Computational Statistics

Assignment 5

Hanbin Liu 11912410

5.1

(a)

Let $x_1, ...x_n$ denote these observations. Since $x_1, ...x_n \sim N(\mu, \sigma^2)$, it then follows that

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = 43.729, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = 79.99.$$

(b)

The MLE of CV is given by $\widehat{\text{CV}} = \hat{\sigma}/\hat{\mu} = 0.20453$. We first generated 20000 bootstrap samples from $N(\hat{\mu}, \hat{\sigma}^2)$, computed 20000 bootstrap replications $\{\widehat{\text{CV}}^*(g)\}_{g=1}^G$ and obtained

$$\bar{\text{CV}}^* = 0.1985, \quad \widehat{\text{Se}}^*(\widehat{\text{CV}}) = 0.03055, \quad [\widehat{\text{CV}}_L^*, \widehat{\text{CV}}_U^*] = [0.138655, 0.258411]$$

R code

```r
## function for calculating the mean, std, two 95% CIs based on column
mean.std.CI <- function(thsample) {
  G <- dim(thsample)[1]
  thmean <- apply(thsample, 2, mean)
  thstd <- sqrt(apply(thsample, 2, var))
  thl <- thmean - 1.96 * thstd
  thu <- thmean + 1.96 * thstd
  thsort <- apply(thsample, 2, sort)
  indexx <- floor(c(0.025 * G, 0.975 * G))
  thL <- (thsort[indexx[1],] + thsort[indexx[1] + 1,]) / 2
  thU <- (thsort[indexx[2],] + thsort[indexx[2] + 1,]) / 2
  results <- c(thmean, thstd, thl, thu, thL, thU)
}


set.seed(1234)
assign5.1.b <- function(G){
  x <- c(32.0, 46.4, 48.1, 27.7, 35.5, 52.6, 66.0, 41.3, 49.9, 36.1, 50.0, 44.7,
         48.2, 36.9, 40.8, 35.1, 63.3, 42.5, 52.4, 40.9, 38.6, 43.2, 41.7, 35.6)
  n <- length(x)
```

```
  mu <- mean(x)
  sigma <- sqrt((n - 1) * var(x) / n)
  CV <- sigma / mu
  CV.sample <- matrix(0, G, 1)

  for (g in 1:G) {
    xstar <- rnorm(n, mean = mu, sd = sigma)
    mustar <- mean(xstar)
    sigmastar <- sqrt((n - 1) * var(xstar) / n)
    CVstar <- sigmastar / mustar
    CV.sample[g, 1] <- CVstar
  }

  M <- mean.std.CI(CV.sample)
  result <- c(M,CV)
  return(result)
}

assign5.1.b(20000)
```

```
## [1] 0.19853286 0.03054995 0.13865497 0.25841076 0.14176579 0.26057329 0.20452563
```

(c)

(c-1)

The parametric bootstrap method. The MLE of the population median θ is $\hat{\theta} = (x_{(12)} + x_{(13)})/2 = 42.1$. We first generated $G = 20000$ bootstrap samples from $N(\hat{\mu}, \hat{\sigma}^2)$, computed 20000 bootstrap replications $\{\hat{\theta}^*(g)\}_{g=1}^{G}$, and obtained

$$\bar{\theta}^* = 43.73, \quad \widehat{\mathrm{Se}}^*(\hat{\theta}) = 2.2319, \quad [\hat{\theta}_L^*, \hat{\theta}_U^*] = [39.3839, 48.1144]$$

R code

```
assign5.1.c1 <- function(G){
  x <- c(32.0, 46.4, 48.1, 27.7, 35.5, 52.6, 66.0, 41.3, 49.9, 36.1, 50.0, 44.7,
         48.2, 36.9, 40.8, 35.1, 63.3, 42.5, 52.4, 40.9, 38.6, 43.2, 41.7, 35.6)
  n <- length(x)
  mu <- mean(x)
  sigma <- sqrt((n - 1) * var(x) / n)
  thMLE <- median(x)
  th.star.sample <- matrix(0, G, 1)
  for (g in 1:G) {
    xstar <- rnorm(n, mean = mu, sd = sigma)
    thstar <- median(xstar)
    th.star.sample[g, 1] <- thstar
  }

  M <- mean.std.CI(th.star.sample)
  result <- c(M, thMLE)
  return(result)
```

```
}

assign5.1.c1(20000)
```

```
## [1] 43.732494  2.231912 39.357947 48.107042 39.383926 48.114363 42.100000
```

(c-2)

The non-parametric bootstrap method. We first generated $G = 20000$ bootstrap samples from the empirical distribution based on $x_1, ..., x_n$, computed 20000 bootstrap replications $\{\hat{\theta}^*(g)\}_{g=1}^G$, and obtained

$$\bar{\theta}^* = 42.57, \quad \widehat{\mathrm{Se}}^*(\hat{\theta}) = 2.1064, \quad [\hat{\theta}_L^*, \hat{\theta}_U^*] = [38.6, 48.1]$$

R code

```
assign5.1.c2 <- function(G){
  x <- c(32.0, 46.4, 48.1, 27.7, 35.5, 52.6, 66.0, 41.3, 49.9, 36.1, 50.0, 44.7,
         48.2, 36.9, 40.8, 35.1, 63.3, 42.5, 52.4, 40.9, 38.6, 43.2, 41.7, 35.6)
  n <- length(x)
  p <- rep(1 / n, n)
  th.star.sample <- matrix(0, G, 1)
  for (g in 1:G) {
    xstar <- sample(x, n, prob = p, replace = T)
    thstar <- median(xstar)
    th.star.sample[g, 1] <- thstar
  }

  M <- mean.std.CI(th.star.sample)
  return(M)
}

assign5.1.c2(20000)
```

```
## [1] 42.572893  2.106401 38.444346 46.701439 38.600000 48.100000
```

5.2

(a)

The observed likelihood function for (ϕ, λ) is

$$L(\phi, \lambda | Y_{\text{obs}}) = [\phi + (1-\phi)e^{-\lambda}]^m \times (1-\phi)^{n-m} \prod_{y_i \notin 0} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

(b)

By exsercise 1.12, we can introduce a latent r.v. Z so that the conditional predictive distribution is

$$f(Z|Y_{\text{obs}}, \phi, \lambda) = \text{Binomial}\Big(Z|m, \frac{\phi}{\phi + (1-\phi)e^{-\lambda}}\Big)$$

and the complete-data likelihood for (ϕ, λ) is given by

$$L(\phi, \lambda) \propto \phi^z[(1-\phi)e^{-\lambda}]^{m-z} \times (1-\phi)^{n-m} \prod_{y_i \notin \mathbb{0}} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$
$$\propto \phi^z(1-\phi)^{n-z}e^{-(n-z)\lambda}\lambda^{\sum_{y_i \notin \mathbb{0}} y_i}$$

Then, the complete-data MLEs are given by

$$\hat{\phi} = \frac{z}{n}, \quad \hat{\lambda} = \frac{\sum_{y_i \notin \mathbb{0}} y_i}{n-z}$$

and the conditional expectation is

$$E(Z|Y_{\text{obs}}, \phi, \lambda) = \frac{m\phi}{\phi + (1-\phi)e^{-\lambda}}$$

Therefore, the EM iteration is

$$\phi^{(t+1)} = \frac{m\phi^{(t)}}{n\phi^{(t)} + n(1-\phi^{(t)})e^{-\lambda^{(t)}}}$$
$$\lambda^{(t+1)} = \frac{\sum_{y_i \notin \mathbb{0}} y_i}{n - \frac{m\phi^{(t)}}{\phi^{(t)}+(1-\phi^{(t)})e^{-\lambda^{(t)}}}}$$

(c)

The algorithm for generating a sample from $Y \sim \text{ZIP}(\phi, \lambda)$ is
Step 1: Draw $U \sim U(0, 1)$ and independently draw $X \sim \text{Poisson}(\lambda)$
Step 2: If $U \leq \phi$, then set $Y = 0$. Otherwise, set $Y = X$.

Bootstrap method:
Step 1: Calculate the MLEs $\hat{\phi}$ and $\hat{\lambda}$ by EM algorithm.
Step 2: Generate a bootstrap sample $\mathbf{y}^* = (y_1^*, ..., y_n^*) \overset{iid}{\sim} \text{ZIP}(\hat{\phi}, \hat{\lambda})$ and compute the corresponding bootstrap replication $\hat{\phi}^*$ and $\hat{\lambda}^*$.
Step 3: Independently repeat Step 2 G times. Obtain G bootstrap replications $\{\hat{\phi}^*(g)\}_{g=1}^G$ and $\{\hat{\lambda}^*(g)\}_{g=1}^G$
Step 4: Compute $\widehat{\text{Se}}(\hat{\phi})$ and $\widehat{\text{Se}}(\hat{\phi})$.

$$\widehat{\text{Se}}(\hat{\phi}) = \sqrt{\frac{1}{G-1}\sum_{g=1}^{G}[\hat{\phi}^*(g) - \bar{\phi}^*]^2}, \quad \bar{\phi}^* = [\hat{\phi}^*(1), +... + \hat{\phi}^*(G)]/G$$

and

$$\widehat{\text{Se}}(\hat{\lambda}) = \sqrt{\frac{1}{G-1}\sum_{g=1}^{G}[\hat{\lambda}^*(g) - \bar{\lambda}^*]^2}, \quad \bar{\lambda}^* = [\hat{\lambda}^*(1), +... + \hat{\lambda}^*(G)]/G$$

Step 5: If $\{\hat{\phi}^*(g)\}_{g=1}^G$ and $\{\hat{\lambda}^*(g)\}_{g=1}^G$ are approximately normallly distributed, the $100(1-\alpha)\%$ bootstrap CI for ϕ and λ are given by

$$[\hat{\phi}_l^*, \hat{\phi}_r^*] = [\bar{\phi}^* - z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\phi}), \bar{\phi}^* + z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\phi})]$$
$$[\hat{\lambda}_l^*, \hat{\lambda}_r^*] = [\bar{\lambda}^* - z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\lambda}), \bar{\lambda}^* + z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\lambda})]$$

Step 6: If the bootstrap CI is beyond the unit interval $[0, 1]$ or the bootstrap replications $\{\hat{\phi}^*(g)\}_{g=1}^G$, $\{\hat{\lambda}^*(g)\}_{g=1}^G$ are non-normally distributed , the $100(1-\alpha)\%$ bootstrap CI for ϕ and λ are given by

$$[\hat{\phi}_L^*, \hat{\phi}_U^*], \quad [\hat{\lambda}_L^*, \hat{\lambda}_U^*],$$

where $\hat{\phi}_L^*$ and $\hat{\phi}_U^*$ are the $(\alpha/2)$G-th and the $(1-\alpha/2)$G-th order statistics of $\{\hat{\phi}^*(g)\}_{g=1}^G$, $\hat{\lambda}_L^*$ and $\hat{\lambda}_U^*$ are the $(\alpha/2)$G-th and the $(1-\alpha/2)$G-th order statistics of $\{\hat{\lambda}^*(g)\}_{g=1}^G$

5.3

(a)

The observed likelihood function for π is

$$L(\pi|Y_{\text{obs}}) = \prod_{i=1}^{n} \left[\frac{1}{1-(1-\pi)^m} \binom{m}{x_i} \pi^{x_i} (1-\pi)^{m-x_i} \right]$$
$$\propto \pi^{n\bar{x}} (1-\pi)^{n(m-\bar{x})} \left[\frac{1}{1-(1-\pi)^m} \right]^n$$

so that the log-likelihood function is given by

$$l(\pi|Y_{\text{obs}}) = n\bar{x}\log(\pi) + n(m-\bar{x})\log(1-\pi) - n\log[1-(1-\pi)^m]$$
$$= n[\bar{x}\log(\pi) + (m-\bar{x})\log(1-\pi) + h(\pi)]$$

where $h(\pi) = -\log[1-(1-\pi)^\pi]$. Define $(1-\pi)^m = e{-}\lambda$, $\lambda = -m\log(1-\pi)$. Let $h(\pi) = -\log(1-e^{-\lambda}) = g(\lambda)$. Since $g'(\lambda) = -e^{-\lambda}/(1-e^{-\lambda})$ and $g''(\lambda) = e^{-\lambda}/(1-e^{-\lambda})^2 > 0$ for all $lambda > 0$. Thus, $g(\lambda)$ is convex. Then we have

$$g(\lambda) \geq g(\lambda^{(t)}) + (\lambda - \lambda^{(t)})g'(\lambda^{(t)}), \ \forall \lambda > 0, \ \lambda^{(t)} > 0$$

or equivalently,

$$h(\pi) \geq h(\pi^{(t)}) - \frac{(1-\pi^{(t)})^m}{1-(1-\pi^{(t)})^m}[-m\log(1-\pi) + m\log(1-\pi^{(t)})]$$
$$= c_0 + \frac{m(1-\pi^{(t)})^m}{1-(1-\pi^{(t)})^m}\log(1-\pi)$$

Now the log-likelihood function is

$$l(\pi|Y_{\text{obs}}) = n[\bar{x}\log(\pi) + (m-\pi)\log(1-\pi) + h(\pi)]$$
$$\geq n\Big[\bar{x}\log(\pi) + (m-\bar{x})\log(1-\pi) + c_0 + \frac{m(1-\pi^{(t)})^m}{1-(1-\pi^{(t)})^m}\log(1-\pi)\Big]$$

Let $Q(\pi|\pi^{(t)}) = n\Big[\bar{x}\log(\pi) + (m-\bar{x})\log(1-\pi) + c_0 + \frac{m(1-\pi^{(t)})^m}{1-(1-\pi^{(t)})^m}\log(1-\pi)\Big]$, then $dQ(\pi|\pi^{(t)})/d\lambda = 0$. Hence, a MM algorithm is given by

$$\pi^{(t+1)} = \frac{\bar{x}[1-(1-\pi^{(t)})^m]}{m}$$

(b)

Bootstrap method:
Step 1: Calculate the MLEs $\hat{\pi}$ by the MM algorithm.
Step 2: Generate a bootstrap sample $\mathbf{x}^* = (x_1^*, ..., x_n^*) \overset{iid}{\sim} \text{ZTB}(m, \hat{\pi})$ and compute the corresponding bootstrap replication $\hat{\pi}^*$.
Step 3: Independently repeat Step 2 G times. Obtain G bootstrap replications $\{\hat{\pi}^*(g)\}_{g=1}^G$
Step 4: Compute $\widehat{\text{Se}}(\hat{\pi})$ and $\widehat{\text{Se}}(\hat{\pi})$.

$$\widehat{\text{Se}}(\hat{\pi}) = \sqrt{\frac{1}{G-1}\sum_{g=1}^{G}[\hat{\pi}^*(g) - \bar{\pi}^*]^2}, \quad \bar{\pi}^* = [\hat{\pi}^*(1), +... + \hat{\pi}^*(G)]/G$$

Step 5: If $\{\hat{\pi}^*(g)\}_{g=1}^G$ are approximately normallly distributed, a $100(1-\alpha)\%$ bootstrap CI for π is given by

$$[\hat{\pi}_l^*, \hat{\pi}_r^*] = [\bar{\pi}^* - z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\pi}), \bar{\pi}^* + z_{\alpha/2}\widehat{\text{Se}}^*(\hat{\pi})]$$

Step 6: If the bootstrap CI is beyond the unit interval $[0, 1]$ or the bootstrap replications $\{\widehat{\pi}^*(g)\}_{g=1}^G$ are non-normally distributed , a $100(1-\alpha)\%$ bootstrap CI for π is given by

$$[\widehat{\pi}^*_L, \widehat{\pi}^*_U]$$

where $\widehat{\pi}^*_L$ and $\widehat{\pi}^*_U$ are the $(\alpha/2)$G-th and the $(1-\alpha/2)$G-th order statistics of $\{\widehat{\pi}^*(g)\}_{g=1}^G$.

5.4

Question Assume that the following observations

$$0.84,\ 8.26,\ 2.07,\ 2.79,\ 5.86,\ 6.98,\ 8.67,\ 9.07,\ 5.65,\ 9.96,$$
$$9.79,\ 4.17,\ 3.49,\ 1.06,\ 0.62,\ 9.89,\ 0.11,\ 9.41,\ 5.79,\ 0.99,$$
$$8.02,\ 5.06,\ 3.96,\ 0.58,\ 2.92,\ 8.63,\ 6.21,\ 1.17,\ 3.09,\ 4.58$$

are from a normal distribution with mean μ and variance σ^2. Use the parametric bootstrap method to find the 95% CIs for the population median.

Solution Let $x_1, ..., x_{30}$ denote these observations. Let θ denote the median, then $\hat{\theta} = x_{(15)} + x_{(16)}/2 = 4.82$. And $\hat{\mu} = \bar{x} = 4.989667$, $\hat{\sigma}^2 = \frac{\sum_{i=1}^{30}(x_i - \bar{x})^2}{30} = 10.43058$. We first generated $G = 20000$ bootstrap samples from $N(\hat{\mu}, \hat{\sigma}^2)$, computed 20000 bootstrap replications $\{\hat{\theta}^*(g)\}_{g=1}^{G}$, and obtained

$$\bar{\theta}^* = 4.9879827, \quad \widehat{\text{Se}}^*(\hat{\theta}) = 0.7236293, \quad [\hat{\theta}_L^*, \hat{\theta}_U^*] = [3.5701312, 6.4027418]$$

R code

```
assign5.4 <- function(G){
  x <- c(0.84, 8.26, 2.07, 2.79, 5.86, 6.98, 8.67, 9.07, 5.65, 9.96,
         9.79, 4.17, 3.49, 1.06, 0.62, 9.89, 0.11, 9.41, 5.79, 0.99,
         8.02, 5.06, 3.96, 0.58, 2.92, 8.63, 6.21, 1.17, 3.09, 4.58)
  n <- length(x)
  mu <- mean(x)
  sigma <- sqrt((n - 1) * var(x) / n)
  thMLE <- median(x)
  th.star.sample <- matrix(0, G, 1)
  for (g in 1:G) {
    xstar <- rnorm(n, mean = mu, sd = sigma)
    thstar <- median(xstar)
    th.star.sample[g, 1] <- thstar
  }

  M <- mean.std.CI(th.star.sample)
  result <- c(M, thMLE)
  return(result)
}

assign5.4(20000)
```

```
## [1] 4.9879827 0.7236293 3.5696693 6.4062962 3.5701312 6.4027418 4.8200000
```