MATH3911 Assignment Two

Simon Cheng, z3333348

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

Question 1

a) We have that $X_1, ..., X_{12}$ are i.i.d. $N(\mu, 4)$ random variables with common density given by

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

The likelihood function is then given by

$$\begin{aligned}
L(x;\mu) &= \prod_{i=1}^{12} f(x_i;\mu) \\
&= \left(\frac{1}{2\sqrt{2\pi}}\right)^{12} e^{-\frac{1}{8}\sum_{i=1}^{12}(x_i-\mu)^2} \\
&= \left(\frac{1}{2\sqrt{2\pi}}\right)^{12} e^{-\frac{1}{8}(\sum_{i=1}^{12} x_i^2 - 2\mu\sum_{i=1}^{12} x_i + 12\mu^2)} \\
&= \left(\frac{1}{2\sqrt{2\pi}}\right)^{12} \left(e^{-\frac{1}{8}\sum_{i=1}^{12} x_i^2}\right)\left(e^{\frac{1}{4}\mu\sum_{i=1}^{12} x_i}\right)\left(e^{-\frac{3}{2}\mu^2}\right).
\end{aligned}$$

For any fixed μ_1 and μ_2, with $\mu_1 < \mu_2$, the likelihood ratio is

$$\begin{aligned}
\frac{L(x;\mu_2)}{L(x;\mu_1)} &= \frac{\left(\frac{1}{2\sqrt{2\pi}}\right)^{12}\left(e^{-\frac{1}{8}\sum_{i=1}^{12} x_i^2}\right)\left(e^{\frac{1}{4}\mu_2\sum_{i=1}^{12} x_i}\right)\left(e^{-\frac{3}{2}\mu_2^2}\right)}{\left(\frac{1}{2\sqrt{2\pi}}\right)^{12}\left(e^{-\frac{1}{8}\sum_{i=1}^{12} x_i^2}\right)\left(e^{\frac{1}{4}\mu_1\sum_{i=1}^{12} x_i}\right)\left(e^{-\frac{3}{2}\mu_1^2}\right)} \\
&= \left(e^{-\frac{3}{2}(\mu_2^2-\mu_1^2)}\right)\left(e^{\frac{1}{4}(\mu_2-\mu_1)\sum_{i=1}^{12} x_i}\right).
\end{aligned}$$

which is a non-decreasing function of $\sum_{i=1}^{12} x_i$ since $\mu_1 < \mu_2$. Therefore, the joint density of $X_1, X_2, ..., X_{12}$ has monotone likelihood ratio in $T(X) = \sum_{i=1}^{12} x_i$.

b) Firstly, we can rewrite the likelihood ratio as

$$\begin{aligned}
L(x;\mu) &= \left(\frac{1}{2\sqrt{2\pi}}\right)^{12} \left(e^{-\frac{1}{8}\sum_{i=1}^{12} x_i^2}\right)\left(e^{\frac{1}{4}\mu\sum_{i=1}^{12} x_i}\right)\left(e^{-\frac{3}{2}\mu^2}\right) \\
&= \left(\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}\mu^2}\right)^{12}\left(\prod_{i=1}^{12} e^{-\frac{1}{8}x_i^2}\right)\left(e^{\frac{1}{4}\mu\sum_{i=1}^{12} x_i}\right).
\end{aligned}$$

We can easily see that this is in the form

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

with $T(X) = \sum_{i=1}^{12} d(X_i) = \sum_{i=1}^{12} X_i$. So by Theorem 2, the UMP unbiased size $\alpha = 0.05$ test for $H_0 : \mu = 1$ versus $H_1 : \mu \neq 1$ is given by

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

where c_1, c_2 satisfy the conditions

$$\begin{aligned} E_{\mu=1}[\varphi^*(X)] &= 0.05 && (1) \\ \left[\frac{\partial}{\partial\mu} E_\mu[\varphi^*(X)]\right]_{\mu=1} &= 0 && (2) \end{aligned}$$

We first note that

$$T(X) = \sum_{i=1}^{12} X_i \sim N(12\mu, 48)$$

and that

$$\frac{T(X) - 12\mu}{2\sqrt{12}} \sim N(0, 1)$$

We can therefore deduce that the power function is given by

$$\begin{aligned} E_\mu[\varphi^*(X)] &= Pr[\varphi^*(X) = 1] \\ &= Pr[T(X) < c_1] + Pr[T(X) > c_2] \\ &= Pr\left[\frac{T(X) - 12\mu}{2\sqrt{12}} < \frac{c_1 - 12\mu}{2\sqrt{12}}\right] + Pr\left[\frac{T(X) - 12\mu}{2\sqrt{12}} > \frac{c_2 - 12\mu}{2\sqrt{12}}\right] \\ &= \Phi\left(\frac{c_1 - 12\mu}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{c_2 - 12\mu}{2\sqrt{12}}\right). \end{aligned}$$

As a result, we have

$$\frac{\partial}{\partial\mu} E_\mu[\varphi^*(X)] = -\frac{\sqrt{12}}{2}\phi\left(\frac{c_1 - 12\mu}{2\sqrt{12}}\right) + \frac{\sqrt{12}}{2}\phi\left(\frac{c_2 - 12\mu}{2\sqrt{12}}\right)$$

where

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

is the standard normal p.d.f. Now, from equation (2), we have

$$\begin{aligned} \left[\frac{\partial}{\partial\mu} E_\mu[\varphi^*(X)]\right]_{\mu=1} &= 0 \\ -\frac{\sqrt{12}}{2}\phi\left(\frac{c_1 - 12}{2\sqrt{12}}\right) + \frac{\sqrt{12}}{2}\phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0 \\ \phi\left(\frac{c_1 - 12}{2\sqrt{12}}\right) &= \phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right). \end{aligned}$$

Since $c_1 \neq c_2$ and the standard normal density is symmetric around zero, we must have

$$\begin{aligned} \frac{c_1 - 12}{2\sqrt{12}} &= -\frac{c_2 - 12}{2\sqrt{12}} \\ c_1 &= 24 - c_2. \end{aligned}$$

Now, substituting this into equation (1) and using the fact that $\Phi(-z) = -\Phi(z)$, we have

$$\begin{aligned}
E_{\mu=1}[\varphi^*(X)] &= 0.05 \\
\Phi\left(\frac{c_1 - 12}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0.05 \\
\Phi\left(\frac{-(c_2 - 12)}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0.05 \\
1 - \Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0.05 \\
1 - \Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0.025 \\
\Phi\left(\frac{c_2 - 12}{2\sqrt{12}}\right) &= 0.975 \\
\frac{c_2 - 12}{2\sqrt{12}} &= \Phi^{-1}(0.975) \approx 1.96
\end{aligned}$$

Therefore the constants c_1 and c_2 are given by

$$\begin{aligned}
c_2 &= 12 + 2\sqrt{12}\Phi^{-1}(0.975) \approx 25.579 \\
c_1 &= 24 - c_1 \approx -1.579
\end{aligned}$$

Finally, the UMP unbiased size $\alpha = 0.05$ test is

$$\varphi^*(X) = \begin{cases} 1, & T(X) < -1.579, T(X) > 25.579 \\ 0, & -1.579 \leq T(X) \leq 25.579 \end{cases}$$

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

$$E_\mu[\varphi^*(X)] = \Phi\left(\frac{-1.579 - 12\mu}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{25.579 - 12\mu}{2\sqrt{12}}\right)$$

Evaluating the power numerically gives the following

$$\begin{aligned}
E_\mu[\varphi^*(X)]_{\mu=0} &= \Phi\left(\frac{-1.579}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{25.579}{2\sqrt{12}}\right) \approx 0.4100 \\
E_\mu[\varphi^*(X)]_{\mu=0.5} &= \Phi\left(\frac{-7.579}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{19.579}{2\sqrt{12}}\right) \approx 0.1393 \\
E_\mu[\varphi^*(X)]_{\mu=1.5} &= \Phi\left(\frac{-19.579}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{7.579}{2\sqrt{12}}\right) \approx 0.1393 \\
E_\mu[\varphi^*(X)]_{\mu=2.5} &= \Phi\left(\frac{-31.579}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{-4.421}{2\sqrt{12}}\right) \approx 0.7383 \\
E_\mu[\varphi^*(X)]_{\mu=4} &= \Phi\left(\frac{-49.579}{2\sqrt{12}}\right) + 1 - \Phi\left(\frac{-22.421}{2\sqrt{12}}\right) \approx 0.9994
\end{aligned}$$

d) The density of the third order statistic $X_{(3)}$ is given by

$$f_{X_{(3)}}(x;\mu) = \frac{12!}{2!9!} f(x;\mu)[F(x;\mu)]^2[1 - F(x;\mu)]^9$$

We know that

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

We now use the fact that $\frac{X_i-\mu}{2} \sim N(0,1)$ for all $i = 1, 2, ..., 12$, to work out

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

Therefore, the density of the third order statistic $X_{(3)}$ is

$$f_{X_{(3)}}(x;\mu) = \left(\frac{330}{\sqrt{2\pi}} e^{-\frac{1}{8}(x-\mu)^2}\right)\left[\Phi\left(\frac{x-\mu}{2}\right)\right]^2 \left[1 - \Phi\left(\frac{x-\mu}{2}\right)\right]^9$$

and under H_0, it is

$$f_{X_{(3)}}(x;1) = \left(\frac{330}{\sqrt{2\pi}} e^{-\frac{1}{8}(x-1)^2}\right)\left[\Phi\left(\frac{x-1}{2}\right)\right]^2 \left[1 - \Phi\left(\frac{x-1}{2}\right)\right]^9.$$

Question 2

a) Firstly, we note that $N_1, ..., N_T$ are i.i.d. random variables with common conditional density being Poisson(λ) so

$$f_{N|\Lambda}(n|\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

The prior distribution of the random parameter Λ is given as

$$\tau(\lambda) = \frac{\lambda^{\alpha-1}e^{-\lambda/b}}{\Gamma(a)b^a}.$$

Then the conditional joint density of $\mathbf{N}$ is

$$\begin{aligned} f_{\mathbf{N}|\Lambda}(\mathbf{n}|\lambda) &= \prod_{i=1}^{T} f_{N|\Lambda}(n|\lambda) \\ &= e^{-\lambda T}\frac{\lambda^{\sum_{i=1}^T n_i}}{\prod_{i=1}^T n_i}. \end{aligned}$$

Combining these two results, we have

$$\begin{aligned} f_{\mathbf{N}|\Lambda}(\mathbf{n}|\lambda)\tau(\lambda) &= \frac{\lambda^{\sum_{i=1}^T n_i+\alpha-1}e^{-\lambda(T+1/b)}}{\Gamma(a)b^a\prod_{i=1}^T n_i} \\ &\propto \lambda^{\sum_{i=1}^T n_i+\alpha-1}e^{-\lambda(T+1/b)} \end{aligned}$$

This means $h(\lambda|\mathbf{N}) \propto \lambda^{\sum_{i=1}^T n_i+\alpha-1}e^{-\lambda(T+1/b)}$ which implies that

$$\lambda|\mathbf{N} \sim \text{Gamma}\left(\sum_{i=1}^T n_i + \alpha, \frac{1}{T+1/b}\right)$$

Now, by using the expected value property of the Gamma function, the Bayes estimator of λ with respect to quadratic loss is given by

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

b) Assuming that $a = 2, b = 2, T = 6$ and using the fact that $\sum_{i=1}^{6} n_i = 12$, we have from part a) that the posterior with respect to λ is

$$h(\lambda|\mathbf{N}) = \frac{\lambda^{13} e^{-\frac{13}{2}\lambda}}{\Gamma(14)(\frac{2}{13})^{14}}.$$

We wish to test the hypothesis $H_0 : \lambda \leq 2$ against $H_1 : \lambda > 2$ using the a Baynesian test with a zero-one loss. Thus, we calculate the posterior probability given the sample as

$$P(\lambda \in \Lambda_0|\mathbf{N}) = \int_0^2 \frac{\lambda^{13} e^{-\frac{13}{2}\lambda}}{\Gamma(14)(\frac{2}{13})^{14}} d\lambda = 0.4269554$$

using the R function

```
> pgamma(2, 14, 13/2)
```

Consequently, we reject H_0 since the posterior probability is less than $\frac{1}{2}$.

Question 3

```
# Question 3
# Part a)

# Function to estimate skewness

myskewness <- function (x) {
n <- length (x)
resid <- x - mean (x)

numer <- sqrt (n) * sum (resid^3)
denom <- (sum (resid^2))^(3/2)
skew <- numer / denom
return (skew)
}

# Function to estimate kurtosis

mykurtosis <- function (x) {
n <- length (x)
resid <- x - mean (x)

numer <- n * sum (resid^4)
denom <- (sum (resid^2))^2
```

```
kurt <- numer / denom - 3
return (kurt)
}

# part b)
# estimate skewness and kurtosis for the vector time in the 'galaxies'
# data set using the functions in a) and built in S-PLUS functions

myskewness (galaxies)
skewness (galaxies, method="moment")
skewness (galaxies)
mykurtosis (galaxies)
kurtosis (galaxies, method="moment")
kurtosis (galaxies)

# part c)
# bootstrap the estimators of skewness and kurtosis from part a)

boot.objs <- bootstrap (galaxies, myskewness (galaxies), B=5000)
boot.objk <- bootstrap (galaxies, mykurtosis (galaxies), B=5000)
summary (boot.objs)
summary (boot.objk)

# part d)
# produce a normal quantile-quantile plot for the vector 'velocities'

qqnorm (time, ylab="Velocities", main="Quantile-Quantile Plot for the Galaxies
Data")
```

```
> myskewness(galaxies)
[1] -0.4338282
> skewness(galaxies, method = "moment")
[1] -0.4338282
> skewness(galaxies)
[1] -0.4419541
> mykurtosis(galaxies)
[1] 2.271297
> kurtosis(galaxies, method = "moment")
[1] 2.271297
> kurtosis(galaxies)
[1] 2.493027
```

b) We can see from the output that the functions "myskewness" and "mykurtosis" give slightly different values to the predefined functions "skewness" and "kurtosis" with default settings. However, we can see by changing the method of the predefined functions to "method = moment", we have the same values.

This can be explained by the fact that our functions calculate the moment estimates whereas the default settings calculate Fisher's g1 and g2 estimates.

c) The output is shown below:

```
> summary(boot.objs)
Call:
bootstrap(data = galaxies, statistic = myskewness(galaxies), B = 5000)

Number of Replications: 5000

Summary Statistics:
      Observed    Mean     Bias     SE
Param  -0.4338 -0.4663 -0.03243 0.4582

Percentiles:
        2.5%    5%    95%  97.5%
Param -1.479 -1.26 0.2254 0.3861

BCa Confidence Intervals:
        2.5%     5%    95%  97.5%
Param -1.253 -1.095 0.4056 0.7491

> summary(boot.objk)
Call:
bootstrap(data = galaxies, statistic = mykurtosis(galaxies), B = 5000)

Number of Replications: 5000

Summary Statistics:
      Observed  Mean    Bias    SE
Param    2.271 2.345 0.07398 1.021

Percentiles:
        2.5%     5%   95% 97.5%
Param 0.6919 0.9015 4.204 4.707

BCa Confidence Intervals:
        2.5%     5%   95% 97.5%
Param 0.6801 0.9095 4.235 4.662
```

From the output, the 95% BCa confidence intervals for the skewness and kurtosis are (-1.253,0.7491) and (0.6801,4.662) respectively.

d) We see that 0 is in the confidence interval for skewness but not kurtosis so we have reason to suspect that the normality of the galaxies data is in doubt. The normal quantile-quantile plot is not close to a straight line, which confirms the fact that the galaxies data may not be normally distributed. We can also see that the skewness quantile-quantile plot seems to be approximately a straight line wheres the kurtosis is clearly not a straight line, again confirming the results in the confidence intervals.

Question 4

```
# Question
# Part i)

set.seed (round (log (3333348)))

# generate observations from the 'idea' distribution
x70 <- runif (70, 0.5, 4)
e70 <- rnorm (70, mean=0, sd=0.2)
y70 <- 2 + 0.8 * x70 + e70

# generate observations from the 'contaminating' distribution (outliers)
x30 <- rnorm(30, mean=5, sd=0.5)
y30 <- rnorm(30, mean=2, sd=0.5)

x <- c (x70, x30)
y <- c (y70, y30)
simuldata <- data.frame (x, y)

# plot the observations, with the least squares, default M-estimate and
default LTS regression lines.
plot (x, y, main="Simulated Data with Regression Line 1")
abline (lm (y~x, simuldata))
text (5.0, 3.9, "LS")

abline (rreg (x,y))
text (6.3, 3.6, "M")

abline (ltsreg.formula (y~x, simuldata))
text (4.0, 3.6, "LTS")

# redraw the plot with the new LTS regression line
plot (x, y, main="Simulated Data with Regression Line 2")
abline (lm (y~x, simuldata))
text (5.0, 3.9, "LS")

abline (rreg (x,y))
text (6.3, 3.6, "M")

abline (ltsreg.formula (y~x, simuldata, quan=70))
text (4.0, 3.6, "LTS")

# find the slope and intercept of the old and new LTS regression lines
summary (ltsreg.formula (y~x, simuldata))
summary (ltsreg.formula (y~x, simuldata, quan=70))
```

The partial S-plus output is shown below:

```
> summary(ltsreg.formula(y ~ x, simuldata))
Method:
[1] "Least Trimmed Squares Robust Regression."

Call:
ltsreg.formula(formula = y ~ x, data = simuldata)

Coefficients:
 Intercept       x
  3.7733   -0.1898

Scale estimate of residuals: 1.179

Robust Multiple R-Squared: 0.0505

Total number of observations:  100

Number of observations that determine the LTS estimate:  90

> summary(ltsreg.formula(y ~ x, simuldata, quan = 70))
Method:
[1] "Least Trimmed Squares Robust Regression."

Call:
ltsreg.formula(formula = y ~ x, data = simuldata, quan = 70)

Coefficients:
 Intercept      x
 2.0467    0.7860

Scale estimate of residuals: 0.2159

Robust Multiple R-Squared: 0.926

Total number of observations:  100

Number of observations that determine the LTS estimate:  70
```

iii) In the second LTS regression line, we have an intercept and slope of 2.0467 and 0.7860, which is pretty close to the true values of 2 and 0.8. This is reflected by the high robust multiple R-squared of 92.6%, which suggests that the fitted regression line explains most of the variation in the response.

In the first LTS regression line, we have an intercept and slope of 3.7733 and -0.1898, which is quite far off from the true values of 2 and 0.8. This is reflected by the low robust multiple R-squared of 5.05%, which suggests that the fitted regression line explains little of the variation in the response.

This contrast can be explained by the robustness of the LTS method to outliers. In standard least squares (LS) regression, the regression line is fitted by minimising the sum of squared residuals. That is, by minimising the objective function

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

where b_0 and b_1 are the intercept and slope of the fitted line. We can see that this method is not robust to outliers as large residuals are reflected by an increase in the function S.

On the other hand, in LTS, the regression line is fitted by minimising the sum of squares of the m smallest residuals, where $m < n$. That is, by minimising the objective function

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

This method is more 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.

Question 5

We know that $X_1, ..., X_n$ are i.i.d. uniform $[0, \theta)$ random variables, with common density function

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

and common cumulative distribution function

$$F(x; \theta) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \le x < \theta \\ 1, & x \ge \theta \end{cases}$$

Before we proceed to solving the problem, we will highlight a useful result, which will make the computation much simpler. We know that if a random variable X is a Beta distribution with parameters α and β, the expected value and variance is given by

$$E[X] = \frac{\alpha}{\alpha + \beta} \text{ and } Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Now, for the special case where F_x is a unifrom distribution on $[0, \theta)$, we have

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

Clearly, $\theta f_{X_{(r)}}(u\theta)$ gives the density of Beta distribution with parameters r and n-r+1. If we let $B \sim Beta(r, n-r+1)$ and use the substitution $u = \frac{y}{\theta}$, we have

$$E[X_{(r)}] = \int_0^\theta y f_{X_{(r)}}(y)dy = \theta \int_0^1 u(\theta f_{X_{(r)}}(u\theta))du = \theta \int_0^1 u f_B(u)du = \theta E[B] = \frac{r}{n+1}\theta$$

Similarly, we have

$$E[X_{(r)}^2] = \int_0^\theta y^2 f_{X_{(r)}}(y)dy = \theta^2 \int_0^1 u^2(\theta f_{X_{(r)}}(u\theta))du = \theta^2 \int_0^1 u^2 f_B(u)du = \theta^2 E[B^2]$$

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

Our expected value and varaince is therefore given by

$$E[X_{(r)}] = \frac{r}{n+1}\theta \text{ and } Var[X_{(r)}] = \frac{r(n-r+1)}{(n+1)^2(n+2)}\theta^2 \tag{3}$$

a) Consider the estimator $T_n = X_{(n)} = \max\{X_1, ..., X_n\}$ of θ. The jackknife estimator is

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

where $T_n^{(i)} = \max\{X_1, ..., X_{i-1}, X_{i+1}, ..., X_n\}$. We have two cases to consider:

(i) If $i = n$, then $T_n^{(n)} = \max\{X_1, ..., X_{n-1}\} = X_{n-1}$ since X_n is excluded and the largest is now X_{n-1}.

(ii) If $i \neq n$, then $T_n^{(i)} = \max\{X_1, ..., X_{i-1}, X_{i+1}, ..., X_n\} = X_n$ since X_n would still be the largest of the whole sample. Putting these two results together we have

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

Now to compute the expectation of the jackknife estimator, we use the first result in (3) to show that

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

Finally, we compute the bias of the jackknife estimator as

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

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

b) Firstly, we make a note that

$$MSE_\theta[JK(T_n)] = Var_\theta[JK(T_n)] + [bias(JK(T_n))]^2.$$

Now, to compute $Var_\theta[JK(T_n)]$, we require $Cov_\theta[X_{(n-1)}, X_n]$, $Var_\theta[X_{(n-1)}]$ and $Var_\theta[X_{(n)}]$. We will now consider the joint density of $X_{(n-1)}$ and $X_{(n)}$ which is given by:

$$\begin{aligned}
f_{X_{(n-1)},X_{(n)}}(u,v;\theta) &= \begin{cases} \frac{n!}{(n-2)!0!0!} f(u;\theta)f(v;\theta)[F(u;\theta)]^{n-2}, & u<v \\ 0, & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{n(n-1)}{\theta^n} u^{n-2}, & u<v \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

Then we have

$$\begin{aligned}
E_\theta[X_{(n-1)}, X_{(n)}] &= \int_0^\theta \int_0^v uv f_{X_{(n-1)},X_{(n)}}(u,v;\theta) du\, dv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v \int_0^v n\, u^{n-1} du\, dv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v[u^n]_{u=0}^v\, dv \\
&= \frac{n-1}{\theta^n} \int_0^\theta v^{n+1}\, dv \\
&= \frac{n-1}{\theta^n} \left[\frac{v^{n+2}}{n+2}\right]_{v=0}^\theta \\
&= \frac{n-1}{n+2}\theta^2
\end{aligned}$$

Thus, we can now compute the required covariance:

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

Now, by again using the result we proved in (3), we have that

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

Furthermore, we know from part a) that

$$JK(T_n) = \frac{2n-1}{n}X_{(n)} - \frac{n-1}{n}X_{(n-1)}$$

so we can calculate its variance by

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

As a result, the mean squared error of the jackknife estimator is given by

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

The mean squared error of the original estimator is given by

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

Clearly for any sample size $n > 1$, we have

$$\begin{aligned}
1 &< n^2 \\
\frac{2\theta^2(n^3+1)}{n^2(n+1)^2(n+2)} &< \frac{2\theta^2(n^3+n^2)}{n^2(n+1)(n+2)}.
\end{aligned}$$

which implies that

$$MSE_\theta[JK(T_n)] < MSE_\theta[T_n].$$

c) We know that the jackknife estimator, $JK(T_n)$ of θ is given by

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

Now, if we assume that T_n and $T_n^{(i)}$ are "regular estimators", then they can be expressed as

$$bias(T_n) = \frac{A}{n} + \frac{B}{n^2} + o(\frac{1}{n^2}) \text{ or } E(T_n) = \theta + \frac{A}{n} + \frac{B}{n^2} + o(\frac{1}{n^2})$$

$$bias(T_n^{(i)}) = \frac{A}{n-1} + \frac{B}{(n-1)^2} + o\left(\frac{1}{n^2}\right) \text{ or } E(T_n^{(i)}) = \theta + \frac{A}{n-1} + \frac{B}{(n-1)^2} + o\left(\frac{1}{n^2}\right)$$

Substituting these expressions in the expected value of the jackknife estimator gives

$$\begin{aligned}
E[JK(T_n)] &= nE[T_n] - \frac{n-1}{n}\sum_{i=1}^{n} E[T_n^{(i)}] \\
&= n\left(\theta + \frac{A}{n} + \frac{B}{n^2}\right) - \frac{n-1}{n}\sum_{i=1}^{n}\left(\theta + \frac{A}{n-1} + \frac{B}{(n-1)^2}\right) + o\left(\frac{1}{n^2}\right) \\
&= n\theta + A + \frac{B}{n} - \frac{n-1}{n}n\left(\theta + \frac{A}{n-1} + \frac{B}{(n-1)^2}\right) + o\left(\frac{1}{n^2}\right) \\
&= \theta + \frac{B}{n} - \frac{B}{n-1} + o\left(\frac{1}{n^2}\right) \\
&= \theta - \frac{B}{n-1} + o\left(\frac{1}{n^2}\right).
\end{aligned}$$

It then follows that

$$bias(JK(T_n)) = -\frac{B}{n-1} + o\left(\frac{1}{n^2}\right)$$

from which it is clear that it does not involve the $\frac{A}{n}$ term. Hence, in general, jackknifing reduces the magnitude of the bias.