

Conditional Independence

CMPUT 261: Introduction to Artificial Intelligence

P&M §8.2

Assignment #1

- **Assignment #1** is due **THURSDAY** at 11:59pm
 - Hand in on eClass
- **Hints:**
 - You do *not* need to draw a picture of the graph; just **fully specify** it
 - Remember, most graphs in practice will be too big to store!
 - Every arc leaving a state corresponds to an action available at that state
 - The destination of a state is the new state from taking the action
 - Different states might have different actions available
- **BONUS MARKS:**
 - I will award **5 assignment bonus marks** to the student who posts the best description of what it means to fully specify a search graph to eClass
 - I will create a superthread for responses, with detailed rules (e.g. deadlines)

Lecture Outline

1. Recap
2. Structure
3. Marginal Independence
4. Conditional Independence

After this lecture, you should be able to:

- Define marginal and conditional independence
- Compute joint probabilities by exploiting marginal and conditional independence
- Compute the minimal number of quantities needed to define a joint distribution given a particular structure / generating process
- Identify marginally or conditionally independent random variables

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** : $P(h \mid e)$

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:** Can I just assign arbitrary numbers in $[0,1]$ to combinations of values?
 - **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**

Generating Process

Example: I keep flipping a fair coin until it come up Heads

- Let S be a random variable that counts how many times I flipped the coin
- Knowing the **process** that **generates** the probabilities gives us a way to **compute** the probabilities rather than explicitly specifying each one individually

Example 2: Same as example 1, except that the coin comes up heads with probability p

Questions:

1. What is $\Pr(S = 1)$?
2. What is $\Pr(S = k)$ (for integer $k > 0$)?
3. How many numbers would I have to assign to **explicitly** describe this distribution?
4. How many numbers would I need to assign to **succinctly** describe the distribution from Example 2?

Marginal Independence

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 $x \in \text{dom}(X)$ and $y \in \text{dom}(Y)$.

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:** What is the conditional 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 $x \in \text{dom}(X)$ and $y \in \text{dom}(Y)$.

Proof:

1. $P(X = x, Y = y) = P(X = x | Y = y)P(Y = y)$ 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 (**why?**)
 - 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

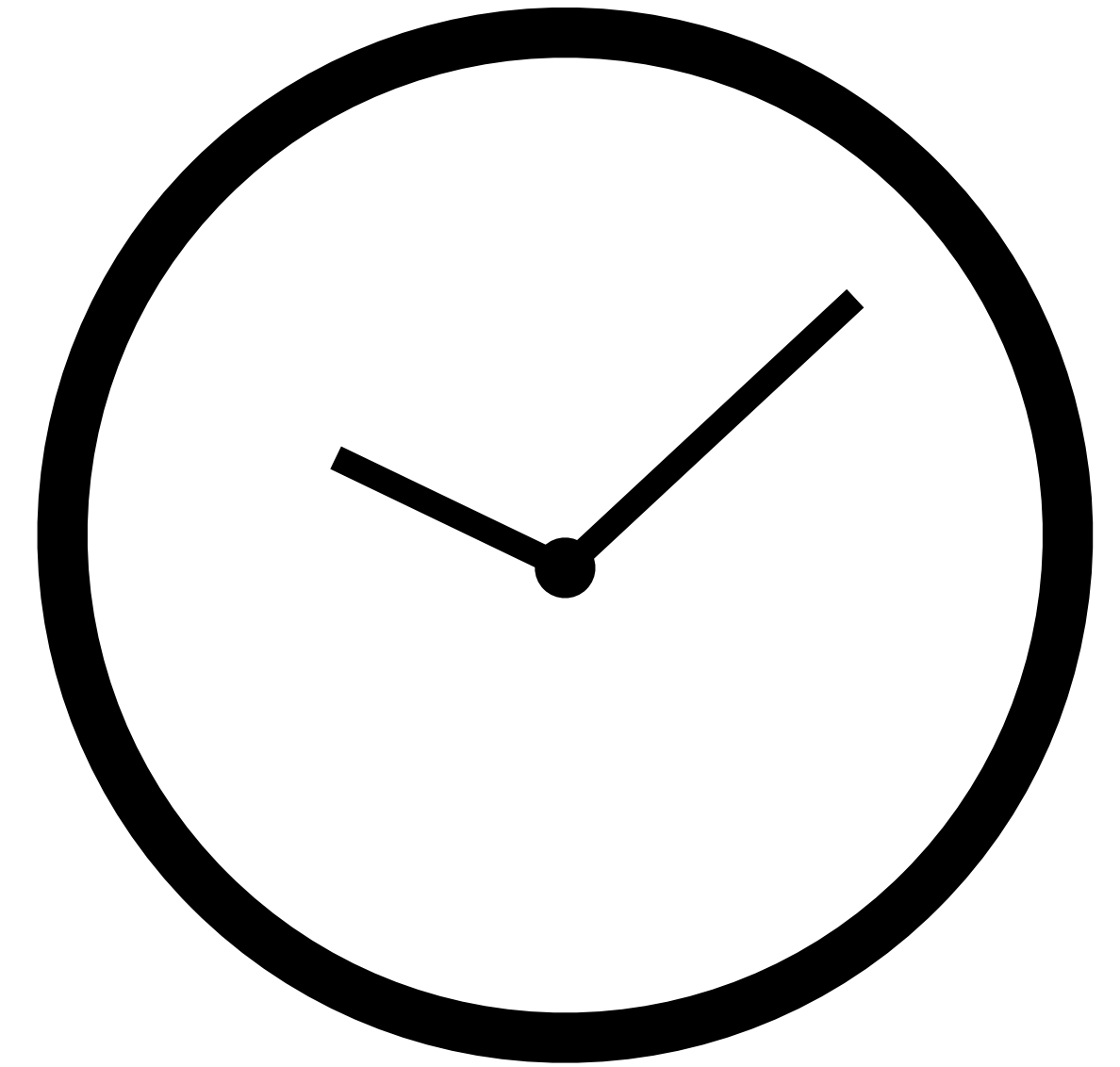
Example:

- 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 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**?

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



Random variables:

A - Time Alice thinks it is

B - Time Bob thinks it is

T - Actual time

Conditional Independence

When conditioning on the value of a **third** variable Z **makes** two variables A, B independent, we say that they are **conditionally independent given 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 $x \in \text{dom}(X)$, $y \in \text{dom}(Y)$, and $z \in \text{dom}(Z)$.

We can write this using the notation $X \perp\!\!\!\perp Y \mid Z$.

I.e., $X \perp\!\!\!\perp Y \mid Z$ iff knowing Z means that learning A gives no additional information about B .

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

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

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

Proof:

$$1. \quad P(X = x, Y = y \mid Z = z) = P(X = x \mid Y = y, Z = z)P(Y = y \mid Z = z) \quad \text{Chain rule}$$

$$2. \quad P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z) \quad \text{Conditional independence}$$



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:

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

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

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

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 tables 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

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

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

Warnings

- 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
 - **Beware:** The **relationship** between marginal and conditional independence is not fixed
- Joint probabilities of (conditionally or marginally) **independent** random variables can be computed by **multiplication**