HW#2

- **Exercise 7.2 (a) (b) Need details and R codes**

Solution (a).

Let $X_1, \cdots, X_n$ be independent and identical samples from $\Gamma(\alpha, \beta)$, that is

$$f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \qquad x > 0,\ \alpha > 0,\ \beta > 0.$$

Suppose α is known, we derive the maximum likelihood estimator (MLE) of β. The likelihood function is

$$L_n(\beta) = \prod_{i=1}^{n} f_{\alpha,\beta}(X_i) = \left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^n \left(\prod_{i=1}^{n} X_i\right)^{\alpha-1} \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} X_i\right).$$

The log-likelihood function is

$$\ell_n(\beta) = \log L_n(\beta) = -n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha-1)\sum_{i=1}^{n}\log X_i - \frac{1}{\beta}\sum_{i=1}^{n} X_i.$$

The MLE can be obtained by solving $\partial\ell_n(\beta)/\partial\beta = 0$ which is equivalent to

$$\frac{1}{\beta^2}\sum_{i=1}^{n} X_i - \frac{n\alpha}{\beta} = 0.$$

Clearly, the MLE is

$$\hat{\beta} = \frac{1}{\alpha n}\sum_{i=1}^{n} X_i.$$

The MLE attains the maximum of log-likelihood function is ensured by examining its second-order derivative

$$\left.\frac{\partial^2}{\partial\beta^2}\ell_n(\beta)\right|_{\beta=\hat{\beta}} = \frac{n\alpha}{\hat{\beta}^2} - \frac{1}{\hat{\beta}^3}\sum_{i=1}^{n} X_i = -(n\alpha)^3\left(\sum_{i=1}^{n} X_i\right)^{-2} < 0.$$

Solution (b).

Suppose both α and β are unknown, we aim to find the MLEs of α and β based on the data in Exercise 7.10 (c). The data consist $n = 14$ samples and are given in Table 1. Now, the log-likelihood function becomes

$$\ell_n(\alpha, \beta) = -n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha-1)\sum_{i=1}^{n}\log X_i - \frac{1}{\beta}\sum_{i=1}^{n} X_i.$$

According to (a), we can replace β by $\Sigma_{i=1}^{n} X_i / n\alpha$. To be specific,

$$\ell_n(\alpha, \beta(\alpha)) = -n\log\Gamma(\alpha) - n\alpha\log\left(\frac{1}{n\alpha}\sum_{i=1}^{n} X_i\right) + (\alpha-1)\sum_{i=1}^{n}\log X_i - n\alpha.$$

Therefore, this problem has been reduced to the maximization of a univariate function. The MLEs can be obtained by solving $\partial\ell_n(\alpha,\beta(\alpha))/\alpha=0$ which is equivalent to

$$\sum_{i=1}^{n}\log X_i - n\psi(\alpha) - n\log\left(\frac{1}{n\alpha}\sum_{i=1}^{n}X_i\right)=0,$$

where $\psi(\alpha)=\partial\log\Gamma(\alpha)/\partial\alpha$ is the digamma function. Since there is no explicit formula for the MLE of α, one needs to use some numerical methods. We suggest applying the Newton-Raphson algorithm. Thus, we require the second-order derivative of $\ell_n(\alpha,\beta(\alpha))$,

$$\frac{\partial^2}{\partial\alpha^2}\ell_n(\alpha,\beta(\alpha))=\frac{n}{\alpha}-n\psi'(\alpha)=0,$$

where $\psi'(\alpha)$ is the trigamma function. Now, we state the Newton-Raphson algorithm.

Algorithm 1 Newton-Raphson algorithm

Step 1. Set initial value $\alpha^{(0)}$.

Step 2. Repeat the Newton-Raphson iteration:

$$\alpha^{(k+1)}=\alpha^{(k)}-\left\{\frac{\partial^2}{\partial\alpha^2}\ell_n(\alpha,\beta(\alpha))\right\}^{-1}\left\{\frac{\partial}{\partial\alpha}\ell_n(\alpha,\beta(\alpha))\right\}\Bigg|_{\alpha=\alpha^{(k)}}.$$

- If $|\alpha^{(k+1)}-\alpha^{(k)}|<10^{-5}$, stop the algorithm and set the MLE as $\alpha^{(k+1)}$.

We apply Algorithm 1 with initial value $\alpha^{(0)}=1$ and it converges in 15 iterations. The result of estimation is $\hat{\alpha}=514.3354$. Then we obtain $\hat{\beta}=\Sigma_{i=1}^{n}X_i/n\hat{\alpha}=0.0449$. Figure 1 reveals that the MLEs attain the maximum of the log-likelihood function. R codes are given in Appendix 1.

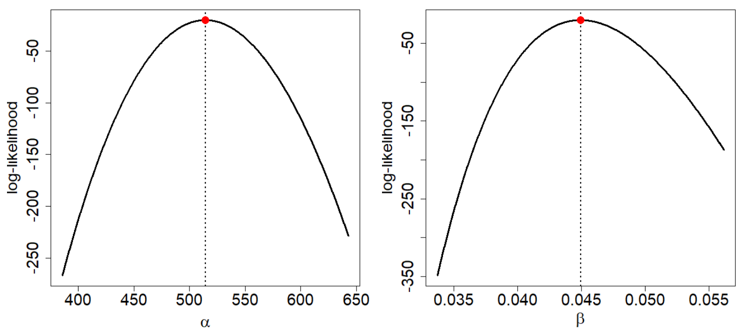

Figure 1. Log-likelihood functions under the gamma distribution based on the cuckoo's egg data. The vertical lines are drawn at $\hat{\alpha}=514.3354$, $\hat{\beta}=0.00449$.

Table1. The length (in millimeters) of cuckoo's egg.

22.0	23.9	20.9	23.8	25.0	24.0	21.7
23.8	22.8	23.1	23.1	23.5	23.0	23.0

- **Exercise 7.6 (a) Is it complete? (b) Draw the figure of the likelihood function to explain your answer (with R codes)**

Solution (a).

Let $X_1, \cdots, X_n$ be independent and identical samples from the probability density function

$$f_\theta(x) = \theta x^{-2}, \qquad 0 < \theta \le x < \infty.$$

This is known as the Pareto type I distribution with the scale and shape parameters being θ and 1, respectively. The joint density is

$$\begin{aligned} f_\theta(x_1, \cdots, x_n) &= \prod_{i=1}^{n} f_\theta(x_i) = \theta^n \prod_{i=1}^{n} x_i^{-2} \mathbf{I}(\theta \le x_i) = \theta^n \mathbf{I}(\theta \le x_{(1)}) \prod_{i=1}^{n} x_i^{-2} \\ &= g_\theta\{T(x_1, \cdots, x_n)\} h(x_1, \cdots, x_n), \end{aligned}$$

where $\mathbf{I}(\cdot)$ is the indicator function, $x_{(1)} = \min(x_1, \cdots, x_n) = T(x_1, \cdots, x_n)$,

$$g_\theta(t) = \theta^n \mathbf{I}(\theta \le t), \qquad h(x_1, \cdots, x_n) = \prod_{i=1}^{n} x_i^{-2}.$$

By the factorization theorem, we obtain $T(X_1, \cdots, X_n) = X_{(1)}$ is a sufficient statistics.

Now, we examine the completeness of $T(X_1, \cdots, X_n) = X_{(1)}$. The cumulative distribution function of T is

$$\begin{aligned} F_T(t) &= \Pr(T \le t) = \Pr(\min(X_1, \cdots, X_n) \le t) = 1 - \Pr(\min(X_1, \cdots, X_n) > t) \\ &= 1 - \Pr(X > t)^n = 1 - \theta^n / x^n. \end{aligned}$$

Its density function is

$$f_T(t) = \frac{\partial}{\partial t} F_T(t) = \frac{n\theta^n}{t^{n+1}}.$$

Suppose that there exist function ϕ such that $E_\theta\{\phi(T)\} = 0$ for all $\theta > 0$. Then, by straightforward calculations, we have for all $\theta > 0$,

$$\begin{aligned} E_\theta\{\phi^+(T)\} = E_\theta\{\phi^-(T)\} &\Rightarrow \int_\theta^\infty \phi^+(t) \frac{n\theta^n}{t^{n+1}} dt = \int_\theta^\infty \phi^-(t) \frac{n\theta^n}{t^{n+1}} dt \\ &\Rightarrow \phi^+(\theta) \frac{n}{\theta} = \phi^-(\theta) \frac{n}{\theta} \\ &\Rightarrow \phi^+(\theta) = \phi^-(\theta) \\ &\Rightarrow \phi(\theta) = 0. \end{aligned}$$

This implies $\phi(t) = 0$ for all $t \ge \theta$. Therefore, we have shown that T is completes hence $T(X_1, \cdots, X_n) = X_{(1)}$ is a completes sufficient statistics.

Solution (b).

According to (a), the likelihood is

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = \theta^n \mathbf{I}(\theta \le X_{(1)}) \prod_{i=1}^{n} X_i^{-2}.$$

One may observe that $L_n(\theta)$ is an increasing function of θ in the range $0 < \theta \le X_{(1)}$. Thus, the MLE is $\hat{\theta} = X_{(1)}$. For illustration, we apply the inverse transform method to generate $n = 100$ samples from the Pareto type I distribution with the scale and shape parameters being 2 and 1, respectively (details of data generation are provided in Appendix 2). Then, we obtain the MLE $\hat{\theta} = X_{(1)} = 2.0293$. Figure 2 plots the likelihood function and it shows that the MLE attains the maximum. R codes are available in Appendix 2.

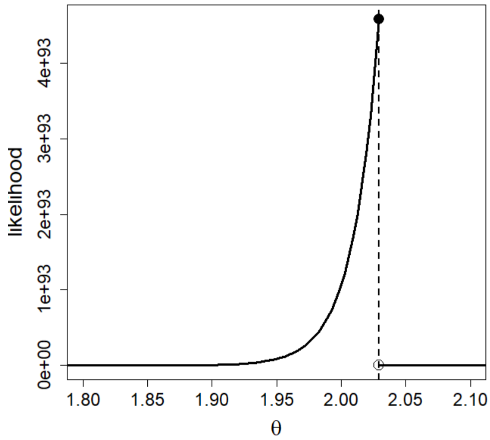

Figure 2. Likelihood function under the Pareto type I distribution based on the generated data. The vertical line is drawn at $\hat{\theta} = 2.0293$.

Solution (c).

The general Pareto type I distribution is defined as $f_\theta(x) = \theta^\alpha \alpha x^{-\alpha-1}$, $x > \theta$, where $\theta > 0$ is a scale parameter and $\alpha > 0$ is a shape parameter. Its k-th moment is derived as

$$E_\theta(X^k) = \int_\theta^\infty x^k f_\theta(x)dx = \theta^\alpha \alpha \int_\theta^\infty x^{k-\alpha-1}dx = \frac{\theta^k \alpha}{\alpha - k}, \qquad \alpha > k.$$

The above formula implies that the k-th moment does not exist if $\alpha \le k$. Under our case $\alpha = 1$, the k-th moment does not exist for all $k \ge 1$. Thus, we cannot find the moments estimator for θ.

- **Exercise 7.8**

Solution (a).

Let X be one observation from $N(0, \sigma^2)$ with probability density function is defined as

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,\ \sigma > 0.$$

Clearly, one has

$$\sigma^2 = \mathrm{var}(X) = E(X^2) - \{E(X)\}^2 = E(X^2).$$

Thus, we obtain X^2 is an unbiased estimator for σ^2.

Solution (b).

For one single observation, the likelihood function is its probability density function. Then, the log-likelihood function is

$$\ell(\sigma) = \log f_\sigma(x) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{x^2}{2\sigma^2}.$$

The MLE is obtained by solving $\partial\ell(\sigma)/\partial\sigma = 0$ which is equivalent to $x^2/\sigma^3 - 1/\sigma = 0$. Clearly, the MLE is $\hat{\sigma} = |X|$. The MLE attains the maximum of log-likelihood function is ensured by examining its second-order derivative

$$\left.\frac{\partial^2}{\partial\sigma^2}\ell_n(\sigma)\right|_{\sigma=\hat{\sigma}} = \frac{1}{\hat{\sigma}^2} - \frac{3X^2}{\hat{\sigma}^4} = -\frac{2}{X^2} < 0.$$

Solution (c).

According to (a), we have

$$E(X^2) = \sigma^2.$$

By the method of moments, if there are n observations (say $X_1, \cdots, X_n$), we estimate σ^2 by

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

However, we only have one observation. Thus, we estimate σ^2 by $\tilde{\sigma}^2 = X^2$. Then it is natural to estimate σ by $\tilde{\sigma} = |X|$.

- **Exercise 7.9**

Let $X_1, \cdots, X_n$ be independent and identical samples from $U(0, \theta)$, that is

$$f_\theta(x) = \frac{1}{\theta}, \qquad 0 \le x \le \theta .$$

Recall that the mean and variance under the uniform distribution are $E(X) = \theta/2$ and $\mathrm{var}(X) = \theta^2/12$, respectively. By the method of moments, we estimate θ by

$$\tilde{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i .$$

One can easily obtain its mean and variance as

$$E(\tilde{\theta}) = \frac{2}{n}\sum_{i=1}^{n} E(X_i) = \theta, \qquad \mathrm{var}(\tilde{\theta}) = \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) = \frac{\theta^2}{3n}.$$

Now, we consider the likelihood-based approach. The likelihood function is

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = \prod_{i=1}^{n}\theta^{-n}\mathbf{I}(0 \le X_i \le \theta) = \mathbf{I}(X_{(n)} \le \theta)\prod_{i=1}^{n}\theta^{-n},$$

where $X_{(n)} = \max(X_1, \cdots, X_n)$. One may observe that $L_n(\theta)$ is a decreasing function of θ in the range $X_{(n)} \le \theta$. Thus, the MLE is $\hat{\theta} = X_{(n)}$. To evaluate its mean and variance, one needs to find the distribution of $X_{(n)}$. The cumulative distribution function of $X_{(n)}$ is

$$\begin{aligned} F_{X_{(n)},\theta}(t) &= \Pr(X_{(n)} \le t) = \Pr(\max(X_1, \cdots, X_n) \le t) = \Pr(X_1, \cdots, X_n \le t) \\ &= \Pr(X \le t)^n = t^n/\theta^n. \end{aligned}$$

Its probability density function is

$$f_{X_{(n)},\theta}(t) = \frac{\partial}{\partial t} f_{X_{(n)},\theta}(t) = nt^{n-1}/\theta^n .$$

Then one can obtain the mean and variance of the MLE as

$$E(\hat{\theta}) = \int_0^\theta t\frac{nt^{n-1}}{\theta^n}dt = \frac{n}{n+1}\theta, \qquad E(\hat{\theta}^2) = \int_0^\theta t^2\frac{nt^{n-1}}{\theta^n}dt = \frac{n}{n+2}\theta^2,$$

$$\mathrm{var}(\hat{\theta}) = E(\hat{\theta}^2) - \{E(\hat{\theta})\}^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

According to the results we obtained, $\tilde{\theta}$ is an unbiased estimator and $\hat{\theta}$ is a biased estimator. Note that the bias of $\hat{\theta}$ is small since $n/(n+1) \to 1$ as $n \to \infty$, however, it may be large if n is small. On the other hand, $\mathrm{var}(\hat{\theta}) < \mathrm{var}(\tilde{\theta})$ for all n and θ. Therefore, the MLE $\hat{\theta}$ is preferred if the sample size is large

- **Detailed answers to Q3 below**

Q3 [+2]

Let $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where μ is restricted to $\mu \le a$ or $\mu \ge b$ for some numbers $a < b$. Assume that σ^2 is known. Hence, the parameter space is $\Theta = (-\infty, a] \cup [b, \infty)$. Obtain the MLE $\hat{\mu}$ (with some figures to explain it).

Solution Q3.

Given σ^2 is known, the likelihood is written as

$$L_n(\mu) = \prod_{i=1}^{n} f_\mu(X_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(X_i - \mu)^2}{2\sigma^2} \right\}.$$

Then the log-likelihood function is

$$\ell_n(\mu) = \log L_n(\mu) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

The MLE is obtain by solving $\partial \ell_n(\mu)/\partial\mu = 0$ and it is the sample mean $\hat{\mu} = \Sigma_{i=1}^{n} X_i / n \equiv \bar{X}_n$. However, the parameter restriction is imposed, it is possible that $\bar{X}_n \notin (-\infty, a] \cup [b, \infty)$. Therefore, additional adjustments are required. The restricted MLE should be

$$\hat{\mu}_{\text{res}} = \begin{cases} \bar{X}_n & \text{if } \bar{X}_n \le a \text{ or } \bar{X}_n \ge b, \\ a & \text{if } a < \bar{X}_n < (a+b)/2, \\ b & \text{if } (a+b)/2 < \bar{X}_n < b, \\ a, b & \text{if } \bar{X}_n = (a+b)/2. \end{cases}$$

Note that if $\bar{X}_n = (a+b)/2$ then there are two MLEs $\hat{\mu}_{\text{res}} = a$ and $\hat{\mu}_{\text{res}} = b$. However, one can simply ignore this case since it happens with 0 probability. Figure 3 illustrates the other three different cases.

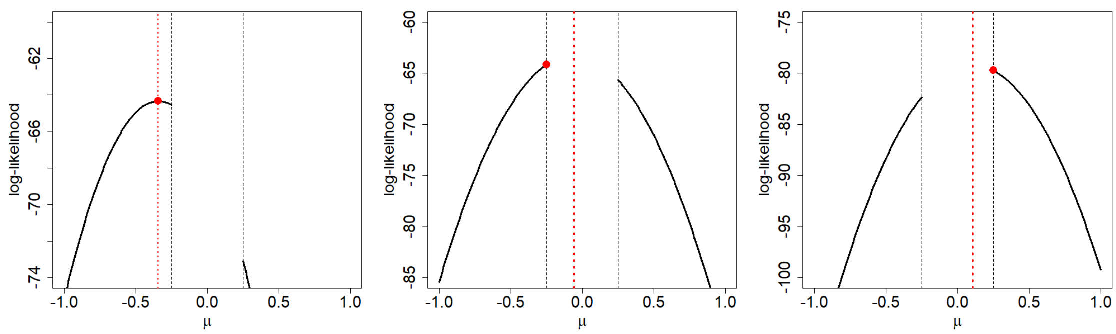

Figure 3. Three different cases of the restricted MLE with $a = -0.25$ and $b = 0.25$. The points denote the restricted MLE $\hat{\mu}_{\text{res}}$ and the dotted vertical lines denote the location of $\bar{X}_n$.

- **Detailed answers to Q4 below**

Q4 [+3]

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Gamma(\alpha, \beta)$ as in HW#1. Let $\psi(\alpha) = \frac{d}{d\alpha}\log\Gamma(\alpha)$ be the digamma function and $\psi'(\alpha)$ be the trigamma function.

(1) Write down the score functions using the sufficient statistics (T_1, T_2).

Solution (1).

According to Exercise 7.2 (a), the joint density for i.i.d. samples from $\Gamma(\alpha, \beta)$ is

$$\begin{aligned} f_{\alpha,\beta}(x_1, \cdots, x_n) &= \prod_{i=1}^{n} f_{\alpha,\beta}(x_i) = \left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^{n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i\right) \\ &= g_{\alpha,\beta}\{T_1(x_1, \cdots, x_n), T_2(x_1, \cdots, x_n)\} h(x_1, \cdots, x_n), \end{aligned}$$

where

$$T_1(x_1, \cdots, x_n) = \prod_{i=1}^{n} x_i, \qquad T_2(x_1, \cdots, x_n) = \sum_{i=1}^{n} x_i,$$

$$g_{\alpha,\beta}(t_1, t_2) = \left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^{n} t_1^{\alpha} \exp\left(-\frac{t_2}{\beta}\right), \qquad h(x_1, \cdots, x_n) = \left(\prod_{i=1}^{n} x_i\right)^{-1}.$$

By the factorization theorem, $(T_1, T_2) = (\Pi_{i=1}^{n} X_i, \Sigma_{i=1}^{n} X_i)$ is a sufficient statistics.

Again, by Exercise 7.2 (b), the log-likelihood function is

$$\ell_n(\alpha, \beta) = -n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha - 1)\sum_{i=1}^{n}\log X_i - \frac{1}{\beta}\sum_{i=1}^{n} X_i.$$

Then the score functions are

$$S_1(\alpha, \beta) = \frac{\partial}{\partial\alpha}\ell_n(\alpha, \beta) = \sum_{i=1}^{n}\log X_i - n\log\beta - n\psi(\alpha),$$

$$S_2(\alpha, \beta) = \frac{\partial}{\partial\beta}\ell_n(\alpha, \beta) = \frac{1}{\beta^2}\sum_{i=1}^{n} X_i - \frac{n\alpha}{\beta}.$$

Now we can express the score functions by using sufficient statistics $(T_1, T_2) = (\Pi_{i=1}^{n} X_i, \Sigma_{i=1}^{n} X_i)$.

$S_1(\alpha, \beta) = \log T_1 - n\log\beta - n\psi(\alpha)$,

$S_2(\alpha, \beta) = \frac{T_2}{\beta^2} - \frac{n\alpha}{\beta}$,

where $T_1 = \Pi_{i=1}^{n} X_i$ and $T_2 = \Sigma_{i=1}^{n} X_i$.

(2) Write down the Hessian matrix $H(\alpha, \beta)$.

Solution (2).

To obtain the Hessian matrix, one requires the second-order derivatives of $\ell_n(\alpha, \beta)$. They are

$$\frac{\partial^2}{\partial \alpha^2}\ell_n(\alpha, \beta) = -n\psi'(\alpha), \qquad \frac{\partial^2}{\partial \alpha \partial \beta}\ell_n(\alpha, \beta) = -\frac{n}{\beta}, \qquad \frac{\partial^2}{\partial \beta^2}\ell_n(\alpha, \beta) = \frac{n\alpha}{\beta^2} - \frac{2}{\beta^3}\sum_{i=1}^{n} X_i .$$

Thus, the Hessian matrix is

$$H(\alpha, \beta) = \begin{bmatrix} \frac{\partial^2}{\partial \alpha^2}\ell_n(\alpha, \beta) & \frac{\partial^2}{\partial \alpha \partial \beta}\ell_n(\alpha, \beta) \\ \frac{\partial^2}{\partial \alpha \partial \beta}\ell_n(\alpha, \beta) & \frac{\partial^2}{\partial \beta^2}\ell_n(\alpha, \beta) \end{bmatrix} = \begin{bmatrix} -n\psi'(\alpha) & -\frac{n}{\beta} \\ -\frac{n}{\beta} & \frac{n\alpha}{\beta^2} - \frac{2}{\beta^3}\sum_{i=1}^{n} X_i \end{bmatrix}.$$

(3) Let $(\hat{\alpha}, \hat{\beta})$ be the solution to $S_1(\alpha, \beta) = S_2(\alpha, \beta) = 0$. Write down $H(\hat{\alpha}, \hat{\beta})$ in terms of $(\hat{\alpha}, \hat{\beta})$.

Solution (3).

Since $(\hat{\alpha}, \hat{\beta})$ is the solution to $S_1(\alpha, \beta) = S_2(\alpha, \beta) = 0$, we have

$$S_2(\hat{\alpha}, \hat{\beta}) = \frac{1}{\hat{\beta}^2}\sum_{i=1}^{n} X_i - \frac{n\hat{\alpha}}{\hat{\beta}} = 0 .$$

This implies

$$n\hat{\alpha} = \frac{1}{\hat{\beta}}\sum_{i=1}^{n} X_i .$$

Then one has the following simplification

$$\left.\frac{\partial^2}{\partial \beta^2}\ell_n(\alpha, \beta)\right|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} = \frac{n\hat{\alpha}}{\hat{\beta}^2} - \frac{2}{\hat{\beta}^3}\sum_{i=1}^{n} X_i = -\frac{n\hat{\alpha}}{\hat{\beta}^2} .$$

Therefore, we obtain

$$H(\hat{\alpha}, \hat{\beta}) = -\begin{bmatrix} n\psi'(\hat{\alpha}) & \frac{n}{\hat{\beta}} \\ \frac{n}{\hat{\beta}} & \frac{n\hat{\alpha}}{\hat{\beta}^2} \end{bmatrix}.$$

Appendix 1. R codes for Exercise 7.2

R codes

```
########## Exercise 7.2 ##########

onedim.score_func = function(Alpha) {

   n = length(x)

   -n*digamma(Alpha)-n*log(sum(x)/(Alpha*n))+sum(log(x))

}

onedim.Hessian_func = function(Alpha) {

   n = length(x)

   -n*trigamma(Alpha)+n/Alpha

}

##### one-dimensional Newton-Raphson algorithm

x = c(22,23.9,20.9,23.8,25,24,21.7,23.8,22.8,23.1,23.1,23.5,23,23)

count      = 0
epsilon = 1e-5
par_old = 1

repeat {

   par_new = par_old-onedim.Hessian_func(par_old)^-1*onedim.score_func(par_old)
   count = count+1

   if (abs(par_new-par_old) < epsilon) {break}
   par_old = par_new

}
par_new
count

Alpha = par_new; Alpha
Beta    = sum(x)/(par_new*length(x)); Beta
```

Appendix 2. Data generation and R codes in Exercise 7.6

Suppose X follows the Pareto type I distribution, its cumulative distribution function is

$$F_X(x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha}, \qquad 0 < \theta \le x < \infty, \ \alpha > 0.$$

By the probability integral transform, we have

$$U = F_X(X) = 1 - \left(\frac{\theta}{X}\right)^{\alpha},$$

where $U \sim U(0,1)$. This implies

$$X = \frac{\theta}{(1-U)^{1/\alpha}}.$$

Thus, one can easily generate random samples $X_1, \cdots, X_n$ from the Pareto type I distribution by transforming random samples $U_i \sim U(0,1)$, $i = 1, 2, \cdots, n$.

R codes

```
########## exercise 7.6 ##########

L_func = function(theta) {theta^n*prod(x)}

n = 100
theta = 2

set.seed(10)
u = runif(n)
x = theta/(1-u)
min(x)

theta_v = seq(1.7,min(x),length.out = 100)
plot(theta_v,sapply(theta_v,L_func),type = "l",xlim = c(1.8,2.1),lwd = 3,
        ylab = "likelihood",xlab = expression(theta),cex.axis = 1.5,cex.lab = 1.8)
lines(seq(min(x),2.2,length.out = 100),rep(0,100),lwd = 3)
points(min(x),L_func(min(x)),pch = 16,cex = 2)
points(min(x),0,cex = 2)
abline(v = min(x),lty = 2,lwd = 2)
```