ProbStat HW4

Maxwell Greene

```r
library(tidyverse)
```

Question #1

(a) Write down a precise sum expression for the probability that the first five rolls give a three at most two times. Use R to compute a numerical answer.

$$
\begin{aligned}
P(A) &= \sum_{s \in A} P(s) \\
&= \sum_{k=0}^{2} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{2} \binom{5}{k} (5/6)^5
\end{aligned}
$$

```r
sum(dbinom(0:2,size=5,prob=1/6))
```

```
## [1] 0.9645062
```

(b) Calculate the probability that the first three does not appear before the fifth roll.

$$
\begin{aligned}
P(A) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \binom{5}{0} (5/6)^5 \approx 0.4019
\end{aligned}
$$

```r
dbinom(0,size=5,prob=1/6)
```

```
## [1] 0.4018776
```

(c) Calculate the probability that the first three appears before the twentieth roll but not before the fifth roll.

This is equivalent to rolling no threes in five rolls followed by rolling at least one three in fifteen rolls.

$$
\begin{aligned}
P(A) &= P(A_1) + P(A_2) \\
&= \binom{5}{0} (1/6)^0 (5/6)^5 * \sum_{k=1}^{15} \binom{15}{k} (1/6)^k (5/6)^{15-k}
\end{aligned}
$$

```r
sum(dbinom(0,size=5,prob=1/6))*sum(dbinom(1:15,size=15,prob=1/6))
```

```
## [1] 0.3757935
```

Question #2

A team of three is chosen randomly from an office with 2 men and 4 women. Let X be the number of women on the team.

(a) Identify the probability distribution X by name with the relevant parameter values stated.

This is a *hypergeometric* distribution.

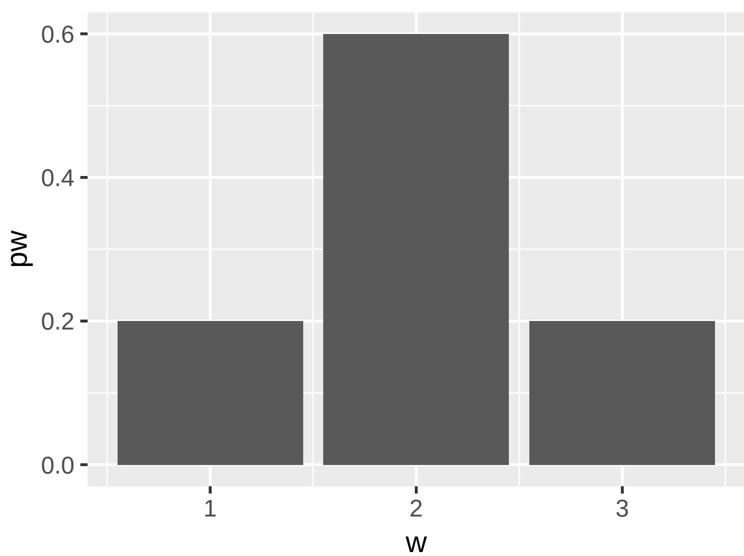
Parameters	
M	number of men total
W	number of women total
N	number of people total
n	number of people chosen for team
w	number of women chosen for team

(b) Give the pmf.

$$P(w) = \frac{\binom{W}{w}\binom{M}{n-w}}{\binom{N}{n}}$$

(c) Plot the pmf (you can use R).

```
M <- 2
W <- 4
n <- 3
w <- 1:3
N <- M+W
hypergeom <- function(M,W,n,w)
{
  return(choose(W,w)*choose(M,n-w)/choose(M+W,n))
}
data <- data.frame(w=w,pw=hypergeom(M,W,n,w)) #dhyper(w,W,M,n)
ggplot(data)+ geom_bar(aes(x=w,y=pw),stat="identity")
```



Question #3

I have an unfair coin for which P(H) = p, where 0 < p < 1. I toss the coin repeatedly until I observe heads for the first time. Let Y be the total number of coin tosses. Find the distribution of Y.

Y is modelled by a *geometric* distribution. That is, the probability that the k^{th} trial is the first success where $P(H) = P(s) = p$, is

$$P(Y = k) = (1 - p)^{k-1}p$$

This makes sense because it requires $(k - 1)$ consecutive failures followed by a single success.

Question #4

Let X be a continuous random variable with the pdf

$$f(t) = \begin{cases} ce^{-t} & \text{if t >= 0} \\ 0 & \text{otherwise} \end{cases}$$

Here c is a positive constant. (a) Find a value for c that guarantees that $\int_{-\infty}^{\infty} f(t)dt = 1$.

$$\begin{aligned} \int_{-\infty}^{\infty} f(t)dt &= c\int_{0}^{\infty} e^{-t}dt \\ &= c\big[-e^{-t}\big]_0^{\infty} \\ &= c\big[0-(-1)\big] = 1 \\ \therefore c = 1 \end{aligned}$$

(b) Find the cdf of X.

(c) Find P(1 < X < 3).

$$\begin{aligned} P(1 < X < 3) &= \int_{1}^{3} e^{-t}dt \\ &= \big[-e^{-t}\big]_1^{3} \\ &= \big[-e^{-3}+e^{-1}\big] \approx 0.318 \end{aligned}$$

```
exp(-1)-exp(-3)
```

```
## [1] 0.3180924
```

Question #5

(a)

```
factorial(12)
```

```
## [1] 479001600
```

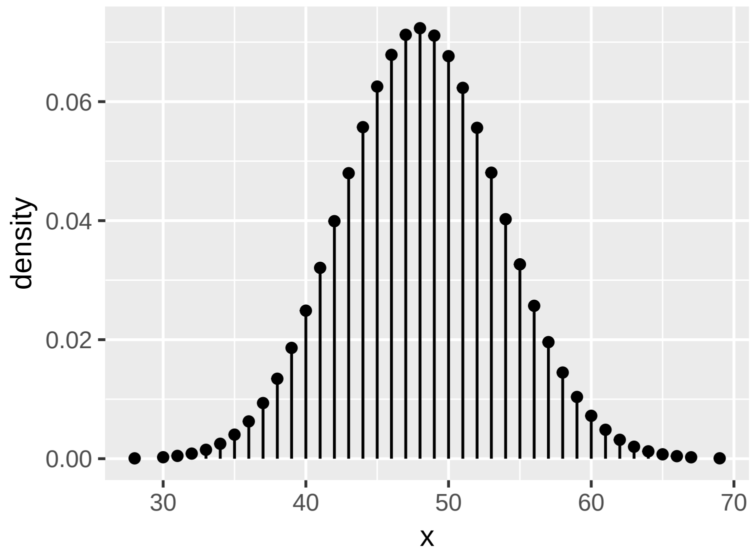
(b)

```
choose(12,5)
```

```
## [1] 792
```

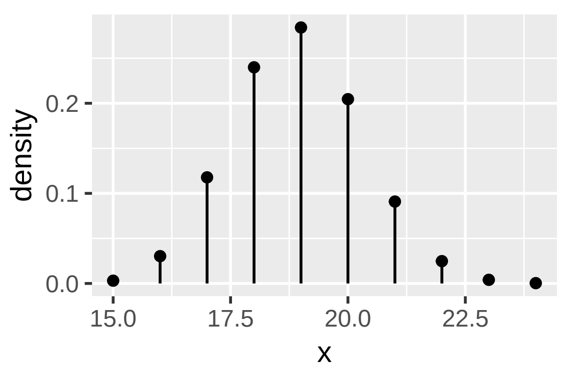
Question #6

```
library(fastR2)
gf_dist(dist="binom", kind = "density",size=130,prob=0.37)
```



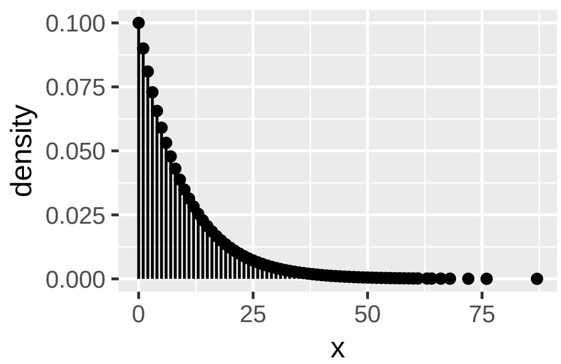
Question #7

```
gf_dist(dist="hyper", kind = "density",m=27,n=13,k=28)
```

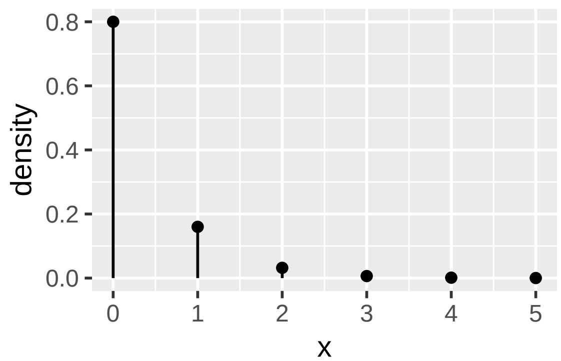


Question #8

```
gf_dist(dist="geom", kind = "density",prob=.1)
```



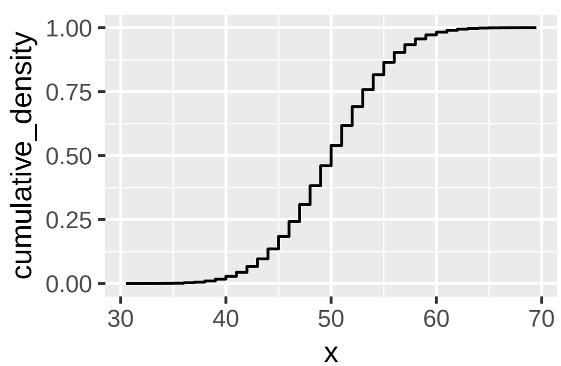

```
gf_dist(dist="geom", kind = "density",prob=.8)
```



Question #9

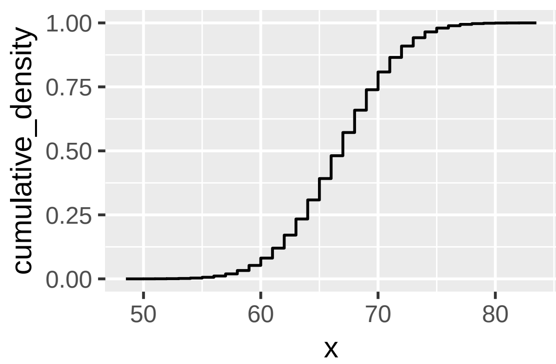
(a)Binomial

```
gf_dist(dist="binom", kind = "cdf",size=100,prob=.5)
```



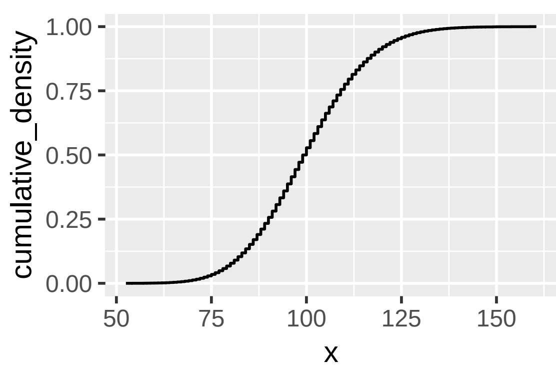
(b)Hypergeometric

```
gf_dist(dist="hyper", kind = "cdf",m=500,n=250,k=100)
```



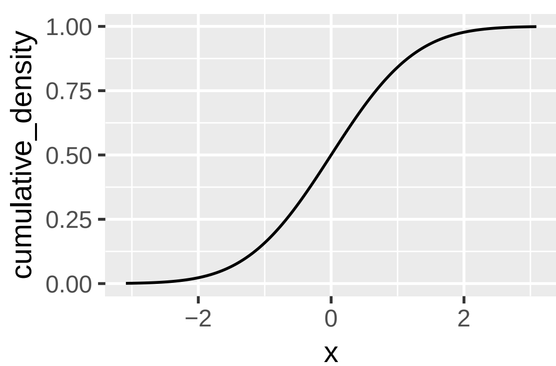
(c)Negative Binomial

```
gf_dist(dist="nbinom", kind = "cdf",size=100,prob=.5)
```



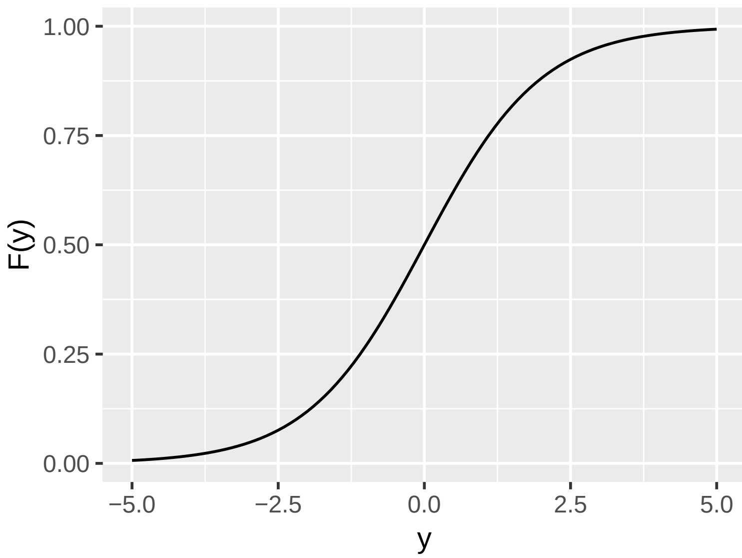
(d)Normal

```
gf_dist(dist="norm", kind = "cdf")
```



Question #10

```
df <- data.frame(y=seq(-5,5,by=0.01))
log_func <- function(y) {1/(1+exp(-y))}
ggplot(data=df,aes(x=y))+stat_function(fun=log_func) + ylab("F(y)")
```



Question (2.7.8)

Five fair dice are rolled. What is the probability that the faces showing constitute a "full house"-that is, three faces show one number and two faces show a second number?

This will be equal to the probability of rolling a number (1) times the rolling the same number two more times (1/6)(1/6) then times the probability of rolling a different number (5/6) times rolling the second number again (1/6).

$$(1)(1/6)(1/6)(5/6)(1/6) = \frac{5}{6^3} \approx .0231$$

```
5/(6^3)
```

```
## [1] 0.02314815
```

Question (3.2.8)

Two lighting systems are being proposed for an employee work area. One requires fifty bulbs, each having a probability of 0.05 of burning out within a month's time. The second has one hundred bulbs, each with a 0.02 burnout probability. Whichever system is installed will be inspected once a month for the purpose of replacing burned-out bulbs. Which system is likely to require less maintenance? Answer the question by comparing the probabilities that each will require at least one bulb to be replaced at the end of thirty days.

System A will be 50 bulbs with $P(A_i) = 0.05$. System B will be 100 bulbs with $P(B_i) = 0.01$.

$$\begin{aligned} P(a > 1 \text{ bulbs blown} \in A) &= \sum_{k=1}^{50} \binom{50}{1} (0.05)^k (0.95)^{50-k} \\ &= 0.923055 \\ P(b > 1 \text{ bulbs blown} \in B) &= \sum_{k=1}^{100} \binom{100}{1} (0.02)^k (0.98)^{100-k} \\ &= 0.8673804 \end{aligned}$$

```r
sum(dbinom(1:50,size=50,prob=0.05))
```

```
## [1] 0.923055
```

```r
sum(dbinom(1:100,size=100,prob=0.02))
```

```
## [1] 0.8673804
```

Question (3.2.28)

Consider an urn with r red balls and w white balls, where $r + w = N$. Draw n balls in order without replacement. Show that the probability of k red balls is hypergeometric

Hypergeometric distribution:

$$P(k \text{ red balls chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

First, let's calculate the number of ways to obtain our selection: we multiply the number of ways to select k red balls from r by the number of ways to select $n-k$ white balls from w, because by selecting k red balls, we are also selecting $n-k$ white balls.This gives $\binom{r}{k}\binom{w}{n-k}$. Since this model considers unordered selections, we should divide this value by the number of ways to choose n balls from N total balls, finally giving the hypergeometric distribution above.

Question (3.3.10)

Urn I and urn II each have two red chips and two white chips. Two chips are drawn simultaneously from each urn. Let X_1 be the number of red chips in the first sample and X_2 the number of red chips in the second sample. Find the pdf of $X_1 + X_2$.

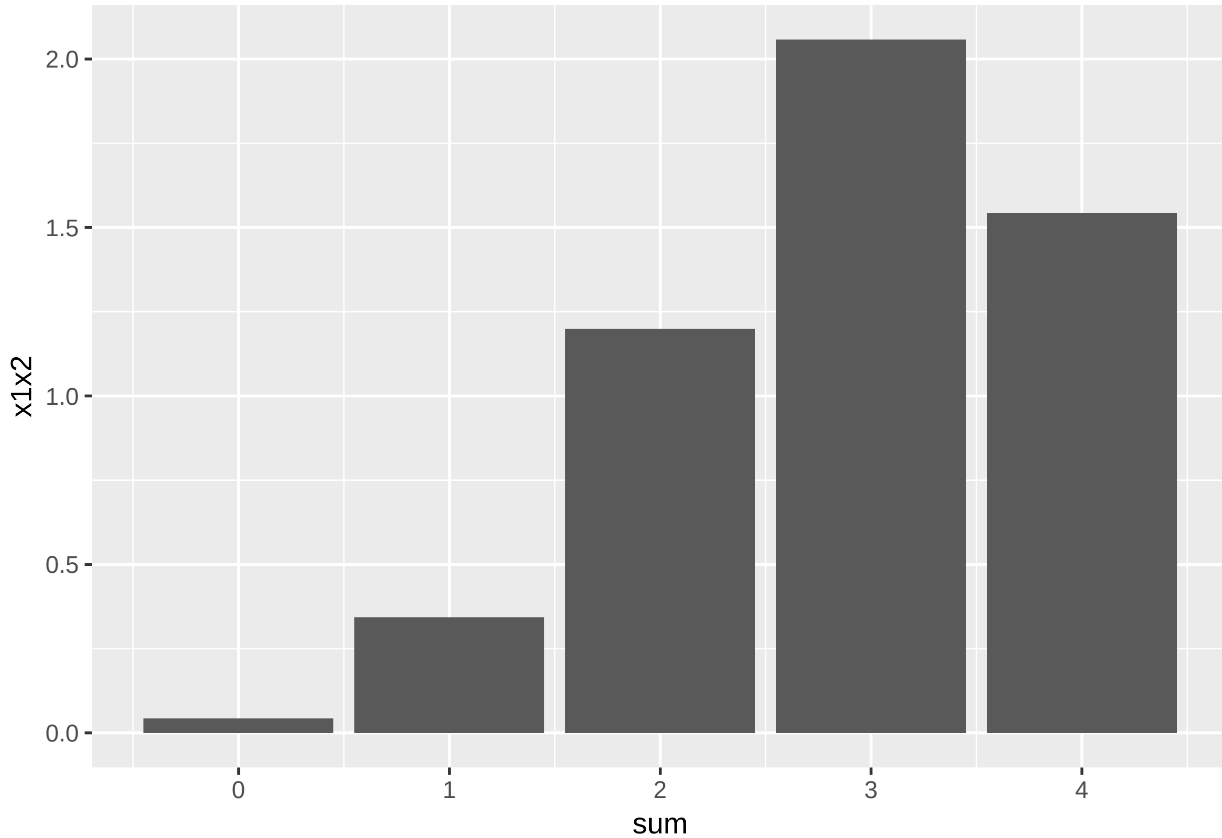
I will utilize the hypergeometric distribution here.

$$\begin{aligned} P(X_1) + P(X_2) = P(X_1, X_2) &= \frac{\binom{r_1}{X_1}\binom{w_1}{n_1-X_1}}{\binom{N_1}{n_1}} + \frac{\binom{r_2}{X_2}\binom{w_2}{n_2-X_2}}{\binom{N_2}{n_2}} \\ &= \frac{\binom{2}{X_1}\binom{2}{2-X_1}}{\binom{4}{2}} + \frac{\binom{2}{X_2}\binom{2}{2-X_2}}{\binom{4}{2}} \\ P(X_1, X_2) &= \frac{\binom{2}{X_1}^2 + \binom{2}{X_2}^2}{6} \end{aligned}$$

```r
px1x2 <- function(x1,x2)
{
  return((choose(4,x1)*choose(4,x2))/choose(8,4))
}
x1 <- 0:2
x1 <- rep(x1,length(x1))
x2 <- 0:2
x2 <- rep(x2,each=length(x1))
data <- data.frame(x1=x1,x2=x2,sum=x1+x2,x1x2=px1x2(x1,x2))
#table(data$sum,data$x1x2)
sum(data$x1x2)
```

```
## [1] 5.185714
```

```
ggplot(data)+ geom_bar(aes(x=sum,y=x1x2),stat="identity")
```



Question (3.4.12)

The cdf for a random variable Y is defined by $F_Y(y) = 0$ for $y < 0$; $F_Y(y) = 4y^3 - 3y^4$ for $0 \leq y \leq 1$; and $FY(y) = 1$ for $y > 1$. Find $P(1/4 \leq Y \leq 3/4)$ by integrating $f_Y(y)$.

Definition 3.4.3:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = P(\{s \in S | Y(s) \leq y\}) = P(Y \leq y)$$

So we want to find $P(1/4 \leq Y \leq 3/4) = F_Y(3/4) - F_Y(1/4)$.

$$\begin{aligned} P(\frac{1}{4} \leq y \leq \frac{3}{4}) &= \int_{-\infty}^{\frac{3}{4}} f_Y(t)dt - \int_{-\infty}^{\frac{1}{4}} f_Y(t)dt \\ &= F_Y(\frac{3}{4}) - F_Y(\frac{1}{4}) \\ &= \left[4(3/4)^3 - 3(3/4)^4\right] - \left[4(1/4)^3 - 3(1/4)^4\right] \\ &= \left[(27/16) - (3^5/4^4)\right] - \left[(1/16) - (3/4^4)\right] \\ &= \frac{27-1}{16} - \frac{3^5+3}{4^4} \approx .664 \end{aligned}$$

```r
(27-1)/(16)-(3^5+3)/(4^4)
```

```
## [1] 0.6640625
```