# L10 – Introduction to Graphical Models

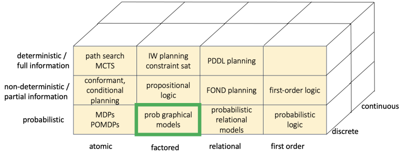
AIMA4e, 12.2-5, 13.1, 13.2.1 or KAlg 2.2-5

# What you should know after this lecture

- Framing of probabilistic inference problem
- How to model a distribution of variables as a factored distribution
- How to represent a factored distribution as a graphical model
- How (and why) to multiply and marginalise out random variables

# Probabilistic reasoning about partially-specified world states

Factored states
Discrete-valued factors
Probability over possible worlds!



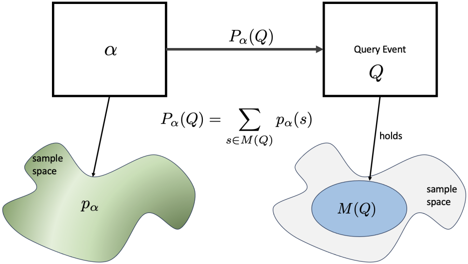

Could have done planning in probabilistic domains first, but the methods we develop now for talking about relationships between random variables will be helpful when we come to that.

# Probability reminder

Informal—but worth studying formally!

- $P(\{\ \}) = 0$
- $P(\mathcal{S}) = 1$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
- In general, $P(E_1 \cap E_2) \neq P(E_1)P(E_2)$
- $P(E_1 \mid E_2) = P(E_1 \wedge E_2)/P(E_2)$

# Probabilistic propositional inference




Given:

- Sample space $\mathcal{S}$ (possible worlds)
- Distribution $p_\alpha$ over $\mathcal{S}$ (pmf or pdf)
- Query $Q \subset \mathcal{S}$

What is $P_\alpha(Q)$?

$$P_\alpha(Q) = \sum_{s \in M(Q)} p_\alpha(s)$$

# Adding information

Conditioning on an event $E$ rules out all the other possible worlds.

$$p_{\alpha|E}(s) = \begin{cases} p_\alpha(s)/P_\alpha(E) & \text{if } s \in E \\ 0 & \text{otherwise} \end{cases}$$

# Probabilistic inference

Given $p_\alpha$ and $Q$ and possible $E$, compute $P_{\alpha|E}(Q)$

Stupidest possible algorithm:

- Enumerate $s \in \mathcal{S}$
- accumulate $p_{\alpha|E}(s)$ if $s \in M(Q)$

Our goal: do this without enumerating $\mathcal{S}$

Idea: use factored representation $\alpha$ to make compact representation of $p_\alpha$, $E$, and $Q$!

# Factored representation

- Random variables $V_1, \ldots, V_n$
- Each $V_i$ has discrete domain of possible values $\Omega_{V_i}$
- Sample space is product $\mathcal{S} = \Omega_{V_1} \times \ldots \times \Omega_{V_n}$
- Sample $s \in \mathcal{S}$ is $(v_1, \ldots, v_n)$ where $v_i \in \Omega_{v_i}$
- $p$ is the joint distribution on $V_1, \ldots, V_n$
- Can use a table $\alpha$ to represent $p$
- Use Boolean expressions over atoms $V = v$ to represent $Q$ and $E$

# Factored representation: example

- Random variables $A, B, C$
- Domains $\Omega = \{0, 1\}$

$\alpha =$

| $a$ | $b$ | $c$ | $p((a, b, c))$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.10 |
| 0 | 0 | 1 | 0.20 |
| 0 | 1 | 0 | 0.05 |
| 0 | 1 | 1 | 0.05 |
| 1 | 0 | 0 | 0.30 |
| 1 | 0 | 1 | 0.05 |
| 1 | 1 | 0 | 0.15 |
| 1 | 1 | 1 | 0.10 |

What is $P_\alpha(A = 1 \mid B = 0 \vee C = 0)$

# Bayes Nets: Compact factored representation of $p$

Define a Bayesian network $\alpha$ :

- Random variables $V_1, \dots, V_n$
- Each $V_i$ has discrete domain of possible values $\Omega_{V_i}$
- Directed acyclic graph $G$ defined on nodes $V_i$
- Parents $pa_G(V_i)$ : set of nodes $V_j$ with edges $(V_j, V_i) \in G$
- For each $V_i$, a conditional probability table (CPT), specifying $P(V_i \mid \text{parents}_G(V_i))$
  - For every assignment $\bar{v}$ to variables in $pa_G(V_i)$
  - and every value $v \in \Omega_{V_i}$
  - specify $P(V_i = v \mid pa_G(V_i) = \bar{v})$

Then for an assignment $s = (v_1, \dots, v_n)$

$$p_\alpha(s) = \prod_i P(V_i = v_i \mid pa_G(V_i) = s[pa_G(V_i)])$$

# Classic example

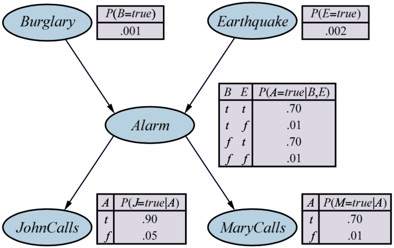



**Figure 13.2** A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters $B$, $E$, $A$, $J$, and $M$ stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

- Non-monotonicity of probability
  - What's $P_\alpha(B = 1)$?
  - What's $P_\alpha(B = 1 \mid M = 1)$?
  - What's $P_\alpha(B = 1 \mid M = 1, E = 1)$?
  - How many params to specify the whole joint as a table?

# Explaining away

Consider the network

```
Battery -> Gauge <- FuelTank
```

Here are some CPTs:

$$\Pr(B = 1) = 0.9$$
$$\Pr(F = 1) = 0.9$$
$$\Pr(G = 1 \mid B = 1, F = 1) = 0.8$$
$$\Pr(G = 1 \mid B = 1, F = 0) = 0.2$$
$$\Pr(G = 1 \mid B = 0, F = 1) = 0.2$$
$$\Pr(G = 1 \mid B = 0, F = 0) = 0.1$$

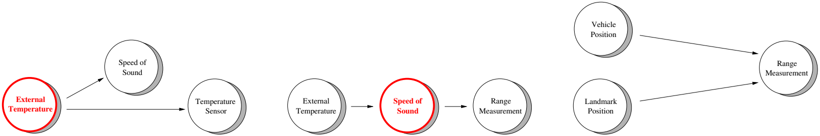
- What is the prior that the tank is empty? $\Pr(F = 0) = 0.1$
- What if we observe the fuel gauge and find that it reads empty? $\Pr(F = 0 \mid G = 0) \approx 0.257$
- Now, what if we find the battery is dead? $\Pr(F = 0 \mid G = 0, B = 0) \approx 0.111$ The probability that the tank is empty has decreased! Finding that the battery is flat explains away the empty fuel tank reading.

# Independence relations

Are we getting something for nothing?

- Independence of random variables: If $P(A = a, B = b) = P(A = a)P(B = b)$ for all $a \in \Omega_a, b \in \Omega_b$, we say that $A$ and $B$ are independent: $A \perp\!\!\!\perp B$.
- Conditional independence: If $P(A = a, B = b \mid C = c, D = d) = P(A = a \mid C = c, D = d)P(B = b \mid C = c, D = d)$ for all $a \in \Omega_A, b \in \Omega_B, c \in \Omega_C, d \in \Omega_d$, we say that $A$ and $B$ are conditionally independent given $C$ and $D$, $A \perp\!\!\!\perp B \mid C, D$.
- Bayes nets get their compactness from independence assumptions encoded in the graph.

# Graph structure encodes independence relations




- Case 1: $P(B|A), P(C|A)$ "outgoing" connection
  - $B \not\perp\!\!\!\perp C$, but $B \perp\!\!\!\perp C \mid A$
- Case 2: $P(B|A), P(C|B)$ "flow" connection
  - $C \not\perp\!\!\!\perp A$, but $C \perp\!\!\!\perp A \mid B$
- Case 3: $P(C|A, B)$ "incoming" connection
  - $A \perp\!\!\!\perp B$, but $A \not\perp\!\!\!\perp B \mid C$

In general $V_i \perp\!\!\!\perp V_j \mid E_1, \ldots, E_K$ if there are no paths from $V_i$ to $V_j$ through outgoing or flow connections that are not blocked by $E$ or through an incoming connection that is enabled by $E$. More about this when we get to factor graphs and Markov blankets.

# Simple inference algorithm

Given a BN, we have a conceptually (but not computationally) simple way to compute the joint

$$p_\alpha(s) = \prod_i P(V_i = v_i \mid \mathrm{pa}_G(V_i) = s[\mathrm{pa}_G(V_i)])$$

We can think of this as multiplying the CPTS in the Bayes net. Informally:

MULTIPLY$(D_1, D_2)$

$\pi$ = table indexed by $\Omega_{\mathrm{vars}(D_1) \cup \mathrm{vars}(D_2)}$
**for** $\bar{v}$ in $\pi$
    $\pi(\bar{v}) = \mathrm{lookup}(\bar{v}, D_1) \cdot \mathrm{lookup}(\bar{v}, D_2)$
**return** $\pi$

# Multiplication example

Given CPTs, $D_1 = P(X_2|X_1)$ and $D_2 = P(X_3|X_1)$, defined over different variable sets:

$D_1 =$

| $X_1$ | $X_2$ | P |
|---|---|---|
| T | T | 0.1 |
| T | F | 0.9 |
| F | T | 0.9 |
| F | F | 0.1 |

$D_2 =$

| $X_1$ | $X_3$ | P |
|---|---|---|
| T | T | 0.9 |
| T | F | 0.1 |
| F | T | 0.1 |
| F | F | 0.9 |

$\textsc{Multiply}(D_1, D_2) =$

| $X_1$ | $X_2$ | $X_3$ | P |
|---|---|---|---|
| T | T | T | $0.1 \times 0.9 = 0.09$ |
| T | T | F | $0.1 \times 0.1 = 0.01$ |
| T | F | T | $0.9 \times 0.1 = 0.09$ |
| T | F | F | $0.9 \times 0.9 = 0.81$ |
| F | T | T | $0.9 \times 0.9 = 0.81$ |
| F | T | F | $0.9 \times 0.1 = 0.09$ |
| F | F | T | $0.1 \times 0.1 = 0.01$ |
| F | F | F | $0.1 \times 0.9 = 0.09$ |

What is the meaning of this multiplication?

$P(X_2|X_1) \times P(X_3|X_1) = P(X_2, X_3|X_1)$.

# Next time

- We would like to avoid computing the whole joint distribution!!
- Algorithms whose complexity depends on the complexity of the network (rather than the product of the domains of all the variables)