MAT7035: Computational Statistics

Suggested Solutions to Assignment 3

2.9 Solution: Note $\boldsymbol{x}_{(i)} = (x_{i1}, \ldots, x_{iq})^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$, we have

$$\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta} = \sum_{j=1}^{q} x_{ij}\theta_j = \sum_{j=1}^{q} x_{ij}\theta_j - \sum_{j=1}^{q} x_{ij}\theta_j^{(t)} + \sum_{j=1}^{q} x_{ij}\theta_j^{(t)}$$

$$= \sum_{j=1}^{q} x_{ij}\theta_j - \sum_{j=1}^{q} x_{ij}\theta_j^{(t)} + \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)} = \sum_{j=1}^{q} x_{ij}(\theta_j - \theta_j^{(t)}) + 1 \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)}$$

$$= \sum_{j \in \mathbb{J}_i} x_{ij}(\theta_j - \theta_j^{(t)}) + \left(\sum_{j \in \mathbb{J}_i} \lambda_{ij} \right) \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)}$$

$$= \sum_{j \in \mathbb{J}_i} \lambda_{ij} \{ \lambda_{ij}^{-1} x_{ij}(\theta_j - \theta_j^{(t)}) + \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)} \}.$$

Since $\{f_i\}$ are concave function, by using the concavity inequality, i.e., the reverse of inequality (e) in Exercise 2.5, we obtain

$$\begin{aligned} f_i(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}) &= f_i \left(\sum_{j \in \mathbb{J}_i} \lambda_{ij} \{ \lambda_{ij}^{-1} x_{ij}(\theta_j - \theta_j^{(t)}) + \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)} \} \right) \\ &\geqslant \sum_{j \in \mathbb{J}_i} \lambda_{ij} f_i \Big(\lambda_{ij}^{-1} x_{ij}(\theta_j - \theta_j^{(t)}) + \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)} \Big) \end{aligned}$$

so that

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta})$$

$$\geqslant \sum_{i=1}^{m} \sum_{j \in \mathbb{J}_i} \lambda_{ij} f_i \Big(\lambda_{ij}^{-1} x_{ij}(\theta_j - \theta_j^{(t)}) + \boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)} \Big) \;\hat{=}\; Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}),$$

and

$$Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{m}\sum_{j\in\mathbb{J}_i}\lambda_{ij}f_i\Big(\lambda_{ij}^{-1}x_{ij}(\theta_j^{(t)}-\theta_j^{(t)})+\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)}\Big)$$

$$= \sum_{i=1}^{m}\sum_{j\in\mathbb{J}_i}\lambda_{ij}f_i(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{m}f_i(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)}) = \ell(\boldsymbol{\theta}^{(t)}|Y_{\text{obs}}).$$

Therefore, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ defined in (2.72) minorizes $\ell(\boldsymbol{\theta}|Y_{\text{obs}})$ at $\boldsymbol{\theta}^{(t)}$.

2.10 Solution: We rewrite (2.61) as

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \frac{a_j}{b_j}$$

for $1 \leqslant j \leqslant q$, where

$$\begin{aligned} a_j &= \sum_{i\in\mathbb{I}_j}\Big\{y_i - \exp(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})\Big\}\,x_{ij} \quad \text{and} \\ b_j &= \sum_{i\in\mathbb{I}_j}\exp(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})\,x_{ij}^2/\lambda_{ij} \\ &= \sum_{i=1}^{m}\exp(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})\,c_i|x_{ij}|, \quad c_i \mathrel{\hat=} \sum_{j'=1}^{q}|x_{ij'}|. \end{aligned}$$

In the form of vectors, we have

$$\begin{pmatrix}\theta_1^{(t+1)}\\ \vdots \\ \theta_q^{(t+1)}\end{pmatrix} = \begin{pmatrix}\theta_1^{(t)}\\ \vdots \\ \theta_q^{(t)}\end{pmatrix} + \begin{pmatrix}a_1\\ \vdots \\ a_q\end{pmatrix}\Bigg/\begin{pmatrix}b_1\\ \vdots \\ b_q\end{pmatrix},$$

or

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \boldsymbol{a}/\boldsymbol{b},$$

where

$$\boldsymbol{a} = \begin{pmatrix}a_1\\ \vdots \\ a_q\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{m}[y_i - \exp(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})]\,x_{i1}\\ \vdots \\ \sum_{i=1}^{m}[y_i - \exp(\boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})]\,x_{iq}\end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1q} & \cdots & x_{mq} \end{pmatrix} \begin{pmatrix} y_1 - \exp(\boldsymbol{x}_{(1)}^\top \boldsymbol{\theta}^{(t)}) \\ \vdots \\ y_m - \exp(\boldsymbol{x}_{(m)}^\top \boldsymbol{\theta}^{(t)}) \end{pmatrix} \\
&= \boldsymbol{X}^\top \{\boldsymbol{y} - \exp(\boldsymbol{X}\boldsymbol{\theta}^{(t)})\},
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{b} &= \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} \exp(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)})\, c_i |x_{i1}| \\ \vdots \\ \sum_{i=1}^{m} \exp(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}^{(t)})\, c_i |x_{iq}| \end{pmatrix} \\
&= \begin{pmatrix} c_1|x_{11}| & \cdots & c_m|x_{m1}| \\ \vdots & \ddots & \vdots \\ c_1|x_{1q}| & \cdots & c_m|x_{mq}| \end{pmatrix} \begin{pmatrix} \exp(\boldsymbol{x}_{(1)}^\top \boldsymbol{\theta}^{(t)}) \\ \vdots \\ \exp(\boldsymbol{x}_{(m)}^\top \boldsymbol{\theta}^{(t)}) \end{pmatrix} \\
&= \begin{pmatrix} |x_{11}| & \cdots & |x_{m1}| \\ \vdots & \ddots & \vdots \\ |x_{1q}| & \cdots & |x_{mq}| \end{pmatrix} \begin{pmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_m \end{pmatrix} \exp(\boldsymbol{X}\boldsymbol{\theta}^{(t)}) \\
&= \boldsymbol{Y}^\top \begin{pmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_m \end{pmatrix} \exp(\boldsymbol{X}\boldsymbol{\theta}^{(t)}).
\end{aligned}
$$

We only need to prove that

$$
\operatorname{diag}(\boldsymbol{Y}\mathbf{1}_q) = \begin{pmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_m \end{pmatrix}. \tag{SA3.1}
$$

In fact,

$$
\boldsymbol{Y}\mathbf{1}_q = \begin{pmatrix} |x_{11}| & \cdots & |x_{1q}| \\ \vdots & \ddots & \vdots \\ |x_{m1}| & \cdots & |x_{mq}| \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^q |x_{1j}| \\ \vdots \\ \sum_{j=1}^q |x_{mj}| \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix},
$$

we immediately obtain (SA3.1).

2.11 Solution: See Example T7.3 in Tutorial 7.

2.12 Solution: The probit regression model is given by

$$Y_i \stackrel{\text{ind}}{\sim} \text{Binomial}(n_i, p_i), \quad p_i = \Phi(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}), \quad i = 1, \ldots, m,$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$. Let $Y_{\text{obs}} = \{y_i\}_{i=1}^m$ denote the observed data, then the observed-data likelihood function of $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}|Y_{\text{obs}}) = \prod_{i=1}^m \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i},$$

and the observed-data log-likelihood function is

$$\begin{aligned}
\ell(\boldsymbol{\theta}|Y_{\text{obs}}) &= c + \sum_{i=1}^m [y_i \log p_i + (n_i - y_i)\log(1-p_i)] \\
&= c + \sum_{i=1}^m \Big\{ y_i \log \Phi(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}) + (n_i - y_i)\log[1 - \Phi(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta})] \Big\} \\
&= c + \sum_{i=1}^m f_i(\boldsymbol{x}_{(i)}^\top \boldsymbol{\theta}),
\end{aligned}$$

where c is a constant,

$$f_i(u) = y_i \log \Phi(u) + (n_i - y_i)\log[1 - \Phi(u)].$$

Note that

$$\begin{aligned}
f_i'(u) &= y_i \frac{\phi(u)}{\Phi(u)} - (n_i - y_i)\frac{\phi(u)}{1 - \Phi(u)} \quad \text{and} \\
f_i''(u) &= y_i \frac{\phi'(u)\Phi(u) - \phi^2(u)}{\Phi^2(u)} - (n_i - y_i)\frac{\phi'(u)[1-\Phi(u)] + \phi^2(u)}{[1-\Phi(u)]^2} \\
&= -\phi(u)\left\{ y_i \frac{u\Phi(u) + \phi(u)}{\Phi^2(u)} + (n_i - y_i)\frac{-u[1-\Phi(u)] + \phi(u)}{[1-\Phi(u)]^2} \right\}, \\
&\hat{=} -\phi(u)\left\{ y_i \frac{A(u)}{\Phi^2(u)} + (n_i - y_i)\frac{B(u)}{[1-\Phi(u)]^2} \right\},
\end{aligned}$$

where $\phi(u)$ is the pdf of $N(0,1)$, $\phi'(u) = -u\phi(u)$,

$$A(u) = u\Phi(u) + \phi(u) \quad \text{and} \quad B(u) = -u[1-\Phi(u)] + \phi(u).$$

We first prove that $f_i''(u) \leqslant 0$ by using the following Lemma 2.1.

Lemma 2.1 *For the standard normal distribution, define the failure (hazard) rate as* $\Psi(u) = \phi(u)/[1-\Phi(u)]$. *Then for all* $u \geqslant 0$, *the following inequalities hold:*

$$u < \frac{\sqrt{u^2+8}+3u}{4} < \Psi(u) < \frac{\sqrt{u^2+4}+u}{2}. \tag{SA3.T2}$$

The proof of this result can be found in Theorem 2.3 of Baricz, Á. (2008). Mills' ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications* **340**, 1362–1370.

(i) When $u \geqslant 0$, we have $A(u) = u\Phi(u) + \phi(u) > 0$ and

$$\begin{aligned} B(u) &= -u[1-\Phi(u)] + \phi(u) = [1-\Phi(u)]\left[-u + \frac{\phi(u)}{1-\Phi(u)}\right] \\ &= [1-\Phi(u)] \cdot [-u + \Psi(u)] \overset{\text{(SA3.T2)}}{>} 0, \end{aligned}$$

so that $f_i''(u) \leqslant 0$. (ii) When $u < 0$, we have $B(u) > 0$ and

$$\begin{aligned} &\Rightarrow \quad -u > 0, \quad \Rightarrow \quad -u \overset{\text{(SA3.T2)}}{<} \Psi(-u), \\ &\Rightarrow \quad 0 < u + \Psi(-u) = u + \frac{\phi(-u)}{1-\Phi(-u)} = u + \frac{\phi(u)}{\Phi(u)}, \\ &\Rightarrow \quad \frac{A(u)}{\Phi(u)} = \frac{u\Phi(u) + \phi(u)}{\Phi(u)} = u + \frac{\phi(u)}{\Phi(u)} > 0, \quad \Rightarrow \quad A(u) > 0, \end{aligned}$$

so that $f_i''(u) \leqslant 0$.

Finally, the DP algorithm is defined by (2.54) and (2.55) in the textbook.

2.13 Solution: (a) The likelihood function is

$$L(\boldsymbol{\theta}, \sigma^2 | Y_{\text{obs}}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - \boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta})^2}{2\sigma^2} \right\},$$

so that the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}, \sigma^2 | Y_{\text{obs}}) = \text{constant} - \frac{m\log(\sigma^2)}{2} - \frac{\sum_{i=1}^{m}(y_i - \boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta})^2}{2\sigma^2}.$$

The MLEs of σ^2 and $\boldsymbol{\theta}$ are given by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m}(y_i - \boldsymbol{x}_{(i)}^{\top}\hat{\boldsymbol{\theta}})^2}{m} \quad \text{and} \quad \hat{\boldsymbol{\theta}} = \arg\max \sum_{i=1}^{m} \Big\{ -(y_i - \boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta})^2 \Big\}.$$

Define $f_i(x) = -(y_i - x)^2$, then

$$f_i'(x) = 2(y_i - x) \quad \text{and} \quad f_i''(x) = -2 < 0,$$

respectively. Thus $f_i(\cdot)$ is strictly concave. From (2.54) and (2.55), we obtain

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \frac{\sum_{i \in \mathbb{I}_j}(y_i - \boldsymbol{x}_{(i)}^{\top}\boldsymbol{\theta}^{(t)})x_{ij}}{\sum_{i \in \mathbb{I}_j} x_{ij}^2/\lambda_{ij}}, \quad 1 \leqslant j \leqslant q.$$

Similar to Exercise 2.10, we have in the form of matrices and vectors

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \boldsymbol{X}^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}^{(t)})/\boldsymbol{Z}^{\top}, \tag{SA3.2}$$

where $\boldsymbol{X} = (x_{ij}) = (\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(m)})^{\top}$, $\boldsymbol{Y} = (|x_{ij}|) = \text{abs}(\boldsymbol{X})$, and $\boldsymbol{Z} = \text{diag}(\boldsymbol{Y}\boldsymbol{1}_q)Y$.

(b) Although we can obtain the MLE of $\boldsymbol{\theta}$ in the closed-form

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}, \tag{SA3.3}$$

when q is very large, $\boldsymbol{X}^{\top}\boldsymbol{X}$ is a $q \times q$ matrix, whose inverse matrix is quite difficult to calculate in a computer. On the other hand, when

$\boldsymbol{X}^\top\boldsymbol{X}$ is almost singular, the calculation of $\hat{\boldsymbol{\theta}}$ in (SA3.3) is numerically unstable or even is not possible. However, the DP algorithm defined in (SA3.2) completely avoided the calculation of an inversion matrix and it has the ascent property, i.e., each iteration will increase the likelihood.

2.14 Proof: Since $\boldsymbol{A}_{m\times m} > 0$, there is an $m \times m$ orthogonal matrix $\boldsymbol{\Gamma}$ satisfying $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{I}_m$ such that

$$\boldsymbol{A} = \boldsymbol{\Gamma}\mathrm{diag}(\theta_1, \ldots, \theta_m)\boldsymbol{\Gamma}^\top,$$

where $\theta_1 \geqslant \cdots \geqslant \theta_m > 0$ be the eigenvalues of $\boldsymbol{A}$. Thus

$$\begin{aligned}
|\boldsymbol{A}| = \det(\boldsymbol{A}) &= |\boldsymbol{\Gamma}| \cdot |\mathrm{diag}(\theta_1, \ldots, \theta_m)| \cdot |\boldsymbol{\Gamma}^\top| = \prod_{i=1}^m \theta_i, \\
\mathrm{tr}\,(\boldsymbol{A}) &= \mathrm{tr}\,[\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\mathrm{diag}(\theta_1, \ldots, \theta_m)] = \sum_{i=1}^m \theta_i, \\
f(\boldsymbol{A}) &= c|\boldsymbol{A}|^{n/2}\exp(-0.5\,\mathrm{tr}\,\boldsymbol{A}) = c\prod_{i=1}^m \theta_i^{n/2}\,\mathrm{e}^{-0.5\theta_i}, \\
\log[f(\boldsymbol{A})] &= \log(c) + 0.5\sum_{i=1}^m [n\log(\theta_i) - \theta_i].
\end{aligned}$$

Let

$$0 = \frac{\partial \log[f(\boldsymbol{A})]}{\partial\theta_i} = 0.5\left(\frac{n}{\theta_i} - 1\right), \quad i = 1, \ldots, m,$$

we have $\theta_i = n$ for $i = 1, \ldots, m$. Thus $f(\boldsymbol{A})$ reaches its maximum when $\theta_1 = \cdots = \theta_m = n$, i.e., $\boldsymbol{A} = n\boldsymbol{I}_m$.

2.25 Solution: (a) Let $f(x) = -\sqrt{x}$. Since $f'(x) = -1/(2\sqrt{x})$ and $f''(x) = 1/(4x^{3/2}) > 0$, the second order Taylor expansion of $f(x)$ around x_0 yields

$$\begin{aligned}
f(x) &= f(x_0) + (x - x_0)f'(x_0) + 0.5(x - x_0)^2 f''(\xi), \quad \xi \in (x, x_0), \\
&\geqslant f(x_0) + (x - x_0)f'(x_0),
\end{aligned}$$

i.e.,

$$-\sqrt{x} \geqslant -\sqrt{x_0} - \frac{1}{2\sqrt{x_0}}(x - x_0), \quad \forall\, x, x_0 > 0. \tag{SA3.4}$$

(b) In (SA3.4), let $x = a^2 + \theta^2$ and $x_0 = a^2 + \theta^{(t)2}$, we have

$$\begin{aligned} -\sqrt{a^2+\theta^2} &\geqslant -\sqrt{a^2+\theta^{(t)2}} - \frac{1}{2\sqrt{a^2+\theta^{(t)2}}}(\theta^2 - \theta^{(t)2}) \\ &= c_1^{(t)} - \frac{\theta^2}{2\sqrt{a^2+\theta^{(t)2}}}, \end{aligned}$$

where $c_1^{(t)}$ is a constant and does not depend on θ. Similarly, in (SA3.4), let $x = b^2 + (c-\theta)^2$ and $x_0 = b^2 + (c - \theta^{(t)})^2$, we have

$$\begin{aligned} -\sqrt{b^2+(c-\theta)^2} &\geqslant -\sqrt{b^2+(c-\theta^{(t)})^2} - \frac{(c-\theta)^2 - (c-\theta^{(t)})^2}{2\sqrt{b^2+(c-\theta^{(t)})^2}} \\ &= c_2^{(t)} - \frac{(c-\theta)^2}{2\sqrt{b^2+(c-\theta^{(t)})^2}}, \end{aligned}$$

where $c_2^{(t)}$ is also a constant and does not depend on θ. Thus, the log-likelihood function is given by

$$\begin{aligned} \ell(\theta) &= -\frac{\sqrt{a^2+\theta^2}}{s_1} - \frac{\sqrt{b^2+(c-\theta)^2}}{s_2} \\ &\geqslant c_3^{(t)} - \frac{\theta^2}{2s_1\sqrt{a^2+\theta^{(t)2}}} - \frac{(c-\theta)^2}{2s_2\sqrt{b^2+(c-\theta^{(t)})^2}} \\ &\hat{=} c_3^{(t)} - \frac{\theta^2}{2a_1^{(t)}} - \frac{(c-\theta)^2}{2b_1^{(t)}} \\ &\hat{=} Q(\theta|\theta^{(t)}), \end{aligned}$$

where $a_1^{(t)} = s_1\sqrt{a^2+\theta^{(t)2}}$ and $b_1^{(t)} = s_2\sqrt{b^2+(c-\theta^{(t)})^2}$. Let

$$0 = \frac{\mathrm{d}Q(\theta|\theta^{(t)})}{\mathrm{d}\theta} = -\frac{\theta}{a_1^{(t)}} + \frac{c-\theta}{b_1^{(t)}},$$

we obtain

$$\theta^{(t+1)} = \frac{c}{1 + b_1^{(t)}/a_1^{(t)}}.$$

(c) The corresponding R code and the output are as follows:

```
function (th0, n)
{# Function Name: CS.SA2.25(th0, n)
   a <- 3
   b <- -1
   cc <- 2
   s1 <- 1
   s2 <- 1.5
   th <- th0
   TH <- rep(0, n)
   for (tt in 1:n) {
      at <- s1 * sqrt(a^2 + th^2)
      bt <- s2 * sqrt(b^2 + (cc - th)^2)
      th <- cc /(1 + bt/at)
      TH[tt] <- th
   }
   return(TH)
 }
```

```
> CS.SA2.25(0, 10)
 [1] 0.9442719 1.1809632 1.2489208 1.2679930 1.2732721
 [6] 1.2747266 1.2751269 1.2752370 1.2752673 1.2752756
```

2.26 Solution: (a) Let $f(x) = -\log(x)$. Since $f'(x) = -1/x$ and $f''(x) = 1/x^2 > 0$, the second order Taylor expansion of $f(x)$ around x_0 yields

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)f'(x_0) + 0.5(x - x_0)^2 f''(\xi), \quad \xi \in (x, x_0), \\ &\geqslant f(x_0) + (x - x_0)f'(x_0), \end{aligned}$$

i.e.,

$$-\log x \geqslant -\log x_0 + (x - x_0)(-x_0^{-1}), \quad \forall\, x, x_0 > 0. \tag{SA3.5}$$

(b) In (SA3.5), let $x = \theta_i + \theta_j$ and $x_0 = \theta_i^{(t)} + \theta_j^{(t)}$, we have

$$-\log(\theta_i + \theta_j) \geqslant -\log(\theta_i^{(t)} + \theta_j^{(t)}) - \frac{\theta_i + \theta_j}{\theta_i^{(t)} + \theta_j^{(t)}} + 1. \tag{SA3.6}$$

Define $y_{i+} = \sum_{j=1}^n y_{ij}$. The log-likelihood function is given by

$$\begin{aligned}
\ell(\boldsymbol{\theta}) \quad &= \quad \sum_{i=1}^n \sum_{j=1}^n y_{ij}\{\log\theta_i - \log(\theta_i + \theta_j)\} \\
&\overset{\text{(SA3.6)}}{\geqslant} \quad \sum_i \sum_j y_{ij} \left\{ \log\theta_i - \log(\theta_i^{(t)} + \theta_j^{(t)}) - \frac{\theta_i + \theta_j}{\theta_i^{(t)} + \theta_j^{(t)}} + 1 \right\} \\
&\hat{=} \quad Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \\
&= \quad \sum_i \log\theta_i \left(\textstyle\sum_j y_{ij}\right) - \sum_i \sum_j \frac{y_{ij}(\theta_i + \theta_j)}{\theta_i^{(t)} + \theta_j^{(t)}} \\
&\quad\quad + \sum_i \sum_j y_{ij} \left\{ 1 - \log(\theta_i^{(t)} + \theta_j^{(t)}) \right\}.
\end{aligned}$$

Let

$$0 = \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\theta_i} = \frac{y_{i+}}{\theta_i} - \sum_{i=1}^n \frac{y_{ij}}{\theta_i^{(t)} + \theta_j^{(t)}} - \sum_{j=1}^n \frac{y_{ij}}{\theta_i^{(t)} + \theta_j^{(t)}},$$

we obtain

$$\theta_i^{(t+1)} = \frac{y_{i+}}{\sum_{i=1}^n \frac{y_{ij}}{\theta_i^{(t)} + \theta_j^{(t)}} + \sum_{j=1}^n \frac{y_{ij}}{\theta_i^{(t)} + \theta_j^{(t)}}}, \quad i = 1, \ldots, n.$$

2.27 Solution: (a) Let $X \sim f(x|\theta, \lambda)$, where

$$f(x|\theta, \lambda) = \frac{\theta}{\lambda}\left(\frac{x}{\lambda}\right)^{\theta-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{\theta}\right], \quad x > 0.$$

When θ is known, we define $Y = X^\theta$ (> 0), then the density of Y is

$$\begin{aligned} f_{_Y}(y|\lambda) &= f(x|\theta,\lambda) \times \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \\ &= \frac{\theta}{\lambda}\left(\frac{x}{\lambda}\right)^{\theta-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{\theta}\right] \times \frac{y^{1/\theta-1}}{\theta} \\ &= \beta\exp(-\beta y), \quad y > 0, \end{aligned}$$

where $\beta = 1/\lambda^\theta > 0$, i.e., $Y \sim \text{Exponential}(\beta)$.

Hence, if $X_1, \ldots, X_m \overset{\text{iid}}{\sim} f(x|\theta,\lambda)$, we have $Y_1, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Exponential}(\beta)$, where $Y_i = X_i^\theta$ for $i = 1, \ldots, m$. Note that the MLE of β is $1/\bar{Y}$, i.e., the MLE of $1/\beta$ is $\bar{Y}$, then the MLE of λ satisfies

$$\lambda^\theta = \frac{1}{m}\sum_{i=1}^{m} x_i^\theta. \tag{SA3.7}$$

(b) The joint density of $X_1, \ldots, X_m \overset{\text{iid}}{\sim} f(x|\theta,\lambda)$ is given by

$$\prod_{i=1}^{m} f(x_i|\theta,\lambda)$$

so that the likelihood function of (θ, λ) is

$$L(\theta,\lambda) = \prod_{i=1}^{m} \frac{\theta}{\lambda}\left(\frac{x_i}{\lambda}\right)^{\theta-1} \exp\left[-\left(\frac{x_i}{\lambda}\right)^{\theta}\right]$$

and the log-likelihood function is

$$\begin{aligned} \ell(\theta,\lambda) &= m\log(\theta) - m\log(\lambda^\theta) + (\theta-1)\sum_{i=1}^{m}\log x_i - \frac{\sum_{i=1}^{m} x_i^\theta}{\lambda^\theta} \\ &\overset{\text{(SA3.7)}}{=} m\log(\theta) - m\log\left(\frac{\sum_{i=1}^{m} x_i^\theta}{m}\right) \end{aligned}$$

$$
\begin{aligned}
& \quad + (\theta-1)\sum_{i=1}^{m}\log x_i - m \\
&= m\log(\theta) - m\log\left(\sum_{i=1}^{m} x_i^{\theta}\right) + (\theta-1)\sum_{i=1}^{m}\log x_i + c \\
&\hat{=} \ell_1(\theta),
\end{aligned}
$$

where c is a constant. The MLE of θ is given by

$$\hat{\theta} = \arg\max_{\theta>0} \ell_1(\theta).$$

We have

$$
\begin{aligned}
\ell_1'(\theta) &= \frac{m}{\theta} - \frac{m\sum_{i=1}^{m} x_i^{\theta}\log x_i}{\sum_{i=1}^{m} x_i^{\theta}} + \sum_{i=1}^{m}\log x_i \quad \text{and} \\
-\ell_1''(\theta) &= \frac{m}{\theta^2} + m\frac{[\sum_{i=1}^{m} x_i^{\theta}(\log x_i)^2]\sum_{i=1}^{m} x_i^{\theta} - (\sum_{i=1}^{m} x_i^{\theta}\log x_i)^2}{(\sum_{i=1}^{m} x_i^{\theta})^2} \\
&= \frac{m}{\theta^2} + \frac{m\sum_{i=1}^{m} x_i^{\theta}(\log x_i)^2}{\sum_{i=1}^{m} x_i^{\theta}} - \frac{m(\sum_{i=1}^{m} x_i^{\theta}\log x_i)^2}{(\sum_{i=1}^{m} x_i^{\theta})^2}.
\end{aligned}
$$

The Newton method can be used to find

$$\theta^{(t+1)} = \theta^{(t)} + [-\ell_1''(\theta^{(t)})]^{-1}\ell_1'(\theta^{(t)}).$$

3.1 Solution: Let

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1}\,\mathrm{e}^{-\beta x},$$

then, $\log[f(x)] = c + (\alpha-1)\log x - \beta x$. Solving

$$0 = \frac{\mathrm{d}\log[f(x)]}{\mathrm{d}x} = \frac{\alpha-1}{x} - \beta,$$

we obtain the mode of $f(x)$, given by $\tilde{x} = (\alpha-1)/\beta$. Then, we have

$$
\begin{aligned}
h(x) &= \frac{\log[f(x)]}{n} = \frac{c + (\alpha-1)\log x - \beta x}{n}, \\
h'(x) &= \frac{\alpha-1}{nx} - \frac{\beta}{n}, \quad \text{and} \quad h''(x) = -\frac{\alpha-1}{nx^2}.
\end{aligned}
$$

Since $\sigma^2 = -1/[nh''(\tilde{x})]$, we obtain

$$\sigma = \sqrt{\frac{\tilde{x}^2}{\alpha - 1}} = \frac{\sqrt{\alpha - 1}}{\beta}.$$

For $\alpha = 5$ and $\beta = 0.5$, we obtain $\tilde{x} = 8$ and $\sigma = 4$. In Table 3.1, we see that the first-order Laplace approximation works better in the central interval of the density, but the accuracy is not very high when the integral is enlarged.

Table 3.1 *The first-order Laplace approximation to gamma integral for $\alpha = 5$ and $\beta = 0.5$*

Interval (a, b)	Exact	Approximation
$(7, 9)$	0.1933414	0.1933507
$(6, 10)$	0.3747700	0.3750458
$(2, 14)$	0.8233485	0.8485588
$(15.987, \infty)$	0.1000051	0.0224544

The corresponding *R* code and output are as follows:

```
f <- function(x){ 0.5^5*x^4*exp(-0.5*x)/gamma(5) }

f1 <- function(f,a,b){integrate(f, lower=a, upper=b)}

f2 <- function(a, b)
{ f(8)*4*sqrt(2*pi)*(pnorm((b-8)/4)-pnorm((a-8)/4)) }

> f1(f,7,9)
0.1933414 with absolute error < 2.1e-15
> f1(f,6,10)
0.37477 with absolute error < 4.2e-15
> f1(f,2,14)
```

```
0.8233485 with absolute error < 9.1e-15
> f1(f,15.987,Inf)
0.1000051 with absolute error < 4.7e-05

> f2(7,9)
[1] 0.1933507

> f2(6,10)
[1]0.3750458

> f2(2,14)
[1]0.8485588

> f2(15.987,Inf)
[1] 0.02245444
```

3.2 Solution: (a) Define $h(x) = \sqrt{2\pi}x/(1+x^2)$ and $f(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-(x-x_0)^2/2}$, we have

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2}\,\mathrm{e}^{-(x-x_0)^2/2}\mathrm{d}x = \int_{-\infty}^{\infty} h(x)f(x)\mathrm{d}x = E\{h(X)\} \,\hat{=}\, \mu,$$

where $X \sim f(x)$. We generate $X_1, \ldots, X_m \overset{\text{iid}}{\sim} f(x)$ and approximate μ by the empirical average

$$\bar{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} h(X_i) = \frac{1}{m}\sum_{i=1}^{m} \frac{\sqrt{2\pi}\,X_i}{1+X_i^2}.$$

(b) Next,

$$\begin{aligned}
\int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\mathrm{d}x &= \int_{-\infty}^{\infty} I(x \leqslant x_0)\cdot\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\mathrm{d}x \\
&= E[I(X \leqslant x_0)] \,\hat{=}\, \mu,
\end{aligned}$$

where $X \sim N(0,1)$. We generate $X_1, \ldots, X_m \stackrel{\text{iid}}{\sim} N(0,1)$ and approximate μ by the empirical average

$$\bar{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} I(X_i \leqslant x_0).$$

The corresponding R codes and outputs are as follows:

```
data1 <- function(n, x_0) { rnorm(n, mean=x_0, sd=1) }

mu1 <- function(n, x_0)
    { mean(sqrt(2*pi)*data1(n,x_0)/(1+data1(n,x_0)^2)) }

data2<-function(n) { rnorm(n,mean=0,sd=1) }

Indicator <- function(x, max=x_0) { as.numeric(x <= max) }

mu2 <- function(n, x_0)
    { mean(Indicator(data2(n),max=x_0)) }

> mu1(1000,-2)
[1] -1.405382

> mu1(10000,-2)
[1] -1.40748

> mu1(10000,0)
[1] -0.01203916

> mu1(10000,2)
[1] 1.444014
```

```
> mu2(1000,-2)
[1] 0.026

> mu2(10000,-2)
[1] 0.0226

> mu2(10000,0)
[1] 0.508

> mu2(10000,2)
[1] 0.9769
```

3.4 Solution: (a) Note that

$$\int_0^1 \cos(\pi x/2)\,\mathrm{d}x = \frac{2}{\pi}\sin(\pi x/2)|_0^1 = \frac{2}{\pi}.$$

Let $f(x) = 1$ denote the pdf of the random variable $X \sim U(0,1)$, then

$$\frac{2}{\pi} = \int_0^1 \cos(\pi x/2)\,\mathrm{d}x = \int_0^1 \cos(\pi x/2)\cdot f(x)\,\mathrm{d}x = E[\cos(\pi X/2)].$$

(b) Let $X \sim U(0,1)$, then

$$\begin{aligned}
& E[\cos^2(\pi X/2)] \\
= {} & \int_0^1 \cos^2(\pi x/2)\,\mathrm{d}x = \frac{2}{\pi}\int_0^1 \cos(\pi x/2)\,\mathrm{d}\sin(\pi x/2) \\
= {} & \frac{2}{\pi}\cos(\pi x/2)\sin(\pi x/2)|_0^1 - \frac{2}{\pi}\int_0^1 \sin(\pi x/2)\,\mathrm{d}\cos(\pi x/2) \\
= {} & \int_0^1 \sin^2(\pi x/2)\,\mathrm{d}x = \int_0^1 [1-\cos^2(\pi x/2)]\,\mathrm{d}x \\
= {} & 1 - E[\cos^2(\pi X/2)],
\end{aligned}$$

i.e., $E[\cos^2(\pi X/2)] = 1/2$. Thus

$$\begin{aligned} \mathrm{Var}\left[\cos(\pi X/2)\right] &= E[\cos^2(\pi X/2)] - [E\cos(\pi X/2)]^2 \\ &= 1/2 - (2/\pi)^2 \doteq 0.095. \end{aligned}$$

(c) Define $Z = \frac{2\cos(\pi Y/2)}{3(1-Y^2)}$, since

$$E(Z) = \int_0^1 \frac{2\cos(\pi y/2)}{3(1-y^2)} \cdot g(y)\,\mathrm{d}y = \int_0^1 \cos(\pi y/2)\,\mathrm{d}y = \frac{2}{\pi},$$

we have

$$\begin{aligned} &\mathrm{Var}\left[\frac{2\cos(\pi Y/2)}{3(1-Y^2)}\right] = \mathrm{Var}(Z) = E[Z - 2/\pi]^2 \\ = &\int_0^1 \left[\frac{2\cos(\pi y/2)}{3(1-y^2)} - \frac{2}{\pi}\right]^2 \cdot g(y)\,\mathrm{d}y \\ = &\int_0^1 \left[\frac{2\cos(\pi y/2)}{3(1-y^2)} - \frac{2}{\pi}\right]^2 \cdot 1.5(1-y^2)\,\mathrm{d}y \\ \doteq &\ \frac{1}{n}\sum_{i=1}^n \left[\frac{2\cos(\pi y_i/2)}{3(1-y_i^2)} - \frac{2}{\pi}\right]^2 \cdot 1.5(1-y_i^2), \end{aligned}$$

where $y_1, \ldots, y_n \overset{\text{iid}}{\sim} U(0,1)$. The corresponding `R` code and the output are as follows:

```
function (n)
{ #Function Name: CS.SA3.4(n=20000)
  y <- runif(n, min=0,max=1)
  z <- 2*cos(pi*y/2)/(3*(1-y^2))
  hy <- (z - 2/pi)^2 * 1.5 *(1-y^2)
  Var <- mean(hy)
  return(Var)
}
```

```
> CS.SA3.4(20000)
[1] 0.00099647
```