CSE 312

Foundations of Computing II

Lecture 8: Chain Rule and Independence

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Rachel Lin, Hunter Schafer

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun's and Anna Karlin's slides for 312 20su and 20au

Announcement

- Pset 3 is out today
- Pset 2 is due tomorrow

Last Class:

- Conditional Probability

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

- Bayes Theorem

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

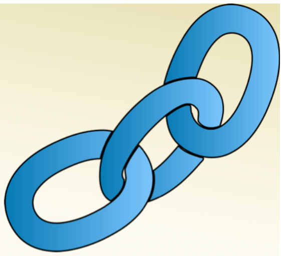
- Law of Total probability

$$\mathbb{P}(F) = \sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i) \quad E_i \text{ partition } \Omega$$

Today:

- Chain Rule
- Independence
- Sequential Process

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 $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$,

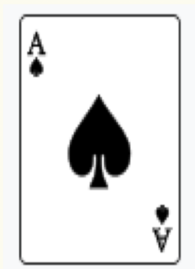
$$\mathbb{P}(\mathcal{A}_1 \cap \dots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2)$$

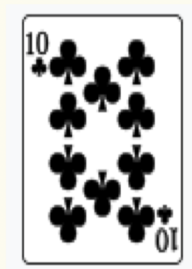
$$\cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_{n-1})$$

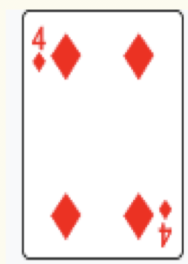
An easy way to remember: We have n tasks and we can do them sequentially, conditioning on the outcome of previous tasks

Chain Rule Example

Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards. (uniform probability space).

What is P (





) = P(**A** ∩ **B** ∩ **C**)?

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

$$\mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A \cap B)$$

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

Independence

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

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

Equivalence:

- 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

Example -- Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.

- A = {at most one T} = {HHH, HHT, HTH, THH}
- B = {at most 2 Heads}= $\{HHH\}^c$

Independent?

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

$$\frac{3}{8} \neq \frac{1}{2} \cdot \frac{7}{8}$$

Poll:

A. Yes, independent

B. No

pollev/rachel312

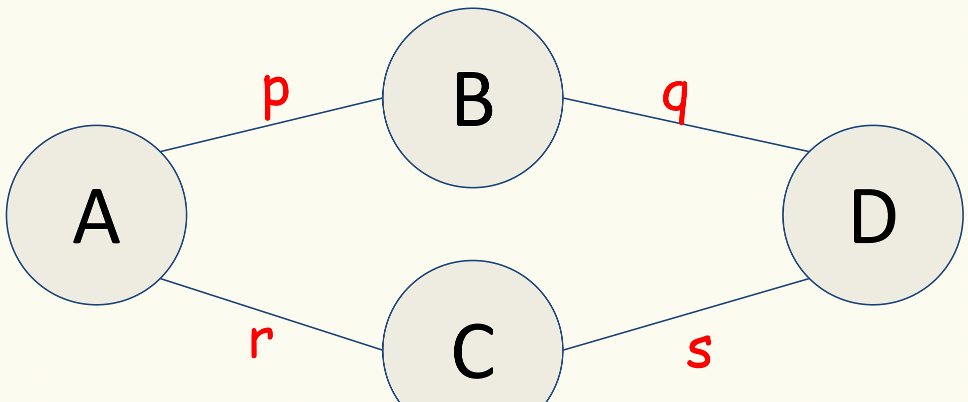
Example – Network Communication

Each link works with the probability given, **independently**.
What's the probability A and D can communicate?

$\mathbb{P}(AD) = \mathbb{P}(AB \cap BD \ or \ AC \cap CD)$

$= \mathbb{P}(AB \cap BD) + \mathbb{P}\ (AC \cap CD) - \mathbb{P}(AB \cap \mathrm{BD} \cap \mathrm{AC} \cap \mathrm{CD})$

$= pq + rs\ - pqrs$



$\mathbb{P}(AB \cap BD) = \mathbb{P}(AB) \cdot \mathbb{P}(BD) = \ pq$

$\mathbb{P}(AC \cap CD) = \mathbb{P}(AC) \cdot \mathbb{P}(CD) = rs$

$\mathbb{P}(AB \cap \mathrm{BD} \cap AC \cap CD) = \mathbb{P}(AC) \cdot \mathbb{P}(CD) \cdot \mathbb{P}(AB) \cdot \mathbb{P}(BD) = pqrs$

Example – Throwing Dies

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) = $\frac{1}{3}$
Pr(Alice wins on 2st round) = $\frac{1}{2} \cdot \frac{1}{3}$
…
Pr(Alice wins on i^{th} round) = ?
Pr(Alice wins on any round) = ?

Often probability space $(\Omega, \mathbb{P})$ is given **implicitly** of the following form, using chain rule and/or independence

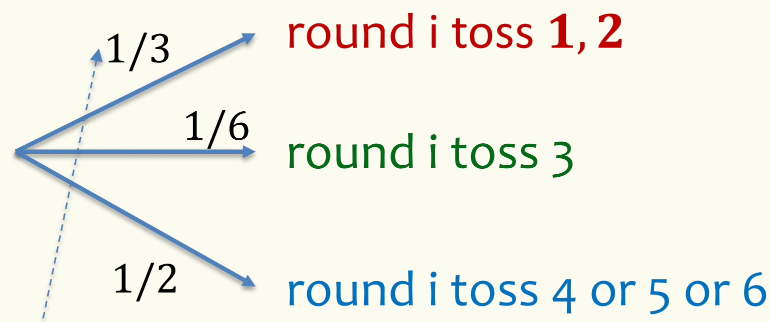
Experiment proceeds in n sequential steps, each step follows some local rules defined by conditional probability and independence.

– Allows for easy definition of experiments where $|\Omega| = \infty$

Sequential Process – Independent case, don't care entire prob. space

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

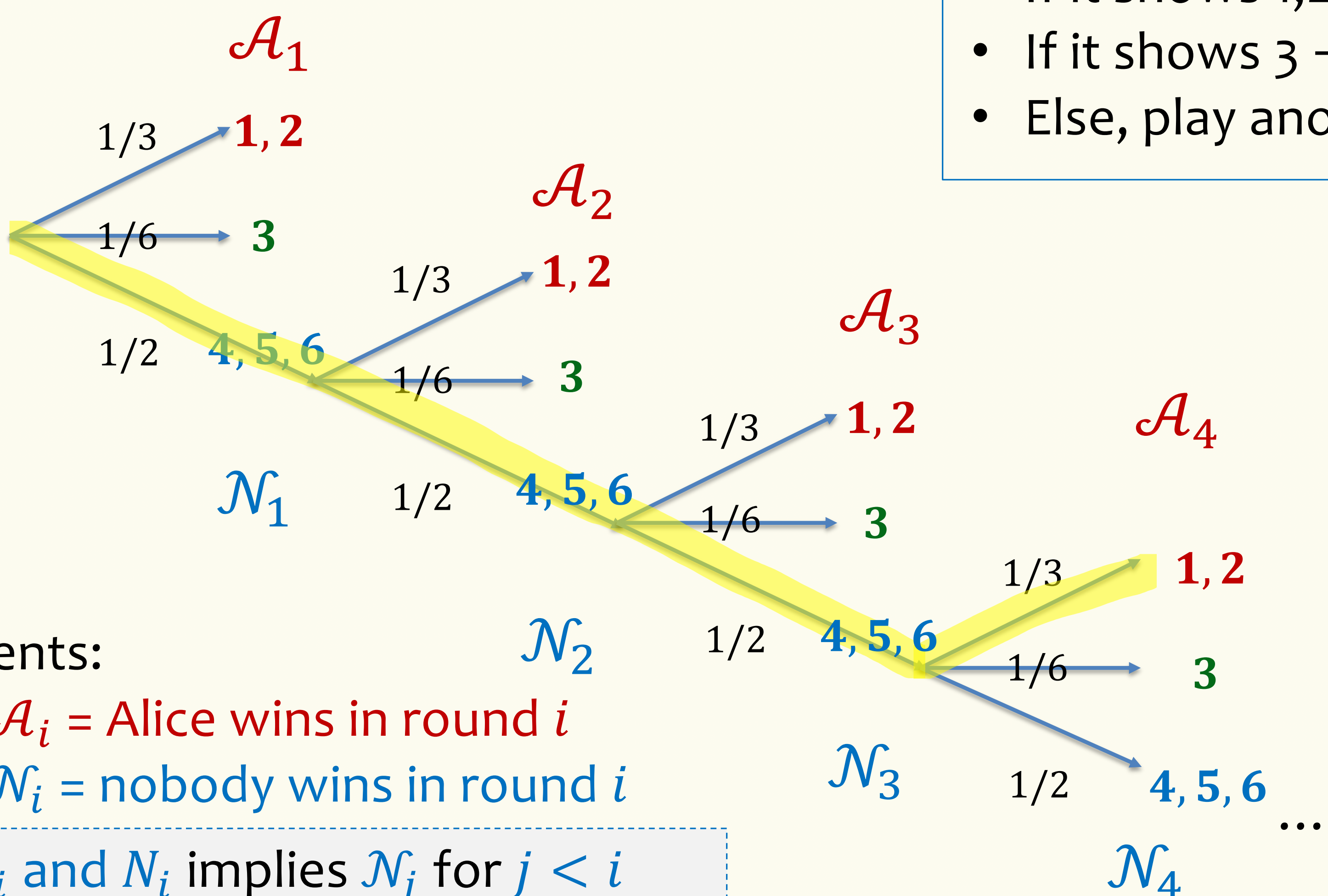


Pr (Alice win | game proceeds to round i) = 1/3

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:

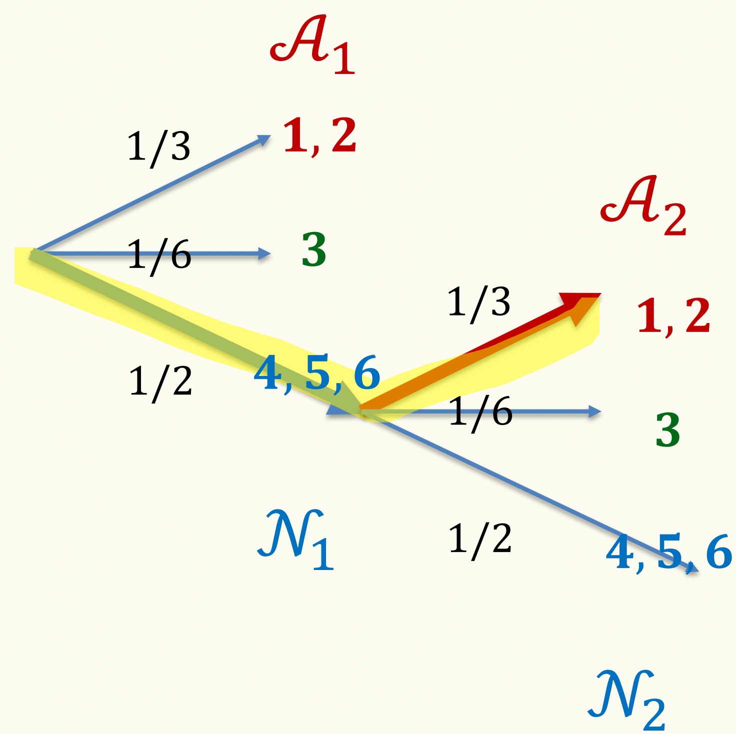
- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in round i

$\mathcal{A}_i$ and $\mathcal{N}_i$ implies $\mathcal{N}_j$ for $j < i$

Sequential Process – Example

Events:

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



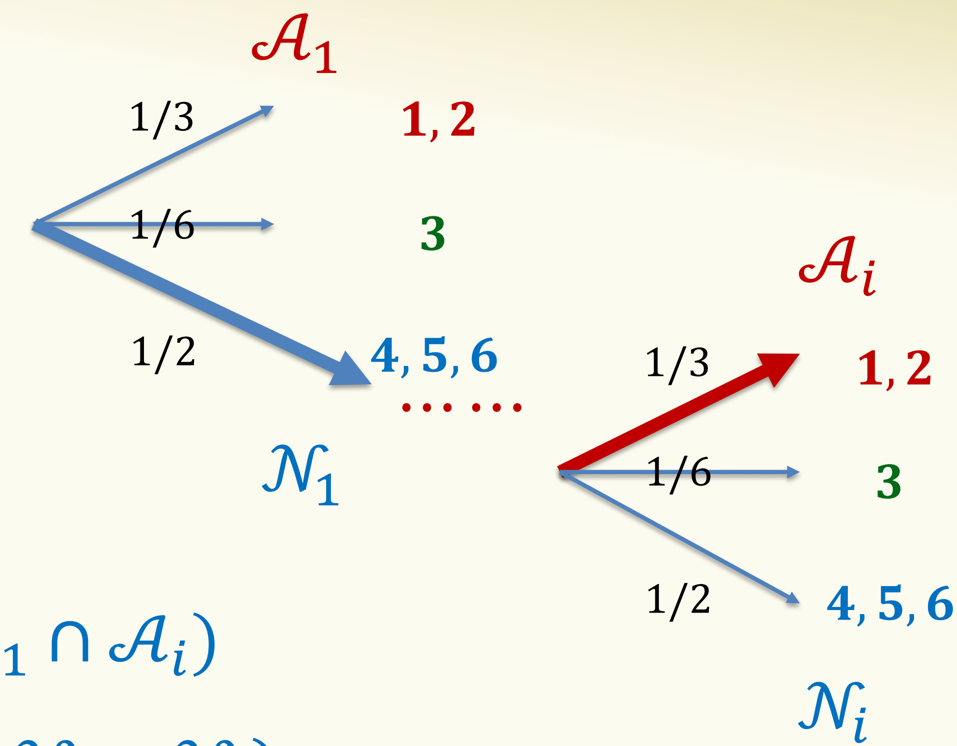
$$\mathbb{P}(\mathcal{A}_2) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{A}_2)$$
$$= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{A}_2 | \mathcal{N}_1)$$
$$= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

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

2nd dice indep of 1st dice

Sequential Process – Example

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

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

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

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

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

Independence – Another Look

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

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

"Equivalently." $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A}).$

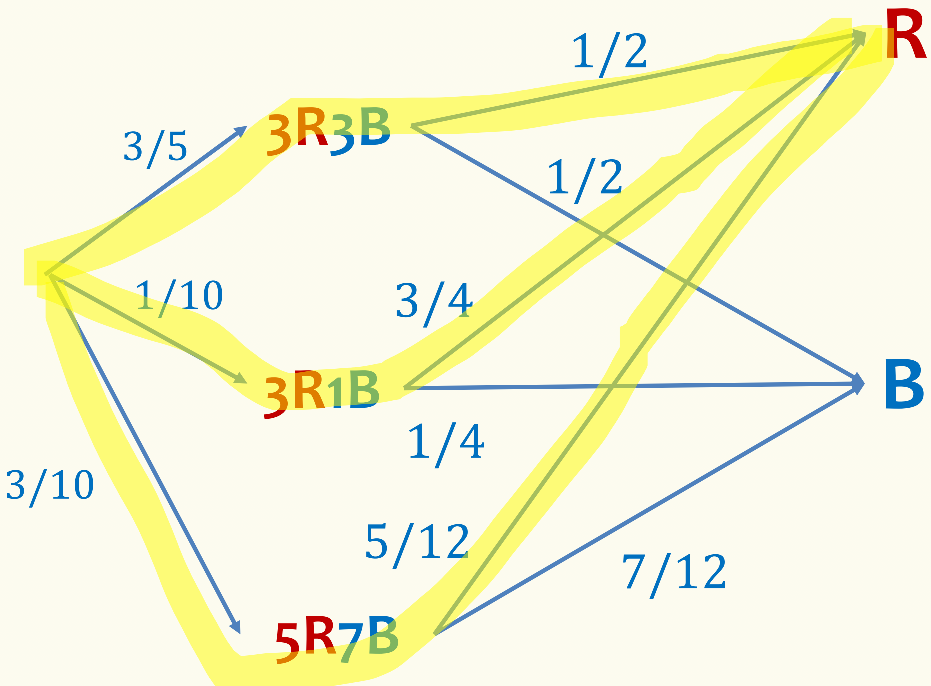
It is important to understand that independence is a property of probabilities of outcomes, not of the root cause generating these events.

Events generated independently → *their probabilities satisfy independence*

~~←~~ *Not necessarily*

This can be counterintuitive!

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 12 **blue** balls w/ probability 3/10

We draw a ball at random from the urn.

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

$\mathbb{P}(\mathbf{3R3B}) \times \mathbb{P}(\mathbf{R} \mid \mathbf{3R3B})$

Are R and 3R3B independent?

Poll: pollev/rachel312

A. Yes, independent B. No

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

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on C if

$$\mathbb{P}(C) \neq 0 \text{ and } \mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid C) = \mathbb{P}(\mathcal{A} \mid C) \cdot \mathbb{P}(\mathcal{B} \mid C).$$

Equivalence:

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

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

Example – Throwing Dies

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?

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

$$= \Pr(H \mid C1)^3 \Pr(C1) + \Pr(H \mid C2)^3 \Pr(C2)$$ Conditional Independence

$$= 0.3^3 \cdot 0.5 + 0.9^3 \cdot 0.5 = 0.378$$

Next Lecture

Next: Random Variables – First encounter

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 20 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

Example. Throwing two dice $\Omega = \{(i, j) \mid i, j \in [6]\}$ $\mathbb{P}\big((i, j)\big) = \frac{1}{36}$.

$$X(i, j) = i + j$$
$$Y(i, j) = i \cdot j$$
$$Z(i, j) = i$$

Random variables!

Random Variables

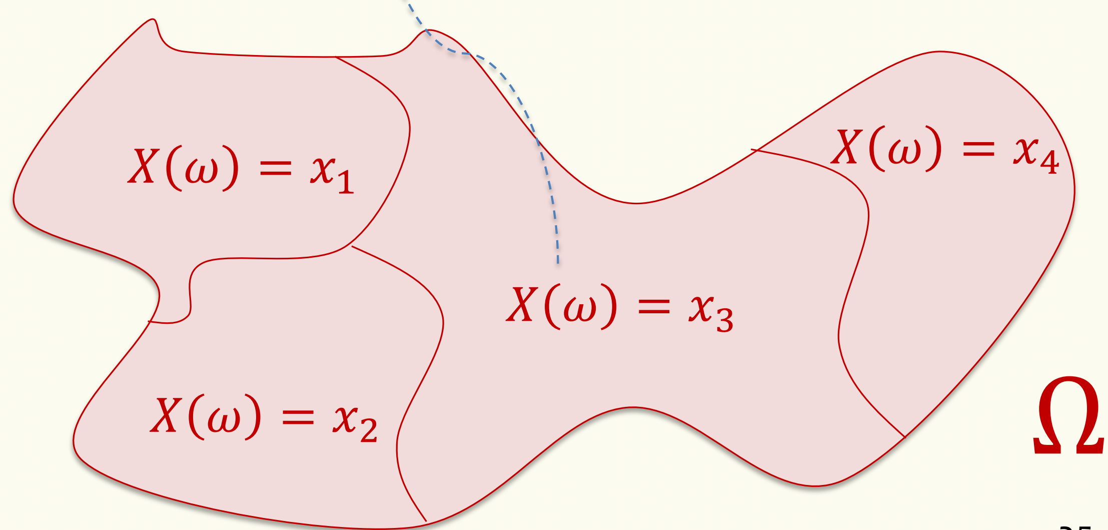
Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

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

We write $\mathbb{P}(X = x) = \mathbb{P}(\{X = x\}) = \mathbb{P}\{\omega \in \Omega \mid X(\omega) = x\}$

Random variables partition the sample space.

$\Sigma_x \mathbb{P}(X = x) = 1$



Random Variables

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

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

We write $\mathbb{P}(X = x) = \mathbb{P}(\{X = x\})$.

Example. $X(i, j) = i + j$

$$\mathbb{P}(X = 4) = \mathbb{P}(\{(1,3), (3,1), (2,2)\}) = 3 \times \frac{1}{36} = \frac{1}{12}$$

$$\mathbb{P}(X = 3) = \mathbb{P}(\{(1,2), (2,1)\}) = 2 \times \frac{1}{36} = \frac{2}{36} = \frac{1}{18}$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{(1,1)\}) = 1 \times \frac{1}{36} = \frac{1}{36}$$

Random Variables

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

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

We write $\mathbb{P}(X = x) = \mathbb{P}(\{X = x\})$.

Example. $Z(i, j) = i$

$$\mathbb{P}(Z = 2) = \mathbb{P}(\{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}) = \frac{1}{6}$$