Probability, Statistics, and Gambling

Sanjit Dandapanthula

October 9, 2022

Warm-up.

Each student begins with 0 points. At each turn, you will individually choose whether to "hit" or "fold". The instructors will flip two coins. If both coins land heads, the game is over. If you fold, you get to keep all of your current points, but you will be unable to play in any more rounds until the end of the game. If you hit and the game does not end, you earn one more point. On the other hand, if you hit and the game ends, you will lose all of your points for the current game. We'll play this game independently three times; anyone who wins at least 10 points total will get a prize!

Introduction.

Today, we will be learning about how to quantify uncertainty in probabilistic games. Then, we'll learn how to strategically gamble in order to maximize our payout in games like the one we just played.

Probability and Statistics.

We want to encode our intuition about probability into mathematical language. Assign the number 1 to heads and the number 0 to tails. In some sense, we expect each flip of the coin to land $\frac{1}{2}$, since it lands 1 half of the time and 0 the other half of the time. A **random variable** X is a set of numbers, together with the probabilities that they occur. For example, when we flip a coin, the outcomes are $\{0, 1\}$ and we have the following probabilities associated to them: $\mathbb{P}[X = 0] = \mathbb{P}[X = 1] = 0.5$.

In general, if a random variable X has outcomes x_i which have a probability $\mathbb{P}(X = x_i)$ of occurring, we define the expected value of X as a "weighted average" of the outcomes (why?):

$$\mathbb{E}[X] = x_1 \cdot \mathbb{P}(X = x_1) + \cdots + x_n \cdot \mathbb{P}(X = x_n).$$

In statistics, the expected value is also sometimes called the mean μ.

Exercise 1. *What is the expected value of one roll of a fair die?*

Exercise 2. *I don't like ones. I'm going to roll one die, but if I roll a one, I'm going to roll one more time. What would you expect me to end up with?*

Exercise 3. *Show that the expectation is* ***linear****; namely, show that if X and Y are random variables and a, b are constants then $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. Linearity of expectation is a really important trick, so keep it in mind!*

Exercise 4. *Mathematicians call the solution to this problem **Jensen's inequality**. Suppose you're stuck on an island and you need food. A game-master approaches you with a cart of bananas (don't question it) and has two games for you. The first game: she can roll a die and you can take bananas equal to the square of her roll. The second game: she can roll two dice and you can take bananas equal to the product of her rolls. Which game should you play to take the most bananas?*

Exercise 5. *This exercise is meant as a challenge; feel free to skip it and move on. Prove **Markov's inequality**: the probability that $X \geq a$ is at most $\frac{\mathbb{E}[X]}{a}$. This is an example of what probabilists call a **tail bound** (why?).*

Another useful quantity in probability and statistics is called the **variance**. The variance is defined by:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The variance measures how far away a random variable tends to be from its expected value (why?). In statistics, the square root of variance is called the **standard deviation** and is sometimes denoted σ (and the variance is called σ^2).

Exercise 6. *What is the variance of a fair n-sided die roll?*

Exercise 7. *Using the linearity of expectation, prove that $\sigma^2 = \mathbb{E}[(X-\mu)^2]$ is an equivalent formula for the variance.*

Exercise 8. *This exercise is meant as a challenge; feel free to skip it and move on. Prove using Markov's inequality and the previous exercise that $\mathbb{P}[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}$ (this is called **Chebyshev's inequality**). This means that X deviates k standard deviations from the mean at most $\frac{1}{k^2}$ of the time. Hint: this statement is the same as $\mathbb{P}[(X-\mu)^2 \geq k^2\sigma^2] \leq \frac{1}{k^2}$.*

Here's another neat probability trick. If a game depends on itself, you can find and solve a **recurrence relation** to compute a probability.

Exercise 9. *This problem is about a **random walk**. I'm standing on the number line at $x = 10$ and* *my dog is at $x = 0$. Every minute, my dog moves right one step with probability p and left one step with probability $1 - p$. How many minutes should I expect my dog to take to find me?*

Exercise 10. *Imagine we're biologists and we're in the lab. You put one bacterium in a Petri dish (or something, I'm not actually a biologist). Each second, the bacterium has a probability p of splitting into two identical bacteria and a probability $1 - p$ of dying. What is the probability that your colony of bacteria eventually dies?*

Gambling Strategies.

First, we're going to learn to sum a geometric series.

Exercise 11. *Find a formula for $1 + 2 + 2^2 + \cdots + 2^n$. This is called a* ***geometric series****.*

Let's say you're playing a game with probability p of winning and $1 - p$ of losing. You can bet on the outcome of the game. If you win, you'll win back twice your initial bet. If you lose, you lose your initial bet. The **martingale betting strategy** is to start by betting one dollar, and to bet double the amount you previously bet each time you lose your money.

Exercise 12. *What is the expected amount of money you can make in "unlimited" time using the martingale betting strategy if you have an "unlimited" store of money to draw on? The answer should surprise you!*

Exercise 13. *This problem is meant to be a challenge. What is the expected amount of money you can make using the martingale betting strategy if you only have \$B in the bank? Hint: use the geometric series formula you derived above.*

There are a few more intricacies to betting strategies. For example, consider the following.

Exercise 14. *This problem is a modeling problem. I'm not going to give you any numbers, so you'll come up with a model that you think makes sense with the information you have and give me an answer. When thinking about real problems, you won't always have complete information; you need to do the best you can with what you have. You and your friend are racing tadpoles. The first tadpole is really fast on some days and really slow on some days. The second tadpole is kinda fast on some days and kinda slow on some days. The third tadpole is very consistently average speed. Which tadpole should you bet on?*

Here's an important note. Let's say you have B dollars in the bank (your life's savings). Let's say I have a weighted coin that lands heads 51% of the time and tails 49% of the time. Suppose I was willing to bet you that the coin lands tails. How much are you betting? You probably don't want to bet your life savings, since you have a 49% chance of losing all your money. You'd rather play it safe and save more money to gamble later.

The lesson: you shouldn't always bet on something just because you're expected to win. The casino is able to get away with this because they have a large store of money, play the game thousands of times with many patrons and make money on expectation (this is called the **law of large numbers**). You probably can't. Kids, don't gamble.

Also, as a note, there is something called the Kelly criterion, which tells you how much of your savings to gamble, given the odds of winning. The proof of the Kelly criterion will be a challenge problem below.

Challenge Problems.

Exercise 15. *A regular deck of cards has 52 cards, 4 of which are aces. Suppose you shuffle the deck and* *start dealing cards one-by-one. How many cards should you expect to flip over before you see the first ace?*

Exercise 16. *Go back to the warm-up and compute the expected return at each turn. Come up with a strategy to maximize your expected earnings if you're allowed to play the game many times.*

Exercise 17. *You're standing in line to watch John speak at an event. The manager of the event says that the first person in line who shares a birthday with someone ahead of them will get a prize. Assuming that birthdays are uniformly distributed among the 365 days of a (non-leap) year, where should you stand in line?*

Exercise 18. *This question will ask you to prove the* ***Kelly criterion****. You may need to use calculus, so it is a challenge problem. Suppose* p *is the probability that you win a game, and* $q = 1 - p$ *is the probability that you lose. Suppose you bet a fraction* f *of your wealth. Then, suppose you'll make* $fb \cdot W$ *dollars if you win and lose* $fa \cdot W$ *dollars if you lose, where* W *is your current wealth. What fraction* f^* *of your wealth should you bet to maximize your earnings? Hint: the rate at which you make money is* $r = (1 + fb)^p(1 - fa)^q$*. Maximize this quantity (perhaps by taking* log *of both sides and doing some calculus).*