Discrete Random Variables Application

Lt Col Ken Horton   Professor Bradley Warner

08 June, 2020

Exercises

1. Suppose we are flipping a fair coin, and the result of a single coin flip is either heads or tails. Let X be a random variable representing the number of flips until the first heads.

a) Is X discrete or continuous? What is the domain/support of X?

b) What values do you *expect* X to take? What do you think is the average of X? Don't actually do any formal math, just think about if you were flipping a regular coin, how long it would take you to get the first heads.

c) Advanced: In `R`, generate 10,000 observations from X. What is the empirical, from the simulation, pmf? What is the average value of X based on this simulation? Create a bar chart of the proportions. Note: Unlike the example in the Notes, we don't have the pmf, so you will have to simulate the experiment.

Note: There are many ways to do this. Below is a description of one approach. It assumes we are extremely unlikely to go past 1000 flips.

- First, let's sample with replacement from the vector c("H","T"), 1000 times with replacement, use `sample()`.
- As we did in the reading, use `which()` and a logical arguement to find the first occurence of a heads.

d) Find the theoretical distribution.

2. Repeat Problem 1, but with a different random variable. Consider Y: the number of coin flips until the *fifth* heads.

3. Suppose you are a data analyst for a large international airport. Your boss, the head of the airport, is dismayed that this airport has received negative attention in the press for inefficiencies and sluggishness. In a staff meeting, your boss gives you a week to build a report addressing the "timeliness" at the airport. Your boss is in a big hurry and gives you no further information or guidance on this task.

Prior to building the report, you will need to conduct some analysis. To aid you in this, create a list of at least three random variables that will help you address timeliness at the airport. For each of your random variables,

a) Determine whether it is discrete or continuous.

b) Report its domain.

c) What is the experimental unit?

d) Explain how this random variable will be useful in addressing timeliness at the airport.

I will provide one example:

Let D be the difference between a flight's actual departure and its scheduled departure. This is a continuous random variable, since time can be measured in fractions of minutes. A flight can be early or late, so domain is any real number. The experimental unit is each individual (non-canceled) flight. This is a useful random variable because the average value of D will describe whether flights take off on time. We could also find out how often D exceeds 0 (implying late departure) or how often D exceeds 30 minutes, which could indicate a "very late" departure.

4. Consider the experiment of rolling two fair six-sided dice. Let the random variable Y be the absolute difference between the two numbers that appear upon rolling the dice.

a) What is the domain/support of Y?

b) What values do you *expect* Y to take? What do you think is the average of Y? Don't actually do any formal math, just think about the experiment.

c) Find the probability mass function and cumulative distribution function of Y.

d) Find the expected value and variance of Y.

e) Advanced: In `R`, obtain 10,000 realizations of Y. In other words, simulate the roll of two fair dice, record the absolute difference and repeat this 10,000 times. Construct a frequency table of your results (what percentage of time did you get a difference of 0? difference of 1? etc.) Find the mean and variance of your simulated sample of Y. Were they close to your answers in part d?

5. Prove the Lemma from the Notes: Let X be a discrete random variable, and let a and b be constants. Then $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$.

6. In the Notes, we saw that $\mathrm{Var}(X) = \mathrm{E}[(X - \mu_X)^2]$. Show that $\mathrm{Var}(X)$ is also equal to $\mathrm{E}(X^2) - [\mathrm{E}(X)]^2$.