Chapter 3

Random Variables

3.1 The general construction

Informally, a random variable X is just a random number – a number whose value is randomly determined. In this class we'll usually denote a random variable with an uppercase Roman letter toward the end of the alphabet: U, V, X, Y, Z, etc. So how did this number become random in the first place? It became random because its value was computed based on the outcome of some underlying random phenomenon.

Consider rolling a pair of fair, four-sided dice. The two faces that you roll are random. The sample space of this outcome is a set of pairs:

$$S = \{s = (x, y) : x = 1, 2, 3, 4; y = 1, 2, 3, 4\}.$$

Now imagine adding up the two faces you rolled. That is, given an outcome $s = (x, y)$ in S, we calculate $X(s) = x + y$. This is a random variable. It's a number whose value is random because it is computed based on the outcome of some random phenomenon. In Table 3.1, I have enumerated the sample space S of the die roll and included the value of X that you get for each of the different outcomes.

Because the outcome of the die roll is random, the value of X is random, and so we might want to answer question about the probability of X achieving certain values. For example, what is the probability that X ends up being equal to 5? Examining the table, there are four ways that can happen: rolling (4, 1), (3, 2), (2, 3), or (1, 4). So the event "X is equal to 5" can be rewritten as "You roll (4, 1), (3, 2), (2, 3), or (1, 4)." This is a disjoint union of equally likely outcomes (remember

	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Table 3.1: Sums of two four-sided die rolls.

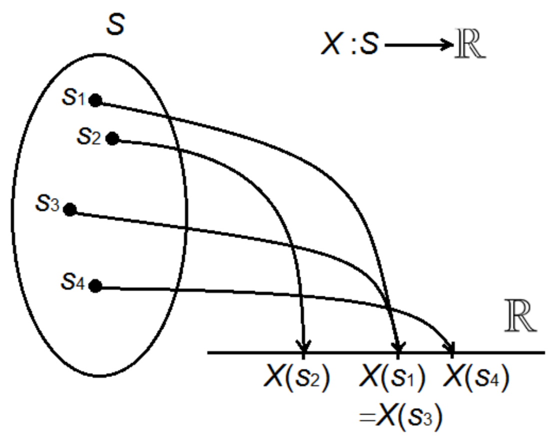


Figure 3.1

the dice are fair), so we know that

$$\begin{aligned} P(X = 5) &= P(\text{rolling } 4, 1) + P(\text{rolling } 3, 2) + P(\text{rolling } 2, 3) + P(\text{rolling } 1, 4) \\ &= 4(1/16) \\ &= 1/4. \end{aligned}$$

What about the probability that X ends up being even? We see that this is the same as saying "X is 2 or 4 or 6 or 8," and again, we're in a world with a finite sample space and equally likely outcomes, so we can get an answer by counting:

$$P(X \text{ even}) = P(X \in \{2, 4, 6, 8\}) = \frac{8}{16} = 1/2.$$

So we see that the probabilities governing the behavior of X are directly inherited from the probabilities governing the behavior of the underlying phenomenon – in this case the outcome of the die roll. Now let's try to make these ideas a bit more precise.

Definition 3.1. A random variable is a function $X : S \to \mathbb{R}$ from a probability space $(S,\ B \subseteq S,\ P_0)$ to the real numbers. So the function takes in an outcome $s \in S$ and returns the value of the variable $X(s) \in \mathbb{R}$.

Let's unpack this a little bit:

- (**base probability space**) The base probability space $(S,\ B \subseteq S,\ P_0)$ represents the underlying random phenomenon that is determining the value of X. In our example, this was the die roll;

- (**random variable**) The random variable itself is a completely deterministic function that takes in outcomes associated with the underlying phenomenon and returns real numbers (Figure 3.1). So X is random because the outcome it receives is random, not because of anything special its doing. Recall in the die roll example that X was just the addition function;

- (**pushforward probability space**) the action of the function X induces a new probability space on the real numbers called the **pushforward probability space**: $(\mathbb{R},\ A \subseteq \mathbb{R},\ P)$. The new probability measure P is called the **probability distribution** of X, and it is defined in terms of the base probability measure P_0 in the following way:

$$P(X \in A) = P_0\left(X^{-1}(A)\right) = P_0\left(\{s \in S : X(s) \in A\}\right). \tag{3.1}$$

  So probabilities in the new space are computed by pulling the set A back to the original space (via the inverse image of X) and applying the original probability measure P_0.

Okay, whatever nerd. The bottom line is this. X is this quantity that is inheriting all of its randomness from the underlying phenomenon that we represent with the base space. The action of X then "pushes this probability forward" to the real numbers. Once you're in the real numbers, you can ignore most of this mishegoss and just work with the random variable itself and with probabilities on $\mathbb{R}$. So at the end of the day, in order to fully specify a random variable, you need to write down two things:

- (**range**) $\text{Range}(X) \subseteq \mathbb{R}$ is the set of all possible values that X can take on. Sometimes to save space I might also call this set fancy, scipt-y $\mathcal{X}$;
- (**distribution**) the probability distribution of X is the probability measure on $\mathbb{R}$ that answers the question "what is the probability that X takes values in the set A?" So the distribution is a function that takes in events $A \subseteq \mathbb{R}$ and spits out probabilities $P(X \in A)$.

3.1.1 Why do we end up "ignoring" most of this?

> "At a certain point in most probability courses, the sample space is rarely mentioned anymore and we work directly with random variables. But you should keep in mind that the sample space is really there, lurking in the background." - Larry Wasserman, *All of Statistics*

Well, first off, we would most definitely *not* ignore it if we were doing proper, mathematically rigorous probability theory. But I grant that we're not quite there, so why do we ultimately get to eschew this formalism and jump right to "range and distribution"?

Consider the following random quantity:

$$X = \text{"the price of Google stock at 2:30pm ET on 2/19/2025."}$$

For all intents and purposes, this is random. It cannot be perfectly predicted. How is its value ultimately determined? Well, for starters, we don't even really know. But we guess that the value of X has something to do with supply and demand in a more-or-less competitive market, so you have millions of buyers across the globe bidding for Google stock, and millions of sellers trying to offload Google stock, and they are making these decisions unilaterally using various kinds of information and strategies, and the bids and asks somehow equilibrate to determine the stock price. But while all of that is going on, the market as a whole is buffeted by random shocks related to severe weather events, global politics, the personal lives of the Google executives, co-movement with the value of other stocks, possible technical glitches at the New York Stock Exchange, and on and on ad infinitum. So the “underlying random phenomenon" that is ultimately driving the value of X is nothing less than the sum total disparate churning across all human civilization.

Writing down the probability space $(S,\ B \subseteq S,\ P_0)$ that captures all of this is nigh impossible, and even if you could, how exactly do you write down the actual functional form of X? The bottom line is, you can’t, and you don’t try. But you still want to study X, so you just skip to the end and say something like “X is a nonnegative continuous random variable with the log-normal distribution." Good enough. Now go lose money.

3.2 Discrete random variables

As we said, in order to fully specify a random variable, we must write down its range and its distribution. So how do you actually write those things down? There are many ways, but in this course we will focus on two very important special cases. Here's the first:

Definition 3.2. A random variable X is **discrete** if its range is a finite or countably infinite set:

$$\text{Range}(X) = \{x_1, x_2, x_3, ...\} \subseteq \mathbb{R}.$$

In this generic setting we will always adopt the convention that $x_1 < x_2 < x_3 < ...$

Because the range of a discrete random variable is countable, to fully specify its distribution, it suffices to list out the individual probabilities of the individual values in the range:

value	probability
x_1	$P(X = x_1)$
x_2	$P(X = x_2)$
$\vdots$	$\vdots$
x_i	$P(X = x_i)$
$\vdots$	$\vdots$

For many of the random variables that we will study, there will be a handy formula for computing these probabilities, and so we can encode this schedule of individual probabilities in a function:

Definition 3.3. The **probability mass function** (**pmf**) of a discrete random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) = \begin{cases} P(X = x) & x \in \text{Range}(X) \\ 0 & \text{else.} \end{cases} \tag{3.2}$$

Naturally, we require that $\sum_{i=1}^{\infty} p_X(x_i) = 1$. If context is clear, we can dispense with the subscript-X and just denote the pmf with $p(x)$.

So the pmf is the function that takes in values in Range(X) and spits out their individual probabilities. Figure 3.2 displays a cartoon of what the pmf of a discrete random variable might look like when you plot it. With this picture, we have a better sense of where this word "distribution" comes from. We have total measure 1 to allocate among the values in the range of X. When you plot the pmf, you see how the probability mass is "distributed" between them.

Having access to the individual probabilities of the individual values in Range(X) is nice, but how does that generalize to an overall rule for computing $P(X \in A)$ for any set $A \subseteq \mathbb{R}$? As a consequence of countable additivity, you can show that the general rule is

$$P(X \in A) = \sum_{x \in A \cap \text{Range}(X)} p_X(x). \tag{3.3}$$

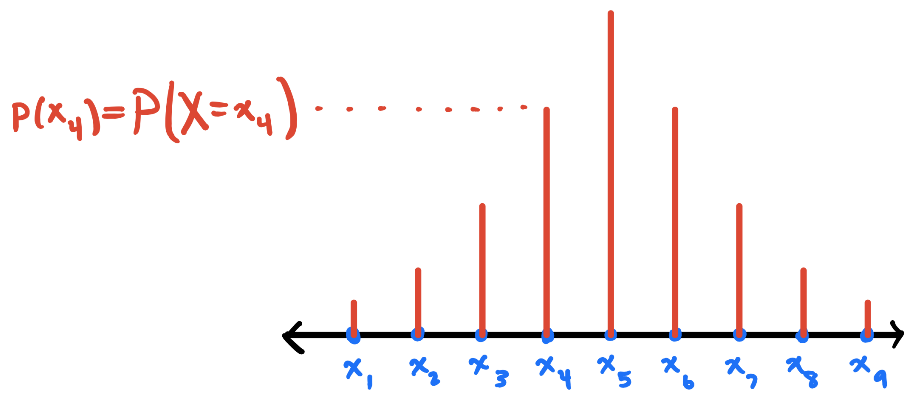


Figure 3.2

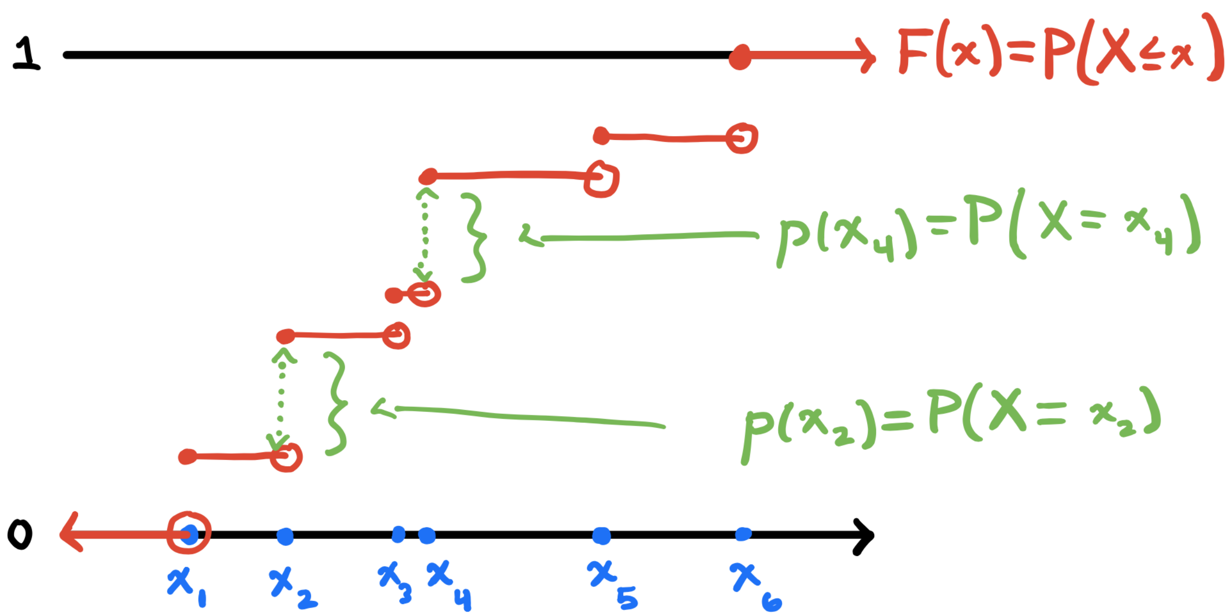


Figure 3.3

So the main property of the pmf that you should internalize is that it is a function that you *sum over* in order to compute the distribution of X.

The pmf is this handy-dandy object that we can write down as a shorthand for specifying the entire distribution of a random variable X. But it's not the only object that would get that job done. Here is an alternative object that you could write down that would also fully characterize the distribution of X:

Definition 3.4. The **cumulative distribution function** (**cdf**) of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = P(X \in (-\infty, x]) = P(X \leq x)$. Once again, you can dispense with the subscript-X if the context is clear.

Figure 3.3 is a cartoon of what the cdf might look like for a discrete random variable. It has several properties that you should note:

- (**non-decreasing**) it may plateau in places, but it never goes down;
- (**right continuous**) it has all of its limits if you approach from the right. This is a consequence of the "$\leq$" in the definition of the cdf;
- (**"starts" at 0; "ends" at 1**) This function has $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. In Figure 3.3, 0 and 1 are eventually achieved as values of the cdf, but this is not necessary.

These properties are universal properties shared by any cdf you will ever see for the rest of your life. For discrete random variables specifically, we observe the additional property that the cdf is piecewise constant (a step function).

The cdf can be expressed in terms of the pmf by adding:

$$F(x) = P(X \in (-\infty, x]) = \sum_{x_i \leq x} p(x_i).$$

The pmf can be expressed in terms of the cdf by subtracting:

$$p(x_i) = P(X = x_i) = F(x_i) - F(x_{i-1}).$$

We will elaborate on the derivation of this later, but the above identity just makes math of the observation that the size of the jumps along the cdf are equal to values of the pmf. With these two expressions, we see that the pmf and cdf are basically two sides of the same coin. They contain fundamentally the same information, and given one, you can always recover the other. Depending on the task at hand, you may prefer to work with one or the other, but writing down either amounts to a complete characterization of the distribution of X.

Once you have written down the range and the distribution of a random variable, you have fully characterized it. You have completely described the probability behavior of the random quantity that you sought to model. That said, the full distribution of a random variable is difficult to "use" in its entirety. So we are interested in extracting from the distribution some simple summaries that describe the "typical" or "average" behavior of the random variable. Here is perhaps the most important one:

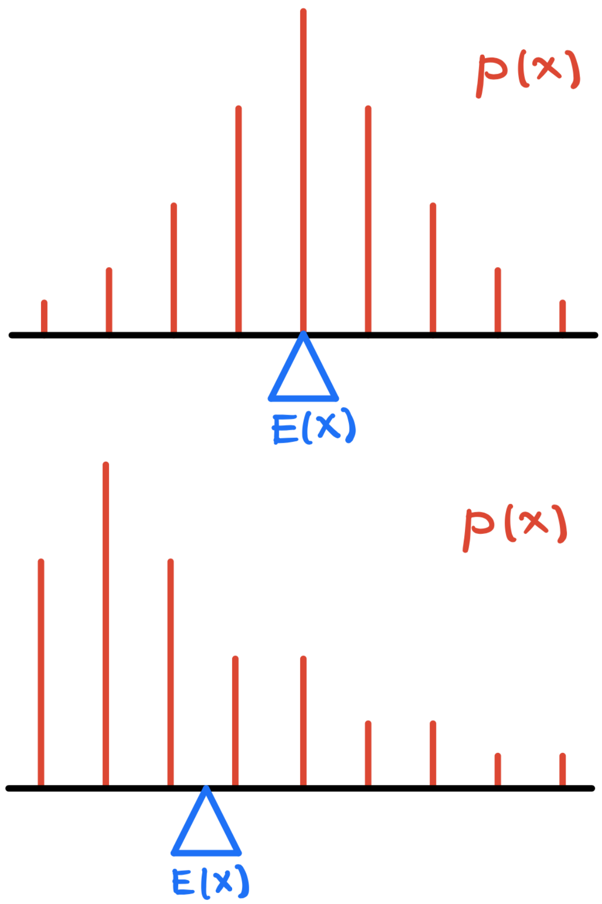


Figure 3.4

Definition 3.5. The **expected value** of a discrete random variable X is the weighted average of the values in its range weighted by their probabilities:

$$E(X) = \sum_{x \in \text{Range}(X)} xP(X = x). \tag{3.4}$$

We also refer to $E(X)$ as the **expectation** or the **mean**.

Informally, the expected value of a random variable is the answer to the question "what is the typical or average value of this random quantity." Because it is random, we cannot perfectly predict exactly what value X will assume in a given instance, but "on average," what will it be?

In addition to this interpretation, the expected value also has a nice visual interpretation in terms of the pmf. It is the center of mass or balance point of the pmf, as seen in Figure 3.4. Notice that in the case of a symmetric pmf, $E(X)$ coincides with the center of the distribution and the peak. In the asymmetric case, you no longer have that in general. Note also that our formula for the expected value is an infinite sum in general, and not all infinite sums converge, so $E(X)$ may not always exist or be finite, and we will see examples of this in Section 3.3.1.

k	$P(X = k)$
2	1/16
3	2/16
4	3/16
5	4/16
6	3/16
7	2/16
8	1/16

Table 3.2: pmf of the sum of two fair, four-sided die.

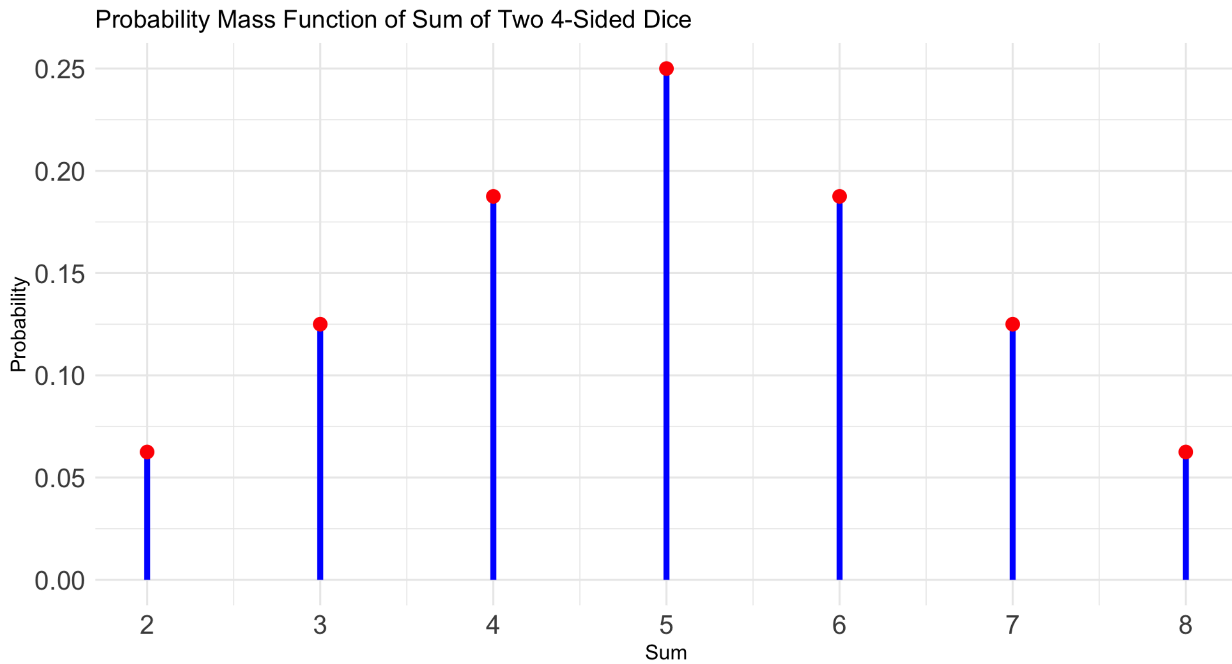
Example 3.1. Recall the example from Section 3.1 where X represents the sum of two fair, four-sided die rolls. From Table 3.1, we see that $\text{Range}(X) = \{2, 3, 4, 5, 6, 7, 8\}$, so this is a discrete random variable with a finite range, and we can fully specify the distribution by simply listing out all of the individual probabilities, as in Table 3.2. Alternatively, we could summarize the contents of this table with the function

$$p_X(x) = \frac{\min(x-1, 9-x)}{16}, \quad x = 2, 3, ..., 7, 8.$$

This is obviously not necessary in this case because the range is small and we can list everything out compactly, but for large or infinite ranges, writing down such a function is really the only way to specify the distribution. Figure 3.5 displays the pmf and cdf of X, and the expected value is

$$E(X) = \sum_{k=2}^{8} kP(X = k) = 2\frac{1}{16} + 3\frac{2}{16} + \cdots + 7\frac{2}{16} + 8\frac{1}{16} = 5.$$

From Figure 3.5, we could see that this distribution is symmetric, and so it is not surprising that $E(X)$ corresponds to the location of the peak.


Probability Mass Function of Sum of Two 4-Sided Dice
Probability
0.25
0.20
0.15
0.10
0.05
0.00
2
3
4
5
6
7
8
Sum

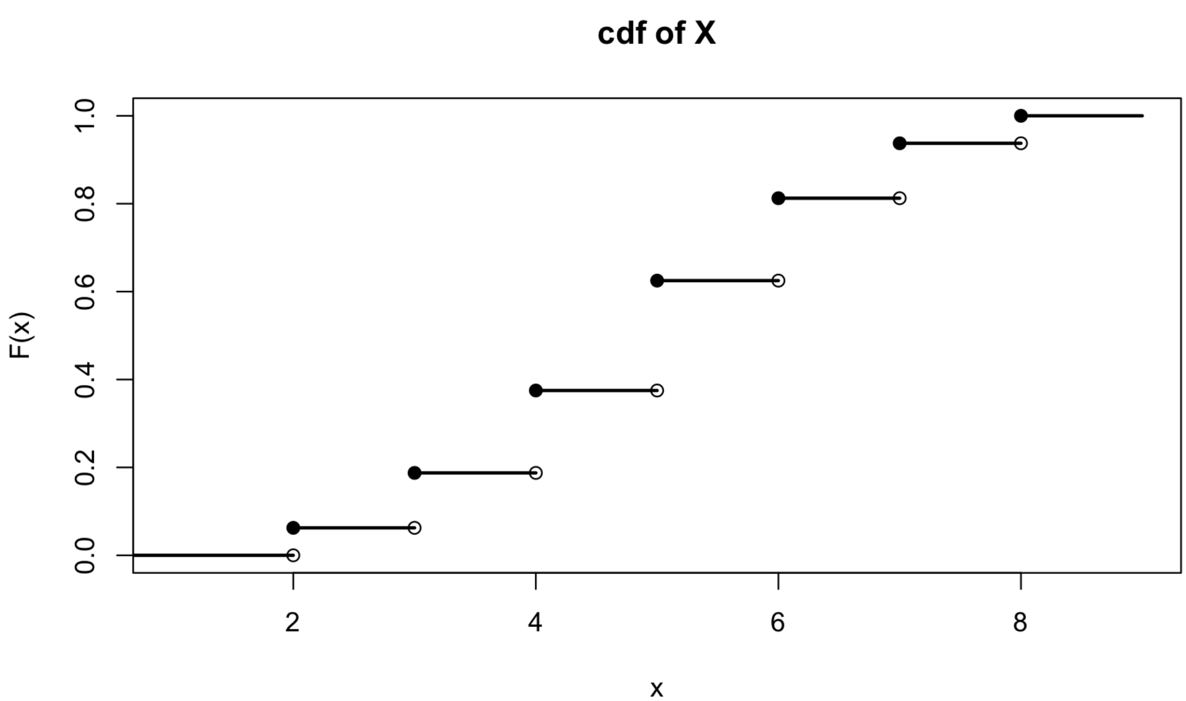

cdf of X
F(x)
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
x

Figure 3.5