Conditional Independence

CMPUT 366: Intelligent Systems

P&M §8.2

Lecture Outline

1. Recap
2. Structure
3. Marginal Independence
4. 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**: $P(h \mid e)$

Recap: 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 | h) is easier to compute than P(h | e).

Bayes' Rule:

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

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? A: 2^k - 1
- 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**?
  - $P(C_1 = \text{heads})$ A: 1/2
- Suppose that C_2, C_3, and C_4 are **tails**
- **Question:** *Now* what is the probability that C_1 is **heads**?
  - $P(C_1 = \text{heads} \mid C_2 = \text{tails}, C_3 = \text{tails}, C_4 = \text{tails})$ A: 1/2

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

Proposition:
If X and Y are marginally independent, then

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

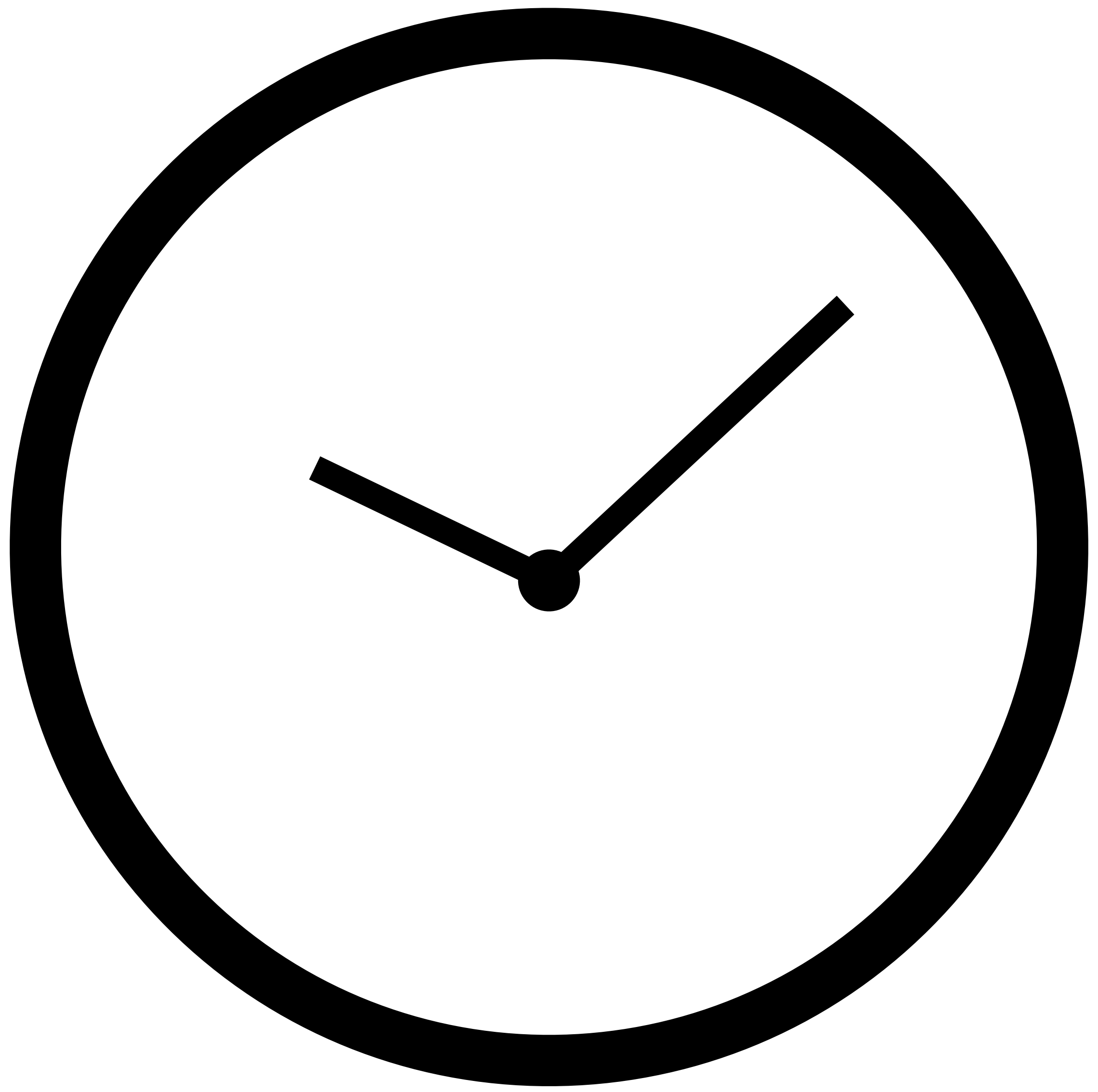
for all values of $x \in \text{dom}(X)$ and $y \in \text{dom}(Y)$

- Instead of storing the **entire joint distribution**, we can store 4 **marginal distributions**, and use them to recover joint probabilties
  - **Question:** How many numbers do we need to assign to fully specify the marginal distribution for a **single** binary variable? A: 1
- If **everything** is independent, learning from observations is hopeless
  - 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
- **Question:** Are Alice and Bob's opinions **independent**?

  $P(A \mid B) \neq P(A)$ A: no

- **Question:** Suppose it is 10:00. Are *A* and *B* **independent**?

  $P(A \mid B, T=10{:}00) = P(A \mid T=10{:}00)$ A: yes

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 $x \in \mathrm{dom}(X)$, $y \in \mathrm{dom}(Y)$, and $z \in \mathrm{dom}(Z)$.

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

Exploiting Conditional Independence

Proposition:
If X and Y are marginally 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 $z \in \text{dom}(Z)$.

- We can again just store **smaller tables** and recover joint distributions by **multiplication**
- **Question:** How many **tables** do we need to store for variables X, Y, Z when X and Y are independent given Z?

A: 3

Caveats

- Sometimes, 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**: $P(C_1\text{=heads} \mid C_2\text{=heads}) = P(C_1\text{=heads})$
    - In fact, C_1 and C_2 are also both **marginally independent of B**: $P(C_1 \mid B\text{=true}) = P(C_1)$
  - 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: $P(C_1\text{=heads} \mid B\text{=true}, C_2\text{=heads}) \neq P(C_1\text{=heads} \mid B\text{=true})$
    - 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**