Statistics for Data Science
Unit 4 Part 1 Homework: Discrete Random Variables

January 21, 2019

1. **Best Game in the Casino**

   You flip a fair coin 3 times, and get a different amount of money depending on how many heads you get. For 0 heads, you get \$0. For 1 head, you get \$2. For 2 heads, you get \$4. Your expected winnings from the game are \$6.

   (a) How much do you get paid if the coin comes up heads 3 times?

   (b) Write down a complete expression for the cumulative probability function for your winnings from the game.

2. **Reciprocal Dice**

   Let X be a random variable representing the outcome of rolling a 6-sided die. Before the die is rolled, you are given two options:

   (a) You get $1/E(X)$ in dollars right away.

   (b) You wait until the die is rolled, then get $1/X$ in dollars.

   Which option is better for you, in expectation?

3. **The Baseline for Measuring Deviations**

   Given any random variable X and a real number t, we can define another random variable $Y = (X - t)^2$. In other words, for any random variable X, we can choose a real number, t, as a baseline and calculate the squared deviation of X away from t.

   You might wonder why we often square deviations (instead of taking an absolute value, or cubing them, etc.). This exercise will shed some light on why this is a natural choice.

   (a) Write down an expression for $E(Y)$ and simplify it as much as you can. Even though we haven't proved this yet, you can use the fact that for any two random variables, A and B, $E(A + B) = E(A) + E(B)$.

   (b) Taking a partial derivative with respect to t, compute the value of t that minimizes $E(Y)$. (Hint: Your answer should be a very familiar value)

   (c) What is the value of $E(Y)$ for this choice of t? (Hint: this should also be a very familiar value)

4. **Optional Advanced Exercise: Heavy Tails**

One reason to study the mathematical foundation of statistics is to recognize situations where common intuition can break down. An unusual class of distributions are those we call *heavy-tailed*. The exact definition varies, but we'll say that a heavy-tailed distribution is one for which not all moments are finite. Consider a random variable M with the following pmf:

$$p_M(x) = \begin{cases} c/x^3, & x \in \{1, 2, 3, ...\} \\ 0, & otherwise. \end{cases}$$

where c is a constant (you can calculate its value if you like, but it's not important).

(a) Is $E(M)$ finite?

(b) Is $var(M)$ finite?

Heavy-tailed distributions may seem odd, but they're not as rare as you might suspect. Researchers argue that the distribution of wealth is heavy-tailed; so is the distribution of computer file sizes, insurance payouts, and area burned by forest fires. These random variables are problematic in that a lot of common statistical techniques don't work on them. For this class, we'll assume that all of our variables don't have heavy-tails.