Raghu_606_wk3_HW

Raghu Ramnath

2/16/2017

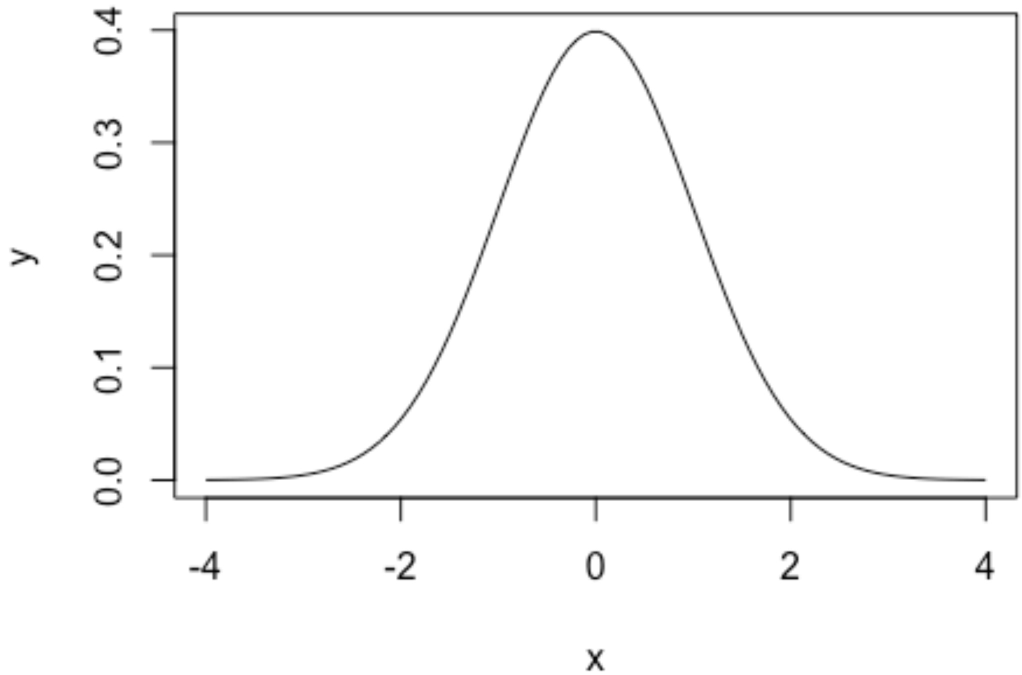
DATA 606 Homework chapter 3 Distribution of Random Variables

3.2 Area under the curve, Part II. What percent of a standard normal distribution N(μ = 0, sd = 1) is found in each region? Be sure to draw a graph.

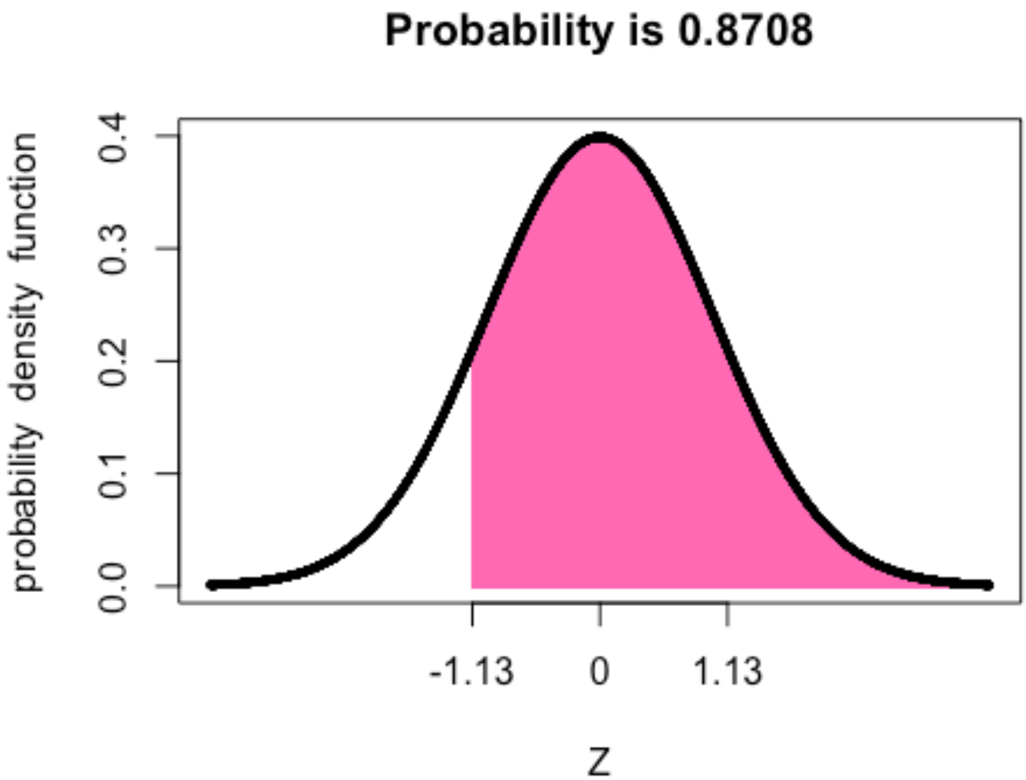
```
require(fastGraph)

## Loading required package: fastGraph

mean<- 0
SD <- 1
x <- seq(-4, 4, length = 100)
y <- dnorm(x, mean, SD)
```

(a) Z > -1.13

```
plot(x,y, type="n")
lines(x, y)
```

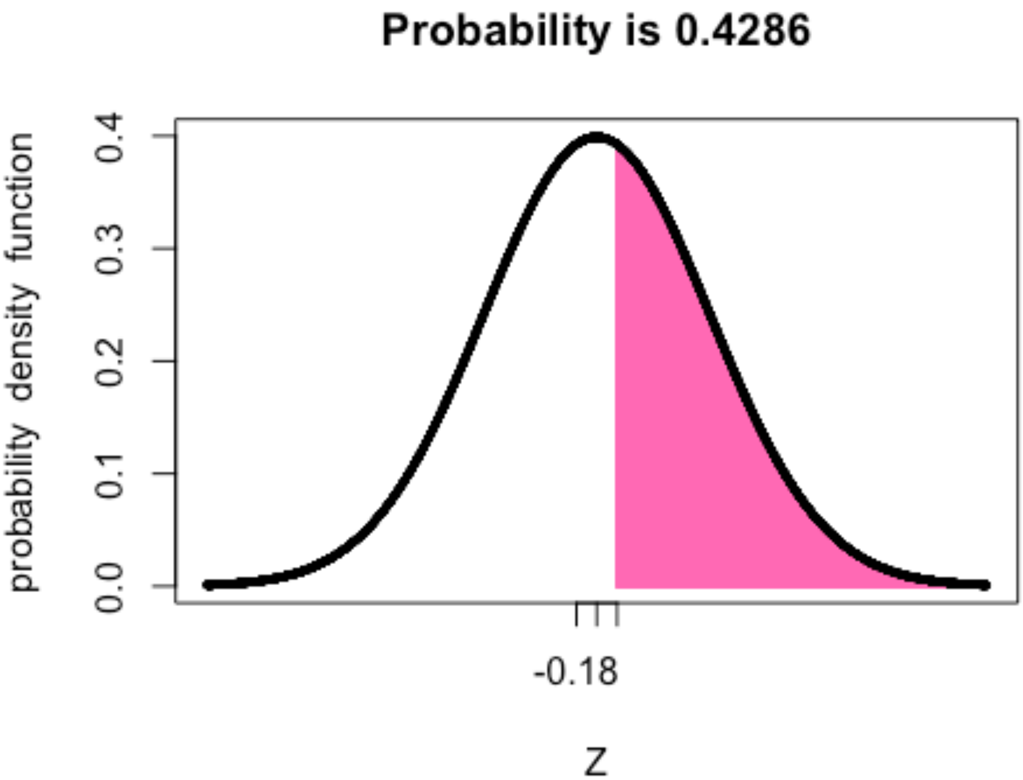



```r
shadeDist(-1.13, lower.tail = FALSE)
```



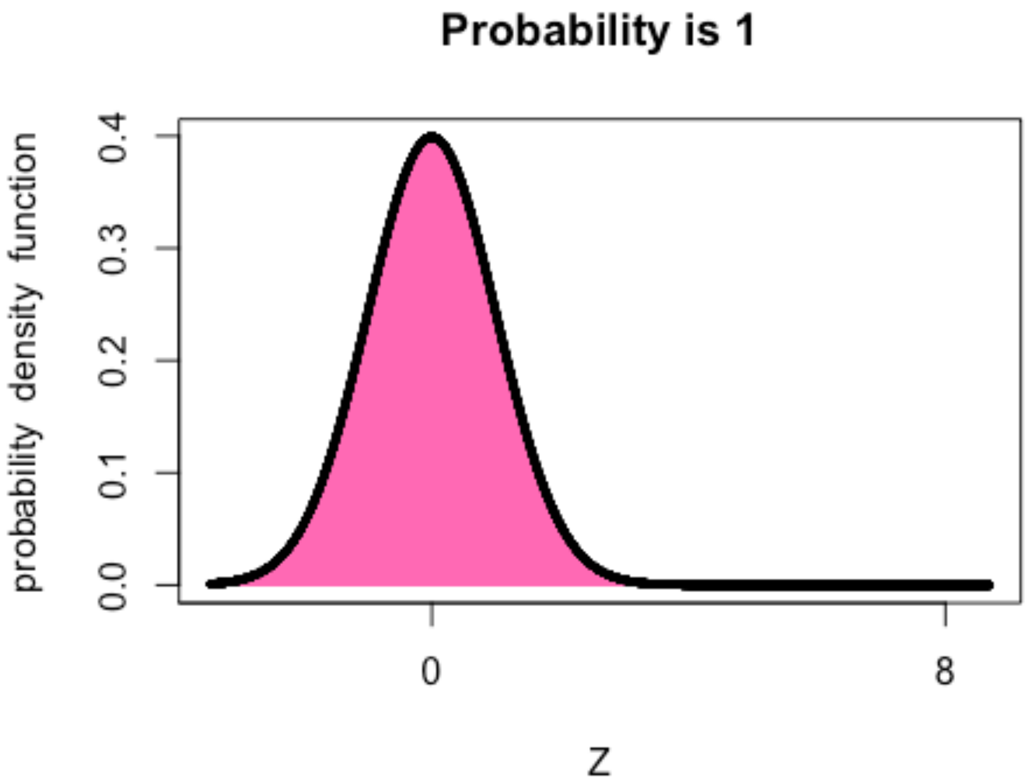
(b) Z > 0.18

```
shadeDist(0.18, lower.tail=FALSE)
```



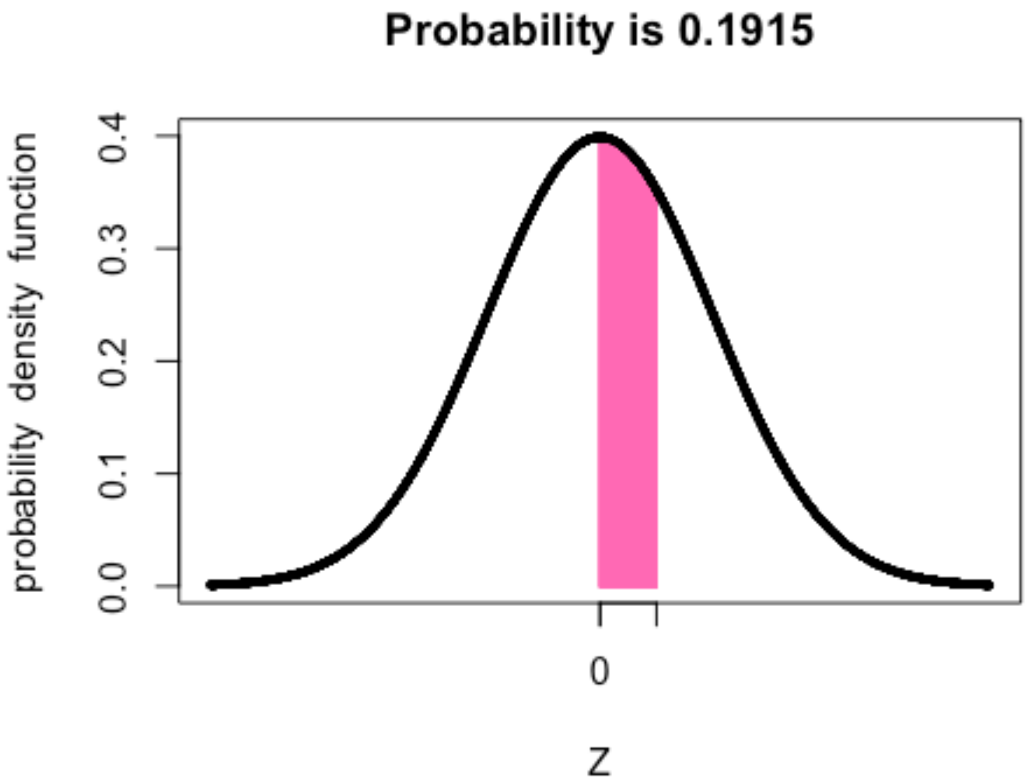
(c) Z > 8

```
shadeDist(8)
```



(d) $|Z| < 0.5$

```
shadeDist( c( 0, 0.5 ),"dnorm" , 0, 1, lower.tail = FALSE )
```



3.4 Triathlon times, Part I.

In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo competed in the Men, Ages 30 - 34 group while Mary competed in the Women, Ages 25 - 29 group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups: • The finishing times of the Men, Ages 30 - 34 group has a mean of 4313 seconds with a standard deviation of 583 seconds. • The finishing times of the Women, Ages 25 - 29 group has a mean of 5261 seconds with a standard deviation of 807 seconds. • The distributions of finishing times for both groups are approximately Normal. Remember: a better performance corresponds to a faster finish.

(a) Write down the short-hand for these two normal distributions.

N(μ=4313, sd=583) for Men and N(μ=5261, sd=807) for women

(b) What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?

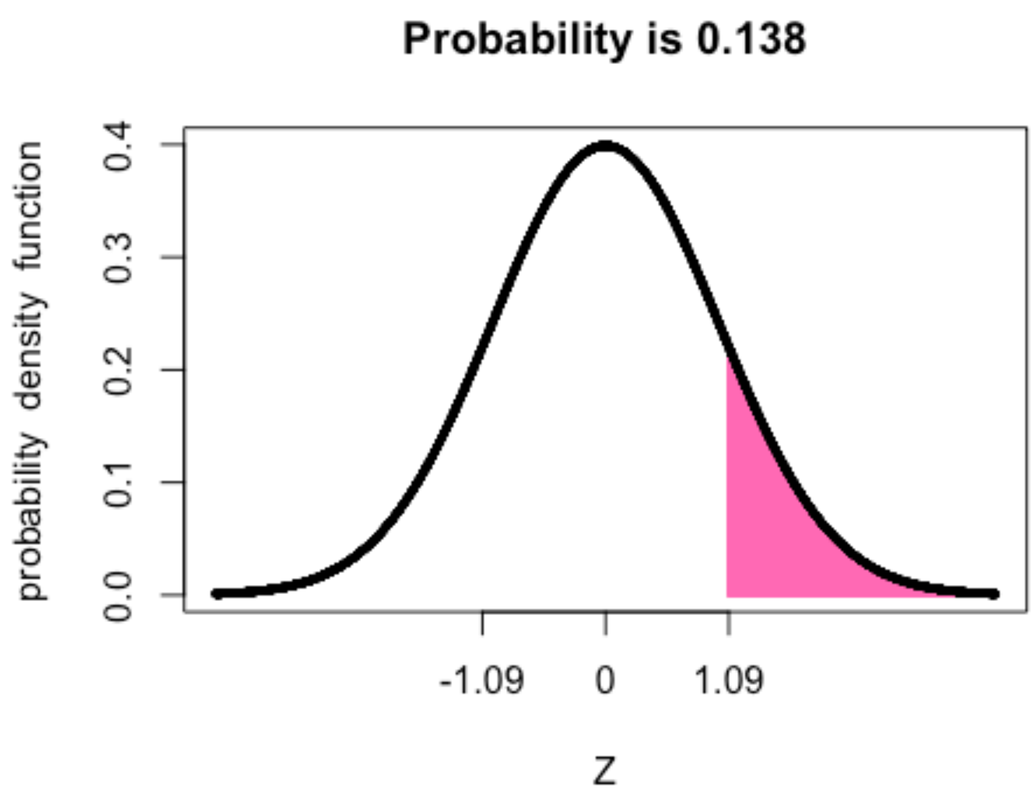
```
Z_Leo <- (4948-4313)/583
Z_Leo

## [1] 1.089194

Z_Mary <- (5513-5261)/807
Z_Mary

## [1] 0.3122677
```

(c) Did Leo or Mary rank better in their respective groups? Explain your reasoning.

ANS: Leo is 1.09 standard deviations above the mean and Mary is 0.31 standard deviation above the mean. So Mary did better in her group than Leo.

(d) What percent of the triathletes did Leo finish faster than in his group?

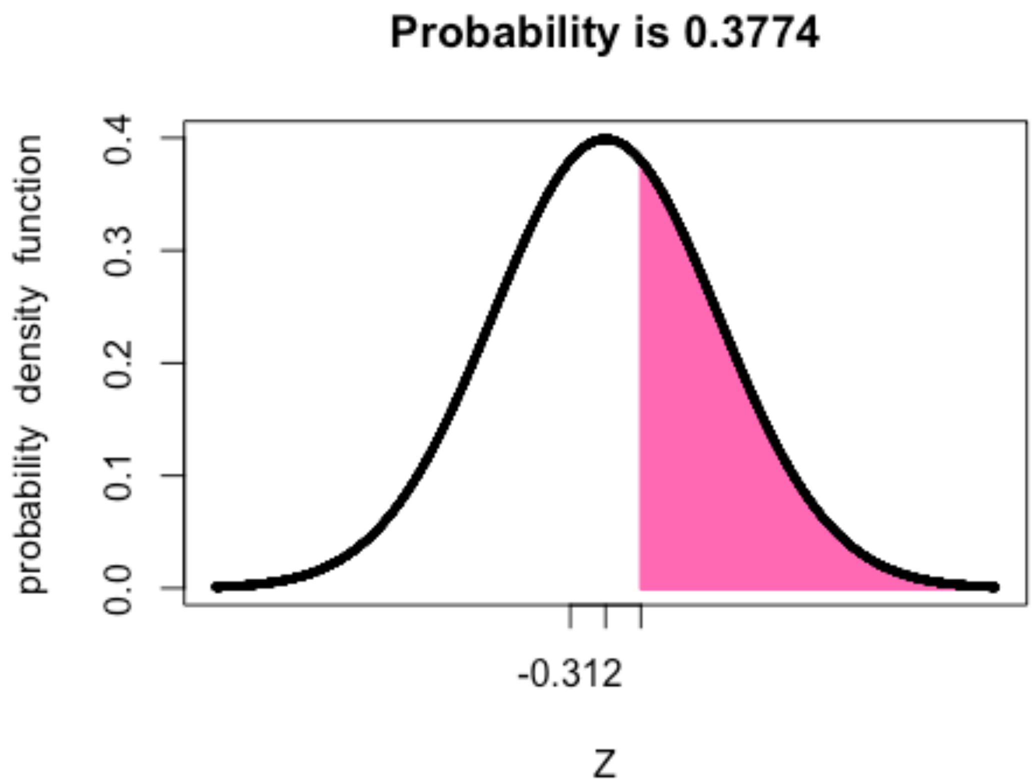
```
pnorm(1.09,lower.tail=FALSE)

## [1] 0.1378566

shadeDist(Z_Leo, lower.tail = FALSE)
```



(e) What percent of the triathletes did Mary finish faster than in her group?

```
pnorm(0.31,lower.tail=FALSE)
```

```
## [1] 0.3782805
shadeDist(Z_Mary, lower.tail = FALSE)
```



(f) If the distributions of finishing times are not nearly normal, would your answers to parts

(e) change? Explain your reasoning.

ANS: If the curve change and skewed the answer would change. The distribution would be shifted from the center and the z-score would be affected. Thus, the z-score evaluation might not be the best for this evaluation since it depends on the mean,

3.18 Heights of female college students.

Below are heights of 25 female college students. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 54, 55, 56, 56, 57, 58, 58, 59, 60, 60, 60, 61, 61, 62, 62, 63, 63, 63, 64, 65, 65, 67, 67, 69, 73 (a) The mean height is 61.52 inches with a standard deviation of 4.58 inches. Use this information to determine if the heights approximately follow the 68-95-99.7% Rule.

```
hgt<-
(c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,65,67,67
,69,73))
sd<-4.58
mean<-61.52
x <- seq(min(hgt), max(hgt), length = length(hgt))
quantile(hgt,0.68)

## 68%
##  63

quantile(x,0.68)

##   68%
## 66.92

quantile(hgt,0.95)

##  95%
## 68.6

quantile(x,0.95)

##   95%
## 72.05

quantile(hgt,0.997)

##  99.7%
## 72.712

quantile(x,0.997)

##  99.7%
## 72.943
```

Quantile of the heights for 68% and 95% standard deviation are different when compared to normal distribution.

(b) Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.

ANS: The data approximately follow the 68-95-99.7% rule. the data is irregular and only approximate a normal distribution.

3.22 Defective rate.

A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others. (a) What is the probability that the 10th transistor produced is the first with a defect?

```r
#Tenth is first defective
tenth <- (0.98^9)*0.02
tenth

## [1] 0.01667496
```

(b) What is the probability that the machine produces no defective transistors in a batch of 100?

```r
1-pgeom(100,.02)

## [1] 0.1299672

#no defective in 100
good <- 0.98^100
good

## [1] 0.1326196
```

(c) On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?

```r
#mean
1/.02

## [1] 50

#standard deviation
sqrt((1-0.02)/0.02^2)

## [1] 49.49747
```

(d) Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation?

```r
#mean
1/.05

## [1] 20

#standard deviation
sqrt((1-0.05)/0.05^2)

## [1] 19.49359
```

(e) Based on your answers to parts (c) and (d), how does increasing the probability of an event a↵ect the mean and standard deviation of the wait time until success?

ANS: Increasing the probability of an event decreases the mean and standard deviation.

3.38 Male children.

While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

(a) Use the binomial model to calculate the probability that two of them will be boys.

```
dbinom(2,3,0.51)

## [1] 0.382347

p <- 0.51
k <- 2
n <- 3

p2of3 <- ( factorial(n) / (factorial(k) * factorial(n-k) )) * p^k * (1-
p)^(n-k)
p2of3

## [1] 0.382347
```

(b) Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers from parts (a) and (b) match.

MMF,MFM, FMM P(A1orA2)=P(A1)+P(A2)

```
(0.51*0.51*0.49)*3

## [1] 0.382347
```

(c) If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a).

ANS: The answers for a and b match but the calculation of b is cumbersome if it has to be done manually for c.

3.42 Serving in volleyball.

A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other. (a) What is the probability that on the 10th try she will make her 3rd successful serve?

```
choose(9,2)*0.15^3*0.85^7
```

```
## [1] 0.03895012
```

(b) Suppose she has made two successful serves in nine attempts. What is the probability that her 10th serve will be successful?

ANS:The probability of success on the 10th serve is the same as the probability of success for the previous 9 servers - 0.15

(c) Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be di↵erent. Can you explain the reason for this discrepancy?

ANS: For b, we are concerned only about the success event whereas for a, it fits the case of a negative binonmial distribution - (probability of kth success on the nth trial).