Gov 2000 - 2. Random Variables and Probability Distributions

Matthew Blackwell
Harvard University
mblackwell@gov.harvard.edu

Where are we? Where are we going?

Eventually, we will be working with actual data—columns of numbers in a spreadsheet. We will want to use these numbers in the spreadsheet to learn about the world and specifically to learn about data that we didn't collect. For instance, we might have a column with a sample of respondents' partisan affiliation, measured on a 1-7 scale. But if we had drawn a different sample from the population of all citizens, we would have seen a different set of partisan affiliations. We will use probability to formalize this uncertainty, which will allow us to use the data we have (our sample) to learn about the distribution of all citizens (the populations).

List experiments for sensitive questions

- A list experiment is one where we ask respondents to tell us how many items on a list they agree with, where the contents of the list are randomized across respondents.
- For instance, perhaps we want to know what proportion of people would be upset by a black family moving in next door to them. What we would do is the following.

- Randomly split the survey into two halves. In the first half, ask respondents how many of the following upset you:

1. the federal government increasing the tax on gasoline;
2. professional athletes getting million-dollar salaries;
3. large corporations polluting the environment.

- The other half received the same prompt, except with an additional item:

1. the federal government increasing the tax on gasoline;
2. professional athletes getting million-dollar salaries;
3. large corporations polluting the environment;
4. a black family moving in next door.

- It turns out that we can use the answers to these questions to figure out what proportion of the population would be upset by a black family moving in next door. But in order to do that, we need to understand random variables.

WHAT ARE RANDOM VARIABLES?

A random variable is a numerical summary of an uncertain event. Imagine that we are calling five people at random and asking them if they support the president and they each either answer "Yes" or "No." Before making the calls, we don't know what the sequence answers will be—the outcome is uncertain. A random variable somehow summarizes these outcomes in a single number. For example, we might define a random variable X to be the number of these respondents that answer "Yes."

A random variable can be defined for any sample space:

Definition 1. A **random variable** (r.v.) is a function that maps from the sample space Ω into the real numbers.

Very simply, random variables are functions that map outcomes of the experiment to numbers. Sometimes this connection is obvious or trivial because the sample space is already a collection of numbers. Other times, we need to construct random variables. Why do we need to introduce these functions? Remember that we said that statistics is the mathematical study of data. In order to use the tools of math to tackle our questions of interest, we are going to need to work with numerical outputs. Working with the original sample space might be incredibly difficult and very application specific. But once we convert these sample spaces into random variables, we can see that very different problems might lead to random variables with very similar properties.

Examples of random variables

- **Coin Flipping** Imagine our experiment was tossing a coin 5 times. The sequence of outcomes of the flips, $\omega = HTHTT$ for example, is not a random variable because it isn't a number. But we could make it into a random variable, X if make it the number of heads in the 5 tosses. Notice how the random variable takes in outcomes from the sample space ($HTHTT$) and converts them into a number, 2.Each sample space can have many different random variables defined on it. For the coin toss, we could also define a variable to be the number of tails flips, Y. In this case, these two variables would be related by $X = 5 - Y$.
- **Voter Turnout** For just one person, the sample space is $\Omega = \{\text{voted,didn't vote}\}$. But again these outcomes can be results of a random variable because they are not numeric. We could define a random variable, X, that converts these outcomes into numbers (called a **Bernoulli** or **binary** random variable):

$$X = \begin{cases} 1 \text{ if voted} \\ 0 \text{ if didn't vote} \end{cases}$$

- **Government duration** Sometimes the sample space is already numeric so creating random variables is more obvious. What if our experiment is how long a government lasts in a parliamentary system? Obviously here the sample space is the set of nonnegative numbers $\Omega = [0, \infty)$. Then our random variable might just be equal to the outcome itself.

We almost always use capital roman letters for the "name" of the random variable such as X. Here that is just shorthand for the number of heads in 5 coin flips. Obviously when we need to do mathematical operations on the variable, its shorthand name X will be easier to use. We will refer to a particular value that X might take with lower case letters, x. So we might write $\mathbb{P}(X = x)$ to be the probability that the number of heads is equal to x. Note that a r.v. is a function, so to be precise, we need to write $X(\omega)$, but we often shorten this to simply X when the underlying sample space is either clear or not important for the discussion.

PROBABILITY DISTRIBUTIONS

It might seem confusing at first that we call these random variables since they deterministically map from the sample space to the real line. Where does the randomness come from? In a nutshell, **uncertainty over outcomes drives uncertainty over random variables**. The randomness in the example of X being whether the person voted or not comes from the randomness of that outcome, not in the mapping of "vote" into 1 and "didn't vote" into 0.

We'll use probability to formalize the uncertainty over what value X will take. That is, let $\mathbb{P}(\omega)$ be the probability of some event (such as $\mathbb{P}(H) = 0.5$). The probability of some value of $X = x$ is just the probability of the events that would lead to $X = x$:

$$\mathbb{P}_X(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$$

Let X be the number of heads in two coin flips of a fair coin. Then we can write out all of the possible outcomes (TT, HT, TH, HH), their probabilities, and the values that the X would take:

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

x	$\mathbb{P}_X(X = x)$
0	1/4
1	1/2
2	1/4

Remember that probabilities on the sample space come from a **data generating process** (DGP)—assumptions about the physical or social world. Assuming that we have independent coin flips induces independent probabilities of 0.5 for each coin flip. Random sampling from a set induces equal probabilities of each object. The DGP, then, will also induce the probability distribution for the random variable.

DISTRIBUTION FUNCTIONS

The **distribution** of a r.v. X describes what values of X are more likely than other values. Above we derived the distribution of simple r.v.s by directly investigating the underlying sample space. It is cumbersome to derive the probabilities of X each time we need them, so it is helpful to have a function that can give us the probability of values or sets of values of X. The most general of these functions is the cdf.

Definition 2. The **cumulative distribution function** or **cdf** of a r.v. X, denoted $F_X(x)$, is defined by:

$$F_X(x) \equiv \mathbb{P}_X(X \leq x).$$

The cdf tells us the probability of a r.v. being less than some given value. For instance, suppose that X was the age of a random selected person in the United States. Then, $F_X(18)$ is the probability that this person is 18 or younger, $\mathbb{P}_X(X \leq 18)$.

This function completely describes the distribution of a random variable. That is, if we have two r.v.s, X and Y and their cdfs are the same at every point, x, $F_X(x) = F_Y(x)$, then they have identical distributions.

Example: random assignment to treatment

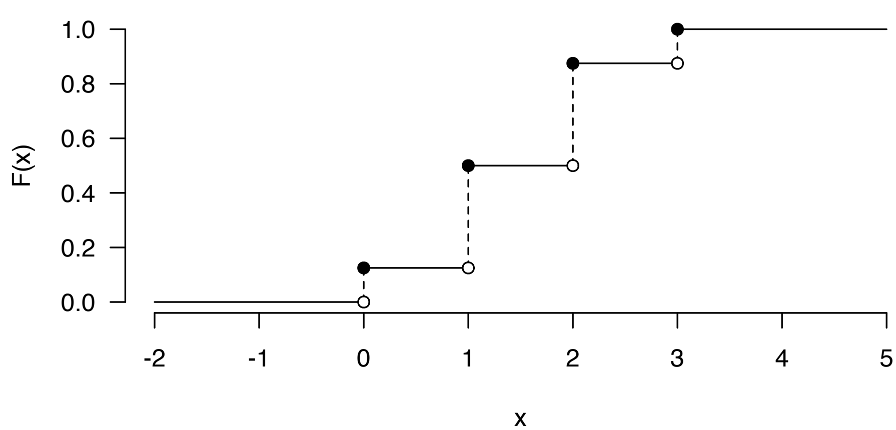
Suppose that we're running a randomized control trial to see if some intervention works—maybe we're random assigning some people to receive GOTV mailer or randomly assigning them to watch negative versus positive ads. Let's say that we did a poor job at recruiting subjects so we only have 3 subjects. Here's our procedure for randomly assignment: flip a coin for each unit independently and assign those with heads to Treatment and those with tails to Control. We'll define X to be the number of treated units:

$$X = \begin{cases} 0 & \text{if } (C, C, C) \\ 1 & \text{if } (T, C, C) \text{ or } (C, T, C) \text{ or } (C, C, T) \\ 2 & \text{if } (T, T, C) \text{ or } (C, T, T) \text{ or } (T, C, T) \\ 3 & \text{if } (T, T, T) \end{cases}$$

We can use the underlying probabilities of the coin flips to calculate the probability of each outcome. First note that $\mathbb{P}(C, T, C) = \mathbb{P}(C)\mathbb{P}(T)\mathbb{P}(C) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$, where the first equality holds by independence of the coin flips and the second by the fair coin assumption. Also, note that this is true for any of the outcomes. Thus, we can use this to determine the cdf for the number of treated units:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

We can plot the cdf of this r.v. as:



Properties of a cdf

The cdf has a few properties that are useful to write out:

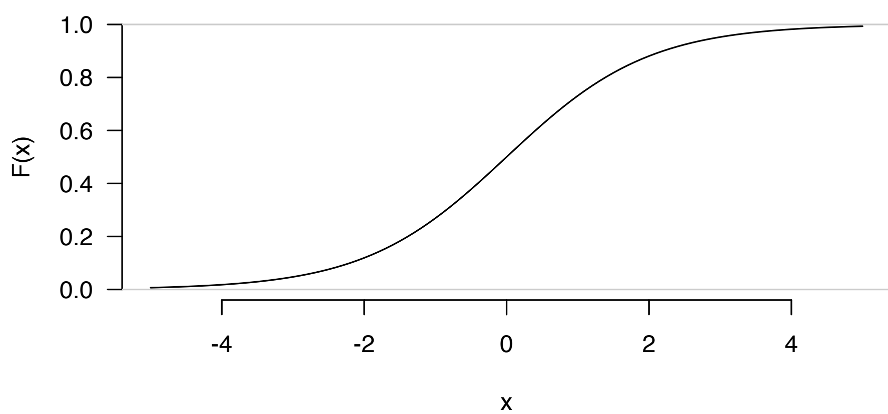
1. never decreases ($F_X(a) \geq F_X(b)$ if $a > b$)
2. limits to 0 toward negative infinity, limits to 1 toward positive infinity,
3. right-continuous (no jumps when we approach a point from the right)

Sometimes a cdf will be continuous function, which indicates that the probability of being less than x is not terribly different than the probability of being less than $x + \varepsilon$, where ε is some very small number. The cdf of the number of treated units is obviously not continuous since, for instance, the probability of X being less than 0.9999 ($F_X(0.999) = 1/8$) is very different than the probability of being less than 1.0001 ($F_X(1.001) = 1/2$).

Here's an example of a continuous cdf:

$$F_X(x) = \frac{1}{1 + e^{-x}}$$

Here is what this cdf looks like:



It is clear from the plot that $F_X(x)$ never decreases and that it is continuous (so it is also right-continuous).

Problem 1. Show that the cdf, $F_X(x) = (1 + e^{-x})^{-1}$ limits to 0 as $x \to -\infty$ and limits to 1 as $x \to \infty$.

Calculating probabilities from the cdf

Since the cdf complete determines the distribution of the r.v., it makes sense that we can use to calculate the probability of any value or set of values that X could take. Obviously, it can give us the probabilities of intervals like $(-\infty, b]$. From the complement rule of probability, we can see that $\mathbb{P}_x(X > x) = 1 - F_X(x)$, which gives us intervals like (a, ∞). Putting these together, we can get the probability of any range of values:

$$\mathbb{P}(a < X \leq b) = F(b) - F(a)$$

You can see that this works by using the following equality which we can prove using the properties of probabilities from last week (noting that the complement of $X \leq a$ is $X \geq a$):

$$\mathbb{P}(X \leq b) = \mathbb{P}(X \leq a) + \mathbb{P}(X \geq a \cap X \leq b)$$

DISCRETE RANDOM VARIABLES

Definition 3. A r.v. is **discrete** if its cdf is a step function, which implies that it only takes a finite or countably infinite number of values with positive probability.

Countably infinite just means that it takes on any integer and there's no (obvious) upper bound to the values that it can take. The most obvious discrete r.v. is the binary r.v., which can only take on two values: 0 and 1. Defining discrete r.v.s with the cdf is slightly more technically correct, but defining it in terms of the countability of its values is more intuitive.

Examples of discrete r.v.s

- Number of Democrats who win election in the Senate
- An indicator of whether two countries go to war
- The number of times a particular word is used in a document

Probability mass function

For a discrete r.v., each possible value of the r.v. has an associated probability of occurring. Go back to our example on voting. For a particular individual, they have some probability of voting $\mathbb{P}(X = 1)$ and some probability of not voting $\mathbb{P}(X = 0)$. But this can generalize as well: we can list out all possible values of the discrete r.v. and also their associated probabilities. This provides a nice summary of the entire distribution of the variable.

Definition 4. The **probability mass function** for a discrete random variable, X, is given by

$$f_X(x_j) = \mathbb{P}(X = x) \qquad \text{for all } x$$

Some properties of the pmf fall out of the properties of probability: $0 \leq f_X(x) \leq 1$ and $\sum_{i=1}^{k} f_X(x_j) = 1$. Given this, we can write the cdf of a discrete r.v. as:

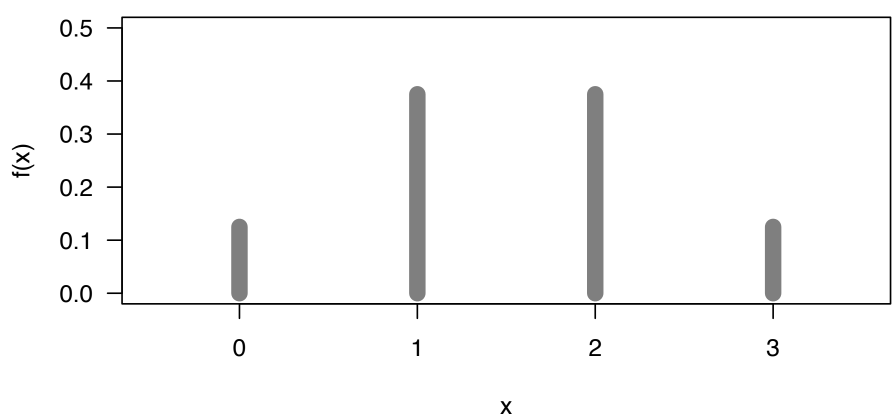
$$F_X(x) = \sum_{x_k \leq x} f_X(x_k),$$

which is just the sum of the pmf for all values of X less than x.

Clearly, the number of treated units r.v. above is discrete. In fact, we can easily compute its pmf:

$$\begin{aligned}
f_X(0) &= \mathbb{P}(X = 0) = \mathbb{P}(C, C, C) = \frac{1}{8} \\
f_X(1) &= \mathbb{P}(X = 1) = \mathbb{P}(T, C, C) + \mathbb{P}(C, T, C) + \mathbb{P}(C, C, T) = \frac{3}{8} \\
f_X(2) &= \mathbb{P}(X = 2) = \mathbb{P}(T, T, C) + \mathbb{P}(C, T, T) + \mathbb{P}(T, C, T) = \frac{3}{8} \\
f_X(3) &= \mathbb{P}(X = 3) = \mathbb{P}(T, T, T) = \frac{1}{8}
\end{aligned}$$

- We could plot this pmf using R:



CONTINUOUS RANDOM VARIABLES

Definition 5. A r.v. is **continuous** if its cdf is continuous, which implies that it can take on every value in some interval of the real line.

Continuous variables might take any value between $-\infty$ to ∞ or they might be positive only or they might be in some interval like $[0, 1]$. The important part is that they contain all real numbers in that interval. If so, then there are an uncountably infinite number of possible realizations. Note that the variables are only approximately continuous—that is, they have a very large number of possible realizations and treating it as continuous is a good approximation.

Examples of continuous random variables

- The length of time between two governments in a parliamentary system
- The proportion of voters who turned out
- Budgets allocations to various government programs

Probability density function

With a continuous r.v., we want to do something similar—describe how likely some set of outcomes are. We might think to take the same approach as with a discrete r.v. and just go through each possible value of X and list out its corresponding probability. This approach breaks down, though, when the number of possible values is uncountable because the number of possible realizations is massive (there is an infinite number of them in any subset of the real line). This means that we have to take the probability of any particular realization (for example, $2.32879873\ldots$) as 0 and instead we will work with the probability of X being in some set B.

Definition 6. The **probability density function** or pdf, $f_X(x)$, for a continuous random variable X is the function that satisfies:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

The pdf of a continuous r.v. can be used to get the probability of any interval on the real line, $B \subset \mathbb{R}$, by simply looking at the area under the pdf for that region:

$$\mathbb{P}(X \in B) = \int_B f_X(x)dx.$$

In particular, we have the following, when $a \leq b$, then we can find the probability of X being between a and b as:

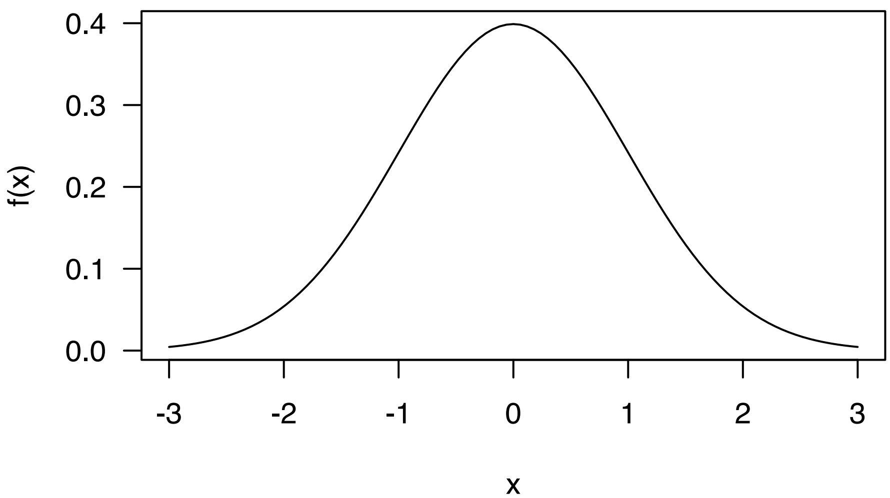
$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

Based on this definition, every pdf will meet two conditions. First, the pdf will be nonnegative, so that $f_X(x) \geq 0$ for all x. Second, the total density must be equal to 1. That is,

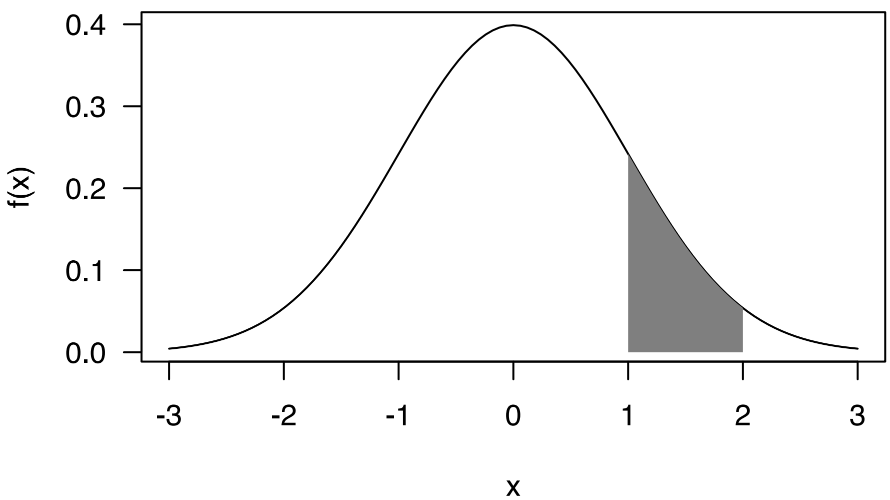
$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

.

The pdf gives us information about how likely various outcomes are. Regions with higher values of the pdf are areas where we are more likely to see a realization of X.



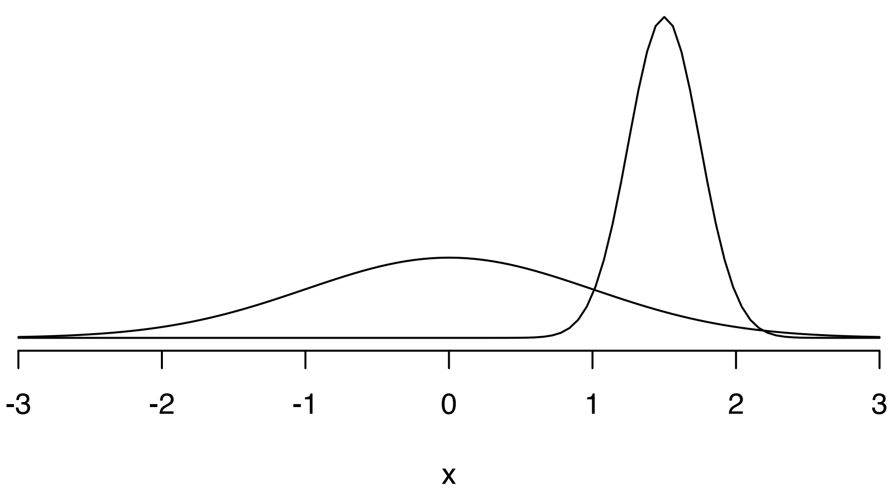
But be careful! The height of the curve here, $f_X(x)$ is not equal to the probability of x occurring—remember that is 0 for a continuous variable. To get the probability that X will fall in some region, we need to take the integral, which corresponds to the area under the pdf curve:



This discussion highlights an important misconception: the pdf, unlike the pmf, might have values greater than or equal to 1. Intuitively, this is because the height of curve is not a probability.

PROPERTIES OF DISTRIBUTIONS

The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution. Thus, it is important to think about some properties of distributions that help summarize them. For instance, if we look at these two distributions, what would we say is the difference between them:



As we've seen, distributions can be these complicated mathematical expressions that are hard to interpret. These two distribution have many differences: one is the probability that each places around -1, another is how much they place around 2, and so forth.

It would be nice, however, to be able to summarize these distributions quickly so that we can have an intuitive understanding of where most of the data will be. To do this, we can think of two features of distribution that are easily quantifiable and informative: the **center** of the distribution and the **spread** of that distribution around its center.

MEASURES OF CENTRAL TENDENCY - EXPECTED VALUE

The central tendency of the distribution is a measure of the "middle" of the distribution. There are a couple of ways we might think about where the middle is. These measures are one-number summaries of the distribution in the sense that they represent our best guess of the value of X before we see it. The measure of central tendency we will focus on in this class is the **expected value**, which is also called the **expectation** or the **mean** of the distribution. We refer to expected value of X as $\mathbb{E}[X]$ or μ.

Motivation - calculating averages

Imagine we had a bunch of numbers an you wanted to calculate the average: (1,1,3,4,4,5). Obviously, you would add them up and divide by the number of items in the sum:

$$\frac{1+1+3+4+4+5}{6} = \frac{18}{6}3$$

Now, you could always have calculated that a slightly different way:

$$\frac{1}{6} \times 1 + \frac{1}{6} \times 1 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5$$

and if we group terms we get:

$$\frac{2}{6} \times 1 + \frac{1}{6} \times 3 + \frac{2}{6} \times 4 + \frac{1}{6} \times 5$$

This last expression is another way to calculate the mean: sum up the values in the set, weighted by their proportion in the set. This form is the exact way that we'll think of the mean.

Definition

As with the distribution, we calculate the expected value differently for discrete and continuous random variables. For both of them, the expected value is a **weighted average** of the realizations weighted by the probability of occurring.

Definition 7. The expected value of a bounded discrete r.v. X is:

$$\mathbb{E}[X] = \sum_{j=1}^{k} x_j f_X(x_j).$$

The expected value of a bounded continuous r.v. X is:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

Sometimes we will calculate the expectation from the distribution directly using this definition (like in the next example). Other times, we'll use some known/given expectation and then think about how transformation of the r.v. would give different expectations. Finally, sometimes we will be able to determine the expectation if we know that the r.v. comes from one of the famous distribution below.

Example - number of treated units

Let's go back to the number of treated units to figure how many units we should expect to be treated in our experiment:

$$\begin{aligned} E[X] = \sum_{j=1}^{k} x_j f(x_j) &= 0 \times f_X(0) + 1 \times f_X(1) + 2 \times f_X(2) + 3 \times f_X(3) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times 18 \\ &= 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = 1.5 \end{aligned}$$

If we look back at the pmf of this distribution, it makes a lot of sense that the answer would 1.5 since that is in the middle of the distribution. This answer brings up an interesting feature of the expected value: it doesn't have to be one of the values that the r.v. can take.

Properties of the expected value

The expected value has a lot of nice properties that make it easy to work with. Both of the key properties of expected values are that they are linear. What does that mean?

- **Additivity:** (expectation of sums are sums of expectations)

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

- **Homoegeneity:** Suppose that a and c are constants. Then,

$$\mathbb{E}[aX + c] = a\mathbb{E}[X] + c$$

- **Law of the Unconscious Statistician**, or LOTUS. If $g(X)$ is a function of a discrete random variable, then

$$E[g(X)] = \sum_x g(x) f_X(x),$$

  which basically says that the expected value of the transformation of the random variable is just the weighted average of the transformed outcomes.

Example: list experiments

- Let's say that Y is the number of items that people say upset them with the additional "black family" item and X be the number of items that upset them with just the 3 baseline items. Then, we could write $Y = X + A$, where $A = 1$ if the black neighbors question upset them and $A = 0$ it did not.
- Then, we know that $E[Y] - E[X] = E[A]$, but can you prove that?
- If A is a Bernoulli r.v., then how can we interpret $E[A]$?

MEASURES OF SPREAD

Now we have some sense of where the middle of the distribution is, but we also want to know how spread out the distribution is around that middle. We'll talk about two of these that are closely related: the variance and the standard deviation.

Definition 8. The **variance** is the average of the squared distances from the mean. We sometimes denote it σ_X^2

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

We sometimes denote it $\sigma_X^2 = \mathbb{V}[X]$. Since the squared distances are always nonnegative, the variance is also always nonnegative. If most of the observations are close to the expected value, then the variance will be closer to 0. If the observations are far from the expected value, then the variance will be higher.

We can use LOTUS from above to calculate the variance for a discrete random variable:

$$\mathbb{V}[X] = \sum_x (x - \mathbb{E}[X])^2 f_X(x)$$

And we can apply the same principle for continuous random variables:

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx$$

Definition 9. The **standard deviation** is just the (positive) square root of the variance: $\sigma_X = \sqrt{\text{Var}[X]}$.

You can interpret this as the average distance from the expected value of the distribution. What is nice about this is that it is in the same units as the original variable, whereas the variance is in squared units. For instance, if X is age, then, σ_X is also in years, whereas σ_X^2 is in years-squared.

Example - number of treated units

Let's go back to the number of treated units to figure out the variance of the number of treated units:

$$\begin{aligned}
\mathbb{V}[X] &= \sum_{j=1}^{k} (x_j - \mathbb{E}[X])^2 f(x_j) \\
&= (0 - 1.5)^2 \times f_X(0) + (1 - 1.5)^2 \times f_X(1) + (2 - 1.5)^2 \times f_X(2) + (3 - 1.5)^2 \times f_X(3) \\
&= (-1.5)^2 \times \frac{1}{8} + (-0.5)^2 \times \frac{3}{8} + 0.5^2 \times \frac{3}{8} + 1.5^2 \times 18 \\
&= 2.25 \times \frac{1}{8} + 0.25 \times \frac{3}{8} + 0.25 \times \frac{3}{8} + 2.25 \times 18 = 0.75
\end{aligned}$$

Exercise: What's the standard deviation for this distribution?

Properties of variances

If a and be are constants, then we have the following properties:

- $\mathbb{V}[b] = 0$
- $\mathbb{V}[aX + b] = a^2 \mathbb{V}[X]$

Variances have slightly different properties than expectations, but there are similar flavors. First, note that the variance of a constant, b, is 0: $\mathbb{V}[b] = 0$. You should use the definition of the variance to convince yourself why that is the case (hint: what's the expected value of a constant?).

FAMOUS DISTRIBUTIONS

We will use distributions to model some population from which our data is a sample. That is, we will assume that there is some population-level distribution of ages or support for the president and that when we see a specific respondent's age or support for the president, we are seeing a single draw from this distribution.

But it's cumbersome to always write out the underlying sample space and then derive the pdf/pmf from it. More often, we will focus on certain **families of distributions** that have a common form of the pdf/pmf up to some **parameters**, which vary across the family. Each family has a story for what processes generate a r.v. from that family. Sometimes, we can match our specific substantive example to these stories to justify modeling a certain r.v. from a specific family.

Bernoulli

Let X be a binary variable with $\mathbb{P}(X = 1) = p$ and, thus, $\mathbb{P}(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a **Bernoulli distribution** with the following pmf:

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}.$$

There are infinite number of Bernoulli distributions, each with a different value p. This collection of distributions is called the family and p is the parameter that varies across the family.

Exercise: Let X be a Bernoulli r.v. with $\mathbb{P}(X = 1) = p$. Use the definition of the expected value to calculate $\mathbb{E}[X]$.

Binomial

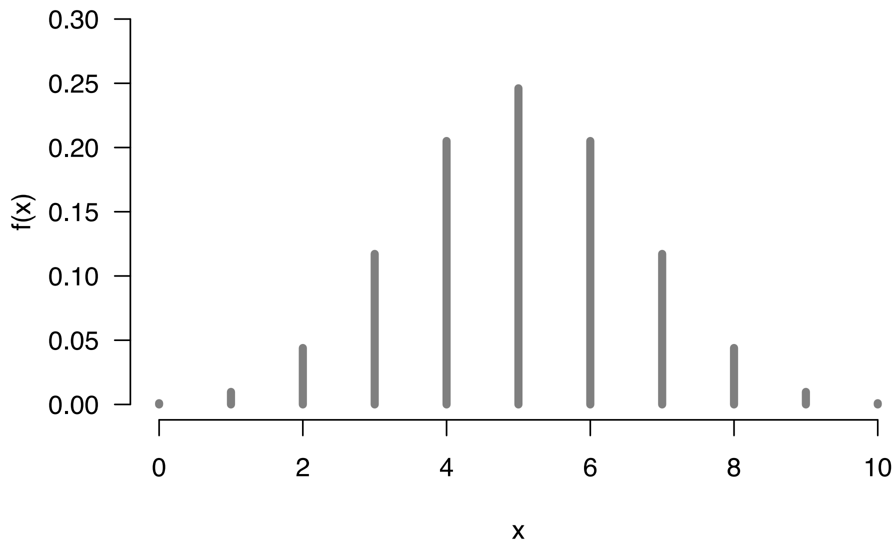
Let X be the number of heads in n independent coin flips with probability p of heads. Then X has a **binomial distribution** written $X \sim \text{Bin}(n, p)$ which has p.m.f.:

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where $\binom{n}{k} = n!/(k!(n-k)!)$ A Binomial r.v., X is equivalent to the sum of n Bernoulli r.v.s each with probability p. That is, suppose that $Z_1, \ldots, Z_n$ are independent (we'll define this formally next week) Bernoulli r.v.s with probability p. Then, you can write

$$X = Z_1 + \ldots + Z_n.$$

The expectation of a Binomial r.v. is $\mathbb{E}[X] = np$ and the variance is $\mathbb{V}[X] = np(1-p)$. One example of a Binomial r.v. with $n = 3$ and $p = 0.5$ number of treated units in the RCT example.



Discrete uniform

- Probably the most famous distribution for a discrete r.v. is the **discrete uniform distribution** that puts equal probability on each value that X can take:

$$f_X(x) = \begin{cases} 1/k & \text{for } x = 1, \dots, k \\ 0 & \text{otherwise} \end{cases}$$

- Note that we can summarize these distributions with one number—with the discrete distribution it's the number of possible outcomes and with the Bernoulli distribution it is probability of variable being 1.

Continuous uniform

A simple example of a continuous distribution is the **continuous uniform distribution** on the $(0, 1)$ interval is the distribution where the probability of an interval is equal to one over its length. We write $X \sim \text{Unif}(0, 1)$ and it has the pdf:

$$f_X(x) = \begin{cases} 1 & \text{for } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

More generally, a r.v. might be uniform over any interval, $[a, b]$, which has the pdf:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Normal distribution

The **normal distribution** is the classic "bell-shaped" curve. It is extremely useful and ubiquitous in statistics. If X has a normal distribution, we write $X \sim N(\mu, \sigma^2)$, where μ is the expected value of the distribution and σ^2 is the variance. The pdf for the Normal distribution is:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

When the mean is 0 and the variance is 1, we call this the **standard normal distribution**. The reason this distribution comes up so much is that many things follow an approximately Normal distribution.