

 Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Probability: Discrete Random Variables > Random Variables > Random Variables Defined

 Bookmark this page

Random Variables Defined

Learning Objective: Distinguish between discrete and continuous random variables

In the previous sections we've learned principles and tools that help us find probabilities of events in general. Now that we've become proficient at doing that, we'll talk about random variables. Just like any other variable, random variables can take on multiple values. What differentiates random variables from other variables is that the values for these variables are determined by a random trial, random sample, or simulation. The probabilities for the values can be determined by theoretical or observational means. Such probabilities play a vital role in the theory behind statistical inference, our ultimate goal in this course.

Introduction

We first discussed variables in the Exploratory Data Analysis portion of the course. A variable is a characteristic of an individual. We also made an important distinction between **categorical variables**, whose values are groups or categories (and an individual can be placed into one of them), and **quantitative variables**, which have numerical values for which arithmetic operations make sense. In the previous two modules, we focused mostly on events which arise when there is a categorical variable in the background: blood type, pierced ears (yes/no), gender, on time delivery (yes/no), side effect (yes/no), etc. Now we will begin to consider quantitative variables that arise when a random experiment is performed. We will need to define this new type of variable.

random variable

(definition) A **random variable** assigns a unique numerical value to the outcome of a random experiment.

A random variable can be thought of as a function that associates exactly one of the possible numerical outcomes to each trial of a random experiment. However, that number can be the same for many of the trials.

Before we go any further, here are some simple examples:

Example: Theoretical

Consider the random experiment of flipping a coin twice. The sample space of possible outcomes is $S = \{HH, HT, TH, TT\}$.

Now, let's **define the variable X to be the number of tails** that the random experiment will produce.

- If the outcome is HH, we have no tails, so the value for X is 0.
- If the outcome is HT, we got one tail, so the value for X is 1.
- If the outcome is TH, we again got one tail, so the value for X is 1.
- Lastly, if the outcome is TT, we got two tails, so the value for X is 2.

As the definition suggests, X is a quantitative variable that takes the possible values of 0, 1, or 2.

It is random because we do not know which of the three values the variable will eventually take. We can ask questions like:

- What is the probability that X will be 2? In other words, what is the probability of getting 2 tails?
- What is the probability that X will be at least 1? In other words, what is the probability of getting at least 1 tail?

As you can see, random variables are not really a new thing, but just a different way to look at the same problem.

Note that if we had tossed a coin three times, the possible values for the number of tails would be 0, 1, 2, or 3. In general, if we toss a coin "n" times, the possible number of tails would be 0, 1, 2, 3, ... , or n.

Example: Observational

Consider getting data from a random sample on the number of ears in which a person wears one or more earrings.

We **define the variable X to be the number of ears** in which a randomly selected person wears an earring.

If the selected person does not wear any earrings, then $X = 0$.

If the selected person wears earrings in either the left or the right ear, then $X = 1$.

If the selected person wears earrings in both ears, then $X = 2$.

As the definition suggests, X is a quantitative variable which takes the possible values of 0, 1, or 2. We can ask questions like:

- What is the probability that a randomly selected person will have earrings in both ears?
- What is the probability that a randomly selected person will not be wearing any earrings in either ear?

NOTE...

We identified the first example as **theoretical** and the second as **observational**. Let's discuss the distinction.

To answer probability questions about a theoretical situation, we only need the principles of probability. However, if we have an observational situation, the only way to answer probability questions is to use the relative frequency we obtain from a random sample.

Here is a different type of example:

Example: Lightweight Boxer

Assume we choose a lightweight male boxer at random and record his exact weight. According to the boxing rules, a lightweight male boxer must weigh between 130 and 135 pounds, so the sample space here is $S = \{\text{All the numbers in the interval } 130\text{-}135\}$. (Note that we can't list all the possible outcomes here.)

We'll **define X to be the weight of the boxer** (again, as the definition suggests, X is a quantitative variable whose value is the result of our random experiment). Here X can take any value between 130 and 135. We can ask questions like:

- What is the probability that X will be more than 132? In other words, what is the probability that the boxer will weigh more than 132 pounds?
- What is the probability that X will be between 131 and 133? In other words, what is the probability that the boxer weighs between 131 and 133 pounds?

What is the difference between the random variables in these examples? Let's see:

- They all arise from a random experiment (tossing a coin twice, choosing a person at random, choosing a lightweight boxer at random).
- They are all quantitative (number of tails, number of ears, weight).

Where they differ is in the type of possible values they can take: In the first two examples, X has three distinct possible values: 0, 1, and 2. You can list them. In contrast, in the third example, X takes any value in the interval 130-135, and thus the possible values of X cover an infinite range of possibilities, and cannot be listed.

A random variable like the one in the first two examples, whose possible values are a list of distinct values, is called a **discrete random variable**. A random variable like the one in the third example, that can take any value in an interval, is called a **continuous random variable**.

Just as the distinction between categorical and quantitative variables was important in Exploratory Data Analysis, the distinction between discrete and continuous random variables is important here, as each one gets a different treatment when it comes to calculating probabilities and other quantities of interest.

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved