Lessons this week will cover Chapter 5 in the textbook.

Lesson 1: Discrete random variables, probability functions and expectation

Key Takeaways

By the end of this lesson you should be able to:

- Differentiate between a discrete and continuous random variable.
- Define and draw a probability function
- Define and draw a cumulative distribution function
- Solve for and interpret the expected value of a discrete random variable
- Solve for and interpret the variance of a discrete random variable
- Solve for the expected value and variance of functions of random variables.

Section 5.1: What is a random variable?

Random Variable: This is a variable that takes on numerical values according to a chance process.

For example: If we flip a coin 25 times and count the number of heads observed then we can define the random variable: X= the number of heads observed in 25 flips.

Notice that we will never know the actual value that X takes on until the experiment is terminated. But we do know what **kind of values** it can take on **(outcomes)** and we know the **distribution** of the random variable i.e the likelihood of observing any one of these values. This causes us to make a distinction between the random variable and the observed value.

<u>Notation</u>:

- Capitalized letters are used to represent **random variables** (r.v), **X**
- The actual value of the random variable **observed** is denoted by a lower-case letter, $\boldsymbol{x}$.

Section 5.2: Discrete vs. Continuous random variables

The variable X= # of heads observed in 25 flips is a numeric variable i.e quantitative in nature but in addition is said to be a **discrete** random variable.

- A **discrete** random variable is one that can take on a ***countable number of possible values***. This can be a finite set or a countably infinite set.
  - In our example of X = # of heads observed in 25 flips we have that $x = 0, 1, \ldots, 25$
  - We can also have X= # of times it rains in a span of 10 years. This is defined such that $x = 0, 1, 2, \ldots$ (Note that there is no upper bound on the number of times it rains, but this is still considered discrete)
- The second type is called a **Continuous** Random Variable. This is a random variable that can take on an ***infinite number of possible values***. They are usually represented by a real number.
  - X = time taken to complete a particular task. This is defined such that $x > 0$
  - X = length of time you wait for a bus to arrive, where buses arrive every 15 minutes. This is defined such that $x \in [0; 15]$.

- In other cases, we could have a random variable X such that $x \in (-\infty, \infty)$

To summarise:

X	
Discrete Random Variable	**Continuous Random Variable**
Range consists of a finite or countably infinite set of values. The random variable can take integer values or, in general, values in a countable set, e.g. 0,1,2,….	Range consists of an infinite set of values. The random variable takes values on some interval of real numbers. For example, (0,1) or $(0,\infty)$ or $(-\infty,\infty)$.

Example 1

Security analysts are professionals who devote full-time efforts to evaluating the investment worth of a narrow list of stocks. The following variables are of interest to the security analysts. Which are discrete and which are continuous?

1. The closing price of a particular stock on the New York Stock Exchange.
2. The number of shares of a particular stock that are traded each business day.

Soln:

1. X = closing price of a particular stock on the New York Stock Exchange. This can be positive or negative taking on any value and so is continuous.
2. X = number of shares of a particular stock that are traded each business day. This is counting the number of whole shares traded and so is discrete.

You Try 1

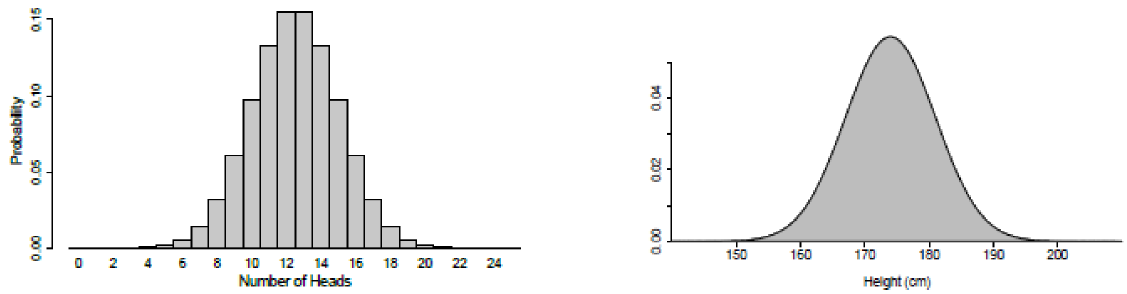
Security analysts are professionals who devote full-time efforts to evaluating the investment worth of a narrow list of stocks. The following variables are of interest to the security analysts. Which are discrete and which are continuous?

1. The percentage change in yearly earnings between 2011 and 2012 for a particular firm.
2. The number of new products introduced per year by a firm.

Before working with a random variable, it is important to determine whether it is discrete or continuous as they are handled differently.

In the Figure below we note that discrete random variables (Figure 5.1 a)) will be modeled using a histogram where the height of the bar represents the probability (y-axis) and the x-axis represents the possible values of X.

In Figure 5.1b) notice that the x-axis are the possible values of X but the y-axis is not labeled. This is because the graph drawn represents what is called the density curve. In Chapter 6 we will discuss this curve in more detail but to summarise for continuous random variables the probability is defined by the area under the curve as opposed to the height of the curve.



(a) The distribution of the number of heads in 25 tosses

(b) The distribution of the height of adult Canadian males

Figure 5.1: A discrete probability distribution and a continuous probability distribution.

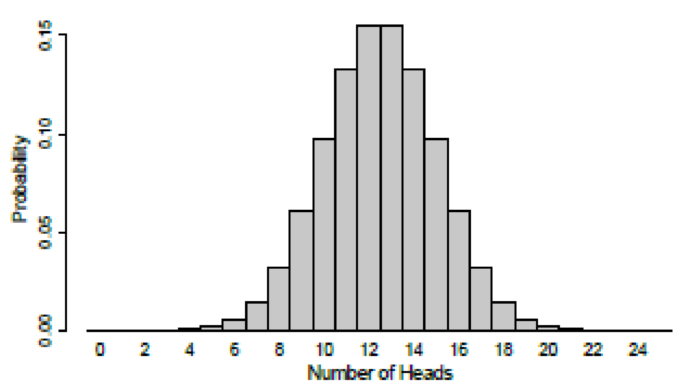
In this Chapter we focus on discrete random variables and in Chapter 6 we will look at continuous random variables.

Section 5.3: Discrete Probability Distributions

To properly define a discrete random variable and to solve for probabilities we use a **Probability mass function (pmf)/ Probability function (pf.), $p(x)$**

For **discrete** random variables: $\boldsymbol{p(x) = \mathrm{P}\ (X = x)}$ represents the probability that X (the r.v.) takes on the value x. This is called a probability function, and it allocates a probability for every value of x.

Visually, it is represented by a **histogram**



(a) The distribution of the number of heads in 25 tosses

Properties of the probability function

$p(x)$ has ALL the same properties as a probability and so we have:

1. $0 \leq p(x) \leq 1$ for ALL values of x
2. $\sum_{all\ x} p(x) = 1$. This

Example 2

A discrete random variable X can assume five possible values: 2, 3, 5, 8, and 10. Its probability function is shown in the table:

$\boldsymbol{x}$	**2**	**3**	**5**	**8**	**10**
$\boldsymbol{p(x) = P(X = x)}$	0.15	0.10	?	0.25	0.25

1. Using the properties of $p(x)$ solve for P(X=5).
2. What is the probability X equals 2 or 10?
3. What is $P(X \leq 8)$?
4. What is $P(X < 8)$?

Soln:

1. $P(X = 5) = 1 - (0.15 + 0.10 + 0.25 + 0.25) = 0.25$
2. $P(X = 2 \cup X = 10) = P(X = 2) + P(X = 10) = 0.15 + 0.25 = 0.4$
3. $P(X \leq 8) = P(X = 2) + P(X = 3) + P(X = 5) + P(X = 8) = 0.75$
4. $P(X < 8) = P(X = 2) + P(X = 3) + P(X = 5) = 0.5$

You Try 2

Consider the following probability function of a random variable X.

$\boldsymbol{x}$	**10**	**20**	**30**	**40**
$\boldsymbol{p(x) = P(X = x)}$	0.2	0.2	0.5	0.1

1. Is this a valid discrete probability function? Justify your response.
2. What is the most likely value of X?
3. What is the conditional probability that X is less than 25, given X is less than 35?

Cumulative Distribution Function

A second function of interest is the cumulative distribution function (cdf) which is defined as:

$$\boldsymbol{F(x) = \mathbf{Pr}(X \leq x)}$$

It represents the cumulative probability, less than or equal to x.

Properties of the cumulative distribution function

1. $F(x)$ is a non-decreasing function of x
2. $0 \leq F(x) \leq 1$ for all x
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, i.e. F(x) starts at 0 and ends at 1

Section 5.3.1: Expectation and Variance of a Discrete Random Variable

Populations vs. Samples: Recall that a population represents the entire group of interest, whereas the sample represents the subset drawn from the population on which we collect information.

Typically, we are interested in a characteristic(s) of the population. These characteristics are called **parameters** and include things like the expected value ($\mu = E(X)$) and the variance (σ^2). Since we are unable to collect information on the entire population the parameters are usually unknown.

Samples drawn from populations are useful because they allow us to calculate an estimate of our unknown parameters of interest. These estimates are called **statistics** and include things like the sample mean ($\bar{x}$) and sample variance (s^2).

So far, we have only dealt with sample statistics since we are unaware of the general behaviour/ distribution of the population and so the best we can do is sample a subset and use it to gain estimates. For random variables since we know the **exact behaviour** (i.e have a probability model) we can solve for parameters.

Expected Value (Mean)

For a discrete random variable, X, if the probability function is known, then we can calculate the population mean/expectation.

Formula: $\boldsymbol{\mu_X = E(X) = \sum_{all\, x} x \times \mathbf{Pr}\,(X = x)}$

This is a theoretical weighted average of all possible values of X since each value is weighted by its probability of occurring.

Note: ***The expected value of X represents the long-run average,*** the average of the X-values that would be observed in an infinite number of trials.

Expected Value of a function of X

Often, we are interested in the expected value of a function of a random variable X, e.g. Log(X), or X^2. The expected value of a function $g(x)$, where X is a discrete random variable, is defined as:

$$\boldsymbol{E[g(X)] = \sum_{all\, x} g(x) \times \mathbf{Pr}\,(X = x)}$$

Variance

The expected value is useful in providing an idea of where the centre of the distribution lies. Recall that in addition to centrality we want an understanding of spread. The population variance gives us a good indication of the average square spread around the mean.

Formula:

1. Best for interpretation purposes:

$$\boldsymbol{\sigma_X^2 = Var(X) = E\left[(X - \mu)^2\right] = \sum_{all\, x} (x - \mu_x)^2 \mathbf{Pr}\,(X = x)}$$

2. Best for calculation purposes:

$$\sigma_X^2 = Var(X) = \left[\sum_{all\,x} x^2 \Pr(X = x)\right] - \mu_X^2 = E(X^2) - [E(X)]^2$$

Both formulas are equivalent, the first allows us to better visualize what the variance is describing. The second formula is typically easier and quicker to calculate.

Note: **Standard Deviation =** $\sigma_X = \sqrt{\sigma_X^2}$

Example 3

A discrete random variable X can assume five possible values: 2, 3, 5, 8, and 10. Its probability function is shown in the table:

x	2	3	5	8	10
$p(x) = P(X = x)$	0.15	0.10	0.25	0.25	0.25

Solve for the expected value and variance of X.

Soln:

$E(X) = (2 \times 0.15) + (3 \times 0.10) + (5 \times 0.25) + (8 \times 0.25) + (10 \times 0.25) = 6.35$

$E(X^2) = (2^2 \times 0.15) + (3^2 \times 0.10) + (5^2 \times 0.25) + (8^2 \times 0.25) + (10^2 \times 0.25) = 48.75$

Therefore $Var(X) = E(X^2) - (E(X))^2 = 48.75 - (6.35)^2 = 8.4275$

You Try 3

Consider the following probability function of a random variable X.

x	10	20	30	40
$p(x) = P(X = x)$	0.2	0.2	0.5	0.1

Solve for the expected value and variance of X

Properties of Expectation and Variance

For constants a and b,

- $E(a + bX) = a + bE(X)$
- $Var(a + bX) = b^2 Var(X)$

These properties demonstrate that the expected value is a **location** parameter, shifting the distribution left and right as we add and subtract from the random variable. It also indicates that the variance is a **scale** parameter, noticing that only the multiplication factor of ‘b’ causes the spread to change, i.e. causes the distribution to stretch out or condense.

Example 4

A friend of yours has a business opportunity in which they have an expected profit of $2M, with a standard deviation of $0.8M. To reduce their risk, they will sell you ¼ of the profits (and possible losses) for a fee of $0.4M. Let X represent your friend’s profit on his original deal.

1. If you take the deal, what is your expected profit?

2. If you take this deal you are opening yourself up to considerable risk. We would like to quantify the risk by solving for the standard deviation of your profit.

Soln:

Based on the description provided your profit is defined as $Y = \frac{1}{4}X - 0.4$

1. $E(Y) = E\left(\frac{1}{4}X - 0.4\right) = \frac{1}{4}E(X) - 0.4 = \left(\frac{1}{4} \times 2\right) - 0.4 = 0.1M$
2. $Var(Y) = Var\left(\frac{1}{4}X - 0.4\right) = \left(\frac{1}{4}\right)^2 Var(X) = \left(\frac{1}{4}\right)^2 \times (0.8)^2 = 0.04$
3. Hence $std.dev(X) = \sqrt{0.04} = 0.2M$

Some additional properties:

If X and Y are two random variables and a and b are any real numbers, then

- $E(aX + bY) = aE(X) + bE(Y)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y)$

Recall: In Chapter 15, $Cov(X,\ Y)$ represents the **covariance** between X and Y. The main interpretation is its **sign**. This is a mathematical way of measuring the relationship between X and Y. When we have the probability function, we are again able to solve for the population covariance:

$$\boldsymbol{Cov(X,\ Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_{all\ x,y} (X - \mu_x)(Y - \mu_Y)\Pr(X = x\ and\ Y = y)}$$

Properties:

- $Cov(X,\ Y) = Cov(Y,\ X)$
- $Cov(X,\ X) = Var(X)$
- $Cov(aX,\ bY) = abCov(X,\ Y)$ where a, b are constants

Consider what happens to the covariance if X and Y are **Independent**. Under this case, $\boldsymbol{Cov(X,Y) = 0}$ (If X and Y are independent that means X does not affect Y and vice versa, meaning that there is NO linear relationship between them, and thus their covariance is 0) and hence

- $Var(X + Y) = Var(X) + Var(Y)$
- $Var(X - Y) = Var(X) + Var(Y)$

We can extend these properties to the sums of more than two random variables.

Let $X_1, X_2, \dots, X_n$ represent n random variables, then: $\boldsymbol{E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)}$

If in addition $X_1, X_2, \dots, X_n$ are **independent** random variables, then: $\boldsymbol{Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)}$

Example 5

Assume we have independent random variables X and Y such that

	Mean	**Std.Dev**
X	10	2

Y	20	5

Solve for the expectation and standard deviation of the following:

1. 3X
2. Y+6
3. X-Y

Soln:

1. $E(3X) = 3E(X) = 3 \times 10 = 30$ and $Var(3X) = 9Var(X) = 9 \times 2^2 = 36 \Rightarrow std.dev = \sqrt{36} = 6$
2. $E(Y+6) = E(Y) + 6 = 20 + 6 = 26$ and $Var(Y+6) = Var(Y) = 5^2 = 25 \Rightarrow std.dev = \sqrt{25} = 5$
3. $E(X-Y) = E(X) - E(Y) = 10 - 20 = -10$ and $Var(X-Y) = Var(X) + Var(Y) = 4 + 25 = 29 \Rightarrow std.dev = \sqrt{29} = 5.38516$

You Try 4

A broker has calculated the expected values of two different financial instruments X and Y. Suppose that E(X) = \$100, E(Y) = \$90, $\sigma_X = \$12$, and $\sigma_Y = \$8$. Assuming the instruments are independent, solve for the following the expected value and standard deviation of the functions below:

1. X+10
2. 5Y
3. X+Y

Model Distributions

In the rest of the chapter, our aim is to <u>identify common types</u> of processes or problems and to develop **probability distributions** that represent them. Many processes or problems tend to have the same structure.

Practice

Chapter 5: 1, 3, 4-8, 10, 36-38, 41, 42, 43, 45, 53, 54

Lesson 2: Bernoulli, Binomial and Geometric distributions

Key Takeaways

By the end of this lesson you should be able to:

- Identify and apply a Bernoulli distribution
- Identify and apply a Binomial distribution
- Identify and apply a Geometric distribution

Section 5.4: The Bernoulli Distribution

Physical Set-up

We conduct an experiment a single time, i.e have 1 trial. In this trial we have two possible mutually exclusive outcomes, labelled **success** vs. **failure**, where:

$$P(success) = p \text{ and hence } P(failure) = 1 - p = q$$

Define $\boldsymbol{X} = \begin{cases} \textbf{1 if a success occurs} \\ \textbf{0 if a failure occurs} \end{cases}$

Such a random variable has a **Bernoulli distribution**, with probability function:

$\boldsymbol{P(X = x) = p^x(1-p)^{1-x}}$**,** For $\boldsymbol{x = 0, 1}$

Example: Flip a coin Once, and X=1 if we observe a Head.

Expectation and Variance

If X is Bernoulli(p) then:

- The expected value of X is $\boldsymbol{E(X)} = \sum_{x=0}^{1} x\, p(x) = \big(0 \times (1-p)\big) + (1 \times p) = \boldsymbol{p}$
- The variance of X is $\boldsymbol{Var(X)} = E(X^2) - \big(E(X)\big)^2 = p - p^2 = \boldsymbol{p(1-p)}$
    - Note that $E(X^2) = \big(0^2 \times (1-p)\big) + (1^2 \times p) = p$

The Bernoulli distribution is a special case of the Binomial where n = 1.

Section 5.5: The Binomial Distribution

This distribution arises when we have an experiment made up of **n independent** Bernoulli trials. Some examples include:

- The number heads observed when we flip a fair coin 50 times.
- The number of contracts that a firm gets among 40 that it bid for.
- The number of people that reply “Yes I will buy the product” among 100 who receive a marketing survey.

In reading these statements the experiments may sound quite different to each other, however if we take a closer look we will notice that they share some common characteristics.

Physical Set-up

- Two outcomes: Success vs. Failure

- Independent trials
- Multiple trials
- Same probability of success (p) with each trial

Example 1

The number heads observed when we flip a fair coin 50 times. Justify that this experiment satisfies the physical set-up of a Binomial distribution.

Soln:

- Two outcomes: Head (success) vs. Tail (failure)
- Independent trials: The outcome in each flip does not affect the outcomes observed in any other flip.
- Multiple trials: we are flipping the coin 50 times (i.e. a value larger than 1).
- Same probability of success (p) with each trial: Flipping a "fair" coin is to say that there is a 50-50 chance of getting a head vs. a tail, hence p = 0.5 for each trial.

You Try 1

The number of people that reply "Yes I will buy the product" among 100 who receive a marketing survey. Justify that this experiment satisfies the physical set-up of a Binomial distribution.

Now that we understand the physical set-up of the Binomial distribution and we know the type of outcomes X can take on let's look at the probability function.

Some New Notation

First, we introduce some tools and notation.

Factorials

n! = n factorial = n(n-1)(n-2)…(2)(1)

For example: 5! = $5 \times 4 \times 3 \times 2 \times 1 = 120$

What does n! represent?

This represents the number of ways we can arrange n objects in n spots where order matters.

In more detail, if you have n spots in total to fill and you have n objects to select from then:

You have **n** choices for the first spot, **n-1** choices for the second spot, …, and **1** for the last spot.

Special Cases:

- 0! = 1
- 1! = 1

Example 2

How many ways can we arrange the numbers 1, 2, 3?

Soln:

In this example we have 3 objects and 3 spots to fill where the numbers can appear in any order (i.e order matters).

Hence, we have $3! = 3 \times 2 \times 1 = 6$ ways

The choose function

Suppose I want to select 3 objects from 4. For example, I want to select 3 letters from the word MATH. **Order does not matter**, i.e. if I select MTA or MAT, I don't count this twice.

How many ways can this be done?

Let's list them out:

MAT, MAH, MTH, ATH- notice any other combination is simply a re-arrangement of one of the subsets listed here. Hence, if order does not matter then we have a total of 4 ways to select 3 letters from the word MATH. This can also be solved for by using the **choose** function.

Notation: $\binom{n}{r}$ = ${}_{n}C_{r}$. This is read "n choose r"

The formula is: $\frac{n!}{r!(n-r)!}$

Using Your Calculator: Try looking for an **nCr** button on your calculator

In R, this is done by using `choose(n,r)`

Some special cases: $\binom{n}{0} = 1, \binom{n}{n} = 1, \binom{n}{1} = n, \binom{n}{r} = \binom{n}{n-r}$

Binomial probability distribution

Define ***X= number of successes observed in n trials***

Then, ***X~Bin(n,p)*** has probability distribution: $\boldsymbol{P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}}$, where $\boldsymbol{x = 0, 1, \ldots, n}$

Example 3

The customer service department for a wholesale electronics outlet claims that 90% of all customer complaints are resolved to the satisfaction of the customer. In order to test this claim, a random sample of 15 customers who have filed complaints is selected. Define X = number of customers among 15 whose complaints were resolved to the customer's satisfaction.

1. Explain why X is a Binomial random variable.

   Give an expression for the following probabilities:
2. $P(X = 14)$
3. $P(X > 14)$
4. $P(X \leq 14)$
5. $P(9 < X \leq 12)$

Soln:

1. Under this experiment X is defined to have:

- **Two outcomes** – Resolved complaints to customer satisfaction vs. not
- **Independent trials** – it is assumed that each customer is unrelated to any other customer and so customer satisfaction is not related to any other customer.
- **Multiple trials** – the question indicates we surveyed 15 customers
- **Same of P(success) = p with each customer** – the question indicates that P(customer complaints is resolved to the satisfaction of the customer) =0.9 for all customers.
- Hence, X~Bin(n = 15, p = 0.9)

2. $P(X=14) = \binom{15}{14}(0.9)^{14}(1-0.9)^{15-14} = 0.34315$
3. $P(X>14) = P(X \geq 15) = P(X=15) = \binom{15}{15}(0.9)^{15}(1-0.9)^{15-15} = 0.20589$
4. $P(X \leq 14) = 1 - P(X>14) = 1 - c = 1 - 0.20589 = 0.79411$
5. $P(9 < X \leq 12) = P(X=10) + P(X=11) + P(X=12)$

$$= \binom{15}{10}(0.9)^{10}(1-0.9)^{15-10} + \binom{15}{11}(0.9)^{11}(1-0.9)^{15-11} + \binom{15}{12}(0.9)^{12}(1-0.9)^{15-12}$$
$$= 0.1818114$$

Expectation and Variance

If X~Bin(n; p) then:

- The expected value of X is $\boldsymbol{E(X) = \mu_X = np}$
- The variance of X is $\boldsymbol{Var(X) = \sigma_X^2 = np(1-p)}$

Example 4

In our customer satisfaction example, among the 15 customers sampled what is the average number that will express satisfaction to how their complaint was resolved. What is the standard deviation?

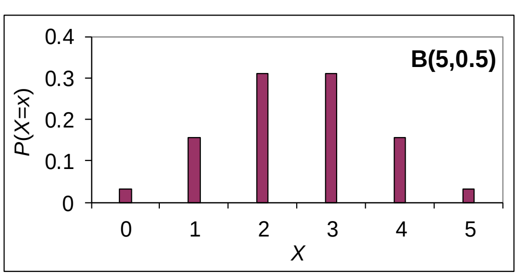
Soln:

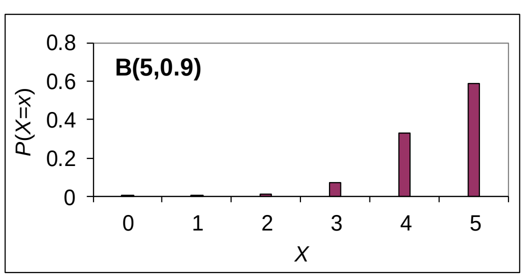
In this example X~Bin(n = 15, p = 0.9) Hence,

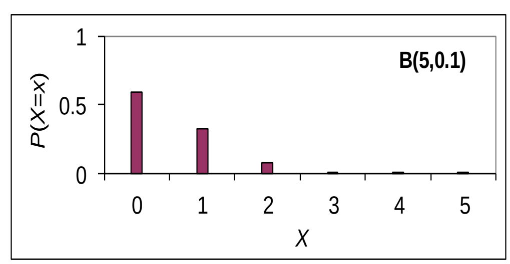
- $E(X) = n \times p = 15 \times 0.9 = 13.5$
- $Var(X) = n \times p \times (1-p) = 15 \times 0.9 \times 0.1 = 1.35$, And $std.dev(X) = \sqrt{1.35} = 1.16895$

Effect of changing *p* when *n* is fixed

Binomial distributions are skewed when *p* is close to 0 or close to 1 (especially if the sample is small). In the figures below notice that all have 5 trials and the distribution is most symmetric when p is at 0.5. The distribution creates tails as p moves to one of the extreme ends.

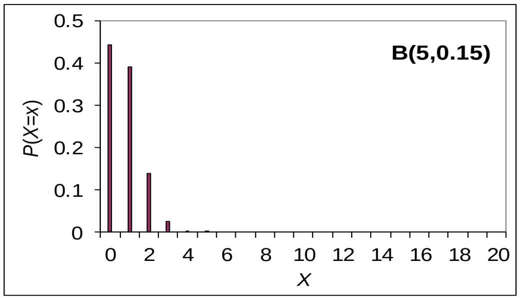


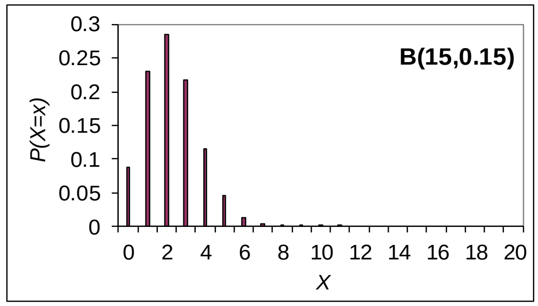


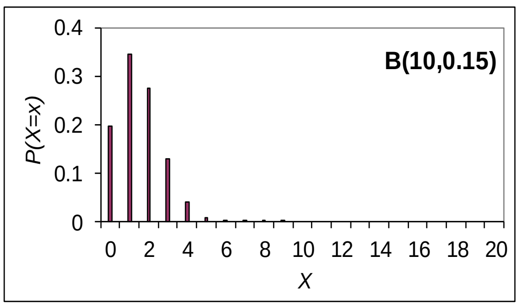


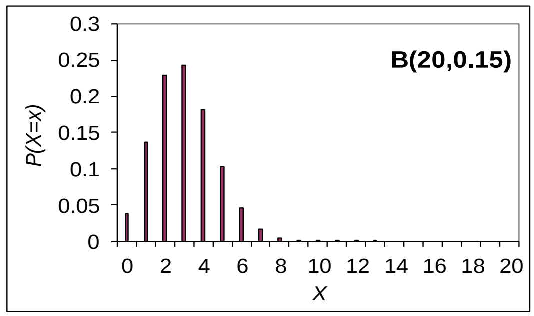
Effect of changing *n* for a fixed value of *p*

In these figures we have a p=0.15 (i.e a value closer to 0 which is why the distribution starts out very right skewed. Notice however that as n increases the distribution starts to spread out more this is occurring since as n increases the variance of the distribution also increases creating the larger spread.



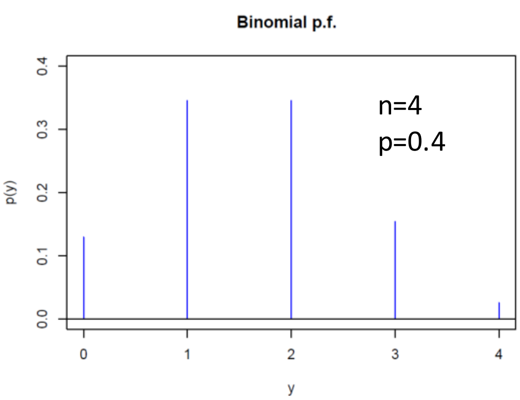


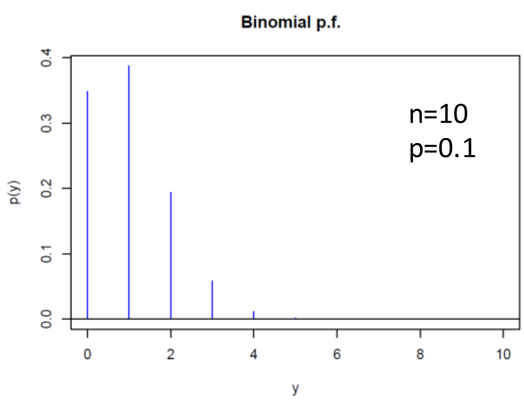


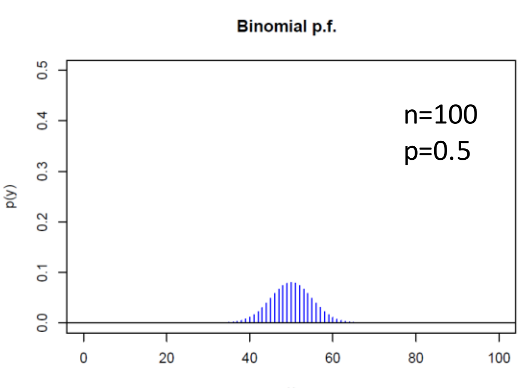


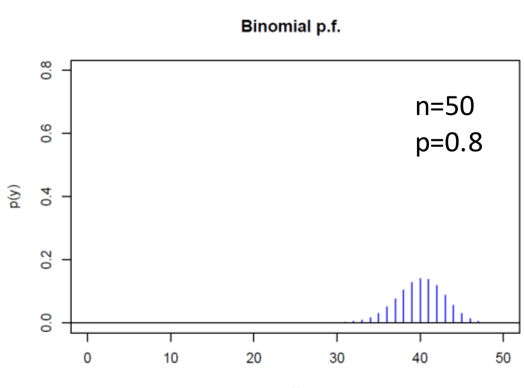
Effect of changing n and p

A few additional examples to show different combinations of n and p. Notice that regardless of the value of p, as n increases the distribution appears to become bell shaped and more symmetric in nature. We will discuss this notice further in a later week.









You Try 2

A venture capital firm has a list of potential investors who have previously invested in new technologies. On average, these investors invest about 5% of the time. A new client of the firm is interested in finding investors for a mobile phone application that enables financial transactions. An analyst at the firm starts by calling 10 potential investors. Let X= number of investors among 10 who are interested.

1. Explain why X is a Binomial random variable.
2. Find the probability that exactly 2 of them will be interested.
3. Find the probability that at least 1 but less than 3 will be interested.
4. Solve for the expected number who will be interested. Solve for the variance.

Section 5.8: Geometric distribution

Physical setup

Again, we have independent Bernoulli trials, each having two possible outcomes (Success vs. Failure). The probability, p, of success is the same each time. However now X represents the total number of trials needed to get the **first success**. This is the number of failures PLUS the one success.

Some examples include:

1. The probability you win a lottery prize in any given week is a constant p. Then the number of weeks till you win a prize for the first time is a geometric distribution.
2. The number of phone calls a telemarketer must make to get their first sale.

Probability function:

There is only <u>one way</u> to arrange $\boldsymbol{x-1}$ failures followed by $\mathbf{1}$ success; this leaves us with

$\mathbf{P}(\boldsymbol{X}=\boldsymbol{x})=(\mathbf{1}-\boldsymbol{p})^{\boldsymbol{x}-\mathbf{1}}\boldsymbol{p}$; Where $x = 1,2,3,\ldots$

Expectation and Variance

If X~Geo(p) then:

- The expected value of X is $\boldsymbol{E(X)=\mu_X=\frac{1}{p}}$
- The variance of X is $\boldsymbol{Var(X)=\sigma_X^2=\frac{1-p}{p^2}}$

We notice that both the Binomial and Geometric models assume:

- Two outcomes in each trial,
- Independent Trails, and
- Each trial has the same probability of success.

Example 5

A company receives 60% of its orders over the internet. Let X = total # of orders received by the company to receive the first order over the internet.

1. Justify why X is a Geometric random variable.
2. What is the probability that the fifth order received is the first internet order?
3. Find the expected number of orders received till the first internet order.

Soln:

1. For the random variable, X we have
    a. **Two outcomes:** Orders received via the internet vs. not via the internet
    b. **Independent trials**: it is assumed that every order is received in a way that is independent of how other orders are received.
    c. **Same P(success) in each trial:** every order has a p = 0.6 to be received over the internet.
    d. And finally, X counts the total number of failures + 1 success. Hence X~Geo(0.6)
2. $P(X = 5) = (1 - 0.6)^4(0.6)^1 = 0.01536$
3. $E(X) = \frac{1}{0.6} = 1.6667$

You Try 3

If the probability is 0.75 that an applicant for a driver's license will pass the road test on any given try, what is the probability that an applicant will finally pass the test on the fourth try?

You Try 4

Suppose a fair six-sided die is rolled repeatedly and, on each roll, you note whether a six occurs. Let X be the total number of rolls to get the first 'six'.

1. What is the distribution of X? Explain how it satisfies the underlying properties.
2. Find the probability that you roll a six on your 10^{th} trial.
3. Solve for the expected total number of trials for a six to occur for the first time.

Practice

Chapter 5: 11-17, 28-30, 39, 40, 57, 58, 62, 66(a-d)

Lesson 3: Poisson random variable and end notes

Key Takeaways

By the end of this lesson you should be able to:

- Identify and apply a Poisson distribution
- Distinguish between the various discrete distributions discussed.

Section 5.7: Poisson Distribution

Fun Fact: Named after the French Mathematician **Simeon Denis Poisson** who introduced it in 1837.

His name is one of the 72 names of French scientists, engineers and mathematicians engraved along the Eiffel Tower in recognition of their contributions.

Examples of random variables modelled by a Poisson distribution:

1. The number of calls arriving at a switchboard in a minute.
2. The number of bankruptcies that are filed in a month.
3. The number of customers arriving to their bank in a day.
4. The number of chocolate chips in a cookie.
5. The number of bacteria per ml of a solution.

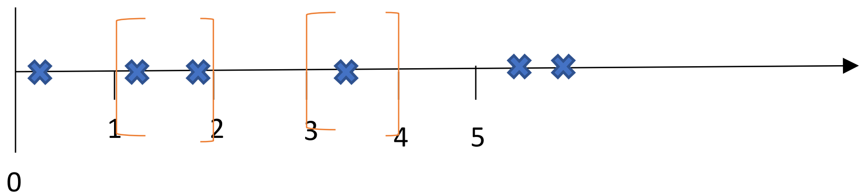
Do you see a pattern with these examples?

Physical setup

Run an experiment such that we have:

1. **Independent:** the # of occurrences in non-overlapping intervals are independent.
2. **Individuality:** Only **one event** can occur at the same time.
3. **Uniformity** (homogeneity): events occur at a uniform rate of λ over time.

Visually:



Consider a timeline starting at time t=0. The process begins to evolve having events occur randomly along the timeline at an average rate of λ per unit of time. Consider two non-overlapping intervals, (1;2] and (3;4]. Since these intervals are **non-overlapping** by property 1, the number of occurrences within them will be independent, i.e. the number of occurrences that happened in (1; 2] will not affect the number of events that will occur in (3; 4]. In addition, notice that these intervals are of the **same length**, hence by property 3 occurrences will happen at the same average rate for each interval.

Let **X = # of times an event occurs in an interval of length t** (this can be an interval of time, area, distance, or volume). Where the event occurs at a rate of λ per unit time.

Then we say **X~Poi(λ)** having probability function: $\mathbf{P}(\boldsymbol{X}=\boldsymbol{x})=\frac{\boldsymbol{e}^{-(\boldsymbol{\lambda t})}(\boldsymbol{\lambda t})^{\boldsymbol{x}}}{\boldsymbol{x}!}$, For $x=0,\ 1,\ 2,\ \ldots$

Alternative formula if we rescale the units such that an interval of length t is now considered 1 unit:

Let $\boldsymbol{\mu}=\boldsymbol{\lambda t}$, Then: $\mathbf{P}(\boldsymbol{X}=\boldsymbol{x})=\frac{\boldsymbol{e}^{-\boldsymbol{\mu}}(\boldsymbol{\mu})^{\boldsymbol{x}}}{\boldsymbol{x}!}$, Where $x=0,\ 1,\ 2,\ \ldots$

And we say X follows a **Poisson Process**

Difference between λ and μ

- $\boldsymbol{\lambda}$ is called the Intensity or Rate of occurrence. It represents the average rate at which events occur per unit of time.
- $\boldsymbol{\mu}=\boldsymbol{\lambda t}$ represents the average number of occurrences in t units of time.

Example 1

A coffee-shop manager wishes to provide prompt service for customers at the drive-through window. The average arrival rate is seven customers per 15-min period. Let X denote the number of customers arriving per 15-min period. Assume X has a Poisson distribution.

1. Find the probability that ten customers arrive in a particular 15-min period.
2. Find the probability that 20 customers arrive in a particular 30-min period.

Soln:

$X\sim Poi(\mu=7$ per 15 min period)

1. $P(X=10)=\frac{e^{-7}7^{10}}{10!}=0.07098$
2. For this we must first adjust the average rate of occurrence, $\mu=\frac{7}{15}\times 30=14$ customer. Then,

   $P(X=20)=\frac{e^{-14}14^{20}}{20!}=0.0286$

Expectation and Variance

If X~Poi($\mu = \lambda$t) then,

- The expected value of X is $\boldsymbol{E(X) = \mu_X = \lambda t}$
- The variance of X is $\boldsymbol{Var(X) = \sigma_X^2 = \lambda t}$

Example 2

For the coffee-shop from Example 1, solve for:

1. Average number of customers to arrive in a 15-min period.
2. Average number of customers to arrive in a 20-min period.
3. Average number of customers to arrive in a 5-min period.
4. Average number of customers to arrive in a 2 hour period.

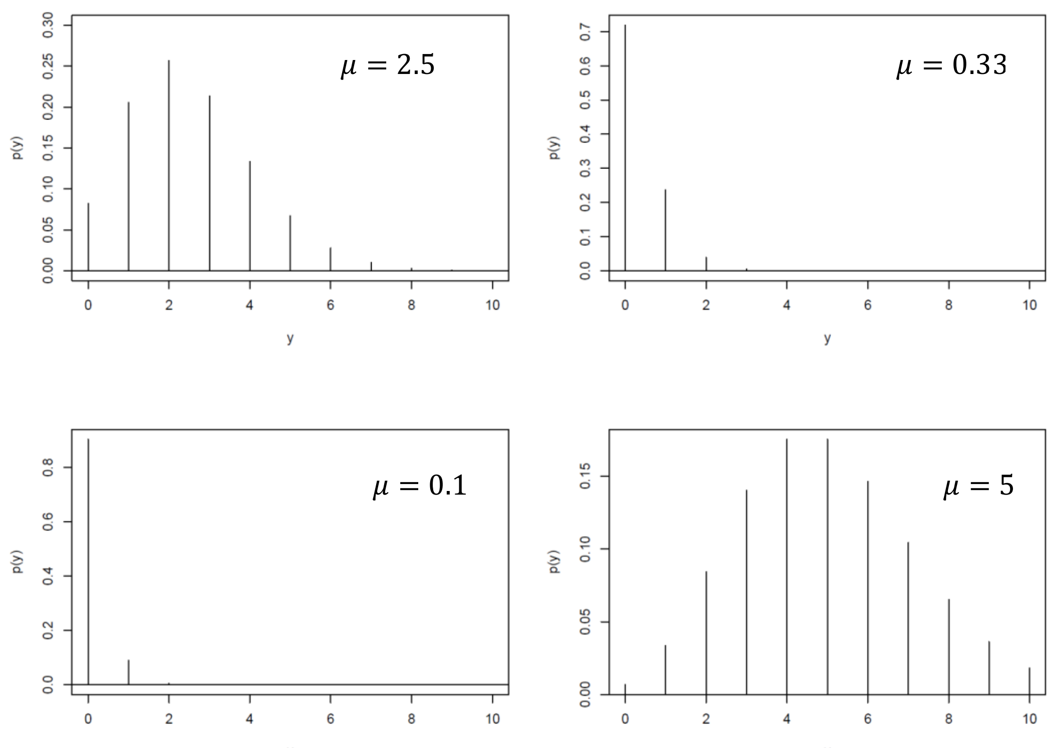
Soln:

$X \sim Poi(\mu = 7$ per 15min period)

1. $E(X) = 7$
2. $E(X) = \frac{7}{15} \times 20 = 9.333$
3. $E(X) = \frac{7}{15} \times 5 = 2.333$
4. $E(X) = \frac{7}{15} \times 120 = 56$ (note here we need to convert 2 hours to minutes so that the **units of the average rates match**). An alternative approach is to re-express 15mins as ¼ of an hour, hence we have $E(X) = \frac{7}{0.25} \times 2 = 56$.

Effect of changing μ

Generally, the larger the value of μ the more spread out the distribution and the shape becomes more symmetric.



Binomial vs. Poisson

When reading a question deciding on whether a random variable, X, is Binomial or Poisson etc. can be tricky. Below are two questions you can ask yourself to help identify which random variable you are working with.

1. Can we specify in advance the maximum value of X?
   Answering this question is helpful since for a Binomial random variable the experiment is repeated a **finite number of times** and this is known in advance (for example flipping a coin 50 times). Remember if
   - If X is BIN(n; p) then $x = 0, 1, \ldots, n$
   - If X is POI(λt) then $x = 0, 1, 2, \ldots$ (i.e no maximum known in advance)
2. Does it make sense to ask how often the event did NOT occur?
   By this we mean are you able to count the number of successes as well as the number of failures? This works for the Binomial distribution since by definition the distribution is **symmetric** (i.e. in 50 flips we can count the number of heads and also the number of tails observed). For a Poisson distribution this is not typically the case, e.g. in counting the number of potholes observed in a stretch of road, here we can count the number of potholes since something is physically there to count. It makes no sense to count its complement (number of no potholes, there is nothing there to count).

- If you answered YES to both, then X is BINOMIAL. Here are some other examples:
  - In looking back at out customer complaint example, we can count the # of satisfied customers and the # of unsatisfied customers.
  - When looking at tracking a disease, we can count the # of diseased or the # of non-diseased subjects in a study.
- If the answer is NO to both, then X is POISSON. Here are some other examples:
  - It makes no sense to ask how often a person does not hiccup.
  - It makes no sense to ask how often a secretary does not make an error when typing.

Example 3

Consider again our coffee-shop example, where customers arrive at an average rate of seven per 15-min interval. The coffee shop can currently serve up to ten customers per 15-min period without significant delay. Based on the Poisson distribution identified in an earlier example the $P(X \leq 10)$ in a 15-min interval can be shown to equal 0.90.

Consider now an 8-hour workday made up of 32, (8*4), non-overlapping 15-min intervals. What is the probability that among these 32 intervals at least 30 have 10 or less customers arrive?

Soln:

This question is dealing with two random variables. The first is X, which we have seen before:

- X = # of customers arriving in a specified interval.

The second random variable is defined by the new question of interest. Here we want:

- Y = # of 15min intervals among 32 that have "10 or less customers arrive".

What is the distribution of Y?

First recognise that Y has a maximum value that we can specify in advance, it is counting the number of trials among 32 that satisfy a particular event of interest, hence $y = 0, 1, 2, \ldots, 32$.

Second recognise that we can count both success and failure. If we see more than 10 customers in a 15-min interval this is considered a failure.

Observing these two characteristics we get the hint that Y is Binomial. Let us run through the underlying assumptions of a Binomial and see if Y meets all the criteria:

- **Two outcomes**: "≤ 10 customers in 15mins" vs. "> 10 customers in 15mins"
- **Independent trials**: Here a trial is a 15min interval, notice we are considering non-overlapping intervals hence by the independence assumption of the Poisson distribution the number of customers to arrive in each interval is independent of every other interval. This implies that whether we get a success or failure in each trial is independent of other trials.
- **Multiple trials**: we have a total of n=32 intervals (trials)
- **Same P(success) in each trial:** This comes from the uniformity assumption of the Poisson distribution. Here P(success) = $P(X \leq 10)$ in a 15-min interval = 0.9 and this is the same regardless of which 15-min interval we are considering.
- Hence **Y~Bin(n = 32, p = 0.9)**

Now what do we want to solve for? We want:

$$\begin{aligned} P(Y \geq 30) &= P(Y = 30) + P(Y = 31) + P(Y = 32) \\ &= \binom{32}{30}(0.9)^{30}(0.1)^2 + \binom{32}{31}(0.9)^{31}(0.1)^1 + \binom{32}{32}(0.9)^{32}(0.1)^0 = 0.3666835 \end{aligned}$$

Poisson as an approximation to the Binomial

To make things a little tricker, under certain assumptions the Poisson distribution can approximate the Binomial distribution well. After a bit of mathematics, we find that the Poisson distribution with $\boldsymbol{\lambda = np}$ closely approximates the Binomial distribution if ***n is large*** $(n \to \infty)$ and **p is small** $(p \to 0)$.

Example 4

Suppose that an automobile parts wholesaler claims that 0.5% of the car batteries in a shipment are defective. A random sample of 200 batteries is taken, and four are found to be defective.

1. Solve for the exact probability that four or more batteries among 200 are found to be defective.
2. Use the Poisson approximation to find the probability that four or more car batteries in a random sample of 200 would be found defective.

Soln:

1. Let X = # of defective batteries among 200.
    - X~Bin(n = 200, p = 0.005)

- $P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 2)] = 0.01868$

2. Notice that n is large and p is small, so here our approximation is such that $\lambda = n \times p = 200 \times 0.005 = 1$

$$\begin{aligned} P(X \geq 4) = 1 - P(X < 4) &= 1 - P(X \leq 3) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 2)] \\ &= 1 - e^{-1}\left[\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!} + \frac{1^3}{3!}\right] = 0.01899 \end{aligned}$$

Notice that the approximation is the same as the exact probability to 3 decimal points. This gets better as ***n gets larger*** $(n \to \infty)$ and **p gets smaller** $(p \to 0)$.

Practice

Chapter 5: 21-27, 47, 48, 51, 64

Summary Table

This information is also in the formula sheet posted in the Course Information folder on Learn.

Distribution	**Random Variable**	**Probability Function**	**Expectation**	**Variance**
Bernoulli	$X = \begin{cases} 1 \text{ if a success occurs} \\ 0 \text{ if a failure occurs} \end{cases}$	$p^x(1-p)^{1-x}$ $x = 0,1$	p	$p(1-p)$
Binomial X~Bin(n,p)	X=# of successes among n trials	$\binom{n}{x} p^x(1-p)^{n-x}$ $x = 1,2,\dots,n$	np	$np(1-p)$
Geometric X~Geo(p)	X= Total # of trials to the 1st success	$(1-p)^{x-1}p$ $x = 1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson X~Poi(λ)	X=# of occurrences within an interval of length of t.	$\frac{e^{-(\lambda t)}(\lambda t)^x}{x!}$ $x = 0,\ 1,\ 2,\dots$	$\mu = \lambda t$	$\mu = \lambda t$

R component

We are back to coding this week (kind of). A discrete random variable, X, is described as a random variable that takes on a countable or infinitely countable set of values.

Recall: We use two functions to describe the behaviour of a discrete random variable:

Probability Function:

The probability that a discrete random variable X, takes on a particular value, x, that is $p(x) = P(X = x)$ is frequently called the probability function (p.f.) or probability mass function (p.m.f.)

Cumulative Distribution Function:

The cumulative distribution function (cdf) for any random variable X is defined as $F(x) = P(X \leq x)$. It defines a nondecreasing function where for a discrete random variable X, the cdf is solved as

$$F(a) = \sum_{x \leq a} f(x)$$

Discrete Random Variable Names in R

Distribution Name	R extension
Binomial	`binom(x,n,p)`
Bernoulli	`Binom(x,1,p)`
Geometric	`geom(x,p)`
Poisson	`pois(x,` μ`)`

To get a full list of the distributions available in R you can use the following command:

`help(Distributions)`

For every distribution there are four commands. The commands for each distribution are prepended with a letter to indicate the functionality:

		Notation for Discrete random variables ('?' Represents value returned)
"d"	returns the height of the probability density function	$p(x) = P(X = x) = ?$
"p"	returns the cumulative density function	$F(x) = P(X \leq x) = ?$
"q"	returns the inverse cumulative density function (quantiles)	$F^{-1}(x) = P(X \leq ?) = q$
"r"	returns randomly generated numbers	

Binomial Probabilities

In R, you would find $p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ probability using: `dbinom(x,n,p)`

Alternatively you can obtain, $\mathrm{F}(x) = P(X \leq x)$ using: `pbinom(x,n,p)`

Geometric Probabilities

In R, you would find $P(X = x) = (1-p)^x p$ using: `dgeom(x,`p`)`

Alternatively you can obtain, $Pr(X \leq x)$ using: `pgeom(x,`p`)`

Poisson Probabilities

In R, you would find $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$ using: `dpois(x,`μ`)`

Alternatively you can obtain, $Pr(X \leq x)$ using: `ppois(x,`μ`)`

I have attached a R file in the Week 5 folder on Learn that shows how to use these functions and how to create some of the plots in this week's lesson document. You will not be tested directly on these commands, but knowing how to use them may make a quiz or exam easier.