Unit 3 Homeowrk: Summarizing Distributions

w203: Statistics for Data Science

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.

1. (3 points) How much do you get paid if the coin comes up heads 3 times?
2. (3 points) Write down a complete expression for the cumulative probability function for your winnings from the game.

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. Either:

a. You get $1/E(X)$ in dollars right away; or,
b. You wait until the die is rolled, then get $1/X$ in dollars.

1. (3 points) Which option is better for you, in expectation? (2 point)
2. (3 points) Knowing who you are, and your tolerance for risk, which of the two options would you take, and why?

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.

1. (3 points) Compute the expected payout from the contract, $E(X) = E(g(T))$.

Great Time to Watch Async

Suppose your waiting time in minutes for the Caltrain in the morning is uniformly distributed on [0, 5], whereas waiting time in the evening is uniformly distributed on [0, 10]. Each waiting time is independent of all other waiting times.

1. If you take the Caltrain each morning and each evening for 5 days in a row, what is your total expected waiting time? (3 points)
2. What is the variance of your total waiting time? (3 points)
3. What is the expected value of the difference between the total evening waiting time and the total morning waiting time over all 5 days? (3 points)
4. What is the variance of the difference between the total evening waiting time and the total morning waiting time over all 5 days? (3 points)

Maximizing Correlation

Show that if $Y = aX + b$ where X and Y are random variables and $a \neq 0$, $corr(X, Y) = -1$ or $+1$. (3 points)

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:

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

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

1. Is $E(M)$ finite?
2. Is $V(M)$ finite?

(Up to 3 bonus points total)

> 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.
>
> In this class, we won't cover heavy tailed distributions in depth, but we want you to become alert to their possibility.

Note: Maximum score on any homework is 100%