

Statistics for Data Science

Unit 4 Homework: Random Variables

January 30, 2017

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. Processing Pasta

A certain manufacturing process creates pieces of pasta that vary by length. Suppose that the length of a particular piece, L , is a continuous random variable with the following probability density function.

$$f(l) = \begin{cases} 0, & l \leq 0 \\ l/2, & 0 < l \leq 2 \\ 0, & 2 < l \end{cases}$$

- (a) Write down a complete expression for the cumulative probability function of L .
- (b) Using the definition of expectation for a continuous random variable, compute the expected length of the pasta, $E(L)$.

3. The Warranty is Worth It

Suppose the life span of a particular (shoddy) server is a continuous random variable, T , with a uniform probability distribution between 0 and 1 year. The server comes with a contract that guarantees you money if the server lasts less than 1 year. In particular, if the server lasts t years, the manufacturer will pay you $g(t) = \$100(1 - t)^{1/2}$. Let $X = g(T)$ be the random variable representing the payout from the contract.

- (a) Compute the expected payout from the contract, $E(X) = E(g(T))$, using the expression for the expectation of a function of a random variable.
- (b) Next, compute $E(X)$ another way, by first characterizing the random variable X . Follow these steps:

- i. First, suppose that you are given a specific value for the payoff, $X = x$, where $\$0 \leq x \leq \100 . What is the value for T that results in this payoff?
- ii. Next, suppose that all you know is that the payoff is less than or equal to some value, $X \leq x$, where again $\$0 \leq x \leq \100 . What does this tell you about the life span of the server? Specifically, write down the condition for T that is equivalent to $X \leq x$.
- iii. Using the condition you just wrote down, what is the probability that $X \leq x$? Write down a complete expression for the cumulative probability function of X .
- iv. Take a derivative to compute the probability density function for X .
- v. Use the pdf of X to compute $E(X)$. If you did everything right, your answer should match what you got in part (a).

4. 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 use properties of expectation to simplify it as much as you can.
- (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 ?

5. Optional Advanced Exercise: Characterizing a Function of a Random Variable

Let X be a continuous random variable with probability density function $f(x)$, and let h be an invertible function where h^{-1} is differentiable. Recall that $Y = h(X)$ is itself a continuous random variable. Prove that the probability density function of Y is

$$g(y) = f(h^{-1}(y)) \cdot \left| \frac{d}{dy} h^{-1}(y) \right|$$