Start of Lecture 19

Last time: Conditional Probability, Independence, Bayes' Theorem

Today: Random Variables, Expectation



Recall

Definition: The conditional probability of an event A given an event B is

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$



Recall

Definition: The conditional probability of an event A given an event B is

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

Note: Rearranging the definition gives us

$$P(A \cap B) = P(A|B)P(B).$$



Recall

Recall: In the COVID example, we had $A = \text{COVID}$ and $B = \text{Positive Test}$.

We computed

$$P(B) = P(B \cap A) + P(B \cap \overline{A}) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

.



Law of Total Probability

More generally, . . .

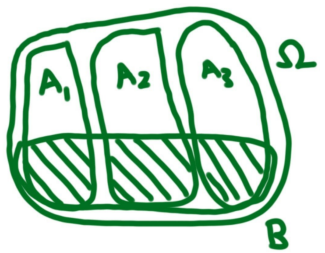


Law of Total Probability

More generally, . . .

Theorem: Let $A_1, \ldots, A_n$ be a partition of the sample space Ω. Then for any event B,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$



Law of Total Probability

More generally, . . .

Theorem: Let $A_1, \ldots, A_n$ be a partition of the sample space Ω. Then for any event B,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

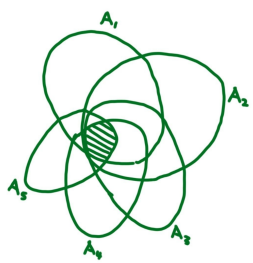
Proof: Since $A_1, \ldots, A_n$ is a partition of Ω, we have that $B \cap A_1, \ldots, B \cap A_n$ is a partition of B. Therefore,

$$\begin{aligned} P(B) &= \sum_{i=1}^{n} P(B \cap A_i) \\ &= \sum_{i=1}^{n} P(B|A_i)P(A_i). \end{aligned}$$



Multiplication Rule

We can apply $P(A \cap B) = P(B)P(A|B)$ many times to a sequence of events.

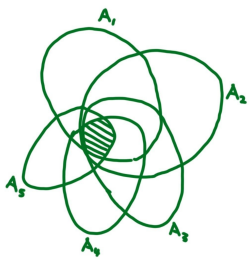


Multiplication Rule

We can apply $P(A \cap B) = P(B)P(A|B)$ many times to a sequence of events.

Theorem: For any events $A_1, \ldots, A_n$,

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = P(A_1) \\ &\cdot P(A_2 \mid A_1) \\ &\cdot P(A_3 \mid A_1 \cap A_2) \\ &\cdots \\ &\cdot P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$



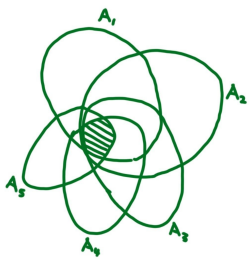
Multiplication Rule

We can apply $P(A \cap B) = P(B)P(A|B)$ many times to a sequence of events.

Theorem: For any events $A_1, \ldots, A_n$,

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = P(A_1) \\ &\cdot P(A_2 \mid A_1) \\ &\cdot P(A_3 \mid A_1 \cap A_2) \\ &\cdots \\ &\cdot P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

Example: A bin contains 4 red, 3 blue, and 2 green balls. I draw 3 balls uniformly at random without replacement. What is the probability of drawing red, then blue, then red?



Multiplication Rule

We can apply $P(A \cap B) = P(B)P(A|B)$ many times to a sequence of events.

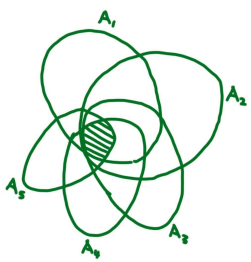
Theorem: For any events $A_1, \ldots, A_n$,

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = P(A_1) \\ &\cdot P(A_2 \mid A_1) \\ &\cdot P(A_3 \mid A_1 \cap A_2) \\ &\cdots \\ &\cdot P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

Example: A bin contains 4 red, 3 blue, and 2 green balls. I draw 3 balls uniformly at random without replacement. What is the probability of drawing red, then blue, then red?

Solution: By the multiplication rule,

$$P(\text{red, blue, red}) = \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{3}{7} = \frac{1}{14}.$$



Multiplication Rule

Corollary: If $A_1, \ldots, A_n$ are events such that each A_i is independent of $A_1 \cap \cdots \cap A_{i-1}$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n).$$



Multiplication Rule

Corollary: If $A_1, \ldots, A_n$ are events such that each A_i is independent of $A_1 \cap \cdots \cap A_{i-1}$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n).$$

Example: I flip a biased coin with probability 0.7 of heads ten times. What is the probability of getting the sequence *HHHHTTTTTT*? What is the probability of getting 4 heads and 6 tails in any order?



Multiplication Rule

Corollary: If $A_1, \ldots, A_n$ are events such that each A_i is independent of $A_1 \cap \cdots \cap A_{i-1}$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n).$$

Example: I flip a biased coin with probability 0.7 of heads ten times. What is the probability of getting the sequence $HHHHTTTTTT$? What is the probability of getting 4 heads and 6 tails in any order?

Solution: The flips are independent, so

$$P(HHHHTTTTTT) = (0.7)^4 \cdot (0.3)^6.$$



Multiplication Rule

Corollary: If $A_1, \ldots, A_n$ are events such that each A_i is independent of $A_1 \cap \cdots \cap A_{i-1}$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n).$$

Example: I flip a biased coin with probability 0.7 of heads ten times. What is the probability of getting the sequence $HHHHTTTTTT$? What is the probability of getting 4 heads and 6 tails in any order?

Solution: The flips are independent, so

$$P(HHHHTTTTTT) = (0.7)^4 \cdot (0.3)^6.$$

There are $\binom{10}{4}$ orders with 4 heads and 6 tails, so

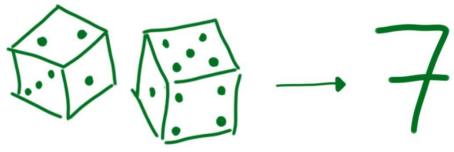
$$P(\text{4 heads, 6 tails}) = \binom{10}{4}(0.7)^4(0.3)^6.$$



Random Variables

Definition: For a probability space (Ω, P), an S-valued random variable X is a function

$$X : \Omega \to S.$$



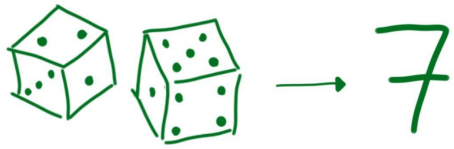
Random Variables

Definition: For a probability space (Ω, P), an S-valued random variable X is a function

$$X : \Omega \to S.$$

Examples: Roll two fair dice: $\Omega = [6] \times [6]$, $P(\omega) = 1/36$.

- **Sum:** $X : \Omega \to \mathbb{Z}$ where $X((a, b)) = a + b$.



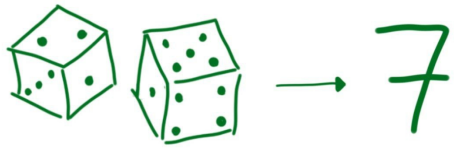
Random Variables

Definition: For a probability space (Ω, P), an S-valued random variable X is a function

$$X : \Omega \to S.$$

Examples: Roll two fair dice: $\Omega = [6] \times [6]$, $P(\omega) = 1/36$.

- **Sum:** $X : \Omega \to \mathbb{Z}$ where $X((a, b)) = a + b$.
- **Maximum:** $Y : \Omega \to \mathbb{Z}$ where $Y((a, b)) = \max\{a, b\}$.



Random Variables

Definition: For a probability space (Ω, P), an S-valued random variable X is a function

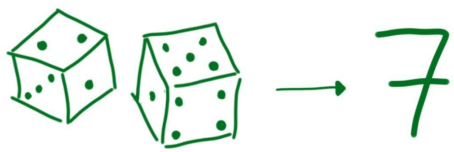
$$X : \Omega \to S.$$

Examples: Roll two fair dice: $\Omega = [6] \times [6]$, $P(\omega) = 1/36$.

- **Sum:** $X : \Omega \to \mathbb{Z}$ where $X((a, b)) = a + b$.
- **Maximum:** $Y : \Omega \to \mathbb{Z}$ where $Y((a, b)) = \max\{a, b\}$.
- **Rolling a six:** $Z : \Omega \to \{0, 1\}$, where

$$Z((a, b)) = \begin{cases} 1, & a = 6 \text{ or } b = 6, \\ 0, & \text{otherwise,} \end{cases}$$

Definition: An indicator random variable is a random variable $X : \Omega \to \{0, 1\}$.



Random Variables: Notation

Notation: We use conditions on random variables to define events:

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\}.$$



Random Variables: Notation

Notation: We use conditions on random variables to define events:

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\}.$$

Example: If X is the sum of two dice, then

$$\{X = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}.$$



Random Variables: Notation

Notation: We use conditions on random variables to define events:

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\}.$$

Example: If X is the sum of two dice, then

$$\{X = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}.$$

We can then write:

- $P(X = 7)$: the probability that the sum is 7.
- $P(Y \geq 3)$: the probability that the maximum roll is at least 3.



Expectation

We want to talk about "average" outcomes. . .



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example: I roll a fair die. Let $X(\omega) = \omega$ be the number rolled. What is $\mathbb{E}[X]$?



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example: I roll a fair die. Let $X(\omega) = \omega$ be the number rolled. What is $\mathbb{E}[X]$? 3.5



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example: I roll a fair die. Let $X(\omega) = \omega$ be the number rolled. What is $\mathbb{E}[X]$? 3.5

Example: I flip three fair coins. Let Y be the number of heads. What is $\mathbb{E}[Y]$?



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example: I roll a fair die. Let $X(\omega) = \omega$ be the number rolled. What is $\mathbb{E}[X]$? 3.5

Example: I flip three fair coins. Let Y be the number of heads. What is $\mathbb{E}[Y]$? 1.5



Expectation

We want to talk about "average" outcomes...

Definition: Let X be a real-valued random variable. The expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example: I roll a fair die. Let $X(\omega) = \omega$ be the number rolled. What is $\mathbb{E}[X]$? 3.5

Example: I flip three fair coins. Let Y be the number of heads. What is $\mathbb{E}[Y]$? 1.5

Fact: We can also sum over the values x that X takes:

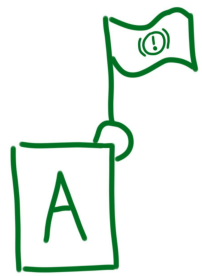
$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x).$$



Expectation of Indicator Variables

Fact: For an indicator random variable A,

$$\mathbb{E}[A] = P(A = 1).$$



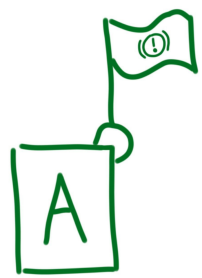
Expectation of Indicator Variables

Fact: For an indicator random variable A,

$$\mathbb{E}[A] = P(A = 1).$$

Proof: Since A only takes the values 0 and 1,

$$\mathbb{E}[A] = 0 \cdot P(A = 0) + 1 \cdot P(A = 1) = P(A = 1).$$



Expectation of Indicator Variables

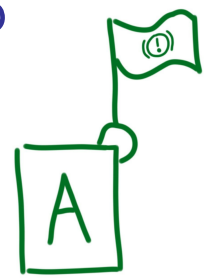
Fact: For an indicator random variable A,

$$\mathbb{E}[A] = P(A = 1).$$

Proof: Since A only takes the values 0 and 1,

$$\mathbb{E}[A] = 0 \cdot P(A = 0) + 1 \cdot P(A = 1) = P(A = 1).$$

Example: I flip four fair coins. Let $A = 1$ if they all come up heads, and $A = 0$ otherwise. What is $\mathbb{E}[A]$?



Expectation of Indicator Variables

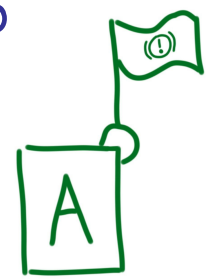
Fact: For an indicator random variable A,

$$\mathbb{E}[A] = P(A = 1).$$

Proof: Since A only takes the values 0 and 1,

$$\mathbb{E}[A] = 0 \cdot P(A = 0) + 1 \cdot P(A = 1) = P(A = 1).$$

Example: I flip four fair coins. Let $A = 1$ if they all come up heads, and $A = 0$ otherwise. What is $\mathbb{E}[A]$? $\frac{1}{16}$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) =$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$.



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\mathbb{E}[X] = 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \cdots$$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\begin{aligned}\mathbb{E}[X] &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \cdots \\ &= p + 2(1-p)p + 3(1-p)^2 p + \cdots\end{aligned}$$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\begin{aligned}\mathbb{E}[X] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots \\ &= p + 2(1-p)p + 3(1-p)^2 p + \cdots \\ &= p + (1-p)\left(2p + 3(1-p)p + 4(1-p)^2 p + \cdots\right)\end{aligned}$$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\begin{aligned}
\mathbb{E}[X] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots \\
&= p + 2(1-p)p + 3(1-p)^2 p + \cdots \\
&= p + (1-p)\left(2p + 3(1-p)p + 4(1-p)^2 p + \cdots\right) \\
&= p + (1-p)\begin{pmatrix} p + 2(1-p)p + 3(1-p)^2 p + \cdots \\ + p + (1-p)p + (1-p)^2 p + \cdots \end{pmatrix}
\end{aligned}$$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\begin{aligned}
\mathbb{E}[X] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots \\
&= p + 2(1-p)p + 3(1-p)^2 p + \cdots \\
&= p + (1-p)\left(2p + 3(1-p)p + 4(1-p)^2 p + \cdots\right) \\
&= p + (1-p)\begin{pmatrix} p + 2(1-p)p + 3(1-p)^2 p + \cdots \\ + p + \ (1-p)p + \ (1-p)^2 p + \cdots \end{pmatrix} \\
&= p + (1-p)\left(\mathbb{E}[X] + 1\right).
\end{aligned}$$



Expectation: Duolingo Streaks

Exercise: Each day, I independently forget my Duolingo lesson with probability p, where $0 < p \leq 1$. Let X count the days until I first forget, including that day.

Write R for remembering and F for forgetting:

$$\Omega = \{F, RF, RRF, RRRF, \ldots\}, \qquad \text{e.g. } X(RRF) = 3.$$

What is $\mathbb{E}[X]$?

Solution: First, note that $P(X = n) = (1-p)^{n-1}p$. So,

$$\begin{aligned}
\mathbb{E}[X] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots \\
&= p + 2(1-p)p + 3(1-p)^2 p + \cdots \\
&= p + (1-p)\left(2p + 3(1-p)p + 4(1-p)^2 p + \cdots\right) \\
&= p + (1-p)\begin{pmatrix} p + 2(1-p)p + 3(1-p)^2 p + \cdots \\ + p + (1-p)p + (1-p)^2 p + \cdots \end{pmatrix} \\
&= p + (1-p)\left(\mathbb{E}[X] + 1\right).
\end{aligned}$$



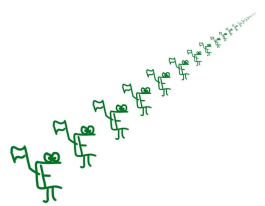
Solving for $\mathbb{E}[X]$ gives $\mathbb{E}[X] = \frac{1}{p}$.

This is how far we got in class.

Linearity of Expectation

Theorem: For random variables X_1, X_2 with finite expectations,

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$



Linearity of Expectation

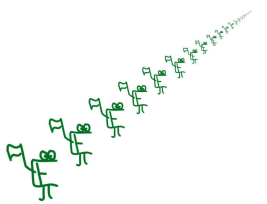
Theorem: For random variables X_1, X_2 with finite expectations,

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

More generally, for random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

The outcomes of the random variables need not be independent!



Linearity of Expectation

Theorem: For random variables X_1, X_2 with finite expectations,

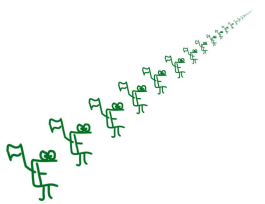
$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

More generally, for random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

The outcomes of the random variables need not be independent!

Example: I roll two fair dice. Let X_1 and X_2 be their values. What is $\mathbb{E}[X_1 + X_2]$?



Linearity of Expectation

Theorem: For random variables X_1, X_2 with finite expectations,

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

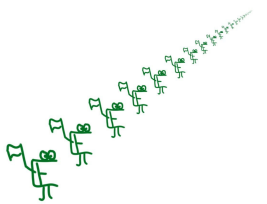
More generally, for random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

The outcomes of the random variables need not be independent!

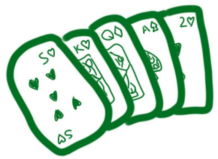
Example: I roll two fair dice. Let X_1 and X_2 be their values. What is $\mathbb{E}[X_1 + X_2]$?

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7.$$



Exercise: Shuffling a Deck

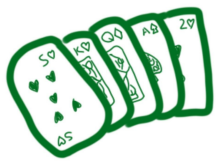
Exercise: I shuffle a deck of 52 cards uniformly at random. How many cards do we expect to end up in their original positions?



Exercise: Shuffling a Deck

Exercise: I shuffle a deck of 52 cards uniformly at random. How many cards do we expect to end up in their original positions?

Solution: Let $X_i = 1$ if card i stays in its original position, and $X_i = 0$ otherwise. Define the total $X := \sum_{i=1}^{52} X_i$.



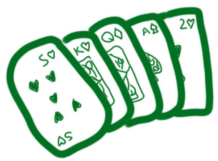
Exercise: Shuffling a Deck

Exercise: I shuffle a deck of 52 cards uniformly at random. How many cards do we expect to end up in their original positions?

Solution: Let $X_i = 1$ if card i stays in its original position, and $X_i = 0$ otherwise. Define the total $X := \sum_{i=1}^{52} X_i$.

Each card is equally likely to land in any of the 52 positions, so

$$\mathbb{E}[X_i] = P(X_i = 1) = \frac{1}{52}.$$



Exercise: Shuffling a Deck

Exercise: I shuffle a deck of 52 cards uniformly at random. How many cards do we expect to end up in their original positions?

Solution: Let $X_i = 1$ if card i stays in its original position, and $X_i = 0$ otherwise. Define the total $X := \sum_{i=1}^{52} X_i$.

Each card is equally likely to land in any of the 52 positions, so

$$\mathbb{E}[X_i] = P(X_i = 1) = \frac{1}{52}.$$

By linearity of expectation,

$$\mathbb{E}[X] = \sum_{i=1}^{52} \mathbb{E}[X_i] = 52 \cdot \frac{1}{52} = 1.$$

The same answer holds for a deck of any size.

