CST 363 - Intro to Data Science

Dr. Glenn Bruns

# Lab: discrete random variables - review

1. Define “random variable” in your own words.

I would say something like: “it’s a mapping from outcomes to numbers”, or, “it’s a way of making outcomes numeric”, or, “it’s a way to give a name to a distribution over some numbers”.

1. Suppose I told you that “X is a random variable with possible values ‘red’ and ‘green’”. What is wrong with this statement?

The values of random variables must be numbers! Otherwise what would it mean to get the expectation of a random variable?

1. Consider the sample space of an experiment in which three coins are flipped. The outcomes are HHH, HHT, etc. Draw a picture of the sample space, showing all outcomes.

Draw an oval containing the 8 possible outcomes.

1. Using the sample 3-coin flip example, let random variable X be the number of heads in the 3 flips. For each possible value of the X, draw a line enclosing the elements of the sample space having that value.

For example, for value 2, draw a circle around the set of outcomes HHT, HTH, THH.

1. Using the 3-coin flip again, what is the P(X = 2)? What is P(X > 1)?

Using naive probability, P(X = 2) is ⅜, because event X = 2 has 3 outcomes, and there are 8 total outcomes. Similarly, event X > 1 has 4 associated outcomes, so P(X > 1) = 4/8 = ½.

1. Let Y be 1 if the first coin of a three-coin flip is heads, and be 0 if the first coin is tails. For each outcome in the sample space, what is X + Y?

For HHH the value is 3 + 1 = 4, for HTH the value is 2 + 1 = 3, for TTH the value is 1 + 0 = 1, etc.

1. What is P(X + Y = 3)?

It is ¼, as there are eight possible outcomes and 2 of them belong to the event X + Y = 3. Those two events are HHH and THH. You can work this out by drawing a table of all the outcomes, figuring out the value of X and Y for each outcome, and then figuring out the value of X + Y for each outcome.

1. Assume that every McDonald's customer buys only one item, and that 20% get a hamburger (250 calories), 50% get a quarter pounder with cheese (540 calories), and 30% get a bacon clubhouse burger (750 calories). Let random variable X be the number of calories consumed by a McDonald's customer. Draw a PMF for the distribution of X.

It has three bars. The bar for 250 has height 0.2, the bar for 540 has height 0.5, and the bar for 750 has height 0.3.

1. Using the figures in 8, how many calories does a typical McDonald’s customer consume per visit?

You can’t use a simple average, since more people buy quarter pounders than cheeseburgers. You need to use a weighted average: 0.2 \* 250 + 0.5 \* 540 + 0.3 \* 750 = 545 calories. If you used a simple average you would get ⅓ \* (250 + 540 + 750) = 513 calories.

1. Now we'll write code to simulate. Create an array of length 1000 representing samples from random variable X of problem 8. What is the average of the values in your array? How does this value compare to the result you got for the previous problem?

import numpy as np

# 10

x = np.random.choice([250,540,750], size=1000, replace=True, p=[0.2, 0.5, 0.3])

x.mean()