Southern University of Science and Technology
Department of Mathematics

MAT 7035 Computational Statistics

11 January (Friday) 2019 Time: 16:30 p.m. – 18:30 p.m.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Answer ALL 6 questions. Marks are shown in square brackets.

1. Use the inversion method to generate a random variable from the following distribution, and write down the algorithm:

   (a) (Zero-truncated Poisson distribution) The *probability mass function* (pmf) is

   $$p_x = \Pr(X = x) = c \cdot \frac{\lambda^x \, \mathrm{e}^{-\lambda}}{x!}, \qquad x = 1, 2, \ldots, \infty,$$

   where $\lambda > 0$ and c is the normalizing constant depending on λ. Denote the value of c by λ before generating this zero-truncated Poisson distribution. [**5 marks**]

   (b) (The Rayleigh distribution) The density is $f(x) = \sigma^{-2} x \exp(-\frac{x^2}{2\sigma^2})$, where $x \geqslant 0$ and $\sigma > 0$. [**5 marks**]

   [**Total: 10 marks**]

2. Suppose that we want to draw random samples from the target density $f(x)$ with support $\mathcal{S}_X$. Furthermore, we assume that there exist an envelope constant $c\,(\geqslant 1)$ and an envelope density $g(x)$ having the same support $\mathcal{S}_X$ such that $f(x) \leqslant cg(x)$ for all $x \in \mathcal{S}_X$.

(a) State the rejection algorithm for generating one random sample X from $f(x)$. **[3 marks]**

(b) Use the uniform density $g(x) = 1$ for $x \in (0, 1)$ as the envelope function to generate a random variable having the beta density

$$f(x) = 30x^2(1-x)^2, \quad x \in (0,1)$$

by the rejection method. **[10 marks]**

(c) How much is the acceptance probability for the current rejection algorithm? **[2 marks]**

[Total: 15 marks]

3. Assume that the binary responses $Y_1, \ldots, Y_n$ are independent, and

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(\pi_i), \\ \pi_i &= 1 - \exp\{-\exp[\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\beta}]\}, \quad 1 \leqslant i \leqslant n, \end{aligned}$$

where $\boldsymbol{x}_{(i)}$ is a known vector of covariates for subject i, and $\boldsymbol{\beta}_{p\times 1}$ is an unknown vector of parameters. [HINT: Let $F(z) = 1 - \exp\{-\mathrm{e}^z\}$, then $f(z) \mathrel{\hat{=}} F'(z) = \mathrm{e}^z \exp\{-\mathrm{e}^z\}$]

(a) Derive the score vector and the observed information matrix.

[10 marks]

(b) Using Newton–Raphson algorithm to find the MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and the estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. **[5 marks]**

[Total: 15 marks]

4. Let a random vector $(Y_1, \ldots, Y_4)^{\top}$ follow the following multinomial distribution:

$$(Y_1, \ldots, Y_4)^{\top} \sim \text{Multinomial}\left(n; \frac{\theta+2}{5}, \frac{1-\theta}{5}, \frac{\theta}{5}, \frac{2-\theta}{5}\right), \qquad (1)$$

where $0 \leqslant \theta \leqslant 1$, $n = \sum_{i=1}^{4} y_i$ and $\{y_i\}_{i=1}^{4}$ denote the corresponding observed values of $\{Y_i\}_{i=1}^{4}$. Let $Y_{\text{obs}} = (y_1, \dots, y_4)^\top = (125, 18, 20, 34)^\top$ be the observed data.

(a) Based on $Y_{\text{obs}} = (y_1, \dots, y_4)^\top$, derive the log-likelihood function $\ell(\theta|Y_{\text{obs}})$, the score function $\nabla\ell(\theta|Y_{\text{obs}})$, the observed information $I(\theta|Y_{\text{obs}})$, and the expected information $J(\theta)$. [HINT: If $(Y_1, \dots, Y_d)^\top \sim \text{Multinomial}(n; p_1, \dots, p_d)$, then $E(X_i) = np_i$ for $i = 1, \dots, d$.] **[8 marks]**

(b) Use the Fisher scoring algorithm to calculate the value of the MLE $\hat{\theta}$ of θ. Furthermore, let $\theta^{(t)}$ denote the t-th approximation of $\hat{\theta}$ in the Fisher scoring algorithm. If the initial value $\theta^{(0)} = 0.5$, calculate $\theta^{(t)}$ for $t = 1, \dots, 7$. **[7 marks]**

(c) Augment the observed data Y_{obs} with two latent variables (Z_1, Z_4) by splitting y_1 and y_4, and the corresponding cell probabilities as follows:

$$\begin{array}{ccccc} \dfrac{\theta+2}{5} & = & \dfrac{2(1-\theta)}{5} & + & \dfrac{3\theta}{5}, \\ \updownarrow & & \updownarrow & & \updownarrow \\ y_1 & = & Z_1 & + & (y_1 - Z_1); \end{array} \tag{2}$$

$$\begin{array}{ccccc} \dfrac{2-\theta}{5} & = & \dfrac{2(1-\theta)}{5} & + & \dfrac{\theta}{5}, \\ \updownarrow & & \updownarrow & & \updownarrow \\ y_4 & = & Z_4 & + & (y_4 - Z_4). \end{array} \tag{3}$$

Use the *expectation–maximization* (EM) algorithm to calculate the value of the MLE $\hat{\theta}$ of θ. Furthermore, let $\theta^{(t)}$ denote the t-th approximation of $\hat{\theta}$ in the EM algorithm. If the initial value $\theta^{(0)} = 0.6556928$, calculate $\theta^{(t)}$ for $t = 1, \dots, 5$. **[10 marks]**

(d) Let the fixed point iteration of the above EM algorithm be represented by $\theta^{(t+1)} = h(\theta^{(t)})$. Derive the expression of $h(\cdot)$ and calculate the value of the convergence rate

$$r \hat{=} \lim_{t\to\infty} \frac{|\theta^{(t+1)} - \hat{\theta}|}{|\theta^{(t)} - \hat{\theta}|}$$

of this EM algorithm. [**5 marks**]

(e) Describe the bootstrap method to obtain a $100(1-\alpha)\%$ bootstrap confidence interval for θ. [**5 marks**]

(f) Let the prior distribution of θ be Beta(a_0, b_0). State the I-step and P-step of the *data augmentation* (DA) algorithm. [**5 marks**]

[**Total: 40 marks**]

5. Let X be a discrete random variable with probability mass function (pmf) $p_i = \Pr(X = x_i)$ for $i = 1, 2, 3$ and Y be a discrete random variable with pmf $q_j = \Pr(Y = y_j)$ for $j = 1, 2, 3$. Given two conditional distribution matrices

$$\mathbf{A} = \begin{pmatrix} 1/6 & 0 & 3/14 \\ 0 & 1/4 & 4/14 \\ 5/6 & 3/4 & 7/14 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1/4 & 0 & 3/4 \\ 0 & 1/3 & 2/3 \\ 5/18 & 6/18 & 7/18 \end{pmatrix},$$

where the (i, j) element of $\mathbf{A}$ is $a_{ij} = \Pr\{X = x_i | Y = y_j\}$ and the (i, j) element of $\mathbf{B}$ is $b_{ij} = \Pr\{Y = y_j | X = x_i\}$.

(a) Find the marginal distribution of X. [**3 marks**]

(b) Find the marginal distribution of Y. [**3 marks**]

(c) Find the joint distribution of (X, Y). [**4 marks**]

[**Total: 10 marks**]

6. Let $Y_{\text{obs}} = \{n_1, n_2, n_3, n_4; m_1, m_2\}$ denote the observed frequencies and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_4)^\top$ be the cell probability vector satisfying $\theta_i \geqslant 0, \theta_1 +$

$\cdots + \theta_4 = 1$. Suppose that the observed-data likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}|Y_{\rm obs}) \propto \left(\prod_{i=1}^{4} \theta_i^{n_i}\right)(\theta_1+\theta_2)^{m_1}(\theta_1+\theta_2+\theta_3)^{m_2}.$$

(a) Use the EM algorithm to find the MLEs of $\boldsymbol{\theta}$. [**5 marks**]

(b) In fact, it is not necessary to employ the EM algorithm. Derive the closed-form expressions for the MLEs of $\boldsymbol{\theta}$. [**5 marks**]

[**Total: 10 marks**]

=== END OF THE PAPER ===

1. [E/U] **Solution**. (a) Note that 0 is truncated from the support. Thus

$$\frac{1}{c} = 1 - \mathrm{e}^{-\lambda} \quad \text{or} \quad c = \frac{1}{1 - \mathrm{e}^{-\lambda}}.$$

We have $p_1 = \frac{\lambda \mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}}$ and the recursive identity between p_{x+1} and p_x is

$$\frac{p_{x+1}}{p_x} = \frac{c \cdot \mathrm{e}^{-\lambda} \dfrac{\lambda^{x+1}}{(x+1)!}}{c \cdot \mathrm{e}^{-\lambda} \dfrac{\lambda^x}{x!}} = \frac{\lambda}{x+1}.$$

The algorithm is the following:
Step 1: generate U from $U(0,1)$;
Step 2: Let $i = 1$, $p = \frac{\lambda \mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}}$ and $F = p$;
Step 3: If $U < F$, set $X = i$ and stop; Otherwise
Step 4: Let $p = \frac{\lambda}{i+1} p$, $F = F + p$, $i = i + 1$ and go back to step 3.

(b) The cdf of the Rayleigh distribution with density

$$f(x) = \sigma^{-2} x \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \geqslant 0, \quad \sigma > 0,$$

is given by

$$\begin{aligned} F(x) &= \int_0^x f(y)\,dy = \int_0^x \sigma^{-2} y \exp\left(-\frac{y^2}{2\sigma^2}\right) dy \\ &= -\exp\left(-\frac{y^2}{2\sigma^2}\right)\Big|_0^x \\ &= 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right). \end{aligned}$$

Thus, $F(X) \stackrel{d}{=} U \sim U(0,1)$ implies

$$X \stackrel{d}{=} F^{-1}(U) = \sigma\sqrt{-2\log(1-U)} \stackrel{d}{=} \sigma\sqrt{-2\log(U)}\,.$$

2. [L/E] **Solution**. (a) THE REJECTION ALGORITHM:

Step 1. Draw $U \sim U(0,1)$ and independently draw $Y \sim g(\cdot)$;

Step 2. If $U \leqslant \frac{f(Y)}{cg(Y)}$, set $X = Y$; otherwise, go to Step 1.

(b) By differentiating the ratio

$$\frac{f(x)}{g(x)} = 30x^2(1-x)^2$$

with respect to x and setting the resultant derivative equal to zero, we obtain the maximal value of this ratio at $x = 1/2$. Hence

$$c = \max_{0<x<1} \frac{f(x)}{g(x)} = 30 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^2 = \frac{15}{8},$$

and

$$\frac{f(x)}{cg(x)} = 16x^2(1-x)^2.$$

The rejection method is as follows:

Step 1. Draw $U_1, U_2 \overset{\text{iid}}{\sim} U(0,1)$;

Step 2. If $U_1 \leqslant 16U_2^2(1-U_2)^2$, set $X = U_2$; otherwise, go to Step 1.

(c) The acceptance probability is $1/c = 8/15 = 0.5333333$.

3. [L/E/U] **Solution**. (a) Define

$$F(z) = 1 - \exp\{-\,\mathrm{e}^z\} \quad \text{and} \quad f(z) = F'(z) = \mathrm{e}^z \exp\{-\,\mathrm{e}^z\}.$$

The likelihood function and the log-likelihood function of $\boldsymbol{\beta}$ are given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}.$$

and

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Big\{ y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i) \Big\}.$$

Since

$$\frac{\partial \pi_i}{\partial \boldsymbol{\beta}} = \frac{\partial F(\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = f(\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}) \boldsymbol{x}_{(i)},$$

the score vector and the observed information matrix are given by

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

respectively.

(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

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

4. [L/U/E] **<u>Solution</u>**. (a) The observed-data likelihood function of θ is

$$\begin{aligned} L(\theta | Y_{\text{obs}}) &= \binom{n}{y_1, \ldots, y_4} \left(\frac{\theta + 2}{5} \right)^{y_1} \left(\frac{1 - \theta}{5} \right)^{y_2} \left(\frac{\theta}{5} \right)^{y_3} \left(\frac{2 - \theta}{5} \right)^{y_4} \\ &\propto (\theta + 2)^{y_1} (1 - \theta)^{y_2} \theta^{y_3} (2 - \theta)^{y_4}, \end{aligned} \tag{4}$$

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

$$\ell(\theta | Y_{\text{obs}}) = y_1 \log(\theta + 2) + y_2 \log(1 - \theta) + y_3 \log \theta + y_4 \log(2 - \theta).$$

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

$$\begin{aligned}
\nabla\ell(\theta|Y_{\text{obs}}) &= \frac{y_1}{\theta+2} - \frac{y_2}{1-\theta} + \frac{y_3}{\theta} - \frac{y_4}{2-\theta}, \\
I(\theta|Y_{\text{obs}}) = -\nabla^2\ell(\theta|Y_{\text{obs}}) &= \frac{y_1}{(\theta+2)^2} + \frac{y_2}{(1-\theta)^2} + \frac{y_3}{\theta^2} + \frac{y_4}{(2-\theta)^2}, \\
J(\theta) = E[-\nabla^2\ell(\theta|Y_{\text{obs}})] &= \frac{n}{5}\left(\frac{1}{\theta+2} + \frac{1}{1-\theta} + \frac{1}{\theta} + \frac{1}{2-\theta}\right),
\end{aligned}$$

respectively.

(b) The Fisher scoring algorithm is given by

$$\theta^{(t+1)} = \theta^{(t)} + [J(\theta^{(t)})]^{-1}\nabla\ell(\theta^{(t)}|Y_{\text{obs}}).$$

If let $\theta^{(0)} = 0.5$, we obtain

$$\begin{aligned}
\theta^{(1)} &= 0.6569597, \\
\theta^{(2)} &= 0.6555983, \\
\theta^{(3)} &= 0.6557001, \\
\theta^{(4)} &= 0.6556925, \\
\theta^{(5)} &= 0.6556931, \\
\theta^{(6)} &= 0.6556930, \\
\theta^{(7)} &= 0.6556930.
\end{aligned}$$

(c) From (2), the distribution of $Z_1|(Y_{\text{obs}}, \theta)$ is

$$f(z_1|Y_{\text{obs}}, \theta) = \text{Binomial}\left(z_1 \middle| y_1, \frac{2(1-\theta)}{\theta+2}\right), \tag{5}$$

for $z_1 = 0, 1, \ldots, y_1$. From (3), the distribution of $Z_4|(Y_{\text{obs}}, \theta)$ is

$$f(z_4|Y_{\text{obs}}, \theta) = \text{Binomial}\left(z_4 \middle| y_4, \frac{2(1-\theta)}{2-\theta}\right), \tag{6}$$

for $z_4 = 0, 1, \ldots, y_4$. The E-step is to compute

$$\begin{aligned} E(Z_1|Y_{\text{obs}}, \theta) &= E(Z_1|y_1, \theta) = y_1 \times \frac{2(1-\theta)}{\theta+2}, \quad \text{and} \\ E(Z_4|Y_{\text{obs}}, \theta) &= E(Z_4|y_4, \theta) = y_4 \times \frac{2(1-\theta)}{2-\theta}. \end{aligned}$$

From (1), the complete-data likelihood function of θ is

$$\begin{aligned} L(\theta|Y_{\text{obs}}, z_1, z_4) &= \binom{n}{z_1, y_1 - z_1, y_2, y_3, z_4, y_4 - z_4} \\ &\quad \times \left[\frac{2(1-\theta)}{5}\right]^{z_1} \left(\frac{3\theta}{5}\right)^{y_1 - z_1} \left(\frac{1-\theta}{5}\right)^{y_2} \left(\frac{\theta}{5}\right)^{y_3} \\ &\quad \times \left[\frac{2(1-\theta)}{5}\right]^{z_4} \left(\frac{\theta}{5}\right)^{y_4 - z_4} \\ &\propto \theta^{y_1+y_3+y_4-z_1-z_4}(1-\theta)^{y_2+z_1+z_4}. \end{aligned} \tag{7}$$

The M-step is to find the complete-data MLE

$$\hat{\theta} = \frac{y_1 + y_3 + y_4 - z_1 - z_4}{n} = 1 - \frac{y_2 + z_1 + z_4}{n}, \quad n = \sum_{i=1}^{4} y_i.$$

The EM algorithm is given by

$$\theta^{(t+1)} = 1 - \frac{y_2 + 2[1-\theta^{(t)}]\{y_1/[\theta^{(t)}+2] + y_4/[2-\theta^{(t)}]\}}{n}. \tag{8}$$

If let $\theta^{(0)} = 0.6556928$, we obtain

$$\begin{aligned} \theta^{(1)} &= 0.6556929, \\ \theta^{(2)} &= 0.6556929, \\ \theta^{(3)} &= 0.6556929, \\ \theta^{(4)} &= 0.6556930, \\ \theta^{(5)} &= 0.6556930. \end{aligned}$$

(d) Let $\theta^{(t+1)} = h(\theta^{(t)})$. From (8), we have

$$h(\theta) = 1 - \frac{y_2 + 2(1-\theta)[y_1/(\theta+2) + y_4/(2-\theta)]}{n},$$

so that

$$h'(\theta) = \frac{2}{n}\left(\frac{y_1}{\theta+2} + \frac{y_4}{2-\theta}\right) + \frac{2(1-\theta)}{n}\left[\frac{y_1}{(\theta+2)^2} - \frac{y_4}{(2-\theta)^2}\right].$$

Thus

$$r = |h'(\hat{\theta})| = 0.7308134.$$

The R code is as follows:

```
function(ind, th0, NumEM1)
{   # Function name: Linkage.model.EM1.EM2(ind, th0, NumEM1)
    # -------------- Input --------------------------------
    # ind    = 1: calculate the MLE by Fisher scoring algorithm
    #        = 2: calculate the MLE by EM algorithm
    #        = 3: calculate the convergence rate of
    #             the EM algorithm
    # th0    = initial value of \theta
    # NumEM1 = the number of iterations in the EM
    # -------------- Output ------------------------------
    # TH = approximates of the posterior mode
    # r  = the convergence rate of the EM algorithm
    # ----------------------------------------------------
    y <- c(125, 18, 20, 34)
    n <- sum(y)
    if (ind == 1) {
       th <- th0
       TH <- matrix(0, NumEM1, 1)
```

```
        for (tt in 1:NumEM1) {
           a <- y[1]/(th+2)-y[2]/(1-th)+y[3]/th-y[4]/(2-th)
           J <- 0.2*n*(1/(th + 2) + 1/(1-th) + 1/th + 1/(2-th))
           th <- th + a/J
           TH[tt] <- th
        }
        return(TH) }
     if (ind == 2) {
        th <- th0
        TH <- matrix(0, NumEM1, 1)
        for (tt in 1:NumEM1) {
           Ez1 <- y[1]*2*(1-th)/(th + 2)
           Ez4 <- y[4]*2*(1-th)/(2 - th)
           th <- 1 - (y[2] + Ez1 + Ez4)/n
           TH[tt] <- th
        }
        return(TH) }
     if (ind == 3) {
        hth <- 0.6556930
        th <- hth
        a <- y[1]/(th + 2) + y[4]/(2-th)
        b <- y[1]/(th + 2)^2 - y[4]/(2-th)^2
        r <- abs((2/n)*a + (2*(1-th)/n)*b)
        return(r) }
}
```

(e) The bootstrap method to obtain a $100(1-\alpha)\%$ bootstrap confidence interval for θ is as follows:

Step 1. For the MLE $\hat{\theta}$ obtained by the Fisher scoring algorithm or the EM algorithm (8), we can generate one bootstrap sample

$$(y_1^*, \ldots, y_4^*)^\top \sim \text{Multinomial}\left(n; \frac{\hat{\theta}+2}{5}, \frac{1-\hat{\theta}}{5}, \frac{\hat{\theta}}{5}, \frac{2-\hat{\theta}}{5}\right).$$

Based on $Y_{\text{obs}}^* = (y_1^*, \ldots, y_4^*)^\top$, we can compute the corresponding bootstrap replication $\hat{\theta}^*$ by using (8).

Step 2. Independently repeating this process (i.e., Step 1) G times, we obtain G bootstrap replications $\{\hat{\theta}^*(g)\}_{g=1}^G$.

Step 3. A $100(1-\alpha)\%$ bootstrap CI for θ is $[\hat{\theta}_L^*,\ \hat{\theta}_U^*]$, where $\hat{\theta}_L^*$ and $\hat{\theta}_U^*$ are the $(\alpha/2)G$-th and the $(1-\alpha/2)G$-th order statistics of $\{\hat{\theta}^*(g)\}_{g=1}^G$.

(f) Let $\theta \sim \text{Beta}(a_0, b_0)$. From (7), the complete-data posterior distribution of θ is

$$\theta|(Y_{\text{obs}}, z) \sim \text{Beta}(a_0 + y_1 + y_3 + y_4 - z_1 - z_4,\ b_0 + y_2 + z_1 + z_4). \quad (9)$$

Therefore, the I-step is to draw z_1 from (5), draw z_4 from (6) and the P-step is to draw θ from (9).

5. [E/L] **Solution**. (a) Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and $\mathcal{Y} = \{y_1, y_2, y_3\}$. By using (SA4.1) with $y_0 = y_3$, the X-marginal is given by

$$\begin{aligned}
\xi_1 &\hat{=} \Pr(X = x_1) = f_X(x_1) \\
&\propto \frac{f_{(X|Y)}(x_1|y_0)}{f_{(Y|X)}(y_0|x_1)} = \frac{\Pr(X = x_1|Y = y_3)}{\Pr(Y = y_3|X = x_1)} \\
&= \frac{a_{13}}{b_{13}} = \frac{3/14}{3/4} = \frac{4}{14}, \\
\xi_2 &\hat{=} \Pr(X = x_2) = f_X(x_2) \\
&\propto \frac{f_{(X|Y)}(x_2|y_0)}{f_{(Y|X)}(y_0|x_2)} = \frac{\Pr(X = x_2|Y = y_3)}{\Pr(Y = y_3|X = x_2)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{a_{23}}{b_{23}} = \frac{4/14}{3/2} = \frac{6}{14}, \\
\xi_3 &\mathrel{\hat{=}} \Pr(X = x_3) = f_X(x_3) \\
&\propto \frac{f_{(X|Y)}(x_3|y_0)}{f_{(Y|X)}(y_0|x_3)} = \frac{\Pr(X = x_3|Y = y_3)}{\Pr(Y = y_3|X = x_3)} \\
&= \frac{a_{33}}{b_{33}} = \frac{7/14}{7/18} = \frac{18}{14}.
\end{aligned}
$$

Note that $\xi_1 + \xi_2 + \xi_3 = 1$, we obtain

$$
\begin{aligned}
\xi_1 &= \frac{4/14}{4/14 + 6/14 + 18/14} = \frac{4}{4+6+18} = \frac{4}{28} = \frac{2}{14}, \\
\xi_2 &= \frac{6/14}{4/14 + 6/14 + 18/14} = \frac{6}{4+6+18} = \frac{6}{28} = \frac{3}{14}, \\
\xi_3 &= \frac{18/14}{4/14 + 6/14 + 18/14} = \frac{18}{4+6+18} = \frac{18}{28} = \frac{9}{14},
\end{aligned}
$$

which are summarized into

X	x_1	x_2	x_3
$\xi_i = \Pr(X = x_i)$	$2/14$	$3/14$	$9/14$

(b) Similarly, letting $x_0 = x_3$ in (SA4.1) yields the following Y-marginal

Y	y_1	y_2	y_3
$\eta_j = \Pr(Y = y_j)$	$3/14$	$4/14$	$7/14$

(c) The joint distribution of (X, Y) is given by

$$
\mathbf{P} = \begin{pmatrix} 1/28 & 0 & 3/28 \\ 0 & 2/28 & 4/28 \\ 5/28 & 6/28 & 7/28 \end{pmatrix}.
$$

6. [E/L] **Solution**. (a) We introduce a latent random variable W to split the term $(\theta_1+\theta_2)^{m_1}$ so that the conditional predictive distribution is

$$W|(m_1,\boldsymbol{\theta}) \sim \text{Binomial}\left(m_1; \frac{\theta_1}{\theta_1+\theta_2}\right),$$

and

$$E(W|m_1,\boldsymbol{\theta}) = \frac{m_1\theta_1}{\theta_1+\theta_2}. \tag{10}$$

Next, we introduce a latent vector $Z = (Z_1, Z_2, Z_3)^\top$ to split the term $(\theta_1+\theta_2+\theta_3)^{m_2}$ so that the conditional predictive distribution is

$$Z|(m_2,\boldsymbol{\theta}) \sim \text{Multinomial}_3\left(m_2; \frac{\theta_1}{\theta_{123}}, \frac{\theta_2}{\theta_{123}}, \frac{\theta_3}{\theta_{123}}\right),$$

where $\theta_{123} \mathrel{\hat{=}} \theta_1+\theta_2+\theta_3$ and $Z_1+Z_2+Z_3 = m_2$. The conditional expectations are given by

$$E(Z_i|m_2,\boldsymbol{\theta}) = \frac{m_2\theta_i}{\theta_1+\theta_2+\theta_3}, \quad i = 1, 2, 3. \tag{11}$$

Note that W is independent of Z, the complete-data likelihood function is given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) \propto \theta_1^{n_1+W+Z_1}\theta_2^{n_2+m_1-W+Z_2}\theta_3^{n_3+Z_3}\theta_4^{n_4}.$$

Taking log, we obtain

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) = \log L(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) = (n_1+W+Z_1)\log\theta_1$$

$$+(n_2+m_1-W+Z_2)\log\theta_2 + (n_3+Z_3)\log\theta_3 + n_4\log\theta_4.$$

Thus, the E-step of the EM algorithm is to compute the conditional expectations (10) and (11), and the M-step of the EM algorithm is to update the complete-data MLEs

$$\begin{aligned}
\hat\theta_1 &= \frac{n_1+W+Z_1}{n+m_1+m_2}, \quad \hat\theta_2 = \frac{n_2+m_1-W+Z_2}{n+m_1+m_2},\\
\hat\theta_3 &= \frac{n_3+Z_3}{n+m_1+m_2}, \quad \hat\theta_4 = \frac{n_4}{n+m_1+m_2},
\end{aligned}$$

by replacing W and Z_i with $E(W|m_1, \theta)$ and $E(Z_i|m_2, \theta)$, where $n = n_1 + n_2 + n_3 + n_4$.

(b) The log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}|Y_{\text{obs}}) &= \sum_{i=1}^{3} n_i \log(\theta_i) + n_4 \log(1 - \theta_1 - \theta_2 - \theta_3) \\ &\quad + m_1 \log(\theta_1 + \theta_2) + m_2 \log(\theta_1 + \theta_2 + \theta_3).\end{aligned}$$

Let

$$0 = \frac{\partial \ell}{\partial \theta_1} = \frac{n_1}{\theta_1} - \frac{n_4}{\theta_4} + \frac{m_1}{\theta_1 + \theta_2} + \frac{m_2}{\theta_1 + \theta_2 + \theta_3}, \tag{12}$$

$$0 = \frac{\partial \ell}{\partial \theta_2} = \frac{n_2}{\theta_2} - \frac{n_4}{\theta_4} + \frac{m_1}{\theta_1 + \theta_2} + \frac{m_2}{\theta_1 + \theta_2 + \theta_3}, \tag{13}$$

$$0 = \frac{\partial \ell}{\partial \theta_3} = \frac{n_3}{\theta_3} - \frac{n_4}{\theta_4} + \frac{m_2}{\theta_1 + \theta_2 + \theta_3}, \tag{14}$$

The tip is to first express each θ_i $(i = 1, 2, 3, 4)$ as a function of an unknown constant ρ and then find this ρ via the constraint

$$1 = \theta_1 + \theta_2 + \theta_3 + \theta_4.$$

Note that (12) − (13) and (12) − (14) will result in

$$\frac{n_1}{\theta_1} = \frac{n_2}{\theta_2} \hat{=} \frac{1}{\rho}, \tag{15}$$

$$\frac{n_1}{\theta_1} + \frac{m_1}{\theta_1 + \theta_2} = \frac{n_3}{\theta_3}, \tag{16}$$

respectively. Therefore, $\theta_1 = n_1\rho$ and $\theta_2 = n_2\rho$. Inserting them into (16), we have

$$\frac{1}{\rho} + \frac{m_1}{n_1\rho + n_2\rho} = \frac{n_3}{\theta_3} \Longrightarrow \theta_3 = \frac{(n_1 + n_2)n_3\rho}{n_1 + n_2 + m_1}. \tag{17}$$

Inserting $\theta_1, \theta_2, \theta_3$ into (14), we obtain

$$\frac{n_4}{\theta_4} = \frac{1}{\rho}\left[1 + \frac{m_1}{n_1+n_2} + \frac{m_2}{n_1+n_2+\dfrac{(n_1+n_2)n_3}{n_1+n_2+m_1}}\right] \hat{=} \frac{c_0}{\rho},$$

or

$$\theta_4 = \frac{n_4\rho}{c_0}. \tag{18}$$

From $1 = \theta_1 + \theta_2 + \theta_3 + \theta_4$, we solve

$$1 = \rho\left[\underbrace{n_1+n_2+\frac{(n_1+n_2)n_3}{n_1+n_2+m_1}}_{c_1} + \frac{n_4}{c_0}\right] = \rho(c_1 + n_4/c_0)$$

and obtain

$$\rho = \frac{c_0}{c_1c_0+n_4} \quad \text{where} \quad c_1 \hat{=} n_1+n_2+\frac{(n_1+n_2)n_3}{n_1+n_2+m_1}. \tag{19}$$

It is easy to check that

$$c_1c_0 + n_4 = n_1+n_2+n_3+n_4+m_1+m_2 \hat{=} n.$$

Thus,

$$\rho = \frac{c_0}{n}. \tag{20}$$

Finally,

$$\hat{\theta}_4 = \frac{n_4}{n}, \quad \hat{\theta}_3 = \frac{(n_1+n_2)n_3\rho}{n_1+n_2+m_1}, \quad \hat{\theta}_1 = n_1\rho, \quad \hat{\theta}_2 = n_2\rho.$$