

Assignment 2: MATH3911

Alexander Kwok¹

^[1]email: z3332883@zmail.unsw.edu.au;

Assignment 2: Version from May 26, 2013

This assignment is my own work. I have read and understood the University Rules in respect to Student Academic Misconduct.

Question 1

- (a) It is given that X_1, X_2, \dots, X_{10} are *i.i.d* distributed random variables from $N(\mu, 4)$ with a density function as follow:

$$f(x; \mu) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{8}}$$

Therefore the likelihood function of X_i is

$$\begin{aligned} L(x; \mu) &= \prod_{i=1}^{10} f(x_i; \mu) \\ &= \left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} (x_i - \mu)^2 \right) \\ &= \left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} (x_i^2 - 2x_i\mu + \mu^2) \right) \\ &= \left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} x_i^2 \right) \exp\left(\frac{1}{4}\mu \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{10}{8}\mu^2 \right) \\ &= \left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} x_i^2 \right) \exp\left(\frac{1}{4}\mu \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{5}{4}\mu^2 \right) \end{aligned}$$

Then the likelihood ratio can be written as

$$\begin{aligned} \frac{L(x; \mu_2)}{L(x; \mu_1)} &= \frac{\left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} x_i^2 \right) \exp\left(\frac{1}{4}\mu_2 \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{5}{4}\mu_2^2 \right)}{\left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} x_i^2 \right) \exp\left(\frac{1}{4}\mu_1 \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{5}{4}\mu_1^2 \right)} \\ &= \exp\left(\frac{1}{4}[\mu_2 - \mu_1] \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{5}{4}(\mu_2^2 - \mu_1^2) \right) \end{aligned}$$

which is a non-decreasing function of $\sum_{i=1}^{10} X_i$ for any fixed μ_1 and μ_2 , with $\mu_1 < \mu_2$. Therefore, the joint density of X_1, X_2, \dots, X_{10} has monotone likelihood ratio in $T(X) = \sum_{i=1}^{10} X_i$.

(b) By rearranging the likelihood ratio as

$$\begin{aligned} L(x; \mu) &= \left(\frac{1}{2\sqrt{2\pi}} \right)^{10} \exp\left(-\frac{1}{8} \sum_{i=1}^{10} x_i^2 \right) \exp\left(\frac{1}{4} \mu \sum_{i=1}^{10} x_i \right) \exp\left(-\frac{5}{4} \mu^2 \right) \\ &= \left(\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8} \mu^2} \right)^{10} \left(\prod_{i=1}^{10} e^{-\frac{1}{8} x_i^2} \right) \exp\left(\frac{1}{4} \mu \sum_{i=1}^{10} x_i \right) \end{aligned}$$

which is in the form

$$(a(\mu))^n \left(\prod_{i=1}^n b(x_i) \right) \exp\left(c(\mu) \sum_{i=1}^n d(x_i) \right)$$

with $T(X) = \sum_{i=1}^{10} d(X_i) = \sum_{i=1}^{10} X_i$.

Therefore, the UMP unbiased size $\alpha = 0.1$ test of $H_0 : \mu = 1$ versus $H_1 : \mu \neq 1$ is

$$\varphi^*(X) = \begin{cases} 1 & \text{if } T(X) < c_1 \text{ or } T(X) > c_2 \\ 0 & \text{if } c_1 \leq T(X) \leq c_2 \end{cases}$$

where the constants $c_i (i = 1, 2)$, satisfy the conditions:

$$E_{\mu=1}[\varphi^*(X)] = 0.1 \quad (1)$$

$$\left[\frac{\partial}{\partial \mu} E_{\mu}[\varphi^*(X)] \right]_{\mu=1} = 0 \quad (2)$$

Note that

$$T(X) = \sum_{i=1}^{10} X_i \sim N(10\mu, 40)$$

and

$$\frac{T(X) - 10\mu}{\sqrt{40}} \sim N(0, 1)$$

Therefore, the power function can be deduced as

$$\begin{aligned} E_{\mu}[\varphi^*(X)] &= P[\varphi^*(X) = 1] \\ &= P[T(X) < c_1] + P[T(X) > c_2] \\ &= P\left(\frac{T(X) - 10\mu}{\sqrt{40}} < \frac{c_1 - 10\mu}{\sqrt{40}} \right) + P\left(\frac{T(X) - 10\mu}{\sqrt{40}} > \frac{c_2 - 10\mu}{\sqrt{40}} \right) \\ &= \Phi\left(\frac{c_1 - 10\mu}{2\sqrt{10}} \right) + 1 - \Phi\left(\frac{c_2 - 10\mu}{2\sqrt{10}} \right) \end{aligned}$$

As a result, we have

$$\begin{aligned} \frac{\partial}{\partial \mu} E_{\mu}[\varphi^*(X)] &= -\frac{10}{2\sqrt{10}} \Phi'\left(\frac{c_1 - 10\mu}{2\sqrt{10}} \right) + \frac{10}{2\sqrt{10}} \Phi'\left(\frac{c_2 - 10\mu}{2\sqrt{10}} \right) \\ &= -\frac{\sqrt{10}}{2} \Phi'\left(\frac{c_1 - 10\mu}{2\sqrt{10}} \right) + \frac{\sqrt{10}}{2} \Phi'\left(\frac{c_2 - 10\mu}{2\sqrt{10}} \right) \end{aligned}$$

where

$$\Phi'(z) = \frac{d}{dz} \Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

which is the standard normal density function.

Use condition (??), we have

$$\begin{aligned} \left[\frac{\partial}{\partial \mu} E_{\mu}[\varphi^*(X)] \right]_{\mu=1} &= 0 \\ -\frac{\sqrt{10}}{2} \Phi' \left(\frac{c_1 - 10}{2\sqrt{10}} \right) + \frac{\sqrt{10}}{2} \Phi' \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0 \\ \phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= \phi \left(\frac{c_1 - 10}{2\sqrt{10}} \right) \end{aligned}$$

Since $c_1 \neq c_2$ and using the fact that density of normal distribution is symmetric around zero, then

$$\begin{aligned} \frac{c_2 - 10}{2\sqrt{10}} &= -\frac{c_1 - 10}{2\sqrt{10}} \\ c_1 &= 20 - c_2 \end{aligned}$$

Then substitute $c_1 = 20 - c_2$ to (??) and using the fact that $\Phi(-z) = 1 - \Phi(z)$,

$$\begin{aligned} E_{\mu=1}[\varphi^*(X)] &= 0.1 \\ \Phi \left(\frac{c_1 - 10}{2\sqrt{10}} \right) + 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.1 \\ \Phi \left(\frac{20 - c_2 - 10}{2\sqrt{10}} \right) + 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.1 \\ \Phi \left(\frac{-c_2 + 10}{2\sqrt{10}} \right) + 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.1 \\ 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) + 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.1 \\ 1 - \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.05 \\ \Phi \left(\frac{c_2 - 10}{2\sqrt{10}} \right) &= 0.95 \\ \frac{c_2 - 10}{2\sqrt{10}} &= \Phi^{-1}(0.95) \\ c_2 &= 10 + 2\sqrt{10} \times \Phi^{-1}(0.95) \end{aligned}$$

Therefore,

$$c_2 \approx 10 + 2\sqrt{10} \times (1.644854) = 20.40297 \text{ and } c_1 = 20 - c_2 \approx -0.40297$$

Then the UMP unbiased size $\alpha = 0.1$ is

$$\varphi^*(X) = \begin{cases} 1 & \text{if } T(X) < -0.40297 \text{ or } T(X) > 20.40297 \\ 0 & \text{if } -0.40297 \leq T(X) \leq 20.40297 \end{cases}$$

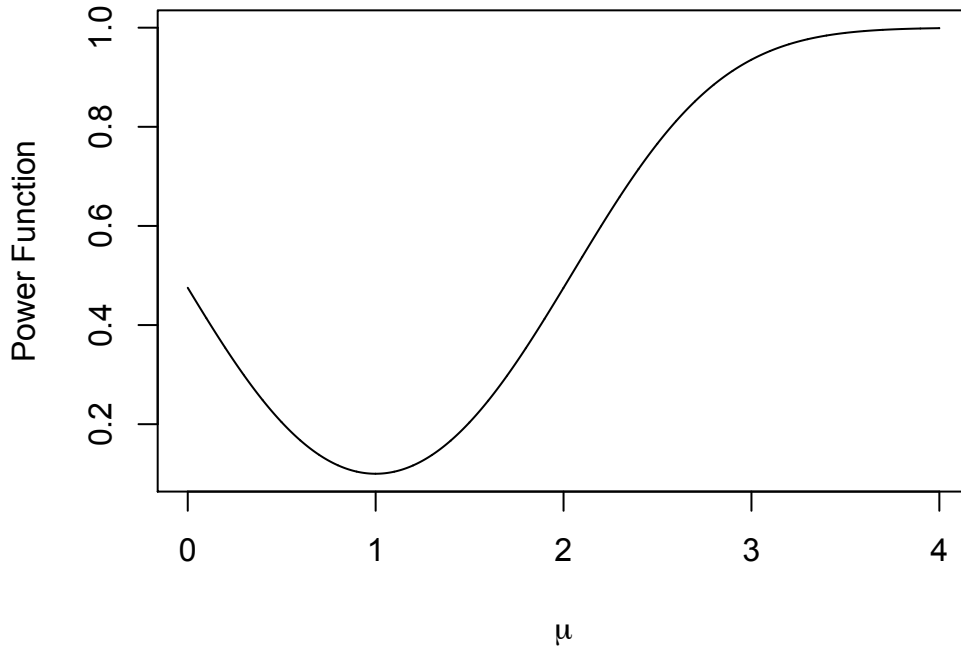
(c) The power function of this test is found in in part (b) as

$$E_{\mu}[\varphi^*(X)] = \Phi\left(\frac{-0.40297 - 10}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297 - 10}{2\sqrt{10}}\right)$$

So by Evaluating the power numerically for $\mu = 0, 0.5, 1.5, 2.5$, and 4, we have the following,

$$\begin{aligned} E_{\mu=0}[\varphi^*(X)] &= \Phi\left(\frac{-0.40297}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297}{2\sqrt{10}}\right) \approx 0.4752263 \\ E_{\mu=0.5}[\varphi^*(X)] &= \Phi\left(\frac{-0.40297 - 5}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297 - 5}{2\sqrt{10}}\right) \approx 0.2039110 \\ E_{\mu=1.5}[\varphi^*(X)] &= \Phi\left(\frac{-0.40297 - 15}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297 - 15}{2\sqrt{10}}\right) \approx 0.2039110 \\ E_{\mu=2.5}[\varphi^*(X)] &= \Phi\left(\frac{-0.40297 - 25}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297 - 25}{2\sqrt{10}}\right) \approx 0.7663720 \\ E_{\mu=4}[\varphi^*(X)] &= \Phi\left(\frac{-0.40297 - 40}{2\sqrt{10}}\right) + 1 - \Phi\left(\frac{20.40297 - 40}{2\sqrt{10}}\right) \approx 0.9990277 \end{aligned}$$

And hence using the above value, we can sketch the power function for all values of μ on the real axis:



(d) To obtain the density of the second order statistic $X_{(2)}$, we use the following identity:

$$f_{X_{(k)}}(x) = k \binom{n}{k} [F(x; \mu)]^{k-1} [1 - F(x; \mu)]^{n-k} f(x; \mu)$$

Therefore

$$\begin{aligned} f_{X_{(2)}}(x) &= 2 \binom{10}{2} [F(x; \mu)]^{2-1} [1 - F(x; \mu)]^{10-2} f(x; \mu) \\ &= 2 \frac{10!}{2!8!} [F(x; \mu)] [1 - F(x; \mu)]^8 f(x; \mu) \\ &= \frac{10!}{8!} [F(x; \mu)] [1 - F(x; \mu)]^8 f(x; \mu) \end{aligned}$$

also it is given that $f(x; \mu) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{8}}$, we also know that

$$\frac{X_i - \mu}{2} \sim N(0, 1)$$

for all $i = 1, 2, \dots, 10$

$$\begin{aligned} F(x; \mu) &= P[X_i < x] \\ &= P\left[\frac{X_i - \mu}{2} < \frac{x - \mu}{2}\right] \\ &= \Phi\left(\frac{x - \mu}{2}\right) \end{aligned}$$

Therefore, the density of the 2nd order statistic $X_{(2)}$ is

$$f_{X_2}(x; \mu) = \frac{90}{2\sqrt{2\pi}} e^{-\frac{1}{8}(x-\mu)^2} \left[\Phi\left(\frac{x - \mu}{2}\right) \right] \left[1 - \Phi\left(\frac{x - \mu}{2}\right) \right]^8$$

and under $H_0 : \mu = 1$

$$f_{X_2}(x; 1) = \frac{45}{\sqrt{2\pi}} e^{-\frac{1}{8}(x-1)^2} \left[\Phi\left(\frac{x - 1}{2}\right) \right] \left[1 - \Phi\left(\frac{x - 1}{2}\right) \right]^8$$

Then by using the following R command:

```
> fun=function(x){(45/sqrt(2*pi))*exp(-(x-1)^2/8)*pnorm((x-1)/2)*(1-pnorm((x-1)/2))^8}
> integrate(fun, -Inf, 0)
0.8635273 with absolute error < 8.3e-09
```

As a result we found that

$$P(X_2 < 0) = 0.8635273$$

Question 2

(a) Firstly we note that N_1, \dots, N_T are all *i.i.d* random variable with common density being $Poisson(\lambda)$ so

$$f_{N_i|\Lambda}(n_i|\lambda) = e^{-\lambda} \frac{\lambda^{n_i}}{n_i!}$$

The conditional joint density of \mathbf{N} is

$$\begin{aligned} f_{N_1, N_2, \dots, N_T|\Lambda}(n_1, n_2, \dots, n_T|\lambda) &= \prod_{i=1}^T f_{N_i|\Lambda}(n_i|\lambda) \\ &= \frac{e^{-\lambda T} \lambda^{\sum_{i=1}^T n_i}}{\prod_{i=1}^T n_i!} \end{aligned}$$

The prior distribution on Λ is $Gamma(a, b)$ with known $a > 0, b > 0$:

$$\tau(\lambda) = \frac{\lambda^{a-1} e^{-\lambda/b}}{\Gamma(a) b^a}, \lambda > 0$$

Using both of the given information from above we can deduce that for $N = (N_1, N_2, \dots, N_T)$ given λ :

$$\begin{aligned} h(\lambda|N) &\propto f(N|\lambda)\tau(\lambda) \\ &= \frac{e^{-\lambda T} \lambda^{\sum_{i=1}^T n_i}}{\prod_{i=1}^T (n_i!)} \frac{\lambda^{a-1} e^{-\lambda/b}}{\Gamma(a) b^a} \\ &= \frac{\lambda^{\sum_{i=1}^T n_i + a - 1} e^{-\lambda(T+1/b)}}{\Gamma(a) b^a \prod_{i=1}^T (n_i!)} \\ h(\lambda|N) &\propto \lambda^{\sum_{i=1}^T n_i + a - 1} e^{-\lambda(T+1/b)} \end{aligned}$$

which means that $\lambda|N \sim Gamma(\sum_{i=1}^T n_i + a, \frac{1}{T+1/b})$.

Then by applying the Expected value property of the Gamma function, the Bayes estimator of λ with respect to quadratic loss is given as follow

$$E[\lambda|N] = \frac{\sum_{i=1}^T n_i + a}{T + 1/b}$$

(b) Assume $a = 3, b = 2$, if the counts within the last seven years were 2, 4, 7, 3, 4, 4, 5, then we also know that $\sum_{i=1}^7 n_i = 2 + 4 + 7 + 3 + 4 + 4 + 5 = 29$

Therefore the estimate of λ for this data

$$E[\lambda|N] = \frac{29 + 3}{7 + \frac{1}{2}} = \frac{64}{15} \approx 4.2666667$$

(c) Assuming $a = 3, b = 2, T = 3$ and using the fact that $\sum_{i=1}^7 n_i = 29$, we can found the posterior distribution with respect to λ as follow

$$h(\lambda|N) = \frac{\lambda^{31} e^{-\lambda/7.5}}{\Gamma(32)(1/7.5)^{32}}$$

To test for the statement "the yearly intensity λ is less than 4", we can construct a hypothesis test as follow

$$H_0 : \lambda < 4 \text{ vs } H_1 : \lambda \geq 4$$

Using the Bayesian testing with a zero-one loss. We then calculate posterior probability given the sample as follow

$$P(\lambda \in \Lambda_0|N) = \int_0^4 \frac{\lambda^{31} e^{-\lambda/7.5}}{\Gamma(32)(1/7.5)^{32}} d\lambda = 0.0.381357$$

which is calculated using the following R code:

```
> pgamma(4,32,15/2)
[1] 0.381357
```

As a result we reject H_0 since the posterior probability is less than $\frac{1}{2}$. Therefore, we can conclude that the statement claimed by the bank is false.

Question 3

```
(a) > #Question 3
> #Part a
> #myskewness
> mkskewness = function(x){
+   n= length(x)
+   res= x - mean(x)
+   numerator = sqrt(n)*sum(res^3)
+   denominator = (sum(res^2))^(3/2)
+   skew=numerator/denominator
+   return(skew)
+ }
> #mykurtosis
> mykurtosis = function(x){
+   n=length(x)
+   res = x-mean(x)
+   numerator = n*sum(res^4)
+   denominator = (sum(res^2))^2
+   kurtosis = numerator/denominator -3
+   return(kurtosis)
+ }

(b) > #Part b)
> time = lung[,2]
> myskewness (time)
[1] 1.086788
> skewness(time)
[1] 1.093999
> kurtosis(time)
[1] 0.9401333
> mykurtosis(time)
[1] 0.8934417
```

By observing the above output, we observe that there is a difference between the output value between "myskewness", "mykurtosis" and "skewness", "kurtosis".

Therefore, we obtain a different value for "myskewness()", "mykurtosis()" compare to "skewness()", "mykurtosis()".

However, by changing the method argument to "moment", "skewness(method = "moment")", "mykurtosis("moment")", we then obtain the same numerical result as "myskewness()", "mykurtosis()".

```
> skewness(time, method="moment")
[1] 1.086788
> kurtosis(time, method="moment")
[1] 0.8934417
```

An explanation of this can be found from the help of the in-built functions. We notice that there is an optional arguments for "method" which can take in either "fisher" for Fisher's g1(skewness) and g2(kurtosis) version, and "moment" for the functional forms of the statistic.

Since the default option is "fisher", then when we obtain the value using "skewness" and "kurtosis" without specifying the method argument, we are essentially calculating Fisher's g1 and g2 estimates. Therefore, when we specify "moment" in the "method" argument, we will obtain the same value as our own functions.

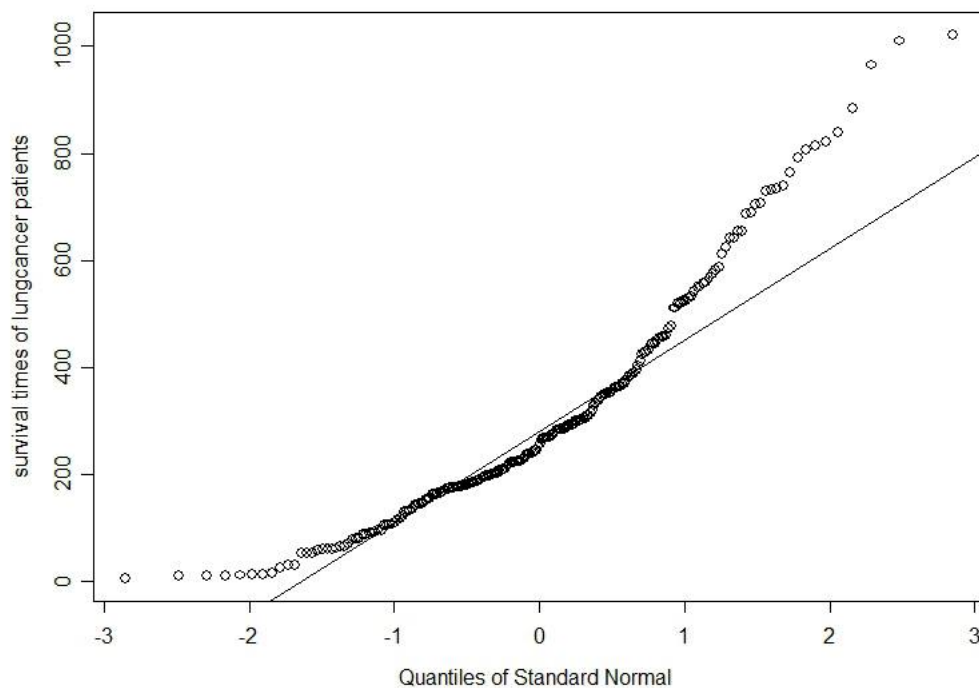
- (c) The code for Bootstrap the $\hat{\gamma}_1$ and $\hat{\gamma}_2$ estimators by using $B = 2000$ replicates is provided below. Then the output containing the 95% confidence interval using the BCa method is also shown below:

```
> #Part c)
> boot.obj1<-bootstrap(time,myskewness(time),B=2000)
> boot.obj2<-bootstrap(time,mykurtosis(time),B=2000)
> summary(boot.obj1)
Call:
bootstrap(data = time, statistic = myskewness(time), B = 2000)
Number of Replications: 2000
Summary Statistics:
Observed Mean Bias SE
Param 1.087 1.074 -0.01233 0.1402
Percentiles:
2.5% 5% 95% 97.5%
Param 0.7974 0.8453 1.312 1.35
BCa Confidence Intervals:
2.5% 5% 95% 97.5%
Param 0.8319 0.87 1.337 1.387
> summary(boot.obj2)
Call:
bootstrap(data = time, statistic = mykurtosis(time), B = 2000)
Number of Replications: 2000
Summary Statistics:
Observed Mean Bias SE
Param 0.8934 0.8552 -0.03819 0.4973
Percentiles:
2.5% 5% 95% 97.5%
Param -0.02609 0.09735 1.715 1.931
BCa Confidence Intervals:
2.5% 5% 95% 97.5%
Param 0.09131 0.2115 1.923 2.159
```

According to the output, the 95% BCa confidence intervals for the skewness is (0.8319, 1.387) and the 95% BCa confidence intervals for the kurtosis is (0.09131, 2.159).

- (d) Since both of the confidence interval for skewness and kurtosis exclude 0, we can conclude that the normality of the lung data is in doubt. The normal quantile-quantile plot is not close to a straight line, which confirms the conclusion that was made before that the lung data may not be normally distributed. We can also observe that the data is skewed as the dot does not form a straight line and the kurtosis is clearly not a straight line, therefore, it confirms with the conclusion again.

Quantile-Quantile Plot for the Lung data

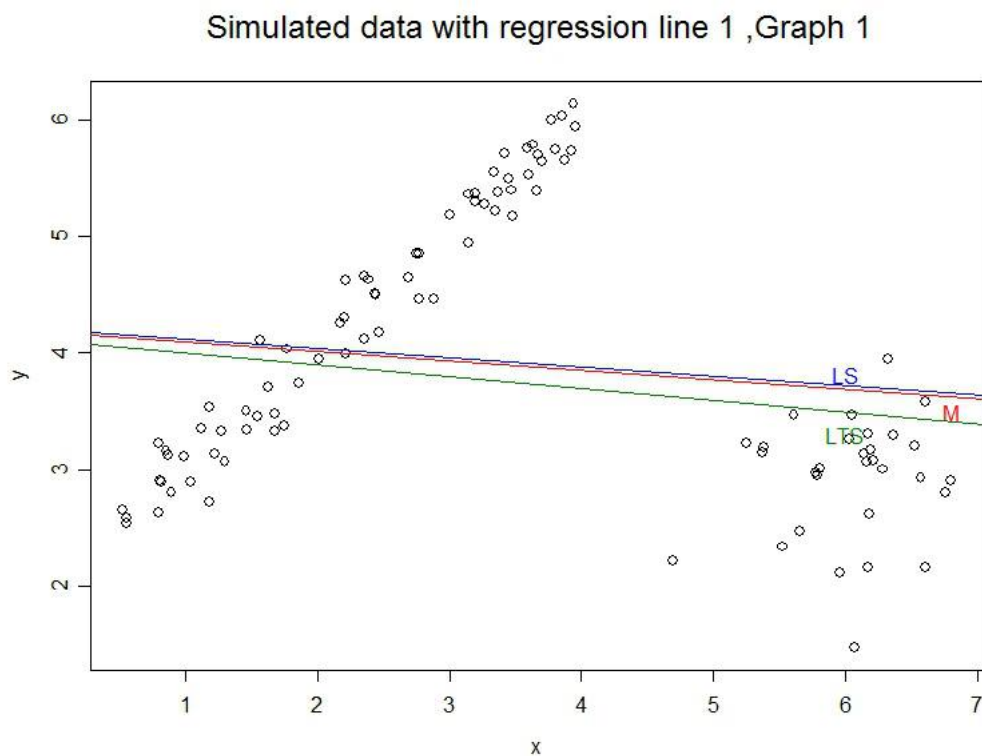


```
> #Part d)
> qqnorm(time, ylab="survival times of lungcancer patients", main="Quantile-Quantile Plot
> for the Lung data")
> qqline(time)
```

Question 4

```
> set.seed(round(log(3332883)))
> x70 <- runif(70, 0.5, 4)
> e70 <- rnorm(70, mean = 0, sd = 0.2)
> y70 <- 2 + x70 + e70
> x30 <- rnorm(30, mean = 6, sd = 0.5)
> y30 <- rnorm(30, mean = 3, sd = 0.5)
> x <- c(x70, x30)
> y <- c(y70, y30)
> simuldata <- data.frame(x, y)
```

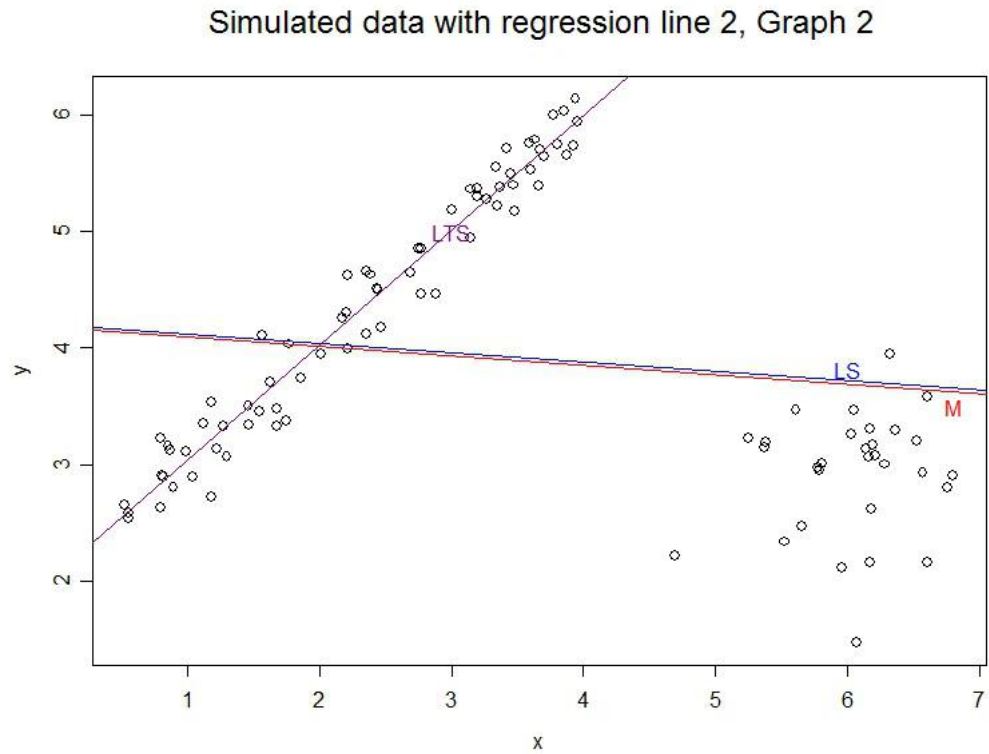
i) Graph 1:



R code:

```
> plot(x, y, main = "Simulated data with regression line 1 ,Graph 1")
> abline(lm(y ~ x, simuldata), col = "blue")
> text(6, 3.82, "LS", col = "blue")
> abline(rreg(x, y), col = "red")
> text(6.8, 3.5, "M", col = "red")
> abline(ltsreg.formula(y ~ x, simuldata), col = "green")
> text(6, 3.3, "LTS", col = "green")
```

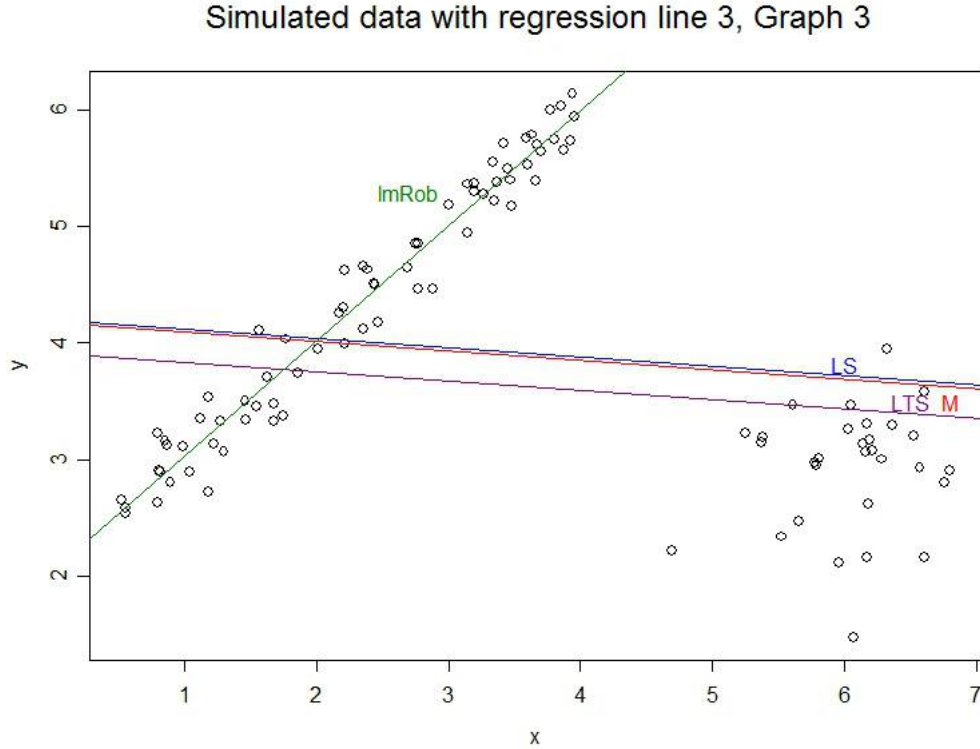
ii) Graph 2:



R code:

```
> #line 2
plot(x, y, main = "Simulated data with regression line 2, Graph 2")
> abline(lm(y ~ x, simuldata), col = "blue")
> text(6, 3.82, "LS", col = "blue")
> abline(rreg(x, y), col = "red")
> text(6.8, 3.5, "M", col = "red")
> abline(ltsreg.formula(y ~ x, simuldata, quan = 70), col = "purple")
> text(3, 5, "LTS", col = "purple")
```

iii) **Graph 3:**



R code:

```
> #lmRob, with residual = 85
plot(x, y, main = "Simulated data with regression line 3, Graph 3")
> abline(lm(y ~ x, simuldata), col = "blue")
> text(6, 3.82, "LS", col = "blue")
> abline(rreg(x, y), col = "red")
> text(6.8, 3.5, "M", col = "red")
> abline(ltsreg.formula(y ~ x, simuldata, quan = 85), col = "purple")
> text(6.5, 3.5, "LTS", col = "purple")
> abline(lmRob(y ~ ., simuldata), col = "green")
> text(2.7, 5.3, "lmRob", col = "green")
```

Comment:

The LTS estimator with 85 residuals consider most of the residuals when plotting the line. In this case, we can observe that , from the graph, using 85 residuals will not deliver a good fit compare to the previous one which only contains 70 residuals.

In LTS, the regression line is fitted by minimising the sum of squares of the m smallest residuals, where $m \leq n$. That is, by minimising the objective function

$$S_{LTS} = \sum_{i=1}^m |e_{(i)}|^2$$

This method is robust to outliers as the largest $n-m$ outliers do not affect the function S_{LTS} .

In part i), all three regression lines were highly influenced by outliers. Even LTS was not sufficiently robust as by default, $m=90$ was chosen, which was too high since 20 clear outliers were included. In part ii), $m=70$ was chosen, which removed the effect of the 30 clear outliers. This resulted in a more accurate representation of the model. In part iii), $m=85$ was chosen , which was higher than 70 and lower than 90, as a result it still include 15 clear outliers. Therefore, it produced lines that were highly

influenced by outliers.

However, the `lmRob` procedure produced a better fit which is the green line in the above graph. As it is similar to the previous line which used 70 residuals, this suggested that the `lmRob` does actually adjust "automatically" to the amount of contamination after starting with a high-breakdown point estimator. Therefore, the claim could be believed to be true that `lmRob` can produce a good fit with its default settings.

Question 5

a) i) Now consider X_1, X_2, \dots, X_n are i.i.d. uniform $[0, \theta)$ random variable, with following density function:

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{for } 0 \leq x < \theta \\ 0 & \text{for otherwise} \end{cases}$$

and we can find its cumulative distribution as follow

$$F(x; \theta) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{\theta} & \text{for } 0 \leq x \leq \theta \\ 1 & \text{for otherwise} \end{cases}$$

Consider the estimator $T_n = X_{(n)} = \max\{X_1, \dots, X_n\}$ of θ . Using the follow identity,

$$f_{X_{(k)}}(y) = k \binom{n}{k} [F(y; \theta)]^{k-1} [1 - F(y; \theta)]^{n-k} f(y; \theta)$$

Since $F(x; \theta)$ is a uniform distribution on $[0, \theta)$, we have

$$\begin{aligned} f_{X_{(k)}}(y) &= k \frac{n!}{k!(n-k)!} \left(\frac{y}{\theta}\right)^{k-1} \left(\frac{y}{\theta}\right)^{n-k} \frac{1}{\theta} & \text{for } y \in [0, \theta) \\ &= \frac{n!}{(k-1)!(n-k)!} \left(\frac{y}{\theta}\right)^{k-1} \left(\frac{y}{\theta}\right)^{n-k} \frac{1}{\theta} \end{aligned}$$

Let $y = u\theta$,

$$f_{X_{(k)}}(u\theta) = \frac{n!}{(k-1)!(n-k)!} u^{k-1} u^{n-k}$$

Compare to the beta distribution $Z \sim \text{Beta}(\alpha, \beta)$

$$f(z) = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!} z^{\alpha-1} (1 - z)^{\beta-1}$$

Therefore

$$\theta f_{X_{(k)}}(u\theta) \sim \text{Beta}(\alpha = 1, \beta = n - k + 1)$$

Then denote $B \sim \text{Beta}(k, n - k + 1)$, we can find

$$\begin{aligned} E[X_{(k)}] &= \int_0^\theta y f_{X_{(k)}}(y) dy \\ &= \int_0^1 u\theta f_{X_{(k)}}(u\theta) \theta du & \text{substitute } y = u\theta \\ &= \theta \int_0^1 u \left(\theta f_{X_{(k)}}(u\theta) \right) du \\ &= \theta \int_0^1 u f_B(u) du \\ &= \theta E[B] \\ &= \frac{k}{k + n - k + 1} \theta \\ &= \frac{k}{n + 1} \theta \end{aligned}$$

Then we can obtain

$$E[X_{(n-1)}] = \frac{n-1}{n+1} \theta$$

Using this fact, we proceed to the next part. First we need to find the $bias(JK(T_n))$. It is known that the jackknife estimator of θ is

$$JK(T_n) = nT_n - \frac{n-1}{n} \sum_{i=1}^n T_n^{(i)}$$

where $T_n^{(i)} = \max\{X_1, \dots, X_{i-1}, X_{i+1}\}$ is the i^{th} jackknife replication of T_n .

Then if $X_i \neq X_{(n)}$, the i^{th} jackknife sample $\{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$ contains $X_{(n)}$. As $X_{(n)}$ was the largest value in the original sample, it must also be the greatest value in the jackknife sample, so $T_n^{(i)} = X_{(n)}$.

If $X_i = X_{(n)}$, the i^{th} jackknife sample does not contain X_n , so the largest value in the jackknife sample must be the second largest value in the original sample. Hence $T_n^{(i)} = X_{(n-1)}$.

Then applying the above 2 cases,

$$\begin{aligned} JK(T_n) &= nT_n - \frac{n-1}{n} \sum_{i=1}^n T_n^{(i)} \\ &= nX_{(n)} - \frac{n-1}{n} [(n-1)X_{(n)} + X_{(n-1)}] \\ &= X_{(n)} \left(n - \frac{(n-1)^2}{n} \right) - \frac{n-1}{n} X_{(n-1)} \\ &= X_{(n)} \left(\frac{n^2 - n^2 + 2n - 1}{n} \right) - \frac{n-1}{n} X_{(n-1)} \\ &= \frac{2n-1}{n} X_{(n)} - \frac{n-1}{n} X_{(n-1)} \end{aligned}$$

$$\begin{aligned} E[JK(T_n)] &= \frac{2n-1}{n} X_{(n)} - \frac{n-1}{n} X_{(n-1)} \\ &= \frac{2n-1}{n} E[X_{(n)}] - \frac{n-1}{n} E[X_{(n-1)}] \end{aligned}$$

applying the result we derived before, we know that

$$E[X_{(n)}] = \frac{n}{n+1} \theta \text{ and } E[X_{(n-1)}] = \frac{n-1}{n+1} \theta$$

Then we get

$$\begin{aligned} E[JK(T_n)] &= \frac{2n-1}{n} \left(\frac{n}{n+1} \right) \theta - \frac{n-1}{n} \left(\frac{n-1}{n+1} \right) \theta \\ &= \frac{\theta}{n+1} \left(2n-1 - \frac{(n-1)^2}{n} \right) \\ &= \frac{\theta}{n(n+1)} \left(2n^2 - n - n^2 + 2n - 1 \right) \\ &= \frac{\theta}{n(n+1)} \left(n^2 + n - 1 \right) \end{aligned}$$

Then

$$\begin{aligned}
bias(JK(T_n)) &= E[JK(T_n)] - \theta \\
&= \frac{\theta}{n(n+1)} \left(n^2 + n - 1 \right) - \theta \\
&= \frac{\theta}{n(n+1)} \left(n^2 + n - 1 - n(n+1) \right) \\
&= \frac{\theta}{n(n+1)} \left(n^2 + n - 1 - n^2 - n \right) \\
&= -\frac{\theta}{n(n+1)}
\end{aligned}$$

Since

$$bias(JK(T_n)) = \left| -\frac{\theta}{n(n+1)} \right| < \left| -\frac{\theta}{(n+1)} \right|$$

Therefore, the bias of $JK(T_n)$ is of smaller magnitude in comparison to $-\frac{\theta}{(n+1)}$.

ii) To find $MSE[T_n]$ and $MSE[JK(T_n)]$. We applied a well known identity

$$MSE[X] = Var[X] + (bias[X])^2$$

Applying this identity, we first decompose $MSE[JK(T_n)]$ as follow

$$MSE[JK(T_n)] = Var[JK(T_n)] + (bias[JK(T_n)])^2$$

When we compute $Var[JK(T_n)] = Var[\frac{2n-1}{n}X_{(n)} - \frac{n-1}{n}X_{(n-1)}]$, it is necessary for us to first obtain $Var[X_{(n)}]$, $Var[X_{(n-1)}]$ and $Cov[X_{(n)}, X_{(n-1)}]$.

Then, first obtain the joint density of $X_{(n)}$ and $X_{(n-1)}$ using the follow identity

$$f_{X_{(j)}, X_{(k)}}(x, y; \theta) = \begin{cases} \binom{n}{k} \binom{k}{j-1} [F_X(x)]^{j-1} [F_X(y) - F_X(x)]^{k-1-j} [1 - F_X(y)]^{n-k} f_X(x) f_X(y) & \text{for } x \leq y \\ 0 & \text{for otherwise} \end{cases}$$

Applying the above identity we then have

$$\begin{aligned}
f_{X_{(n-1)}, X_{(n)}}(u, v; \theta) &= \begin{cases} \binom{n}{n} \binom{n}{n-2} [F_X(u; \theta)]^{n-2} f_X(u; \theta) f_X(v; \theta) & \text{for } u \leq v \\ 0 & \text{for otherwise} \end{cases} \\
&= \begin{cases} \frac{n!}{n!(n-n)!} \frac{n!}{(n-2)!(n-n+2)!} [F_X(u; \theta)]^{n-2} f_X(u; \theta) f_X(v; \theta) & \text{for } u \leq v \\ 0 & \text{for otherwise} \end{cases} \\
&= \begin{cases} \frac{n!}{(n-2)!} [F_X(u; \theta)]^{n-2} f_X(u; \theta) f_X(v; \theta) & \text{for } u \leq v \\ 0 & \text{for otherwise} \end{cases} \\
&= \begin{cases} n(n-1) \left(\frac{u}{\theta}\right)^{n-2} \frac{1}{\theta} \frac{1}{\theta} & \text{for } u \leq v \\ 0 & \text{for otherwise} \end{cases} \\
f_{X_{(n-1)}, X_{(n)}}(u, v; \theta) &= \begin{cases} \frac{n(n-1)}{\theta^n} u^{n-2} & \text{for } u \leq v \\ 0 & \text{for otherwise} \end{cases}
\end{aligned}$$

Now we can proceed to evaluate $E[X_{(n-1)}, X_{(n)}]$

$$\begin{aligned}
E[X_{(n-1)}, X_{(n)}] &= \int_0^\theta \int_0^v uv f_{X_{(n-1)}, X_{(n)}}(u, v; \theta) dudv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v \int_0^v nu^{n-1} dudv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v [u^n]_0^v dv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v^{n+1} dv \\
&= \frac{n-1}{\theta^n} \frac{v^{n+2}}{n+2} \Big|_0^\theta \\
&= \frac{(n-1)\theta^{n+2}}{(n+2)\theta^n} \\
&= \frac{n-1}{n+2} \theta^2
\end{aligned}$$

Recall what we derive before $E[X_{(n)}] = \frac{n}{n+1}\theta$ and $E[X_{(n-1)}] = \frac{n-1}{n+1}\theta$ then

$$\begin{aligned}
Cov[X_{(n-1)}, X_{(n)}] &= E[X_{(n-1)}, X_{(n)}] - E[X_{(n)}]E[X_{(n-1)}] \\
&= \frac{n-1}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)\left(\frac{n-1}{n+1}\theta\right) \\
&= (n-1)\theta^2 \left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right) \\
&= \frac{(n-1)\theta^2}{(n+1)^2(n+2)} \left((n+1)^2 - n(n+2)\right) \\
&= \frac{(n-1)\theta^2}{(n+1)^2(n+2)} \left(n^2 + 2n + 1 - n^2 - 2n\right) \\
&= \frac{(n-1)\theta^2}{(n+1)^2(n+2)}
\end{aligned}$$

To evaluate $Var[X_{(n)}]$ we can break it down to

$$\begin{aligned}
Var[X_{(n)}] &= E[X_{(n)}^2] - E[X_{(n)}]^2 \\
&= \int_0^\theta y^2 f_{X_{(n)}}(y) dy - (\theta E[B])^2 && \text{using part a i) } E[X_{(n)}] = \theta E[B] \\
&= \theta^2 \int_0^1 u^2 (\theta f_{X_{(n)}}(u\theta)) du - (\theta E[B])^2 && \text{substitute } y = u\theta \\
&= \theta^2 \int_0^1 u^2 (\theta f_B(u)) du - (\theta E[B])^2 && \text{where } B \sim \text{Beta}(k, n-k+1) \\
&= \theta^2 E[B^2] - (\theta E[B])^2 \\
&= \theta^2 (E[B^2] - E[B]^2) \\
&= \theta^2 Var(B) \\
&= \theta^2 \frac{k(n-k+1)}{(n+1)^2(n+2)} && \text{since it's } X_{(n)}, \text{ then } k = n \\
Var[X_{(n)}] &= \frac{n\theta^2}{(n+1)^2(n+2)}
\end{aligned}$$

Similarly we can also get

$$Var[X_{(n-1)}] = \frac{2(n-1)\theta^2}{(n+1)^2(n+2)}$$

Finally using all the result we derived above to obtain the $Var[JK(T_n)]$

$$\begin{aligned} Var[JK(T_n)] &= Var\left[\frac{2n-1}{n}X_{(n)} - \frac{n-1}{n}X_{(n-1)}\right] \\ &= \left(\frac{2n-1}{n}\right)^2 Var[X_{(n)}] + \left(\frac{n-1}{n}\right)^2 Var[X_{(n-1)}] - 2\left(\frac{2n-1}{n}\right)\left(\frac{n-1}{n}\right)Cov[X_{(n-1)}, X_{(n)}] \\ &= \frac{n(2n-1)^2\theta^2}{n^2(n+1)^2(n+2)} + \frac{2(n-1)^3\theta^2}{n^2(n+1)^2(n+2)} - \frac{2(2n-1)(n-1)^2\theta^2}{n^2(n+1)^2(n+2)} \\ &= \frac{n(2n-1)^2 + (n-1)^2[2(n-1) - 2(2n-1)]}{n^2(n+1)^2(n+2)}\theta^2 \\ &= \frac{n(2n-1)^2 + (n-1)^2[2n-2-4n+2]}{n^2(n+1)^2(n+2)}\theta^2 \\ &= \frac{(2n-1)^2 + (n-1)^2[-2]}{n(n+1)^2(n+2)}\theta^2 \\ &= \frac{4n^2 - 4n + 1 - 2n^2 + 4n - 2}{n(n+1)^2(n+2)}\theta^2 \\ &= \frac{2n^2 - 1}{n(n+1)^2(n+2)}\theta^2 \end{aligned}$$

Now we can see that

$$\begin{aligned} MSE[JK(T_n)] &= Var[JK(T_n)] + (bias[JK(T_n)])^2 \\ &= \frac{2n^2 - 1}{n(n+1)^2(n+2)}\theta^2 + \left(-\frac{\theta}{n(n+1)}\right)^2 \\ &= \frac{2n^2 - 1}{n(n+1)^2(n+2)}\theta^2 + \frac{\theta^2}{n^2(n+1)^2} \\ &= \frac{\theta^2}{n^2(n+1)^2} \left(\frac{n(2n^2 - 1)}{(n+2)} + 1\right) \\ &= \frac{\theta^2}{n^2(n+1)^2(n+2)} \left(2n^3 - n + n + 2\right) \\ &= \frac{\theta^2 2(n^3 + 1)}{n^2(n+1)^2(n+2)} \end{aligned}$$

Then we can also find $MSE(T_n)$ using the result from above

$$\begin{aligned}
MSE[T_n] &= Var[T_n] + (bias[T_n])^2 \\
&= Var[X_{(n)}] + (bias[X_{(n)}])^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)} + (E[X_{(n)}] - \theta)^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{n}{n+1}\theta - \theta\right)^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{n - n - 1}{n+1}\right)^2 \theta^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2} \theta^2 \\
&= \left(n + n + 2\right) \frac{\theta^2}{(n+1)^2(n+2)} \\
&= \frac{2(n+1)\theta^2}{(n+1)^2(n+2)} \\
&= \frac{2(n^3 + n^2)\theta^2}{n^2(n+1)^2(n+2)}
\end{aligned}$$

Therefore for any sample size $n > 1$:

$$\begin{aligned}
1 &< n^2 \\
n^3 + 1 &< n^3 + n^2 \\
2(n^3 + 1)\theta^2 &< 2(n^3 + n^2)\theta^2 \\
\frac{2(n^3 + 1)\theta^2}{n^2(n+1)^2(n+2)} &< \frac{2(n^3 + n^2)\theta^2}{n^2(n+1)^2(n+2)} \\
MSE[JK(T_n)] &< MSE[T_n]
\end{aligned}$$

Hence, jackknifing has also reduced the MSE of the estimator T_n .

- b) It is given that the bootstrap-bias adjustment of T_n happens via $2T_n - E_{\hat{F}}[T_n^*]$ where the expectation is calculated under the empirical distribution.

In our case, we have $n = 3$, this distribution puts equal weights of $1/3$ at $X_{(1)}, X_{(2)}$ and $X_{(3)}$.

Hence, to calculate $E_{\hat{F}}[T_3^*]$ we need to consider all the possible outcome of resampling from the original set $\{X_{(1)}, X_{(2)}, X_{(3)}\}$:

$$T_3^* = \begin{cases} X_{(1)} = \max\{X_{(1)}, X_{(1)}, X_{(1)}\} & \text{with probability } \left(\frac{1}{3}\right)^3 \\ \\ X_{(2)} = \begin{cases} \max\{X_{(1)}, X_{(1)}, X_{(2)}\} \\ \max\{X_{(1)}, X_{(2)}, X_{(1)}\} \\ \dots\dots\dots \\ \max\{X_{(2)}, X_{(1)}, X_{(2)}\} \\ \max\{X_{(2)}, X_{(2)}, X_{(2)}\} \end{cases} & \text{with probability } \left(\frac{1}{3}\right)^3 \\ \\ X_{(3)} = \begin{cases} \max\{X_{(1)}, X_{(1)}, X_{(3)}\} \\ \max\{X_{(1)}, X_{(2)}, X_{(3)}\} \\ \dots\dots\dots \\ \max\{X_{(2)}, X_{(2)}, X_{(3)}\} \\ \max\{X_{(2)}, X_{(3)}, X_{(3)}\} \\ \max\{X_{(3)}, X_{(3)}, X_{(3)}\} \end{cases} & \text{with probability } \left(\frac{1}{3}\right)^3 \end{cases}$$

Therefore, we can split it to 3 cases according to the above distribution

Case 1: $X_{(1)}$

Since the only way to obtain $X_{(1)}$ is to have all $X_{(1)}$ in the resampled set. Therefore there's only 1 possible set that will give us $X_{(1)}$.

Case 2: $X_{(2)}$

Any resample sets that contains $X_{(1)}$'s and $X_{(2)}$'s but not $X_{(3)}$ will give us $X_{(2)}$. Therefore, there are 2^3 ways to rearrange $X_{(1)}$ and $X_{(2)}$ in the sets of 3 with replacement. However, we need to exclude the case $\{X_{(1)}, X_{(1)}, X_{(1)}\}$ since it will not give us $X_{(2)}$. Hence, there will be in total $(2^3 - 1)$ possible sets that will give us $X_{(2)}$.

Case 3: $X_{(3)}$

Any resample sets that contains $X_{(1)}$'s and $X_{(2)}$'s and $X_{(3)}$ will give us $X_{(3)}$. Therefore, there are 3^3 ways to rearrange $X_{(1)}$, $X_{(2)}$ and $X_{(3)}$ in the sets of 3 with replacement. However, we need to exclude the cases where $X_{(3)}$ is not included in other words those that will give us $X_{(2)}$ or $X_{(1)}$ as a result. Hence, there will be in total $(3^3 - 2^3 - 1 + 1)$ possible sets that will give us $X_{(2)}$.

Finally we can compute

$$\begin{aligned} E_{\hat{F}}[T_3^*] &= (1) \left(\frac{1}{3}\right)^3 X_{(1)} + (2^3 - 1) \left(\frac{1}{3}\right)^3 X_{(2)} + (3^3 - 2^3 + 1) \left(\frac{1}{3}\right)^3 X_{(3)} \\ &= \frac{1}{27} X_{(1)} + \frac{7}{27} X_{(2)} + \frac{19}{27} X_{(3)} \end{aligned}$$

Hence the bootstrap bias-adjusted estimator of θ is

$$\begin{aligned} 2T_3 - E_{\hat{F}}[T_3^*] &= 2X_{(3)} - E_{\hat{F}}[T_3^*] \\ &= 2X_{(3)} - \left(\frac{1}{27} X_{(1)} + \frac{7}{27} X_{(2)} + \frac{19}{27} X_{(3)}\right) \\ &= \left(2 - \frac{19}{27}\right) X_{(3)} - \frac{7}{27} X_{(2)} - \frac{1}{27} X_{(1)} \\ &\approx 1.2963 X_{(3)} - 0.25926 X_{(2)} - 0.037 X_{(1)} \end{aligned}$$

Hence we can calculate the bias of the bootstrap bias-adjusted estimator of θ is then

$$\begin{aligned} \text{bias}(\hat{\theta}_{\text{bootstrap bias-adjusted}}) &= E[2T_3 - E_{\hat{F}}[T_3^*]] - \theta \\ &\approx E[1.2963 X_{(3)} - 0.25926 X_{(2)} - 0.037 X_{(1)}] - \theta \\ &\approx 1.2963 E[X_{(3)}] - 0.25926 E[X_{(2)}] - 0.037 E[X_{(1)}] - \theta \end{aligned}$$

From 5 a i), we have derived $E[X_{(k)}] = \frac{k}{n+1} \theta$ and $n = 3$ in our case, therefore we know that

$$E[X_{(1)}] = \frac{1}{4} \theta, E[X_{(2)}] = \frac{2}{4} \theta, E[X_{(3)}] = \frac{3}{4} \theta$$

Then substitute the above expected value to obtain the magnitude of the bias

$$\begin{aligned} |\text{bias}(\hat{\theta}_{\text{bootstrap bias-adjusted}})| &\approx \left| 1.2963 \frac{3}{4} \theta - 0.25926 \frac{2}{4} \theta - 0.037 \frac{1}{4} \theta - \theta \right| \\ &\approx |(-0.166655) \theta| = (0.166655) \theta \end{aligned}$$

Comparing to the magnitude of the bias of the original estimator T_3

$$\begin{aligned} |\text{bias}(\hat{\theta}_{\text{original}})| &= |E(X_{(3)}) - \theta| = \left| \theta \left[\frac{3}{4} - 1 \right] \right| \\ &= |(-0.25) \theta| = (0.25) \theta \\ &> (0.166655) \theta = |\text{bias}(\hat{\theta}_{\text{bootstrap bias-adjusted}})| \end{aligned}$$

Hence the bootstrap-based bias adjustment has reduced the magnitude of the bias of the original estimator T_3 .