Computational Statistics

Assignment 3

Hanbin Liu 11912410

2.9

Since $\{f_i(\cdot)\}_{i=1}^m$ are strictly concave functions, $Q(\theta|\theta^{(t)})$ is concave. Then by using the concavity inequality, we have

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

since $\sum_{j\in\mathbb{J}_i}\lambda_{ij}=1$ and $x_{(i)}^\top\theta=\sum_{j\in\mathbb{J}_i}\lambda_{ij}\Big\{\lambda_{ij}^{-1}x_{ij}(\theta_j-\theta_j^{(t)})+x_{(i)}^\top\theta^{(t)}\Big\}$. Thus,

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

and

$$Q(\theta^{(t)}|\theta^{(t)})=\sum_{i=1}^m\sum_{j\in\mathbb{J}_i}\lambda_{ij}f_i(x_{(i)}^\top\theta^{(t)})=\sum_{i=1}^m f_i(x_{(i)}^\top\theta^{(t)})=l(\theta^{(t)}).$$

This finishes the proof. $\square$

2.10

We have

$$\mathbf{X}^\top(\mathbf{y}-e^{\mathbf{X}\theta^{(t)}})=\begin{pmatrix}x_{(1)} & \dots & x_{(m)}\end{pmatrix}\begin{pmatrix}y_1-e^{x_{(1)}^\top\theta^{(t)}}\\ \vdots\\ y_m-e^{x_{(m)}^\top\theta^{(t)}}\end{pmatrix}$$

$$\mathbf{Y}\mathbf{1}_q=\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\mathbf{1}_q\\ \vdots\\ \mathrm{abs}(x_{(m)}^\top)\mathbf{1}_q\end{pmatrix},\quad \mathrm{diag}(\mathbf{Y}\mathbf{1}_q)=\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\mathbf{1}_q & & \\ & \ddots & \\ & & \mathrm{abs}(x_{(m)}^\top)\mathbf{1}_q\end{pmatrix}$$

$$\mathbf{Z}=\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\mathbf{1}_q & & \\ & \ddots & \\ & & \mathrm{abs}(x_{(m)}^\top)\mathbf{1}_q\end{pmatrix}\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\\ \vdots\\ \mathrm{abs}(x_{(m)}^\top)\end{pmatrix}=\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\mathbf{1}_q\cdot\mathrm{abs}(x_{(1)}^\top)\\ \vdots\\ \mathrm{abs}(x_{(m)}^\top)\mathbf{1}_q\cdot\mathrm{abs}(x_{(m)}^\top)\end{pmatrix}$$

$$\mathbf{Z}^\top e^{\mathbf{X}\theta^{(t)}}=\begin{pmatrix}\mathrm{abs}(x_{(1)}^\top)\mathbf{1}_q\cdot\mathrm{abs}(x_{(1)}) & \cdots & \mathrm{abs}(x_{(m)}^\top)\mathbf{1}_q\cdot\mathrm{abs}(x_{(m)})\end{pmatrix}\begin{pmatrix}e^{x_{(1)}^\top\theta^{(t)}}\\ \vdots\\ e^{x_{(m)}^\top\theta^{(t)}}\end{pmatrix}$$

Let $A_{q\times1}^{(t)}$ denote $\mathbf{X}^\top(\mathbf{y}-e^{\mathbf{X}\theta^{(t)}})$, $B_{q\times1}^{(t)}$ denote $\mathbf{Z}^\top e^{\mathbf{X}\theta^{(t)}}$ and $A_{q\times1}^{(t)}=\begin{pmatrix}a_1^{(t)} & \dots & a_q^{(t)}\end{pmatrix}^\top$,$B_{q\times1}^{(t)}=\begin{pmatrix}b_1^{(t)} & \dots & b_q^{(t)}\end{pmatrix}^\top$. It then follows that

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

Thus, it suffices to show that $\sum_{i\in\mathbb{I}_j}\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}x_{ij} = a_j^{(t)}$ and $\sum_{i\in\mathbb{I}_j}\exp(x_{(i)}^\top\theta^{(t)})x_{ij}^2/\lambda_{ij} = b_{ij}^{(t)}$. Note that

$$A^{(t)} = \begin{pmatrix} x_{11} & \cdots & x_{m1} \\ x_{12} & \cdots & x_{m2} \\ \vdots & \ddots & \vdots \\ x_{1q} & \cdots & x_{mq} \end{pmatrix} \begin{pmatrix} y_1 - e^{x_{(1)}^\top\theta^{(t)}} \\ \vdots \\ y_m - e^{x_{(m)}^\top\theta^{(t)}} \end{pmatrix},$$

which implies that $a_j^{(t)} = \sum_{i=1}^q\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}x_{ij} = \sum_{i\in\mathbb{I}_j}\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}x_{ij}$. Similarly, by the definition of $B^{(t)}$, we have

$$\begin{aligned} b_j^{(t)} &= \sum_{i=1}^q\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}|x_{ij}|\sum_{j=1}^q|x_{ij}| \\ &= \sum_{i=1}^q\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}|x_{ij}|^2\frac{\sum_{j'\in\mathbb{J}_i}|x_{ij'}|}{|x_{ij}|} \\ &= \sum_{i=1}^q\left\{y_i - \exp(x_{(i)}^\top\theta^{(t)})\right\}|x_{ij}|^2/\lambda_{ij}, \end{aligned}$$

which completes the proof. □

2.12

The log-likelihood function is given by

$$\begin{aligned} l(\theta|Y_{\text{obs}}) &= \sum_{i=1}^m\log\binom{n_i}{y_i} + \sum_{i=1}^m y_i\log p_i + \sum_{i=1}^m(n_i - y_i)\log(1-p_i) \\ &= \sum_{i=1}^m\log\binom{n_i}{y_i} + \sum_{i=1}^m y_i\log\Phi(x_{(i)}^\top\theta) + \sum_{i=1}^m(n_i-y_i)\log(1-\Phi(x_{(i)}^\top\theta)) \\ &= c + \sum_{i=1}^m\left\{y_i\log\Phi(x_{(i)}^\top\theta) + (n_i-y_i)\log(1-\Phi(x_{(i)}^\top\theta))\right\} \\ &= c + \sum_{i=1}^m f_i(x_{(i)}^\top\theta), \end{aligned}$$

where

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

Noting that

$$f_i'(u) = y_i\frac{\phi(u)}{\Phi(u)} - (n_i - y_i)\frac{\phi(u)}{1-\Phi(u)}$$

and

$$-f_i''(u) = y_i\frac{u\phi(u)\Phi(u) + \phi^2(u)}{\Phi^2(u)} + (n_i - y_i)\frac{\phi^2(u) - u\phi(u) + u\phi(u)\Phi(u)}{(1-\Phi^2(u))^2}$$

Thus, the DP algorithm is given by

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \tau_j^2(\theta^{(t)})\sum_{i\in\mathbb{I}_j}f_i'(x_{(i)}^\top\theta^{(t)})x_{ij},$$

where

$$\tau_j^2(\theta) = \left[\sum_{i\in\mathbb{I}_j}\{-f''(x_{(i)}^\top\theta)\}x_{ij}^2/\lambda_{ij}\right]^{-1}$$

and $\mathbb{I}_j, \lambda_{ij}$ are the definitions in textbook. □

2.13

(a)

The likelihood function is

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

so that the log-likelihood function is given by

$$l(\theta,\sigma^2|Y_{\rm obs}) = -\frac{m}{2}\log(2\pi\sigma^2) - \sum_{i=1}^m \frac{(y_i - x_{(i)}^\top\theta)^2}{2\sigma^2}$$

Note that the MLEs of σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^m (y_i - x_{(i)}^\top\hat{\theta})^2}{m}$$

This means that if we can find $\hat{\theta}$, then $\hat{\sigma}^2$ is known. Thus, for $l(\theta,\sigma^2|Y_{\rm obs})$, we can first view σ^2 as a constant and to find the MLE of θ. We have

$$\hat{\theta} = \text{argmax } l(\theta|\sigma^2, Y_{\rm obs}) = \text{argmax } \sum_{i=1}^m -(y_i - x_{(i)}^\top\theta)^2 = \text{argmax } \sum_{i=1}^m f_i(x_{(i)}^\top\theta),$$

where $f_i(u) = -(y_i-u)^2$. Since $f_i'(u) = 2(y_i-u)$ and $-f_i''(u) = 2$, it follows that

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \frac{\sum_{i\in\mathbb{I}_j}(y_i - x_{(i)}^\top\theta^{(t)})x_{ij}}{\sum_{i\in\mathbb{I}_j} x_{ij}^2/\lambda_{ij}}$$

By 2.10, a matrix form is given by

$$\theta^{(t+1)} = \theta^{(t)} + \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\theta^{(t)})/\mathbf{Z}^\top,$$

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

(b)

(i) q is very large:
The dimension of $\mathbf{X}^\top\mathbf{X}$ is $q\times q$, so the inverse of $\mathbf{X}^\top\mathbf{X}$ is quite difficult to calculate. However, the DP algorithm avoid the calculation of an inverse matrix.
(ii) $\mathbf{X}^\top\mathbf{X}$ is almost singular:
The calculation of $\hat{\theta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ is numerically unstable or even impossible. However, the DP algorithm has the ascent property. The likeihood will increase in each iteration. □

2.14

Since $\mathbf{A}$ is positive definite, $\mathbf{A}$ is invertible. Note that

$$\frac{\partial|\mathbf{A}|}{\partial\mathbf{A}} = |\mathbf{A}|(\mathbf{A}^{-1})^\top \quad \text{and} \quad \frac{\partial \text{tr}\mathbf{A}}{\partial\mathbf{A}} = \mathbf{I}_m$$

We have

$$\begin{aligned}
\frac{\partial f(\mathbf{A})}{\partial\mathbf{A}} &= c\frac{n}{2}|\mathbf{A}|^{\frac{n}{2}-1}\frac{\partial|\mathbf{A}|}{\partial\mathbf{A}}\exp(-0.5\text{tr}\mathbf{A}) + c|\mathbf{A}|^{\frac{n}{2}}\exp(-0.5\text{tr}\mathbf{A})\times(-0.5)\frac{\partial\text{tr}\mathbf{A}}{\partial\mathbf{A}} \\
&= \frac{nc}{2}|\mathbf{A}|^{\frac{n}{2}-1}|\mathbf{A}|(\mathbf{A}^{-1})^\top\exp(-0.5\text{tr}\mathbf{A}) - \frac{c}{2}|\mathbf{A}|^{\frac{n}{2}}\exp(-0.5\text{tr}\mathbf{A})\mathbf{I_m} \\
&= \frac{1}{2}c|\mathbf{A}|^{\frac{n}{2}}\exp(-0.5\text{tr}\mathbf{A})\Big[n(\mathbf{A}^{-1})^\top - \mathbf{I}_m\Big]
\end{aligned}$$

$\mathbf{A}$ is positive definite and c is positive, then $\frac{1}{2}c|\mathbf{A}|^{\frac{n}{2}}\exp(-0.5\mathrm{tr}\mathbf{A}) > 0$ and $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = 0$ iff $n(\mathbf{A}^{-1})^\top = \mathbf{I}_m$ iff $\mathbf{A} = n\mathbf{I}_m$. However, $\mathbf{A} = n\mathbf{I}_m$ may be a saddle point. Let $\mathbf{A} = \mathbf{I}_m$, then $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} > 0$. Let $\mathbf{A} = 2n\mathbf{I}_m$, then $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} < 0$. Therefore, $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}$ take both positive and negative values, whichi implies that $\mathbf{A} = n\mathbf{I}_m$ must be a maximum or minimum.
Note that $f(n\mathbf{I}_m) = cn^{\frac{mn}{2}}\exp(-\frac{mn}{2})$ and $f(\mathbf{I}_m) = c\exp(-\frac{m}{2}) < cn^{\frac{mn}{2}}\exp(-\frac{mn}{2})$. Therefore,

$$n\mathbf{I}_m = \mathrm{argmax}_{\mathbf{A}>0} f(\mathbf{A}).$$

$\square$

2.25

(a)

Let $f(x) = -\sqrt{x}$, which is a convex function. Then

$$f(x) - f(x_0) \geq f'(x_0)(x - x_0), \ \forall x, x_0 > 0$$

That is,

$$-\sqrt{x} \geq -\sqrt{x_0} - (x - x_0)/(2\sqrt{x_0}), \ \forall x, x_0 > 0$$

(b)

Let

$$Q(\theta|\theta^{(t)}) = -\frac{\sqrt{a^2 + (\theta^{(t)})^2}}{s_1} - \frac{\theta^2 - (\theta^{(t)})^2}{2s_1\sqrt{a^2 + (\theta^{(t)})^2}} - \frac{\sqrt{b^2 + (c - \theta^{(t)})^2}}{s_2} - \frac{(c-\theta)^2 - (c - \theta^{(t)})^2}{2s_2\sqrt{b^2 + (c - \theta^{(t)})^2}}$$

Then, by (a), we have

$$Q(\theta|\theta^{(t)}) \leq -\frac{\sqrt{a^2 + \theta^2}}{s_1} - \frac{\sqrt{b^2 + (c-\theta)^2}}{s_2} = l(\theta)$$

and $Q(\theta^{(t)}|\theta^{(t)}) = -\frac{\sqrt{a^2 + (\theta^{(t)})^2}}{s_1} - \frac{\sqrt{b^2 + (c - \theta^{(t)})^2}}{s_2} = l(\theta^{(t)})$. Therefore, an MM algorithm is given by

$$\theta^{(t+1)} = \underset{\theta}{\mathrm{argmax}} \ Q(\theta|\theta^{(t)})$$

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

$$\theta^{(t+1)} = \frac{cs_1\sqrt{a^2 + (\theta^{(t)})^2}}{s_1\sqrt{a^2 + (\theta^{(t)})^2} + s_2\sqrt{b^2 + (c - \theta^{(t)})^2}}$$

(c)

Since $a = 3, b = -1, c = 2, s_1 = 1, s_2 = 1.5$, we have

$$\theta^{(t+1)} = \frac{2\sqrt{9 + (\theta^{(t)})^2}}{\sqrt{9 + (\theta^{(t)})^2} + 1.5\sqrt{1 + (2 - \theta^{(t)})^2}}$$

The initial value is $\theta^{(0)} = 0$, then

$$\theta^{(1)} = \frac{2\sqrt{9+(\theta^{(0)})^2}}{\sqrt{9+(\theta^{(0)})^2}+1.5\sqrt{1+(2-\theta^{(0)})^2}} = 0.9442719$$
$$\theta^{(2)} = \frac{2\sqrt{9+(\theta^{(1)})^2}}{\sqrt{9+(\theta^{(1)})^2}+1.5\sqrt{1+(2-\theta^{(1)})^2}} = 1.1809632$$
$$\theta^{(3)} = \frac{2\sqrt{9+(\theta^{(2)})^2}}{\sqrt{9+(\theta^{(2)})^2}+1.5\sqrt{1+(2-\theta^{(2)})^2}} = 1.2489208$$
$$\theta^{(4)} = \frac{2\sqrt{9+(\theta^{(3)})^2}}{\sqrt{9+(\theta^{(3)})^2}+1.5\sqrt{1+(2-\theta^{(3)})^2}} = 1.2679930$$
$$\theta^{(5)} = \frac{2\sqrt{9+(\theta^{(4)})^2}}{\sqrt{9+(\theta^{(4)})^2}+1.5\sqrt{1+(2-\theta^{(4)})^2}} = 1.2732721$$

R codes:

```
theta <- rep(0,6)
for (i in 2:6){
  theta[i] <- 2*sqrt(9+theta[i-1]^2) / (sqrt(9+theta[i-1]^2) + 1.5*sqrt(1+(2-theta[i-1])^2))
}
theta
```

```
## [1] 0.0000000 0.9442719 1.1809632 1.2489208 1.2679930 1.2732721
```

□

2.26

(a)

Since $f(x) = -\log x$ is a convex function, it follows that

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0), \ \forall x, x_0 > 0$$

That is,

$$-\log x \geq -\log x_0 + (x - x_0)(-x_0^{-1}), \ \forall x, x_0 > 0$$

(b)

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n}\prod_{j=1}^{n}\left(\frac{\theta_i}{\theta_i+\theta_j}\right)^{y_{ij}}$$

so that the log-likelihood function is given by

$$l(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\log\theta_i + y_{ij}\Big[-\log(\theta_i+\theta_j)\Big]$$

Let

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\log\theta_i + y_{ij}\Big[-\log(\theta_i^{(t)}+\theta_j^{(t)}) + (\theta_i+\theta_j-\theta_i^{(t)}-\theta_j^{(t)})(-(\theta_i^{(t)}+\theta_j^{(t)})^{-1})\Big]$$

Then, by the inequality above, we have

$$Q(\theta|\theta^{(t)}) \le l(\theta)$$

and

$$Q(\theta^{(t)}|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\log\theta_i^{(t)} + y_{ij}\Big[-\log(\theta_i^{(t)}+\theta_j^{(t)})\Big] = l(\theta^{(t)})$$

Thus, an MM algorithm is given by

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

Solving $\frac{\partial Q(\theta|\theta^{(t)})}{\partial\theta_k} = 0$ yields that

$$\theta_k^{(t+1)} = \frac{\sum_{i=1}^{n} y_{ki}}{\sum_{i=1}^{n}\frac{y_{ki}+y_{ik}}{\theta_k^{(t)}+\theta_i^{(t)}}}$$

□

2.27

(a)

The likelihood function is given by

$$\begin{aligned} L(\theta,\lambda) &= \prod_{i=1}^{m}\frac{\theta}{\lambda}\Big(\frac{x}{\lambda}\Big)^{\theta-1}\exp\Big\{-\Big(\frac{x}{\lambda}\Big)^{\theta}\Big\} \\ &= \theta^m\lambda^{-m\theta}\exp\Big\{-\sum_{i=1}^{m}\Big(\frac{x_i}{\lambda}\Big)^{\theta}\Big\}\prod_{i=1}^{m}x_i^{\theta-1} \end{aligned}$$

so that the log-likelihood function is

$$l(\theta,\lambda) = m\log\theta - m\theta\log\lambda - \sum_{i=1}^{m}\Big(\frac{x_i}{\lambda}\Big)^{\theta} + (\theta-1)\log x_i$$

If θ is known, then by solving

$$0 = \frac{\partial l(\theta,\lambda)}{\partial\lambda} = -\frac{m\theta}{\lambda} + \sum_{i=1}^{m}\frac{\theta x_i^{\theta}}{\lambda^{\theta+1}}$$

we have $\lambda^{\theta} = \sum_{i=1}^{m} x_i^{\theta}/m$.

(b)

Since the MLE of λ satisfies $\lambda^{\theta} = \sum_{i=1}^{m} x_i^{\theta}/m$, the log-likelihood function can be written as

$$\begin{aligned} l(\theta,\lambda) &= m\log\theta - m\log\lambda^{\theta} - \sum_{i=1}^{m}\frac{x_i^{\theta}}{\lambda^{\theta}} + (\theta-1)\log x_i \\ &= m\log\theta - m\log\Big(\sum_{i=1}^{m}x_i^{\theta}/m\Big) - \sum_{i=1}^{m}\frac{x_i^{\theta}}{(\sum_{i=1}^{m}x_i^{\theta}/m)} + (\theta-1)\log x_i \\ &= m\log\theta - m\log\Big(\sum_{i=1}^{m}x_i^{\theta}\Big) + (\theta-1)\sum_{i=1}^{m}\log(x_i) + m\log m - m \end{aligned}$$

The MLE of θ is given by

$$\hat\theta = \underset{\theta>0}{\operatorname{argmax}}\, l(\theta,\lambda) = \underset{\theta>0}{\operatorname{argmax}}\ m\log\theta - m\log\Big(\sum_{i=1}^{m}x_i^{\theta}\Big) + (\theta-1)\sum_{i=1}^{m}\log(x_i)$$

since $m \log m - m$ is a constant. To apply Newton method, we should find $\nabla l_1(\theta)$ and $I(\theta)$.

$$\nabla l_1(\theta) = \frac{m}{\theta} - m\frac{1}{\sum_{i=1}^m x_i^\theta}\sum_{i=1}^m x_i^\theta \log x_i + \sum_{i=1}^m \log(x_i) = \frac{m}{\theta} - \frac{m\sum_{i=1}^m x_i^\theta \log x_i}{\sum_{i=1}^m x_i^\theta} + \sum_{i=1}^m \log(x_i)$$

$$I(\theta) = -\nabla^2 l_1(\theta) = \frac{m}{\theta^2} + m\frac{\Big(\sum_{i=1}^m (\log x_i)^2 x_i^\theta\Big)\Big(\sum_{i=1}^m x_i^\theta\Big) - \Big(\sum_{i=1}^m x_i^\theta \log x_i\Big)^2}{\Big(\sum_{i=1}^m x_i^\theta\Big)^2}$$

Thus, the Newton method is

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} + I^{-1}(\theta^{(t)})\nabla l_1(\theta^{(t)}) \\ &= \theta^{(t)} + \frac{(b^{(t)}\theta^{(t)})^2}{(b^{(t)})^2 + \Big(a^{(t)}b^{(t)} - (c^{(t)})^2\Big)(\theta^{(t)})^2}\left(\frac{1}{\theta^{(t)}} - \frac{c^{(t)}}{b^{(t)}} + \frac{\sum_{i=1}^m \log x_i}{m}\right),\end{aligned}$$

where $a^{(t)} = \Big(\sum_{i=1}^m (\log x_i)^2 x_i^{\theta^{(t)}}\Big)$, $b^{(t)} = \Big(\sum_{i=1}^m x_i^{\theta^{(t)}}\Big)$, $c^{(t)} = \Big(\sum_{i=1}^m x_i^{\theta^{(t)}} \log x_i\Big)$. $\square$

3.1

Note that

$$\int_a^b \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx = \frac{1}{\Gamma(\alpha)}\int_{a\beta}^{b\beta} x^{\alpha-1}e^{-x}\, dx$$

If we let $f(x) = x^{\alpha-1}e^{-x}$, then $h(x) = \frac{1}{n}\log f(x) = \frac{(\alpha-1)\log x - x}{n}$, $h'(x) = \frac{1}{n}(\frac{\alpha-1}{x} - 1)$ and $h''(x) = \frac{1-\alpha}{nx^2}$. Solving $h'(x) = 0$ yields that $\tilde{x} = \alpha - 1$. Moreover, $\sigma^2 = -1/\{nh''(\tilde{x})\} = \alpha - 1$. Thus, by using the first–order Laplace approximation (3.7), we have

$$\begin{aligned}\int_{a\beta}^{b\beta} x^{\alpha-1}e^{-x}\, dx &= f(\tilde{x})\sqrt{2\pi}\sigma\left\{\Phi\Big(\frac{b\beta - \tilde{x}}{\sigma}\Big) - \Phi\Big(\frac{a\beta - \tilde{x}}{\sigma}\Big)\right\} \\ &= (\alpha-1)^{\alpha-1}e^{-(\alpha-1)}\sqrt{2\pi(\alpha-1)}\left\{\Phi\Big(\frac{b\beta - \alpha + 1}{\sqrt{\alpha-1}}\Big) - \Phi\Big(\frac{a\beta - \alpha + 1}{\sqrt{\alpha-1}}\Big)\right\}\end{aligned}$$

so that

$$\int_a^b \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}\, dx = \frac{1}{\Gamma(\alpha)}(\alpha-1)^{\alpha-1}e^{-(\alpha-1)}\sqrt{2\pi(\alpha-1)}\left\{\Phi\Big(\frac{b\beta - \alpha + 1}{\sqrt{\alpha-1}}\Big) - \Phi\Big(\frac{a\beta - \alpha + 1}{\sqrt{\alpha-1}}\Big)\right\}$$

The approximate results:

$$\begin{aligned}(a,b) = (7,9): &\quad I = \frac{1}{\Gamma(5)}4^4e^{-4}\sqrt{8\pi}\left\{\Phi\Big(\frac{9\times 0.5 - 4}{2}\Big) - \Phi\Big(\frac{7\times 0.5 - 4}{2}\Big)\right\} \approx 0.1933507 \\ (a,b) = (6,10): &\quad I = \frac{1}{\Gamma(5)}4^4e^{-4}\sqrt{8\pi}\left\{\Phi\Big(\frac{10\times 0.5 - 4}{2}\Big) - \Phi\Big(\frac{6\times 0.5 - 4}{2}\Big)\right\} \approx 0.3750458 \\ (a,b) = (2,14): &\quad I = \frac{1}{\Gamma(5)}4^4e^{-4}\sqrt{8\pi}\left\{\Phi\Big(\frac{14\times 0.5 - 4}{2}\Big) - \Phi\Big(\frac{2\times 0.5 - 4}{2}\Big)\right\} \approx 0.8485588 \\ (a,b) = (15.987,\infty): &\quad I = \frac{1}{\Gamma(5)}4^4e^{-4}\sqrt{8\pi}\left\{\Phi\Big(\frac{\infty\times 0.5 - 4}{2}\Big) - \Phi\Big(\frac{15.987\times 0.5 - 4}{2}\Big)\right\} \approx 0.02245444\end{aligned}$$

R codes:

```r
approximate <- function(a,b){
  if(b=='infty'){
    I <- (32*sqrt(8*pi)/(3*exp(4)))*(1 - pnorm(a/4 - 2))
  }else{
    I <- (32*sqrt(8*pi)/(3*exp(4)))*(pnorm(b/4 - 2) - pnorm(a/4 - 2))
  }
  return(I)
}
approximate(7,9)
```

```
## [1] 0.1933507
```

```r
approximate(6,10)
```

```
## [1] 0.3750458
```

```r
approximate(2,14)
```

```
## [1] 0.8485588
```

```r
approximate(15.987,'infty')
```

```
## [1] 0.02245444
```

Exact results:

```r
pgamma(9,5,0.5)-pgamma(7,5,0.5)
```

```
## [1] 0.1933414
```

```r
pgamma(10,5,0.5)-pgamma(6,5,0.5)
```

```
## [1] 0.37477
```

```r
pgamma(14,5,0.5)-pgamma(2,5,0.5)
```

```
## [1] 0.8233485
```

```r
1-pgamma(15.987,5,0.5)
```

```
## [1] 0.1000051
```

□

3.2

For $\int_{-\infty}^{\infty} \frac{x}{1+x^2} e^{-(x-x_0)^2/2}\,dx$, let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-x_0)^2/2}, \quad h(x) = \sqrt{2\pi}\frac{x}{1+x^2}$$

Then $f(x)$ is a density of normal distribution. If we can generate $x_1, ..., x_n \overset{iid}{\sim} f(x)$, then by using the classical Monte Carlo integration, we have

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2} e^{-(x-x_0)^2/2}\,dx = E(h(X)) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{2\pi}x_i}{1+x_i^2}$$

For $\int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$, let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx, \quad h(x) = I(x < x_0)$$

Then $f(x)$ is a density of normal distribution. If we can generate $x_1, ..., x_n \overset{iid}{\sim} f(x)$, then by using the classical Monte Carlo integration, we have

$$\int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = E(h(X)) \approx \frac{1}{n}\sum_{i=1}^{n} I(x_i < x_0)$$

$\square$

3.4

Let $f(x) = 1, 0 < x < 1$ be the density of uniform distribution on $(0, 1)$ and $h(x) = \cos(\pi x/2), 0 < x < 1$. Then the integral $\int_0^1 \cos(\pi x/2)\,dx = \int_0^1 h(x)f(x)\,dx = E(h(X))$, where $X \sim f(x)$. Moreover,

$$\begin{aligned}
\mathrm{Var}\{\cos(\pi X/2)\} &= \mathrm{Var}(h(X)) = E(h(X))^2 - (E(h(X)))^2 \\
&= \int_0^1 \cos^2(\pi x/2)\,dx - (2/\pi)^2 \\
&= 1/2 - (2/\pi)^2 \approx 0.095
\end{aligned}$$

Let $\varphi(y) = \frac{2\cos(\pi y/2)}{3(1-y^2)}, 0 < y < 1$, then

$$\int_0^1 \varphi(y)g(y)\,dy = \int_0^1 \cos(\pi y/2)\,dy = 2/\pi$$

That is, $E(\varphi(Y)) = 2/\pi$. The variance of $\varphi(Y)$ is given by

$$\mathrm{Var}(\varphi(Y)) = E\Big(\varphi(Y) - E(\varphi(Y))\Big)^2 = E\Big(\varphi(Y) - 2/\pi\Big)^2$$

If we can generate $y_1, ..., y_n \overset{iid}{\sim} f(y)$, then by using the classical Monte Carlo integration, we have

$$E\Big(\varphi(Y) - 2/\pi\Big)^2 = \int_0^1 \left(\frac{2\cos(\pi y/2)}{3(1-y^2)} - \frac{2}{\pi}\right)^2 \frac{3(1-y^2)}{2}\,dy \approx \frac{1}{n}\sum_{i=1}^{n} \left(\frac{2\cos(\pi y_i/2)}{3(1-y_i^2)} - \frac{2}{\pi}\right)^2 \frac{3(1-y_i^2)}{2}$$

By R codes, we know that $\mathrm{Var}(\varphi(Y)) \approx 0.00099$.

R codes:

```
myvar <- function(n){
  y <- runif(n,0,1)
  a <- 2*cos(pi*y/2) / (3*(1-y^2))
  b <- (a-2/pi)^2
  var <- sum(b*1.5*(1-y^2))/n
  return(var)
}

myvar(10000)
```

```
## [1] 0.0009979755
```

```
myvar(1000000)
```

```
## [1] 0.0009920382
```

□