STAT 8003 Group K:

Homework 8

Lu Li, Andrew Powell, and Shade V. Gabriel

November 13, 2014

1. Let $X = \{x_1, ..., x_N\}$ be a random sample from an exponential distribution with the following density function.

$$f(x|\theta) = \theta e^{-\theta x} \tag{1}$$

The observed data is the following.

$$D = \{1.07, 0.88, 0.66, 0.55, 1.15, 0.65, 3.45, 3.55, 3.51, 0.48\} \tag{2}$$

(a) Find an exact pivot.

Since $X_1, X_2, \ldots, X_n \sim Exp(\theta)$ then $\sum_i^n X_i \sim \Gamma(n, \theta)$;

Since $X \sim \Gamma(\alpha, \beta)$ and for any $c > 0$, $cX \sim \Gamma(\alpha, \frac{\beta}{c})$ then

$$\frac{\theta}{n} \sum_i^n X_i \sim \Gamma(n, \theta \times \frac{n}{\theta}) \Rightarrow \theta \bar{X}_n \sim \Gamma(n, n);$$

Since the distribution does not depend on θ, $g(x, \theta) = \theta \bar{X}_n$ is a pivotal quantity.

(b) Use the pivot to construct the 95% confidence interval for θ.

$\frac{qGamma(0.025,10,10)}{\bar{X}_n} < \theta < \frac{qGamma(0.975,10,10)}{\bar{X}_n}$

(c) Apply your interval to this data set.

```
x <- c(1.07, 0.88, 0.66, 0.55, 1.15, 0.65, 3.45, 3.55, 3.51, 0.48)
n <- length (x)
alpha <- 0.05
qgamma(c(alpha / 2, 1 - alpha / 2), shape = n, rate = n)/mean(x)

#equivalent to qchisq(c(alpha / 2, 1 - alpha / 2), df = 2*n)
#/(2*sum(x))

 ######################
 # 0.3006513 1.0711476 #
 ######################
```

Listing 1: R code

2. Consider an i.i.d. sample of random variables with following density function.

$$f(x|\sigma) = \frac{1}{2\sigma}e^{-\frac{|x|}{\sigma}} \tag{3}$$

Use the approximate pivot method to construct a $100(1-\sigma)$ % confidence interval of σ.

Since $f(x|\sigma)$ is really the Laplace distribution whose location parameter is equal to 0 and scale parameter is equal to σ, the following relationship can be made.

$$\frac{2\sum_{i=1}^{n}|x_i|}{\sigma} \sim \chi^2(2n) \tag{4}$$

Let $q(p)$ be the quantile for the distribution $\chi^2(2n)$. Using Equation 4 and $q(p)$, the CI with 95 % confidence is the following.

$$\text{CI} = \left[\frac{2\sum_{i=1}^{n}|x_i|}{q(0.975)}, \frac{2\sum_{i=1}^{n}|x_i|}{q(0.025)}\right] \tag{5}$$

3. A sample of students from an introductory psychology class were polled regarding the number of hours they spent studying for the last exam. All students anonymously submitted the number of hours on a 3 by 5 card. There were 24 individuals in the one section of the course polled. The data was used to make inferences regarding the other students taking the course. The data is shown below.

$$D = \{4.5, 7.5, 22, 9, 7, 10.5, 14.5, 15, 9, 19, 9, 3.5, 8, 11, 2.5, 5, 9, 8.5, 7.5, 18, 20, 14, 20, 8\} \tag{6}$$

(a) Obtain a confidence interval based on central limit theorem.

```
#### 3(a)

y <- c(4.5, 7.5, 22, 9, 7, 10.5, 14.5, 15, 9, 19, 9, 3.5, 8, 11,
    2.5, 5, 9, 8.5, 7.5, 18, 20, 14, 20, 8)
n <- length (y)
alpha <- 0.05
mu.Low <- mean(y) - qnorm( 1-alpha/2 )* sqrt(var(y))/sqrt(n)
mu.Upp <- mean(y) + qnorm( 1-alpha/2 )* sqrt(var(y))/sqrt(n)
CI <- c(mu.Low, mu.Upp)

####################
#8.676935 13.156398#
####################
```

Listing 2: R code

(b) Obtain a confidence interval based on T-distributions.

```
#### 3(b)

mu.Low <- mean(y) - qt( 1-alpha/2, n-1 )* sqrt(var(y))/sqrt(n)
mu.Upp <- mean(y) + qt( 1-alpha/2, n-1 )* sqrt(var(y))/sqrt(n)
CI <- c(mu.Low, mu.Upp)

####################
#8.552726 13.280607#
####################
```

Listing 3: R code

(c) Obtain a confidence interval based on bootstrapping with $B = 10,000$.

```
#### 3(c)
 #I tried two packages with the sam seed and each gives a different
 #answer.
 #The second approach/package is what we used in class.
library(boot)
args(boot)

set.seed(1)
mean.fun <- function(y, i)
{     m <- mean(y[i])
      n <- length(i)
      v <- (n-1)*var(y[i])/n^2
      c(m, v)
```

```
}
y.boot <- boot(y, mean.fun, R = 10000)
boot.ci (y.boot, type = "all")

 ################################################################
 #Intervals :                                                   #
 #  Level        Normal              Basic            Studentized   #
 #   95%    ( 8.75, 13.11 )    ( 8.71, 13.02 )    ( 8.79, 13.53 ) #
 #               Percentile            BCa                         #
 #          ( 8.81, 13.12 )    ( 8.92, 13.29 )                     #
 ################################################################

library(bootstrap)
args(bootstrap)

set.seed(1)
y.boot2 <- bootstrap(y,nboot=10000,theta=mean)
quantile(y.boot2$thetastar,c(.025,.975))

##################
#  2.5%   97.5%  #
# 8.8125 13.1250 #
##################

```

Listing 4: R code

4. The Poisson distribution has been used by traffic engineers as a model for light traffic, based on the rationale that if the rate is approximately constant and the traffic is light (so the individual cars move independently of each other), the distribution of counts of cars in a given time interval or space area should be nearly Poisson. The following table shows the number of right turns during 300 3-min intervals at a specific intersection.

n	Frequency
0	14
1	30
2	36
3	68
4	43
5	43
6	30
7	14
8	10
9	6
10	4
11	1
12	1
13+	0

(a) Use the pivot method to construct a $(1-\alpha)$ confidence interval of the rate.

The MLE estimate of the parameter λ is simply $\bar{Y}$, where $Y = \{y_1, ..., y_n\}$.

$$\text{CI} = \left[\bar{Y} - z(\alpha/2)\sqrt{\frac{\bar{Y}}{n}}, \bar{Y} + z(\alpha/2)\sqrt{\frac{\bar{Y}}{n}}\right] \tag{7}$$

(b) Use variance stabilization method to construct a $(1-\alpha)$ confidence interval of the rate.

$$\begin{aligned}
x_i &\sim \text{Poisson}(x_i) \\
\text{Var}(x_i) &= \lambda_i \\
&= f^2(\lambda x_i) \\
f(z) &= z^{\frac{1}{2}} \\
h(z) &= \frac{1}{f(z)} \\
&= \frac{1}{z^{\frac{1}{2}}} \\
\int h(z)dz &= \int \frac{1}{z^{\frac{1}{2}}} dz \\
&= \int z^{-\frac{1}{2}} dz \approx z^{\frac{1}{2}} \\
&\text{so } h(x_i) \approx x_i^{\frac{1}{2}} \text{ can be used to stabilize variance}
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\text{For large values of } n \text{, the following is true} \\
\sqrt{\frac{n}{\lambda}}(\hat{\lambda} - \lambda) &\to N(0,1) \\
&\text{Let } h(z) = \lambda^{\frac{1}{2}} \text{ and } h'(z) = \frac{1}{2}\lambda^{-\frac{1}{2}} \\
\frac{\sqrt{\frac{n}{\lambda}}(\sqrt{\hat{\lambda}} - \sqrt{\lambda})}{\frac{1}{2}\lambda^{-\frac{1}{2}}} &\to N(0,1) \\
\text{CI} &= \left[\left(\sqrt{\hat{\lambda}} + \frac{z(\frac{\alpha}{2})}{2\sqrt{n}}\right)^2, \left(\sqrt{\hat{\lambda}} - \frac{z(\frac{\alpha}{2})}{2\sqrt{n}}\right)^2\right]
\end{aligned} \tag{9}$$

(c) Plug in the data and calculate the 95 % C by both methods. Which one do you prefer?

```
#### 4(c)
z <- c(rep(0,14),rep(1,30),rep(2,36),rep(3,68),rep(4,43),rep(5,43),
       rep(6,30),rep(7,14),rep(8,10),rep(9,6),rep(10,4),11,12)
n <- length (z)
alpha <- 0.05
mu.Low <- mean(z) - qnorm( 1-alpha/2 )* sqrt(mean(z))/sqrt(n)
mu.Upp <- mean(z) + qnorm( 1-alpha/2 )* sqrt(mean(z))/sqrt(n)
CI <- c(mu.Low, mu.Upp)

 ##################
 # 2.5%    97.5%  #
 #3.67005 4.11661 #
 ##################

CI2 <- c((sqrt(mean(z)+qnorm(alpha/2)/2/sqrt(n)))^2,
    (sqrt(mean(z)-qnorm(alpha/2)/2/sqrt(n)))^2)

 ##################
 # 2.5%    97.5%  #
 #3.83675 3.94991 #
 ##################

 #CI2 is better as its range is closer compare to CI

```

Listing 5: R code