Conditional Independence

CMPUT 366: Intelligent Systems

P&M §8.2

Logistics & Assignment #1

- **Assignment #1** is due Feb 8 11:55pm (**next week**)
- Office hours have begun!
  - Not mandatory; for getting help from TAs
  - New Monday office hours: **6:00-7:00pm** Mountain time
  - Python refresher **TODAY** (will be recorded)

Lecture Outline

1. Recap
2. Operating on Conditional Probabilities
3. Expected Value
4. Structure
5. Marginal Independence
6. Conditional Independences

Recap: Probability

- **Probability** is a numerical measure of **uncertainty**
  - **Not** a measure of **truth**
- **Semantics:**
  - A **possible world** is a **complete assignment** of values to variables
  - Every possible world has a probability
  - Probability of a **proposition** is the sum of probabilities of **possible worlds** in which the statement is **true**

Recap: Conditional Probability

- When we receive **evidence** in the form of a proposition e, it **rules out** all of the possible worlds in which e is **false**
  - We set those worlds' probability to **0**, and **rescale** remaining probabilities to sum to **1**
- Result is probabilities **conditional on e**: $P(h \mid e)$

Chain Rule

Definition: conditional probability

$$P(h \mid e) = \frac{P(h, e)}{P(e)}$$

- We can run this **in reverse** to get

$$P(h, e) = P(h \mid e) \times P(e)$$

Definition: chain rule

$$\begin{aligned} P(\alpha_1, \ldots, \alpha_n) &= P(\alpha_1) \times P(\alpha_2 \mid \alpha_1) \times \cdots \times P(\alpha_n \mid \alpha_1, \ldots, \alpha_{n-1}) \\ &= \Pi_{i=1}^{n} P(\alpha_i \mid \alpha_1, \ldots, \alpha_{i-1}) \end{aligned}$$

Bayes' Rule

- From the **chain rule**, we have

$$
\begin{aligned}
P(h, e) &= P(h \mid e)P(e) \\
&= P(e \mid h)P(h)
\end{aligned}
$$

- **Often**, $P(e \mid h)$ is easier to compute than $P(h \mid e)$.

Bayes' Rule:

$$\underbrace{P(h \mid e)}_{\text{Posterior}} = \frac{\overbrace{P(e \mid h)}^{\text{Likelihood}}\ \overbrace{P(h)}^{\text{Prior}}}{\underbrace{P(e)}_{\text{Evidence}}}$$

Expected Value

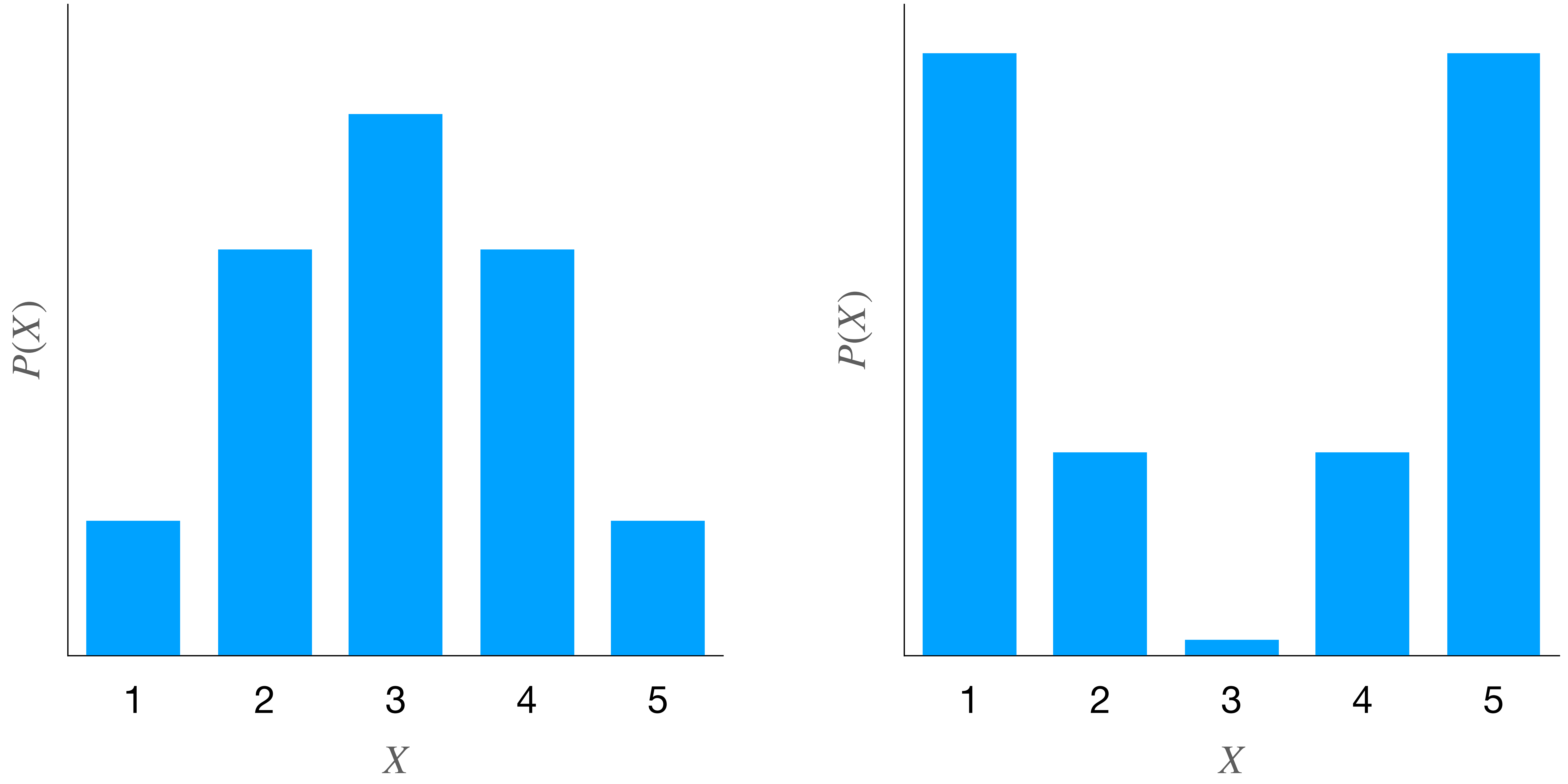
- The **expected value** of a **function** f on a random variable is the weighted **average** of that function over the domain of the random variable, **weighted** by the **probability** of each value:

$$\mathbb{E}\left[f(X)\right] = \sum_{x \in dom(X)} P(X = x) f(x)$$

- The **conditional expected value** of a **function** f is the average value of the function over the domain, weighted by the **conditional probability** of each value:

$$\mathbb{E}\left[f(X) \mid Y = y\right] = \sum_{x \in dom(X)} P(X = x \mid Y = y) f(x)$$

Expected Value Examples



$\mathbb{E}[X] = 3$

$\mathbb{E}[X^2] \simeq 10$

$\mathbb{E}[X] = 3$

$\mathbb{E}[X^2] \simeq 12$

Unstructured Joint Distributions

- Probabilities are not fully **arbitrary**
  - **Semantics** tell us probabilities given the joint distribution.
  - Semantics alone do not restrict probabilities **very much**
- In general, we might need to **explicitly** specify the entire **joint distribution**
  - **Question:** How many numbers do we need to assign to fully specify a joint distribution of k binary random variables?
- We call situations where we have to explicitly enumerate the entire joint distribution **unstructured**

Structure

- Unstructured domains are very hard to reason about
- Fortunately, most real problems are generated by some sort of **underlying process**
  - This gives us **structure** that we can exploit to represent and reason about probabilities in a more **compact** way
  - We can **compute** any required joint probabilities based on the process, instead of specifying every possible joint probability explicitly
- Simplest kind of structure is when random variables don't **interact**

Marginal Independence

When the value of one variable **never** gives you information about the value of the other, we say the two variables are **marginally independent.**

Definition:

Random variables X and Y are **marginally independent** iff

1. $P(X = x \mid Y = y) = P(X = x)$, and
2. $P(Y = y \mid X = x) = P(Y = y)$

for all values of $x \in dom(X)$ and $y \in dom(Y)$.

Marginal Independence Example

- I flip four **fair coins**, and get four results: C_1, C_2, C_3, C_4
- **Question:** What is the probability that C_1 is **heads**?
  - $P(C_1 = heads)$
- Suppose that C_2, C_3, and C_4 are **tails**
- **Question:** *Now* what is the probability that C_1 is **heads**?
  - $P(C_1 = heads \mid C_2 = tails, C_3 = tails, C_4 = tails)$
  - Why?

Properties of Marginal Independence

Proposition:

If X and Y are marginally independent, then

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all values of $x \in dom(X)$ and $y \in dom(Y)$.

Proof:

1. $P(X = x, Y = y) = P(X = x \mid Y = y)P(Y = y)$ Chain rule
2. $P(X = x, Y = y) = P(X = x)P(Y = y)$ Marginal independence

∎

Exploiting Marginal Independence

C_1	P
H	0.5

C_2	P
H	0.5

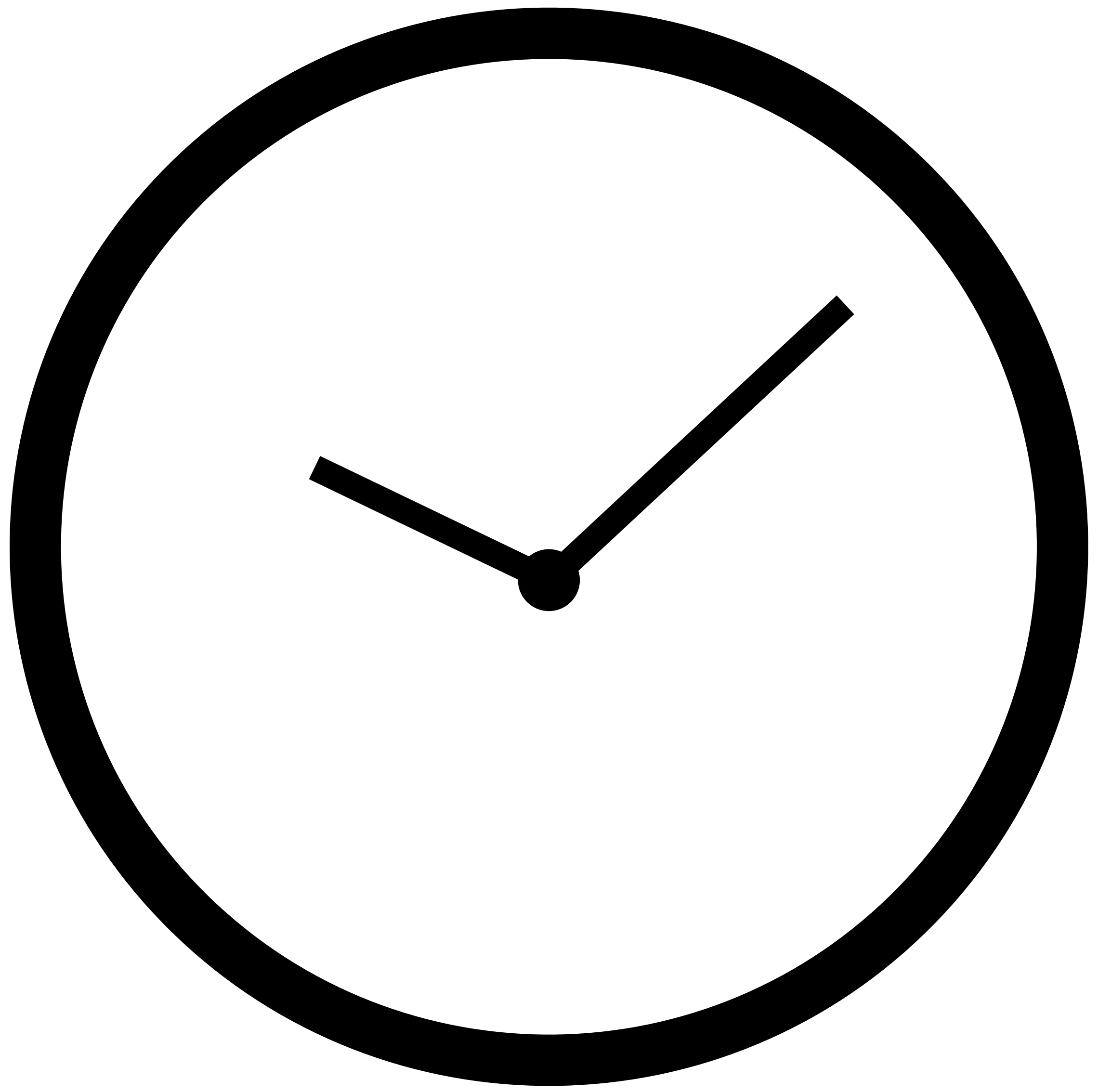
C_3	P
H	0.5

C_4	P
H	0.5

- Instead of storing the **entire joint distribution**, we can store 4 **marginal distributions**, and use them to recover joint probabilities
    - **Question:** How many numbers do we need to assign to fully specify the marginal distribution for a **single** binary variable?
- If **everything** is independent, learning from observations is hopeless (**why?**)
    - But also if **nothing** is independent
    - The **intermediate** case, where many variables are independent, is ideal

C_1	C_2	C_3	C_4	P
H	H	H	H	0.0625
H	H	H	T	0.0625
H	H	T	H	0.0625
H	H	T	T	0.0625
H	T	H	H	0.0625
H	T	H	T	0.0625
H	T	T	H	0.0625
H	T	T	T	0.0625
T	H	H	H	0.0625
T	H	H	T	0.0625
T	H	T	H	0.0625
T	H	T	T	0.0625
T	T	H	H	0.0625
T	T	H	T	0.0625
T	T	T	H	0.0625

Clock Scenario



Example:

- I have a stylish but impractical clock with no number markings
- Two students, Alice and Bob, both look at the clock at the same time, and form opinions about what time it is
  - Their opinion of the time is **directly affected** by the actual time
  - They don't talk to each other, so Alice's opinion of the time is not directly affected by Bob's opinion of the time (& vice versa)
- **Question:** Are A and B **marginally independent**?

$$P(A \mid B) \neq P(A)$$

- **Question:** If we know it is 10:09. Are A and B **independent**?

$$P(A \mid B, T = \text{10:09}) = P(A \mid T = \text{10:09})$$

Random variables:

A - Time Alice thinks it is

B - Time Bob thinks it is

T - Actual time

Conditional Independence

When knowing the value of a **third** variable Z **makes** two variables A, B independent, we say that they are **conditionally independent given** Z (or **independent conditional on** Z).

Definition:
Random variables X and Y are **conditionally independent given *Z*** iff

$$P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)$$

for all values of $x \in dom(X)$, $y \in dom(Y)$, and $z \in dom(Z)$.
We write this using the notation $X \perp\!\!\!\perp Y \mid Z$.

Clock example: A and B are conditionally independent given T.

Properties of Conditional Independence

Proposition:

If X and Y are conditionally independent given Z, then

$$P(X = x, Y = y \mid Z) = P(X = x \mid Z)P(Y = y \mid Z)$$

for all values of $x \in dom(X)$, $y \in dom(Y)$, and $z \in dom(Z)$.

Proof:

1. $P(X = x, Y = y \mid Z) = P(X = x \mid Y = y, Z = z)P(Y = y \mid Z)$ Chain rule
2. $P(X = x, Y = y \mid Z) = P(X = x \mid Z)P(Y = y \mid Z)$ Conditional independence

∎

Properties of Conditional Independence

Question: Is conditional independence **commutative**?

- i.e., If $X \perp\!\!\!\perp Y \mid Z$, is it also true that $Y \perp\!\!\!\perp X \mid Z$?

Proof:

$$X \perp\!\!\!\perp Y \mid Z \iff P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$ previous result

$$\iff P(Y, X \mid Z) = P(Y \mid Z)P(X \mid Z)$$ commutativity of multiplication

$$\iff Y \perp\!\!\!\perp X \mid Z \quad \blacksquare$$

Exploiting Conditional Independence

If X and Y are marginally independent given Z, then we can again just store **smaller tables** and recover joint distributions by **multiplication.**

- **Question:** How many **tables** do we need to store in order to be able to compute the joint distribution of X, Y, Z when X and Y are independent given Z?
  - i.e., how many table to be able to compute $P(X = x, Y = y, Z = z)$ for every combination of x, y, z?

Preview: In the upcoming lectures, we will see how to efficiently exploit **complex structures** of conditional independence

Simplified Clock Example

A	*T*	P(A | T)
12	1	0.25
1	1	0.50
2	1	0.25
1	2	0.25
2	2	0.50
3	2	0.25
2	3	0.25
3	3	0.50
4	3	0.25

⋮

B	*T*	P(B | T)
12	1	0.25
1	1	0.5
2	1	0.25
1	2	0.25
2	2	0.5
3	2	0.25
2	3	0.25
3	3	0.5
4	3	0.25

⋮

T	P(T)
1	0
2	1/10
3	1/10
4	1/10
5	1/10
6	1/10
7	1/10
8	1/10
9	1/10
10	1/10
11	1/10
12	0

$$
\begin{aligned}
& P(A=1, B=2, T=2) \\
&= P(A=1 \mid T=2) P(B=2 \mid T=2) P(T=2) \\
&= 0.25 \times 0.5 \times 0.10 \\
&= 0.0125
\end{aligned}
$$

$$
\begin{aligned}
& P(A=1, B=2, T=1) \\
&= P(A=1 \mid T=1) P(B=2 \mid T=1) P(T=1) \\
&= 0.5 \times 0.25 \times 0.0 \\
&= 0
\end{aligned}
$$

Caveats

- Often, when two variables are **marginally** independent, they are also **conditionally** independent given a third variable
  - E.g., coins C_1, and C_2 are marginally independent, **and also** conditionally independent given C_3: Learning the value of C_3 does not make C_2 any more informative about C_1.
- This is **not always true**
  - Consider another random variable: B is true if both C_1 and C_2 are the **same** value
  - C_1 and C_2 are **marginally independent**: $P(C_1 = heads \mid C_2 = heads) = P(C_1 = heads)$
    - In fact, C_1 and C_2 are also both **marginally independent of *B***: $P(C_1 \mid B = true) = P(C_1)$
  - But if I know the value of B, then knowing the value of C_1 tells me **exactly** what the value of C_2 must be: $P(C_1 = heads \mid B = true, C_2 = heads) \neq P(C_1 = heads \mid B = true)$
    - C_1 and C_2 are **not conditionally independent given *B***

Summary

- **Unstructured** joint distributions are **exponentially** expensive to represent (and operate on)
- **Marginal and conditional independence** are especially important forms of **structure** that a distribution can have
  - Vastly **reduces the cost** of representation and computation
  - **Caveat:** The **relationship** between marginal and conditional independence is not fixed
- Joint probabilities of (conditionally or marginally) **independent** random variables can be computed by **multiplication**