Chapter 3

Random Variables

So far we have explored *random experiments* that produce *events* with certain *probabilities*. *Random Variables* introduce a new way to examine the events for an experiment.

Example 1: Consider an experiment where two three-sided dice are rolled. The sample space for the experiment is as follows.

$$S = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$$

From the last chapter we know how to calculate the probability of any event occurring. We could find $P(\text{roll a one})$ or $P(\text{roll an even})$.

Now consider the number of ones that are rolled in each outcome; there could be zero, one, or two. The number of ones rolled is a **random variable** - it varies randomly between iterations of the experiment.

Notice that **every** event in the sample space can be characterized by the number of ones it has. Random variables are often denoted with capitol letters. Let's let X be the number of ones rolled. (**Note:** $P(X = x)$ stands for the probability the random variable X takes on the value x, this often written shorthand as $p(x)$)

$$\begin{aligned} P(X=0) &= P(22)+P(23)+P(32)+P(33)=\frac{4}{9} \\ P(X=1) &= P(12)+P(13)+P(21)+P(31)=\frac{4}{9} \\ P(X=2) &= P(11)=\frac{1}{9} \end{aligned}$$

Also notice that the probability of rolling one, two, **or** three 1's is 1 as this is the probability of **any** event occurring. (Remember we can add the probabilities as events are mutually exclusive).

$$\begin{aligned} P(X=0,\ \text{or}\ X=1,\ \text{or}\ X=2) &= P(X=0)+P(X=1)+P(X=2) \\ &= \frac{4}{9}+\frac{4}{9}+\frac{1}{9} \\ &= 1 \end{aligned}$$

3.1 Types of Random Variables

There are both *qualitative* and *quantitative* random variables. We will explore the various types of random variables in more detail in later chapters. Let's start by focusing only on quantitative (numeric) random variables. There are two different types of quantitative random variables we need to be familiar with.

A ***Discrete Random Variable*** is a variable that can take on a *countable* number of values. Consider some examples.

- Number of baskets made
- Number of pets
- Number of movies watched

A ***Continuous Random Variable*** is a variable that can take on a *non-countable* or *infinite* number of values. Consider some examples.

- Height
- Age
- Weight

We will explore both types of variables throughout this chapter. Let's start with discrete random variables.

3.2 Discrete Random Variables

Discrete Random Variables
A **discrete random variable** is a numerical quantity that varies between iterations of an experiment and can only take on a *countable* number of values. Some examples are number skittles eaten, and number of cats owned.

We often denote a random variable with the a capitol letter, like X. We denote the probability X takes on a particular value, x, as $P(X = x)$ or $p(x)$. The function that takes values of x and returns respective probabilities is referred to as the **probability mass function**, or **p.m.f.**

Properties of Discrete Random Variables
For all values of x

$$0 \leq p(x) \leq 1$$

$$\sum_{\text{All } x} p(x) = 1$$

This just means the probability of **any** event occurring, so it seems intuitive that is should be 1.

3.2.1 Probability Distributions for Discrete Random Variables

Example 1: Suppose you flip two coins, let X be the number of heads that appear. X can take on the values zero, one, or two. Let's start by writing the sample space.

$$S : \{TT, HT, TH, HH\}$$

(a) What is $p(0)$?

It is clear that $p(0) = \frac{1}{4} = 0.25$

(b) What is $p(1)$?

It is clear that $p(1) = \frac{2}{4} = \frac{1}{2} = 0.5$

(c) What is $p(2)$?

It is clear that $p(2) = \frac{1}{4} = 0.25$

The **probability distribution** for a random variable refers to the summary of probabilities for all possible values the variable can take on.

We have already shown the distribution through formulas summarizing $p(0)$, $p(1)$, and $p(2)$. We have two other methods for visualizing the distribution; tables, and graphs.

(d) Visualize the probability distribution for X using a **table**.

x	0	1	2
$P(X = x)$	0.25	0.5	0.25

We may also be interested in the *cumulative probability* at each value of x. or $P(X \leq x)$ for all values of x. We can also represent this using a table.

x	0	1	2
$P(X \leq x)$	0.25	0.75	1.00

(e) Visualize the *probability distribution* for X using a **graph**.

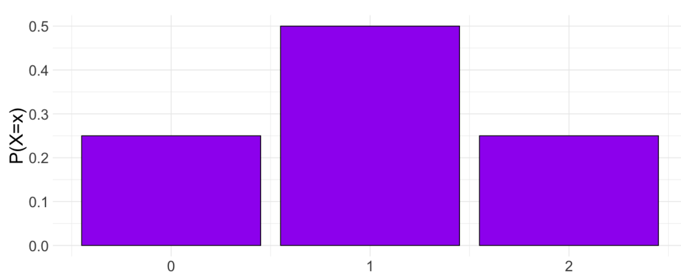

Figure 3.1: Probability Distribution for X

Notice the area of the blocks represents probabilities, and the area of all blocks must sum to one.

(f) Visualize the *cumulative probability distribution* for X using a **graph**.

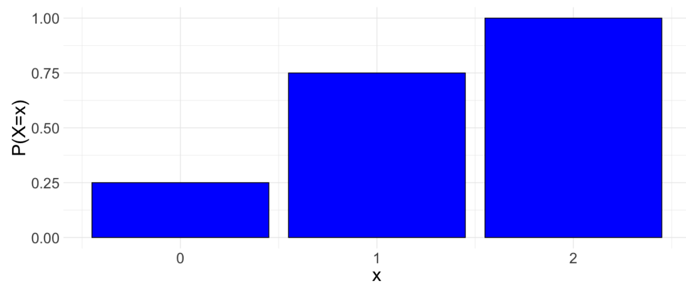

Figure 3.2: Cumulative Probability Distribution for X

Our distribution represents the *theoretical probability* of every outcome in a random experiment. Our distribution does not describe for certain what will happen, but instead what will happen *on average*.

3.2.2 The Mean of a Discrete Random Variable

The **mean** or **expected value** of a discrete random variable X is denoted $E(X)$ or μ_x and describes the average value of X if the experiment is conducted a large (theoretically infinite) amount of times.

Example 2: A platoon of stormtroopers are gathered for training. In the Death Star's shooting range the troopers are firing blasters at targets from 10 meters away. Stormtroopers are notoriously bad shots. Suppose that single a stormtrooper makes one out of five shots.

Suppose each stormtrooper is given 3 shots. Let X be the number of targets that the stormtrooper blasts.

(a) Estimate the expected value of X, $E(X)$, using simulation

1. Here we want to simulate the number of shots the Stormtrooper get's out of three. The possible values for X are 0, 1, 2, and 3. We know that the stormtrooper makes 1 out of 5 shots so there is a 0.2 or 20% chance the trooper will make their shot.
2. Here we are assuming that the stormtroopers make 20% of shots, and that each shot is independent.
3. We will assign the numbers $0 - 19$ for targets blasted and assign the numbers $20 - 99$ for targets missed.
4. Next we will run simulations:

Simulation Number	Simulation	X
1	76 21 91	0
2	57 36 38	0
3	54 18 00	2
4	05 56 54	1

5. Next we estimate $E[X]$ to be the average of X for the simulations. $E[X] \approx \frac{3}{4} = 0.75$. Just four simulations is a pretty bogus approximation, running 100000 simulations in R gives $E[X] \approx 0.60046$

(b) Show the probability distribution table for X

x	0	1	2	3
$P(X=x)$	0.512	0.384	0.096	0.008

- $P(X=0) = (0.80)(0.80)(0.80) = 0.512$
- $P(X=1) = (0.20)(0.80)(0.80) + (0.80)(0.20)(0.80) + (0.80)(0.80)(0.20) = 0.384$
- $P(X=2) = (0.80)(0.20)(0.20) + (0.20)(0.80)(0.20) + (0.20)(0.20)(0.80) = 0.096$
- $P(X=3) = (0.20)(0.20)(0.20) = 0.008$

But how do we calculate the theoretical expectation, $E(X)$, without simulation? Suppose you run n simulations for the random variable x.

(c) How many of the n simulations would you expect to result in zero blasted targets?

As there is a 0.512 chance of stormtroopers blasting zero targets we would expect $0.512n$ of simulations to result in $X = 0$

(d) How many of the n simulations would result in one blasted target?

We would expect $0.384n$ of simulations to result in $X = 1$

(e) How many of the n simulations would result in two blasted targets?

We would expect $0.096n$ of simulations to result in $X = 2$

(f) How many of the n simulations would result in three blasted target?

We would expect $0.008n$ of simulations to result in $X = 3$

(g) Now calculate $E(X)$ by taking the average value for X over the n simulations

$$\begin{aligned} E[X] &= \frac{0.512n(0) + 0.384n(1) + 0.096n(2) + 0.008n(3)}{n} \\ &= 0.512(0) + 0.384(1) + 0.096(2) + 0.008(3) \\ &= 0.6 \end{aligned}$$

Notice how the n terms cancel out. After the cancellation we have stumbled upon the formula for expected value.

Expected Value for a Discrete Random Variable
The expected value for a discrete random variable X is denoted $E(x)$ or μ_X and describes the average value for X over a large number of trials.

$$\begin{aligned} E[X] &= x_1 \cdot p(x_1) + x_2 \cdot p(x_2) + \cdots x_n \cdot p(x_n) \\ &= \sum_{i=1}^{n} x_i \cdot p(x_i) \end{aligned}$$

This is also sometimes notated as

$$E[X] = \sum_{\text{all } x} x \cdot p(x)$$

They both mean the same thing, simply take the sum of all possible values of x multiplied by their respective probabilities.

Properties of Discrete Random Variables

$$E(c) = c$$

For some constant c.

$$E(aX) = aE(X)$$

For some constant a. Also note that expect

$$E(aX + c) = aE(X) + c$$

Example 3: Suppose that a carnival game costs \$2 to play once and offers a \$0 payout 55% of the time, a \$2 dollar pay out 30% of the time , and a \$3 pay out 15% of the time. What are your expected winnings if you played the game a large number of times?

Here we let W be the winnings for a round of the carnival game. Then we summarize the probability distribution in a table

w	-2	0	1
$P(W = w)$	0.55	0.30	0.15

Now we may compute the expected value for W

$$\begin{aligned} E[W] &= -2(0.55) + 0(0.3) + 1(0.15) \\ &= -0.95 \end{aligned}$$

This means if you played the game a theoretically infinite number of times you would expect to loose an average of \$0.95 a game.

What if you played this game a large number of times? Would would the distribution of your winnings look like?

Example 4: Suppose you flip two coins and let X be the number of heads that appear. The probability distribution for X is shown below. What is $E(X)$?

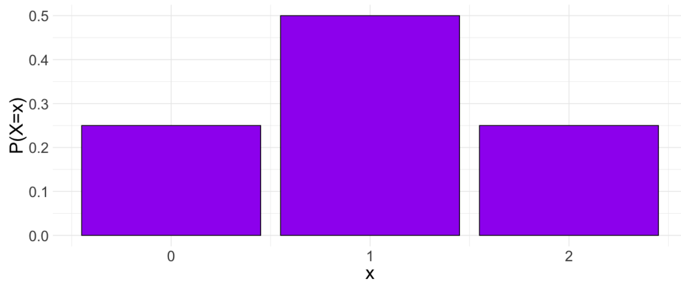

Figure 3.3: Probability Distribution for X

In this example $E(X)$ represents the average number of heads that appear if the two coins are flipped a large number of times. Let's use the formula here

$$\begin{aligned} E(X) &= \sum_{i=1}^{3} x_i \cdot p(x_i) \\ &= x_1 \cdot p(x_1) + x_2 \cdot p(x_2) + x_3 \cdot p(x_3) \\ &= 0(0.25) + 1(0.5) + 2(0.25) \\ &= 1 \end{aligned}$$

This means that we would expect an average of 1 head over a large amount of trials.

3.2.3 The Variance of a Discrete Random Variable

The **variance** of a discrete random variable X is denoted $Var(X)$ or σ_x^2 and represents how spread out all values of X are about the mean of X.

Example 5: In a game of poker you are dealt five cards from a standard deck. Let Y be the number of red cards that appear in a randomly dealt hand.

(a) Determine the probability distribution for Y.

y	0	1	2	3	4	5
$P(Y=y)$	0.025310124	0.149559824	0.325130052	0.325130052	0.149559824	0.025310124

- $P(Y=0) = \frac{\binom{26}{0}\binom{26}{5}}{\binom{52}{5}} = 0.025310124$
- $P(Y=1) = \frac{\binom{26}{1}\binom{26}{4}}{\binom{52}{5}} = 0.149559824$
- $P(Y=2) = \frac{\binom{26}{2}\binom{26}{3}}{\binom{52}{5}} = 0.149559824$
- $P(Y=3) = \frac{\binom{26}{3}\binom{26}{2}}{\binom{52}{5}} = 0.325130052$
- $P(Y=4) = \frac{\binom{26}{4}\binom{26}{1}}{\binom{52}{5}} = 0.149559824$
- $P(Y=5) = \frac{\binom{26}{5}\binom{26}{0}}{\binom{52}{5}} = 0.025310124$

(b) Calculate $E(Y)$

$$\begin{aligned} \mu_x &= 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + 4 \cdot p(5) + 6 \cdot p(6) \\ &= 2.5 \end{aligned}$$

(c) Estimate $Var(Y)$ using simulation

1. Here we are interested in the number of red cards that appear in our hand. There may be 0, 1, 2, 3, 4, or 5.
2. We are assuming that there are 13 cards in a deck of 52. As we are sampling *without replacement*. It means that we may not use the same number twice.
3. We will assign the numbers 0-25 to red cards, and 26-99 and other cards.
4. Now let's run some simulations.

Simulation Number	Simulation	Y	$Y-\mu_Y$	$(Y-\mu_Y)^2$
1	41 12 25 35 44	2	-0.5	0.25
2	36 47 06 04 28	2	-0.5	0.25
3	12 17 14 38 06	4	1.5	2.25
4	34 26 10 15 03	3	0.5	0.25

5. To estimate the variance of Y we take the average of the simulations for $(Y-\mu_y)^2$. Our four simulations give an estimated variance of $Var(Y) \approx 0.75$. In R 10000 simulations gives $Var(Y) \approx$ 1.15164

Suppose now that you simulated the random variable n times.

(d) How many of the n simulations would you expect result in no red cards being drawn?

Out of the n simulations we would expect $np_0 = n \cdot 0.025310124$ of simulations to result in zero red cards being drawn.

(e) Calculate $Var(Y)$ by taking the average of $(Y-\mu_Y)^2$ over the n simulations.

We expect np_i observed values of each y_i, $i = 1, 2, 3, 4, 5$. This tells us how many terms we expect to have for $(Y-\mu_Y)^2$ in each category.

$$\begin{aligned} Var(Y) &= \frac{np_0(0-\mu_Y)^2 + np_1(1-\mu_Y)^2 + \cdots + np_5(5-\mu_Y)^2}{n} \\ &= p_0(0-\mu_Y)^2 + p_1(1-\mu_Y)^2 + \cdots + p_5(5-\mu_Y)^2 \\ &= 1.151961 \end{aligned}$$

Notice here that the n terms simply cancel out. This leaves us with a mathematical definition of variance.

$$Var(Y) = E[(Y-\mu_Y)^2]$$

The units for variance is the square of the original units for the random variable. In the past example it went from (number of red cards) to (number of red cards)^2. We will talk about why we square each difference in more detail in upcoming chapters. To *standardize* the units, we take the square root of variance which leaves us with the **standard deviation**.

The standard deviation for a random variable Y is denoted σ_Y or $SD(Y)$.

$$SD(Y) = \sqrt{Var(Y)}$$

You can see how the σ notation is useful

$$\sigma_Y = \sqrt{\sigma_Y^2}$$

Another Useful Formula

We may also expand $Var(Y) = E[(Y - \mu_y)^2]$ to create another simple formula that can be used to calculate variance.

$$\begin{aligned} Var(Y) &= E[(Y - \mu_Y)^2] \\ &= E[Y^2 - 2\mu_Y Y + \mu_Y^2] \\ &= E[Y^2] - E[2\mu_Y Y] + E[\mu_Y^2] \\ &= E[Y^2] - 2\mu_Y E[Y] + \mu_Y^2 \end{aligned}$$

Next recall that $E[Y] = \mu_y$

$$\begin{aligned} &= E[Y^2] - 2\mu_Y^2 + \mu_Y^2 \\ &= E[Y^2] - \mu_Y^2 \\ &= E[Y^2] - E[Y]^2 \end{aligned}$$

Variance for a Discrete Random Variable
The **variance** of a discrete random variable X is denoted $Var(X)$ or σ_x^2 and represents how spread out all values of X are about the mean of X.

$$Var(X) = \sigma_X^2 = E[(X - \mu_X)^2] = E[X^2] - E[X]^2$$

Standard Deviation also describes how spread out the values of X are about the mean, but it uses the **same units** as X.

$$SD(X) = \sigma_X = \sqrt{Var(X)}$$

Example 6: Lets return to our clumsy stormtrooper example. We let X be the number of shots out of three a stormtrooper blasts while training. The probability distribution table is shown below. We also shows that $E(X) = 0.6$.

x	0	1	2	3
$P(X = x)$	0.512	0.384	0.096	0.008

(a) Calculate $Var(X)$
We know $Var(X) = E(X^2) - E(X)^2$. We already have $E(X)^2 = 0.6^2 = 0.36$. Next lets find $E(X^2)$.

x^2	0	1	4	9
$P(X = x)$	0.512	0.384	0.096	0.008

$$\begin{aligned} E(X^2) &= 0(0.512) + 1(.384) + 4(0.096) + 9(0.008) \\ &= 0.84 \end{aligned}$$

Now we have $Var(X) = E(X^2) - E(X^2) = 0.84 - 0.36 = 0.48$

(b) Calculate $SD(X)$
Simply take $SD(X) = \sqrt{Var(X)} = \sqrt{0.48} = 0.6928203$

Common Discrete Random Variables

What we will do next is classify several different types of discrete random variables. We will also look at the expected value and variance for our different types of random variables.

3.3.1 Bernoulli Random Variable

Bernoulli random variables are the simplest of all. They describe an experiment that results in one of two outcomes; a *success* that occurs with probability p and a *failure* that occurs with probability $q = 1 - p$.

Example 1: Suppose a basketball player makes 4 out of 5 baskets and takes one single shot. Let X be the number of shots the player makes.

(a) What is $P(X = 0)$

This is simply the probability of missing $P(X = 0) = 1 - 0.85 = 0.15$

(b) What is $P(X = 1)$

This is simply the probability of making the shot $P(X = 1) = 0.85$

(c) What is $E(X)$?

Here we have $E(X) = 0(0.15) + 1(0.85)$. Notice that for any Bernoulli random variable, X, we have

$$E(X) = 0(1 - p) + 1p = p$$

(d) What is $Var(X)$? First of $E(X)^2 = p^2$ and $E(X^2) = 0(1 - p) + 1(p) = p$.

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= p^2 - p \\ &= p(1 - p) \end{aligned}$$

For this particular example we have $Var(X) = (0.85)(0.15)$

Bernoulli Random Variables: Let X be a Bernoulli Random Variable

1. **Probability Mass Function:** $P(X = x) = p(1 - p)^{x-1}, \quad x = 0, 1$
2. **Expected Value:** $E(X) = p$
3. **Variance:** $Var(X) = p(1 - p)$

3.3.2 Binomial Random Variables

Binomial random variables are a simple extension of Bernoulli random variables. It describes an experiment where multiple *independent* Bernoulli trials are performed.

Example 2: Now let's suppose a basketball player makes 4 out of 5 baskets but takes three shots. Now let X be the number of shots the player makes.

(a) What is $P(X = 1)$? Here need to think not only of the probability of missing all shots but the order in which it can happen. There are three scenarios we are interested in

Scenario	Outcome	Probability
1	Success, Failure, Failure	(0.85)(0.15)(0.15)
2	Failure, Success, Failure	(0.15)(0.85)(0.15)
3	Failure, Failure, Success	(0.15)(0.15)(0.85)

We are interested in scenario 1, 2, **or** 3, so we add the probabilities

$$\begin{aligned} P(X = 1) &= (0.85)(0.15)(0.15) + (0.15)(0.85)(0.15) + (0.15)(0.15)(0.85) \\ &= 3(0.85)(0.15)^2 \\ &= 0.057375 \end{aligned}$$

Notice for 3 shots the number of ways you can make 1 shot and 2 failures is $\binom{3}{1}$. This is the same as the number of unique arrangements you can have for the letters SFF.

(b) What is $P(X = 2)$

Let's try to know be a little bit more creative with our solution now. We know there will be $\binom{3}{2} = 3$ ways that a person can score two baskets and miss one. So we have:

$$\begin{aligned} P(X = 2) &= 3(0.85)^2(0.15) \\ &= 0.325125 \end{aligned}$$

(c) What is $P(X > 0)$?

Now we are interested in any scenario where the player makes more than no shots. We know that all probabilities must sum to one

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$$

If we want $P(X = 1) + P(X = 2) + P(X = 3)$ we can calculate the three probability of the three events **or** you can find $1 - P(X = 0)$ (much easier!).

$$\begin{aligned} P(X > 0) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= 1 - P(X = 0) \\ &= 1 - (0.15)^3 \\ &= 0.996625 \end{aligned}$$

In other words theres a very good chance the player will score 1, 2 or 3 shots in the experiment.

(d) Create a probability distribution table for X.

x	0	1	2	3
$P(X = x)$	0.003375	0.057375	0.325125	0.614125

- $P(X = 0) = (0.15)^3 = 0.003375$
- $P(X = 1) = 3(0.85)(0.15)^2 = 0.057375$
- $P(X = 2) = 3(0.85)^2(0.15) = 0.325125$
- $P(X = 3) = (0.85)^3 = 0.614125$

(e) What is $E(X)$?

Let's use our formula for expectation.

$$\begin{aligned} E(X) &= 0(0.003375) + 1(0.057375) + 2(0.325125) + 3(0.614125) \\ &= 2.55 \end{aligned}$$

For binomial random variables, there is a nice shortcut that we may use. $E(X) = np$. We will leave the proof for this as a **challenge** exercise.

(f) What is $Var(X)$?

We know that $E(X)^2 = (2.55)^2$, then lets find $E(X^2)$

$$\begin{aligned} E(X^2) &= 0(0.003375) + 1(0.057375) + 4(0.325125) + 9(0.614125) \\ &= 6.885 \end{aligned}$$

Now we may find variance

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= 6.885 - (2.55)^2 \\ &= 0.3825 \end{aligned}$$

There is also a shortcut formula we can use to determine the variance of a binomial random variable $Var(X) = np(1 - p)$. Proving this is left as a **challenge** exercise.

Binomial Random Variables Let X be a binomial random variable with parameters n, and p. This can also be denoted as $X \sim \text{Binomial}(n, p)$

1. **Probability Mass Function**: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$
2. **Expected Value:** $E(X) = np$
3. **Variance:** $Var(X) = np(1 - p)$

Example 3: Suppose that an employee is late to work approximately two days per year.

(a) What is the probability that an employee is late 4 days in a given year?

Let's let X be the number of days the employee is late in the year. X is a binomial random variable with $n = 365$ and $p = \dfrac{1}{365}$. We have

$$\begin{aligned} P(X=5) &= \binom{365}{5}\left(\frac{1}{365}\right)^5\left(\frac{364}{365}\right)^{360} \\ &= 0.00301952041 \end{aligned}$$

Remember that we are assuming on each day the employee has a $\frac{1}{365}$ chance of being late, and that the chance of being late on each day is independent.

(b) What is the probability that the employee is late at least 1 day in a given year?

Here we are in interested in every value of X with the exception of 0.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X=0) \\ &= 1 - \left(\frac{364}{365}\right)^{365} \\ &= 0.63262507565 \end{aligned}$$

This seems relatively likely.

(c) What is the probability that the employee is late for one day in a year or four days in a year?

Now we simply take the sum of $P(X = 1)$ and $P(X = 4)$

$$\begin{aligned} P(X=1 \cup X=4) &= P(X=1) + P(X=4) \\ &= \binom{365}{1}\left(\frac{1}{365}\right)^1\left(\frac{364}{365}\right)^{364} + \binom{365}{4}\left(\frac{1}{365}\right)^4\left(\frac{364}{365}\right)^{361} \\ &= 0.36838419242 + 0.01522306703 \\ &= 0.3836073 \end{aligned}$$

(d) On average, how many days a year would you expect the employee to be late.

We know that $E(X) = np$. so $E(X) = 365\left(\frac{1}{365}\right) = 1$. We expect the employee to be late one day per year.

(e) Let X be the number of days the employee is late in a year. What is $Var(X)$?

We know that $Var(X) = np(1-p) = 365\left(\frac{1}{365}\right)\left(\frac{364}{365}\right) = 0.9972603$

3.3.3 Geometric Random Variables

Next up we will explore geometric random variables. Geometric random variables also involve independent Bernoulli trials with probability of success p and probability of failure $(1-p)$. A geometric random variable X denotes the number of trials until the first success.

Example 4: 5% of the trade federations battle droids are manufactured with defective targeting systems. You are an engineer on Geonosis who is randomly inspecting droids.

(a) What is the probability the 4th droid you inspect was manufactured with a defective targeting system?

Let's let X be the number of droids it takes to inspect before finding a defective droid. X is a geometric random variable. We are interested in a scenario where you *fail* three times, and then *succeed* on the fourth time.

$$\begin{aligned} P(X=2) &= (0.95)(0.95)(0.95)(0.05) \\ &= (0.95)^3(0.05) \\ &= 0.04286875 \end{aligned}$$

(b) What is the probability you need to inspect two or more new droids before you find one with a defective targeting system?

Here we are interested in $P(X \geq 2)$. Notice that X could take on any integer value above two.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 1) \\ &= 1 - 0.05 \\ &= 0.95 \end{aligned}$$

Let X be the number of new droids it takes to randomly inspect before you find one with a defective targeting system.

(c) Why is it difficult to visualize the probability distribution for X in a table?

X can take on any positive whole number value so the table would extend forever.

(d) Use simulation to estimate $E(X)$

1. Here we are interested in simulating the number of new droids it takes to randomly inspect before finding a defective one.
2. We are assuming there is a $\frac{5}{100}$ chance of finding a defective choice each trial. We are also assuming that probability of selecting a defective droid is independent between trials.
3. We will assign the numbers 0-4 to finding a defective droid. We will assign the numbers 5-99 to finding a functioning droid.
4. Now we will run simulations by randomly selecting numbers.

   Simulation 1: 72 1

   Simulation 2: 74 26 59 91 27 34 62 44 26 18 16 50 24 67 29 15 28 79 78 40 92 5

   Simulation 3: 6 38 9 15 72 11 61 45 60 99 9 74 62 91 35 42 70 88 89 42 11 63 9 8 73 39 83 15 44 80 11 61 55 19 6 8 19 34 4

   $\vdots$

5. On the first simulation it took inspecting 2 droids to find a defect, on the second it took 22, and on the third it took 39. Using our 3 simulations we would estimate $E(X)$ to be 21 droids. Running this same simulation 100000 times in R gives $E(X) \approx 20.13459$ droids.

There is a shortcut to calculate the theoretical expected value for a geometric random variable. $E(X) = \frac{1}{p}$. The proof involves a thorough understanding of calculus and geometric series and is left as a **challenge exercise**. In this example $E(X) = \frac{1}{0.05} = 20$ droids.

(e) Use simulation to estimate $Var(X)$

1. Here we want to simulate how each value of X deviates from the mean of 20 droids, and square each result.
2. We hold the same assumptions as the last problem.
3. We will also use the same number designation as the last problem to classify defective and functional droids.
4. We will now run several simulations.

Simulation 1: 72 1

Simulation 2: 74 26 59 91 27 34 62 44 26 18 16 50 24 67 29 15 28 79 78 40 92 5

Simulation 3: 6 38 9 15 72 11 61 45 60 99 9 74 62 91 35 42 70 88 89 42 11 63 9 8 73 39 83 15 44 80 11 61 55 19 6 8 19 34 4

Simulation Number	X	$X-\mu_X$	$(X-\mu_X)^2$
1	2	-18	324
2	22	2	4
3	39	19	1521

We now average the results for $(X-\mu_X)^2$ over all simulations. We get $Var(X) \approx 616.3333$. This is fairly far off from the theoretical variance. Our formula for the theoretical variance is $Var(X) = \frac{1-p}{p^2}$. For this example we have $Var(X) = 380$ droids^2.

(f) $Var(X) = 380$, calculate $SD(X)$

We simply take $SD(X) = \sqrt{Var(X)} = \sqrt{380} = 19.49359$ droids.

Geometric Random Variables Let X be a geometric random variable with parameter p. This can also be denoted $X \sim \text{Geometric}(p)$.

1. **Probability Mass Function:** $P(X=x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$

2. **Expected Value:** $E(X) = \dfrac{1}{p}$

3. **Variance:** $Var(X) = \dfrac{1-p}{p^2}$

Example 5: Suppose there is a 20% chance that the next new person you meet plays the piano.

(a) What is the probability that you don't meet someone who plays piano until the fifth person you meet?

Let's let X be the number of people you meet until you meet someone who plays the piano. X is a geometric random variable. $P(X=5) = (0.8)^4(0.2) = 0.08192$

(b) What is the probability that the eighth person you meet plays the piano?

Here we have $P(X=8) = (0.8)^7(0.2) = 0.04194304$

(c) How many people would you expect to meet until you find someone who plays the piano?

Here we'd like $E(X) = \frac{1}{0.20} = 5$ people. We would expect to have to meet 5 people before encountering someone who plays the piano.