CSC 384
Introduction to Artificial Intelligence

Uncertainty 2
Introducing Bayesian Networks

Alice Gao and Randy Hickey

Winter 2023

Learning Goals

By the end of the lecture, you should be able to

1. Verify a conditional or unconditional independence relationship using the definitions.
2. Explain why we can represent a joint distribution compactly by using independence relationships.
3. Describe components of a Bayesian network.
4. Compute a joint probability given a Bayesian network.

Outline

- Unconditional & Conditional Independence
- Bayesian Networks

UNCONDITIONAL & CONDITIONAL INDEPENDENCE

Unconditional Independence

X and Y are (unconditionally) independent if and only if

$P(x|y) = P(x|\neg y) = P(X)$

$$P(X|Y) = P(X)$$
$$P(Y|X) = P(Y)$$
$$P(X \wedge Y) = P(X)P(Y)$$

Observing the value of X

does not change our belief in Y.

Conditional Independence

X and Y are conditionally independent given Z iff

$$P(X|Y \land Z) = P(X|Z)$$
$$P(Y|X \land Z) = P(Y|Z)$$
$$P(X \land Y|Z) = P(X|Z)P(Y|Z)$$

If you **already know the value of Z**,
observing the value of X
does not change our belief in Y.

Independence v.s. Conditional Independence

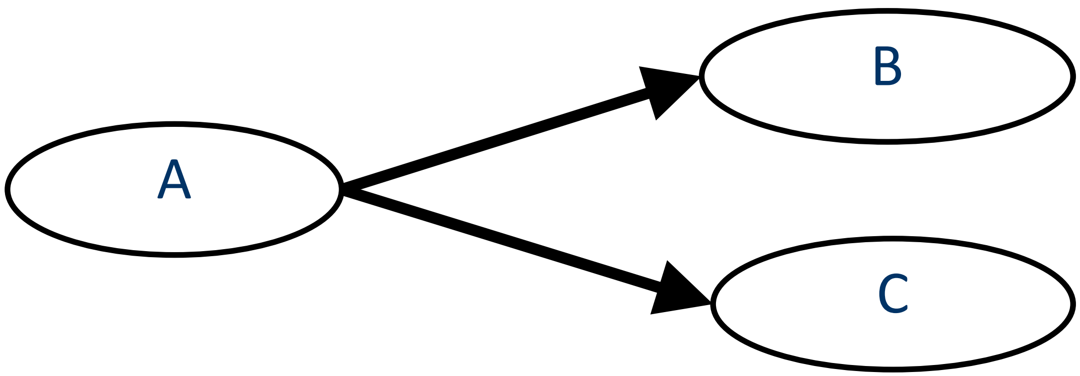
- Independence $\nrightarrow$ Conditional independence, and
- Conditional independence $\nrightarrow$ Independence.

- Exercise: Construct counter-examples for both.

Conditional independence $\nrightarrow$ Independence.

Example:

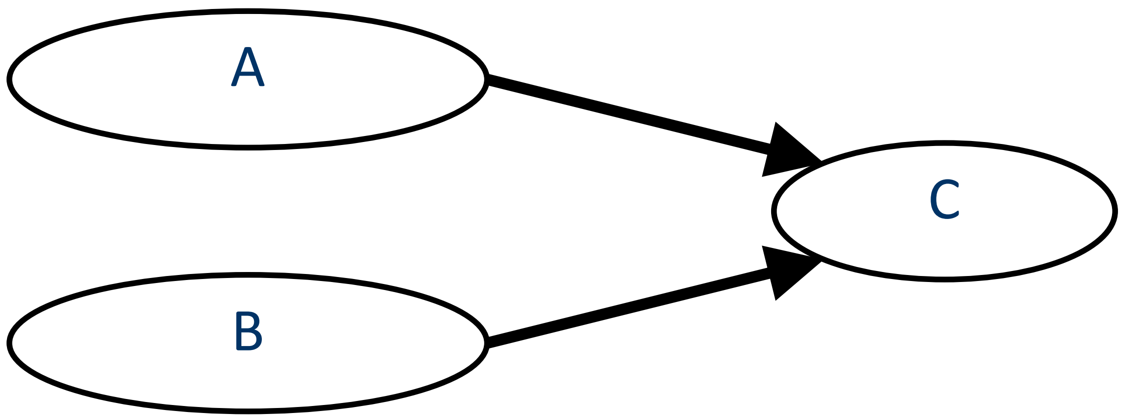
- $P(A) = 0.3, P(B|A) = 0.6, P(C|A) = 0.2$
- B and C are conditionally independent given A.
- But B and C are not independent.



Independence $\nrightarrow$ Conditional Independence.

Example:

- $P(a) = 0.2, P(b) = 0.3$
- $P(c|a \wedge b) = 0.4, P(c|a \wedge \neg b) = 0.5,$
- $P(c|\neg a \wedge b) = 0.6, P(c|\neg a \wedge \neg b) = 0.7$
- A and B are independent.
- But A and B are not independent given C.



Computational Impact of Independence

How many probabilities do we need to specify the joint distribution of X, Y and Z?

1. X, Y and Z are NOT independent.

2. X, Y and Z are independent.

Computational Impact of Independence

If the variables have an independence relationship, we can represent the joint dist. with fewer probabilities.

How many probabilities do we need to specify the joint distribution of X, Y and Z?

1. X, Y and Z are NOT independent.
Must define $2^3 - 1 = 7$ probabilities.
e.g. $P(x \land y \land z)$, $P(x \land y \land \neg z)$, $P(x \land \neg y \land z)$, etc.

2. X, Y and Z are independent.
Must define 3 probabilities ($P(x)$, $P(y)$, $P(z)$) .
e.g. $P(x \land y \land \neg z) = P(x)P(y)P(\neg z)$

BAYESIAN NETWORKS

Why Bayesian Networks?

Example: The Holmes Scenario

- 6 random variables:
  - Earthquake, Radio, Burglary, Alarm, Watson, and Gibbon.
- How many probs do I need to define the joint dist.?

We can compute any probability using the joint dist., but

- Unnatural and tedious to specify the probabilities.
  - What is $P(e \wedge r \wedge b \wedge a \wedge w \wedge g)$? 🤔
- Quickly becomes intractable as # of variables grows.

Why Bayesian Networks?

A Bayesian network

- Is a compact version of the joint distribution, and
- Takes advantage of the unconditional and conditional independence among the variables.

A Bayesian Network for the Holmes Scenario

There are many possible networks for the Holmes scenario.

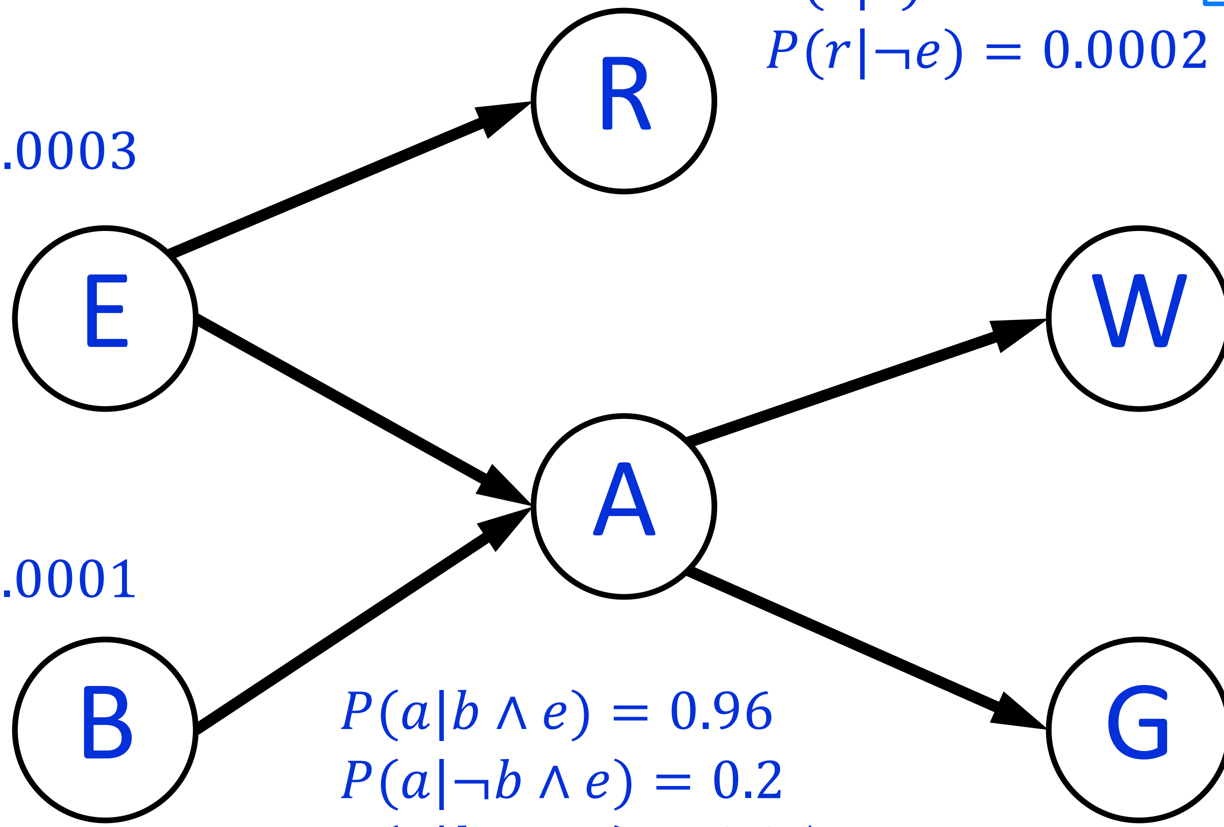
$2^6 - 1 = 63$ probs

$\Rightarrow 12$ probs.

$P(r|e) = 0.9$
$P(r|\neg e) = 0.0002$

$P(e) = 0.0003$

$P(b) = 0.0001$



$P(w|a) = 0.8$
$P(w|\neg a) = 0.4$

$P(g|a) = 0.4$
$P(g|\neg a) = 0.04$

$P(a|b \wedge e) = 0.96$
$P(a|\neg b \wedge e) = 0.2$
$P(a|b \wedge \neg e) = 0.95$
$P(a|\neg b \wedge \neg e) = 0.01$

Components of a Bayesian Network

- A direct acyclic graph
- Each node represents a random variable.
- $X \rightarrow Y$ means that X is a parent of Y.
- Each node X_i has a conditional probability distribution $P(X_i|Parents(X_i))$ quantifying the effects of X_i's parents on X_i.

The Semantics of Bayesian Networks

1. Representing the joint distribution.

2. Encoding the (un)conditional independence relationships.

The Semantics of Bayesian Networks

1. Representing the joint distribution.

Calculate any joint probability given a Bayesian network.

2. Encoding the (un)conditional independence relationships.

Determine any (un)conditional independence relationship using d-separation.

Calculating a Joint Probability

Calculate any joint probability using the formula

$$P(X_n \wedge \cdots \wedge X_1) = \prod_{i=1}^{n} P(X_i | Parents(X_i))$$

Calculating a Joint Probability

Q1: What is the probability that

- The alarm IS going,
- NEITHER a burglary NOR an earthquake is occurring,
- Both Watson and Gibbon ARE calling, and
- There is NO radio report of an earthquake?

Calculating a Joint Probability

Q1: What is the probability that

- The alarm IS going,
- NEITHER a burglary NOR an earthquake is occurring,
- Both Watson and Gibbon ARE calling, and
- There is NO radio report of an earthquake?

Answer:

$$
\begin{aligned}
&P(a \wedge \neg b \wedge \neg e \wedge w \wedge g \wedge \neg r) \\
&= P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)P(w|a)P(g|a)P(\neg r|\neg e) \\
&= (1 - 0.0001)(1 - 0.0003)(1 - 0.0002)(0.01)(0.4)(0.8) \\
&= 0.0032
\end{aligned}
$$

Calculating a Joint Probability

Q2: What is the probability that

- The alarm is NOT going,
- NEITHER a burglary NOR an earthquake is occurring,
- NEITHER Watson NOR Gibbon is calling, and
- There is NO radio report of an earthquake?

(A) 0.57   (B) 0.67   (C) 0.77   (D) 0.87   (E) 0.97

Calculating a Joint Probability

Q2: What is the probability that

- The alarm is NOT going,
- NEITHER a burglary NOR an earthquake is occurring,
- NEITHER Watson NOR Gibbon is calling, and
- There is NO radio report of an earthquake?

(A) 0.57 (B) 0.67 (C) 0.77 (D) 0.87 (E) 0.97

Answer:

$$P(\neg a \wedge \neg b \wedge \neg e \wedge \neg w \wedge \neg g \wedge \neg r)$$
$$= P(\neg a | \neg b \wedge \neg e) P(\neg b) P(\neg e) P(\neg w | \neg a) P(\neg g | \neg a) P(\neg r | \neg e)$$
$$= (1 - 0.0001)(1 - 0.0003)(1 - 0.0002)(1 - 0.01)(1 - 0.4)(1 - 0.04)$$
$$\approx 0.57$$