Overview

- Expected Value of a Random Variable
- When Expected Value Isn’t Enough
- Variance
Expected Value of a Random Variable

The **Expected Value** of discrete random variable $X$ taking values in $\{x_1, x_2, \cdots, x_n\}$ is:

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

Why is this of interest? One important example:

- Suppose $I$ is the indicator variable for event $E$ (so $I = 1$ is event $E$ occurs, $I = 0$ otherwise).
- Then $E[I] = 1 \times P(E) + 0 \times (1 - P(E)) = P(E)$.
- The expected value of $I$ is the probability that event $E$ occurs.
Expected Value of a Random Variable

Some more intuition:

- Suppose we repeat our experiment $N$ times. Let $\omega_k$ be the outcome of the $k$'th trial and $X_k = X(\omega_k)$.
- For $N$ large, think of $NP(X = x_i)$ as the number of times that $X_k = x_i$ as $k = 1, 2, \cdots, N$ (we’ll come back to this later).
- Then:

\[
\frac{1}{N} \sum_{k=1}^{N} X_k \approx \frac{1}{N} \left( x_1 NP(X = x_1) + x_2 NP(X = x_2) + \cdots + x_n NP(X = x_n) \right)
\]

\[
= \sum_{i=1}^{n} x_i P(X = x_i) = E[X]
\]

Example:

- Suppose each experiment is one pull at the lever of a slot machine and $X_k$ is the winnings for the $k$'th pull.
- Then $\sum_{k=1}^{N} X_k$ is the total winnings over $N$ goes and $\frac{1}{N} \sum_{k=1}^{N} X_k$ is the average winnings per go.
- But $\frac{1}{N} \sum_{k=1}^{N} X_k = E[X]$, so $E[X]$ is the average winnings per go taken over many trials.
Expected Value of a Random Variable

- Expected value is also called: mean, average, expectation, first moment.
- The sum is over all values of $X$.
- Values $x_i$ for which $P(X = x_i) = 0$ don’t contribute to the expected value.
- Values of $x_i$ with higher probability $P(X = x_i)$ contribute more.
Who cares?

Suppose we have to choose an action. Once we have chosen we get a “reward” with some probability. Which action should be choose?

- Often a reasonable answer is: the one with the best expected outcome

- Delayed gratification. Suppose I have to choose between two courses. One takes 1 year but with probability 0.5 will allow me to earn €5000 extra per year for next 5 years. The other is fast, takes 1 week, but with probability 0.05 earns an extra €5000 extra per year. Which should I choose?

- Option 1: cost is 1 year, say €10K earnings missed. Reward is €5000 for next 5 years with probability 0.5 and €0 with probability 0.5. Expected reward is $-10000 + 25000 \times 0.5 + 0 \times 0.5 = 2500$.

- Option 2: cost is 1 week, say €10K/52 earnings missed. Reward is €5000 over next 5 years with probability 0.05 and €0 with probability 0.95. Expected reward is $-10000/52 + 25000 \times 0.05 + 0 \times 0.95 \approx 1000$. 
Who cares?

- Suppose we play the lottery and pay €1 per play. There are two possible outcomes, win or lose. Random variable $X$ is $10^3$ if win, -1 (price of ticket) if lose. $P(X = 1M) = p = 1/10^6$, $P(X = -1) = 1 - p$. If we play the lottery many times, our average return is $E[X] = 10^3/10^6 - 1 \times (1 - 1/10^6) = -0.999$. If we don’t play the lottery then our average return is 0, but higher than −0.999.

- I run an investment bank. With probability 0.99 I make profit of €1000. With probability (1-0.99)=0.01 I lose €100 billion. My expected return is $0.99 \times 1000 - 0.01 \times 100 \times 10^9 = -9,999,010$. 

Lying With Statistics

The expected value is an example of a summary statistic, that tries to communicate information about a dataset as simply as possible. But great care is needed.

Example:
- School has 3 classes with 5, 10 and 150 students
- Randomly choose a class with equal probability
- Random variable $X$ is the size of the chosen class
- What is $E[X]$ ?

$$E[X] = 5 \times \frac{1}{3} + 10 \times \frac{1}{3} + 150 \times \frac{1}{3} = 165/3 = 55$$
Lying With Statistics

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a student with equal probability
- Random variable $Y$ is the size of the class that the student is in
- What is $E[Y]$?

$$E[Y] = 5 \times \frac{5}{165} + 10 \times \frac{10}{165} + 150 \times \frac{150}{165} \approx 137$$

$E[Y]$ is student perception of class size but $E[X]$ is what’s usually reported.
Appears in other ways e.g. if try to estimate interval between buses (also length time bias\(^1\)).

---

\(^1\)https://en.wikipedia.org/wiki/Length_time_bias
Expected Value of a Random Variable

\[ E[aX + b] = aE[X] + b. \]  Proof:

\[
E[aX + b] = \sum_x (ax + b)p(x)
\]

\[
= \sum_x axp(x) + \sum_x bp(x)
\]

\[
= a \sum_x xp(x) + b \sum_x p(x)
\]

\[ = aE[X] + b \]
Expected Value of a Random Variable

**Linearity:**
- Take two random variables $X$ and $Y$
- $E[X + Y] = E[X] + E[Y]$, where $a$, $b$ are real valued.
- **Proof:**

$$E[X + Y] = \sum_x \sum_y (x + y)P(X = x \text{ and } Y = y)$$

$$= \sum_x \sum_y xP(X = x \text{ and } Y = y) + \sum_y \sum_x yP(X = x \text{ and } Y = y)$$

$$\overset{(a)}{=} \sum_x xP(X = x) + \sum_y yP(Y = y)$$

$$= E[X] + E[Y]$$

\( (a) \) Remember marginal distributions:
$$\sum_y P(X = x \text{ and } Y = y) = P(X = x)$$

- More generally, $E[\sum_i X_i] = \sum_i E[X_i]$
Expected Value of a Random Variable

- Bernoulli random variable, $X \sim Ber(p)$:

  \[
  E[X] = p \\
  Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)
  \]

- Binomial random variable, $X \sim Bin(n, p)$. Sum of $n Ber(p)$ independent random variables so:

  \[
  E[X] = np \\
  Var(X) = np(1 - p)
  \]
Expected Value of a Random Variable

- Take two independent random variables $X$ and $Y$
- $E[XY] = E[X]E[Y]$
- Proof:

\[
E[XY] = \sum_x \sum_y xy P(X = x \text{ and } Y = y)
\]
\[
= \sum_x \sum_y xy P(X = x) P(Y = y)
\]
\[
= \sum_x x P(X = x) \sum_y y P(Y = y)
\]
\[
= E[X]E[Y]
\]
Expected Value of a Random Variable

- Expected value is the first moment of random variable $X$, $E[X] = \sum_{i=1}^{n} x_i p(x_i)$.
- $N$’th moment of $X$ is $E[X^N] = \sum_{i=1}^{n} x_i^N p(x_i)$, will see a use for this shortly.
When Expected Value Isn’t Enough

Game description:
- We have a fair coin
- We keep flipping (perhaps infinitely many times) until we reach the first tails
- Random variable $N =$ number of flips before first tails (so $N$ is the number of consecutive heads)
- You win $2^N$ euros at the end

How much would you pay to play?
- Random variable $X =$ your winnings
- $E[X] = \left(\frac{1}{2}\right)^1 \times 2^0 + \left(\frac{1}{2}\right)^2 \times 2^1 + \left(\frac{1}{2}\right)^2 \times 2^2 + \cdots$
- $E[X] = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1}2^i = \sum_{i=0}^{\infty} \frac{1}{2} = \infty$
- Pay €100K each time and play 10 times. Any takers?

---

2https://en.wikipedia.org/wiki/St._Petersburg_paradox
Demo
Gamblers Ruin

Roulette. 18 red, 18 black, 1 green (37 total).

- Bet on red, \( p = \frac{18}{37} \) to win €1 otherwise \( 1 - p \) you lose €1
- Bet €1
- If win then stop, if lose then double bet and repeat
- Random variable \( X \) is winnings on stopping

\[
E[Z] = p \times 1 + (1 - p)p \times (2 - 1) + (1 - p)^2 p \times (4 - 2 - 1) + \cdots
\]

\[
= \sum_{i=0}^{\infty} (1 - p)^i p(2^i - \sum_{j=1}^{i-1} 2^j) = 1
\]

- Expected winnings are \( > 0 \) so why don’t we play infinitely often?
- You have finite money! Usually also a max bet.

\(^3\text{https://en.wikipedia.org/wiki/Martingale\_\(\text{(betting\_system)}\)}}
Variance

- All have the same expected value, $E[X] = 3$
- But “spread” is different
- Variance is a statistic that quantifies “spread”
Variance

Let $X$ be a random variable with mean $\mu$. The variance of $X$ is $\text{Var}(X) = E[(X - \mu)^2]$.

- Discrete random variable taking values in $D = \{x_1, x_2, \ldots, x_n\}$.

$$\text{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

with $\mu = E[X] = \sum_{i=1}^{n} x_i p(x_i)$

- Example. Flip coin, $X = 1$ if heads, 0 otherwise.
  $$E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}.$$  
  $$\text{Var}(X) = (1 - \frac{1}{2})^2 \times \frac{1}{2} + (0 - \frac{1}{2})^2 \frac{1}{2} = \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4}.$$ 

- Variance is mean squared distance of $X$ from the mean $\mu$
- $\text{Var}(X) \geq 0$
- Standard deviation is square root of variance $\sqrt{\text{Var}(X)}$. 

Variance

Discrete random variable taking values in $D = \{x_1, x_2, \cdots, x_n\}$. Variance:

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

$$= \sum_{i=1}^{n} (x_i^2 - 2x_i \mu + \mu^2)p(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 p(x_i) - 2 \sum_{i=1}^{n} x_i p(x_i) \mu + \mu^2 \sum_{i=1}^{n} p(x_i)$$

$$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2$$

$$= E[X^2] - (E[X])^2$$
Variance

Var(aX + b) = a^2 Var(X). Proof:

\[ Var(aX + b) = E[(aX + b)^2] - E[aX + b]^2 \]
\[ = E[a^2 X^2 + 2abX + b^2] - (aE[X] + b)^2 \]
\[ = a^2 E[X^2] + 2abE[X] + b^2 - a^2 E[X]^2 - 2abE[X] - b^2 \]
\[ = a^2 E[X^2] - a^2 E[X]^2 \]
\[ = a^2 (E[X^2] - E[X]^2) = a^2 Var(X) \]

(recall \( E[aX + b] = aE[X] + b \)).
Variance

For independent random variables $X$ and $Y$ then

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. Proof.

$\text{Var}(X + Y) = E[(X + Y)^2] - E[X + Y]^2$

$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$


$= \text{Var}(X) + \text{Var}(Y)$

(recall $E[XY] = E[X]E[Y]$ when $X$ and $Y$ are independent)
Anscombe’s Quartet

The variance is another example of a summary statistic, this time one that indicates the spread in a data set. But great care is again needed.

All four datasets also have:
- $E[X] = 9$, $\text{Var}(X) = 11$
- $E[Y] \approx 7.50$, $\text{Var}(Y) \approx 4.12$
- Take home message: plot the data, don’t just rely on summary statistics.

source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet