Random Variables

This chapter discusses random variables.

What are Random Variables?

In probability theory, we use *random variables* to represent the outcomes of random processes. Don't confuse them with variables you define in your code or the variables you read about in algebra class. These variables are traditionally denoted by upper case letters. We call them random because they represent the outcomes of random processes such as flipping a coin, throwing dice or the possibility of rain on a given day. As an example, we define two random variables below:

$$X = Number\ of\ heads\ in\ 3\ flips\ of\ a\ coin$$

$$Y = Faceup\ value\ of\ a\ dice\ greater\ than\ 2\ on\ a\ roll\ of\ dice$$

Note that in both the above cases X and Y can take on different values depending on the outcome of the experiment.

What are Discrete Random Variables

The two random variables we discussed are examples of *discrete* random variables. Discrete means *distinct* or *separate* values. We call the discussed variables discrete because these variables can take on a *countable* number of values. Sometimes the list of countable values can be infinite, but the random variables will be discrete for as long as you can count them (even if that means counting forever).

The other type of random variables are continuous ones. For instance, the weight of your friend can be between 10lbs and 500lbs. However, if you pick any two values in that range, there is an infinite number of values between the two picked values. If you pick 120lbs and 121lbs, there is an infinite number of values between those two numbers that a random variable can take on e.g. 120.01, 120.001, 12.0001 so and so forth. The litmus test to determine if a random variable is continuous is to pick two values and verify

if an infinite number of values exist between them. If you can't list or count all the possible values then you are likely dealing with a continuous random variable. In the case of a discrete random variable, you can count a finite number of values between your two chosen values.

In the study of algorithms, we will only concern ourselves with discrete random variables and their distributions, but its important to understand the distinction between the two types.

What are Indicator Random Variables

The indicator variable for an event A is a variable having value 1 if the A happens, and 0 otherwise. The number of times something happens can be written as a sum of indicator variables. For example, we may define I_6 as an indicator variable when 6 appears face-up on a roll of a fair die. If we roll the die three times, we can express the number of times 6 appears on the die as:

$$I = I_{6_1} + I_{6_2} + I_{6_3}$$

We could have also declared a random variable representing the number of times 6 appears in three rolls of a die. However, sometimes it is easier to reason about problems using indicator random variables than complicated random variables.