Probability

https://cs.pomona.edu/classes/cs140/

Outline

Topics and Learning Objectives

- Review probability concepts
- Discuss linearity of expectations
- Discuss indicator variables

Exercise

- Linearity of expectations

Extra Resources

- https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectations

Linearity of expectation is the property that the
expected value of the sum of random variables is equal to the
sum of their individual expected values,
regardless of whether they are independent.

The expected value of a random variable is essentially
a weighted average of possible outcomes.

Definitions

- Space of all possible outcomes is Ω
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
- Sum of the probability of all outcomes is 1 ($\sum p_i = 1$)
- Let X be our random variable for the value of some outcome
  - It is a mapping (function) from Ω to a real-value ($\mathbb{R}$)
- The expected value of X (E[X]) is a weighted sum of the outcomes
- Let x_i be the value of a single possible outcome

$$E[X] = \sum_{i=1}^{|\Omega|} x_i \cdot p_i$$

Example

What is the expected value for a single die roll?

- Space of all possible outcomes is Ω
  - {1, 2, 3, 4, 5, 6}
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
  - $p_i = \frac{1}{6}$ for all outcomes
- Sum of the probability of <u>all</u> outcomes is 1 ($\sum p_i = 1$)
  - $\sum_{i=1}^{6} p_i = 6 \cdot \frac{1}{6} = 1$
- Let X be our random variable for the value of some outcome
  - X can be any of the values in Ω
- The expected value of X (E[X]) is a weighted sum of the outcomes

$$E[X] = \sum_{i=1}^{|\Omega|} x_i \cdot p_i$$

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

Example

What is the expected sum for rolling two dice?

- Space of all possible outcomes is Ω
  - {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
  - p_i depends on how many times each outcome can occur
- Sum of the probability of all outcomes is 1 ($\sum p_i = 1$)
  - $\sum_{i=2}^{12} p_i = 1$
- Let X be our random variable for the value of some outcome
  - X can be any of the values in Ω
- The expected value of X (E[X]) is a weighted sum of the outcomes
  - $E[X] = 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + \cdots$

Sum of Rolling Two Dice

What is p_5?

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

$$E[\sum X_i]$$

$$= \sum_{i=1}^{6} \sum_{j=1}^{6} (i+j) * p_{i+j}$$
$$= 2(\tfrac{1}{36}) + 3(\tfrac{2}{36}) + \cdots + 7(\tfrac{6}{36}) + 12(\tfrac{1}{36})$$
$$= 7$$

$$\sum E[X_i]$$ This is easier to calculate

$$= \sum_{i=1}^{|\Omega|} x_i \cdot p_i + \sum_{i=1}^{|\Omega|} x_i \cdot p_i$$
$$= 3.5 + 3.5$$
$$= 7$$

Linearity of Expectations

Let X_1, X_2, …, X_n be random variables defined for the same space

$$\sum E[X_i] = E[\sum X_i]$$

Linearity of Expectations

Let X_1, X_2, …, X_n be random variables defined for the same space

$$\sum E[X_i] = E[\sum X_i]$$

You are going to flip 10 coins. If you end up with x heads you will be paid \$1 * x. What is your expected payout?

Example

Expected payout

- Space of all possible outcomes is *Ω*
  - {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
  - p_i depends on how many ways you can get the outcome
- Sum of the probability of all outcomes is 1 ($\sum p_i = 1$)
  - $\sum_{i=0}^{10} p_i = 1$
- Let *X* be our random variable for the value of some outcome
  - X can be any of the values in *Ω*
- The expected value of *X* (E[X]) is a weighted sum of the outcomes
  - $E[X] = 0\left(\frac{1}{1024}\right) + 1\left(\frac{10}{1024}\right) + \cdots$

Expected Payout

$$E[\sum X_i] = \sum E[X_i]$$

You are going to flip 10 coins. If you end up with x heads you will be paid $1 * x. What is your expected payout?

- You might be tempted to do the following:

$$E[X] = \sum_{i=1}^{|\Omega|} x_i \cdot p_i = 0 \cdot p_0 + 1 \cdot p_1 + \cdots + 10 \cdot p_{10}$$

- But we can use linearity of expectations to make the problem easier

Expected Payout

$$E[\sum X_i] = \sum E[X_i]$$

You are going to flip 10 coins. If you end up with x heads you will be paid $1 * x. What is your expected payout?

- Instead, we can treat X like a sum of random variables

$$X = X_1 + X_2 + \cdots + X_{10}$$

- Now we just find the expected value of X_i

$$E[X_i] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

Expected Payout

$$E[\sum X_i] = \sum E[X_i]$$

You are going to flip 10 coins. If you end up with x heads you will be paid $1 * x. What is your expected payout?

- Now, the expected value of X is the sum of the expected values of X_i

$$E[X] = E[X_1 + X_2 + \cdots + X_{10}] = 10 \cdot \frac{1}{2} = 5$$

- So, the expected payout is $5

Trick question for the day

The expected value for the amount of rain on Saturday and Sunday is 2 inches and 3 inches, respectively. There is a 50% chance of rain on Saturday. If it rains on Saturday, there is a 75% chance of rain on Sunday, but if it does not rain on Saturday, then there is only a 50% chance of rain on Sunday.

What is the expected value (in inches) for the total amount of rain over the weekend?

Indicator Variables

An indicator variable is a random variable that takes the value 1 for some desired outcome, and the value 0 for all other outcomes.

This technique is useful when the random variable is counting the number of occurrences of simple events.

This will come in handy for our analysis of **Quicksort**.

What is the expected number of 5s when you roll one die?

- Space of all possible outcomes is Ω
  - {0, 1}
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
  - $p_0 = \frac{5}{6}$ and $p_1 = \frac{1}{6}$
- Sum of the probability of <u>all</u> outcomes is 1 ($\sum p_i = 1$)
  - $\sum_{i=0} p_i = p_0 + p_1 = 1$
- Let X be our random variable for the value of some outcome
  - X can be any of the values in Ω
- The expected value of X (E[X]) is a weighted sum of the outcomes
  - $E[X] = 0\left(\frac{5}{6}\right) + 1\left(\frac{1}{6}\right) = \frac{1}{6}$

Exercise

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Did a roll have a sum of exactly 5 or 6?

- Space of all possible outcomes is Ω
- Probability of an individual outcome is p_i (all values of $p_i \geq 0$)
- Sum of the probability of all outcomes is 1 ($\sum p_i = 1$)
- Let X be our random variable for the value of some outcome
- The expected value of X (E[X]) is a weighted sum of the outcomes

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Distinct Colors Problem

A box contains a yellow ball, an orange ball, a green ball, and a blue ball.

You randomly select 4 balls from the box (with replacement).

What is the expected number of distinct ball colors that you will select?

Distinct Colors Problem

A box contains a yellow ball, an orange ball, a green ball, and a blue ball.

You randomly select 4 balls from the box (with replacement).

What is the expected number of distinct ball colors that you will select?

- You could do this by directly computing the probabilities.
- For example, what is the probability that 1 color is selected?

p(select 1 color) = # of ways to do this / # of total possibilities
p(select 1 color) = 4 / (4*4*4*4)

- What is the probability that 2 colors are selected?
- This gets pretty difficult, even for this simple case.

Distinct Colors Problem

A box contains a yellow ball, an orange ball, a green ball, and a blue ball.

You randomly select 4 balls from the box (with replacement).

$\boxed{\text{What is the expected number of \underline{distinct} ball colors that you will select?}}$

- What does your intuition tell you is a *reasonable* number?
  - Exactly 1
  - Between 1 and 2
  - Exactly 2
  - Between 2 and 3
  - Exactly 3
  - Between 3 and 4
  - Exactly 4

```python
#!/usr/bin/env python3

from argparse import ArgumentParser
from random import choices


def run_trial(colors, draw):
    """Run a single trial of drawing balls of different colors

    Choose "draw" number of balls for the set of
     colors. Convert this to a set to eliminate
     duplicates, and then take the length of the
     set.

    Args:
        colors : a list of different colors (or numbers)
        draw   : the number of balls to draw from the list

    Return:
        return the number of distinct colors
    """
    return len(set(choices(colors, k=draw)))


if __name__ == "__main__":
    argument_parser = ArgumentParser(
        description="Run an experiment to count the number of distinct colors drawn from a box."
    )
    argument_parser.add_argument("--num_trials", type=int, default=1000)
    argument_parser.add_argument("--num_colors", type=int, default=4)
    argument_parser.add_argument("--draw_count", type=int, default=4)

    args = argument_parser.parse_args()

    colors = list(range(args.num_colors))
    average_distinct_colors = (
        sum(run_trial(colors, args.draw_count) for _ in range(args.num_trials))
        / args.num_trials
    )

    print("Average distinct colors:", average_distinct_colors)
```

```
❯ ./distinct_colors.py
Average distinct colors: 2.72
```

Number of distinct colors selected is denoted by $X_{\#}$

Let X_y denote the random variable that a yellow ball is selected

$$X_y = \begin{cases} 0 & \text{if } \textbf{no} \text{ yellow ball is selected} \\ 1 & \text{otherwise} \end{cases}$$

Indicator variable

Then

$$X_{\#} = X_y + X_o + X_g + X_b$$

and

$$E[X_{\#}] = E[X_y] + E[X_o] + E[X_g] + E[X_b]$$

Linearity of Expectations!

and since selecting each ball has the same probability

$$E[X_y] = E[X_o] = E[X_g] = E[X_b]$$

so

$$E[X_{\#}] = 4E[X_y] = 4E[X_o] = 4E[X_g] = 4E[X_b]$$

Now we just need to calculate $E[X_y]$

- What are the possible outcomes?

The expected value of an indicator variable is the same as its probability!

$$E[X_y] = \sum_{i=1}^{|\Omega|} x_i \cdot p_i = 1 \cdot p(X_y = 1) + 0 \cdot p(X_y = 0) = p(X_y = 1)$$

What is the probability that at least one yellow ball is selected?

Difficult to calculate. Must take into account 1 yellow, 2 yellow, 3, yellow, or 4 yellow

Instead, we will calculate the <u>complementary probability</u>.

$$p(X_y = 1) = 1 - p(X_y = 0)$$

What is the probability that **no** yellow ball is selected?

$$p(X_y = 0) = p(\bar{y})p(\bar{y})p(\bar{y})p(\bar{y}) = \left(\frac{3}{4}\right)^4$$

Now, what is the probability that at least one yellow ball is selected

$$p(X_y = 1) = 1 - p(X_y = 0) = 1 - \left(\frac{3}{4}\right)^4$$

Now, we have $\mathrm{E}[\mathrm{X}_y]$

$$E[X_y] = p(X_y = 1) = 1 - \left(\frac{3}{4}\right)^4$$

Now we can calculate the expected number of distinct colors

$$\mathrm{E}[X_\#] = 4 \cdot E[X_y] = 4 \cdot \left(1 - \left(\frac{3}{4}\right)^4\right) \sim 2.734$$