MAT7035: Computational Statistics

Suggested Solutions to Assignment 4

4.1 Solution. See 24.2•–24.5• on pages 97–98 in Tian GL and Jiang XJ (2021). *Mathematical Statistics*. Science Press, Beijing, P.R. China.

4.2 Solution. Note that

$$\begin{aligned} f(x|y) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-y)^2}{2}\right\} \quad \text{and} \\ f(y|x) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\}, \end{aligned}$$

from the sampling-wise IBF, we obtain

$$f_X(x) \propto \frac{f(x|y)}{f(y|x)} = 1, \quad -\infty < x < \infty,$$

indicating that $f_X(x)$ does not exist.

4.3 Solution. Note that

$$\begin{aligned} f(x|y) &= c_1^{-1}(y) \exp\left\{-\frac{[x-\mu_1-\rho\sigma_1\sigma_2^{-1}(y-\mu_2)]^2}{2\sigma_1^2(1-\rho^2)}\right\}, \ a_1 \leqslant x \leqslant b_1, \\ f(y|x) &= c_2^{-1}(x) \exp\left\{-\frac{[y-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_1)]^2}{2\sigma_2^2(1-\rho^2)}\right\}, \ a_2 \leqslant y \leqslant b_2, \end{aligned}$$

where

$$\begin{aligned} c_1(y) &= \int_{a_1}^{b_1} \exp\left\{-\frac{[x-\mu_1-\rho\sigma_1\sigma_2^{-1}(y-\mu_2)]^2}{2\sigma_1^2(1-\rho^2)}\right\} \mathrm{d}x, \\ &= \int_{a_1^*}^{b_1^*} \mathrm{e}^{-z^2/2} \sqrt{\sigma_1^2(1-\rho^2)}\, \mathrm{d}z \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{2\pi\sigma_1^2(1-\rho^2)}\,\Big[\Phi(b_1^*)-\Phi(a_1^*)\Big],\\
a_1^* &= \frac{a_1-\mu_1-\rho\sigma_1\sigma_2^{-1}(y-\mu_2)}{\sqrt{\sigma_1^2(1-\rho^2)}},\\
b_1^* &= \frac{b_1-\mu_1-\rho\sigma_1\sigma_2^{-1}(y-\mu_2)}{\sqrt{\sigma_1^2(1-\rho^2)}},\\
c_2(x) &= \int_{a_2}^{b_2}\exp\left\{-\frac{[y-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_1)]^2}{2\sigma_2^2(1-\rho^2)}\right\}\mathrm{d}y,\\
&= \int_{a_2^*}^{b_2^*}\mathrm{e}^{-z^2/2}\sqrt{\sigma_2^2(1-\rho^2)}\,\mathrm{d}z\\
&= \sqrt{2\pi\sigma_2^2(1-\rho^2)}\,\Big[\Phi(b_2^*)-\Phi(a_2^*)\Big],\\
a_2^* &= \frac{a_2-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_1)}{\sqrt{\sigma_2^2(1-\rho^2)}},\\
b_2^* &= \frac{b_2-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_1)}{\sqrt{\sigma_2^2(1-\rho^2)}}.
\end{aligned}
$$

By using the point-wise IBF, we have

$$
\{f_X(x)\}^{-1} = \int_{a_2}^{b_2}\frac{f(y|x)}{f(x|y)}\,\mathrm{d}y
$$

$$
\begin{aligned}
&= c_2^{-1}(x)\int_{a_2}^{b_2}\frac{c_1(y)\exp\left\{-\frac{[y-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_1)]^2}{2\sigma_2^2(1-\rho^2)}\right\}}{\exp\left\{-\frac{[x-\mu_1-\rho\sigma_1\sigma_2^{-1}(y-\mu_2)]^2}{2\sigma_1^2(1-\rho^2)}\right\}}\,\mathrm{d}y\\
&= c_2^{-1}(x)\int_{a_2}^{b_2}c_1(y)\exp\left\{\frac{(x-\mu_1)^2}{2\sigma_1^2}-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}\mathrm{d}y\\
&= c_2^{-1}(x)\,\mathrm{e}^{\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{a_2}^{b_2}c_1(y)\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}\mathrm{d}y\\
&\propto c_2^{-1}(x)\,\mathrm{e}^{\frac{(x-\mu_1)^2}{2\sigma_1^2}},
\end{aligned}
$$

i.e.,

$$\begin{aligned} f_X(x) &\propto \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} c_2(x) \\ &\propto \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \Big[\Phi(b_2^*) - \Phi(a_2^*)\Big], \quad a_1 \leqslant x \leqslant b_1. \end{aligned}$$

By symmetry, we obtain

$$\begin{aligned} f_Y(y) &\propto \mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} c_1(y) \\ &\propto \mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \Big[\Phi(b_1^*) - \Phi(a_1^*)\Big], \quad a_2 \leqslant y \leqslant b_2. \end{aligned}$$

4.4 Solution. (a) By applying the formula

$$f_X(x) \propto \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)}, \tag{SA4.1}$$

and setting $y_0 = b/2$, the marginal distribution of X is given by

$$f_X(x) \propto \frac{1 - \exp(-bx)}{x} \mathrel{\hat{=}} h(x), \quad 0 \leqslant x < b < \infty. \tag{SA4.2}$$

We first prove

$$h(x) \leqslant b \quad \text{for any } x \in [0, b). \tag{SA4.3}$$

For any continuous and twice differentiable function $g(x)$ with $g''(x) > 0$, the second order Taylor expansion of $g(x)$ around x_0 is

$$\begin{aligned} g(x) &= g(x_0) + (x - x_0)g'(x_0) + \frac{(x-x_0)^2}{2} g''(\xi) \\ &\geqslant g(x_0) + (x - x_0)g'(x_0), \end{aligned}$$

where ξ is a point between x and x_0. Now let $g(x) = \mathrm{e}^{-bx}$ and $x_0 = 0$. Since $g'(x) = -b\,\mathrm{e}^{-bx}$ and $g''(x) = b^2\,\mathrm{e}^{-bx} > 0$ for any $x \in [0, b)$, we have $\mathrm{e}^{-bx} \geqslant 1 - bx$, or $b \geqslant (1 - \mathrm{e}^{-bx})/x = h(x)$, implying (SA4.3). From (SA4.3), we obtain

$$\int_0^b h(x)\mathrm{d}x \leqslant \int_0^b b\,\mathrm{d}x = b^2 < \infty,$$

which implies $f_X(x)$ exists.

(b) If let $b = \infty$, then from (SA4.2), $f_X(x) \propto 1/x$, $0 \leqslant x < \infty$. Obviously, $f_X(x)$ is not a density.

4.5 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

$$\begin{aligned}
\xi_1 &\hat{=} \Pr(X = x_1) = f_X(x_1) \\
&\propto \frac{f_{(X|Y)}(x_1|y_0)}{f_{(Y|X)}(y_0|x_1)} = \frac{\Pr(X = x_1|Y = y_3)}{\Pr(Y = y_3|X = x_1)} \\
&= \frac{a_{13}}{b_{13}} = \frac{3/14}{3/4} = \frac{4}{14}, \\
\xi_2 &\hat{=} \Pr(X = x_2) = f_X(x_2) \\
&\propto \frac{f_{(X|Y)}(x_2|y_0)}{f_{(Y|X)}(y_0|x_2)} = \frac{\Pr(X = x_2|Y = y_3)}{\Pr(Y = y_3|X = x_2)} \\
&= \frac{a_{23}}{b_{23}} = \frac{4/14}{3/2} = \frac{6}{14}, \\
\xi_3 &\hat{=} \Pr(X = x_3) = f_X(x_3) \\
&\propto \frac{f_{(X|Y)}(x_3|y_0)}{f_{(Y|X)}(y_0|x_3)} = \frac{\Pr(X = x_3|Y = y_3)}{\Pr(Y = y_3|X = x_3)} \\
&= \frac{a_{33}}{b_{33}} = \frac{7/14}{7/18} = \frac{18}{14}.
\end{aligned}$$

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

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

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

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

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

4.6 Solution. (a) Note that $z_i = y_i$ for $i = r+1, \ldots, m$, the complete-data likelihood is given by

$$\begin{aligned} L(\theta|Y_{\text{obs}}, \boldsymbol{z}) &= f(y_1, \ldots, y_m|\theta) = \prod_{i=1}^{m} f(y_i|\theta) \\ &= \prod_{i=1}^{m} \text{Exponential}(y_i|\theta) = \prod_{i=1}^{m} \theta\,\mathrm{e}^{-\theta y_i} \\ &= \theta^m \exp\left\{-\theta \sum_{i=1}^{m} y_i\right\} = \theta^m \exp\left\{-\theta\left[\sum_{i=1}^{r} y_i + \sum_{i=r+1}^{m} z_i\right]\right\} \\ &= \theta^m \exp\left\{-\theta(y^* + \mathbf{1}^\top \boldsymbol{z})\right\}, \end{aligned}$$

where $y^* \hat{=} \sum_{i=1}^{r} y_i$.

(b) Since the prior distribution of θ is $\text{Gamma}(\alpha_0, \beta_0)$, the complete-data posterior density

$$\begin{aligned} p(\theta|Y_{\text{obs}}, \boldsymbol{z}) &\propto L(\theta|Y_{\text{obs}}, \boldsymbol{z}) \times \text{Gamma}(\theta|\alpha_0, \beta_0) \\ &= \theta^m \exp\left\{-\theta(y^* + \mathbf{1}^\top \boldsymbol{z})\right\} \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1}\,\mathrm{e}^{-\theta\beta_0} \\ &\propto \theta^{m+\alpha_0-1} \exp\left\{-\theta(y^* + \mathbf{1}^\top \boldsymbol{z} + \beta_0)\right\}, \end{aligned}$$

i.e.,

$$\theta|(Y_{\text{obs}}, \boldsymbol{z}) \sim \text{Gamma}(m + \alpha_0,\ y^* + \mathbf{1}^\top \boldsymbol{z} + \beta_0). \tag{SA4.4}$$

On the other hand,

$$f(y_{r+1}, \ldots, y_m|\theta) = \prod_{i=r+1}^{m} f(y_i|\theta) = \prod_{i=r+1}^{m} \theta\, \mathrm{e}^{-\theta y_i},$$

so that the conditional predictive density is given by

$$f(\boldsymbol{z}|Y_{\text{obs}}, \theta) = \prod_{i=r+1}^{m} f(z_i|c_i, \theta)$$

$$= \prod_{i=r+1}^{m} a_i^{-1} \cdot \theta\, \mathrm{e}^{-\theta z_i} \cdot I_{(z_i > c_i)} = \prod_{i=r+1}^{m} \theta\, \mathrm{e}^{-\theta(z_i - c_i)} \cdot I_{(z_i > c_i)},$$

where the normalizing constant a_i is given by

$$a_i = \int_{c_i}^{\infty} \theta\, \mathrm{e}^{-\theta z_i}\, \mathrm{d}z_i = -\,\mathrm{e}^{-\theta z_i}|_{c_i}^{\infty} = \mathrm{e}^{-\theta c_i}.$$

In other words, $Z_i|(c_i, \theta)$ follows an exponential distribution with mean $1/\theta$ truncated at c_i. Define $W_i = Z_i - c_i$ for $i = r+1, \ldots, m$, then

$$f(w_i|c_i, \theta) = f(z_i|c_i, \theta) \left| \frac{\mathrm{d}z_i}{dw_i} \right| = \theta\, \mathrm{e}^{-\theta w_i} \cdot I_{(w_i > 0)};$$

i.e., $W_i|(c_i, \theta) \sim \text{Exponential}(\theta) = \text{Gamma}(1, \theta)$. Therefore, we have

$$\mathbf{1}^\top \mathbf{z} - c.|(Y_{\text{obs}}, \theta) = \sum_{i=r+1}^{m} (Z_i - c_i)|(Y_{\text{obs}}, \theta)$$

$$= \sum_{i=r+1}^{m} W_i|(Y_{\text{obs}}, \theta) \sim \text{Gamma}(m - r, \theta).$$

(c) The Gibbs sampler is as follows.

- First, given $\theta^{(t)}$, we draw $W^{(t)}$ from $\text{Gamma}(m - r, \theta^{(t)})$.

- Let $W^{(t)} = \mathbf{1}^{\top}\mathbf{z}^{(t)} - c.$ or $\mathbf{1}^{\top}\mathbf{z}^{(t)} = W^{(t)} + c.$.
- From (SA4.4), then we draw $\theta^{(t+1)}$ from Gamma$(m + \alpha_0,\ y^* + \mathbf{1}^{\top}\mathbf{z}^{(t)} + \beta_0)$.
- Repeat this process until convergence.

4.7 Solution. (a) First, we need to find the joint density:

$$\begin{aligned}
f(Y_{\text{obs}}, \boldsymbol{\lambda}, \beta) &= f(Y_{\text{obs}}|\boldsymbol{\lambda}, \beta) \cdot f(\boldsymbol{\lambda}|\beta) \cdot f(\beta) \\
&= f(Y_{\text{obs}}|\boldsymbol{\lambda}) \cdot f(\boldsymbol{\lambda}|\beta) \cdot f(\beta) \\
&= f(N_1, \ldots, N_m|\boldsymbol{\lambda}) \cdot f(\lambda_1, \ldots, \lambda_m|\beta) \cdot f(\beta) \\
&= \left\{\prod_{i=1}^{m} \frac{(\lambda_i t_i)^{N_i}}{N_i!}\, \mathrm{e}^{-\lambda_i t_i}\right\} \cdot \left\{\prod_{i=1}^{m} \frac{\beta^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_i^{\alpha_0 - 1}\, \mathrm{e}^{-\lambda_i \beta}\right\} \\
&\quad \times \frac{b_0^{a_0}}{\Gamma(a_0)} \beta^{a_0 - 1}\, \mathrm{e}^{-\beta b_0}.
\end{aligned}$$

Therefore, we obtain

$$f(\boldsymbol{\lambda}|Y_{\text{obs}}, \beta) \propto f(Y_{\text{obs}}, \boldsymbol{\lambda}, \beta) \propto \prod_{i=1}^{m} \lambda_i^{N_i + \alpha_0 - 1}\, \mathrm{e}^{-\lambda_i(t_i + \beta)};$$

i.e., $f(\boldsymbol{\lambda}|Y_{\text{obs}}, \beta) = \prod_{i=1}^{m}$ Gamma$(\lambda_i|N_i + \alpha_0,\ t_i + \beta)$. Similarly, we have

$$f(\beta|Y_{\text{obs}}, \boldsymbol{\lambda}) \propto f(Y_{\text{obs}}, \boldsymbol{\lambda}, \beta) \propto \beta^{a_0 + m\alpha_0 - 1}\, \mathrm{e}^{-\beta(b_0 + \sum_{i=1}^{m} \lambda_i)};$$

i.e., $f(\beta|Y_{\text{obs}}, \boldsymbol{\lambda}) = $ Gamma$(\beta|a_0 + m\alpha_0,\ b_0 + \sum_{i=1}^{m} \lambda_i)$.

(b)

4.8 Solution. R code is as follows:

```
function(th0)
{
     # Function name: exercise4.8(th0=0.8):
```

```
y <- c(14, 0, 1, 5)
a <- b <- 1
q <- rep(0, y[1] + 1)
for(k in 1:(y[1] + 1)) {
     zk <- k - 1
     q[k] <- dbinom(zk, y[1], th0/(th0 + 2))/
             dbeta(th0, a+y[4]+zk, b+y[2]+y[3])
}
p <- q/sum(q)
return(q, p)}
```

The corresponding output is as follows:

```
> exercise4.8(0.8)
$q:
 [1] 3.2695e-003 1.7165e-002 4.3389e-002 6.9422e-002
 [5] 7.8100e-002 6.5083e-002 4.1303e-002 2.0230e-002
 [9] 7.6705e-003 2.2372e-003 4.9351e-004 7.9759e-005
[13] 8.9204e-006 6.1756e-007 1.9955e-008

$p:
 [1] 9.3829e-003 4.9260e-002 1.2452e-001 1.9923e-001
 [5] 2.2413e-001 1.8678e-001 1.1853e-001 5.8057e-002
 [9] 2.2013e-002 6.4205e-003 1.4163e-003 2.2890e-004
[13] 2.5600e-005 1.7723e-006 5.7269e-008
```

We can see that our $\{q_k(0.8)\}_{k=1}^{15}$ are different from $\{q_k(0.5)\}_{k=1}^{15}$, but $\{p_k\}_{k=1}^{15}$ retain the same.

4.9 Solution. <u>The P-step</u>. The complete-data likelihood function for

(ϕ, λ) is

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

Since the joint prior distribution of (ϕ, λ) is

$$\begin{aligned} \pi(\phi, \lambda) &= \mathrm{Beta}(\phi | a_0, b_0) \times \mathrm{Gamma}(\lambda | \alpha_0, \beta_0) \\ &= \frac{\phi^{a_0-1}(1-\phi)^{b_0-1}}{B(a_0, b_0)} \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} \, \mathrm{e}^{-\beta_0 \lambda}, \end{aligned}$$

the joint posterior distribution of (ϕ, λ) is

$$p(\phi, \lambda | Y_{\rm com}) \propto \phi^{n - n\bar{z} + a_0 - 1} (1-\phi)^{n\bar{z} + b_0 - 1} \times \lambda^{n\bar{x} + \alpha_0 - 1} \, \mathrm{e}^{-(n+\beta_0)\lambda},$$

where $\bar{z} = (1/n) \sum_{i=1}^{n} z_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$, or

$$p(\phi, \lambda | Y_{\rm com}) = p(\phi | Y_{\rm com}) \times p(\lambda | Y_{\rm com})$$

$$= \mathrm{Beta}(\phi | n - n\bar{z} + a_0, n\bar{z} + b_0) \times \mathrm{Gamma}(\lambda | n\bar{x} + \alpha_0, n + \beta_0).$$

The I-step. From (2.2) and (2.3) in the Solution to Exercise 2.21, we have

$$Z_i | (Y_i = y_i, \phi, \lambda) \sim \begin{cases} \text{Bernoulli}\,(p_0), & \text{if } y_i = 0, \\ \text{Degenerate}(1), & \text{if } y_i > 0, \end{cases} \tag{SA4.5}$$

where

$$p_0 = \frac{(1-\phi)\,\mathrm{e}^{-\lambda}}{\phi + (1-\phi)\,\mathrm{e}^{-\lambda}}. \tag{SA4.6}$$

From (2.5) and (2.6) in the Solution to Exercise 2.21, we have

$$\begin{aligned} X_i | (Y_i = 0, \phi, \lambda) &\sim \mathrm{ZIP}(p_0, \lambda) \quad \text{and} \\ X_i | (Y_i = y_i > 0, \phi, \lambda) &\sim \mathrm{Degenerate}(y_i). \end{aligned}$$

4.10 Solution. (a) The conditional predictive distribution is

$$Z|(Y_{\rm obs}, \theta) \sim \text{Binomial}\left(y_1, \frac{\theta}{\theta+2}\right). \tag{4.1}$$

The complete-data likelihood function of θ is

$$L(\theta|Y_{\rm obs}, z) \propto \theta^{z+y_4}(1-\theta)^{y_2+y_3}.$$

When the prior distribution of θ is Beta(a_0, b_0), the complete-data posterior density of θ is

$$\theta|(Y_{\rm obs}, z) \sim \text{Beta}(z + y_4 + a_0, y_2 + y_3 + b_0), \tag{4.2}$$

so that the complete-data posterior mode is

$$\tilde{\theta} = \frac{z + y_4 + a_0 - 1}{z + y_2 + y_3 + y_4 + a_0 + b_0 - 2}.$$

Replacing z by $E(Z|Y_{\rm obs}, \theta^{(t)}) = y_1\theta^{(t)}/[\theta^{(t)} + 2]$, we have the following EM iteration:

$$\theta^{(t+1)} = \frac{y_1\theta^{(t)}/[\theta^{(t)} + 2] + y_4 + a_0 - 1}{y_1\theta^{(t)}/[\theta^{(t)} + 2] + y_2 + y_3 + y_4 + a_0 + b_0 - 2}. \tag{SA4.7}$$

Let $Y_{\rm obs} = (y_1, \dots, y_4)^{\top} = (125,\ 18, 20, 34)^{\top}$ and $\theta^{(0)} = 0.5$, using (SA4.7), we obtain

$$\begin{array}{lllllllll} \theta^{(1)} & = & 0.6082474, & \theta^{(4)} & = & 0.6267773, & \theta^{(7)} & = & 0.6268214, \\ \theta^{(2)} & = & 0.6243211, & \theta^{(5)} & = & 0.6268156, & \theta^{(8)} & = & 0.6268215, \\ \theta^{(3)} & = & 0.6264889, & \theta^{(6)} & = & 0.6268207, & \theta^{(9)} & = & 0.6268215, \end{array}$$

The R code is as follows:

```
function(ind, th0, NumEM1)
{   # Function name: Linkage.model.EM1.EM2(ind, th0, NumEM1)
```

```
# -------------- Input --------------------------------
# ind    = 1: calculate the posterior mode via (SA4.7)
#        = 2: calculate the convergence rate of
#             the 1-st EM algorithm via (4.7)
#        = 3: calculate the posterior mode via (4.11)
#        = 4: calculate the convergence rate of
#             the 2-nd EM algorithm via (4.14)
# th0    = initial value of \theta, th0 = 0.5
# NumEM1 = the number of iterations in the 1-th & 2-nd EM
# -------------- Output -------------------------------
# TH = approximates of the posterior mode
# r1 = the convergence rate of the 1-st EM algorithm
# r2 = the convergence rate of the 2-nd EM algorithm
# ----------------------------------------------------
y <- c(125, 18, 20, 34)
N <- sum(y)
a0 <- b0 <- 1
if (ind == 1) {
   th <- th0
   TH <- matrix(0, NumEM1, 1)
   for (tt in 1:NumEM1) {
      Ez <- y[1]*th/(th + 2)
      a <- a0+b0-2
      th <- (Ez + y[4] + a0 - 1)/(Ez + y[2]+y[3]+y[4]+a)
      TH[tt] <- th
   }
   return(TH) }
if (ind == 2) {
   tth <- 0.6268215
   b <- (N*tth + 2*(N-y[1]))^2
```

```
        r1 <- abs(2*y[1]*(y[2] + y[3])/b)
        return(r1) }
    if (ind == 3) {
        th <- th0
        TH <- matrix(0, NumEM1, 1)
        for (tt in 1:NumEM1) {
            Ez <- 3*y[1]*th/(th + 2)
            th <- (Ez + y[4] + a0 - 1)/(N+a0+b0-2)
            TH[tt] <- th
        }
        return(TH) }
    if (ind == 4) {
        tth <- 0.6268215
        r2 <- abs(6*y[1]/(N*(tth + 2)^2))
        return(r2) }
}
```

(b) Let $a_0 = b_0 = 1$, the first EM iteration defined by (SA4.7) can be rewritten as

$$\theta^{(t+1)} = h_1(\theta^{(t)}), \tag{4.3}$$

where the fixed-point function is given by

$$h_1(\theta) = \frac{y_4 + y_1\theta/(\theta+2)}{y_4 + y_1\theta/(\theta+2) + y_2 + y_3} = \frac{(y_1+y_4)\theta + 2y_4}{N\theta + 2(N-y_1)},$$

and $N = \sum_{i=1}^4 y_i$. It is easy to derive

$$h_1'(\theta) = \frac{2y_1(y_2+y_3)}{[N\theta + 2(N-y_1)]^2}. \tag{4.4}$$

Let $\tilde\theta$ denote the mode of θ. The first-order Taylor expansion of $h_1(\theta)$

around $\tilde{\theta}$ yields

$$h_1(\theta) = h_1(\tilde{\theta}) + (\theta - \tilde{\theta})h_1'(\tilde{\theta}) + \frac{(\theta - \tilde{\theta})^2}{2!}h_1''(\xi), \tag{4.5}$$

where ξ is a point between θ and $\tilde{\theta}$. Thus, we have

$$\begin{aligned} \theta^{(t+1)} &\overset{(4.3)}{=} h_1(\theta^{(t)}) \\ &\overset{(4.5)}{=} h_1(\tilde{\theta}) + (\theta^{(t)} - \tilde{\theta})h_1'(\tilde{\theta}) + \frac{(\theta^{(t)} - \tilde{\theta})^2}{2}h_1''(\xi^{(t)}) \\ &\overset{(4.3)}{=} \tilde{\theta} + (\theta^{(t)} - \tilde{\theta})h_1'(\tilde{\theta}) + \frac{(\theta^{(t)} - \tilde{\theta})^2}{2}h_1''(\xi^{(t)}), \end{aligned} \tag{4.6}$$

where $\xi^{(t)}$ is a point between $\theta^{(t)}$ and $\tilde{\theta}$. Therefore, the convergence rate of the first EM algorithm is given by

$$\begin{aligned} r_1 &= \lim_{t\to\infty} \frac{|\theta^{(t+1)} - \tilde{\theta}|}{|\theta^{(t)} - \tilde{\theta}|} \\ &\overset{(4.6)}{=} \lim_{t\to\infty} |h_1'(\tilde{\theta}) + 0.5(\theta^{(t)} - \tilde{\theta})h_1''(\xi^{(t)})| \\ &= |h_1'(\tilde{\theta})| \qquad (4.7) \\ &\overset{(4.4)}{=} \left|\frac{2y_1(y_2 + y_3)}{[N\tilde{\theta} + 2(N - y_1)]^2}\right|. \qquad (4.8) \end{aligned}$$

In general $r_1 \in [0, 1)$. When $r_1 = 0$, the algorithm is said to converge *super-linearly*. Thus, we prefer a smaller r_1; in other words, the smaller the convergence rate, the faster the algorithm converges.

In this example, $r_1 = 0.1327787$.

(c) Given $\theta = \theta^{(t)}$, the I-step is to draw $Z^{(t)}$ from (4.1). Given $Z = Z^{(t)}$, the P-step is to draw $\theta^{(t+1)}$ from (4.2).

(d) The conditional predictive distribution is given by

$$Z|(Y_{\text{obs}}, \theta) \sim \text{Binomial}\left(y_1, \frac{3\theta}{\theta+2}\right). \tag{4.9}$$

The likelihood of the complete-data $\{Y_{\text{obs}}, Z\} = \{z, y_1 - z, y_2, y_3, y_4\}$ is

$$\begin{aligned}
L(\theta|Y_{\text{obs}}, z) &= \binom{n}{z, y_1 - z, y_2, y_3, y_4} \left(\frac{3\theta}{4}\right)^z \left[\frac{2(1-\theta)}{4}\right]^{y_1 - z} \\
&\quad \times \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4} \\
&\propto \theta^{z+y_4}(1-\theta)^{y_1-z+y_2+y_3}.
\end{aligned}$$

When the prior of θ is Beta(a_0, b_0), the complete-data posterior density of θ is

$$\theta|(Y_{\text{obs}}, z) \sim \text{Beta}(z + y_4 + a_0, y_1 - z + y_2 + y_3 + b_0), \tag{4.10}$$

so that the complete-data posterior mode is

$$\tilde{\theta} = \frac{z + y_4 + a_0 - 1}{N + a_0 + b_0 - 2}, \qquad N = \sum_{i=1}^{4} y_i.$$

Replacing z by $E(Z|Y_{\text{obs}}, \theta^{(t)}) = 3y_1\theta^{(t)}/[\theta^{(t)}+2]$, we have the 2-nd EM iteration:

$$\theta^{(t+1)} = \frac{3y_1\theta^{(t)}/[\theta^{(t)}+2] + y_4 + a_0 - 1}{N + a_0 + b_0 - 2}. \tag{4.11}$$

Let $Y_{\text{obs}} = (y_1, \ldots, y_4)^\top = (125, 18, 20, 34)^\top$ and $\theta^{(0)} = 0.5$, using (4.11), we obtain

$$\begin{array}{lllllllll}
\theta^{(1)} & = & 0.5532995, & \theta^{(10)} & = & 0.6264501, & \theta^{(19)} & = & 0.6268197, \\
\theta^{(2)} & = & 0.5850885, & \theta^{(11)} & = & 0.6266165, & \theta^{(20)} & = & 0.6268205, \\
\theta^{(3)} & = & 0.6034240, & \theta^{(12)} & = & 0.6267084, & \theta^{(21)} & = & 0.6268210, \\
\theta^{(4)} & = & 0.6137962, & \theta^{(13)} & = & 0.6267591, & \theta^{(22)} & = & 0.6268212, \\
\theta^{(5)} & = & 0.6195991, & \theta^{(14)} & = & 0.6267871, & \theta^{(23)} & = & 0.6268213, \\
\theta^{(6)} & = & 0.6228256, & \theta^{(15)} & = & 0.6268025, & \theta^{(24)} & = & 0.6268214, \\
\theta^{(7)} & = & 0.6246135, & \theta^{(16)} & = & 0.6268110, & \theta^{(25)} & = & 0.6268214, \\
\theta^{(8)} & = & 0.6256022, & \theta^{(17)} & = & 0.6268157, & \theta^{(26)} & = & 0.6268215, \\
\theta^{(9)} & = & 0.6261485, & \theta^{(18)} & = & 0.6268183, & \theta^{(27)} & = & 0.6268215.
\end{array}$$

(e) Let $a_0 = b_0 = 1$, the second EM iteration defined by (4.11) can be rewritten as

$$\theta^{(t+1)} = h_2(\theta^{(t)}), \tag{4.12}$$

where the fixed-point function is given by

$$h_2(\theta) = \frac{y_4 + 3y_1\theta/(\theta+2)}{N},$$

and $N = \sum_{i=1}^4 y_i$. It is easy to derive

$$h_2'(\theta) = \frac{\mathrm{d}h_2(\theta)}{\mathrm{d}\theta} = \frac{6y_1}{N(\theta+2)^2}. \tag{4.13}$$

The convergence rate of the second EM algorithm is given by

$$r_2 = |h_2'(\tilde{\theta})| \stackrel{(4.13)}{=} \left| \frac{6y_1}{N(\tilde{\theta}+2)^2} \right|. \tag{4.14}$$

In this example, $r_2 = 0.5517393$.

4.11 Solution. Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{ZTP}(\lambda)$ and $Y_{\text{obs}} = \{x_i\}_{i=1}^n$ denote the observed data. 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 $Y_{\rm com} = \{y_i\}_{i=1}^n = \{z_i, x_i\}_{i=1}^n$, where $\{Y_i\}_{i=1}^n \stackrel{\rm iid}{\sim} \text{Poisson}(\lambda)$, $\{Z_i\}_{i=1}^n \stackrel{\rm iid}{\sim} \text{Bernoulli}(1 - \mathrm{e}^{-\lambda})$ and $\{Z_1, \ldots, Z_n\} \perp\!\!\!\perp \{X_1, \ldots, X_n\}$.

<u>Complete-data posterior distribution</u>: The complete-data likelihood function of λ is

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

When the prior distribution of λ is $\text{Gamma}(\alpha_0, \beta_0)$, i.e.,

$$\pi(\lambda) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\lambda^{\alpha_0 - 1}\,\mathrm{e}^{-\beta_0\lambda}, \quad \lambda > 0,$$

the complete-data posterior distribution is

$$p(\lambda|Y_{\rm com}) \propto \lambda^{\sum_{i=1}^n y_i}\,\mathrm{e}^{-n\lambda} \cdot \lambda^{\alpha_0 - 1}\,\mathrm{e}^{-\beta_0\lambda} = \lambda^{n\bar{y} + \alpha_0 - 1}\,\mathrm{e}^{-(n+\beta_0)\lambda},$$

i.e., $\lambda|Y_{\rm com} \sim \text{Gamma}(n\bar{y} + \alpha_0, n + \beta_0)$, where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{1}{n}\sum_{i=1}^n z_i x_i.$$

The I-step of the DA algorithm is to independently generate $Z_i = z_i$ from

$$Z_i|(Y_{\rm obs}, \lambda) = Z_i|\lambda \sim \text{Bernoulli}(1 - \mathrm{e}^{-\lambda}),$$

and the P-step is to generate λ from $\text{Gamma}(n\bar{y} + \alpha_0, n + \beta_0)$.