STAT 323 Assignment 2

Instructor: Claudia Marie Mahler
Name: Yifeng Pan
UCID: 30063828

Summer 2019

1 Let $X_1, X_2, \ldots, X_{20}$ repensent a random sample where each X_i is a normally distributed with a mean $\mu = 100$ and unknown variance σ^2. From what you know about the distribution of $\frac{(n-1)s^2}{\sigma^2}$, find an interval that would contain σ^2 95% of the time.

$\mu = 100, n = 20, df = 19$

$$\begin{aligned}
0.95 &= P(a \leq \sigma^2 \leq b) \\
&= P(1/a \geq 1/\sigma^2 \geq 1/b) \text{ they are all positive.} \\
&= P((n-1)s^2/a \geq (n-1)s^2/\sigma^2 \geq (n-1)s^2/b) \\
&= P(19s^2/a \geq \chi^2_{19} \geq 19s^2/b) \\
&= P(19s^2/b \leq \chi^2_{19} \leq 19s^2/a) \\
19s^2/b &= qchisq(0.025, 19) \\
19s^2/a &= qchisq(0.95 + 0.025, 19) \\
a &= \frac{19s^2}{qchisq(0.975, 19)} \\
b &= \frac{19s^2}{qchisq(0.025, 19)}
\end{aligned}$$

The 95% interval for σ^2 would be $(\frac{19s^2}{qchisq(0.975,19)}, \frac{19s^2}{qchisq(0.025,19)})$.

2 Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a population whose density function is given by:

$$f(x) = \begin{cases} 3\beta^3 x^{-4} & \textbf{for } \beta \leq x \\ 0 & \textbf{elsewhere} \end{cases}$$

where $\beta > 0$ is unknown. Consider the estimator $\hat{\beta} = \min(X_1, X_2, \ldots, X_n)$. Derive the bias of the estimator $\hat{\beta}$.

$F_X(x) = \int_\beta^x 3\beta^3 x^{-4} dx = (1 - \frac{\beta^3}{x^3})$. So $\hat{\beta} = n[1 - F_X(x)]^{n-1} f_X(x) = n[1 - (1 - \frac{\beta^3}{x^3})]^{n-1} 3\beta^3 x^{-4} = n(\frac{\beta^3}{x^3})^{n-1} 3\beta^3 x^{-4} = 3n\frac{\beta^{3n-3}}{x^{3n-3}}\beta^3 x^{-4} = 3n\frac{\beta^{3n}}{x^{3n+1}}$.

$$\begin{aligned}
E(\hat{\beta}) &= E(3n\frac{\beta^{3n}}{x^{3n+1}}) \\
&= \int_\beta^\infty 3n\frac{\beta^{3n}}{x^{3n+1}} x dx \\
&= 3n\beta^{3n} \int_\beta^\infty \frac{1}{x^{3n}} dx \\
&= 3n\beta^{3n}(-\beta^{-3n+1}/(1-3n))) \\
&= \beta 3n/(3n-1) \\
B(\hat{\beta}) &= E(\hat{\beta}) - \beta \\
&= \beta 3n/(3n-1) - \beta \\
&= \beta 3n/(3n-1) - (3n\beta - \beta)/(3n-1) \\
&= \frac{\beta}{3n-1}
\end{aligned}$$

3 Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population that can be modeled by the following probability model:

$$f_X(x) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha}, 0 < x < \theta, \alpha > 0$$

a Find the probability density function of $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$.

$$\begin{aligned} F_X(x) &= \int_0^x \frac{\alpha x^{\alpha-1}}{\theta^\alpha} dx \\ &= \frac{x^\alpha}{\theta^\alpha} \end{aligned}$$

$$\begin{aligned} X_{(n)} &= n(F(x))^{n-1} f(x) \\ &= n(\frac{x^\alpha}{\theta^\alpha})^{n-1} \frac{\alpha x^{\alpha-1}}{\theta^\alpha} \\ &= an\frac{x^{\alpha n-1}}{\theta^{\alpha n}} \text{ for } 0 < x < \theta, \alpha > 0, n \in \mathbb{Z}^+ \end{aligned}$$

b Is $X_{(n)}$ an unbiased estimator for θ? If not, suggest a function of $X_{(n)}$ that is an unbiased estimator of θ.

$$\begin{aligned} E(X_{(n)}) &= \int_0^\theta an\frac{x^{\alpha n}}{\theta^{\alpha n}} dx \\ &= \frac{\alpha\theta n}{\alpha n + 1} \end{aligned}$$

$$\begin{aligned} B(X_{(n)}) &= \frac{\alpha\theta n}{\alpha n + 1} - \theta \\ &= \frac{\alpha\theta n}{\alpha n + 1} - \frac{\alpha\theta n + \theta}{\alpha n + 1} \\ &= \frac{\theta}{\alpha n + 1} \text{ which is biased.} \end{aligned}$$

An unbiased estimator of θ would be : $\frac{X_{(n)}(\alpha n+1)}{\alpha n}$. Proof:

$$\begin{aligned} B(\frac{X_{(n)}(\alpha n + 1)}{\alpha n}) &= E(\frac{X_{(n)}(\alpha n + 1)}{\alpha n}) - \theta \\ &= \frac{\alpha n + 1}{\alpha n} E(X_{(n)}) - \theta \\ &= \frac{\alpha n + 1}{\alpha n} \frac{\alpha\theta n}{\alpha n + 1} - \theta \\ &= \theta - \theta \\ &= 0 \text{ which is unbiased.} \end{aligned}$$

4 Let $X_1, X_2, \ldots, X_n$ be a random sample of size n where $X_i \sim Exponential(\beta)$ and Let $Y = \sum_{i=1}^{n} X_i$.

a Show that $Z = \frac{2Y}{\beta}$ is a pivotal quantity.

(1) Z is a function of β. (2) β is the only unknown.

$$\begin{aligned}
m_Z(t) &= E(\exp(\frac{t2Y}{\beta})) \\
&= E(\exp(\frac{t2(\sum_{i=1}^{n} X_i)}{\beta})) \\
&= E(\exp(\frac{t2X}{\beta}))^n \text{ as they are indepentent samples} \\
&= m_X(\frac{2t}{\beta})^n \\
&= (1 - \beta\frac{2t}{\beta})^{(-1)^n} \\
&= (1 - \beta\frac{2t}{\beta})^{-n} \\
&= (1 - 2t)^{-n}
\end{aligned}$$

Which is the MGF of a chi-square distro with $df = 2n$. (3) So the distro of Z does not depend on β.

b Use your finds from a) to construct a 95% confidence interval for β (hint: this will be interms of Y and a chi-square distribution).

$$\begin{aligned}
0.95 &= P(a \leq Z \leq b) \text{ where } a = qchisq(0.025, 2n), b = qchisq(.95 + 0.025, 2n). \\
&= P(a \leq 2Y/\beta \leq b) \\
&= P(1/a \geq \beta/2Y \geq 1/b) \text{ As they are all positive.} \\
&= P(2Y/a \geq \beta \geq 2Y/b)
\end{aligned}$$

The 95% confidence interval for β would be $(\frac{2Y}{qchisq(.95+0.025,2n)}, \frac{2Y}{qchisq(.025,2n)})$.

5 The file NHLSalaries1314.R contains data that resulted from a random sample of $n = 70$ professional hockey players who have contracts with NHL teams. Load the data into R to answer the following questions.

```r
#x = c(3110000,875000, ... , 550000) #The full size 70 vector is omitted.
conf_level = 0.95

library(EnvStats)

mx = mean(x)
sdx = sd(x)
cat("Mean: ", mx, '\n')
cat("SD: ", sdx, '\n')

CImean = t.test(x, conf.level = conf_level)$conf.int
cat("Confidence Inteval for mu: (", CImean, ")\n")

CIv = varTest(x, conf.level = conf_level)$conf.int
cat("Confidence Inteval for sigma: (", sqrt(CIv), ")\n")
```

Output:

```
Mean:  2141214
SD:  1911294
Confidence Inteval for mu: ( 1685482 2596946 )
Confidence Inteval for sigma: ( 1638777 2293373 )
```

a Find the sample mean and sample standard deviation of the data.

Mean: 2141214
SD: 1911294

b Find a 95% confidence interval for μ, the average salary of an NHL player for the 2013-2014 season. Interpret the meaning of this interval in the context of the data.

95% confidence Inteval for μ: (1685482, 2596946)
This means that we are 95% confident that the true mean for all NHL players for the 2013 to 2014 season is between 1685482 and 2596946.

c Find a 95% confidence interval for σ, the standard deviation of the distribution of NHL player salaries for the 2013-2014 season

95% confidence Inteval for σ: (1638777, 2293373)

6 What is the normal body temperature for healthy humans? A random sample of 130 healthy human body temperatures yielded an average temperature of 98.25 degrees and a standard deviation of 0.73 degrees.

a Find a 99% confidence interval for the average body temperature of healthy people.

$n = 130, \overline{x} = 98.25, s = 0.73$
99% confidence interval:

$$\begin{aligned}\overline{x} \pm t_{a/2,n-1}\frac{s}{\sqrt{n}} &= 98.25 \pm t_{0.005,129}\frac{0.73}{\sqrt{130}} \\ &= 98.25 \pm qt(0.005, 129)\frac{0.73}{\sqrt{130}} \\ &\approx 98.25 \pm 0.1674\end{aligned}$$

b Does the interval you obtained in part a) contain the value 98.6 degrees, the accepted average temperature cited by physicians? What conclusions can you draw?

No, the interval in part a) is too low.
Possible conclusions in order of likelihood:

1. The data collected was wrong, likely didn't account for something.
2. My calculations in a) are wrong.
3. Nothing is wrong, and this is an extreme edge case of an 'random' sample. The $1 - 2 * pt(-((98.6 - 98.25) * sqrt(130)/0.73), 129) \approx 99.99998\%$ confidence interval would contain 98.6 degrees. Altho this is statistical negligeable.
4. The accepted conventions by physicians are wrong.

7 The Environmental Protection Agency (EPA) has collected data on LC50 measurements (chemical concentrations that kill 50% of test animals) for certain chemicals likely to be found in freshwater rivers and lakes. For certain species of fish, the LC50 measurements (in parts per million) for DDT in 12 experiments were as follows: $16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9$. Using this data and assuming that LC50 measurements have an approximately normal distribution, construct a 98% confidence interval for the mean LC50 for DDT.

Using the script from q5. Replacing line 1 with $x = c(16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9)$, and replacing line 2 with $conf\_level = .98$ results in the following output:

```
Mean:  9
SD:  6.424385
Confidence Inteval for mu: ( 3.959158 14.04084 )
Confidence Inteval for sigma: ( 4.285091 12.19355 )
```

98% confidence interval for mean LC50: (3.959158, 14.04084)

8 In a recent survey of $n = 1005$ Canadians between the ages of 18 and 34, the polling company Ipsos found that 723 indicated that they "owe it to their parents to keep them comfortable in their retirement." Find a 99% confidence interval for the proportion of all Canadians aged 18 to 34 who hold the same sentiment.

$n = 1005, \hat{p} = 723/1005$, About a normal distro. 99% confidence interval:

$$
\begin{aligned}
\hat{p} \pm Z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 723/1005 \pm Z_{0.005}\sqrt{\frac{(723/1005)(1-723/1005)}{1005}} \\
&\approx 0.719403 \pm qnorm(0.005)0.01417244 \\
&\approx 0.719403 \pm 0.03650578 \\
&\approx (68.29\%, 75.59\%)
\end{aligned}
$$

9 Owing to the variability of trade-in allowance, the profit per new car sold by an automobile dealer varies from car to car. The profits per sale (in hundreds of dollars) tabulated for the past week were $2.1, 3.0, 1.2, 6.2, 4.5,$ and 5.1. Assume the profits are normally distributed.

Using the script from q5. Replacing line 1 with $x = c(2.1, 3.0, 1.2, 6.2, 4.5, 5.1)$, results in the following output:

```
Mean:  3.683333
SD:  1.905168
Confidence Inteval for mu: ( 1.683982 5.682685 )
Confidence Inteval for sigma: ( 1.189221 4.672643 )
```

a Find a 95% confidence interval for μ, the mean profit of a new car sold by the automobile dealer.

95% confidence Inteval for μ: (1.683982, 5.682685)

b Find a 95% confidence interval for σ, the standard deviation of the profit of a new car sold by the automobile dealer.

95% confidence Inteval for σ: (1.189221, 4.672643)