CS 486/686
Probabilities

Yuntian Deng

Lecture 6

RN 12.2–12.3 · PM 8.1

Reminders

DUE TONIGHT Thu May 28, 11:59 PM

Chat 2 (Heuristic search & CSPs) on Chrysalis.

OUT TODAY Due Tue Jun 2

Chat 3 (Probabilities) — released today.

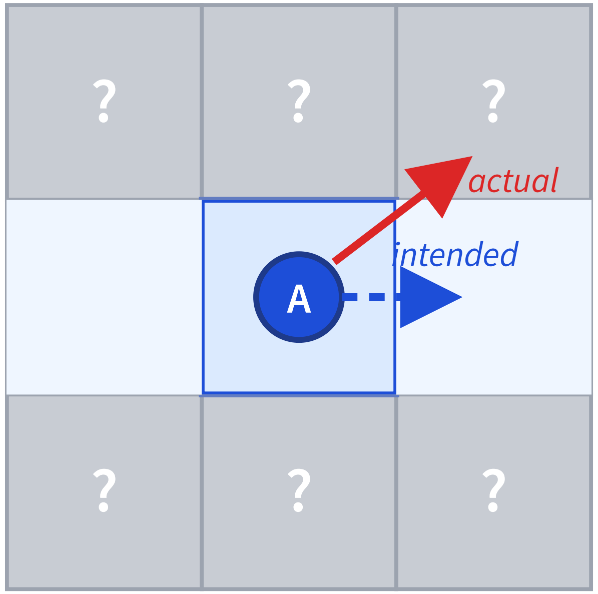
Learning goals

Calculate prior, posterior, and joint probabilities using:

- the **sum rule** (marginalization),
- the **product rule** (conditioning),
- the **chain rule** (joint factorization),
- and **Bayes' rule** (flipping a conditional).
- Then put it all together with a **universal 3-step recipe**.

Why handle uncertainty?

- Can't see the whole state — partial observation.
- Actions may not produce their intended outcome.
- **Reason about uncertainty** and decide anyway.



Visible cells are clear; the rest are hidden. Actions drift.

Frequentists vs Bayesians

Two camps for the formal measure of uncertainty.

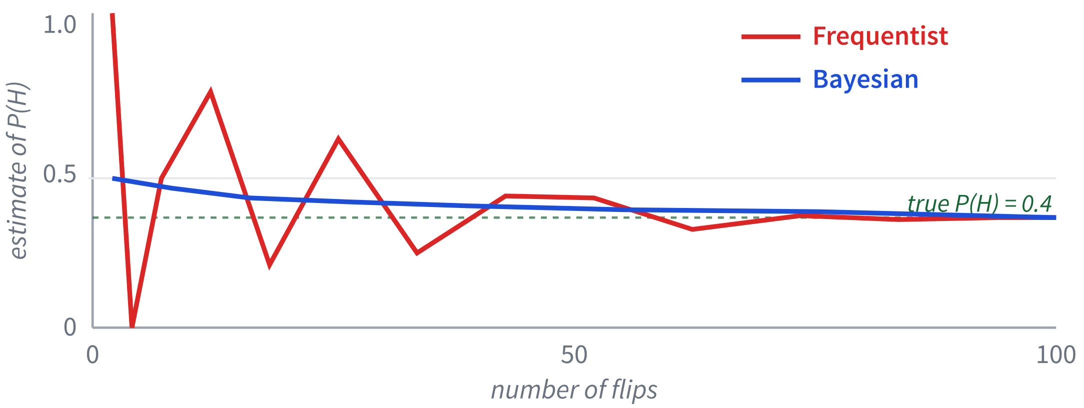
Frequentist

Long-run frequency. *Objective.* After n H, m T: $P(\mathrm{H}) = \frac{n}{n+m}$. Needs data.

Bayesian

Degree of belief: prior + evidence. Laplace 1H/1T: $P(\mathrm{H}) = \frac{1+n}{2+n+m}$.

Works with no data.

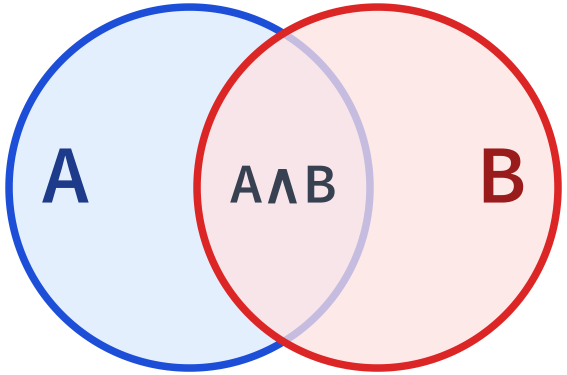


Random variables

- A *domain* of values + a *distribution* over them.
- **e.g.** $A =$ "alarm sounding": domain $\{T, F\}$, $P(A{=}T) = 0.1, P(A{=}F) = 0.9$
  .
- **Boolean shorthand:** $P(A) \equiv P(A{=}T)$; $P(\neg A) \equiv P(A{=}F)$.

Axioms of probability

- $0 \leq P(A) \leq 1$
- $P(\text{true}) = 1;\ P(\text{false}) = 0$
- **Inclusion-exclusion:** $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



$0 < P(A) < 1$ means we're *uncertain*, not that A is partly true.

Joint probability distribution

- **Atomic event:** an assignment to every variable.
- **Joint distribution:** a probability for every atomic event.

$$\text{e.g. } P(\text{weather}, \text{temperature})$$

	Hot	Mild	Cold
Sunny	0.10	0.20	0.10
Cloudy	0.05	0.35	0.20

Probabilities sum to 1 across the table.

Prior vs posterior

- **Prior** $P(X)$ – belief in X with no other info.
- **Posterior** $P(X \mid Y)$ – belief in X given evidence Y.

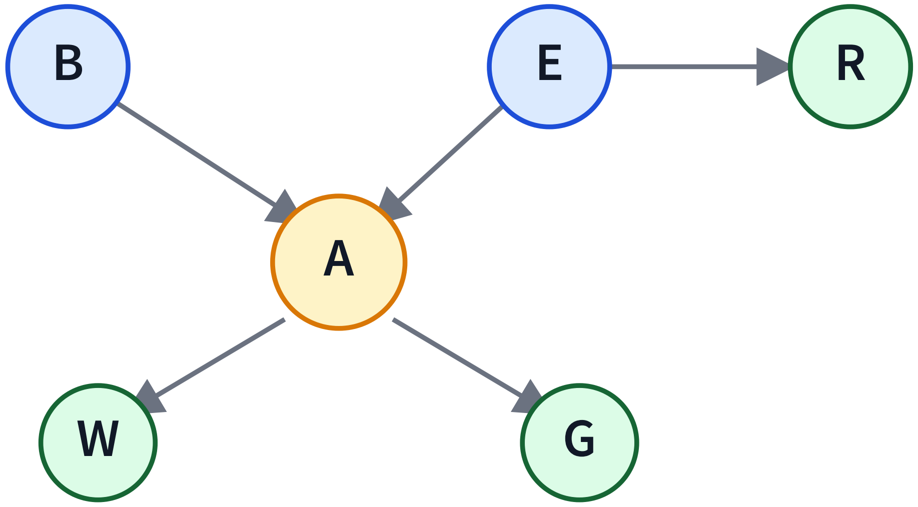
Running example: Holmes' burglar alarm

Mr. Holmes has a burglar alarm and two flaky neighbours (Watson, Gibbon) who call when they hear it. The alarm also misfires on earthquakes; earthquakes are reported on the radio.

6 Boolean variables ⇒ $2^6 = 64$ joint probabilities.

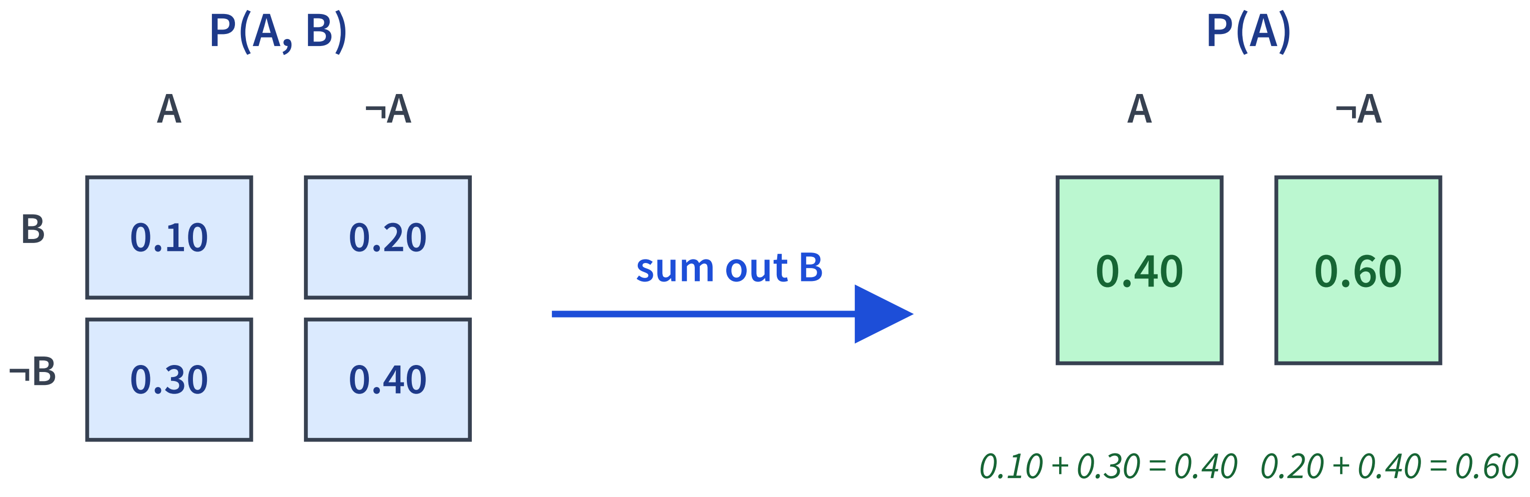
Random variables

- B: burglary
- E: earthquake
- A: alarm sounding
- W: Watson calling
- G: Gibbon calling



The sum rule (marginalization)

From a joint, get a probability over a *subset* by summing out the rest.



In general: $P(A) = \sum_b P(A \wedge B{=}b)$

Marginalizing the Holmes joint

$P(A, W, G)$ over Watson W, Gibbon G, and the alarm A:

	A		$\neg A$	
	G	$\neg G$	G	$\neg G$
W	0.032	0.048	0.036	0.324
$\neg W$	0.008	0.012	0.054	0.486

Example — $P(\neg A \wedge W)$: sum out G within the highlighted cells.

$$P(\neg A \wedge W) = 0.036 + 0.324 = \mathbf{0.36}$$

Quiz pack: marginalization

Q1. $P(\neg A \wedge W)$?

> $= 0.036 + 0.324 = \mathbf{0.36}$.

Q2. $P(A \wedge \neg G)$?

> $= 0.048 + 0.012 = \mathbf{0.06}$.

Q3. $P(\neg A)$?

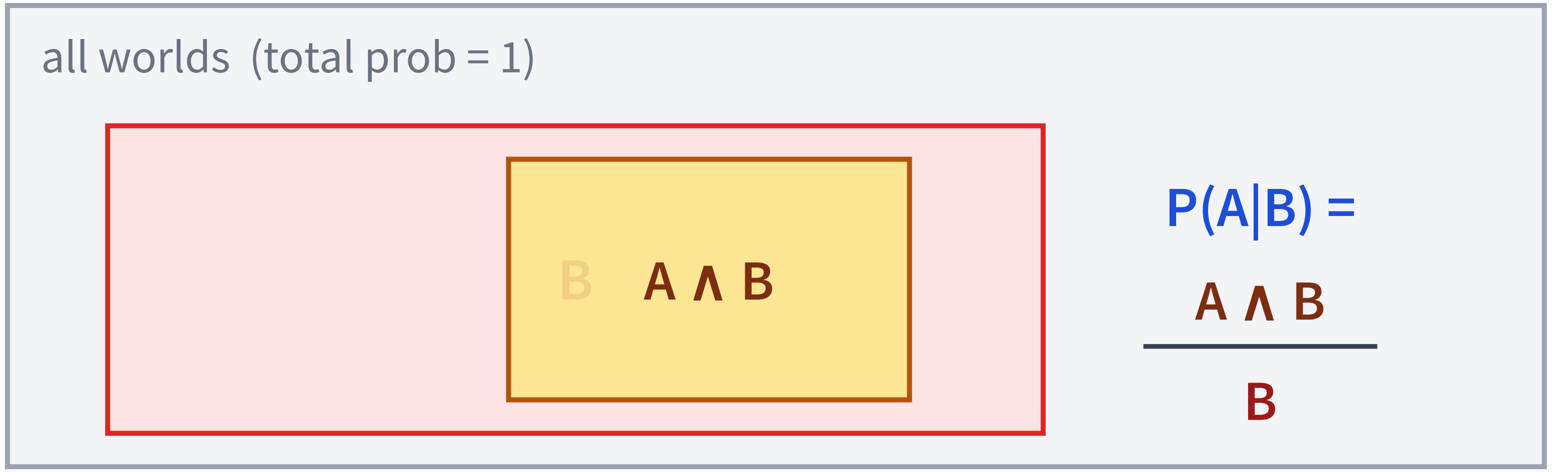
> $= 0.036 + 0.324 + 0.054 + 0.486 = \mathbf{0.9}$.

	A		$\neg A$	
	G	$\neg G$	G	$\neg G$
W	0.032	0.048	0.036	0.324
$\neg W$	0.008	0.012	0.054	0.486

Conditional probability and the product rule

$P(A|B)$ = fraction of the B-world where A also holds.

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Multiply both sides by $P(B)$:

$$P(A \wedge B) = P(A|B)\,P(B)$$

Q: $P(W \mid \neg A)$?

From the previous slides: $P(\neg A \wedge W) = 0.36, P(\neg A) = 0.9.$

Q. Watson calls, given the alarm is NOT going.

A. 0.2

B. 0.4

C. 0.6

D. 0.8

> **B.** $P(W|\neg A) = \dfrac{P(\neg A \wedge W)}{P(\neg A)} = \dfrac{0.36}{0.9} = \mathbf{0.4}.$

Q: $P(\neg G \mid A)$?

From the previous slides: $P(A \wedge \neg G) = 0.06, P(\neg A) = 0.9$ (so $P(A) = 0.1$).

Q. Gibbon does NOT call, given the alarm is going.

A. 0.2

B. 0.4

C. 0.6

D. 0.8

C. $P(\neg G|A) = \dfrac{P(A \wedge \neg G)}{P(A)} = \dfrac{0.06}{0.1} = \mathbf{0.6}$.

The chain rule (joint factorization)

The product rule, repeated.

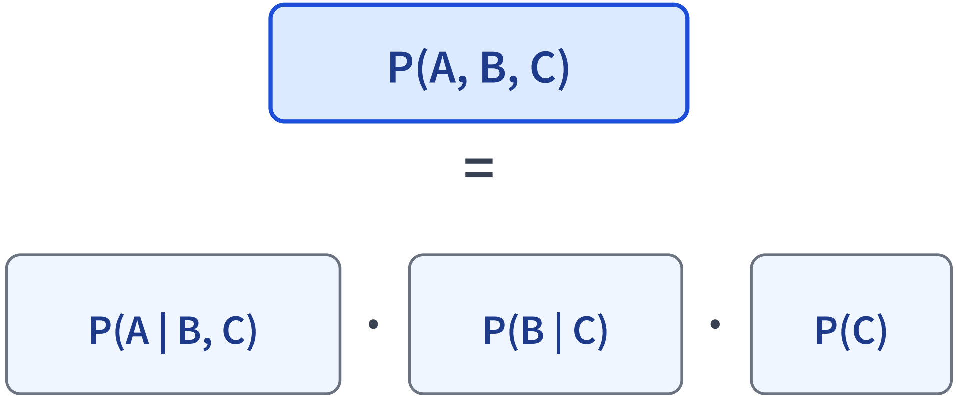
2 vars: $P(A, B) = P(A|B)\,P(B)$

3 vars:

$$P(A, B, C) = P(A|B, C)\,P(B|C)\,P(C)$$

In general:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$



condition each on its predecessors
(any ordering works)

Q: $P(A \wedge W \wedge \neg G)$?

Given: $P(A) = 0.1, P(W|A) = 0.9, P(G|A \wedge W) = 0.3.$

Q. Apply the chain rule to the conjunction.

A. 0.027

B. 0.037

C. 0.054

D. 0.063

D. $P(A) \cdot P(W|A) \cdot P(\neg G|A \wedge W) = 0.1 \times 0.9 \times 0.7 = \mathbf{0.063}.$

Q: $P(\neg A \wedge \neg W \wedge \neg G)$?

Given: $P(A) = 0.1, P(W|\neg A) = 0.4, P(G|\neg A \wedge \neg W) = 0.1.$

Q. Apply the chain rule, this time to the all-negated conjunction.

A. 0.486

B. 0.324

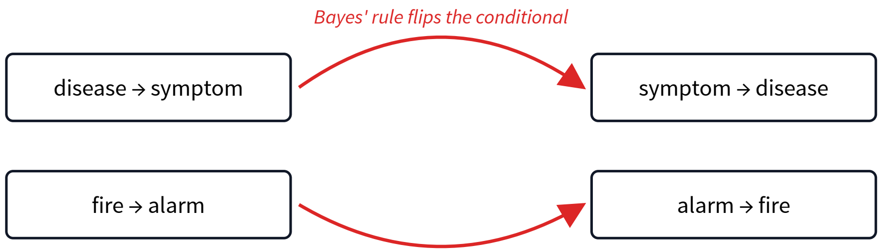
C. 0.216

D. 0.054

> **A.** $P(\neg A) \cdot P(\neg W|\neg A) \cdot P(\neg G|\neg A \wedge \neg W) = 0.9 \times 0.6 \times 0.9 = \mathbf{0.486}$.

Why flip conditionals?

Models give us **causal** knowledge. Diagnosis needs **evidential** reasoning.



We know $P(\text{symptom}|\text{disease})$ and want $P(\text{disease}|\text{symptom})$.

Bayes' rule

Bayes' rule. $$P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$

$$\frac{\text{Likelihood } P(Y \mid X) \times \text{Prior } P(X)}{\text{Evidence } P(Y)}$$

Posterior

P(X | Y)

=

Likelihood

P(Y | X)

×

Prior

P(X)

Evidence

P(Y)

normaliser

Derive it from the product rule: $P(X, Y) = P(X|Y)\,P(Y) = P(Y|X)\,P(X)$

→ divide by $P(Y)$.

Q: $P(\neg A \mid W)$?

Given: $P(A) = 0.1, P(W|A) = 0.9, P(W|\neg A) = 0.4.$

Q. Alarm NOT going, given Watson calls.

A. 0.2

B. 0.45

C. 0.8

D. 0.9

C. Marginalize: $P(W) = P(A)P(W|A) + P(\neg A)P(W|\neg A) = 0.09 + 0.36 = 0.45.$

Bayes: $P(\neg A|W) = \dfrac{P(\neg A)P(W|\neg A)}{P(W)} = \dfrac{0.36}{0.45} = \mathbf{0.8}.$

Q: $P(A \mid \neg G)$?

Given: $P(A) = 0.1, P(G|A) = 0.3, P(G|\neg A) = 0.1.$

Q. Alarm going, given Gibbon does NOT call.

A. 0.03

B. 0.08

C. 0.30

D. 0.70

B. Marginalize: $P(\neg G) = P(A)P(\neg G|A) + P(\neg A)P(\neg G|\neg A) = 0.07 + 0.81 = 0.88.$

Bayes: $P(A|\neg G) = \dfrac{P(A)P(\neg G|A)}{P(\neg G)} = \dfrac{0.07}{0.88} \approx \mathbf{0.08}.$

A universal approach

Three rules — one per kind of probability you need.

1. **To calculate a conditional probability** — write it as a ratio of two joint probabilities *(product rule, reversed)*: $P(X|Y) = \dfrac{P(X \wedge Y)}{P(Y)}$.
2. **To calculate a partial joint** (some variables missing) — sum over the missing variables to get full joints *(sum rule, reversed)*.
3. **To calculate a full joint** (all variables) — factor as a product of conditionals *(chain rule)*.

Worked example: $P(A|C)$

Given:

$$P(A) = 0.6$$
$$P(B|A) = 0.4,\ P(\neg B|\neg A) = 0.2$$
$$P(C|A,B) = 0.1,\ P(C|\neg A,B) = 0.2,\ P(C|A,\neg B) = 0.5,\ P(C|\neg A,\neg B) = 0.8$$

Step 1. Convert the conditional into a ratio of joints:

$$P(A|C) = \frac{P(A \wedge C)}{P(C)} = \frac{P(A \wedge C)}{P(A \wedge C) + P(\neg A \wedge C)}$$

Worked example: step 2 (marginalize)

From step 1: $P(A|C) = \dfrac{P(A \wedge C)}{P(A \wedge C) + P(\neg A \wedge C)}$. The numerator and denominator are partial joints — sum out the missing variable B.

Step 2. Sum out B:

$$P(A \wedge C) = P(A \wedge B \wedge C) + P(A \wedge \neg B \wedge C)$$
$$P(\neg A \wedge C) = P(\neg A \wedge B \wedge C) + P(\neg A \wedge \neg B \wedge C)$$

Now every term on the right is a full-joint probability, ready for the chain rule.

Worked example: step 3 + answer

Apply the chain rule $P(A \wedge B \wedge C) = P(C|A \wedge B)\,P(B|A)\,P(A)$ to each full joint.

$$P(A \wedge B \wedge C) = 0.1 \times 0.4 \times 0.6 = 0.024$$

$$P(A \wedge \neg B \wedge C) = 0.5 \times 0.6 \times 0.6 = 0.180$$

$$P(\neg A \wedge B \wedge C) = 0.2 \times 0.8 \times 0.4 = 0.064$$

$$P(\neg A \wedge \neg B \wedge C) = 0.8 \times 0.2 \times 0.4 = 0.064$$

$$P(A|C) = \frac{0.024 + 0.180}{0.024 + 0.180 + 0.064 + 0.064} = \frac{0.204}{0.332} \approx \mathbf{0.614}$$

Four rules at a glance

Rule	Formula	Use when
Sum	$P(A) = \sum_b P(A \wedge B{=}b)$	marginalizing out variables
Product	$P(A \wedge B) = P(A\|B)\, P(B)$	defining or applying a conditional
Chain	$P(X_1 \wedge \cdots \wedge X_n) = \prod_i P(X_i\|X_1 \wedge \cdots \wedge X_{i-1})$	computing a full joint from conditionals
Bayes'	$P(X\|Y) = \dfrac{P(Y\|X)\, P(X)}{P(Y)}$	flipping causal → evidential

Learning goals (recap)

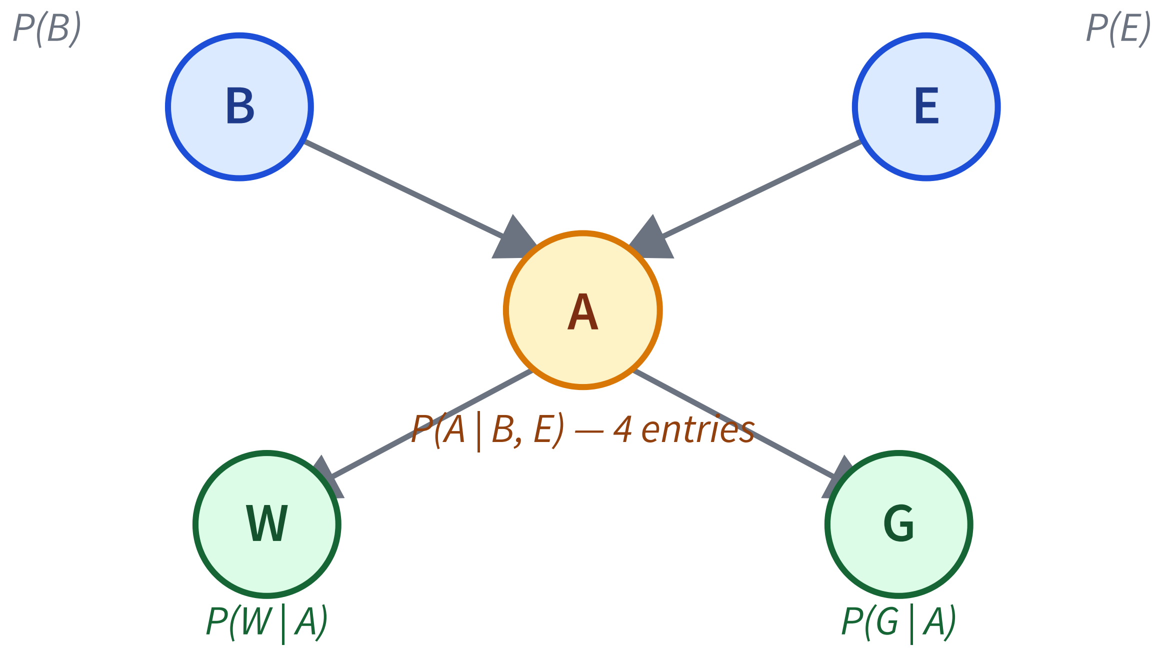
Calculate prior, posterior, and joint probabilities using:

- ✓ the **sum rule**
- ✓ the **product rule**
- ✓ the **chain rule**
- ✓ **Bayes' rule**
- ✓ the **universal 3-step recipe**

Next: independence and Bayes nets

Chain rule gives full joints — but 2^n numbers is too many.

Conditional independence + **Bayesian networks** → far fewer numbers.



5 variables: 32 joint numbers → ~10 CPT entries.