

# Probability and Statistics for Engineers Homework Two

## TMATH 390

Ben Foster\*  
Instructor: Julia Eaton

April 16, 2015

---

\*Institute of Technology, University of Washington Tacoma

## Contents

First Problem	1
Second Problem	2
Third Problem	3
Fourth Problem	4
Fifth Problem	5
Sixth Problem	5
Seventh Problem	6
Eighth Problem	8
Ninth Problem	9
Tenth Problem	9
Eleventh Problem	12
Lab 2 Printouts	13

## Problem 1

On her way to work, a friend of ours must pass through ten traffic signals. Suppose that in the long run, she encounters a red light at 40% of these signals and that whether any particular signal is red is independent of whether any other one is red. (Note: what distribution could you model this using? Is there an analogy of the red/green outcome with another variable that has a binary outcome?)

- (a) On what proportion of days will our friend encounter at most two red lights?
- (b) On what proportion of days will our friend encounter at least five red lights?
- (c) On what proportion of days will our friend encounter between three and five (inclusive) red lights?

**Note:** We should model this using a binomial distribution. We could compare this with the flip of a coin if the probability of getting heads in the long run was 40% since each individual coin flip result is also independent of one another and it is a binary answer. Also note that we should not use the poisson approximation because we have a relatively large  $n$  and a relatively small  $p$ .

**Answer to a** Using R, we would determine the answer by typing in the command `pbinom(2, 10, 0.4)` since we know that this is a binomial distribution. Here's how to do it by hand using the equation  $p(x) = \frac{n!}{(n-x)!x!} * p^x * (1-p)^{n-x}$  where  $n$  = the number of cases there are,  $x$  = the number of successful encounters that day and  $p$  = the probability of getting a success:

$$p(x \leq 2) = \sum_{x=0}^2 \frac{10!}{(10-x)!x!} * 0.4^x * (1-0.4)^{10-x}$$

After doing the math required, we get an answer of: 0.167.

**Answer to b** Using R, we would use a similar command as the first problem: `sum(dbinom(5:10, 10, 0.4))`. By hand we would also use a similar method, except this time, we would use the equation:

$$p(x \geq 5) = \sum_{x=5}^{10} \frac{10!}{(10-x)!x!} * 0.4^x * (1-0.4)^{10-x}$$

After doing the math, we get an answer of 0.367.

**Answer to c** Using R, we would use a similar command as the first and second problems: `sum(dbinom(3:5, 10, 0.4))`. By hand, we would also use a similar method, except this time, we would use the equation:

$$p(3 \leq x \leq 5) = \sum_{x=3}^5 \frac{10!}{(10-x)!x!} * 0.4^x * (1-0.4)^{10-x}$$

After doing the math, we get an answer of 0.666.

## Problem 2

Suppose that the number of drivers who travel between a particular origin and destination during a designated time period has a Poisson distribution with parameter  $\lambda = 8$ . In the long run, in what proportion of time periods will the number of drivers

- (a) Be at most 4?
- (b) Exceed 8?
- (c) Between 6 and 10 (inclusive)?

**Answer to a** To solve this in R, we would input the command: `ppois(4,8)`. By hand, we would use the equation  $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$  where  $\lambda = 8$ , the average number of drivers in this particular region and  $x$  = the number of drivers and then sum it all together:

$$p(x \leq 4) = \sum_{x=0}^4 \frac{e^{-8}8^x}{x!}$$

Once you sum up all of the necessary equations, then you get an answer of: 0.0996

**Answer to b** To solve this in R, we would use a similar command as part a: `1 - ppois(8, 8)`. By hand we would use the same equation except this time we are going to add together the Poisson mass functions with  $x \leq 8$  and then subtract that answer from 1 in order to get the proportion with  $x > 8$ :

$$p(x \leq 8) = \sum_{x=0}^8 \frac{e^{-8}8^x}{x!} = 0.593$$
$$p(x > 8) = 1 - p(x \leq 8) = 0.407$$

After doing the math, we can see that the answer is 0.407.

**Answer to c** To solve this in R, we would use a similar command as part a and part b: `sum(dpois(6:10))`. By hand, we would use the same equation except this time we are going to just sum from  $x = 6$  to  $x = 10$ .

$$p(6 \leq x \leq 10) = \sum_{x=6}^{10} \frac{e^{-8}8^x}{x!}$$

After doing the math, we get an answer of: 0.625.

### Problem 3

Let  $x$  denote the distance (m) that an animal moves from its birth site to the first territorial vacancy it encounters. Suppose that for banner-tailed kangaroo rats,  $x$  has an exponential distribution with parameter  $\lambda = 0.01386$ .

- (a) What proportion of distances are at most 80 m?
- (b) What proportion of distances are at least 60 m?
- (c) What is the median distance?

**Answer to a** To find the answer using R (which is the preferred method), you would input the command: `sum(dexp(0:80, 0.01386))`. To do this by hand, we have to solve the exponential distribution equation:

$$p(x \leq 80) = \sum_{x=0}^{80} 0.01386 * e^{-0.01386*x}$$

Which equals 0.679.

**Answer to b** To find this answer using R, you would input a similar command: `1 - sum(dexp(0:60, 0.01386))`. To do this by hand we would have to solve the equation:

$$p(x \geq 60) = \sum_{x=60}^{\infty} 0.01386 * e^{-0.01386*x}$$

After doing the math, you get 0.435.

**Answer to c** To find this answer using R, you would input the command: `qexp(0.5, 0.01386)`. To do this by hand you would have to solve the equation:

$$p(x) = \sum_{x=0}^{med} 0.01386 * e^{-0.01386*x} = \frac{1}{2}$$

Where  $med$  is the median. After doing the math, you get an answer of 50.01.

## Problem 4

Suppose that 10% of all bits transmitted through a digital communication channel are erroneously received and that whether any particular bit is erroneously received is independent of whether any other bit is erroneously received. Consider sending a very large number of messages, each consisting of 20 bits.

- (a) What proportion of these messages will have at most 2 erroneously received bits?
- (b) What proportion of these messages will have at least 5 erroneously received bits?

**Answer to a** This problem can be solved using the Binomial Distribution equation:  $p(x) = \frac{n!}{(n-x)!x!} * p^x * (1-p)^{n-x}$  where  $n$  = the number of bits in each message,  $p$  = the probability of a bit is erroneously received and  $x$  = the number of bits erroneously received. To do this problem in R, we would input the command: `sum(dbinom(0:2,20,0.1))`. To do this by hand, we can input the variables to get the equation:

$$p(x \leq 2) = \sum_{x=0}^2 \frac{20!}{(20-x)!x!} * 0.1^x * (1-0.1)^{20-x}$$

From this, we get the answer:  $p(x \leq 2) = 0.677$ .

**answer to b** This problem is similar to the one above. Since we are finding the proportion of messages that have *at least* 5 erroneous bits, then it would be easier to just find the proportion of message with less than 5 erroneously received bits and then subtract one. In R, all we would have to do is input the command: `sum(dbinom(5:20,20,0.1))`. To do this by hand, we would use the equation:

$$p(x \geq 5) = 1 - \left( \sum_{x=0}^4 \frac{20!}{(20-x)!x!} * 0.1^x * (1-0.1)^{20-x} \right)$$

From this, we get the answer:  $p(x \geq 5) = 0.0432$ .

## Problem 5

Consider writing onto a computer disk and then sending it through a certifier that counts the number of missing pulses. Suppose this number  $x$  has a Poisson distribution with parameter  $\lambda = 0.2$ .

- (a) What proportion of disks have exactly one missing pulse?
- (b) What proportion of disks have at least two missing pulses?

**Answer to a** To answer this problem in R, we would input the command: `dpois(1, 0.2)`. By hand, we would use the equation

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Where  $\lambda = 0.2$  and  $x$  = the number of missing pulses. Filling in the variables of the equation, we end up with an answer of 0.164.

**Answer to b** To answer part b in R, we would take a somewhat similar approach, except this time, we would type in the command: `1 - sum(dpois(0:1, 0.2))`. We subtracted the value from 1 because we're looking for the proportion of *at least* two missing pulses. By hand, we would just use the equation:

$$p(x \geq 2) = \sum_{x=0}^1 \frac{e^{-\lambda} \lambda^x}{x!}$$

From this, we get an answer of 0.0175.

## Problem 6

The bursting strength of wine bottles of a certain type is normally distributed with parameters  $\mu = 250$  psi and  $\sigma = 30$  psi. If these bottles are shipped 12 to a carton, in what proportion of cartons will at least one of the bottles have a bursting strength exceeding 300 psi? Hint: think of a bottle as a success, or "heads"  $H$  if its bursting strength exceeds 300 psi.

**Answer** The answer to this question is a little complicated. Following the hint, we would use a binomial distribution to show the proportion of cartons where at least one wine bottle has burst. However, we need the probability that a wine bottle will burst in a carton. We can find this by using the normal distribution and finding the proportion of bottles that burst above 300 psi. Therefore, in R, we would input the command: `1 - dbinom(0, 12, (1 - pnorm(300, 250, 30)))`.

To do this by hand, we would first have to use the equation  $z = \frac{x-\mu}{\sigma}$  so that we can get the current normal distribution to a standard normal distribution and then we can just use our handout to get the value for when  $x = 300$ . From here, we use the number we got from the normal distribution, subtract it from 1, and then use that for the probability in the equation:

$$p(x \geq 1) = 1 - \frac{12!}{(12-0)!0!} * 0.478^0 * (1 - 0.478)^{12-0}$$

Where we said  $n = 12$ ,  $x = 0$  and  $p = (1 - \text{pnorm}(300, 250, 30))$ . From this, we get an answer of 0.444.

## Problem 7

Here are some data: 115.8, 115.2, 114.6, 115.9, 116.4.

- (a) Compute the mean of the deviations from the sample mean. (Note: the deviation of observation  $x_i$  from the mean  $\bar{x}$  is  $(x_i - \bar{x})$ .)  
 (b) Compute the sample standard deviation using the defining formula.  
 (c) Compute it using the computational formula.  
 Setting up tables like this will help:

For parts (a) and (b):

$i$	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	115.8		
2	115.2		
3	114.6		
4	115.9		
5	116.4		
$n = 5$	$\sum_{i=1}^5 x_i =$ $\bar{x} =$	$\sum_{i=1}^5 (x_i - \bar{x}) =$ divide the above by $n$ for (a)	$\sum_{i=1}^5 (x_i - \bar{x})^2 =$ divide the above by $n - 1$ for (b)

For part (c): Combine the two sums you computed using the "computation" formula to get (c).

$i$	$x_i$	$(x_i)^2$
1	115.8	
2	115.2	
3	114.6	
4	115.9	
5	116.4	
	$\sum_{i=1}^5 x_i =$ $(\sum_{i=1}^5 x_i)^2 =$	$\sum_{i=1}^5 (x_i)^2 =$

**Answer to a and b** To find the answers to these two, we will fill in the first table above. Also be sure to take the square root of  $\frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{4}$  so that we get the sample standard deviation: From the

$i$	$x_i$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$
1	115.8	0.22	0.0484
2	115.2	0.38	0.1444
3	114.6	0.98	0.9604
4	115.9	0.32	0.1024
5	116.4	0.82	0.6724
$n = 5$	$\sum_{i=1}^5 x_i = 577.9$ $\bar{x} = 115.58$	$\sum_{i=1}^5 (x_i - \bar{x}) = 2.72$ (a) = 0.544	$\sum_{i=1}^5 (x_i - \bar{x})^2 = 1.928$ (b) = 0.694



table we can see that the mean of the deviations from the sample mean is -0.72 and the sample standard deviation is 1.13.

**Answer to c** To find the answer to c, we will fill in the second table above.

$i$	$x_i$	$(x_i)^2$
1	115.8	13409.64
2	115.2	13271.04
3	114.6	13133.16
4	115.9	13432.81
5	116.4	13548.96
	$\sum_{i=1}^5 x_i = 577.9$ $(\sum_{i=1}^5 x_i)^2 = 333968.41$	$\sum_{i=1}^5 (x_i)^2 = 66795.61$

The computational form for the standard deviation is:

$$\begin{aligned}s_x &= \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^5 x_i^2 - \frac{1}{n} \left( \sum_{i=1}^5 x_i \right)^2 \right)} \\&= \sqrt{\frac{1}{5-1} \left( 66795.61 - \frac{333968.41}{5} \right)} \\&= 0.694\end{aligned}$$

## Problem 8

Recall the uniform distribution, with the the density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

The corresponding density "curve" has constant height over the interval from  $a$  to  $b$ .

Compute the mean of the uniform distribution in terms of  $a$  and  $b$ .

**Answer** Since the mean is the average of all the numbers and a uniform distribution has the same height throughout the entire graph, then the mean distribution in terms of  $a$  and  $b$  is:

$$\frac{a+b}{2}$$

Here's a more in-depth way of describing it:

$$\begin{aligned} E[X] &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \left[ \frac{x^2}{2} \Big|_a^b \right] \\ &= \frac{1}{2(b-a)} (b^2 - a^2) \\ E[X] &= \frac{b+a}{2} \end{aligned}$$

## Problem 9

Find the mean and the variance of the number of under-inflated tires, if the mass density is  $p(0) = 0.4, p(1) = p(2) = p(3) = 0.1, p(4) = 0.3$ . For what proportion of such cars will the number of under-inflated tires be within 1 standard deviation of the mean?

**Answer** In order to find the mean of the number of under-inflated tires, we have to use the equation:

$$\begin{aligned} \text{mean} = \mu_x &= \sum_{k=0}^n k * p(k) \\ &= 0 * p(0) + 1 * p(1) + 2 * p(2) + 3 * p(3) + 4 * p(4) \\ &= 1.8 \end{aligned}$$

In order to find the variance of the number of under-inflated tires, we have to use the equation:

$$\begin{aligned} \text{variance} = \sigma^2 &= \sum_{k=0}^n (k - \mu_x)^2 * p(k) \\ &= (0 - 1.8)^2 * p(0) + (1 - 1.8)^2 * p(1) + (2 - 1.8)^2 * p(2) + (3 - 1.8)^2 * p(3) + (4 - 1.8)^2 * p(4) \\ &= 2.96 \end{aligned}$$

With standard deviation being  $\sigma = \sqrt{\text{variance}} = 1.72$ . Now we have to find the proportion of cars such that the number of inflated tires will be within 1 standard deviation of the mean. This means we need to find the proportion of cars that have between 0.08 and 3.52 under-inflated tires. Since we cant have 0.08 of a tire or 3.52 of a tire, then we are going to round down and assume that we are searching for proportion of cars with between 0 and 3 under-inflated tires, inclusive. The proportion of cars with between 0 and 3 under-inflated tires is:  $p(0) + p(1) + p(2) + p(3) = 0.7$ .

## Problem 10

Recall from Exercise 2 the data on the concentration ( $EU/mg$ ) in settled dust for one sample of urban homes and another of farm homes: (a) Determine the medians, quartiles, and IQRs for the

U:	6.0	5.0	11.0	33.0	4.0	5.0
	80.0	18.0	35.0	17.0	23.0	
F:	4.0	14.0	11.0	9.0	9.0	8.0
	4.0	20.0	5.0	8.9	21.0	9.2
	3.0	2.0	0.3			

two samples.

(b) Are there any outliers in either sample? Any extreme outliers?

(c) Construct a comparative boxplot and use it as a basis for comparing and contrasting the two samples. Use R.

**Answer to a** Using R, finding the medians, quartiles, and iQRs for these two samples is very easy. All we have to do is put in the series of commands in R:

```
> u <- c(6.0, 5.0, 11.0, 33.0, 4.0, 5.0, 80.0, 18.0,
35.0, 17.0, 23.0)
> u <- sort(u)
> quantile(u)
> IQR(u)

> f <- c(4.0, 14.0, 11.0, 9.0, 9.0, 8.0, 4.0, 20.0,
5.0, 8.9, 21.0, 9.2, 3.0, 2.0, 0.3)
> f <- sort(f)      > quantile(f)
> IQR(f)
```

By entering these commands, I got the following answers:

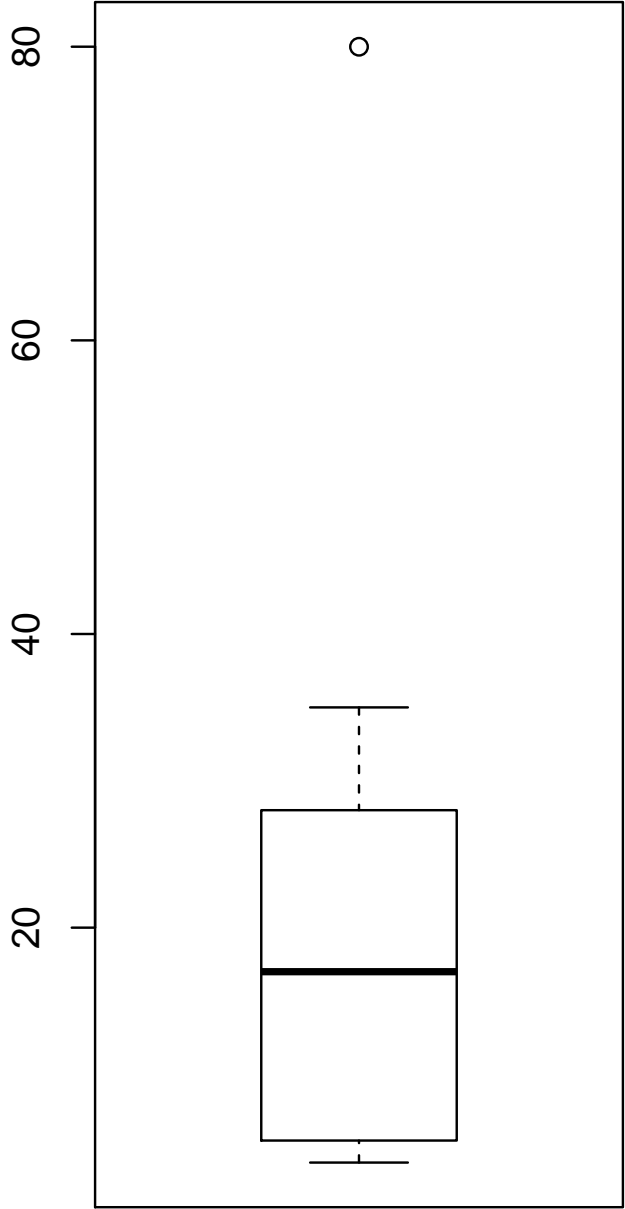
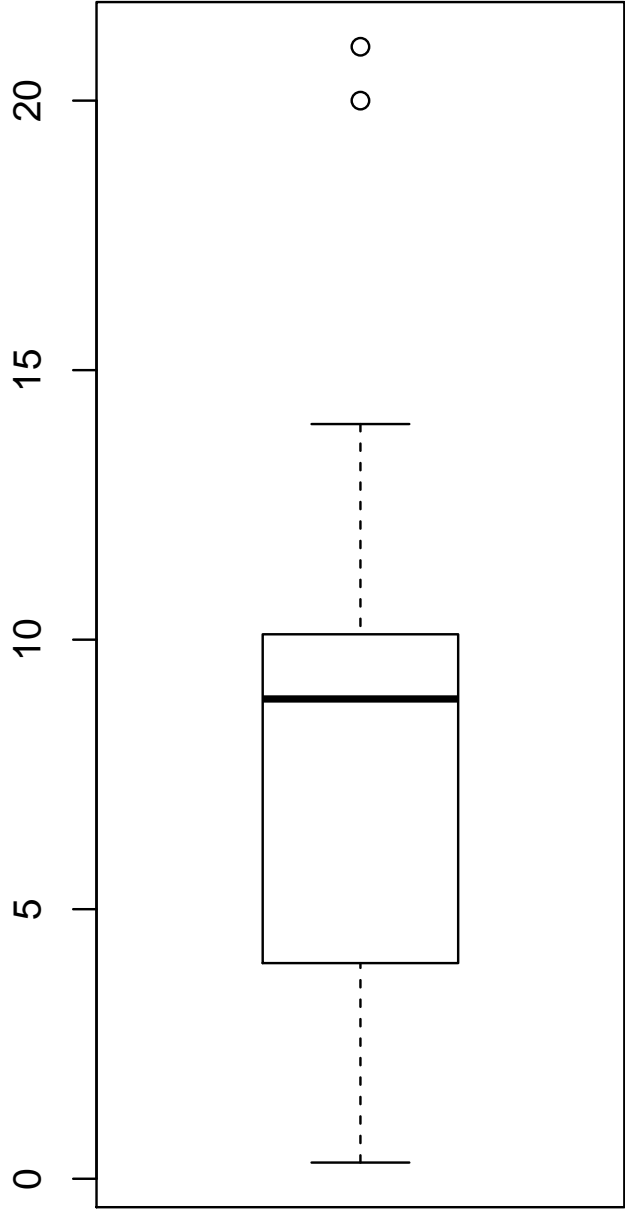
	Min	$Q_1$	Med.	$Q_3$	Max	IQR
U	4.0	5.5	17.0	28.0	80.0	22.5
F	0.3	4.0	8.9	10.1	21.0	6.1

**Answer to b** Since we are doing all of this in R, then it would be easier to tell if there are outliers by constructing the boxplot first and analyzing the data. To do this, we just enter the following commands:

```
> par(mfrow=c(1,2))
> boxplot(u)
> boxplot(f)
```

From the boxplots on the next page, we can see that there is only one extreme outlier for U which is 80.0 and two somewhat extreme outliers for F which are 20.0 and 21.0.

**Answer to c** We already built the boxplots in the previous part. Looking at the two can be a little confusing at first, but once you realize that F has a much smaller scale than U, then you'll see that U has a wider spread than F. U also has a much higher median, a larger interquartile range and a larger maximum. From this, we can see that the concentration of ( $EU/mg$ ) in settled dust around Urban homes is generally much higher than around Farm homes.



## Problem 11

Find the expected value and variance of the exponential distribution with parameter  $\lambda$ . Show work.

**Answer** Find  $E[X]$  and  $V[X]$ . with  $f(x) = \lambda e^{-\lambda x}$ .

$$E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx$$

Begin by doing integration by parts:  $\begin{array}{ll} u = x & dv = \lambda e^{-\lambda x} \\ du = dx & v = \frac{\lambda e^{-\lambda x}}{-\lambda} \end{array}$

$$= x(-e^{-\lambda x})|_0^{\infty} - \int_0^{\infty} -e^{\lambda x} dx$$

$$= x e^{-\lambda x}|_0^{\infty} + \frac{e^{-\lambda x}}{-\lambda}|_0^{\infty}$$

$$x e^{-\lambda x}|_0^{\infty} = (\lim_{b \rightarrow \infty} -b e^{-\lambda b}) - 0 \text{ and } \frac{e^{-\lambda x}}{-\lambda}|_0^{\infty} = \frac{-1}{\lambda} [\lim_{b \rightarrow \infty} e^{-\lambda b} - 1] = \frac{-1}{\lambda}(-1)$$

$$\text{Therefore : } E[X] = \frac{1}{\lambda}$$

Now we want to find  $V[X] = E[X^2] - (E[X])^2$ .

$$E[X^2] = \int_0^{\infty} x^2 f(x) dx$$

$$E[X^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx$$

Once again, do integration by parts:  $\begin{array}{ll} u = x^2 & dv = \lambda e^{-\lambda x} \\ du = 2x & v = \frac{\lambda e^{-\lambda x}}{-\lambda} \end{array}$

$$= \lambda \left[ \frac{-1}{\lambda} x^2 e^{-\lambda x} - \int 2x \frac{e^{-\lambda x}}{-\lambda} dx \right]$$

$$= \lambda \left[ \frac{1}{\lambda} \left[ -x^2 e^{-\lambda x} + 2 \int x e^{-\lambda x} dx \right] \right]$$

$$= -x^2 e^{-\lambda x} + 2 \left[ \frac{-1}{\lambda} x e^{-\lambda x} - \int \frac{e^{-\lambda x}}{-\lambda} dx \right]$$

$$= -x^2 e^{-\lambda x} + 2 \left[ \frac{-1}{\lambda} x e^{-\lambda x} + \frac{1}{\lambda} \int e^{-\lambda x} dx \right]$$

$$= -x^2 e^{-\lambda x} + \frac{2}{\lambda} \left[ -x e^{-\lambda x} + \frac{e^{-\lambda x}}{-\lambda} \right]$$

$$-e^{-\lambda x} \left( x^2 + \frac{2}{\lambda} \left( x + \frac{1}{\lambda} \right) \right) \Big|_0^{\infty}$$

$$= 0 + 1 \left( \frac{2}{\lambda} * \frac{1}{\lambda} \right)$$

$$E[X^2] = \frac{2}{\lambda^2}$$

FINALLY

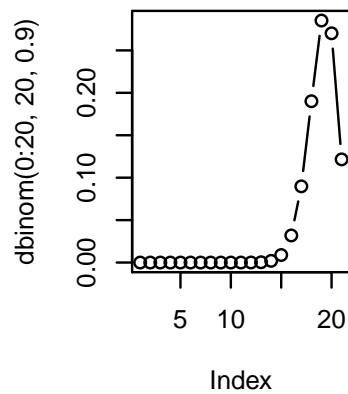
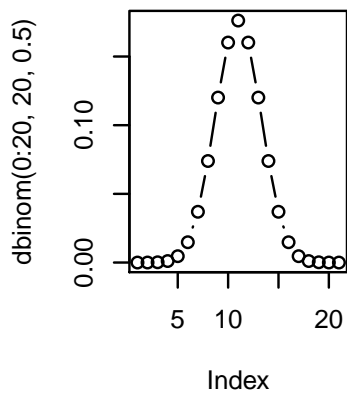
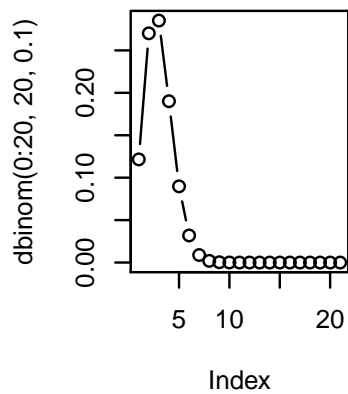
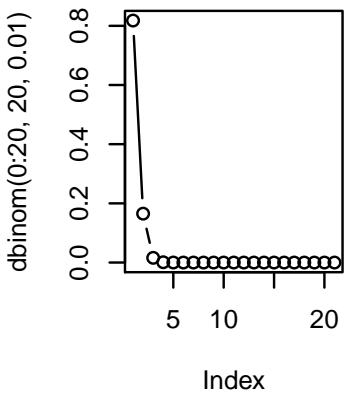
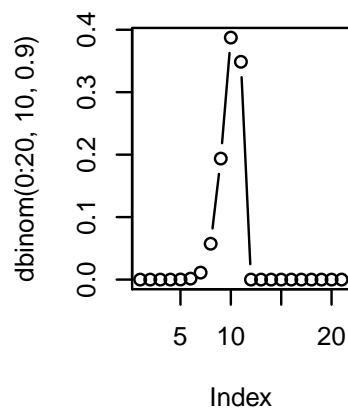
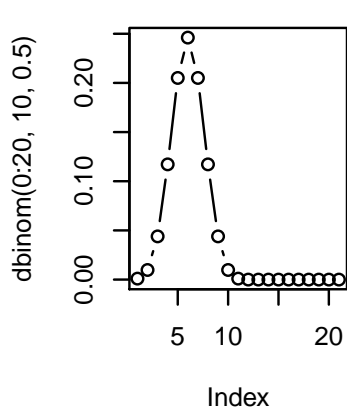
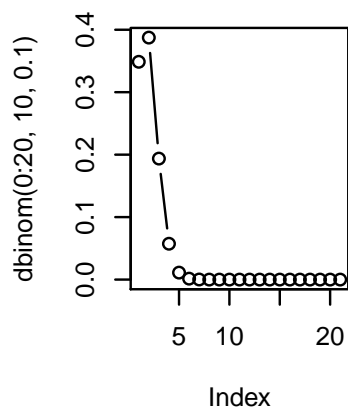
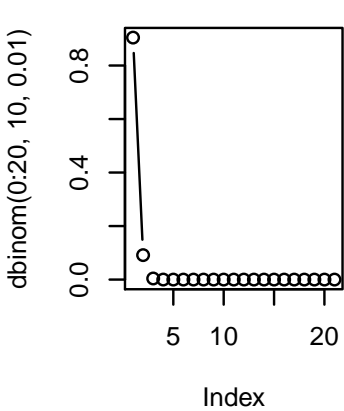
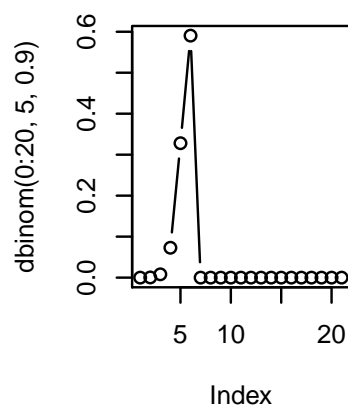
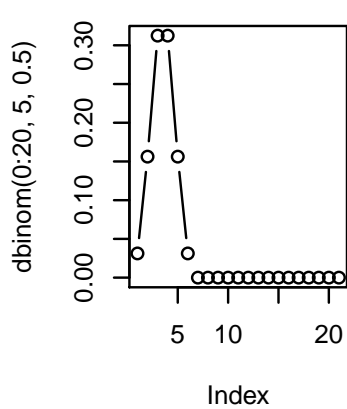
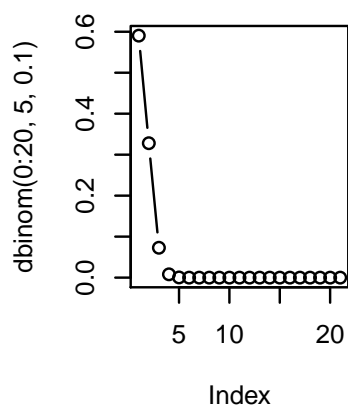
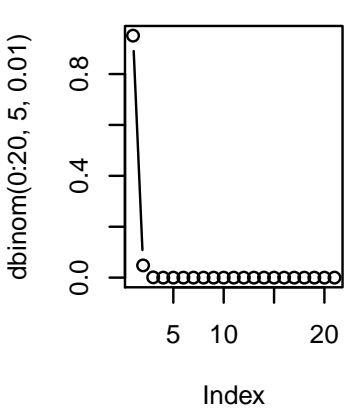
$$V[X] = E[X^2] - (E[X])^2$$

$$V[X] = \frac{2}{\lambda^2} - \left( \frac{1}{\lambda} \right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

$$V[X] = \frac{1}{\lambda^2}$$

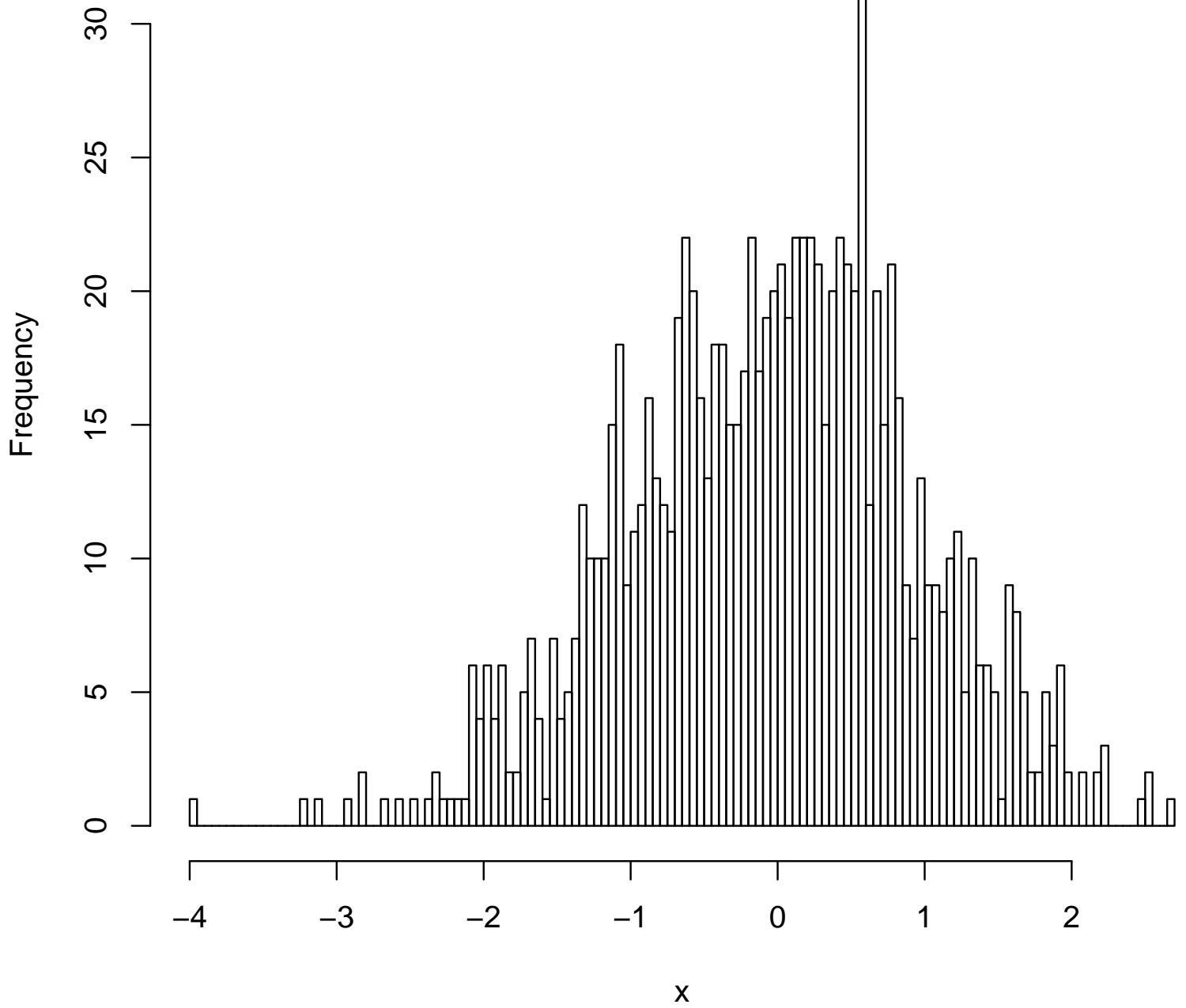
## Lab 2 Printouts

Lab 2 printouts are attached.

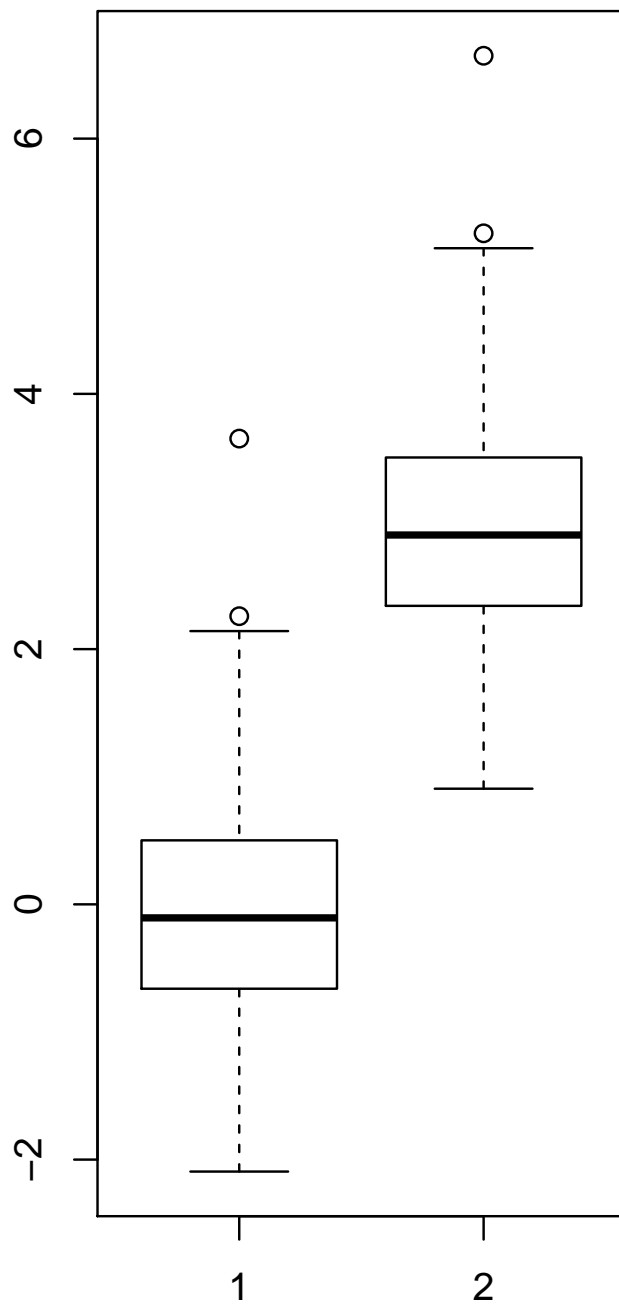
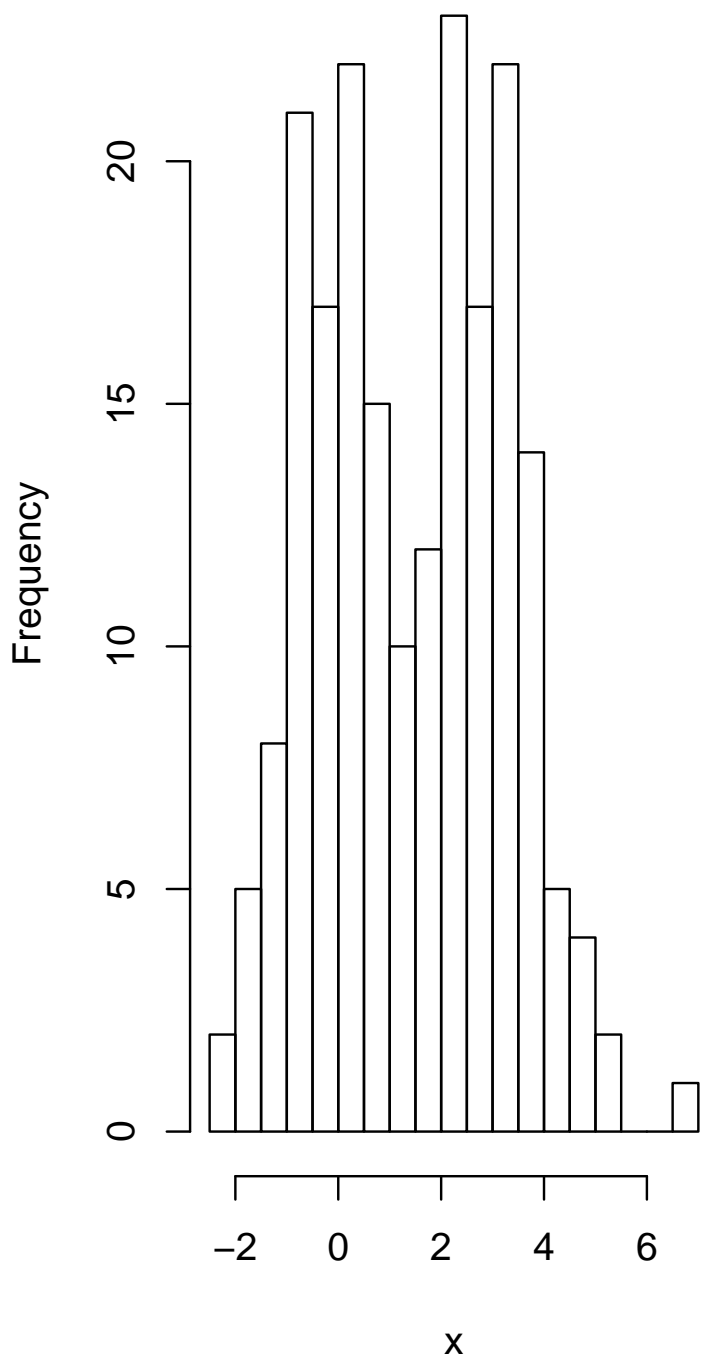




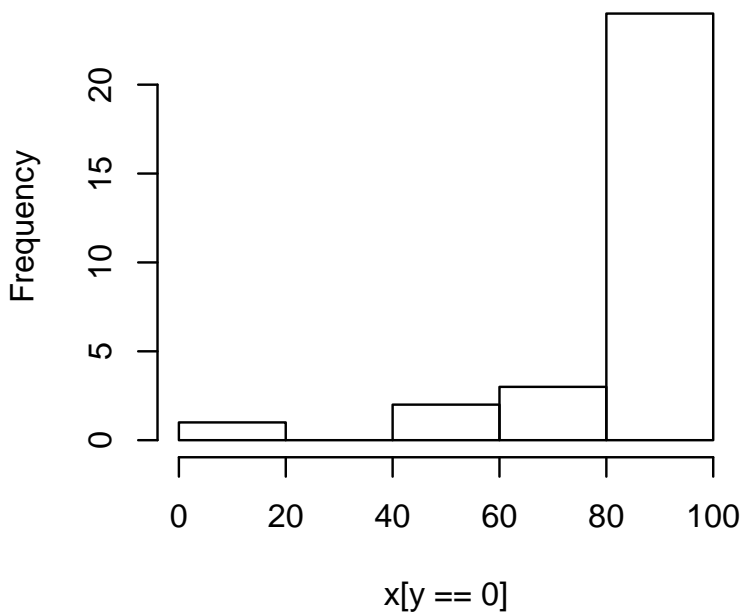
**Histogram of x**



**Histogram of x**



**Histogram of x[y == 0]**



**Histogram of x[y == 1]**

