CSE 312

Foundations of Computing II

Lecture 7: More on independence; start random variables

Announcement: Concept check break this weekend!

Agenda

- Recap
- Sometimes Independence Occurs for Nonobvious Reasons
- Independence As An Assumption
- Conditional Independence
- New Topic: Random Variables

Bayes Theorem with Law of Total Probability

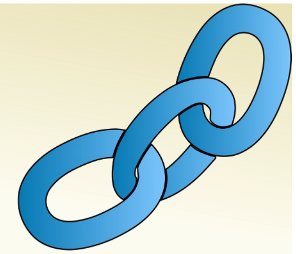
Bayes Theorem with LTP: Let $E_1, E_2, \ldots, E_n$ be a partition of the sample space, and F, G events. Then,

$$P(G|F) = \frac{P(F|G)P(G)}{P(F)} = \frac{P(F|G)P(G)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

Simple Partition: In particular, if E is an event with non-zero probability, then

$$P(G|F) = \frac{P(F|G)P(G)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$

Chain Rule



$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} \quad \Rightarrow \quad \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

Theorem. (Chain Rule) For events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$,

$$\mathbb{P}(\mathcal{A}_1 \cap \cdots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2)$$

$$\cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{n-1})$$

An easy way to remember: We have n events and we can evaluate their probabilities sequentially, conditioning on the occurrence of previous events.

Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are (statistically) **independent** if

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}) \cdot \mathbb{P}(\mathcal{B}).$$

Alternatively,

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(B)$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

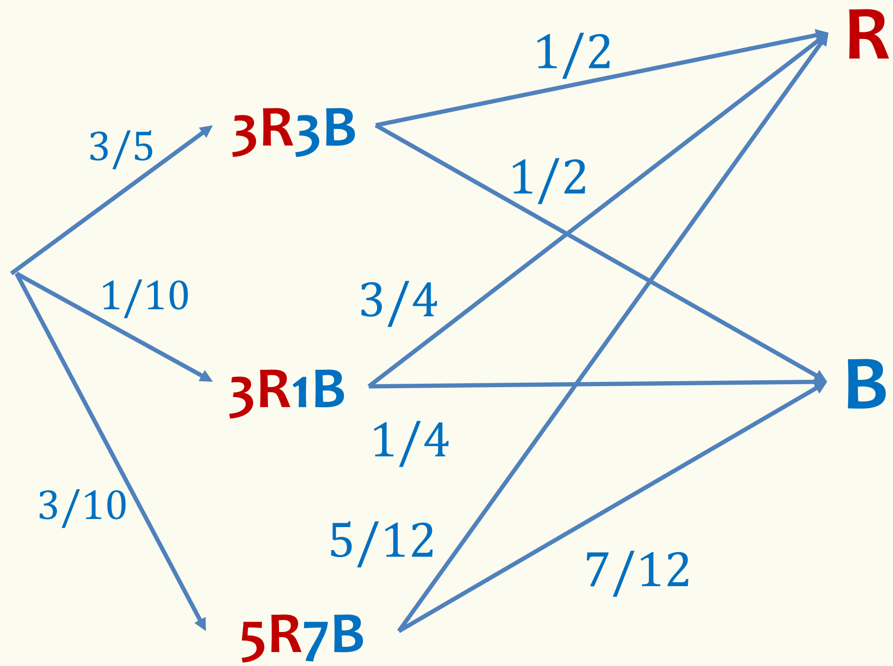
"The probability that $\mathcal{B}$ occurs after observing $\mathcal{A}$" -- Posterior
= "The probability that $\mathcal{B}$ occurs" -- Prior

Agenda

- Recap
- Sometimes Independence Isn't Obvious
- Independence As An Assumption
- Conditional Independence

- New Topic: Random Variables

Sequential Process



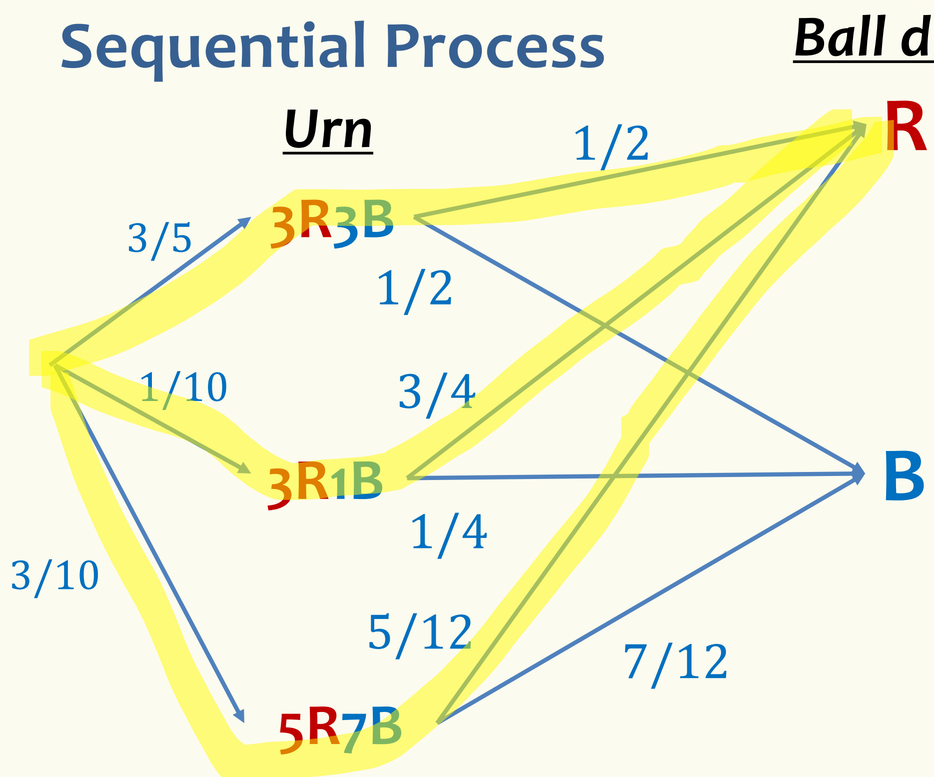
Setting: An urn contains:

- 3 **red** and 3 **blue** balls w/ probability 3/5
- 3 **red** and 1 **blue** balls w/ probability 1/10
- 5 **red** and 7 **blue** balls w/ probability 3/10

We draw a ball at random from the urn.

Are **R** and **3R3B** independent?

Sequential Process



Setting: An urn contains:

- 3 **red** and 3 **blue** balls w/ probability 3/5
- 3 **red** and 1 **blue** balls w/ probability 1/10
- 5 **red** and 7 **blue** balls w/ probability 3/10

We draw a ball at random from the urn.

Are **R** and **3R3B** independent?

$$P(\mathbf{R}) = \frac{3}{5} \times \frac{1}{2} + \frac{1}{10} \times \frac{3}{4} + \frac{3}{10} \times \frac{5}{12} = \frac{1}{2}$$

$$P(\mathbf{R} \mid \mathbf{3R3B}) = \frac{1}{2}$$

Independent! $P(\mathbf{R}) = P(\mathbf{R} \mid \mathbf{3R3B})$

Agenda

- Recap
- Sometimes Independence Occurs for Nonobvious Reasons
- Independence As An Assumption
- Conditional Independence

- New Topic: Random Variables

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

A : event that the main chute doesn't open $\mathbb{P}(A) = 0.02$
B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

  A : event that the main chute doesn't open $\mathbb{P}(A) = 0.02$
  B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

Assuming independence doesn't justify the assumption! Both chutes could fail because of the same rare event e.g., freezing rain.

Corollaries of independence of two events

- Example: A sky diver has two chutes

  A : event that the main chute doesn't open $\mathbb{P}(A) = 0.02$
  B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- What is the chance that **both open** assuming independence?

Agenda

- Recap
- Sometimes Independence Occurs for Nonobvious Reasons
- Independence As An Assumption
- Conditional Independence
- New Topic: Random Variables

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on $\mathcal{C}$ if $\mathbb{P}(\mathcal{C}) \neq 0$ and $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid \mathcal{C}) = \mathbb{P}(\mathcal{A} \mid \mathcal{C}) \cdot \mathbb{P}(\mathcal{B} \mid \mathcal{C})$.

Plain Independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}) \cdot \mathbb{P}(\mathcal{B}).$$

Equivalence:

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{B})$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on $\mathcal{C}$ if $\mathbb{P}(\mathcal{C}) \neq 0$ and $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid \mathcal{C}) = \mathbb{P}(\mathcal{A} \mid \mathcal{C}) \cdot \mathbb{P}(\mathcal{B} \mid \mathcal{C})$.

Equivalence:

- If $\mathbb{P}(\mathcal{A} \cap \mathcal{C}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B} | \mathcal{A} \cap \mathcal{C}) = \mathbb{P}(\mathcal{B} \mid \mathcal{C})$
- If $\mathbb{P}(\mathcal{B} \cap \mathcal{C}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A} | \mathcal{B} \cap \mathcal{C}) = \mathbb{P}(\mathcal{A} \mid \mathcal{C})$

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin twice independently. What is the probability both tosses heads?

$$\Pr(HH) = \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2)$$ LTP

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

$$\Pr(HH) = \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2)$$ LTP

$$= \Pr(H \mid C2)^2 \Pr(C1) + \Pr(H \mid C2)^2 \Pr(C2)$$ Conditional Independence

$$= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45$$

$$\Pr(H) = \Pr(H \mid C1)\Pr(C1) + \Pr(H \mid C2)\Pr(C2) = 0.6$$

New topic: random variables

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

Random Variables (Idea)

Often: We want to **capture quantitative properties** of the outcome of a random experiment, e.g.:

- *What is the total of two dice rolls?*
- *What is the number of coin tosses needed to see the first head?*
- *What is the number of heads among 5 coin tosses?*

Random Variables

Definition. A **random variable (RV)** for a probability space $(\Omega, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is called its range/support Ω_X

Example. Number of heads in 2 independent coin flips $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

RV Example

20 balls labeled 1, 2, … , 20 in an urn

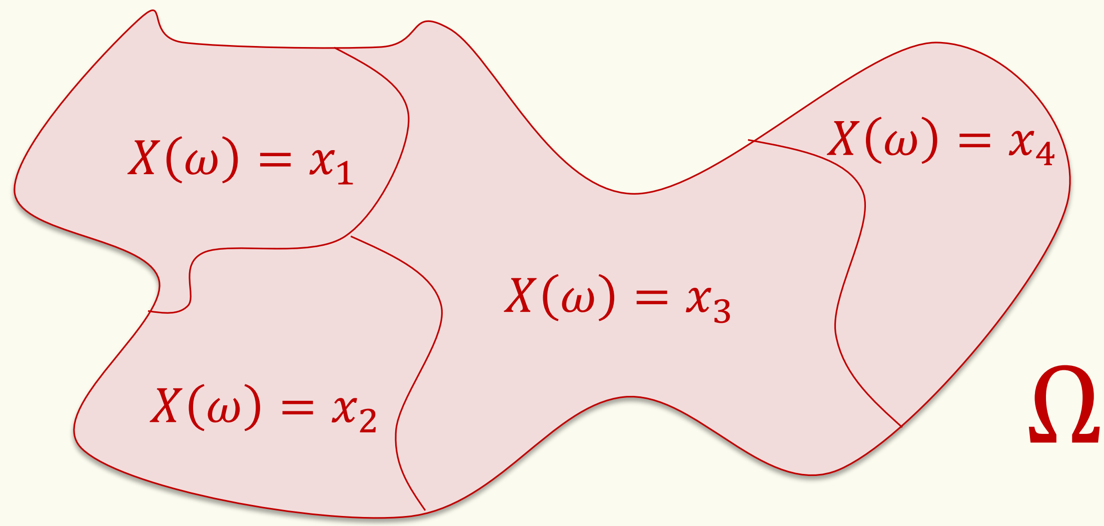
- Draw a subset of 3 uniformly at random
- Let $X = \text{maximum of the 3 numbers on the balls}$
  - Example: $X(2,7,5) = 7$
  - Example: $X(15,3,8) = 15$
- What is $|\Omega_X|$?

Agenda

- Random Variables
- Probability Mass Function (pmf)
- Cumulative Distribution Function (CDF)

Probability Mass Function (PMF)

Random variables partition the sample space.



Example: 20 balls labeled 1, 2, …, 20 in a bin

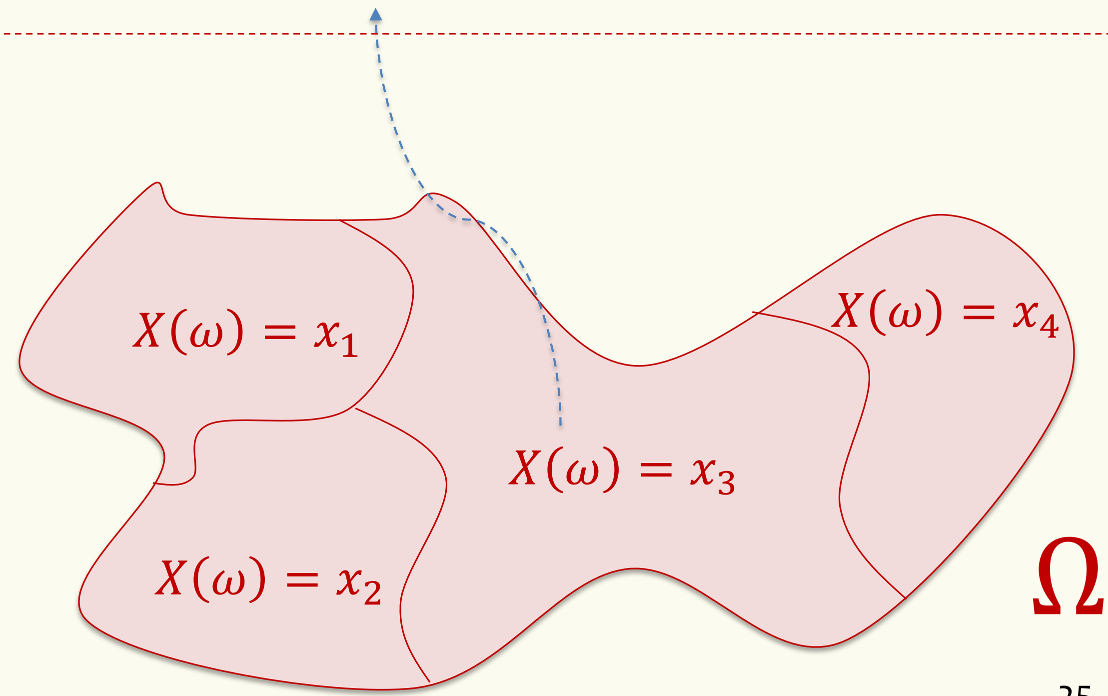
Draw a subset of 3 uniformly at random

Let $X =$ maximum of the 3 numbers on the balls

Probability Mass Function (PMF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

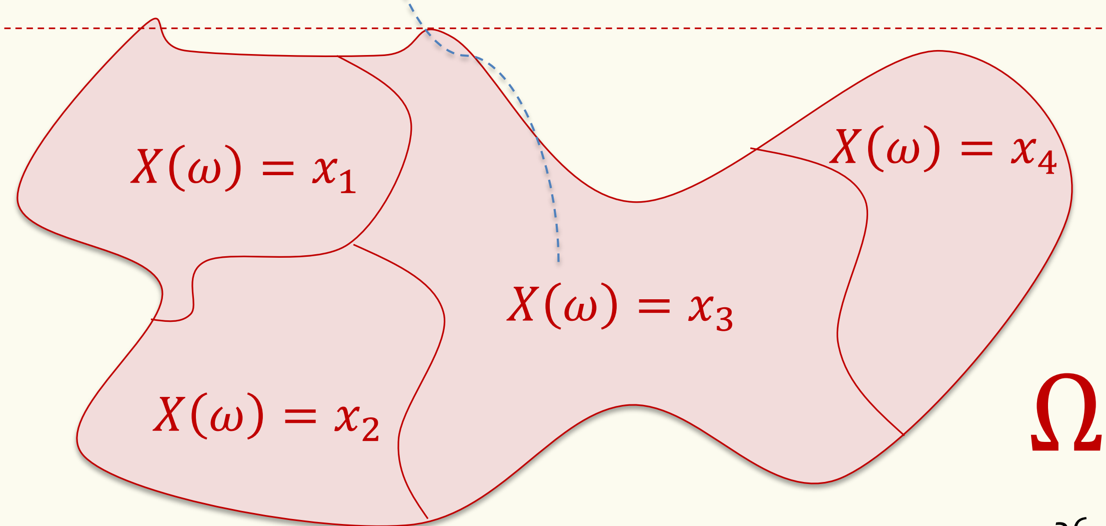


Probability Mass Function (PMF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

We write $\mathbb{P}(X = x) = \mathbb{P}(\{X = x\}) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$ where $\mathbb{P}(X = x)$ is the probability mass function (PMF) of X



$$\sum_{x \in \Omega_X} \mathbb{P}(X = x) = 1$$

Probability Mass Function (PMF)

Definition. For a RV $X\colon \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

We write $\mathbb{P}(X = x) = \mathbb{P}(\{X = x\}) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$ where $\mathbb{P}(X = x)$ is the probability mass function (PMF) of X

You also see this notation (e.g. in book):

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{p}_{\boldsymbol{X}}(\boldsymbol{x})$$

Probability Mass Function

Flipping two independent coins

$\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

$X = \text{number of heads in the two flips}$

$X(HH) = 2$ $\quad X(HT) = 1$ $\quad X(TH) = 1$ $\quad X(TT) = 0$

$\Omega_{\mathrm{X}} = \{0, 1, 2\}$

What is $Pr(X = k)$?

Probability Mass Function

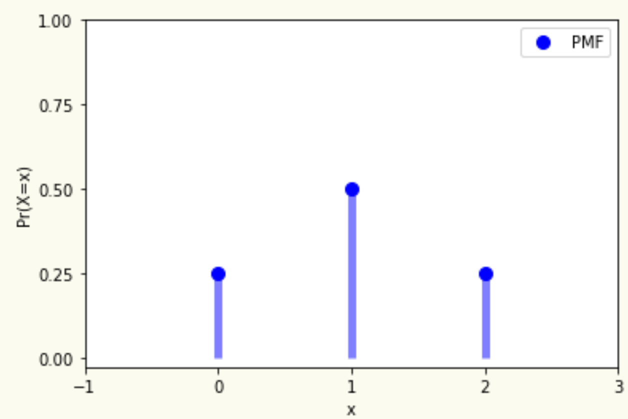
Flipping two independent coins $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

$X =$ number of heads in the two flips

$$X(HH) = 2 \qquad X(HT) = 1 \qquad X(TH) = 1 \qquad X(TT) = 0$$

$$\Omega_{\mathrm{X}} = \{0, 1, 2\}$$

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



RV Example

20 balls labeled 1, 2, …, 20 in an urn

- Draw a subset of 3 uniformly at random
- Let $X = \text{maximum of the 3 numbers on the balls}$

What is $Pr(X = 20)$?

Agenda

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

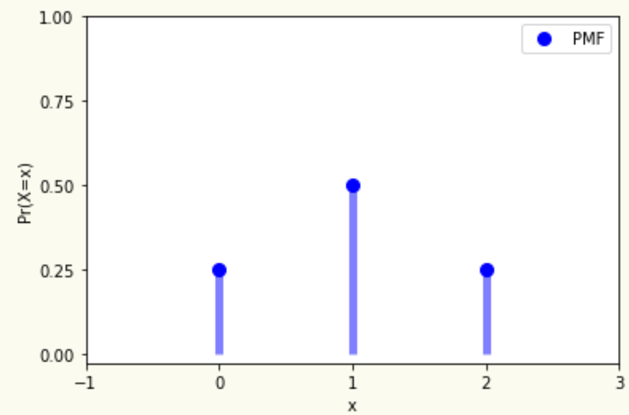
Cumulative Distribution Function (CDF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X specifies for any real number x, the probability that $X \leq x$.

$$\mathrm{F}_X(x) = \Pr(X \leq x)$$

Go back to 2 coin flips, where X is the number of heads

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



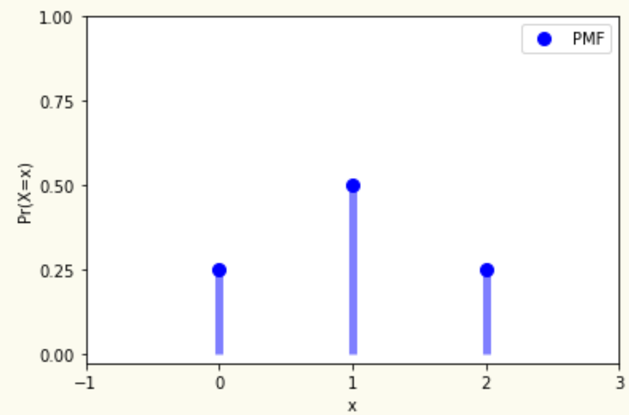
Cumulative Distribution Function (CDF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of where X specifies for any real number x, the probability that $X \leq x$.

$$\mathrm{F}_X(x) = \Pr(X \leq x)$$

Go back to 2 coin clips, where X is the number of heads

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \end{cases} \qquad F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \leq x < 1 \\ \frac{3}{4}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$



Example: Returning Homeworks

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\mathbf{Pr}(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	
1/6	1, 3, 2	
1/6	2, 1, 3	
1/6	2, 3, 1	
1/6	3, 1, 2	
1/6	3, 2, 1	


gifak-net
h_2
h_1
b
a
S=a*h₁
S=b*h₂

If time permits….

Often probability space (Ω, P) is given **implicitly** via sequential process

- *Experiment proceeds in n sequential steps, each step follows some local rules defined by the chain rule and independence*
- *Natural extension:* Allows for easy definition of experiments where $|\Omega| = \infty$

Example – Throwing A Die Repeatedly

Alice and Bob are playing the following game.

A 6-sided die is thrown, and each time it's thrown, regardless of the history, it is equally likely to show any of the six numbers

If it shows 1, 2 → Alice wins.
If it shows 3 → Bob wins.
Otherwise, play another round

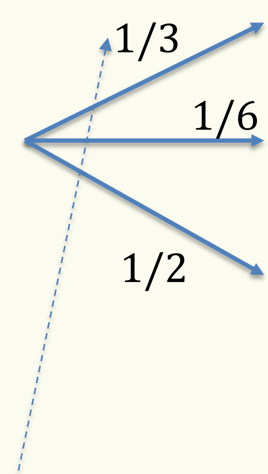
What is Pr(Alice wins on 1st round) =
Pr(Alice wins on 2st round) =
…
Pr(Alice wins on i^{th} round) = ?
Pr(Alice wins) = ?

Sequential Process – defined in terms of independence

A 6-sided die is thrown, and each time it's thrown, regardless of the history, it is equally likely to show any of the six numbers

Local Rules: In each round, toss a die

- If it shows 1,2 → Alice wins
- If it shows 3 → Bob wins
- Else, play another round



$\boldsymbol{i_{12}}$ = i th round toss 1 or 2

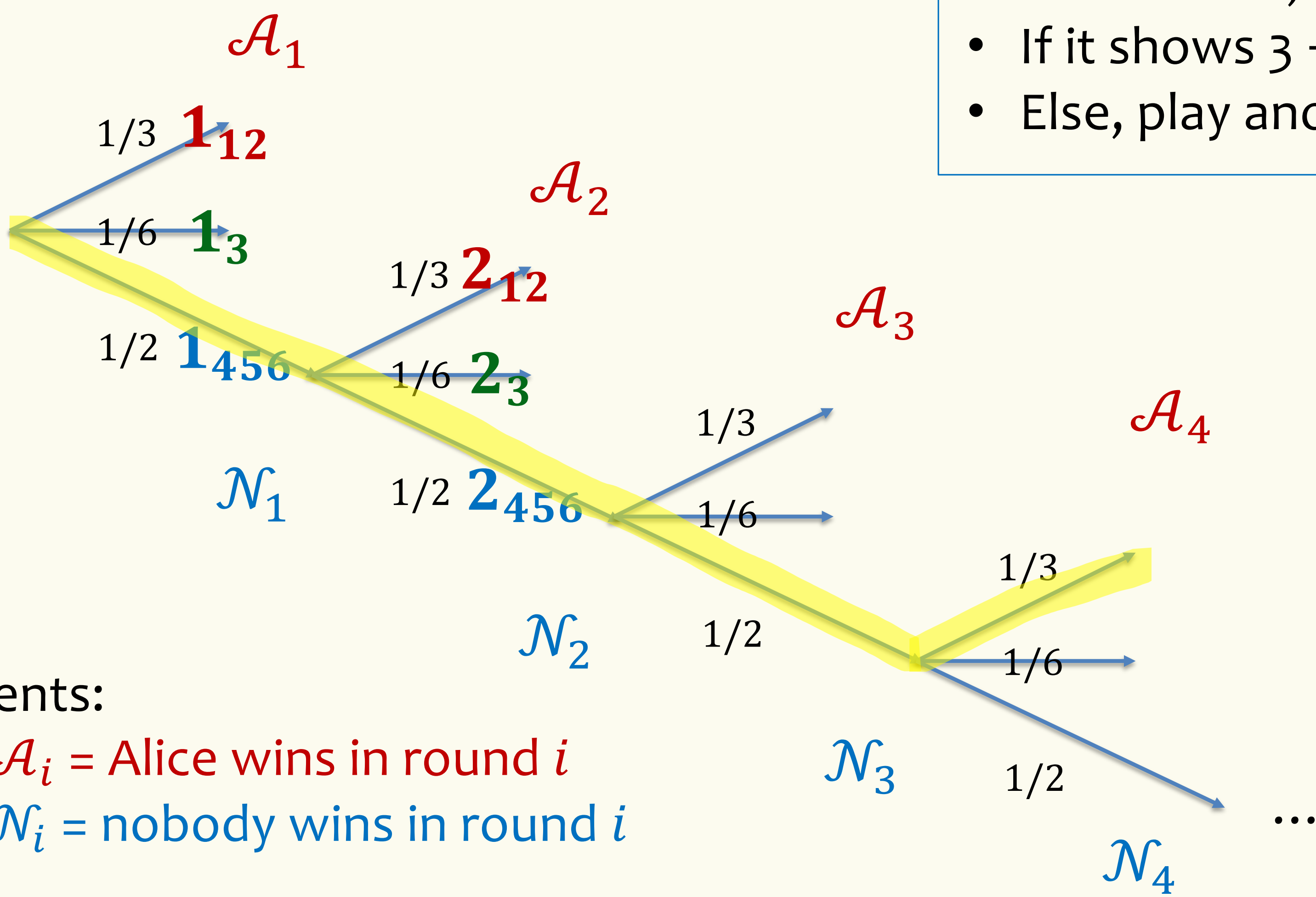
$\boldsymbol{i_{3}}$ = i th round toss 3

$\boldsymbol{i_{456}}$ = i th round toss 4 or 5 or 6

Pr (Alice wins on i -th round | nobody won in rounds $1..i$-1) = 1/3

Sequential Process – Example

Local Rules: In each round
- If it shows 1,2 → Alice wins
- If it shows 3 → Bob wins
- Else, play another round



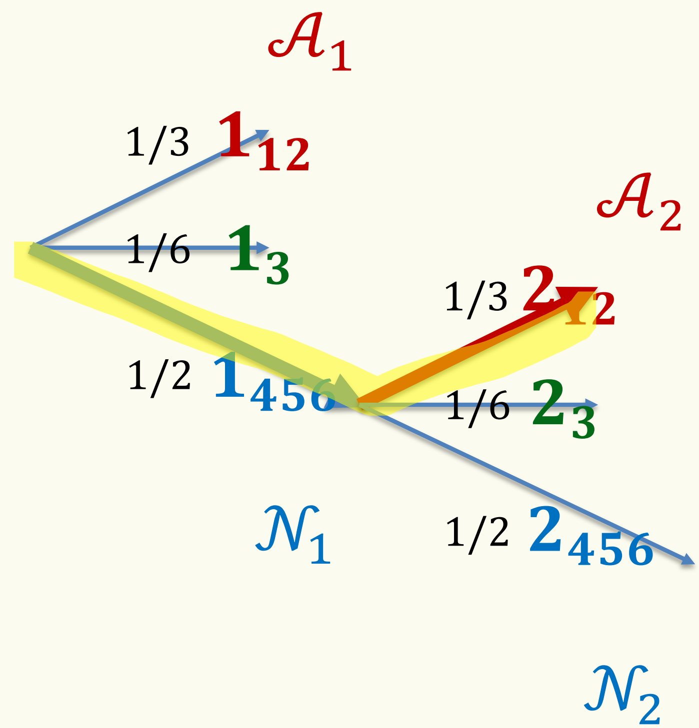
Events:
- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in round i

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds 1..i

$$\mathbb{P}(\mathcal{A}_2) = \mathbb{P}(\mathcal{N}_1 \cap \mathcal{A}_2)$$

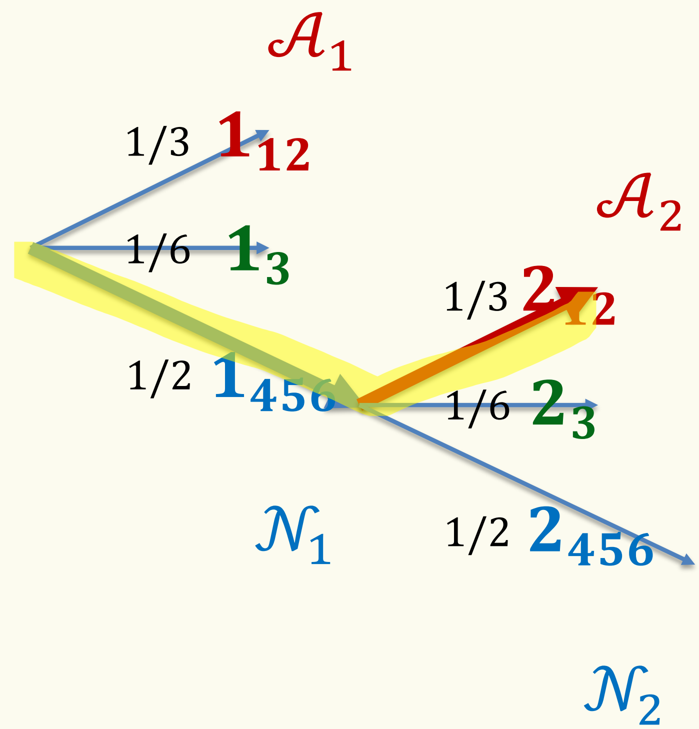


2nd roll indep of 1st roll

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds 1..i



$$\mathbb{P}(\mathcal{A}_2) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{A}_2)$$
$$= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{A}_2 | \mathcal{N}_1)$$
$$= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

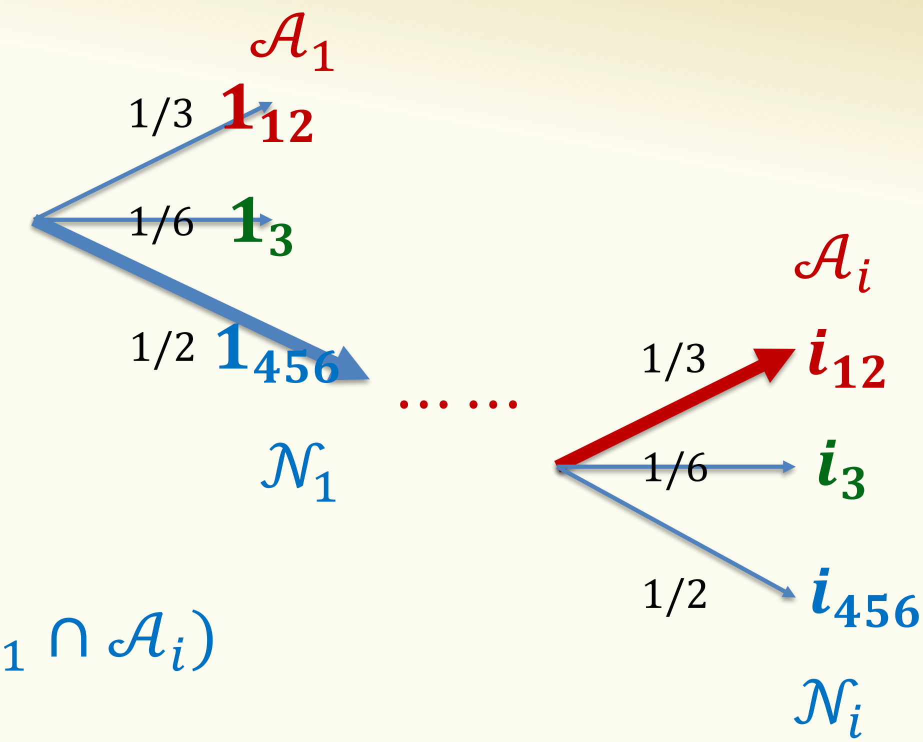
The event $\mathcal{A}_2$ implies $\mathcal{N}_1$, and this means that $\mathcal{A}_2 \cap \mathcal{N}_1 = \mathcal{A}_2$

2nd roll indep of 1st roll

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds $1..i$

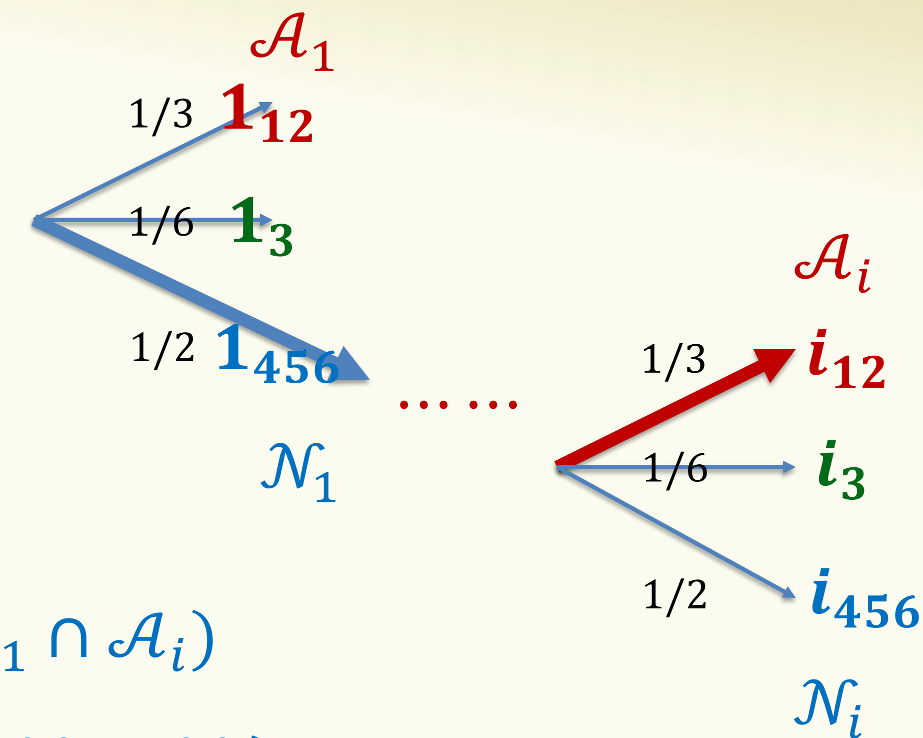


$$\mathbb{P}(\mathcal{A}_i) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1} \cap \mathcal{A}_i)$$

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in round $1..i$



$$\mathbb{P}(\mathcal{A}_i) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1} \cap \mathcal{A}_i)$$

$$= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_2 | \mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_3 | \mathcal{N}_1 \cap \mathcal{N}_2)$$

$$\cdots \times \mathcal{P}(\mathcal{N}_{i-1} | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1}) \times \mathcal{P}(\mathcal{A}_i | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1})$$

$$= \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$$

Sequential Process -- Example

$\mathcal{A}_i$ = Alice wins in round i $\quad \mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

What is the probability that Alice wins?

Sequential Process -- Example

$\mathcal{A}_i$ = Alice wins in round i $\mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

What is the probability that Alice wins?

$$\mathbb{P}(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots) = \sum_{i=1}^{\infty} \mathbb{P}(\mathcal{A}_i)$$

All $\mathcal{A}_i$'s are disjoint.

$$\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3} = \frac{1}{3} \times 2 = \frac{2}{3}$$

Fact. If $|x| < 1$, then $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$.