98

STAT 8003, Homework 5

Group #8

Members: Anastasia Vishnyakova, Nooreen Dabbish, Yinghui Lu

Oct. 10, 2013

Problem 1. Consider an i.i.d. sample of random variables with density function

$$f(x|\sigma) = \frac{1}{2\sigma}\exp(-\frac{|x|}{\sigma})$$

.

a). Find the method of moments estimate of σ:

First, we'd like to find the first moment of X, that is $E(X)$.

$$E(X) = \int_{-\infty}^{\infty} x\frac{1}{2\sigma}\exp(-\frac{|x|}{\sigma}\mathrm{d}x$$

It is obvious that inside the integral it is an odd function, thus

$$E(X) = 0$$

Then we calculate the second moment of X as follows,

$$E(X^2) = \int_{-\infty}^{\infty} x^2\frac{1}{2\sigma}\exp(-\frac{|x|}{\sigma})\mathrm{d}x$$

It is also obvious that inside the integral it is an even function, so we can rewrite it as

$$
\begin{aligned}
E(X^2) &= 2\int_0^{\infty} x^2 \frac{1}{2\sigma}\exp(-\frac{x}{\sigma})\mathrm{d}x \\
&= \int_0^{\infty} x^2 \frac{1}{\sigma}\exp(-\frac{x}{\sigma})\mathrm{d}x
\end{aligned}
$$

Solving this by integral by part, we get

$$E(X^2) = 2\sigma^2$$

Set

$$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

we will get

$$\sigma = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2}$$

Hence the mom estimate of σ is $\sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2}$.

b). Find the maximum likelihood estimate of σ.

Maximum likelihood estimation

In this approach, a likelihood estimator is constructed using the sampled datapoints. This estimator is a function of the distribution function parameters.

1. Find the likelihood function. This is the product of the density functions for i.i.d. samples:

$$\begin{aligned}\mathcal{L}(\theta|x_1,\ldots,x_n) &= f(x_1|\theta)\times\ldots\times f(x_n|\theta)\\ &= \prod_{i=1}^{n} f(x_i|\theta)\end{aligned}$$

2. To make the multiplication of n terms more tractable, use the log-likelihood function:

$$\ell(\theta|x_1,\ldots,x_n) = \ln\mathcal{L}(\theta|x_1,\ldots,x_n) = \sum_{i=1}^{n}\ln f(x_i|\theta)$$

3. Can find maxima using calculus—solving for points where ℓ' is zero and ℓ'' is negative. This gives the values of theta with the greatest likelihood.

$$
\begin{aligned}
\mathcal{L}(\sigma|x_1,\ldots,x_n) &= \prod_{i=1}^{n} f(x_i|\sigma) \\
&= \prod_{i=1}^{n} \frac{1}{2\sigma}\exp\left(-\frac{|x_i|}{\sigma}\right) \\
&= \left(\frac{1}{2\sigma}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{|x_i|}{\sigma}\right) \\
\ell = \ln\mathcal{L} &= \ln\left(\left(\frac{1}{2\sigma}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{|x_i|}{\sigma}\right)\right) \\
&= n\ln\left(\frac{1}{2\sigma}\right) + \sum_{i=1}^{n} \ln\exp\left(-\frac{|x_i|}{\sigma}\right) \\
&= n\ln\left(\frac{1}{2\sigma}\right) - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i| \\
&= -n\ln 2\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|
\end{aligned}
$$

Now we take the derivative of the log-likelihood

$$
\begin{aligned}
\ell' &= \frac{\mathrm{d}}{\mathrm{d}\sigma}\left(-n\ln 2\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|\right) \\
&= -\frac{n}{\sigma} + \frac{\sum\limits_{i=1}^{n}|x_i|}{\sigma^2} \\
\ell'' &= \frac{\mathrm{d}}{\mathrm{d}\sigma}\left(-\frac{n}{\sigma} + \frac{\sum\limits_{i=1}^{n}|x_i|}{\sigma^2}\right) \\
&= \frac{n}{\sigma^2} - \frac{2\sum\limits_{i=1}^{n}|x_i|}{\sigma^3}
\end{aligned}
$$

Solving for $\ell\prime = 0$ in order to optimize

$$\begin{aligned}
0 &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} |x_i|}{\sigma^2} \\
&= -n\sigma + \sum_{i=1}^{n} |x_i| \\
n\sigma &= \sum_{i=1}^{n} |x_i| \\
\hat{\sigma} &= \frac{1}{n}\sum_{i=1}^{n} |x_i|
\end{aligned}$$

Second derivative test—a maxima will have a negative second derivative

$$\ell'' = \frac{n}{\sigma^2} - \frac{2\sum_{i=1}^{n} |x_i|}{\sigma^3}$$

$$\ell'' < 0 \implies \sigma < \frac{2\sum_{i=1}^{n} |x_i|}{n}$$

$$\text{Obviously } \frac{1}{n}\sum_{i=1}^{n} |x_i| < \frac{2}{n}\sum_{i=1}^{n} |x_i|$$

$$\text{Thus } \ell'' < 0 \text{ when } \sigma = \frac{1}{n}\sum_{i=1}^{n} |x_i|$$

Hence the mle estimate of σ is $\frac{1}{n}\sum_{i=1}^{n} |x_i|$.

c). Use the pivot method to construct a $(1-\alpha)\%$ confidence interval of σ:

From (b) we know that the mle estimate of σ is $\frac{1}{n}\sum\limits_{i=1}^{n} |x_i|$, which is $\overline{|x|}$, we then try to find the distribution of $|X|$. Let $Y = g(X) = |x|$, then Y can be divided into two partitions,

$$\begin{aligned}
g_1(X) &= x \qquad x \in [0, +\infty) \\
g_2(X) &= -x \quad x \in (-\infty, 0)
\end{aligned}$$

then we have

$$
\begin{aligned}
f_Y(y) &= f_X(g_1^{-1}(y)|g_1^{'-1}(y)| + f_X(g_2^{-1}(y)|g_2^{'-1}(y)| \\
&= \frac{1}{2\sigma}\exp{(-\frac{y}{\sigma})} + \frac{1}{2\sigma}\exp{(-\frac{y}{\sigma})} \\
&= \frac{1}{\sigma}\exp{(-\frac{y}{\sigma})} \qquad y \in [0, +\infty) \\
E(Y) &= E(|X|) \\
&= \int_0^\infty y\frac{1}{\sigma}(\exp{(-\frac{y}{\sigma})}\mathrm{d}y \\
&= \sigma \\
Var(Y) &= Var(|X|) \\
&= E(Y^2) - E^2(Y) \\
&= 2\sigma^2 - \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

As X_is are $i.i.d$, $|X_i|$s are also $i.i.d$, then we have

$$
\begin{aligned}
E(\overline{|X|}) &= E(\frac{1}{n}\sum_{i=1}^{n}|X_i|) \\
&= \frac{1}{n}n\sum_{i=1}^{n}E(|X_i|) \\
&= \sigma \\
Var(\overline{|X|}) &= \frac{1}{n^2}Var(\sum_{i=1}^{n}|X_i|) \\
&= \frac{1}{n}Var(|X_i|) \\
&= \frac{1}{n}\sigma^2
\end{aligned}
$$

We then denote $\hat{\sigma}^2$ as the mle estimate of σ^2, and$\hat{\sigma}^2 = (\frac{1}{n}\sum\limits_{i=1}^{n}|x_i|)^2$.

Thus, according to central limit theorem, we have

$$\frac{\overline{|X|} - \sigma}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim AN(0,1)$$

Let the $100(1-\frac{\alpha}{2})\%$-th quantile of standard normal distribution to be $Z_{\frac{\alpha}{2}}$, then we have

$$p\left(-Z_{\frac{\alpha}{2}} \leq \frac{\overline{|x|} - \sigma}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \leq Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\overline{|x|} - \sqrt{\frac{\hat{\sigma}^2}{n}} Z_{\frac{\alpha}{2}} \leq \sigma \leq \overline{|x|} + \sqrt{\frac{\hat{\sigma}^2}{n}} Z_{\frac{\alpha}{2}}$$

Hence, using the above method, we construct a confidence interval of

$$L = \overline{|x|} - \sqrt{\frac{\hat{\sigma}^2}{n}} Z_{\frac{\alpha}{2}}$$

$$U = \overline{|x|} + \sqrt{\frac{\hat{\sigma}^2}{n}} Z_{\frac{\alpha}{2}}$$

Problem 2. (Same setting as Problem 2 of Homework 4).

a). Use the pivot method to construct a $(1-\alpha)$% confidence interval of the rate.
First, we denote the true value of rate by λ, and the mom estimate of λ by $\hat{\lambda}$. And we denote the number of right turn during a given interval by X_i, where $i = 1, 2 \cdots n$, and the total number of intervals having been recorded by n. Then we have

$$\hat{\lambda} = E(X) = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

After plugging in the data given in the instruction, we got $\hat{\lambda} = 3.89$.
As X_i conform to poisson distribution, and X_is are $i.i.d$, then we have

$$
\begin{aligned}
E(\hat{\lambda}) &= E(\bar{X}) \\
&= \frac{\sum_{i=1}^{n} E(X_i)}{n} \\
&= E(X) \\
&= \lambda \\
Var(\hat{\lambda}) &= Var(\bar{X}) \\
&= (\frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{Var(X)}{n} \\
&= \frac{\lambda}{n}
\end{aligned}
$$

In this problem, the sample size is 300, which could be considered as a large sample size in statistics, thus we use central limit theorem to get the asympototic distribution of $\hat{\lambda}$ to $N(0,1)$, that is,

$$
\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda}} \sim AN(0,1)
$$

Let the $100(1 - \frac{\alpha}{2})\%$-th quantile of standard normal distribution to be $Z_{\frac{\alpha}{2}}$, and plugging in $\hat{\lambda}$ to replace λ, then we have

$$
\begin{aligned}
P(-Z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\hat{\lambda}}} \leq Z_{\frac{\alpha}{2}}) = 1 - \alpha \\
\hat{\lambda} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}
\end{aligned}
$$

Hence, using the above pivot method, we construct a confidence interval of

$$L = \hat{\lambda} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{n}}$$

$$U = \hat{\lambda} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{n}}$$

b). Use variance stabilization method to construct confidence interval of the rate. From what has been proved above, we have

$$\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda}} \sim AN(0, 1)$$

Suppose

$$a_n(W_n - b) \sim AN(0, 1)$$

Then, using Taylor expansion we got

$$a_n\{g(W_n - g(b)\} \sim AN(0, \{g'(b)\}^2$$

and it also can be reformulated as

$$\frac{a_n\{g(W_n - g(b)\}}{|g'(b)|} \sim AN(0, 1)$$

Set a_n to $\sqrt{\frac{n}{\lambda}}$, then we have

$$\frac{\sqrt{n}\{g(W_n) - g(b)\}}{\sqrt{\lambda}|g'(b)|} \sim AN(0, 1)$$

Since $E(\hat{\lambda}) = \lambda$, we have

$$\frac{\sqrt{n}\{g(\hat{\lambda}) - g(\lambda)\}}{\sqrt{\lambda}|g'(\lambda)|} \sim AN(0,1)$$

Make $g'(\lambda) = \lambda^{-\frac{1}{2}}$, by integration, we obtain

$$g(\lambda) = 2\lambda^{\frac{1}{2}}$$

Thus,

$$2\sqrt{n}(\hat{\lambda}^{\frac{1}{2}} - \lambda^{\frac{1}{2}}) \sim AN(0,1)$$

Let the $100(1-\frac{\alpha}{2})\%$-th quantile of standard normal distribution to be $Z_{\frac{\alpha}{2}}$, then we have

$$P(-Z_{\frac{\alpha}{2}} \leq 2\sqrt{n}(\hat{\lambda}^{\frac{1}{2}} - \lambda^{\frac{1}{2}}) \leq Z_{\frac{\alpha}{2}}) = 1 - \alpha$$
$$(\hat{\lambda}^{\frac{1}{2}} - \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}})^2 \leq \lambda \leq (\hat{\lambda}^{\frac{1}{2}} + \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}})^2$$

Using the variance stablization method above to construct the confidence interval, we have

$$L = (\hat{\lambda}^{\frac{1}{2}} - \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}})^2$$
$$U = (\hat{\lambda}^{\frac{1}{2}} + \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}})^2$$

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

plugging in the data of $\hat{\lambda} = 3.89$, $Z_{2.5} = 1.96$, and $n = 300$, we got

$$L = 3.89 - \sqrt{\frac{3.89}{300}} \times 1.96 \approx 3.67$$
$$U = 3.89 + \sqrt{\frac{3.89}{300}} \times 1.96 \approx 4.11$$

for the pivot method.

And

$$L = (\sqrt{3.89} - \frac{1.96}{2 \times \sqrt{300}})^2 \approx 3.67$$
$$U = (\sqrt{3.89} + \frac{1.96}{2 \times \sqrt{300}})^2 \approx 4.12$$

for the variance stablization method.
From the results we can see the confidence interval generated using the two methods are very similar. That is because the sample size is large (n=300).
Looking at the confidence interval obtained in a) and b), it is suggested that the confidence interval constructed using the pivot method might be more influnced by the change of sample size and the mean of the sample. Thus the variance stabilization method seems more stable than the pivot method, especially when the sample size is small.
We then tried to use R to do simulation of these two methods. First we set sample size to 300 and repeated the experiment for 100 times, and we obtained the coverage of confidence interval using pivot methods and variance stabilization method are the same (98%), and the average half-width of the two are also the same. But when we changed the sample size to 6, and repeated it also for 100 times, then the confidence interval constructed by pivot method gave a 95% coverage, while the confidence interval by variance stabilization method gave a 97% coverage, and the average half-widths are also the same. It means that when the sample size is large, either method will achieve a good coverage of the true value, but when sample size is small, variance stabilization might be more acurate.
Below is the simulation results and the R code.

```
R code for large sample size:
```

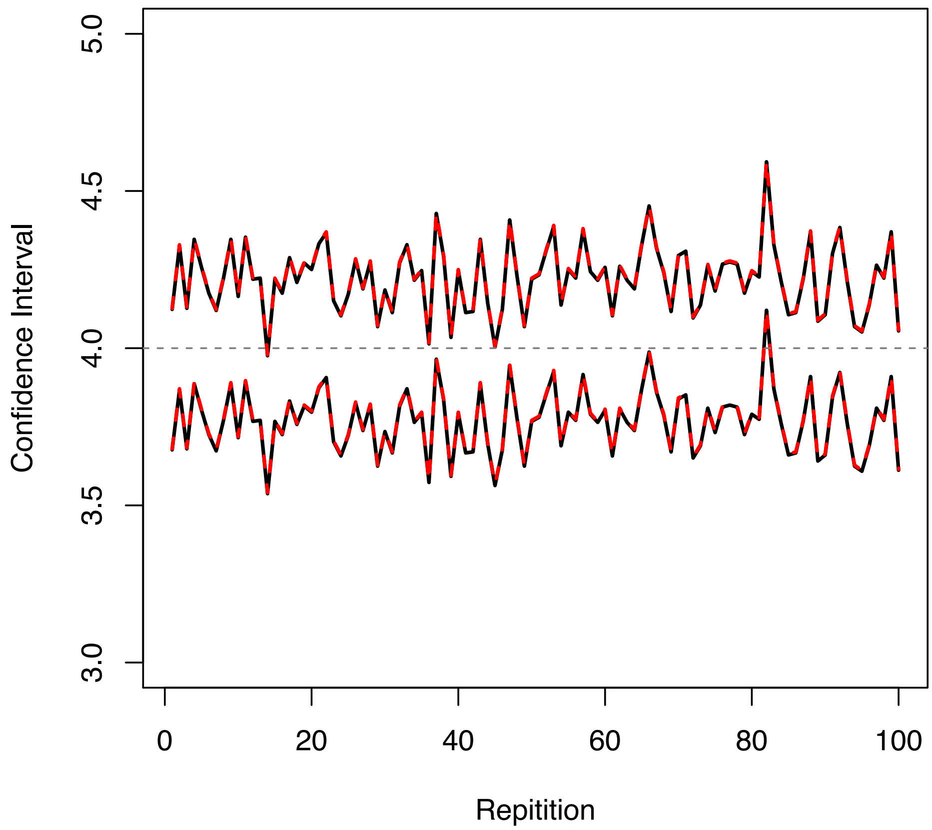

Figure 1: Silmulation of the confidence interval of Poisson distribution with large sample size

```
set.seed(34567)
n = 300
repit = 100

# initialize the vector to store CIs
ci1 = matrix(0,nrow=repit,ncol=2)
ci2 = matrix(0,nrow=repit,ncol=2)
l=4

for (ii in 1:repit){
  #cat("ii = ", ii, "\n")
  y = rpois(n,lambda=l)
```

```
  # y is poisson observation

  mean.y = mean(y)
  a = qnorm(1-0.025)

  ci1[ii,1] = mean.y - a*sqrt(mean.y/n)
  ci1[ii,2] = mean.y + a*sqrt(mean.y/n)

  ci2[ii,1] = (sqrt(mean.y)-(a/(2*sqrt(n))))^2
  ci2[ii,2] = (sqrt(mean.y)+(a/(2*sqrt(n))))^2
}

plot(1:repit,ci1[,1],ylim=c(3,5),xlab="Repitition",
     ylab="Confidence Interval",type="l",lwd=2)
lines(1:repit,ci1[,2],lwd=2)
lines(1:repit,ci2[,1],col=2,lty=2,lwd=2)
lines(1:repit,ci2[,2],col=2,lty=2,lwd=2)
abline(h=l,lty=2,col="gray50")

ct1 = sum((ci1[,1]<l)&(ci1[,2]>l))
ct2 = sum((ci2[,1]<l)&(ci2[,2]>l))

hw1 = mean((ci1[,2]-ci1[,1])/2)
hw2 = mean((ci2[,2]-ci2[,1])/2)

cat("coverage 1 = ", ct1/repit,"\n")
cat("coverage 2 = ", ct2/repit,"\n")

cat("half-width 1 = ", hw1,"\n")
cat("half-width 2 = ", hw2,"\n")
```

R code for small sample size:

```
set.seed(34567)
n = 6
repit = 100

# initialize the vector to store CIs
ci1 = matrix(0,nrow=repit,ncol=2)
ci2 = matrix(0,nrow=repit,ncol=2)
l=4
```

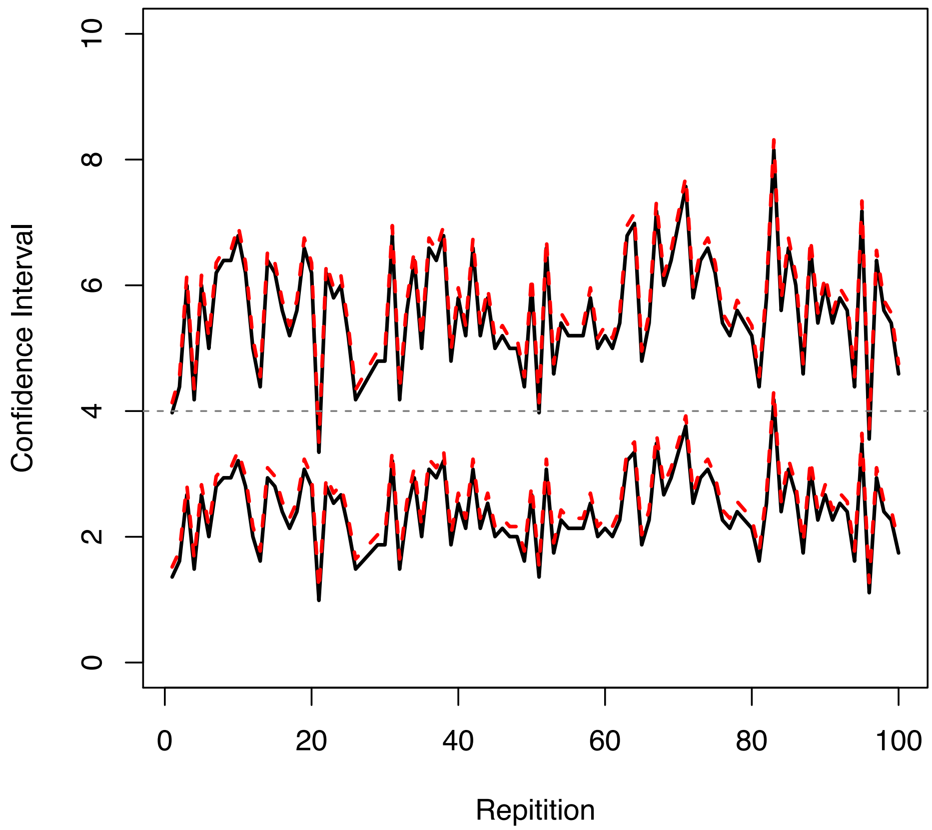

Figure 2: Silmulation of the confidence interval of Poisson distribution with small sample size

```
for (ii in 1:repit){
  #cat("ii = ", ii, "\n")
  y = rpois(n,lambda=l)
  # y is poisson observation

  mean.y = mean(y)
  a = qnorm(1-0.025)

  ci1[ii,1] = mean.y - a*sqrt(mean.y/n)
  ci1[ii,2] = mean.y + a*sqrt(mean.y/n)

  ci2[ii,1] = (sqrt(mean.y)-(a/(2*sqrt(n))))^2
```

```
  ci2[ii,2] = (sqrt(mean.y)+(a/(2*sqrt(n))))^2
}

plot(1:repit,ci1[,1],ylim=c(1,7),xlab="Repitition",
     ylab="Confidence Interval",type="l",lwd=2)
lines(1:repit,ci1[,2],lwd=2)
lines(1:repit,ci2[,1],col=2,lty=2,lwd=2)
lines(1:repit,ci2[,2],col=2,lty=2,lwd=2)
abline(h=l,lty=2,col="gray50")

ct1 = sum((ci1[,1]<l)&(ci1[,2]>l))
ct2 = sum((ci2[,1]<l)&(ci2[,2]>l))

hw1 = mean((ci1[,2]-ci1[,1])/2)
hw2 = mean((ci2[,2]-ci2[,1])/2)

cat("coverage 1 = ", ct1/repit,"\n")
cat("coverage 2 = ", ct2/repit,"\n")

cat("half-width 1 = ", hw1,"\n")
cat("half-width 2 = ", hw2,"\n")
```

Problem 3. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ and σ are unknown. How should the constant c be chosen so that the interval $(\infty < \mu < \bar{X} + c)$ is a 95% confidence interval for μ; that is, c should be chosen so that

$$p(-\infty < \mu \leq \bar{X} + c) = 0.95$$

.

Because μ and σ are unknown, we will use the sample variance s^2 as the estimate of the variance σ^2, and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{s} \sim T_{n-1}$$

Let q be the $.95$ quantile fo t-distribution with n-1 degrees of freedom, then we have

$$P(-\infty < \frac{\sqrt{n}(\bar{X} - \mu)}{s} < q) = .95$$

$$-\infty < \mu < \bar{X} + \frac{sq}{\sqrt{n}} = \bar{X} + c$$

$$c = \frac{sq}{\sqrt{n}}$$

Problem 4. 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.

$$\mu = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

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

```
> #Look up 95% interval in normal :
> qnorm(0.025)
[1] -1.959964
```

$$P(-1.96 \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq 1.96)$$

$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leq \mu \leq \bar{X} + 1.96\sqrt{\sigma^2/n)}$$

Then, $L = \bar{X} - 1.96\sqrt{\sigma^2/n}$; $U = \bar{X} + 1.96\sqrt{\sigma^2/n)}$

b). Obtain a condence interval based on the t-distributions.

```
> #Look up 95% interval in t-distribution:
> qt(0.025,df=n-1)
[1] -2.068658
```

Don't plug in the data here.

$$P(-2.068658 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{s} \leq 2.068658)$$

$$P(\bar{X} - 2.068658\sqrt{s^2/n} \leq \mu \leq \bar{X} + 2.068658\sqrt{s^2/n})$$

Then, $L = \bar{X} - 2.068658\sqrt{s^2/n}$; $U = \bar{X} + 2.068658\sqrt{s^2/n})$

c). Plug in the data, are these two intervals similar? What conclusions can you draw based on the condence intervals.

The following R commands were used to do the calculations:

```
> X <- c(4.5, 22, 7, 14.5, 9, 9, 3.5, 8, 11, 7.5, 18, 20,+
+ 7.5, 9, 10.5, 15, 19, 2.5, 5, 9, 8.5, 14, 20, 8)
> n <-length(X)
> X_bar <- sum(X)/n
> s <- sqrt((sum((X-X_bar)^2))/(n-1))
> sigma <- sqrt((sum((X-X_bar)^2))/(n))
> #Calculate confidence intervals based on Central Limit Theorem (CLM)
> s_e_CLT <- abs((qnorm(0.025)∗sigma)/sqrt(n))
>
> #Overlay distribution of X with normal curve
> h <- hist(X, breaks=n/2, col= "lightgray", xlab="Hours of Study")
> xfit<-seq(min(X),max(X),length=n)
> yfit <- dnorm(xfit,mean = X_bar,sd = s) ∗diff(h$mids[1:2])∗n
> lines(xfit, yfit, col="blue", lwd=2)
>
> #Calculate confidence intervals based on t-distribution
> s_e_T <- abs(qt(0.025,df=n-1)∗s/sqrt(n))
>
> L_CLT <- X_bar - s_e_CLT
> U_CLT <- X_bar + s_e_CLT
> L_T <- X_bar - s_e_T
> U_T <- X_bar + s_e_T
>
> #Mean hours of study for the exam:
> X_bar
[1] 10.91667
> #Confidence interval based central limit theorem:
> L_CLT
[1] 8.724093
```

```
> U_CLT
[1] 13.10924
> #Confidence interval based on t-distribution:
> L_T
[1] 8.552726
> U_T
[1] 13.28061
```

A histogram of data with normal curve with data mean and standard deviation is shown below.

T-based confidence interval is wider compared to confidence intervals based on the Central Limit Theorem because t-distribution has "fatter" tails compared to the standard normal distribution when n is low, like in our case. Therefore, t-distribution-derived confidence interval is greater than standard normal-derived confidnece interval, as evidenced by our calculations. For this problem t-distribution is more appropriate due to the small n. T-distirbutionis designed to give us a better interval estimate of the mean when we have a small sample size. In our case n=24, which is less than 30, hence, t-distirbution with 23 degrees of freedom might be more accurate than the central limit theorem method.

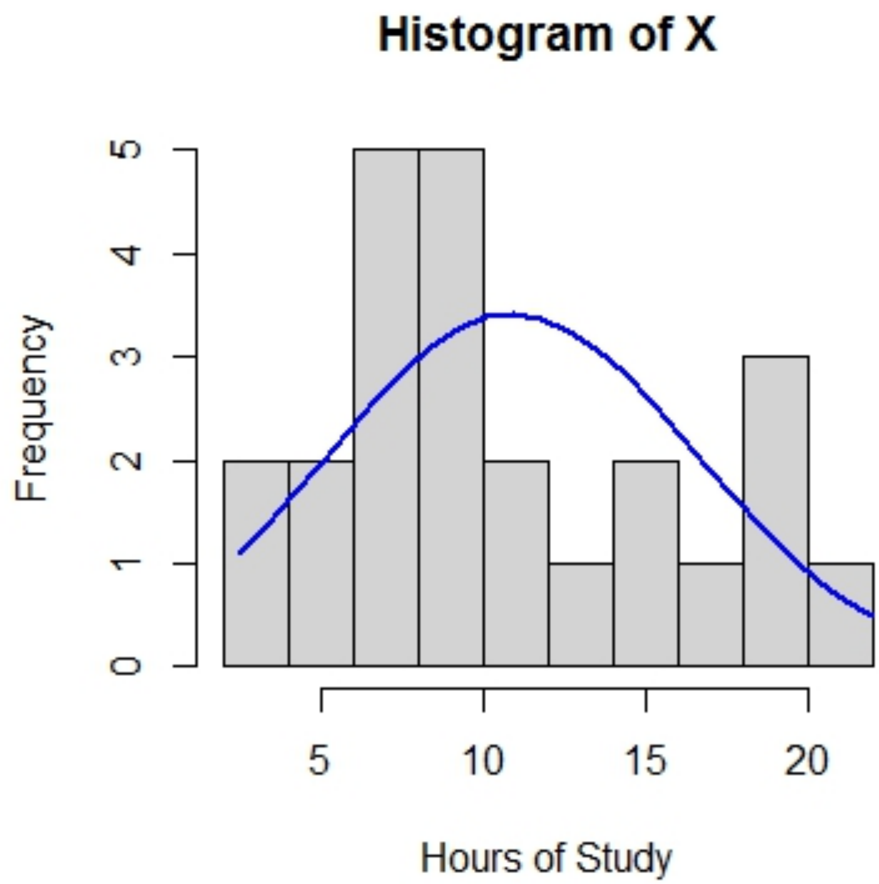

Figure 3: Histogram of Number of hours spent studying with normal curve overlay