CS544 Homework 4

Phillip Escandon - [escandon@bu.edu](mailto:escandon@bu.edu)

16 February, 2021

Table of Contents

## 1. Binomial Distribution - Perfect Score

Suppose a student has 40% chance of scoring a perfect score in an exam with randomly selected questions. Each student will be provided 5 attempts.

For the Binomial Distribution - look for an event that only has two outcomes - a success/ failure, 1/0, etc.

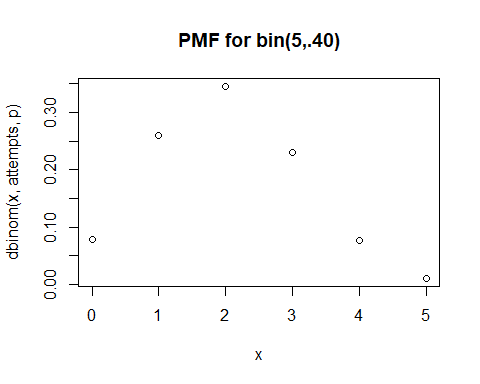
### a. Compute and plot the probability distribution

for the number of perfect scores over the 5 attempts. PMF and CDF

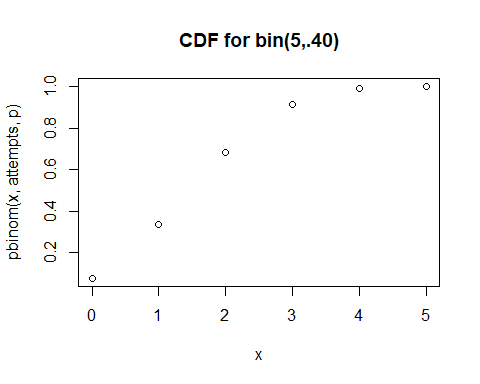
# Probability of success  
p <- .40  
# number of attempts / trials  
attempts <-5  
# number of successes  
successes <- 5  
  
bi<-dbinom(successes,attempts,p)  
  
# Probability Mass Function - shows the distribution of successes  
x <- 0:attempts  
paste('Probability of 5 perfect scores is',bi)

## [1] "Probability of 5 perfect scores is 0.01024"

plot(x,dbinom(x,attempts,p),main = "PMF for bin(5,.40)")



plot(x,pbinom(x,attempts,p),main = "CDF for bin(5,.40)")



### b. Probability of exactly 2 perfect scores

# Probability of success  
p<- .40  
# number of attempts / trials  
attempts<-5  
# number of successes  
successes <- 2  
bi2 <- dbinom(successes,attempts,p)  
  
# Calculation by formula  
# Equation: f(x) = {n x} p^x(1-p)^(n-x)  
paste('Probability of exactly two perfect scores is ', bi2 )

## [1] "Probability of exactly two perfect scores is 0.3456"

The manual calculation will use the above formula to verify the previous calculated probability

# n = attempts   
n <- 5  
# p = Probability  
p<- .4  
# number of attempts / trials  
x<- 2  
  
# Binomial manual Calculation  
fx <- choose(n,x)\* (p\*\*x) \* ((1-p)\*\*(n-x))  
  
paste("f(2) is ",fx)

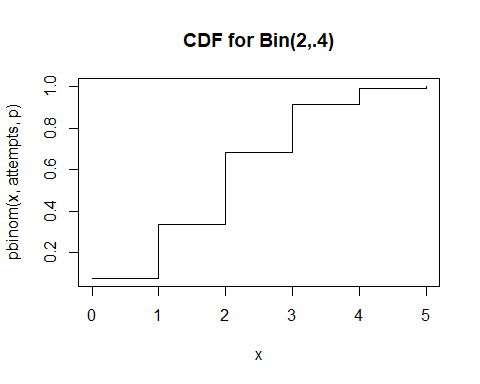
## [1] "f(2) is 0.3456"

### c. Probability of at least 2 perfect scores

# P(X > 1) at least 2 successes  
# Same as 1 - P(X <= 1)  
# 1 - pbinom(1, size = n, prob = p)  
# or  
# sum(dbinom(2:n, size = n, prob = p))  
  
  
# Probability of success  
p<- .40  
# number of attempts / trial  
attempts<- 5  
  
# number of successes  
successes <- 1  
  
  
cdf<-1- pbinom(successes,attempts,p)  
# check  
# sum(dbinom(2:n, size = n, prob = p))  
  
paste("Probability of at least 2 perfect scores is ",cdf)

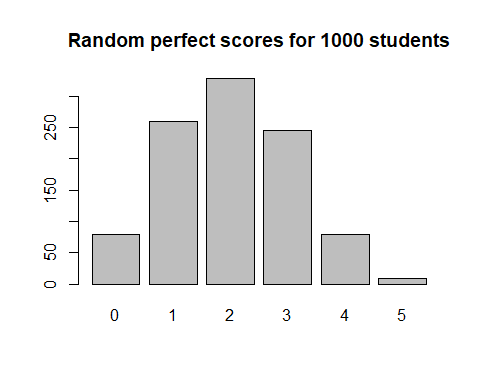
## [1] "Probability of at least 2 perfect scores is 0.66304"

x<- 0:attempts  
# CDF Plot  
plot(x,pbinom(x,attempts,p),type = "s",main = "CDF for Bin(2,.4)")

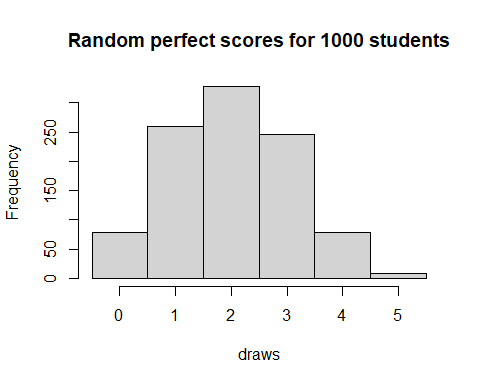


### d. Perfect scores for 1000 sutdents

# Probability of success  
p<- .40  
n<- 5  
successes<- 5  
# draw 1000 students  
draws <- rbinom(1000,size = n, prob = p)  
barplot(table(draws),main = "Random perfect scores for 1000 students")



br<- (0:(n+1)) - .5  
hist(draws,br,main = "Random perfect scores for 1000 students")



Question 2. Negative Binomial Distribution

## 2. Negative Binomial Distibution -

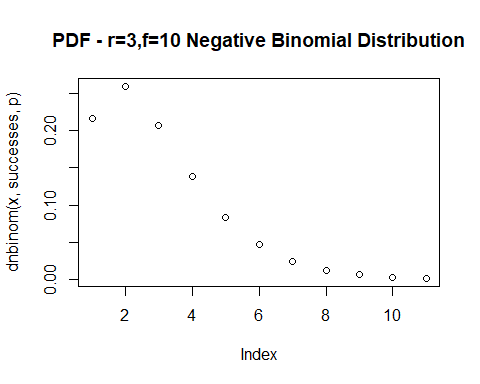
Student has 60 percent chance of a perfect score. Suppose a student has 60% chance of scoring a perfect score in an exam with randomly selected questions. The student has to repeatedly take the exam until they achieve three perfect scores.

### a.

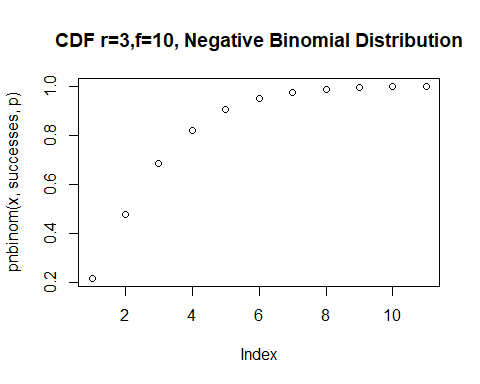
# Need 3 perfect scores - will max 10 fails before quitting.  
# Probability  
p<- .60  
# Number of success as 'r'  
successes<- 3  
# number of failures  
fails <- 10  
# this is the negative binomial distribution. failures before a success.  
  
# args are usually (success, attempts, prob) this time..fails   
(dnbinom(fails,successes,p))

## [1] 0.00149485

x <- 0:fails  
plot(dnbinom(x,successes,p),main = "PDF - r=3,f=10 Negative Binomial Distribution")



plot(pnbinom(x,successes,p),main = "CDF r=3,f=10, Negative Binomial Distribution")



### b. What is the probability 3 perfect scores with exactly 4 failures

successes <- 3  
fails <- 4  
p <- .6  
  
dnb <- dnbinom(fails,successes,p)  
paste('Probability of three perfect scores with 4 failures is',dnb)

## [1] "Probability of three perfect scores with 4 failures is 0.082944"

Perform the same calculation without the R function

# cannot get the example neg binomial by hand using matrix mult   
# Left the args as r,x,p for a simplified choose equation  
  
r <- 3 # successes  
x <- 4 # fails  
p <- .6  
# f(4)  
choose(6,2) \* (.6\*\*3) \* ((1-.6)\*\*4)

## [1] 0.082944

choose((r+x)-1,r-1) \* (p\*\*r)\* ((1-p)\*\*x)

## [1] 0.082944

#Check  
#dnbinom(4,3,.6) == choose(6,2) \* (.6\*\*3) \* ((1-p)\*\*4)

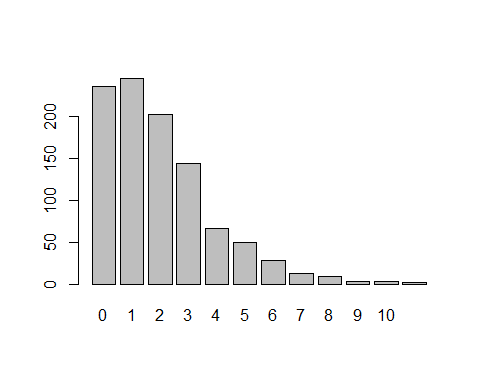
### c. 3 Perfect scores with AT MOST 4 failures

# Apply the CDF here  
# P(X <= 4) at most 4 failures  
r <- 3  
fails <- 4  
p <- .6  
cd <- pnbinom(fails,r,p)  
paste('Probability of 3 Perfect scores with at most 4 failures:',cd)

## [1] "Probability of 3 Perfect scores with at most 4 failures: 0.903744"

### d. Simulate 3 perfect scores for 1000 students. Show a barplot

r <- 3  
fails <- 4  
p <- .6  
x <- rnbinom(1000, size = r, prob = p)  
barplot(table(x))



Question 3.

## 3. Hypergeometric Distribution

M <- 60 # Of interest  
N <- 40 # Not of interest  
K <- 20 # Sample Size

### a.

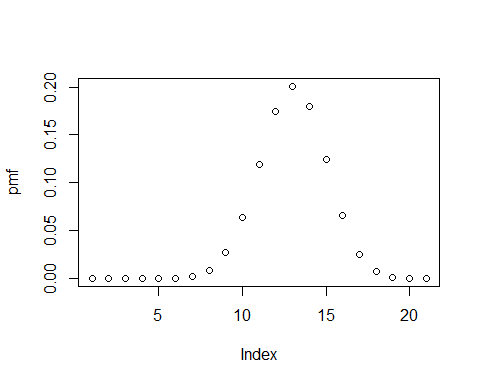
M <- 60 # Of interest  
N <- 40 # Not of interest  
K <- 20 # Sample Size  
dhyper(20, m = M, n = N, k = K)

## [1] 7.820848e-06

pmf <- dhyper(0:K,m=M,n=N,k=K)  
pmf

## [1] 2.571843e-10 1.469625e-08 3.744203e-07 5.665142e-06 5.718253e-05  
## [6] 4.098844e-04 2.167658e-03 8.670631e-03 2.666993e-02 6.376259e-02  
## [11] 1.192361e-01 1.748329e-01 2.007847e-01 1.797234e-01 1.242206e-01  
## [16] 6.530452e-02 2.550958e-02 7.137815e-03 1.346167e-03 1.526019e-04  
## [21] 7.820848e-06

plot(pmf)



### b. Exactly 10 multiple choice out of the 20

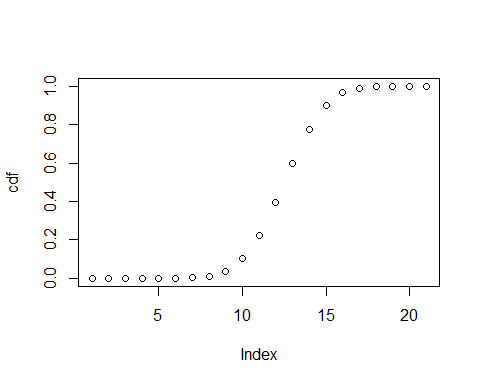
# Review this portion - I don't know  
M <- 60 # Of interest  
N <- 40 # Not of interest  
K <- 20 # Sample Size  
phyper(10,m=M,n=N,k=K)

## [1] 0.22098

cdf <- phyper(0:K,m=M,n=N,k=K)  
cdf

## [1] 2.571843e-10 1.495343e-08 3.893737e-07 6.054516e-06 6.323704e-05  
## [6] 4.731214e-04 2.640779e-03 1.131141e-02 3.798134e-02 1.017439e-01  
## [11] 2.209800e-01 3.958129e-01 5.965976e-01 7.763209e-01 9.005415e-01  
## [16] 9.658460e-01 9.913556e-01 9.984934e-01 9.998396e-01 9.999922e-01  
## [21] 1.000000e+00

plot(cdf)



# P(X = 10) 10 questions of interest out of 20  
  
a <- choose(M,10) \* choose(N,10) / choose(M+N, 20)  
paste("The probability that a student will have exactly 10 multiple choice quesions out of 20 is ",a)

## [1] "The probability that a student will have exactly 10 multiple choice quesions out of 20 is 0.11923605235456"

### c. At least 10 multiple choice out of 20 questions

# P(X > 2) at least 3 faulty chips  
  
1 - phyper(2, m = M, n = N, k = K)

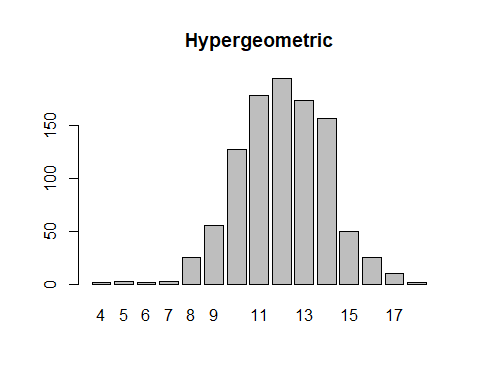
## [1] 0.9999996

# or  
phyper(2, m = M, n = N, k = K, lower.tail = FALSE)

## [1] 0.9999996

### d.

x <- rhyper(1000, m = M, n = N, k = K )  
barplot(table(x),main = 'Hypergeometric')



Question 4.

## 4. Poisson Distribution

# On average 10 questions per day for prof  
lmda <- 10

### a. Exactly 8 questions per day. Prob?

dp <- dpois(8,lambda = lmda)  
paste("Probability of exactly 8 questions per day : ",dp)

## [1] "Probability of exactly 8 questions per day : 0.11259903214902"

### b. At most 8 questions per day. Prob?

pp <- ppois(8,lambda = lmda)  
paste("Probability of at most 8 questions per day : ",pp)

## [1] "Probability of at most 8 questions per day : 0.332819678750719"

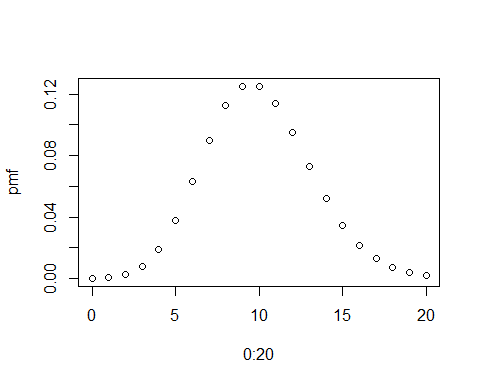
### c. 6 to 12 questions inclusive

pinc <- ppois(12,lambda = lmda) - ppois(5,lambda = lmda)  
paste("Probability of 6-12 questions inclusive is : ",pinc)

## [1] "Probability of 6-12 questions inclusive is : 0.724470513515842"

### d. PMF for the first 20 questions

pmf <- dpois(0:20,lambda = lmda)  
plot(0:20,pmf)



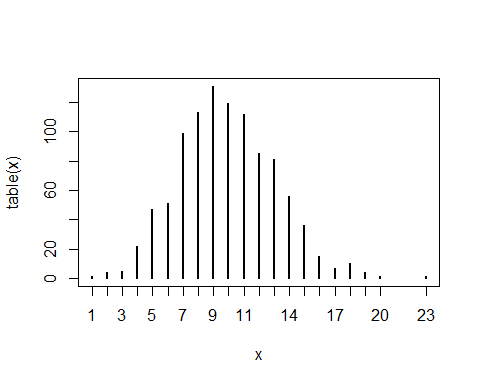
### e. Course runs for 50 days. How many ?’s per day?

Show barplot. Show boxplot. What can you infer?

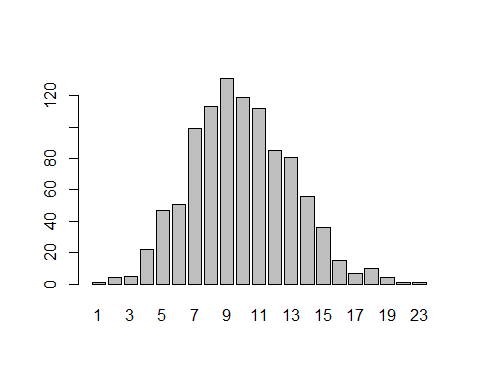
rpois(50, lambda = lmda)

## [1] 7 12 7 5 12 13 12 14 11 8 9 12 8 9 7 9 9 6 9 12 16 7 10 11 11  
## [26] 9 5 9 13 8 8 6 12 8 14 12 9 11 11 6 8 5 10 13 11 8 10 11 13 11

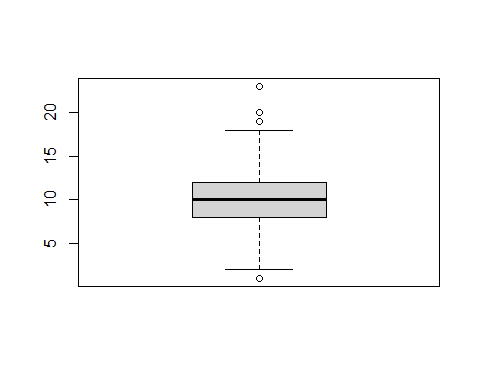
x <- rpois(1000, lambda = lmda)   
plot(table(x))



barplot(table(x))



boxplot(x)



paste("From looking at the box and bar plots you can infer that there are three high outliers, the average number of questions would be about 11 questions per day")

## [1] "From looking at the box and bar plots you can infer that there are three high outliers, the average number of questions would be about 11 questions per day"

Question 5

Suppose that visitors at a theme park spend an average of $100 on souvenirs. Assume that the money spent is normally distributed with a standard deviation of 10 dollars.

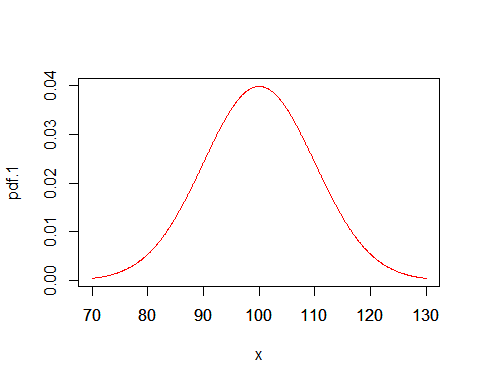
## 5. Normal Distribution

# Ave day $100, standard dev. = $10  
# Remember 68,95,99  
mu <- 100  
sigma <- 10

### a. Plot the PDF of a distribution

covering 3 standard deviations of either side of the mean.

# Ave day $100, standard dev. = $10  
# Remember 68,95,99  
#x <- seq(0,200,1)  
mu <- 100  
sigma <- 10  
x <- seq(mu-3\*sigma,mu+3\*sigma,.1)  
  
  
pdf.1 <- dnorm(x,mu,sigma)  
plot(x,pdf.1,type = "l",col="red")  
 axis(side = 1, at = c(mu-3\*sigma,mu-2\*sigma,mu-sigma,mu,mu+sigma,mu+2\*sigma,mu+3\*sigma),   
 labels = TRUE)



### b. Chance of someone spending more than $120

mu <- 100  
sigma <- 10  
  
# Chance of spending more than $120 so this should be 100 + 2 sigma  
# upper bound  
ub <- 1- pnorm(mu + 2\*sigma, mu ,sigma)  
  
paste("Chance of someone spending more than $120 is ", ub)

## [1] "Chance of someone spending more than $120 is 0.0227501319481792"

### c. Visitor spends between $80 and $90

mu <- 100  
sigma <- 10  
# Spending between 80 and 90 dollars  
b <- pnorm(90,mu,sigma) - pnorm(80,mu,sigma)  
paste("Probability of someone spending between $80 and 90$ is ",b)

## [1] "Probability of someone spending between $80 and 90$ is 0.135905121983278"

### d. Chance of spending within 1 STD,2 STD or 3 STD

mu <- 100  
sigma <- 10  
  
# Spending between 1 STD, 2 STD and 3 STD  
# should be roughly 68, 95 and 99  
std3 <- pnorm(mu + 3\*sigma, mean = mu, sd = sigma) -  
 pnorm(mu - 3\*sigma, mean = mu, sd = sigma)  
paste("Chance of spending with 3 STD: ",std3)

## [1] "Chance of spending with 3 STD: 0.99730020393674"

std2 <- pnorm(mu + 2\*sigma, mean = mu, sd = sigma) -  
 pnorm(mu - 2\*sigma, mean = mu, sd = sigma)  
paste("Chance of spending with 2 STD: ",std2)

## [1] "Chance of spending with 2 STD: 0.954499736103642"

std1 <- pnorm(mu + sigma, mean = mu, sd = sigma) -  
 pnorm(mu - sigma, mean = mu, sd = sigma)  
paste("Chance of spending with 1 STD: ",std1)

## [1] "Chance of spending with 1 STD: 0.682689492137086"

#100\* (pnorm(c(1,2,3)) - pnorm(c(-1,-2,-3)))

### e. Between what two values will the middle 90% of the money spent will fall?

# ? took a while to figure this .  
#todo format for $  
upperLimit <- qnorm(.90,100,10)  
lowerLimit <- qnorm(.10,100,10)  
paste("The middle 90% of the money will be spent between $",lowerLimit," and $",upperLimit)

## [1] "The middle 90% of the money will be spent between $ 87.184484344554 and $ 112.815515655446"

### f. Top 5% of spenders get a T shirt

What’s the min you must spend to get a shirt?

sp <- qnorm(.95,mean=100,sd=10)  
# todo: format for $  
paste("The minimum you must spend to get a shirt is $",sp)

## [1] "The minimum you must spend to get a shirt is $ 116.448536269515"

### g. Show plot for 10,000 visitors with the distribution

y <- rnorm(10000, mean = 100, sd = 10)  
y <- round(y)  
#table(y)  
plot(table(y), type="h")

