CS 486/686 Independence and Bayesian Networks I

Yuntian Deng

Lecture 7

RN 12.4, 13.1–13.2 · PM 8.2–8.3

Reminders

DUE TONIGHT Tue Jun 2, 11:59 PM

Chat 3 (Probabilities) on Chrysalis.

Learning goals

- Tell if two variables are **independent**, or **conditionally independent** given a third
- Derive a **compact representation** of a joint distribution using independence
- Describe the components of a **Bayesian network**
- Compute joint probabilities *from* a Bayes net
- Explain independence in the three key structures

Why Bayes nets? The 2^n problem

A joint distribution over n Boolean variables needs 2^n probabilities.

Variables	Probabilities
6 (Holmes)	64
10	1,024
20	~ 1 million
30	~ 1 billion

Independence lets us factor this huge joint into tiny pieces.

Unconditional independence

Independence. X and Y are *(unconditionally) independent* iff (any, and therefore all):

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

Learning Y does **not** change our belief about X.

Two probabilities (the priors) suffice to recover four joint probabilities.

Conditional independence

Conditional independence. X and Y are *conditionally independent given* Z iff (any, and therefore all):

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

Once we know Z, learning Y doesn't change our belief about X.

> **Important:** independence does *not* imply conditional independence, and vice versa. We'll see both on later slides.

Quiz pack: compact representation

Three Boolean variables A, B, C. How many probabilities specify the joint?

Q1. No independence assumed.

> **7.** Chain rule: $P(A) + P(B|A) + P(C|A \wedge B) = 1 + 2 + 4 = 7$.

Q2. A, B, C all mutually independent.

> **3.** Just the priors: $P(A), P(B), P(C) = 3$.

Q3. A, B conditionally independent given C.

> **5.** $P(C) + P(A|C) + P(B|C) = 1 + 2 + 2 = 5$.

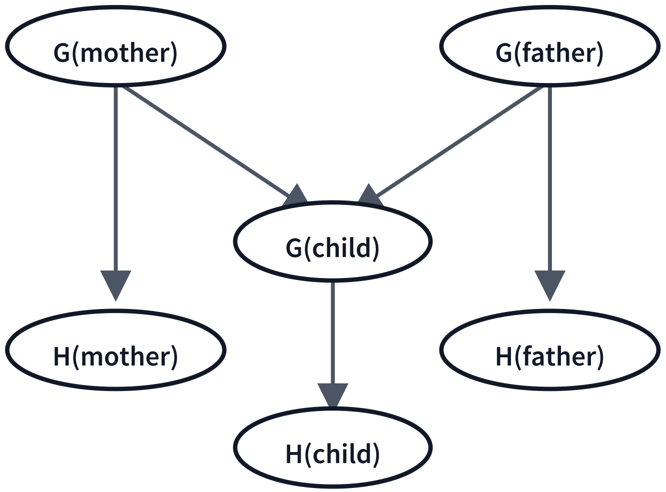
Bayes net = DAG + CPTs

A **Bayesian network** is a *directed acyclic graph* (DAG) in which:

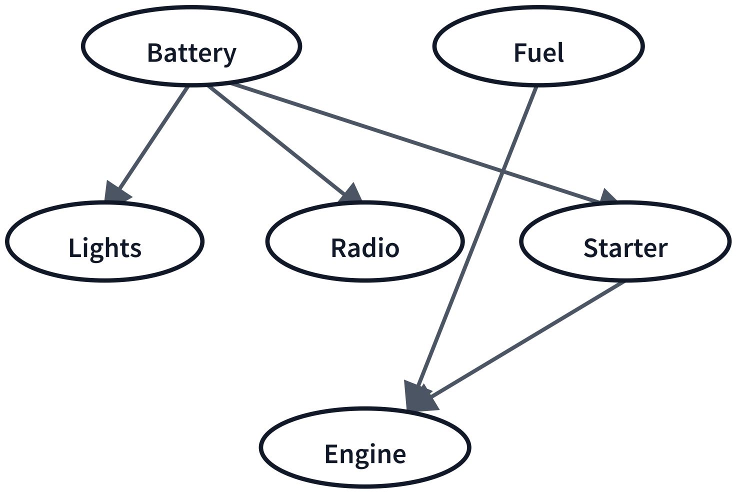
- each node is a random variable,
- each edge $X \to Y$ says "X directly affects Y" $-$ X is a *parent* of Y,
- each node X_i carries a conditional probability table $P(X_i \mid \mathrm{Parents}(X_i))$.

Examples of Bayes nets

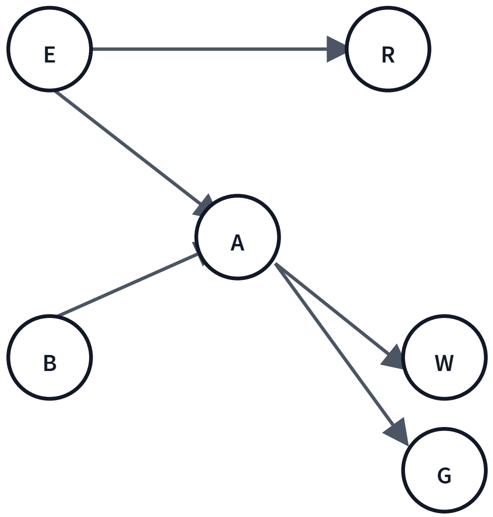
Inheritance of handedness



Car diagnostic network



The Holmes Bayes net



P(E) = 0.0003

P(B) = 0.0001

P(R | ·)
E: 0.9   ¬E: 0.0002

P(A | ·)
B∧E: 0.96   B∧¬E: 0.95
¬B∧E: 0.20   ¬B∧¬E: 0.01

P(W | ·)
A: 0.80   ¬A: 0.40

P(G | ·)
A: 0.40   ¬A: 0.04

From 64 numbers to 12

Representation	Probabilities needed
Full joint over 6 Boolean variables	$2^6 = 64$
Holmes Bayes net: $1 + 1 + 2 + 4 + 2 + 2$	**12**

From those 12 numbers we can reconstruct every one of the 64 joint probabilities.

For larger networks, the savings are exponential.

Joint = product of conditionals

Bayes-net factorization. For a Bayes net over variables $X_1, \ldots, X_n$:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Parents}(X_i)).$$

Each variable is conditionally independent of its non-descendants given its parents.

Worked example: a joint from the Holmes BN

Goal: $P(\neg B \wedge \neg E \wedge A \wedge \neg R \wedge G \wedge W)$. All neighbours call but neither a burglary nor an earthquake actually happened.

Step 1. Factor by the BN structure (using parents):

$$= P(\neg B) \cdot P(\neg E) \cdot P(A|\neg B \wedge \neg E) \cdot P(\neg R|\neg E) \cdot P(G|A) \cdot P(W|A)$$

Step 2. Plug in numbers:

$$= 0.9999 \times 0.9997 \times 0.01 \times 0.9998 \times 0.40 \times 0.80$$

$$\approx \mathbf{3.2 \times 10^{-3}}$$

Q: a different all-negatives joint

Q. What is $P(\neg B \wedge \neg E \wedge \neg A \wedge \neg R \wedge \neg G \wedge \neg W)$?

A. 0.5699

B. 0.6699

C. 0.7699

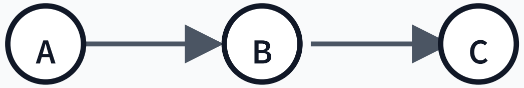
D. 0.8699

E. 0.9699

A — 0.5699. By the same factorization, each factor is now a "no event happened" version:
$= 0.9999 \times 0.9997 \times (1 - 0.01) \times (1 - 0.0002) \times (1 - 0.04) \times (1 - 0.40) \approx 0.5699.$

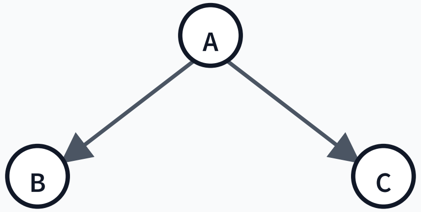
Three key structures

Chain



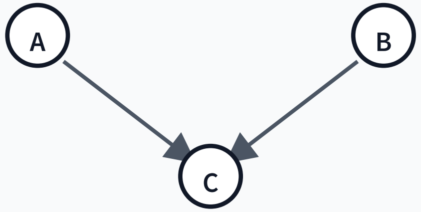
A and C talk *through* B.

Common cause



B and C share a hidden cause A.

Common effect



A and B both cause C (a v-structure).

Each structure has its own independence pattern. Let's check.

Chain: $B \to A \to W$

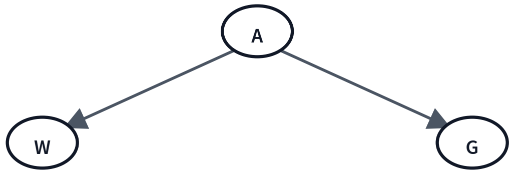
Q1. Are B and W *unconditionally* independent?

> **No.** Knowing B makes the alarm more likely, which makes W more likely. Information flows through A.

Q2. Are B and W *conditionally* independent given A?

> **Yes.** Once we know whether the alarm is going, W only depends on A; B adds nothing more. *The middle node "blocks" the path.*

Common cause: $W \leftarrow A \rightarrow G$



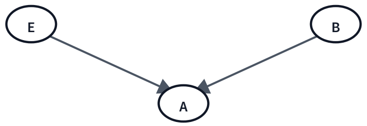
Q1. Are W and G *unconditionally* independent?

> **No.** If W is calling, the alarm is more likely going, so G is more likely too. The shared cause A couples them.

Q2. Are W and G *conditionally* independent given A?

> **Yes.** Once A is known, the two callers are independent observers. *The common cause "blocks" the path when conditioned.*

Common effect: $E \to A \leftarrow B$



Q1. Are E and B *unconditionally* independent?

> **Yes.** Earthquakes and burglaries don't influence each other — they only share the same downstream alarm.

Q2. Are E and B *conditionally* independent given A?

> **No — the opposite.** Given the alarm, learning of an earthquake makes a burglary less likely: **explaining away**. Conditioning on the common effect *creates* a dependence.

> **The v-structure is the surprise:** conditioning on the middle node *opens* a path instead of blocking it.

The three rules

Structure	Are endpoints independent?	Given the middle node?
Chain $A \to B \to C$	No	Yes *blocked*
Common cause $B \leftarrow A \to C$	No	Yes *blocked*
Common effect $A \to C \leftarrow B$	Yes	No *opens up*

In the chain and common-cause structures, knowing the middle node *blocks* information flow. In the v-structure, knowing the middle node *opens* it (explaining away).

Learning goals (recap)

- ✓ Tell if two variables are independent or conditionally independent
- ✓ Derive a compact representation using independence
- ✓ Describe the components of a Bayesian network
- ✓ Compute joint probabilities from a Bayes net
- ✓ Explain independence in the three key structures

Next: d-separation

The three structures cover paths of length 2. For arbitrary paths in a Bayes net, we generalize to **d-separation** — the rule for reading off all (conditional) independences from any DAG.