Statistical Methods for Data Science

HW 5 Solution

Exercise 9.3

(a) (Maximum likelihood) The likelihood function is

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & \text{if } a \le x_1, \ldots, x_n \le b, \\ \\ 0, & \text{otherwise.} \end{cases}$$

It can also be written as

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n}, & \text{if } a \le \min\{x_1, \ldots, x_n\} \text{ and } \max\{x_1, \ldots, x_n\} \le b, \\ \\ 0, & \text{otherwise.} \end{cases}$$

For maximization, we only need to focus on the region where the likelihood function is positive. In this region, the function is monotonically increasing in a and monotonically decreasing in b. Therefore, it is maximized when a is at its largest possible value and b is at its smallest possible value. This implies that the maximum likelihood estimators are:

$$\hat{a} = \min\{x_1, \ldots, x_n\},\ \hat{b} = \max\{x_1, \ldots, x_n\}.$$

(b) (Maximum likelihood). The likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}.$$

Taking its log gives the log-likelihood function as

$$\log\{L(\lambda)\} = \sum_{i=1}^{n} (\log(\lambda) - \lambda x_i).$$

Next, we obtain the likelihood equation as

$$0 = \frac{\partial}{\partial \lambda} \log\{L(\lambda)\} = \sum_{i=1}^{n} (\frac{1}{\lambda} - x_i) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

Solving this equation with respect to λ gives the MLE as

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}.$$

(e) (Maximum likelihood) In this case, the log-likelihood function $\log\{L(\mu, \sigma)\}$ is given by equation (12.1) in the textbook. To maximize this function, our strategy is to first fix σ and maximize the function with respect to μ alone. Then, we will plug in this maximizing value in the log-likelihood function, making it a function of σ alone. Thereafter, we will maximize this function with respect to σ.

For a fixed σ, $\log\{L(\mu, \sigma)\}$ is maximized with respect to μ when $\mu = \hat{\mu} = \bar{x}$, regardless of the value of σ. Next, substituting $\mu = \bar{x}$ into $\log\{L(\mu, \sigma)\}$, we get $\log\{L(\bar{x}, \sigma)\}$, which a function of σ alone. Using the differentiation technique, we can see that this function is maximized with respect to σ at $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$. Therefore, it follows that

$$\hat{\mu} = \bar{X},\ \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

the MLEs of μ and σ.

Exercise 9.7

(a) $\bar{X} \pm z_{0.05}\frac{\sigma}{\sqrt{n}} = 37.7 \pm (1.645)\frac{9.2}{\sqrt{100}} = 37.7 \pm 1.5$ or $[36.2, 39.2]$.

Exercise 9.8

(a) $\bar{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 42 \pm (1.96)\frac{5}{\sqrt{64}} = 42 \pm 1.225$ or $[40.775, 43.225]$.

Exercise 9.9

(a) The standard deviation is unknown. Therefore, the confidence interval is

$$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

where $\alpha = 1 - 0.90 = 0.10, n = 3, t_{n-1,\alpha/2} = t_{2,0.05} = 2.920$. Also,

$$\bar{X} = \frac{30+50+70}{3} = 50, \quad S = \sqrt{\frac{(30-50)^2 + (50-50)^2 + (70-50)^2}{3-1}} = \sqrt{\frac{800}{2}} = 20.$$

We now have the interval as

$$50 \pm (2.920)\frac{20}{\sqrt{3}} = 50 \pm 33.7 \text{ or } [16.3, 83.7].$$

Exercise 9.10

(a) We have: $\hat{p} = \frac{24}{200} = 0.12$. For $\alpha = 1 - 0.96 = 0.04$, $z_{\alpha/2} = z_{0.02} = 2.054$. The confidence interval is:

$$\hat{p} \pm z_{0.02}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.12 \pm (2.054)\sqrt{\frac{0.12(1-0.12)}{200}} = 0.12 \pm 0.047 \text{ or } [0.073, 0.167].$$

Exercise 9.12

(a) The standard deviation is known, therefore we construct a z-interval as:

$$\bar{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 0.62 \pm (1.96)\frac{0.2}{\sqrt{52}} = 0.62 \pm 0.054 \text{ or } [0.566, 0.674].$$

Exercise 9.17 For candidate A's estimate, $\hat{p}_1 = 0.45$, the margin of error is:

$$z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} = 1.96\sqrt{\frac{0.45 \times 0.55}{900}} = 0.0325 \text{ or } 3.25\%.$$

For candidate B's estimate, $\hat{p}_2 = 0.35$, the margin of error is:

$$z_{0.025}\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{m}} = 1.96\sqrt{\frac{0.35 \times 0.65}{900}} = 0.0312 \text{ or } 3.12\%.$$

For candidate A's lead, $\hat{p}_1 - \hat{p}_2 = 0.10$, the margin of error is:

$$z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} = 1.96\sqrt{\frac{0.45 \times 0.55}{900} + \frac{0.35 \times 0.65}{900}} = 0.045 \text{ or } 4.50\%.$$

Exercise 9.18

(a) The problems asks to assume that the two variances are equal. We will additionally assume that the data are normally distributed. Therefore, the appropriate confidence interval for $\mu_1 - \mu_2$ to use in this case is:

$$\bar{X} - \bar{Y} \pm t_{n+m-2,\alpha/2}\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}.$$

From the data, we have the following:

$$n = 14, m = 20, \bar{X} = 50, \bar{Y} = 40.2, S_X^2 = 58, S_Y^2 = 63.33,$$

and also

$$S_P^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = 61.1625.$$

Therefore, the 95% confidence interval for $\mu_1 - \mu_2$ is:

$$50 - 40.2 \pm (2.037)\sqrt{61.1625(\frac{1}{14} + \frac{1}{20})} = 9.8 \pm 5.55 \quad \text{or} \quad [4.25, 15.35].$$

Exercise 9.21 Given that $[a, b]$ is a $100(1-\alpha)\%$ confidence interval for σ^2, we have $P\{a \le \sigma^2 \le b\} = 1 - \alpha$.
Since $a \le \sigma^2 \le b \Leftrightarrow \sqrt{a} \le \sigma \le \sqrt{b}$,
$\Rightarrow P\{\sqrt{a} \le \sigma \le \sqrt{b}\} = 1 - \alpha$
$\Rightarrow [\sqrt{a}, \sqrt{b}]$ is a $(1-\alpha)100\%$ confidence interval for σ.

Additional 1. Using maximum likelihood method:

$$\begin{aligned} L(\theta) &= f_\theta(2)f_\theta(0)f_\theta(0)f_\theta(1)f_\theta(0)f_\theta(1)f_\theta(1)f_\theta(0)f_\theta(2)f_\theta(1) \\ &= 0.5^4 \times \theta^4 \times (0.5-\theta)^2. \end{aligned}$$

Therefore, the log-likelihood function is $\log\{L(\theta)\} = 4\log(0.5) + 4\log(\theta) + 2\log(0.5-\theta)$. Taking its derivative with respect to θ and equating it to 0, we get:

$$\frac{\partial}{\partial\theta}\log\{L(\theta)\} = 0 + \frac{4}{\theta} + \frac{2}{0.5-\theta}(-1) = 0.$$

Solving this equation, gives the MLE of θ as $\hat{\theta} = \frac{1}{3}$.

Additional 2. Using maximum likelihood method:

$$\begin{aligned} L(p) &= f_p(2)f_p(0)f_p(0)f_p(0)f_p(3) \\ &= (1-6p)^3 \times (3p)^0 \times (2p)^1 \times p^1 \\ &= (1-6p)^3 \times 2 \times p^2 \end{aligned}$$

Therefore, the log-likelihood function is $\log\{L(p)\} = 3log(1-6p) + log(2) + 2log(p)$. Taking its derivative with respect to p and equating to 0, we get:

$$\frac{\partial}{\partial p}\log\{L(p)\} = \frac{3}{1-6p}(-6) + 0 + \frac{2}{p} = 0.$$

Solving this equation, gives the MLE of p as $\hat{p} = \frac{1}{15} = 0.067$.

Additional 3 Refer to MLE in Exercise 9.3 (a). Given that lifetimes follow $U[3, b]$ distribution, we have

$$f(x_i) = \begin{cases} \frac{1}{b-3}, & \text{if } 3 \le x_i \le b, \\ 0, & \text{otherwise,} \end{cases}$$

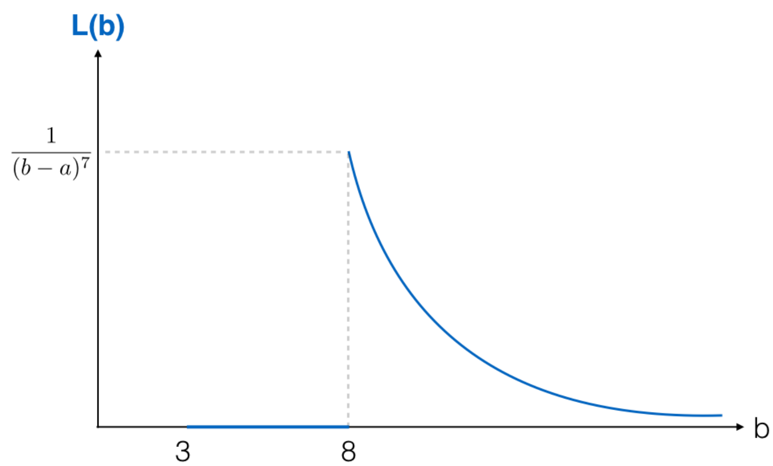
for $i = 1, \ldots, n = 7$. Therefore, the likelihood function of b is,

$$L(b) = \begin{cases} \frac{1}{(b-3)^7}, & \text{if } \max\{x_1, \ldots, x_7\} \leq b, \\ 0, & \text{otherwise,} \end{cases}$$

where we have that all the observations are greater than or equal to 3. Since $\max\{x_1, \ldots, x_7\} = 8$, the likelihood function can be written as

$$L(b) = \begin{cases} \frac{1}{(b-3)^7}, & \text{if } b \geq 8, \\ 0, & \text{otherwise.} \end{cases}$$

It is a decreasing function of b on $[8, \infty]$ and is 0 otherwise. Either we can argue directly or we can see from the plot of $L(b)$ below that the likelihood function is maximized at $b = 8$. Therefore, the MLE of b is $\hat{b} = 8$.




```
##################################################################
# R code for HW 5 Exercises from Chapter 9
##################################################################

########
#9.7(a)#
########
#> 37.7 + c(-1,1) * qnorm(1-(1-0.9)/2) * 9.2/sqrt(100)
#[1] 36.18673 39.21327
#>

########
#9.8(a)#
########
#> 42 + c(-1,1) * qnorm(1-(1-0.95)/2) * 5/sqrt(64)
```

```
#[1] 40.77502 43.22498
#>

########
#9.9(a)#
########
x <- c(30, 50, 70)
n <- length(x)
df <- n - 1
xmean <- mean(x)
xsd <- sd(x)

#> xmean + c(-1,1) * qt(1-(1-0.9)/2, df) * xsd/sqrt(n)
#[1] 16.28291 83.71709
#>

#########
#9.10(a)#
#########
p.hat <- 24/200
n <- 200

#> p.hat + c(-1,1) * qnorm(1-(1-0.96)/2) * sqrt(p.hat*(1-p.hat)/n)
#[1] 0.07280844 0.16719156
#>

#########
#9.12(a)#
#########
#> 0.62 + c(-1,1) * qnorm(1-(1-0.95)/2) * 0.2/sqrt(52)
#[1] 0.5656404 0.6743596
#>

######
#9.17#
######
p1.hat <- 0.45
n <- 900

#> qnorm(1-(1-0.95)/2) * sqrt(p1.hat*(1-p1.hat)/n)
#[1] 0.03250233
#>

p2.hat <- 0.35
m <- 900

#> qnorm(1-(1-0.95)/2) * sqrt(p2.hat*(1-p2.hat)/m)
#[1] 0.03116144
#>
```

```
#> qnorm(1-(1-0.95)/2) * sqrt(p1.hat*(1-p1.hat)/n + p2.hat*(1-p2.hat)/m)
#[1] 0.04502707
#>

#########
#9.18(a)#
#########
x <- c(56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58)
y <- c(53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36,
39, 45)
n <- length(x)
m <- length(y)
xmean <- mean(x)
ymean <- mean(y)
xvar <- var(x)
yvar <- var(y)
pooled.var <- ((n - 1) * xvar + (m - 1) * yvar)/(n + m - 2)

# > xmean - ymean + c(-1,1) * qt(1-(1-0.95)/2, df = n + m - 2) *
# sqrt(pooled.var*((1/n) + (1/m)))
# [1]  4.248889 15.351111
# >

########################################################################
```