Jane Street: Probability and Markets
Practice Problems

Evan Zimmerman

May 31, 2024

Question 1: Proabability, Counting, and Independence

Part (a): What's the probability that when we roll 2 d6s and add up the results, we get 7?

For this problem we can think of all the different possible combinations that add up to 7. {1,6},{2,5},{3,4},{4,3},{5,2}{6,1}. The probability of any of these combinations occurring is $1/36$ because there is a $1/6$ chance of rolling a given number on the first die and a $1/6$ chance of rolling a given number on the second die, so $1/6 * 1/6 = 1/36$. We have 6 different possible combinations so the solution is $\underline{1/36 * 6 = 6/36 = 1/6}$.

Part (b): What's the probability of getting a sum of 10 when you roll 2 d6s? Hint: there's no way to flip [red 5, blue 5]!

For this problem we can think of all the different possible combinations that sum to 10: {4,6}, {5,5}, {6,4}. The probability of rolling any of these 3 combinations is $\underline{3/36 = 1/12}$, following the probabilistic reasoning given above.

Part (c): What's the probability that you draw 2 face cards from a deck of cards, if you put the first card back in before your second draw ("with replacement") ?

A deck of cards has 12 face cards (3 per suite). Since we are drawing with replacement, we can think of the draws as independent. P(face card on first draw AND second draw) = P(face card on first draw) * P(face card on second draw) = $\underline{12/52 * 12/52 = 3/13 * 3/13 = 9/169}$

Part (d): What's the probability that you draw 2 clubs from a deck of cards, if you don't put the first card back in ("without replacement")?

There are 13 clubs in a deck of cards. P(drawing 2 clubs) = P(drawing the first club) * P(drawing the second club) = $\underline{13/52 * 12/51 = 1/4 * 12/51 = 1/4 * 4/17 = 1/17}$
*Note: Since we are drawing without replacement, when we calculate the probability of drawing the second club, we assume we have one less club in the deck.

Part (e): What's the probability that the product of 2 rolls of a d6 is odd? 3 rolls?

What does it take to get an odd product? We can not multiply by 2 or 4 or 6. Let's think of possible combinations: (1,1), (1,3), (1,5) and the reverse (excluding (1,1)) = 5/36, (3,3) and (3,5) (and reverse) are odd, so that gives 3/36 more combinations, and then (5,5) is odd as well = 1/36. So the probability that the product of 2 rolls of a d6 is odd is 9/36 = 1/4.
OR <u>p(not rolling an even number) = 3/6 * 3/6 = 1/4</u> .
In the case that there are 3 rolls, as long as we do not roll an even number, we will get an odd product:
<u>p(not rolling an even number in 3 rolls) = $3/6 * 3/6 * 3/6 = 1/8$.</u>

Question 2: Random Variables

Part (a): If B and R are independent random variables, are B+R and B*R random variables?

Yes they are random variables. We can make random variables out of random variables.

Question 3: Expected Value

Part (a): What is the expected value if the die is reweighed so that 6 occurs half the time?

Let X = the number of pips on the die. $p(X = 6) = 1/2, p(X = 1, 2, 3, 4, 5) = 1/2 \div 5 = 1/10$ respectively (uniformly distributed). So $\underline{E[X] = 6 * 1/2 + \sum_{i=1}^{5} 1/10 * i = 3 + 1/10 + 2/10 + 3/10 + 4/10 + 5/10 = 4.5}$

Part (b): What would the expected value be if the 6 is replaced by a 12?

$\underline{E[X] = 12 * 1/2 + \sum_{i=1}^{5} 1/10 * i = 3 + 1/10 + 2/10 + 3/10 + 4/10 + 5/10 = 7.5}$

Part (c): What's an example of two random variables X and Y such that E[X*Y] doesn't equal E[X]*E[Y]

$E[X * Y] \neq E[X] * E[Y]$ when X and Y are dependent. Let X be a random variable that takes values 0 and 1 with equal probability, and Y = X. Y is entirely dependent on X. E[X] * E[Y] = 0.5 * 0.5 = 0.25, but $E[XY] = 0^2 * 0.5 + 1^2 * 0.5 = 0.5$

Question 4: Confidence Intervals

Part (a): What is a 95% confidence interval for the number of taxis in New York City?

Hypothetical data assumptions

- Sample mean ($\bar{x}$) = 10,000 taxis
- Sample standard deviation (s) = 2,000 taxis
- Sample size (n) = 30

Steps to calculate 95% Confidence Interval

1. Determine the critical value
   - For a 95% confidence level and a reasonably large sample size, use the z-value for the standard normal distribution.
   - The critical value (z*) for a 95% confidence interval is approximately 1.96.
2. Calculate the standard error of the mean (SE)
   $SE = \frac{s}{\sqrt{n}} = \frac{2000}{\sqrt{30}} = 363.15$
3. Calculate the margin of error (ME)
   $ME = z^* * SE = 1.96 * 363.15 = 715.69$
4. Construct the confidence interval
   $10,000 \pm 715.69$

Based on the hypothetical data, the 95% confidence interval for the number of taxis in New York City is approximately: <u>(9284,10716)</u>

Real World Application:

1. Collect a random sample of the number of taxis.
2. Calculate the sample mean and standard deviation.
3. Determine the sample size.
4. Use the appropriate critical value (z* or t-value based on sample size).

5. Calculate the standard error and margin of error.
6. Construct the confidence interval.

This method assumes the sample is representative of the population and that the data follows a normal distribution, which is reasonable given a sufficiently large sample size (Central Limit Theorem).

Question 5: Conditional Probability

Part (a): What is the expected value of a d6, if I tell you the outcome is at least 4?

The outcome could be 4,5, or 6 with equal probability (1/3 probability for each outcome). X = value of a d6.
$\underline{E[X|X \geq 4] = 1/3 * 4 + 1/3 * 5 + 1/3 * 6 = 5}$

Part (b): If I flip 3 coins, what's the probability that I get exactly 2 heads given that I get at least 1 tail?

Let X = Number of heads in 3 coin flips. $P(X = 2|3 - X \geq 1)$. Let's use the formula. There are 8 possible outcomes by the way.

$$\begin{aligned} P(X = 2|3 - X \geq 1) &= p(X = 2 \text{ and } 3 - X \geq 1)/p(3 - X \geq 1) \\ &= 3/8 \div 7/8 \\ &= \underline{3/7} \end{aligned}$$

For the numerator, there are 3 different permutations that lead to 2 heads and at least 1 tail. For the denominator, there is only 1 outcome where there is not at least 1 tail.

Part (c): What's the expected value of a d6, given that the outcome is prime?

Let X = value of a d6. The prime values of a die are 2,3, and 5. Assume the probability of rolling any of the prime values follows a uniform distribution.

$$\begin{aligned} E[X|\text{X is prime}] &= E[X|X = 2, 3, \text{or } 5] \\ &= 1/3 * 2 + 1/3 * 3 + 1/3 * 5 \\ &= \underline{10/3} \end{aligned}$$

Question 6: Market Making

Part (a): Consider a contract that pays out a dollar for every pip on the outcome of a d6 roll. You'll get $1 if you roll a 1, $2 if you roll a 2, etc. Why does it make sense to make a market "3 at 4, 10 up."

"3 at 4, 10 up" means you are willing to bid $3 for 10 contracts and sell at $4 for 10 contracts. By setting the buy price at $3 and the sell price at $4, the market maker ensures a profit margin. If a market participant buys the contract for $4 and another sells it for $3, the market maker profits from the spread.

The pricing "3 at 4, 10 up" makes sense because it aligns with the expected value of the contract while providing the market maker with a reasonable profit margin and compensating for the risk involved. The bid price ($3) is slightly below the expected value, and the ask price ($4) is slightly above it, creating a balanced market for trading this contract.

Part (b): Make a market on the outcome of a 20-sided die roll (d20).

The expected value for a 20-sided die roll is 10.5. Following the reasoning outlined in the previous answer, a balanced market would be 10 at 11, N up. We want to set a bid price below the expected value and a sell price above the expected value to generate a profit margin.

Part (c): Make a market on the outcome of the maximum of 3 d6 rolls.

First find the probability distribution of the maximum of 3 rolls. Let the maximum of 3 rolls be X. We can use the CDF. For a single roll Y, the CDF is $P(Y \leq x) = \frac{x}{6}$.

The probability that 3 rolls is less than or equal to x is:

$$P(X \leq x) = (\frac{x}{6})^3$$

Building on this, we can calculate the probabilities for each of the possible values of the maximum of the 3 die rolls (X).

$$\begin{aligned}
P(X = x) &= P(X \leq x) - P(X \leq x - 1) \\
P(X = 1) &= (\frac{1}{6})^3 \\
P(X = 2) &= (\frac{2}{6})^3 - (\frac{1}{6})^3 \\
&\dots \\
P(X = 6) &= 1 - (\frac{5}{6})^3
\end{aligned}$$

Then we can use the expected value formula: $E[X] = \sum_{i=1}^{6} i * P(X = i) = 4.958$. So we will set a bid price below the expected value and an ask price above it. Our market could be 4.50 at 5.50, N up.

Part (d): The outcome of a d6, with a choice to re-roll once if you don't like the first roll for any reason (but you can't go back to the first roll if the second turns out worse). Hint: who gets to decide to re-roll? Does it matter?

It absolutely matters who gets to decide the re-roll. If the player gets to decide, they will re-roll if they get a 3 or lower (if they roll a 3 they have a 50% chance of increasing on the next roll). If the player rolls a 4, they are more likely to lose money if they re-roll. There is a 50% chance they roll a 4,5 or 6 initially, and a 50% chance they don't and they will re-roll. Calculating the expectation: $1/2 * 5 + 1/2 * 3.5 = 4.25$. With an expected winnings of \$4.25, we could set our market to \$4 at \$4.50, N up.

If the house gets to decide the re-roll, the expected winnings are different. The house will make the player re-roll if they roll a 4,5, or 6 because there is a greater than 50% chance of rolling less. This means the player has a 50% chance of re-rolling, or a 50% chance of sticking with a 1,2, or 3. Calculating the expectations: $0.50 * 2 + 0.50 * 3.5 = 2.75$. We could set our market to \$2.50 at \$3.00, N up.

Part (e): The temperature in an hour

To make a market on the temperature in an hour, we need to calculate the expected value of the temperature in an hour. Temperature is a continuous random variable unlike the random variables in the previous cases. If we knew the pdf of the temperature, we could calculate the expected value through integration, and then make our market by setting our bid slighlty below the expected value and our offer slightly above.

We could also calculate expected value by collecting data on temperature and making a forecast based on historical data.

Part (f): 1,000,000 times the outcome of a six sided die. How would you feel if I hit your bid and the roll ended up a 1?

If X is the outcome of the die roll. Then our random variable in this case is $1000000 * X$. $E[1000000 * X] = 1000000 * E[X] = 3,500,000$. In this case we could make our market at \$3,000,000 at \$4,000,000, 10 up. If you hit the bid, and the roll ended up being a one we could lose a lot of money.

Over a large number of trades, the law of large numbers helps ensure that the actual payouts converge to the expected value. In this case, it would be best to manage risk through diversification. Unexpected outcomes, such as a roll of 1, can lead to substantial losses, but these are offset by the profitable trades when the outcomes are higher.

Question 7: Problem Solving Tools

Part (a): Suppose you flip a coin until you see heads. How many flips do you expect to need?

Think of this recursively. You expend 1 flip, and then you're in a situation where there's a 1/2 chance you're done (no more flips!), and a 1/2 chance you have to start all over again. If you have to re flip, you still expect the same number of flips as you did before this flip, because coin flips are independent of each other. When a process has the same distribution no matter what successes or failures you had earlier, we consider it memoryless, and often times it helps to think about it recursively.

This gives us: $E[flips] = 1 + 1/2 * 0 + 1/2 * E[flips]$ **Solving the above gives us E[flips] = 2**. (In fact, in general, if something has a probability p, you'll expect to need 1 / p trials to get your first success.)

A situation like this follows the geometric distribution. This distribution describes the number of trials needed to get the first success in a series of independent and identically distributed Bernoulli trials. Each trial has two possible outcomes: success (with probability p) and failure (with probability 1-p). $E[X] = 1/p$

Part (b): How many flips to you need to see 3 heads in a row? Hint: you can either write one big recursive equation, or figure out how many flips for 2 heads first and use that.

Let X = the number of flips needed to see 3 heads in a row. We are trying to solve for E[X]. We can best model this problem with a tree diagram:

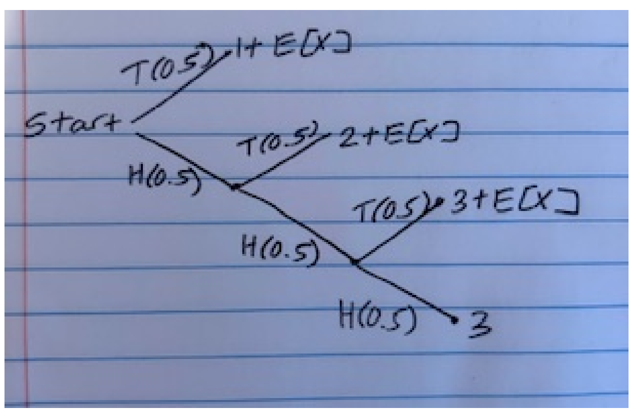

Figure 1: Enter Caption

The key idea with this tree diagram is that every time we flip tails, we have to restart. So it is the number of flips up until that point plus the expected value of flipping 3 consecutive heads.

$$\begin{aligned} E[X] &= \frac{1}{2}(1 + E[X]) + \frac{1}{4}(2 + E[X]) + \frac{1}{8}(3 + E[X]) + \frac{1}{8}(3) \\ E[X] &= \frac{7}{8}x + \frac{14}{8} \\ 8E[X] &= 7E[X] + 14 \\ E[X] &= 14 \end{aligned}$$

Let's consider the general case. $f(n)$ is the expected length for n consecutive tosses. Assume we are at $f(n-1)$. From this state, there is a 50% chance we toss heads (total tosses = $f(n-1)+1$) and there is a 50% chance we toss tails (total tosses = $f(n-1)+1+f(n)$).

$$
\begin{aligned}
f(n) &= \frac{1}{2}(f(n-1)+1) + \frac{1}{2}(f(n-1)+1+f(n)) \\
2f(n) &= f(n-1)+1+f(n-1)+1+f(n) \\
f(n) &= 2f(n-1)+2 \\
f(n) &= 2*(2(f(n-2)+2)+2 \\
&\ldots \\
f(n) &= 2^{n-1}*f(1)+2^n-2 \\
f(n) &= 2^{n+1}-2
\end{aligned}
$$

Note: Kind of related to *negative binomial distribution*