Assignment: Probability, Random Variables and Simulation

Due 2/26 11:59pm

February 12, 2021

Problem 1

An elementary school is offering 3 language classes: one in Spanish, one in French, and one in German. The classes are open to any of the 100 students in the school. There are 28 students in the Spanish class, 26 in the French class, and 16 in the German class. There are 12 students that are in both Spanish and French, 4 that are in both Spanish and German, and 6 that are in both French and German. In addition, there are 2 students taking all 3 classes.

(a) If a student is chosen randomly, what is the probability that he or she is not in any of the language classes?

(b) If a student is chosen randomly, what is the probability that he or she is taking exactly one language class?

(c) If 2 students are chosen randomly, what is the probability that at least 1 is taking a language class?

Problem 2

Suppose that 5 percent of men and .25 percent of women are color blind. A color-blind person is chosen at random. What is the probability of this person being male? Assume that there are an equal number of males and females. What if the population consisted of twice as many males as females?

Problem 3

A salesman has scheduled two appointments to sell encyclopedias. His first appointment will lead to a sale with probability .3, and his second will lead independently to a sale with probability .6. Any sale made is equally likely to be either for the deluxe model, which costs $1000, or the standard model, which costs $500. Determine the probability mass function of X, the total dollar value of all sales.

Problem 4

Suppose the cdf of continuous random variable X is given by:

$$F(x) = \begin{cases} 1 - \exp(-x) & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

Calculate the density, expectation and variance of X.

Problem 5

The Cantor distribution is the classic example of a singular distribution - it is neither discrete nor continuous! We construct Cantor random variables as follows: Let $\{X_i, i = 1, 2, ...\}$ (for i in the positive integers) be an infinite sequence of independent random variables that are 0 with probability 0.5 and 1 with probability 0.5.

Then

$$Y = 2\sum_{i=1}^{\infty} \frac{X_i}{3^i}$$

has a cantor distribution. We can simulate (approximately) from this distribution by only taking the sum up to $i = m$.

Write a function that, given n (the number of samples we want), and m (as defined above) samples an approximation of the cantor distribution. Do not use any loops.

Generate a large number of samples from this distribution, and plot the histogram using seaborn in a 5 different subplots for m = 10, 25, 50, and 100, 250. What changes with different m?

Problem 6

Given a discrete random variable X with a finite set of possible values $\{x_1, x_2, ..., x_n\}$, and pmf p, we can sample from X by generating a uniform random variable U, and then letting $X = x_i$ if

$$U \in [\sum_{j=0}^{i-1} p(j), \sum_{j=0}^{i} p(j))$$

. Write a function that takes an array of possible values, an array with the corresponding pmf values, and n (sample size), and returns a sample from X. Do not use a for loop. Hint: if your pmf array is 1d, use np.tile to make a comparison with U broadcastable. This is an annoying exercise in dealing with the dimensions of numpy arrays.

Problem 7

A Cauchy distribution can be sampled using scipy.stats.cauchy.rvs(). If I generate a sample of n cauchy random variables, $X_1, ..., X_n$, I can define the running sample mean $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ and sample variance, $S_m^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X}_m)^2$. Plot the running sample mean and sample variance. Do they converge?