DSCC/CSC/TCS 462 Statistics Assignment 2

Aradhya Mathur

2022-10-03

1. Consider random variables X and Y. Calculate $Var(3X + 2Y)$ given the following information. (Hint: At some point, you may need to use the fact that variance cannot be negative.)

    - $E(3X + 2) = 8$
    - $E(4X + 2Y) = 14$
    - $E(2Y(X + 1)) = 28$
    - $E(X^2Y^2) = 144$
    - $Cov(X^2, Y^2) = 36$
    - $E(X^2 + 2Y^2) = 33$

1) Given:

$$E(3X+2) = 8 \quad \text{-- (1)}$$
$$E(4X+2Y) = 14 \quad \text{-- (2)}$$
$$E(2Y(X+1)) = 28 \quad \text{-- (3)}$$
$$E(X^2 Y^2) = 144 \quad \text{-- (4)}$$
$$Cov(X^2, Y^2) = 36 \quad \text{-- (5)}$$
$$E(X^2 + 2Y^2) = 33 \quad \text{-- (6)}$$

To find: $var(3X+2Y)$.

We know, $E(ax+b) = aE(x) + b$

$$E(3X+2) = 3E(x) + 2 = 8$$
$$3E(x) = 6$$
$$E(x) = 2 \;//$$

Also, $E(ax+by) = aE(x) + bE(y)$.

$$E(4X+2Y) = 14$$
$$4E(x) + 2E(Y) = 14$$
$$4 \times 2 + 2E(Y) = 14$$
$$2E(Y) = 6$$
$$E(Y) = 3 \;//$$

Now

$$Var(3X+2Y) = 9Var(x) + 4Var(Y) + 12cov(X,Y)$$

because

$$Var(X+Y) = Var(x) + Var(Y) + 2cov(X,Y)$$

and $Var(ax+bY) = a^2 var(x) + b^2 var(Y) + 2ab\,cov(X,Y)$.

$E(2XY + 2Y) = 28$

$2E(XY) + 2E(Y) = 28$

$E(Y) = 3$

$2E(XY) + 6 = 28$

$E(XY) = 11$

$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$

$= 11 - 2\times 3$

$= 11 - 6 = 5$

$\mathrm{Cov}(X,Y) = 5$

$\mathrm{Cov}(X^2, Y^2) = E(X^2Y^2) - E(X^2)E(Y^2)$

$\cancel{144} = 36 = 144 - E(X^2)E(Y^2)$

$\therefore E(X^2)E(Y^2) = 108 \quad - (7)$

Also

$E(X^2 + 2Y^2) = 33$

$E(X^2) + 2E(Y^2) = 33 \quad - (8)$

Solving (7) & (8)

$$E(X^2) = 33 - 2E(Y^2).$$

let $E(y^2) = z$.

$$(33 - 2z) \cdot z = 108$$

$$2z^2 - 33z + 108 = 0$$

$$z = 12 \text{ or } 4.5.$$

$E(Y^2) = 12$	4.5 X
$E(X^2) = 9$	24

$E(Y^2) = 12$ & $E(X^2) = 9$

Calculating $\quad var(Y) = E(Y^2) - [E(Y)]^2$

$$= 12 - 9 = 3$$

$$var(X) = E(X^2) - [E(X)]^2$$

$$= 9 - 4 = 5$$

$$V(3X + 2Y) = 9\,var(X) + 4\,var(X) + 12\,cov(X, Y)$$

$$= 9 \times 5 + 4 \times 3 + 12 \times 5$$

$$= 45 + 12 + 60$$

$$= 117$$

$$Var(3X + 2Y) = 117$$

2. The density function of X is given by $f_X(x) = ax^3 + bx + 2/3$ for $xin[0, 1]$, and $E(X) = 7/15$.

   a. Find a and b.

   b. Calculate the CDF, $F(X)$.

   c. Calculate $Pr(X > 0.75)$

   d. Suppose $Y = 1.5X + 2$. Calculate $E(Y)$.

A2). $f_x(X) = ax^3 + bx + 2/3 \qquad x \in [0,1]$

$$E(X) = \frac{7}{15}$$

a) We know.

$$\int_n^m f(x)\,dx = 1$$

$$\int_n^m f(x)\,x\,dx = E(X).$$

Using first.

$$\int_0^1 \left(ax^3 + bx + \frac{2}{3}\right)dx = 1$$

$$\left[\frac{ax^4}{4} + \frac{bx^2}{2} + \frac{2x}{3}\right]_0^1 = 1$$

$$\frac{a}{4} + \frac{b}{2} + \frac{2}{3} = 1$$

Taking LCM

$$3a + 6b = 4 \quad -(1)$$

Using second.

b). $\int_0^1 (ax^3 + bx + \frac{2}{3})x\,dx = \frac{7}{15}$

$\int_0^1 (ax^4 + bx^2 + \frac{2}{3}x)\,dx = \frac{7}{15}$

$\left[\frac{ax^5}{5} + \frac{bx^3}{3} + \frac{x^2}{3}\right]_0^1 = \frac{7}{15}$

$\frac{a}{5} + \frac{b}{3} + \frac{1}{3} = \frac{7}{15}$

Taking LCM

$3a + 5b = 2. \ -②$

from ① & ②.

~~b~~

$3a + 6b = 4$

$-(3a + 5b = 2)$

$b = 2$

$a = -8/3$ //.

b) $F(x) = \int_0^x \left(-\frac{8}{3}u^3 + 2u + \frac{2}{3}\right) du$

$= -\frac{2}{3}x^4 + x^2 + \frac{2}{3}x.$

CDF //

$$F(x) = \begin{cases} 0 & x < 0 \\ -\frac{2}{3}x^4 + x^2 + \frac{2}{3}x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

c). $P(x > 0.75)$

$= 1 - P(x \leq 0.75)$

$= 1 - \int_0^{0.75} \left(-\frac{8}{3}x^3 + 2x + \frac{2}{3}\right) dx$

$= 1 - \left[-\frac{2}{3}x^4 + x^2 + \frac{2}{3}x\right]_0^{0.75}$

$= 1 - \left[-0.2109375 + 0.5625 + 0.5\right]$

$= 0.148375$

d) $Y = 1.5X + 2.$ $\quad E(x) = \frac{7}{15}$

$E(Y) = \frac{3}{2} E(x) + 2.$

$= \frac{3}{2} \times \frac{7}{15} + 2.$

$= \frac{27}{10} = \underline{2.7}$

3. The distribution of battery life of MacBook laptops is normally distributed with a mean of 8.1 hours and a standard deviation of 1.3 hours. The distribution of Dell laptops is normally distributed with a mean of 6.8 hours with a standard deviation of 0.9 hours.

   a. Calculate the probability that a randomly selected MacBook laptop battery lasts more than 9 hours.

```r
#Pr(Z >= z)
x = 9
mean = 8.1
sd = 1.3
z = (x-mean) / sd
ans = 1 - pnorm(z)
ans
```

```
## [1] 0.2443721
```

0.2443721 is the probability that a randomly selected MacBook laptop battery lasts more than 9 hours.

```
b. Calculate the probability that a randomly selected Dell laptop battery
lasts between 6 and 8 hours.
```

```r
#between 6 and 8
x1 = 8
x2 = 6
mean = 6.8
sd = 0.9
z1 = (x1-mean) / sd
z2 = (x2-mean) / sd
ans = pnorm(z1) - pnorm(z2)
ans
```

```
## [1] 0.7217574
```

0.7217574 is the probability that a randomly selected Dell laptop battery lasts between 6 and 8 hours.

```
c. How long must a MacBook laptop battery last to be in the top 3%?
```

```r
#top 3%
mean = 8.1
sd = 1.3
ans = qnorm(0.97,mean,sd)
ans
```

```
## [1] 10.54503
```

```
10.54503  hrs long must a MacBook laptop battery last to be in the top 3%

d. How long must a Dell laptop battery last to be at the 30th percentile?
```

```r
mean = 6.8
sd = 0.9
ans = qnorm(0.3,mean,sd)
ans
```

```
## [1] 6.32804
```

```
6.32804 hrs long must a Dell laptop battery last to be at the 30th percentile

e. Calculate the probability that a randomly selected MacBook laptop lasts
longer than the 25th percentile of Dell laptops.
```

```r
mean = 6.8
sd = 0.9
x = qnorm(0.25,mean,sd)
mean2 = 8.1
sd2 = 1.3
z = (x-mean2)/sd2
ans = 1 - pnorm(z)
ans
```

```
## [1] 0.9288058
```

```
0.9288058 is the probability that a randomly selected MacBook laptop lasts
longer than the 25th percentile of Dell laptops.

f. A randomly selected laptop has a battery life of at least 8.5 hours.
Calculate the probability of this laptop being a MacBook and the probability
of it being a Dell.
```

```r
x = 8.5
mean1 = 8.1
sd1 = 1.3
z1 = (x-mean1)/sd1
ans1 = 1 - pnorm(z1)
ans1
```

```
## [1] 0.3791582
```

```r
mean2 = 6.8
sd2 = 0.9
z2 = (x-mean2)/sd2
ans2 = 1 - pnorm(z2)
ans2
```

```
## [1] 0.02945336
```

0.3791582 is the the probability of this laptop being a MacBook and 0.02945336 is the probability of it being a Dell.

```
4. Payton applies for 12 jobs, each of which he has a 70% chance of getting

a job offer for. Assume that job offers are independent of each other.



a. How many job offers is Payton expected to receive?
```

```r
x=12
p=0.7
n=12
ans = x*p
ans
```

```
## [1] 8.4
```

8 Jobs

```
b. Calculate the probability that Payton receives job offers from all 12
places.
```

```r
x=12
p=0.7
n=12
ans = dbinom(x,n,p)
ans
```

```
## [1] 0.01384129
```

```
0.01384129 is the probability that Payton receives job offers from all 12
places.

c. Calculate the probability that Payton receives between 5 and 7
(inclusive, i.e., 5, 6, or 7) job offers.
```

```r
x1=5
x2=6
x3=7
p=0.7
n=12
ans = dbinom(x1,n,p) + dbinom(x2,n,p) + dbinom(x3,n,p)
ans
```

```
## [1] 0.2668552
```

```
0.2668552 is the probability that Payton receives between 5 and 7
(inclusive, i.e., 5, 6, or 7) job offers.

d. Calculate the probability that Payton receives strictly more than
9 job offers.
```

```r
x1=10
x2=11
x3=12
p=0.7
n=12
ans = dbinom(x1,n,p) + dbinom(x2,n,p) + dbinom(x3,n,p)
ans
```

```
## [1] 0.2528153
```

```
x = 9
n = 12
p = 0.7
ans = 1-pbinom(x,n,p)
ans
```

```
## [1] 0.2528153
```

```
0.2528153 is the probability that Payton receives strictly more than
9 job offers. Both methods work

e. Calculate the probability that Payton receives strictly fewer than 3
job offers.
```

```
x1=0
x2=1
x3=2
p=0.7
n=12
ans = dbinom(x1,n,p) + dbinom(x2,n,p) + dbinom(x3,n,p)
ans
```

```
## [1] 0.0002063763
```

```
x = 2
n = 12
p = 0.7
ans = pbinom(x,n,p)
ans
```

```
## [1] 0.0002063763
```

```
0.0002063763 is the probability that Payton receives strictly fewer than 3
job offers. Both methods work.


f. Calculate the variance of the number of job offer Payton is expected
to receive.
```

```
p=0.7
n=12
var = n*p*(1-p)
var
```

```
## [1] 2.52
```

2.52 is the variance of the number of job offer Payton is expected to receive.

5. Suppose a company has three email accounts, where the number of emails received at each account follows a Poisson distribution. Account A is expected to receive 4.2 emails per hour, account B is expected to receive 5.9 emails per hour, and account C is expected to received 2.4 emails per hour. Assume the three accounts are independent of each other.

a. Calculate the variance of emails received for each of the three accounts.

```r
varA = 4.2
varB = 5.9
varC = 2.4
# Variance will be same as mean
```

varA = 4.2 ;varB = 5.9 ; varC = 2.4

```
b. Calculate the probability that account A receives at least 8 emails in
an hour.
```

```r
x = 8
lamda = 4.2
ans = 1-ppois(x-1,lamda)
ans
```

```
## [1] 0.06394334
```

0.06394334 is the probability that account A receives at least 8 emails in an hour.

```
c. Calculate the probability that account B receives exactly 4 emails in
an hour.
```

```r
x = 4
lambda = 5.9
ans = dpois(x,lambda)
ans
```

```
## [1] 0.1383118
```

0.1383118 is the probability that account B receives exactly 4 emails in an hour.

```
d. Calculate the probability that account C receives at most 3 emails in
an hour.
```

```r
x = 3
lambda = 2.4
ans = ppois(x,lambda)
ans
```

```
## [1] 0.7787229
```

```
0.7787229 is the probability that account C receives at most 3 emails in
an hour.

e. Calculate the probability that account B receives between 2 and 4 emails
in an hour.
```

```r
#Inclusive of 2,3 and 4
lambda = 5.9
x1=2
x2=3
x3=4
ans = dpois(x1,lambda) + dpois(x2,lambda) + dpois(x3,lambda)
ans
```

```
## [1] 0.2797626
```

```r
#Inclusive of 2,3
lambda = 5.9
x1=2
x2=3
ans = dpois(x1,lambda) + dpois(x2,lambda)
ans
```

```
## [1] 0.1414508
```

```r
#Inclusive of 3 and 4
lambda = 5.9

x2=3
x3=4
ans =  dpois(x2,lambda) + dpois(x3,lambda)
ans
```

```
## [1] 0.2320826
```

```r
#Inclusive of only 3
lambda = 5.9
x2=3
ans =  dpois(x2,lambda)
ans
```

```
## [1] 0.09377074
```

```
0.2797626 is the probability that account B receives between 2 and 4 emails
in an hour considering we include 2 and 4.

f. Calculate the probability that the company receives more than 10 emails
total in an hour. (Hint: the sum of Poisson random variables is also Poisson
distributed. Determine $\lambda$ by doing $\E(A+B+C)$.)
```

```r
l1 = 4.2
l2 = 5.9
l3 = 2.4
lambda = l1+l2+l3
x=10
ans = 1-ppois(x,lambda)
ans
```

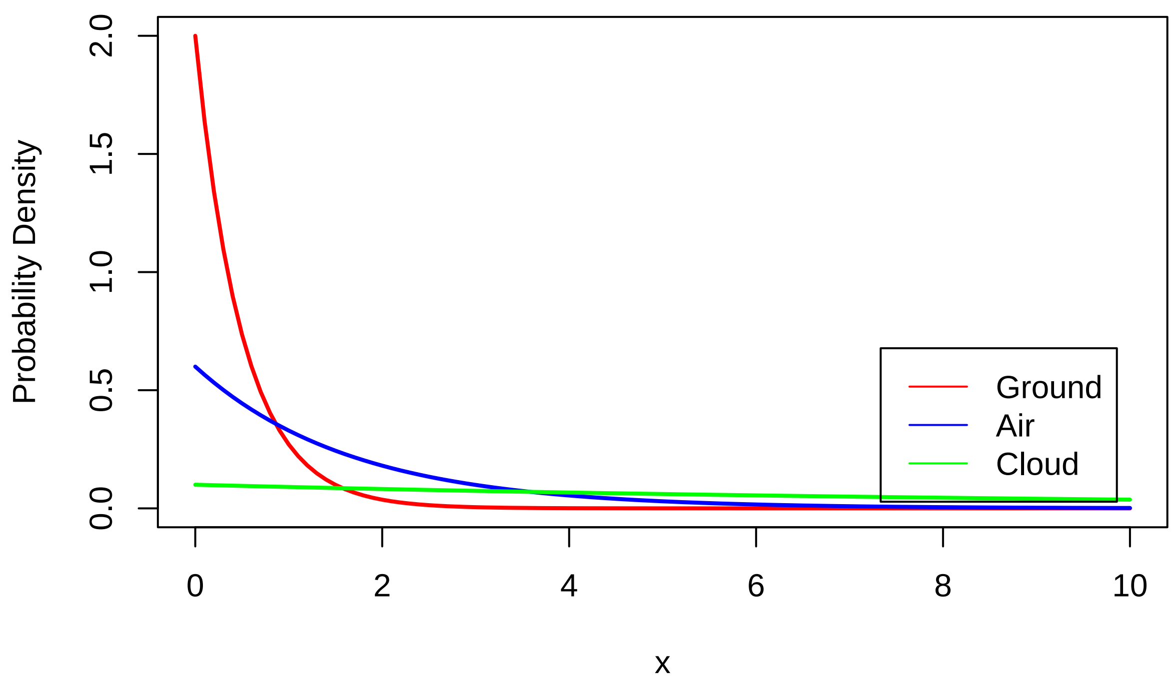
```
## [1] 0.7029253
```

0.7029253 is the probability that the company receives more than 10 emails total in an hour.

6. Suppose that we are interested in the length of time before the next lightning strike. There are three types of lightning we are interested in: cloud-to-ground (G), cloud-to-air (A), and cloud-to-cloud (C). For all types of lightning, the length of time before the next strike is distributed according to an exponential distribution, but the exponential distribution has a different parameter for each type of lightning. In particular, $\lambda_G = 2$, $\lambda_A = 0.6$, and $\lambda_C = 0.1$.

    a. On a single plot, visualize the PDFs over the range $x \in [0, 10]$ for each of these exponential distributions. It may be helpful to use the function "dexp" in R.

```
curve(dexp(x,rate=2),from=0, to=10, lwd=2, col='red', ylab = "Probability Density")
curve(dexp(x,rate=0.6),from=0, to=10, lwd=2, col='blue', add = TRUE)
curve(dexp(x,rate=0.1),from=0, to=10, lwd=2, col='green', add = TRUE)
legend('bottomright',inset=0.05,c("Ground","Air","Cloud"),lty=1,col=c("red","blue","green"))
```




```
b. What are $E(G)$, $E(A)$, and $E(C)$, as well as $Var(G)$, $Var(A)$,
and $Var(C)$?
```

```
lam_g = 2
lam_a = 0.6
lam_c = 0.1
```

```r
Eg = 1/lam_g
Ea = 1/lam_a
Ec = 1/lam_c
VarG = 1/(lam_g)^2
VarA = 1/(lam_a)^2
VarC = 1/(lam_c)^2

Eg
```

```
## [1] 0.5
```

```r
Ea
```

```
## [1] 1.666667
```

```r
Ec
```

```
## [1] 10
```

```r
VarG
```

```
## [1] 0.25
```

```r
VarA
```

```
## [1] 2.777778
```

```r
VarC
```

```
## [1] 100
```

$E(G) = 0.5$, $E(A) = 1.666667$, and $E(C)$ =10 $Var(G) = 0.25$, $Var(A)$=2.777778, and $Var(C)$=100

```
c. Suppose that we repeatedly sample collections of $n = 100$ observations
from the distribution of cloud-to-ground ($G$) lightning strike timings.
What is the mean and variance of this sample distribution?
```

```r
ans = rexp(100, rate=2)
mean(ans)
```

```
## [1] 0.4341005
```

```r
var(ans)
```

```
## [1] 0.175653
```

```
Mean is 0.4664007
Variance is 0.2616048
```

```r
#Considering lambda for cloud-to-ground.
lamG = 2
varG = 1/(lamG)^2
varG
```

```
## [1] 0.25
```

```r
meanG = 1/lamG
meanG
```

```
## [1] 0.5
```

```r
n=100
Sample_Variance = varG/n
Sample_Variance
```

```
## [1] 0.0025
```

```
d. Now, let us examine the empirical sampling distribution of cloud-to-ground
($G$) lightning. For each value of $n = \{10, 100, 1000\}$, sample $n$ points
from the exponential distribution from $G$ a grand total of $m = 5000$ times,
and record the mean of each sample. You should end up with three different
sets of 5000 sample means. For each of these sets, report the sample mean
and sample standard deviation. Comment on how the observed values line up
with what you would expect in theory. It may be useful to use the R function
``rexp()".
```

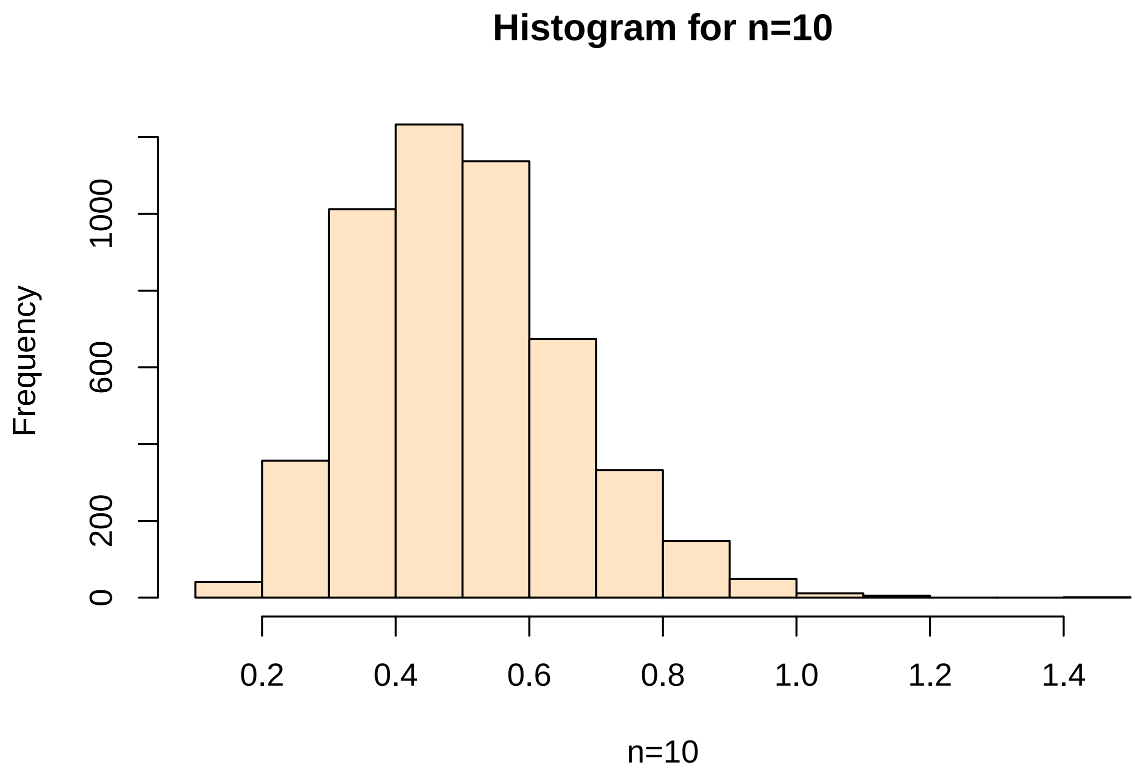
```r
#For n = 10
set.seed(5000 *10)
ss10 = NULL
for (i in 1:5000) {
  ss10 <- c(ss10 , mean(rexp(10,2)))
}
mean(ss10)
```

```
## [1] 0.5018218
```

```r
sd(ss10)
```

```
## [1] 0.1565412
```

```r
hist(ss10, main="Histogram for n=10", xlab="n=10", col="bisque")
```



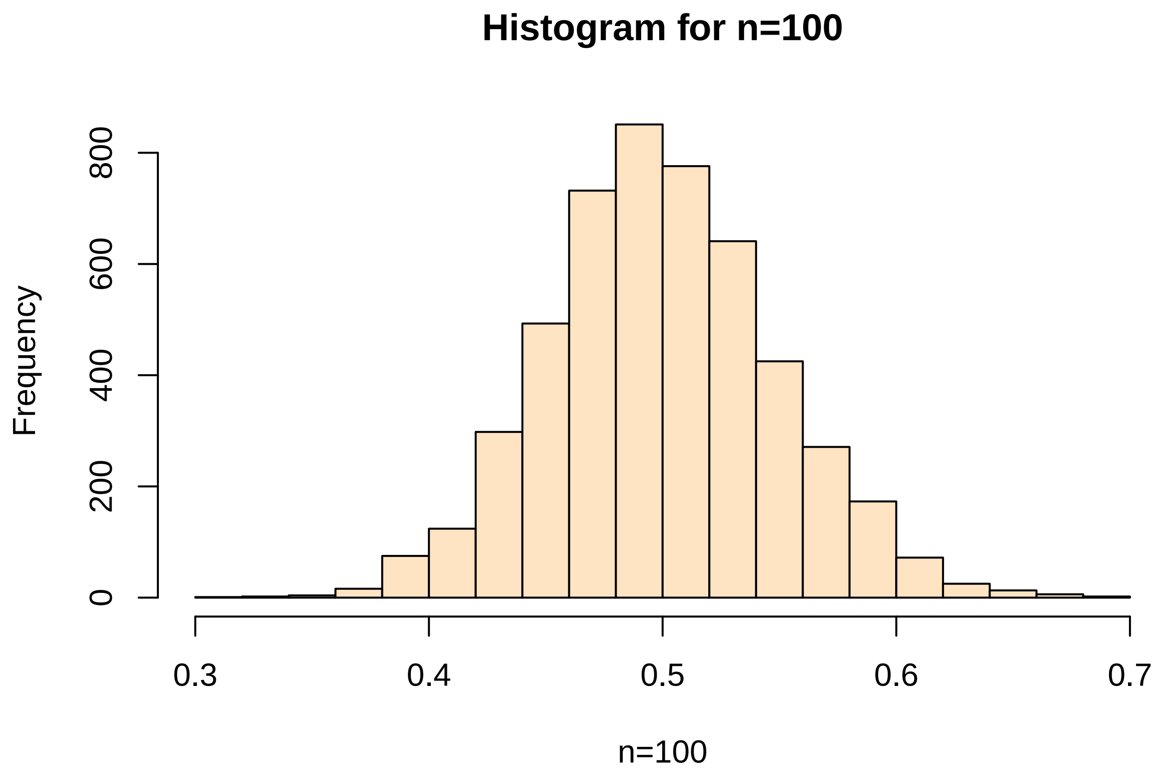

```r
#For n = 100
set.seed(5000 *100)
ss100 = NULL
for (i in 1:5000) {
  ss100 <- c(ss100 , mean(rexp(100,2)))
}
mean(ss100)
```

```
## [1] 0.499667
```

```r
sd(ss100)
```

```
## [1] 0.04881372
```

```r
hist(ss100, main="Histogram for n=100", xlab="n=100", col="bisque")
```



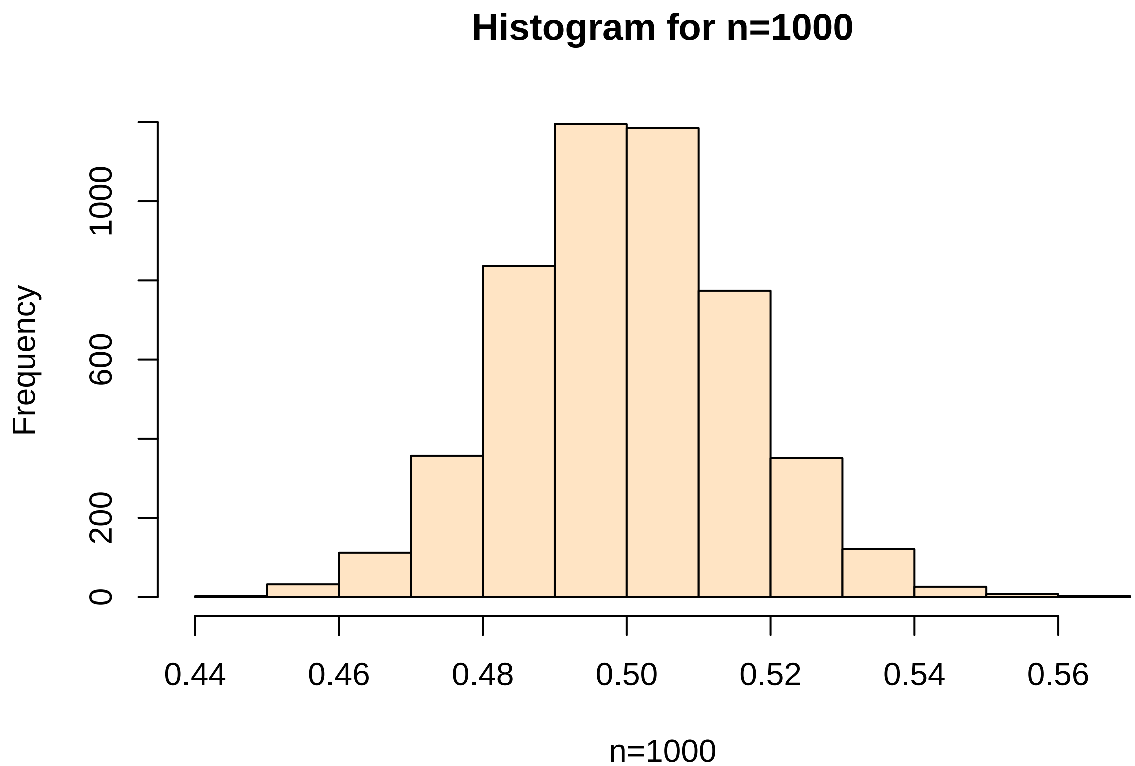

```r
#For n = 1000
set.seed(5000 *1000)
ss1000 = NULL
for (i in 1:5000) {
  ss1000 <- c(ss1000 , mean(rexp(1000,2)))
}
mean(ss1000)
```

```
## [1] 0.499878
```

```r
sd(ss1000)
```

```
## [1] 0.01594702
```

```r
hist(ss1000, main="Histogram for n=1000", xlab="n=1000", col="bisque")
```



For n=10 the histogram is skewed and according to the theory we know for n<30, shape won't be normal. As the n increases to 100 and 1000, we can observed shape of distribution is normal. Also we know that further the population is from normal, the larger the sample size we require to ensure normality of the sampling distribution. Here, we have taken random population so it will require large sample size like n=100 and 1000 to ensure normality of sampling distribution. Also, we can observe that sd is decreases as n increasesas n increases which means increase in accuracy. Also theoretical mean was 0.50 which is almost equal to mean we got from 3 histograms. For n=1000, we observed bimodal histogram.

7. Suppose that we would like to get an idea of how much coffee is consumed by the entire University of Rochester each day. We take a sample of 100 days and find that the average amount of coffee consumed by the University of Rochester per day is 580 gallons.

   a. Assume that coffee consumption comes from a normal distribution with $\sigma = 90$. Find a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day.
   b. Assuming the same information as part a, suppose that we now only want a upper-bound confidence interval. Calculate a one-sided 95% upper-bound confidence interval for the average amount of coffee consumed by the University of Rochester each day.
   c. Now, suppose that we do not know the variance of the true distribution of coffee consumption. However, in our sample, we see that $s = 80$. Find a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day.
   d. Assuming the same information as part c, suppose that we now only want a upper-bound confidence interval. Calculate a one-sided 95% upper-bound confidence interval for the average amount of coffee consumed by the University of Rochester each day.
   e. Assuming the same information as part a (i.e., known population variance), calculate the number of samples needed in order to get a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day of length 16.

A7). a) $\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$

$\sigma = 90$

$n = 100.$

$z_{\alpha/2} = 1.96$

$\bar{x} = 580$

a). $\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 580 - 1.96 * \frac{90}{\sqrt{100}}$

$= 580 - 1.96 * 9$

$= 562.36.$

$\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 580 + 1.96 * \frac{90}{\sqrt{100}}$

$= 580 + 1.96 * 9$

$= 597.64$

Conf interval = $(562.36, 597.64)$

b) $\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$

$z_{\alpha} = 1.645$

$\bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}} = 580 + 1.64 \times \frac{90}{10}$

$= 594.805$

$(-\infty, 594.805)$.

c) $s = 80$

t-test 1.98 for degree of freedom 99

$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} = 580 - 1.98 \times \frac{80}{\sqrt{100}}$

$= 580 - 1.98 \times 8$

$= 564.16$

$\bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}} = 580 + 1.98 \times \frac{80}{\sqrt{100}}$

$= 580 + 1.98 \times 8$

$= 595.84$

$(564.16, 595.84)$

d) $\left(-\infty, \ \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right)$

$t_\alpha = 1.66$

$\bar{x} + t_\alpha \frac{s}{\sqrt{n}} = 580 + 1.66 \times \frac{80}{\sqrt{100}}$

$= 580 + 1.66 \times 8$

$= 593.28$

$(-\infty, 593.28)$

e) $l = 16$; $m = \frac{l}{2}$

$m = 8$

$n = \left\lceil z^2_{\alpha/2} \frac{\sigma^2}{m^2} \right\rceil$

$= \left\lceil 1.96^2 \times \frac{90^2}{8^2} \right\rceil = \lceil 486.2025 \rceil$

$= \underline{\underline{487}}$

Number of samples $\approx \underline{487}$

Short Answers:

- About how long did this assignment take you? Did you feel it was too long, too short, or reasonable?
  Around 7-8 hours, it was reasonable
- Who, if anyone, did you work with on this assignment? No one
- What questions do you have relating to any of the material we have covered so far in class?
  Binomial, Poisson, Exponential Distributions; Central Limit Theorem ; Confidence Interval