ON AVERAGE WHEN PEOPLE SAY AVERAGE THEY MEAN MEAN.

WHEN I SAY AVERAGE I MEAN MEDIocre.

I'M SOMEWHERE IN THE MIDDLE.
SUMMARY STATISTICS

- Motivating example: household income in the US

- **Approach 1**: do a census and obtain all the incomes
  - **Pros**: provides the most information, can perform any analysis
  - **Cons**: difficult/expensive to obtain, we may only be interested in a summary to begin with

- **Approach 2**: obtain a suitable summary of the incomes
SUMMARY STATISTICS

- Motivating example: household income in the US
- We want to provide some information about the incomes
- **Examples**: minimum, maximum, median, average, ...
- Each of them is useful, a combination of them even better
- Most of them are difficult to estimate (min, max, median)
- The average value can be estimated very well using a small number of samples
SUMMARY STATISTICS

- Let $X$ be a discrete random variable
  - Example: $X = \text{income of a random household}$
- One of the most important summary statistics for $X$ is its “average” value (the expectation or mean of $X$)
- Plan for today: definition and properties of expectation
- Later: how to estimate expectation by sampling
EXPECTATION

- **Definition:** Let $X$ be a discrete random variable. The expected value (or mean) of $X$ is

$$
\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot Pr(\{\omega\})
$$

- Expectation is a weighted average, where each outcome is weighted by its probability
**EXPECTATION**

- **Definition:** Let $X$ be a discrete random variable. The expected value (or expectation or mean) of $X$ is

$$
\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot Pr(\{\omega\})
$$

$X =$ number obtained when rolling a die

$$
\mathbb{E}(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{7}{2}
$$
EXAMPLE: SPINNER

- Spin the dial of the spinner. Let $Y$ be the number of the region where it stopped.
  - Range$(Y) = \{1, 2, 3, 4\}$
  - $\Pr(Y = 1) = \frac{1}{2}$, $\Pr(Y = 2) = \frac{1}{4}$, $\Pr(Y = 3) = \Pr(Y = 4) = \frac{1}{8}$
  - What value does $Y$ take on average?
EXAMPLE: SPINNER

- Suppose we spin the dial \( N \) times, where \( N \) is huge. Then we expect to see about \( \frac{N}{2} \) ones, \( \frac{N}{4} \) twos, \( \frac{N}{8} \) threes, and \( \frac{N}{8} \) fours.

- If we add them up and divide by \( N \), we get:
EXAMPLE: ROULETTE

- 38 slots: 18 black, 18 red, 2 green.

- If we bet $1 on red, we get $2 back if red comes up. What’s the expected value of our winnings?
By rearranging the terms in the definition, we obtain the following equivalent way to write the expectation:

\[ \mathbb{E}(X) = \sum_{a \in \text{range}(X)} a \cdot Pr(X = a) \]
EXAMPLE: ROULETTE

- If we bet $1 on red, we get $2 back if red comes up. What’s the expected value of our winnings?

- Let $X$ be the value of winnings

$$
\mathbb{E}(X) = \sum_{a \in \text{Range}(X)} a \cdot \Pr(X = a)
$$
The expectation of a continuous random variable is defined similarly to the discrete case, only that we replace the sum by an integral.

Let $X$ be a random variable with PDF $f$.

**discrete:** $\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x \cdot f(x)$

**continuous:** $\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$
EXAMPLE: UNIFORM RANDOM VARIABLE

- Let $X$ be a value chosen uniformly at random from the interval $[a, b]$
- The PDF is $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
Example: Darts

- Suppose we throw a dart at a circular board of radius r
- Let D be the distance from the center of the board to the hitting point of the dart
- The PDF is $f(x) = \frac{2x}{r^2}$ for $x \in [0, r]$
EXAMPLE: COIN TOSS

- Toss a coin that comes up Heads with probability $p$

- Let $X = \begin{cases} 
1 & \text{if Heads} \\
0 & \text{if Tails}
\end{cases}$
EXAMPLE: UNIFORM RANDOM VARIABLE

- Let X be a value chosen uniformly at random from the set 
  \( \{x_1, x_2, \ldots, x_n\} \)
Let $X$ be a non-zero random variable

True or false: $\mathbb{E}\left(\frac{1}{X}\right) = \frac{1}{\mathbb{E}(X)}$

(A) True  (B) False
ASK THE AUDIENCE

- **Experiment:** toss a fair coin until we obtain Heads
- Let $X$ be the number of tosses

The expectation of $X$ is

(A) 1.5  (B) 2  (C) 4  (D) infinite
ASK THE AUDIENCE

- **Experiment:** toss a fair coin until we obtain Heads
- Let $X$ be the number of tosses
LINEARITY OF EXPECTATION

- **Theorem:** For any random variables $X$ and $Y$
  \[ \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \]

- The theorem holds even if $X$ and $Y$ are dependent!

- More generally:
  \[ \mathbb{E}(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1\mathbb{E}(X_1) + c_2\mathbb{E}(X_2) + \cdots + c_n\mathbb{E}(X_n) \]
  for any constants $c_1, \ldots, c_n$ and random variables $X_1, \ldots, X_n$

\[ \mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b \quad \text{for any constants } a, b \text{ and r.v. } X \]
Sums of Random Variables

- **Experiment**: roll a fair die twice
- Let $X$ and $Y$ be the two numbers obtained
- Let $S$ be the sum of the two rolls, i.e., $S = X + Y$
- If $X$ and $Y$ are independent, we can find $\mathbb{E}(S)$ using the definition and a fair bit of work (try it as an exercise)
- It turns out that there is a much easier way:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \frac{7}{2} + \frac{7}{2} = 7$$
Infinite monkey theorem:

Given enough time, a hypothetical chimpanzee typing at random would, as part of its output, almost surely produce one of Shakespeare's plays (or any other text).

https://en.wikipedia.org/wiki/Infinite_monkey_theorem
Suppose that the monkey types on a 26-letter keyboard. Each letter is chosen independently and uniformly at random from the 26 letters. Suppose that the monkey types $n = 1,000,000$ letters. Let $X = \text{number of times the sequence } \text{"proof" } \text{appears.}$ What is $\mathbb{E}(X)$?
RANDOM PROOFS
FUNCTIONS OF RANDOM VARIABLES

- We saw earlier several examples of random variables that are functions of another random variable

- Examples:
  \[ Y = 3X + 1 \quad Y = X^2 \quad Y = |X| \]

- We also saw how the PDF of \( Y \) is related to the PDF of \( X \)

- This leads to the following result for computing expectation:

  \[
  \begin{align*}
  \text{X is discrete:} & \quad \mathbb{E}(g(X)) = \sum_{x \in \mathbb{R}} g(x) \cdot f_X(x) \\
  \text{X is continuous:} & \quad \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, dx
  \end{align*}
  \]
FUNCTIONS OF RANDOM VARIABLES

Let $X$ be a random variable that is uniform over $[0, 1]$.

\[
f_X(x) = \begin{cases} 
1 & \text{if } x \in [0,1] \\
0 & \text{otherwise} 
\end{cases}
\]

\[
\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) \, dx = \int_{0}^{1} x^2 \, dx = \frac{x^3}{3} \bigg|_{0}^{1} = \frac{1}{3}
\]