

Statistics for Data Science

Unit 4 Part 1 Homework: Discrete Random Variables

Brad Andersen
W203 Section 4

January 30, 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?**

Because:

- Our coin-flipping activities are divided into three separate *trials*
- Each trial flip has only two outcomes: heads or tails
- Each trial flip is independent of the others
- The probability of heads or tails is consistent from one trial flip to the next

our coin-flipping activities can be considered a **binomial experiment**.

Payout is based upon the number of heads resulting in the experiment: 0, 1, 2 or 3. Probabilities can be calculated using the binomial distribution probability theorem, as follows:

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

where x represents the number of heads results, n represents the number of trials in the experiment, and p represents the probability of the outcome.

For example, calculating the probability that the experiment's results include no heads,

$$\begin{aligned} P(X = 0) &= b(0; 3, 0.5) = \binom{3}{0} p^0 (1 - 0.5)^{3-0} \\ &= \frac{3!}{0!(3-0)!} \times 0.5^0 \times 0.5^3 \\ &= \frac{6}{6} \times 1 \times 0.125 \\ &= 1 \times 1 \times 0.125 \\ &= 0.125 \end{aligned}$$

Therefore, the probability mass function for the experiment is as follows:

x	0	1	2	3
$p(x)$	0.125	0.375	0.375	0.125

Payout from the results of the experiment can be considered a function of the discrete random variable X , itself the results of the coin-flipping experiment. If $h(X)$ is this function, it too is a random variable such that $Y = h(X)$. The derived probability mass function for Y is as follows:

y	\$0	\$2	\$4	\$ m
$p(y)$	0.125	0.375	0.375	0.125

where m represents the winnings for experiment results of three heads, the winnings to be calculated. Given that we know the expected winnings from the experiment are \$6, winnings for results of three heads (here represented using m) can be calculated as follows:

$$\begin{aligned}
 E(Y) &= E[h(X)] = \sum_D h(x) \times p(x) \\
 &= \$6 = (\$0)(0.125) + (\$2)(0.375) + (\$4)(0.375) + (m)(0.125) \\
 &= \$6 = \$0 + \$0.75 + \$1.50 + (m)(0.125) \\
 &= \$3.75 = (m)(0.125) \\
 &= \$30 = m
 \end{aligned}$$

Therefore, winnings for experiment results of three heads are \$30.

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

$$F(y) = \begin{cases} 0.125 & 0 \leq y < 1 \\ 0.500 & 1 \leq y < 2 \\ 0.875 & 2 \leq y < 3 \\ 1 & 3 \leq y \end{cases}$$

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?

The probability mass function for a fair six-sided die is as follows:

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

Therefore, the expected value of random variable X can be calculated as follows:

$$\begin{aligned}
 E(X) &= \sum_{x \in D} x \times p(x) \\
 &= (1)(1/6) + (2)(1/6) + (3)(1/6) + (4)(1/6) + (5)(1/6) + (6)(1/6) \\
 &= 1/6 + 1/3 + 1/2 + 2/3 + 5/6 + 1 \\
 &= 21/6 \\
 &= 3.50
 \end{aligned}$$

Given that the expected value of X is 3.50, $1/E(X)$ equals approximately 0.2857. Rounding to the nearest cent, one would receive \$0.29 if option a) were chosen prior to the rolling of the die.

Given that the die is fair, each of its six sides has an equal probability of appearing when the die is rolled; the probability that any side appears is therefore $1/6$, as shown in the probability mass function, above. If one were to receive $1/X$ dollars representing the roll of the die, payouts for each of the die's six sides (rounded to the nearest cent) are as follows:

Die roll	1	2	3	4	5	6
Calculated payout	\$1.00	\$0.50	\$0.33	\$0.25	\$0.20	\$0.17

The payout prior to the rolling of the die (\$0.29) is equidistant from the payouts associated with rolling a 3 (\$0.33) and a 4 (\$0.25). Given that the probability of rolling any one side of the die is equivalent, one would be just as likely to roll a die value associated with a payout *greater than* \$0.29 as *less than* \$0.29. However, payouts associated with die rolls of 1 or 2 are further from \$0.29 than payouts associated with die rolls of 5 or 6. Therefore, one could potentially maximize gains (or minimize losses) by selecting option b), waiting until the die is rolled and collecting $1/X$ dollars.

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 , $E(A + B) = E(A) + E(B)$.**

Given $Y = (X - t)^2$, expression $E(Y)$'s simplification is as follows:

$$\begin{aligned}
 E[h(X)] &= \sum_D h(x) \times p(x) \\
 &= E[(X - t)(X - t)] \\
 &= E[X^2 - 2Xt + t^2] \\
 &= E[X^2] - E[2Xt] + E[t^2] \\
 &= E[X^2] - 2E[Xt] + E[t^2] \\
 &= \underline{\underline{E[X^2] - 2tE[X] + t^2}}
 \end{aligned}$$

- (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)**

The partial derivative with respect to t can be calculated as follows:

$$\begin{aligned}
 \frac{\delta f}{\delta t}(E[X^2] - 2tE[X] + t^2) &= \frac{\delta f}{\delta t}(E[X^2]) + \frac{\delta f}{\delta t}(2tE[X]) + \frac{\delta f}{\delta t}(t^2) \\
 &= 0 - 2E[X] + 2t \\
 &= 2t - 2E[X]
 \end{aligned}$$

Knowing that Y (i.e. $(X - t)^2$) must be greater than or equal to 0, calculating t to minimize $E(Y)$ is as follows:

$$\begin{aligned}
 0 &= 2t - 2E[X] \\
 -2t &= -2E[X] \\
 t &= \underline{\underline{E[X]}}
 \end{aligned}$$

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

Substituting $E[X]$ for t , calculating $E(Y)$ is as follows:

$$\begin{aligned}
 E[Y] &= E[X^2] - 2tE[X] + t^2 \\
 &= E[X^2] - 2(E[X])(E[X]) + E[X^2] \\
 &= 2E[X^2] - 2E[X^2] \\
 &= \underline{\underline{0}}
 \end{aligned}$$

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:

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

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?**

No - neither $E(M)$ nor $var(M)$ is finite. Given that:

- x is within the set of all integers greater than 0, and
- $p_M(x)$ is an exponential function

The probability of x decreases at an exponentially smaller rate as the value of x increases. Therefore, the expected value of M will change without bound. Because the mean of M never ceases to change, neither does its variance.