DATA 252/DATA 551
Modeling and Simulation

Lecture 1: Discrete Random Variables

January 13, 2020

Random variables and sample spaces

Consider a random experiment like tossing a fair coin. This is a random experiment because the outcome is unknown. We can represent this unknown outcome with X, which is called a random variable.

Sample space: the sample space is the set (i.e., collection) of *all* possible outcomes.

- Sample space of tossing a fair coin:

Random variables and sample spaces

- Suppose you toss a fair coin three times. Let X be the number of heads you get. What is the sample space?

- Let Y be the number of tosses until you get the first head. What is the sample space?

Discrete random variables: A random variable is called discrete if the sample space is finite (e.g.: X above) or countably infinite (e.g.: Y above).
Counterexample: weight of a randomly selected person, waiting time until the next bus, etc. These are called *continuous* random variables.

Probability mass function

To fully define a random variable, we need to specify its distribution. In the discrete case, this is done by **assigning probability to each outcome in the sample space**.

- Let X be the outcome of rolling one fair die.
  - Sample space:
  - Corresponding probabilities:

The function $P(x), x = 1, \ldots, 6$ is called the probability mass function for the discrete random variable X.

A valid probability mass function must satisfy the following basic conditions:

i. $P(x) \geq 0$ for all x in the sample space.
ii. $\sum_{x \in \textit{sample space}} P(x) = 1$.

Simulation

We simulate the process of rolling one fair die to verify the probabilities, using `https://www.random.org/dice/`. What can we *expect* from the result?

- Rolling once:
- Rolling 10 times:
- Rolling many, many times (1000? 10000?):

Simulation often requires **repeating** an experiment for **a large number of times**. The result is more trustworthy and can *better* estimate the truth because the summary of data (called a "statistic") has less variation.

Simulation

Simulation has limitations **when the sample space is very large** or **when some outcomes in the sample space has very low chance of occurring**.

- Imagine you have a possibly unfair die with 1000 side and you want to estimate the probability of rolling a one.
- If you don't repeat rolling the die enough times, the outcome of 1 might not occur at all.

Example 1

Let X be the number of heads in two coin tosses. The coins are assumed fair.

a. What is the probability mass function of X?

b. Verify that this probability mass function satisfy the two conditions.

c. How can you simulate from this probability mass function to verify the probabilities empirically?

Example 2

(Example 1.3 from *Introduction to Probability*) Simulation is especially useful when the probabilities are hard to calculate mathematically. Suppose we roll a pair of fair dice. A gambler bets that, in 10 rolls, a pair of six would show up. What is his chance of winning? Describe how you would run a simulation to estimate his chance of winning.

(Can you calculate the exact probability mathematically?)

Example 3

(Example 1.4 from *Introduction to Probability*) You play a game with me as follows. A fair coin is tossed. If a head shows up, I give you a dollar. If a tail shows up, you give me a dollar.

a. Suppose we play this game 10 times. Let X be your net winning after 10 times. Can you design a simulation to study the distribution of X?

b. Suppose you keep playing this game *until you win $5*. Let Y be the length of the game you have to play. Can you design a simulation to study the distribution of Y?