CSE 312

Foundations of Computing II

Lecture 7: Chain Rule and Independence

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

Announcements

- No concept check today!
- Section tomorrow is **important** with new content that you will need on pset 3, problem 7. Bring your laptops.
- I have to be out of town (and will be largely unreachable) Thursday-Saturday – Aleks will give Friday's lecture!
- Quiz 1 out later next week. Will cover material from the first two problem sets.

Friday 10/8: Bayes Theorem with Law of Total Probability

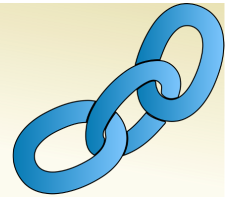
Bayes Theorem with LTP: Let $E_1, E_2, \dots, E_n$ be a partition of the sample space, and F and event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{P(F|E_1)P(E_1)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

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

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

Monday 10/10: Chain Rule



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

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

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

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

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

Monday: 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

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

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 1^{st} round) = $\frac{1}{3}$

...

Pr(Alice wins on i^{th} round) = ?

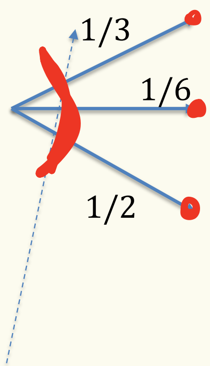
Pr(Alice wins) = ?

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

- 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

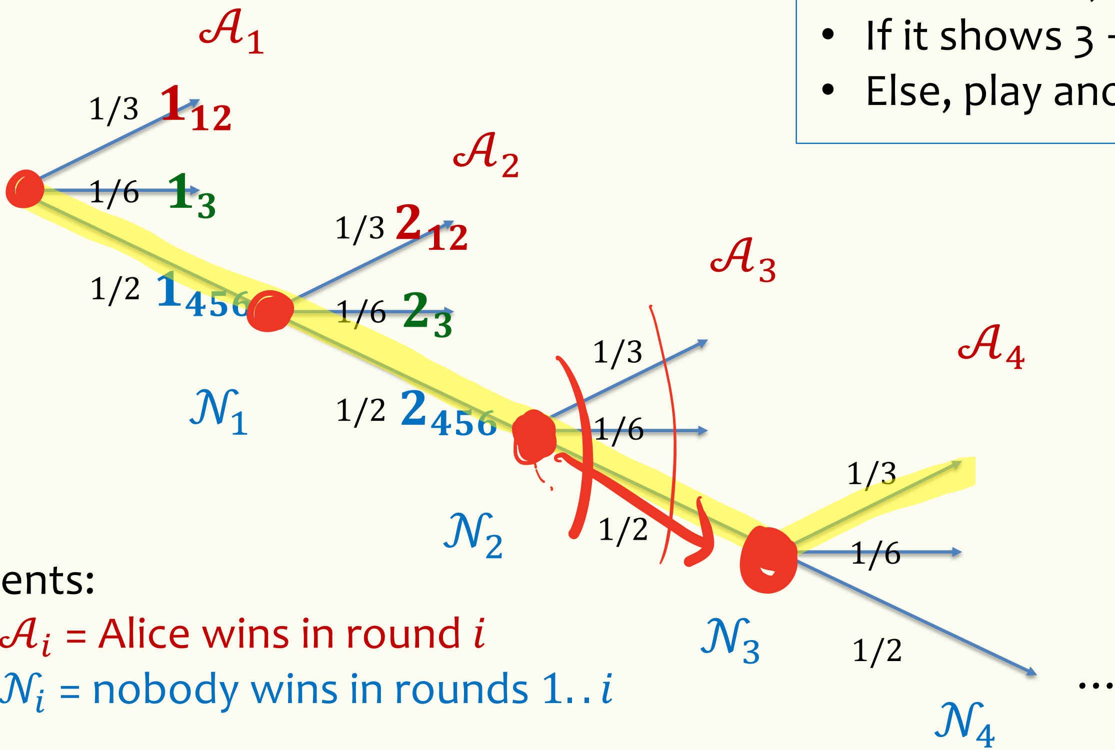
$\boldsymbol{i_{3}}$ = i th round toss 3

$\boldsymbol{i_{456}}$ = 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 rounds $1..i$

$Pr(N_i \mid N_{i-1}) = \frac{1}{2}$

Sequential Process – Example

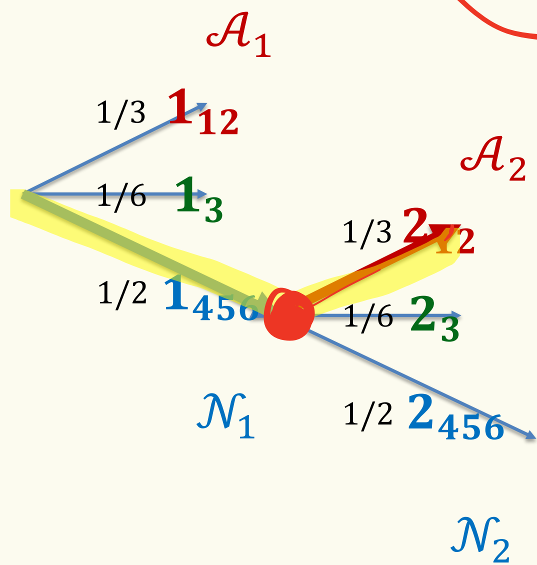
Events:

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

$$\mathbb{P}(\mathcal{A}_2) = \Pr(N_1 \cap A_2)$$

$$= \Pr(N_1)\Pr(A_2 \mid N_1)$$

$$= \frac{1}{2} \cdot \frac{1}{3}$$

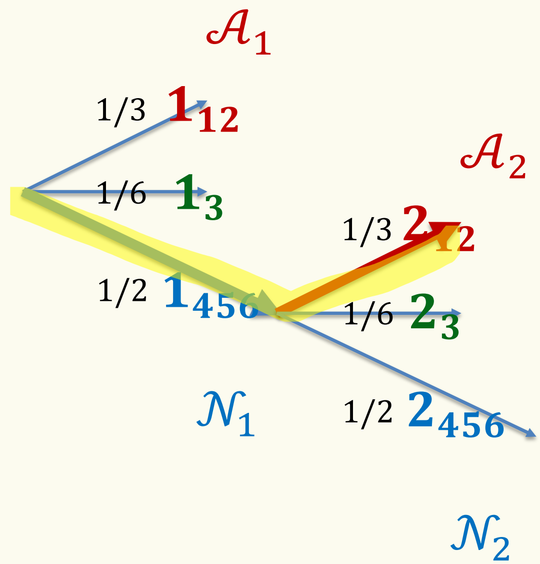


2nd roll indep of 1st roll

Sequential Process – Example

Events:

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



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

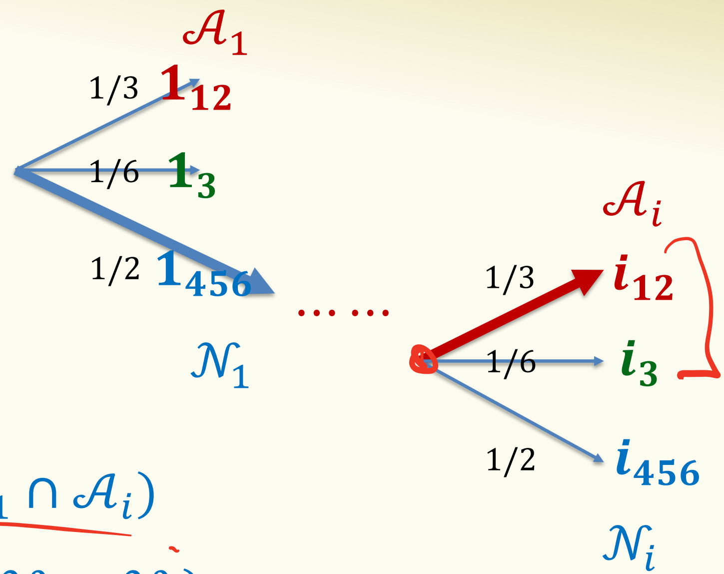
The event $\mathcal{A}_2$ implies $\mathcal{N}_1$, and this means that $\mathcal{A}_2 \cap \mathcal{N}_1 = \mathcal{A}_2$

2^{nd} roll indep of 1^{st} roll

Sequential Process – Example

Events:

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



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

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

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

chain rule

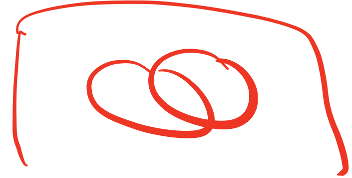
$$= \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$$

Sequential Process -- Example

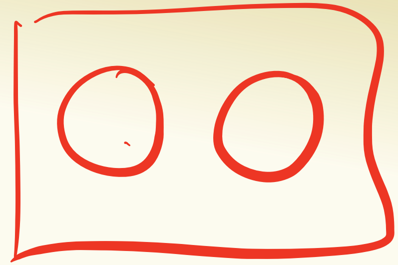
$\mathcal{A}_i$ = Alice wins in round i $\quad \mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

What is the probability that Alice wins?

$$\Pr(\underline{A_1} \cup \underline{A_2} \cup \underline{A_3} \cup \cdots\cdots) = \sum_{i=1}^{\infty} \Pr(A_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} \frac{1}{3}$$



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.*

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

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

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

$$\boxed{\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i}} = \frac{1}{1-\frac{1}{2}} = 2$$

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

all indep

A, B indep $\Rightarrow$ A, B^c; A^c, B; A^c, B^c

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

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 $\qquad \mathbb{P}(A) = 0.02$

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

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

$$\boxed{\begin{aligned} Pr(A \cup B) &= 1 - P(A^c \cap B^c) \\ &= 1 - Pr(A^c)\,P(B^c) \end{aligned}}$$

$$\begin{aligned} Pr(A^c \cap B^c) &= 1 - Pr(A \cup B) \\ &= 1 - (Pr(A) + P(B) - P(A \cap B)) \\ &= 1 - Pr(A) - P(B) + P(A)P(B) \\ &= (1 - Pr(A))(1 - P(B)) \end{aligned}$$

$Pr(A^c \cup B^c) = 1 - P(A \cap B)$
$= 1 - 0.02 \cdot 0.1$

Showed
$\Rightarrow \boxed{Pr(A^c \cap B^c) = P(A^c) P(B^c)}$

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.

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Independence – Another Look

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

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

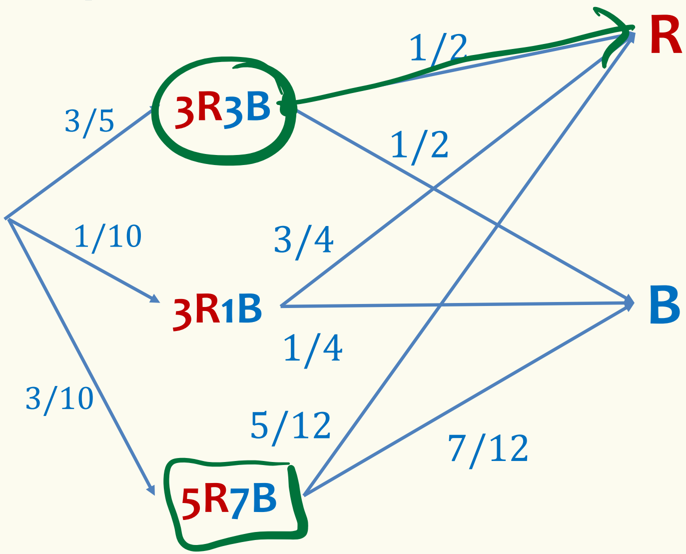
"Equivalently." $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$.

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

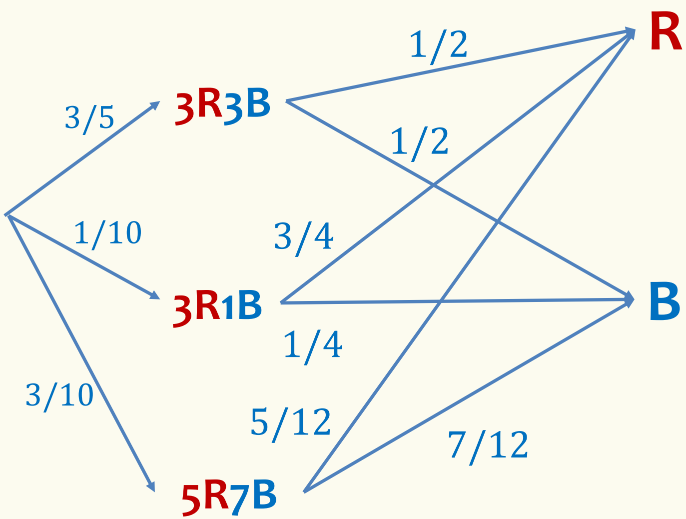
We draw a ball at random from the urn.

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

$$P(R) = P(R \mid 3R3B)\,P(3R3B) + P(R \mid 3R1B)\,P(3R1B) + P(R \mid 5R7B)\,P(5R7B)$$

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

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.

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

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

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

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

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Conditional Independence

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

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

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 we get all heads?

$$\Pr(HH) = \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2) \qquad \text{LTP}$$

$$= P(H|C1)\,P(H|C_1)\,P(C1) + \left(P(H|C2)\right)^2 P(C_2)$$

$$= 0.3^2 \tfrac{1}{2} + 0.9^2 \tfrac{1}{2} =$$

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

$$= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = \boxed{0.45}$$

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