# Lab 5: The Binomial and Poisson Distributions

**For this lab, it will be helpful to have a copy of the knitted version of this document to answer the questions as much of it is written using mathematical notation that may be difficult to read when the document is not knitted.**

## Lab Goals

1. The purpose of this lab is to explore the following distributions:

- Binomial distribution
- Poisson distribution

2. You will be asked to compute probabilities of events using the pmf of these distributions both

- By hand using the formula
- Using R functions

Emphasis is on identifying random events that can be modeled using these distributions and learning how to calculate probabilities using these distributions by hand and in R.

### Some Notation

Your solutions to the problems below must include the formula used for each calculation. Here is some helpful notation you can copy, edit, and paste as needed. This notation is **not** R code. The notation between dollar signs in the Rmd file is in something called `latex` which lets you write mathematical expressions nicely. Without latex you'd get, for example, (x^2+y)/z, whereas with latex you get $\frac{x^2+y}{z}$. Remember that code chunks are for R code only, not `latex`.

1. For a binomial random variable, $X$, the pmf is given by $P(X=k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$ for $k=0,1,\ldots,n$.
2. For a Poisson random variable $X$, the pmf is given by $P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$ for $k=0,1,2,\ldots$.

## Some R Code

1. `exp()` is the exponential function in R, so for instance, $e^3$ ($e \approx 2.718$ to the 3rd power) is

```
exp(3)
```

```
## [1] 20.08554
```

2. `factorial()` is the R function for computing factorials, so for instance, $3! = 3 \times 2 \times 1$ is

```r
factorial(3)
```

```
## [1] 6
```

3. In R the following operators can be used

```r
2+2  # + for addition
```

```
## [1] 4
```

```r
3-2  # - for subtraction
```

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

```r
3*2  # * for multiplication
```

```
## [1] 6
```

```r
4/4  # / for division
```

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

```r
2^3  # ^ for exponents
```

```
## [1] 8
```

4. Keep in mind that R follows the order of operations. So, when evaluating 3 + 6*8. Multiplication will be performed before addition.

```r
3+6*8
```

```
## [1] 51
```

To have the addition evaluated first, you must use parenthesis.

```r
(3+6)*8
```

```
## [1] 72
```

R will evaluate an expression in the following order:

1) Parenthesis
2) Exponents
3) Multiplication
4) Division

5) Addition
6) Subtraction

5. Here is some code you can copy, paste, and edit in code chunks to calculate probabilities in this document for the binomial and Poisson distributions. DO NOT CHANGE THESE CODE CHUNKS TO `eval=TRUE`. Variables in them have not been defined, and it will not run.

a) To calculate $P(X = k)$ for a binomial distribution use the following code where you have specified `n`,`k`, and `p`.

```
P_k = choose(n,k)*(p^k)*(1-p)^(n-k)
```

b) To calculate P(X = k) for the Poisson distribution use the following code where you have defined `lambda` and `k`.

```
P_k = (exp(-lambda)*lambda^(k))/factorial(k)
```

## Properties of Random Variables

For any random variable X,

1. E(cX)=cE(X) for any constant, c
2. E(X+c) = E(X) + c for any constant, c
3. Var(cX) = $c^2$Var(X). Note, this one makes more sense if you think about the standard deviation, which is the square root of the variance. It says that the standard deviation of cX is |c| times the standard deviation of X.
4. Var(X + c) = Var(X)

These rules actually make sense if you think about an example. Suppose we are playing a board game in which you roll a single die. Let X represent the value rolled. We know that $E[X] = 3.5$.

### Problem 1

a) Suppose in this game, you go forward 2X spaces when you roll X. On average how many spaces forward do you go? (Which of the properties did you use?)

*By the first property, we have that E[2X]=2E[X] and thus on average we will go ahead 7=2(3.5) spaces.*

b) For the same rules as above, what is the standard deviation of the number of spaces you move ahead? Your answer can be expressed in terms of "SD(X)," the standard deviation of X. (Which of the properties did you use?)

*By the third property, SD(number of spaces forward) = SD(2X) = 2SD(X).*

c) Now suppose the rules are different. If we roll X, then we go forward X+4 spaces. What is the expected number of spaces that you go forward?

*By the second property,* $E[X+4] = 3.5 + 4 = 7$

d) Using the rules in (c), what is the standard deviation for the number of spaces you go forward?

*By the fourth property, the standard deviation is unchanged by adding 4. This makes sense since adding 4 doesn't affect how variable the outcome is.*

One more property:

For any random variables X and Y,

E(X+Y) = E(X) + E(Y)

Suppose you roll two dice and X and Y are random variables representing the values. What is the expected sum of the two dice?

*3.5+3.5=7*

## Binomial Distribution

A binomial random variable, X~Binomial(n,p), is characterized by the following:

1. X = Number of "successes" out of n "trials"
2. n is fixed in advance
3. The trials are independent of each other
4. For each trial there is a probability of success, p, that is the same for each trial

Some examples:

1. X = Number of hits in n "at bats"
2. X = Number of cars with defective airbags out of n manufactured
3. X = Number of days you wake up on time in a week

The probability mass function for X~Binomial(n,p) is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k \in \{0, 1, 2, ..., n\}$. Here, $\binom{n}{k}$ is the number of ways to choose k items from n and is defined as $\binom{n}{k} := \frac{n!}{k!(n-k)!}$

The mean and variance of a binomial distribution are

$E(X) = \mu = np$

$Var(X) = \sigma^2 = np(1-p)$

**Problem 2**

The proportion of Norwegians in their early 40s who drink at least 1 cup of coffee per day is about p = 0.894. Suppose that we take a random sample of 50 Norwegians from this age group. Let X = the number of Norwegians in the sample that drink at least 1 cup of coffee a day. In your answers to the questions below, include the correct notation for the probability being asked for in terms of the random variable X.

a) What is the expected value of the number of Norwegians that drink at least 1 cup of coffee a day in the sample?

*$X \sim$ Binomial(50,.894). The mean of this distribution is $np$ =*

```
n=50
p=.894
n*p
```

```
## [1] 44.7
```

b) What is the standard deviation of the number of Norwegians that drink at least 1 cup of coffee a day in the sample?

*$X \sim$ Binomial(50,.894). The standard deviation of this distribution is $(np(1-p))^{1/2}$ =*

```
n=50
p=.894
(n*p*(1-p))^(.5)
```

```
## [1] 2.176741
```

For questions 1.(c)-(e), calculate the required probabilities using R as only a calculator. Include all formulas used.

c) What is the probability that exactly 48 Norwegians in the sample drink at least 1 cup of coffee a day?

$P(X = 48) = 50!/48!2!\ .894^{48}(1 - .894)^2$ =

```
n=50
p=.894
k=48
choose(n,k)*(p^k)*(1-p)^(n-k)
```

```
## [1] 0.06352559
```

d) What is the probability that exactly 2 Norwegians in the sample do not drink at least 1 cup of coffee a day?

*Let $Y$ = Number of Norwegians in the sample that do not drink at least 1 cup of coffee a day. $Y \sim$ Binomial(50, 1-.894). $P(Y = 2) = P(X = 48) = 50!/48!2!\ .894^{48}(1 - .894)^2$ =*

```r
choose(n,k)*(p^k)*(1-p)^(n-k)
```

```
## [1] 0.06352559
```

e) What is the probability that more than 2 Norwegians in the sample do not drink at least 1 cup of coffee a day?

*P(Y > 2) = 1 - P(Y ≤ 2) = 1 - (P(Y=0) + P(Y=1) + P(Y=2)) =*

```r
PX0 = p^50
PX1 = 50*(1-p)*p^49
PX2 = choose(n,2)*((1-p)^2)*p^(n-2)
1-PX0-PX1-PX2
```

```
## [1] 0.9109174
```

f) The following simulates 10,000 observations from a Binomial(50,.894) distribution. `numCoffee` is the simulated observations. The last three lines of code use the simulated values to approximate the answers to (a)-(c).

```r
set.seed(1)
num.simulations = 10000
numCoffee = rbinom(num.simulations, size=50, p=0.894)
# check a
mean(numCoffee)
```

```
## [1] 44.6968
```

```r
# check b
sd(numCoffee)
```

```
## [1] 2.194801
```

```r
# check c
mean(numCoffee == 48)
```

```
## [1] 0.0644
```

### Problem 3

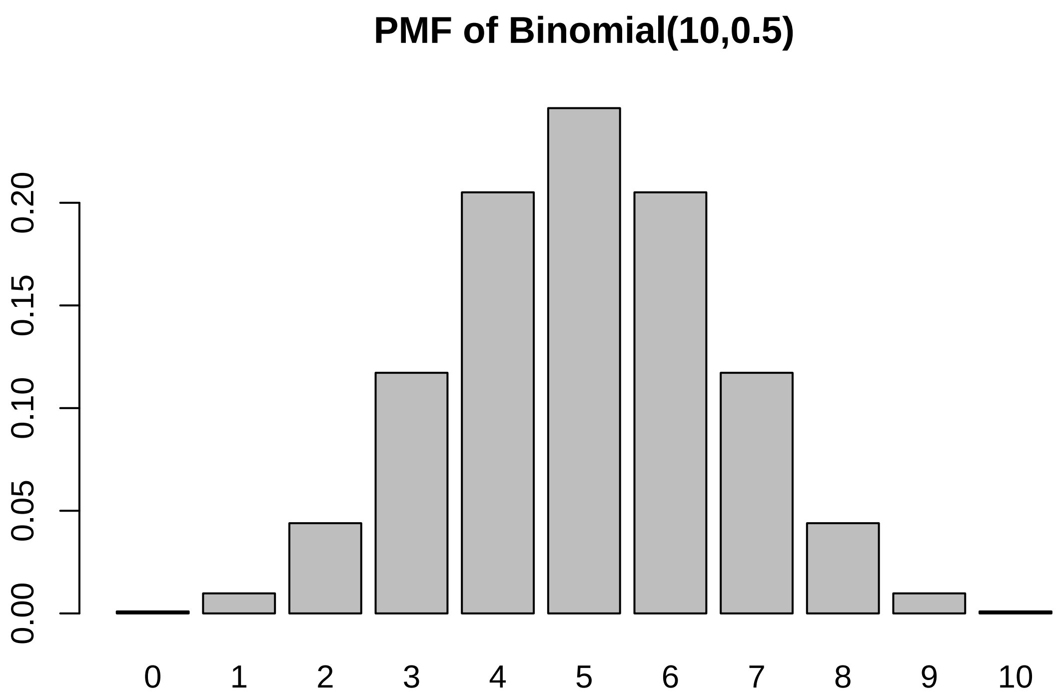
R has built in functions to calculate probabilities and quantiles for the binomial distribution.

The `dbinom(k,n,p)` is the probability mass function for the binomial distribution, i.e. $P(X = k) =$ `dbinom(k,n,p)`. For this function, the following need to be specified:

- `k`, the number of successes associated with the desired probability
- `n`, the number of trials
- `p`, the probability of success for a single trial

The height of each bar in the following illustration is $P(X = k)$ for X~Binomial(10,.5) and k= 0,..,10.

```r
barplot(dbinom(0:10,10,.5),names.arg = 0:10, main="PMF of Binomial(10,0.5)")
```




The `pbinom(k,n,p)` function in R is the probability distribution function for the binomial distribution, i.e. $P(X \leq k)$ =`pbinom(k,n,p)`. Here `k`, `n`, and `p` are as specified above.

The `pbinom(k,n,p)` function can also be used to find the upper tail probabilities, i.e. $P(X > k)$, by using the optional parameter `lower.tail=FALSE`. Thus, $P(X > k) =$ `pbinom(k,n,p,lower.tail=FALSE)`

For 3. (a)-(c), repeat the probability calculations for 2. (c)-(e) using these R functions.

a)

```r
k=48
dbinom(48,n,p)
```

```
## [1] 0.06352559
```

b)

```r
k=2
dbinom(2,n,(1-p))
```

```
## [1] 0.06352559
```

c)

```r
1 - pbinom(2,n,(1-p))
```

```
## [1] 0.9109174
```

The `qbinom(p_quant,n,p)` function in R is the quantile function for the binomial distribution. It finds the smallest value of k such that $P(X \leq k) \geq$ `p_quant`. Here `k`, `n`, and `p` are as specified above. `p_quant` is such that the $100 \times$ `pquant` percentile of the binomial distribution is given by `qbinom(p_quant,n,p)`.For instance, specifying `p_quant = .5` will return the median of the binomial distribution.

d) Find the 25th and 75th percentiles of the binomial distribution specified in problem 2.

```
p_quant=.25
qbinom(p_quant,n,p)
```

```
## [1] 43
```

```
p_quant=.75
qbinom(p_quant,n,p)
```

```
## [1] 46
```

e) Find the smallest number such that the probability of the number of Norwegians that drink at least 1 cup of coffee a day in the sample is greater than this number is less than or equal to 10%.

*This is the 90th percentile of Binomial(50,.894) =*

```
p_quant=.90
qbinom(p_quant,n,p)
```

```
## [1] 47
```

## Poisson Distribution

A poisson random variable, X~Poisson($\lambda$), is characterized by the following:

1. X = Number of rare events that occur over a fixed amount of time or space
2. Events are independent
3. The maximum number of events that occur is not fixed

Some examples:

1. X = Number of tornadoes in a particular area over a year
2. X = Number of raindrops that fall on a particular square inch of roof during a one-second interval of time
3. X = Number of people that arrive at a train station during a 5 minute interval of time

For X~Poisson(λ) and any $k \in \{0, 1, 2, ...\}$, the following probability mass function defines P(X=k).

$P(X = k) = e^{-\lambda}\lambda^k/k!$

The mean and variance of a Poisson distribution are

$E(X) = \mu = \lambda$

$Var(X) = \sigma^2 = \lambda$

**Problem 4**

In the summer months, northern Minnesota has a thriving mosquito population. When the average person steps outside in the summer, the number of mosquito bites he receives every hour follows a Poisson(10) distribution if he has not applied mosquito repellant. If he has applied mosquito repellant, the number of mosquito bites he receives every hour has a Poisson(3) distribution. Assume a randomly chosen Minnesotan has just stepped outside without applying mosquito repellant. Let

X = number of mosquito bites this person receives in the first hour.

Y = number of mosquito bites the person receives in the second hour.

a) What is the probability in the first hour outside this individual does not get bit by a mosquito?

$X \sim$ *Poisson(10)*, $P(X = 0) = e^{-10}10^0/0! =$

```
(exp(-10)*10^(0))/factorial(0)
```

```
## [1] 4.539993e-05
```

b) What is $P(10 < X \leq 12)$?

$P(10 < X \leq 12) = P(X = 11) + P(X = 12) = e^{-10}10^{11}/11! + e^{-10}10^{12}/12!$

```
PK11 = (exp(-10)*10^(11))/factorial(11)

PK12 = (exp(-10)*10^(12))/factorial(12)

PK11 + PK12
```

```
## [1] 0.2085167
```

c) If this Minnesotan plans to stay outside for 2 hours in total and he applies mosquito repellant after the first hour, how many mosquito bites can he expect to receive?

*Let X = # of mosquito bites in the first hour and Y = # mosquito bites in the second hour E(X+Y) = E(X) + E(Y) =10 + 3*

```
10+3
```

```
## [1] 13
```

d) After spending two hours outdoors, this Minnesotan will treat his mosquito bites with a topical ointment. For complete relief from the allergic reaction caused by the mosquito bites, the ointment must be applied to every mosquito bite 3 times. How many times in total will he expect to apply the ointment after the 2 hours to alleviate the allergic reaction of every bite?

$E(3X + 3Y) = 3E(X) + 3E(Y) = 3 \times 10 + 3 \times 3 =$

```
3*10 + 3*3
```

```
## [1] 39
```

**Problem 5**

R also has built in functions to calculate probabilities and quantiles for the Poisson distribution.

The `dpois`$(k, \lambda)$ is the probability mass function for the Poisson distribution, i.e. $P(X = k) =$ `dpois`$(k, \lambda)$. For this function, the following need to be specified:

- `k`, the number of successes associated with the desired probability
- $\lambda$, the mean and variance of the Poisson distribution

The `ppois`$(k, \lambda)$ function in R is the cumulative probability distribution function for the Poisson distribution, i.e. $P(X \leq k) =$ `ppois`$(k, \lambda)$. Here `k` and $\lambda$ are as specified above.

Similar to the `pbinom()` function, the `ppois`$(k, \lambda)$ function can also be used to find the upper tail probabilities, i.e. $P(X > k)$, by using the optional parameter `lower.tail=FALSE`. Thus, $P(X > k) =$ `ppois(`k,$\lambda$,`lower.tail=FALSE)`

For 5. (a)-(b), repeat the probability calculations for 4. (a)-(b) using these R functions.

a)

```
dpois(0,10)
```

```
## [1] 4.539993e-05
```

b)

```
ppois(12, 10) - ppois(10, 10)
```

```
## [1] 0.2085167
```

```
## equivalently
dpois(11, 10) + dpois(12, 10)
```

```
## [1] 0.2085167
```

The `qpois`(p_quant,$\lambda$) function in R is the quantile function for the Poisson distribution. It finds the smallest value of k such that $P(X \leq k) \geq$ `p_quant`. Here `k` and $\lambda$ are as specified above. `p_quant` is such that the $100 \times$ `pquant` percentile of the Poisson distribution is given by `qpois`(p_quant,$\lambda$).

c) Find the median of the first Poisson distribution specified in problem 4.

```
qpois(.5,10)
```

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