbol{y})\} \\ &= \operatorname{diag}\left(\frac{E(Y_1)}{\theta_1^2}, \ldots, \frac{E(Y_{n-1})}{\theta_{n-1}^2} \right) + \frac{E(Y_n)}{\theta_n^2} \mathbf{1}_n \mathbf{1}_n^\top \\ &= N \left\{ \operatorname{diag}(1/\theta_1, \ldots, 1/\theta_{n-1}) + \theta_n^{-1} \mathbf{1}_n \mathbf{1}_n^\top \right\}. \end{aligned}$$

2.16 Solution: Let $\mathbf{x} = (X_1, \ldots, X_n)^\top$, then the joint density of $\mathbf{x}$ is

$$f(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2/w_i}} \exp\left[-\frac{\{x_i - \mu_i(\boldsymbol{\theta})\}^2}{2\sigma^2/w_i} \right]$$

so that the log-likelihood function for $(\boldsymbol{\theta}, \sigma^2)$ is given by

$$\ell(\boldsymbol{\theta}, \sigma^2|\boldsymbol{x}) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n w_i\{x_i - \mu_i(\boldsymbol{\theta})\}^2.$$

Since $E(X_i) = \mu_i(\boldsymbol{\theta})$ and $\mathrm{Var}(X_i) = \sigma^2/w_i$, the score, the observed information, and the expected information are given by

$$
\begin{aligned}
\nabla \ell(\boldsymbol{\theta}, \sigma^2|\boldsymbol{x}) &= \begin{pmatrix} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n} w_i\{x_i - \mu_i(\boldsymbol{\theta})\}\nabla\mu_i(\boldsymbol{\theta}) \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n} w_i\{x_i - \mu_i(\boldsymbol{\theta})\}^2 \end{pmatrix}, \\
-\nabla^2 \ell(\boldsymbol{\theta}, \sigma^2|\boldsymbol{x}) &= \begin{pmatrix} \boldsymbol{A} & \boldsymbol{a} \\ \boldsymbol{a}^\top & b \end{pmatrix}, \\
\boldsymbol{A} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i \left\{ \nabla\mu_i(\boldsymbol{\theta})[\nabla\mu_i(\boldsymbol{\theta})]^\top - [x_i - \mu_i(\boldsymbol{\theta})]\nabla^2\mu_i(\boldsymbol{\theta}) \right\}, \\
\boldsymbol{a} &= \frac{1}{\sigma^4}\sum_{i=1}^{n} w_i\{x_i - \mu_i(\boldsymbol{\theta})\}\nabla\mu_i(\boldsymbol{\theta}) \\
b &= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n} w_i\{x_i - \mu_i(\boldsymbol{\theta})\}^2, \\
\boldsymbol{J}(\boldsymbol{\theta}, \sigma^2) &= E\{-\nabla^2\ell(\boldsymbol{\theta}, \sigma^2|\boldsymbol{x})\} = \begin{pmatrix} E(\boldsymbol{A}) & E(\boldsymbol{a}) \\ E(\boldsymbol{a}^\top) & E(b) \end{pmatrix} \\
&= \begin{pmatrix} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n} w_i\nabla\mu_i(\boldsymbol{\theta})[\nabla\mu_i(\boldsymbol{\theta})]^\top & \mathbf{0}_q \\ \mathbf{0}_q^\top & \dfrac{n}{2\sigma^4} \end{pmatrix}.
\end{aligned}
$$

Note that $\boldsymbol{J}(\boldsymbol{\theta}, \sigma^2)$ is a block-diagonal matrix, then $\boldsymbol{\theta}$ and σ^2 can be estimated separately. In fact, let the second component of the score vector be equal to zero, we have the following explicit solution to σ^2 when $\boldsymbol{\theta}$ is given:

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

Namely, when $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, the MLE of σ^2 is given by

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

On the other hand, we can use the Fisher scoring algorithm to obtain the MLEs of $\boldsymbol{\theta}$ by: for $t \geqslant 0$,

$$\begin{aligned}
\boldsymbol{\theta}^{(t+1)} &= \boldsymbol{\theta}^{(t)} + \{\boldsymbol{J}(\boldsymbol{\theta}^{(t)})\}^{-1}\nabla\ell(\boldsymbol{\theta}^{(t)}|\boldsymbol{x}) \\
&= \boldsymbol{\theta}^{(t)} + \left[\sum_{i=1}^{n} w_i \nabla\mu_i(\boldsymbol{\theta}^{(t)})\{\nabla\mu_i(\boldsymbol{\theta}^{(t)})\}^{\top}\right]^{-1} \\
&\qquad \times \sum_{i=1}^{n} w_i\{x_i - \mu_i(\boldsymbol{\theta}^{(t)})\}\nabla\mu_i(\boldsymbol{\theta}^{(t)}).
\end{aligned}$$

2.17 Solution: (a) Let $\boldsymbol{y} = (y_1, \ldots, y_n)^{\top}$, the likelihood function and the log-likelihood function of $\boldsymbol{\beta}$ are given by

$$\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{y}) &= \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \quad \text{and} \\
\ell(\boldsymbol{\beta}) &\hat{=} \log[L(\boldsymbol{\beta}|\boldsymbol{y})] = \sum_{i=1}^{n}\Big\{y_i\log(p_i) + (1-y_i)\log(1-p_i)\Big\}.
\end{aligned}$$

Let $\phi(\cdot)$ denote the density of $N(0,1)$. Since

$$\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{\partial \Phi(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \phi(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta})\boldsymbol{x}_{(i)},$$

the score vector and the observed information matrix are

$$\begin{aligned}
\nabla\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right)\phi(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta})\boldsymbol{x}_{(i)} \quad \text{and} \\
-\nabla^2\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\boldsymbol{x}_{(i)}\left[\left\{\frac{y_i}{p_i^2} + \frac{1-y_i}{(1-p_i)^2}\right\}\phi(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta})\right. \\
&\qquad \left. + \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right)\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta}\right]\phi(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta})\boldsymbol{x}_{(i)}^{\top}.
\end{aligned}$$

(b) The Newton-Raphson (NR) algorithm updates

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \{-\nabla^2\ell(\boldsymbol{\beta}^{(t)})\}^{-1}\nabla\ell(\boldsymbol{\beta}^{(t)})$$

The estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = \{-\nabla^2 \ell(\hat{\boldsymbol{\beta}})\}^{-1}.$$

2.18 Solution: The observed-data likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}}) = \left(\prod_{i=1}^{4} \theta_i^{n_i}\right) \times (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}}.$$

By writing $n_{12} = Z_1 + Z_2$ with $Z_2 \equiv n_{12} - Z_1$ and $n_{34} = Z_3 + Z_4$ with $Z_4 \equiv n_{34} - Z_3$, a natural latent vector $(Z_1, Z_3)^\top$ can be introduced so that the likelihood function for the complete-data $\{Y_{\text{obs}}, Z_1, Z_3\}$ is

$$L(\boldsymbol{\theta}|Y_{\text{obs}}, Z_1, Z_3) \propto \prod_{i=1}^{4} \theta_i^{n_i + Z_i}.$$

Thus, the MLEs of $\boldsymbol{\theta}$ based on the complete data are given by

$$\hat{\theta}_i = \frac{n_i + Z_i}{N}, \quad i = 1, 2, 3, 4, \tag{SA2.1}$$

where $N = n_1 + n_2 + n_3 + n_4 + n_{12} + n_{34}$.

On the other hand, note that when Y_{obs} and $\boldsymbol{\theta}$ are given, Z_1 and Z_3 are independent binomially distributed. Thus, the conditional predictive distribution is

$$\begin{aligned} f(Z_1, Z_3|Y_{\text{obs}}, \boldsymbol{\theta}) &= \text{Binomial}(Z_1|n_{12},\ \theta_1/(\theta_1 + \theta_2)) \\ &\quad \times \text{Binomial}(Z_3|n_{34},\ \theta_3/(\theta_3 + \theta_4)). \end{aligned}$$

Thus, the E-step of the EM algorithm is to compute the conditional expectations

$$E(Z_1|Y_{\text{obs}}, \boldsymbol{\theta}) = \frac{n_{12}\theta_1}{\theta_1 + \theta_2} \quad \text{and} \quad E(Z_3|Y_{\text{obs}}, \boldsymbol{\theta}) = \frac{n_{34}\theta_3}{\theta_3 + \theta_4},$$

and the M-step is to update (SA2.1) by replacing Z_1 and Z_3 with $E(Z_1|Y_{\text{obs}}, \boldsymbol{\theta})$ and $E(Z_3|Y_{\text{obs}}, \boldsymbol{\theta})$, respectively.

2.19 Solution: The observed-data likelihood function of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|Y_{\rm obs}) &= \left(\prod_{i=1}^{4} \theta_i^{n_i}\right) \times (\theta_1+\theta_2)^{n_{12}}(\theta_3+\theta_4)^{n_{34}} \\ &\quad \times(\theta_1+\theta_3)^{n_{13}}(\theta_2+\theta_4)^{n_{24}}. \end{aligned}$$

By writing

$$\begin{aligned} n_{12} &= Z_1+Z_2 \quad \text{with} \quad Z_2 \equiv n_{12}-Z_1, \\ n_{34} &= Z_3+Z_4 \quad \text{with} \quad Z_4 \equiv n_{34}-Z_3, \\ n_{13} &= W_1+W_3 \quad \text{with} \quad W_3 \equiv n_{13}-W_1, \\ n_{24} &= W_2+W_4 \quad \text{with} \quad W_4 \equiv n_{24}-W_2, \end{aligned}$$

a natural latent vector $Z = (Z_1, Z_3, W_1, W_2)^{\top}$ can be introduced so that the likelihood function for the complete-data $\{Y_{\rm obs}, Z\}$ is

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

Thus, the MLEs of $\boldsymbol{\theta}$ based on the complete data are given by

$$\hat{\theta}_i = \frac{n_i+Z_i+W_i}{N}, \quad i=1,2,3,4, \tag{SA2.2}$$

where $N = n_1+n_2+n_3+n_4+n_{12}+n_{34}+n_{13}+n_{24}$.

On the other hand, note that when $Y_{\rm obs}$ and $\boldsymbol{\theta}$ are given, Z_1, Z_3, W_1 and W_2 are independent binomially distributed. Thus, the conditional predictive distribution is

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

Thus, the E-step of the EM algorithm is to compute the conditional expectations

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

and the M-step is to update (SA2.2) by replacing Z_1, Z_3, W_1 and W_2 with $E(Z_1|Y_{\rm obs},\boldsymbol{\theta})$, $E(Z_3|Y_{\rm obs},\boldsymbol{\theta})$, $E(W_1|Y_{\rm obs},\boldsymbol{\theta})$ and $E(W_2|Y_{\rm obs},\boldsymbol{\theta})$, respectively.

2.20 Solution: (a) Since $Y \sim$ Bernoulli(θ), we have

$$E(Y) = \theta \quad \text{and} \quad E(Y^2) = \theta.$$

On the other hand, from $U \sim$ Poisson(λ), we obtain

$$E(U) = \lambda \quad \text{and} \quad E(U^2) = \text{Var}(U) + (EU)^2 = \lambda + \lambda^2.$$

Let

$$\begin{aligned} \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i &= E(X) = E(Y) + E(U) = \theta + \lambda, \\ \Delta \hat{=} \frac{1}{n}\sum_{i=1}^n X_i^2 &= E(X^2) = E(Y^2) + E(U^2) + 2E(YU) \\ &= \theta + \lambda + \lambda^2 + 2\theta\lambda \\ &= (\theta+\lambda) + \lambda\{\theta + (\theta+\lambda)\}, \end{aligned}$$

we obtain the moment estimators as

$$\hat{\lambda}^{\rm M} = \frac{\Delta - \bar{X}}{\hat{\theta}^{\rm M} + \bar{X}} \quad \text{and} \quad \hat{\theta}^{\rm M} = \sqrt{\bar{X}(1+\bar{X}) - \Delta}.$$

(b) We consider two cases. If $x = 0$, since $X = Y + U$, then

$$\Pr(X = x) = \Pr(Y = 0, U = 0) = (1-\theta)\,{\rm e}^{-\lambda}. \tag{SA2.3}$$

If $x \geqslant 1$, then

$$\begin{aligned}
\Pr(X = x) &= \Pr(Y + U = x) = \sum_{y=0}^{1} \Pr(Y = y, U = x - y) \\
&= \sum_{y=0}^{1} \theta^y (1-\theta)^{1-y} \cdot \frac{\lambda^{x-y}}{(x-y)!}\, \mathrm{e}^{-\lambda} \\
&= (1-\theta)\frac{\lambda^x}{x!}\, \mathrm{e}^{-\lambda} + \theta \frac{\lambda^{x-1}}{(x-1)!}\, \mathrm{e}^{-\lambda}. \qquad \text{(SA2.4)}
\end{aligned}$$

The distribution of X is a special case of the Charlier series distribution, see (A.10) in Appendix A.1.4.

(c) We consider two cases. If $x = 0$, then

$$\begin{aligned}
\Pr(Y = 0|X = x) &= \Pr(Y = 0|X = 0) = \frac{\Pr(Y = 0, X = 0)}{\Pr(X = 0)} \\
&= \frac{\Pr(Y = 0, U = 0)}{\Pr(X = 0)} \overset{\text{(SA2.3)}}{=} 1. \qquad \text{(SA2.5)}
\end{aligned}$$

In other words, $Y|(X = 0) \sim \text{Degenerate}(0)$. If $x \geqslant 1$, then

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

where $\Pr(X = x)$ is given by (SA2.4). Similarly,

$$\Pr(Y = 1|X = x) = \frac{\theta \frac{\lambda^{x-1}}{(x-1)!}\, \mathrm{e}^{-\lambda}}{\Pr(X = x)}.$$

That is,

$$Y|(X = x) \sim \text{Bernoulli}\left(\frac{\theta \frac{\lambda^{x-1}}{(x-1)!}\, \mathrm{e}^{-\lambda}}{\Pr(X = x)} \right), \qquad \text{(SA2.6)}$$

where $\Pr(X = x)$ is given by (SA2.4).

(d) The observed data are $Y_{\text{obs}} = \{X_i\colon i = 1, \dots, n\}$. The complete data are $Y_{\text{com}} = \{(Y_i, U_i)\colon i = 1, \dots, n\}$. Since $Y_i \perp\!\!\!\perp U_i$, the complete-data MLEs of θ and λ are

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} U_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i). \tag{SA2.7}$$

The E-step is to compute the conditional expectations $E(Y_i|X_i, \theta, \lambda)$ for $i = 1, \dots, n$. Without loss of generality, we assume that $X_i = 0$ for $i = 1, \dots, m$ and $X_i \geqslant 1$ for $i = m+1, \dots, n$. From (SA2.5), we have

$$E(Y_i|X_i, \theta, \lambda) = 0, \quad i = 1, \dots, m. \tag{SA2.8}$$

From (SA2.6), we have (for $i = m+1, \dots, n$)

$$E(Y_i|X_i, \theta, \lambda) = \frac{\theta \frac{\lambda^{X_i - 1}}{(X_i - 1)!}\,\mathrm{e}^{-\lambda}}{(1-\theta)\frac{\lambda^{X_i}}{X_i!}\,\mathrm{e}^{-\lambda} + \theta \frac{\lambda^{X_i - 1}}{(X_i - 1)!}\,\mathrm{e}^{-\lambda}}. \tag{SA2.9}$$

Thus, the M-step is to compute $\hat{\theta}$ and $\hat{\lambda}$ given in (SA2.7) by replacing Y_i with $E(Y_i|X_i, \theta, \lambda)$ specified by (SA2.8) and (SA2.9).

2.21 Solution: (a) The pmf of $Y \sim \text{ZIP}(\phi, \lambda)$ is given by (1.36) in Exercise 1.12; i.e.,

$$\begin{aligned} f(y|\phi, \lambda) &= \{\phi + (1-\phi)\,\mathrm{e}^{-\lambda}\} \cdot I(y=0) \\ &\quad + \left\{(1-\phi)\frac{\mathrm{e}^{-\lambda}\lambda^y}{y!}\right\} \cdot I(y>0). \end{aligned} \tag{2.1}$$

Since Z only takes the value 0 or 1, we have

$$\begin{aligned} \Pr(Z=1|Y=y) &= \frac{\Pr(Z=1, Y=y)}{\Pr(Y=y)} = \frac{\Pr(Z=1, X=y)}{f(y|\phi,\lambda)} \\ &= \frac{(1-\phi)\,\mathrm{e}^{-\lambda}\lambda^y/y!}{f(y|\phi,\lambda)} \mathrel{\hat{=}} p_y, \end{aligned}$$

so that

$$p_0 = \frac{(1-\phi)\,\mathrm{e}^{-\lambda}}{\phi + (1-\phi)\,\mathrm{e}^{-\lambda}} \qquad \text{and} \qquad p_y = 1 \ \text{for} \ y > 0. \tag{2.2}$$

Therefore,

$$Z|(Y=y) \sim \begin{cases} \text{Bernoulli}\,(p_0), & \text{if } y=0, \\ \text{Degenerate}(1), & \text{if } y>0. \end{cases} \tag{2.3}$$

(b1) We first find the conditional distribution of $X|(Y=y=0)$. As

$$\begin{aligned}
\Pr(X=x|Y=0) &= \frac{\Pr(X=x, Y=0)}{\Pr(Y=0)} \\
&= \frac{\Pr(X=0, Y=0)}{f(0|\phi,\lambda)} I_{(x=0)} + \frac{\Pr(X=x, Z=0)}{f(0|\phi,\lambda)} I_{(x>0)} \\
&= \frac{\Pr(X=0)}{f(0|\phi,\lambda)} I_{(x=0)} + \frac{\phi\Pr(X=x)}{f(0|\phi,\lambda)} I_{(x>0)} \quad [\because \{X=0\} \subseteq \{Y=0\}] \\
&= \frac{\mathrm{e}^{-\lambda}}{\phi+(1-\phi)\,\mathrm{e}^{-\lambda}} I_{(x=0)} + \frac{\phi}{\phi+(1-\phi)\,\mathrm{e}^{-\lambda}} \cdot \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} I_{(x>0)} \\
&\overset{(2.2)}{=} [p_0+(1-p_0)\,\mathrm{e}^{-\lambda}] I_{(x=0)} + \left[(1-p_0)\frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}\right] I_{(x>0)},
\end{aligned} \tag{2.4}$$

by comparing (2.4) with (2.1), we have

$$X|(Y=0) \sim \text{ZIP}(p_0, \lambda). \tag{2.5}$$

(b2) We then find the conditional distribution of $X|(Y=y>0)$. Note that

$$\begin{aligned}
&\Pr(X=x|Y=y) \\
&= \frac{\Pr(X=x, Y=y)}{\Pr(Y=y)} \qquad [\because y>0 \Rightarrow x=y>0 \;\&\; Z=1] \\
&= \frac{\Pr(X=y, Z=1)}{f(y|\phi,\lambda)} \overset{(2.1)}{=} \frac{(1-\phi)\Pr(X=y)}{(1-\phi)\,\mathrm{e}^{-\lambda}\lambda^y/y!} = 1,
\end{aligned}$$

implying that

$$X|(Y=y>0) \sim \text{Degenerate}(y). \tag{2.6}$$

(c)

$$E(Y) = E(ZX) = E(Z) \cdot E(X) = (1-\phi)\lambda. \tag{2.7}$$

(d1) The M-step. The complete-data likelihood function for (ϕ, λ) is

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

so that complete-data MLEs are

$$\phi = \frac{n - \sum_{i=1}^{n} z_i}{n} \quad \text{and} \quad \lambda = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{2.8}$$

(d2) The E-step. Based on (2.3), we calculate

$$E(Z_i | Y_i = y_i, \phi, \lambda) = p_0 \cdot I_{(y_i=0)} + 1 \cdot I_{(y_i>0)}, \quad i = 1, \dots, n.$$

Based on (2.5) and (2.6), for $i = 1, \dots, n$, we calculate

$$E(X_i | Y_i = y_i, \phi, \lambda) \stackrel{(2.7)}{=} (1-p_0)\lambda \cdot I_{(y_i=0)} + y_i \cdot I_{(y_i>0)}.$$

2.22 Solution: (a) We only need to prove that $Y \sim \text{Poisson}(\lambda)$. To check this, we note that

$$\Pr(Y=0) = \Pr(ZX=0) = \Pr(Z=0) = \mathrm{e}^{-\lambda},$$

and by independency

$$\Pr(Y=y) = \Pr(Z=1, X=y) = \frac{1}{c} \cdot c \, \frac{\lambda^y \, \mathrm{e}^{-\lambda}}{y!}, \quad y > 0.$$

(b1) For each X_i, we introduce the latent variable Z_i to obtain the complete datum $Y_i = Z_i X_i$ via (2.65) in Exercise 2.22. Thus, the complete data are $Y_{\text{com}} = \{y_i\}_{i=1}^n = \{z_i, x_i\}_{i=1}^n$, where $\{Y_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$, $\{Z_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(1/c)$ and $\{Z_1, \dots, Z_n\} \perp\!\!\!\perp \{X_1, \dots, X_n\}$.

<u>The M-step</u>: The complete-data likelihood function of λ is

$$L(\lambda|Y_{\text{com}}) = \prod_{i=1}^{n} \Pr(Y_i = y_i) = \prod_{i=1}^{n} \frac{\lambda^{y_i}\,\mathrm{e}^{-\lambda}}{y_i!}. \tag{2.9}$$

Thus, the complete-data MLE of λ is

$$\lambda = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} z_i x_i}{n}. \tag{2.10}$$

<u>The E-step</u>: The E-step is to replace $\{z_i\}_{i=1}^{n}$ by their conditional expectations

$$\begin{aligned} E(Z_i|Y_{\text{obs}}, \lambda) &= E(Z_i|X_i, \lambda) \\ &= E(Z_i) \qquad\qquad [\because Z_i \perp\!\!\!\perp X_i] \\ &= \frac{1}{c} = 1 - \mathrm{e}^{-\lambda}, \quad i = 1, \ldots, n. \end{aligned} \tag{2.11}$$

By combining (2.10) with (2.11), we obtain the EM iteration:

$$\lambda^{(t+1)} = \bar{x}\{1 - \exp(-\lambda^{(t)})\}. \tag{2.12}$$

(b2) The observed-data likelihood function is

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

so that

$$\begin{aligned} \ell(\lambda|Y_{\text{obs}}) &= \log L(\lambda|Y_{\text{obs}}) \propto n\bar{x}\log\lambda - n\lambda - n\log(1-\,\mathrm{e}^{-\lambda}) \\ &= n\{\bar{x}\log\lambda - \lambda + g(\lambda)\}, \end{aligned}$$

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

(b3) Since $g'(\lambda) = -\mathrm{e}^{-\lambda}/(1-\mathrm{e}^{-\lambda})$ and $g''(\lambda) = \mathrm{e}^{-\lambda}/(1-\mathrm{e}^{-\lambda})^2 > 0$ for all $\lambda > 0$. Thus, $g(\lambda)$ is a strictly convex function. Applying the second-order Taylor expansion, we have

$$g(\lambda) \geqslant g(\lambda_0) + (\lambda - \lambda_0)g'(\lambda_0), \quad \forall \lambda > 0, \quad \lambda_0 > 0.$$

(b4) Let $\lambda_0 = \lambda^{(t)}$, we have

$$\begin{aligned} & \ell(\lambda|Y_{\text{obs}}) = n\{\bar{x}\log\lambda - \lambda + g(\lambda)\} \\ \geqslant\ & n\{\bar{x}\log\lambda - \lambda + g(\lambda_0) + (\lambda-\lambda_0)g'(\lambda_0)\} \\ =\ & n\left\{\bar{x}\log\lambda - \lambda - \log(1-\mathrm{e}^{-\lambda^{(t)}}) - (\lambda-\lambda_0)\frac{\mathrm{e}^{-\lambda^{(t)}}}{1-\mathrm{e}^{-\lambda^{(t)}}}\right\} \\ \hat{=}\ & Q(\lambda|\lambda^{(t)}). \end{aligned}$$

Let $\mathrm{d}Q(\lambda|\lambda^{(t)})/\mathrm{d}\lambda = 0$, we have the following MM algorithm

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

2.23 Solution: Let $\mathbf{x}_i = (X_{i1}, \dots, X_{in})^\top \overset{\text{iid}}{\sim} \text{Dirichlet}_n(\boldsymbol{a})$ on $\mathbb{T}_n$ and $\boldsymbol{x}_i = (x_{i1}, \dots, x_{in})^\top$ be the observations of $\mathbf{x}_i$ for $i = 1, \dots, m$. The likelihood function of $\boldsymbol{a} = (a_1, \dots, a_n)^\top$ for the observed data $Y_{\text{obs}} = \{\boldsymbol{x}_i\}_{i=1}^m$ is

$$L(\boldsymbol{a}|Y_{\text{obs}}) = \prod_{i=1}^m \left\{\frac{\Gamma(a_1 + \cdots + a_n)}{\Gamma(a_1)\cdots\Gamma(a_n)} \prod_{j=1}^n x_{ij}^{a_j - 1}\right\}$$

so that the log-likelihood function is

$$\ell(\boldsymbol{a}|Y_{\text{obs}}) = m\left\{\log\Gamma(a_+) - \sum_{j=1}^n \log\Gamma(a_j) + \sum_{j=1}^n (a_j - 1)\log G_j\right\},$$

where $a_+ = \sum_{j=1}^n a_j$, and

$$G_j = \left(\prod_{i=1}^m x_{ij}\right)^{1/m}, \quad j = 1, \dots, n,$$

denote the geometric means of the n variables.

<u>The gradient and Hessian matrix</u>. It is easy to verify that the gradient and the Hessian matrix are given by

$$\boldsymbol{g} = \nabla \ell(\boldsymbol{a}|Y_{\text{obs}}) = m \begin{pmatrix} \psi(a_+) - \psi(a_1) + \log G_1 \\ \vdots \\ \psi(a_+) - \psi(a_n) + \log G_n \end{pmatrix},$$

and

$$\begin{aligned} \boldsymbol{H} &= \nabla^2 \ell(\boldsymbol{a}|Y_{\text{obs}}) \\ &= m \begin{pmatrix} \psi'(a_+) - \psi'(a_1) & \psi'(a_+) & \cdots & \psi'(a_+) \\ \psi'(a_+) & \psi'(a_+) - \psi'(a_2) & \cdots & \psi'(a_+) \\ \vdots & \vdots & \ddots & \vdots \\ \psi'(a_+) & \psi'(a_+) & \cdots & \psi'(a_+) - \psi'(a_n) \end{pmatrix} \\ &= b \cdot \mathbf{1}_n \mathbf{1}_n^\top + \boldsymbol{B}, \end{aligned}$$

where $b \hat{=} m\psi'(a_+)$, $\boldsymbol{B} \hat{=} -m\ \text{diag}(\psi'(a_1), \dots, \psi'(a_n))$, $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ and $\psi'(x)$ are digamma and trigamma functions.

<u>The iteration</u>. Note that $\boldsymbol{H}$ does not depend on the observed data Y_{obs}. Hence, the Fisher information matrix

$$\boldsymbol{J}(\boldsymbol{a}) = \boldsymbol{I}(\boldsymbol{a}|Y_{\text{obs}}) = -\boldsymbol{H};$$

that is, the Newton–Raphson algorithm is identical to the Fisher scoring algorithm:

$$\boldsymbol{a}^{(t+1)} = \boldsymbol{a}^{(t)} + \boldsymbol{I}^{-1}(\boldsymbol{a}^{(t)}|Y_{\text{obs}}) \nabla \ell(\boldsymbol{a}^{(t)}|Y_{\text{obs}}) = \boldsymbol{a}^{(t)} - \boldsymbol{H}^{-1}\boldsymbol{g},$$

where

$$\boldsymbol{H}^{-1} = \boldsymbol{B}^{-1} - \frac{1}{b^{-1} + \mathbf{1}_n^\top \boldsymbol{B}^{-1} \mathbf{1}_n} \boldsymbol{B}^{-1} \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{B}^{-1}.$$

2.24 Solution: The pmf of $X_i \sim \text{BBinomial}(n_i, \alpha, \beta)$ is

$$\text{BBinomial}(x_i|n_i, \alpha, \beta) = \binom{n_i}{x_i} \frac{B(x_i + \alpha, n_i - x_i + \beta)}{B(\alpha, \beta)},$$

for $x_i = 0, 1, \ldots, n_i$.

(a) The likelihood function of (α, β) for the observed data $Y_{\text{obs}} = \{x_i\}_{i=1}^m$ is

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^m \binom{n_i}{x_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n_i)} \cdot \frac{\Gamma(\alpha+x_i)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta+n_i-x_i)}{\Gamma(\beta)} \\ &= \prod_{i=1}^m \binom{n_i}{x_i} \frac{\prod_{j=0}^{x_i-1}(\alpha+j) \prod_{j=0}^{n_i-x_i-1}(\beta+j)}{\prod_{j=0}^{n_i-1}(\alpha+\beta+j)}, \end{aligned}$$

so that the log-likelihood function of (α, β) is

$$\begin{aligned} \ell(\alpha, \beta) &= c + \sum_{i=1}^m \left\{ \sum_{j=0}^{x_i-1} \log(\alpha+j) + \sum_{j=0}^{n_i-x_i-1} \log(\beta+j) \right\} \\ &\quad - \sum_{i=1}^m \sum_{j=0}^{n_i-1} \log(\alpha+\beta+j), \end{aligned}$$

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

(b) Apply the discrete Jensen's inequality (see Exercise 2.5(d)) to the concave function $\log(\cdot)$, for any $u, v > 0$, we have

$$\begin{aligned} & \log(u+v) \\ = \ & \log\left(\frac{u^{(t)}}{u^{(t)}+v^{(t)}} \cdot \frac{u^{(t)}+v^{(t)}}{u^{(t)}} u + \frac{v^{(t)}}{u^{(t)}+v^{(t)}} \cdot \frac{u^{(t)}+v^{(t)}}{v^{(t)}} v \right) \\ \geqslant \ & \frac{u^{(t)}}{u^{(t)}+v^{(t)}} \log\left(\frac{u^{(t)}+v^{(t)}}{u^{(t)}} u \right) \\ & + \frac{v^{(t)}}{u^{(t)}+v^{(t)}} \log\left(\frac{u^{(t)}+v^{(t)}}{v^{(t)}} v \right). \end{aligned} \tag{2.13}$$

Let $u = \alpha$ and $v = j$, we have

$$\log(\alpha + j) \geqslant \frac{\alpha^{(t)}}{\alpha^{(t)} + j} \log\left(\frac{\alpha^{(t)} + j}{\alpha^{(t)}}\alpha\right) + \frac{j}{\alpha^{(t)} + j} \log\left(\frac{\alpha^{(t)} + j}{j} j\right), \tag{2.14}$$

where $\alpha^{(t)}$ denotes the t-th approximate of the MLE $\hat{\alpha}$.

(c) Apply the support superplane inequality (see Exercise 2.5(c)) to the convex function $-\log(\cdot)$, we have

$$-\log(u) \geqslant -\log(u^{(t)}) - (u - u^{(t)})/u^{(t)}. \tag{2.15}$$

Let $u = \alpha + \beta + j$ and $u^{(t)} = \alpha^{(t)} + \beta^{(t)} + j$, we obtain

$$-\log(\alpha + \beta + j) \geqslant -\log(\alpha^{(t)} + \beta^{(t)} + j) - \frac{\alpha + \beta - \alpha^{(t)} - \beta^{(t)}}{\alpha^{(t)} + \beta^{(t)} + j}, \tag{2.16}$$

where $\beta^{(t)}$ denotes the t-th approximate of the MLE $\hat{\beta}$.

(d) Combining (2.14) with (2.16), we can find the surrogate function

$$\begin{aligned} & Q(\alpha, \beta | \alpha^{(t)}, \beta^{(t)}) \\ = \ & c_1 + \sum_{i=1}^{m} \left\{ \sum_{j=0}^{x_i - 1} \frac{\alpha^{(t)}}{\alpha^{(t)} + j} \log(\alpha) + \sum_{j=0}^{n_i - x_i - 1} \frac{\beta^{(t)}}{\beta^{(t)} + j} \log(\beta) \right\} \\ & - \sum_{i=1}^{m} \sum_{j=0}^{n_i - 1} \frac{\alpha + \beta}{\alpha^{(t)} + \beta^{(t)} + j}, \end{aligned}$$

which minorizes $\ell(\alpha, \beta)$ at $(\alpha, \beta) = (\alpha^{(t)}, \beta^{(t)})$. Solving the system of two equations

$$\frac{\partial Q(\alpha, \beta | \alpha^{(t)}, \beta^{(t)})}{\partial \alpha} = 0 \quad \text{and} \quad \frac{\partial Q(\alpha, \beta | \alpha^{(t)}, \beta^{(t)})}{\partial \beta} = 0,$$

we obtain MM iterates:

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

where

$$\gamma^{(t)} = \sum_{i=1}^{m}\sum_{j=0}^{n_i-1}\frac{1}{\alpha^{(t)}+\beta^{(t)}+j}.$$