Bayesian Networks and Inference

Bayesian Networks

Today:

- Bayesian Networks
- How do we perform exact inference on Bayesian Networks?
- How do we reason about independence in Bayesian Networks?

Review

Bayesian Network

Binary Random Variables X_1, X_2, X_3

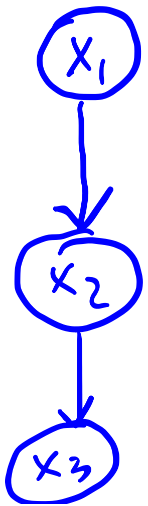
How many independent parameters to specify joint distribution? 7

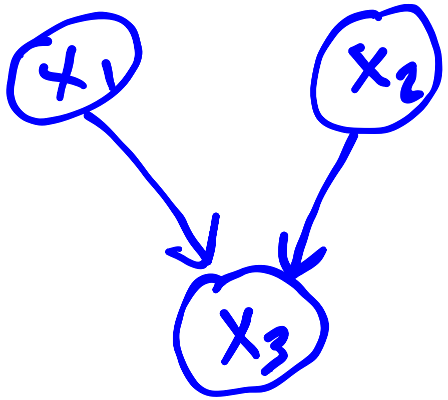
For n binary R.V.s, $2^n - 1$ independent parameters specify the joint distribution.

In general

$$\prod_{i=1}^{n} |\text{support}(X_i)| - 1$$

Bayesian Network: Directed Acyclic Graph (DAG) that represents a **joint probability distribution**



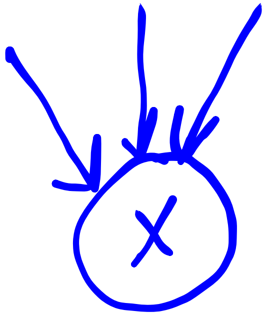


- Node: Random Variable
- Edges encode:

$$P(X_{1:n}) = \prod_{i=1}^{n} P(X_i \mid \text{pa}(X_i))$$

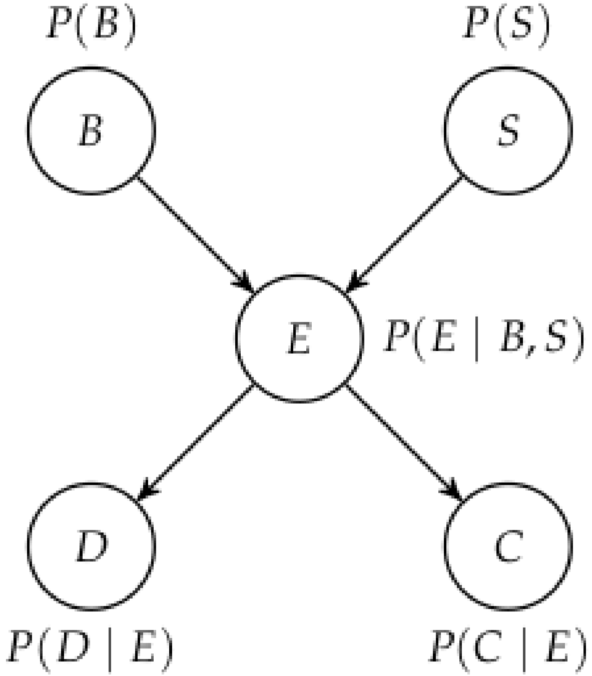
Full Story: "Causality: Models, Reasoning and Inference" by Judea Pearl

Counting Parameters

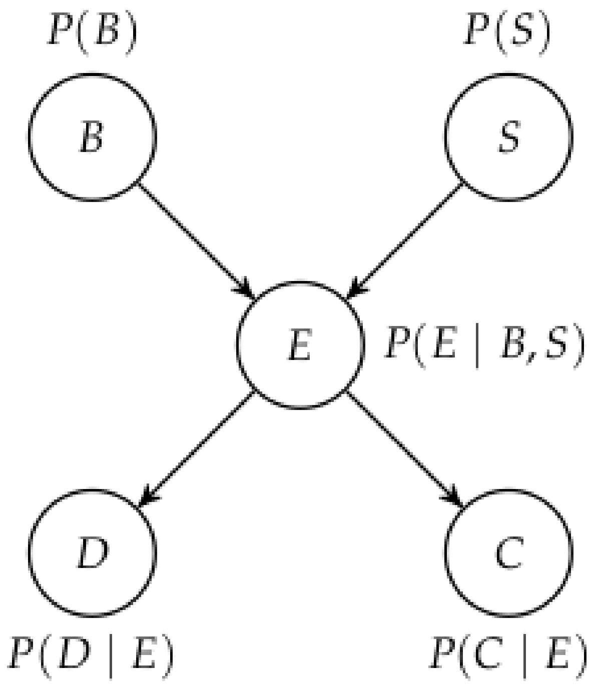


For discrete R.V.s:

$$\dim(\theta_X) = (|\text{support}(X)| - 1) \prod_{Y \in Pa(X)} |\text{support}(Y)|$$



Inference



B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

Inputs

- Bayesian network structure
- Bayesian network parameters
- Values of *evidence variables*

Outputs

- Posterior distribution of *query variables*

Given that you have detected a trajectory deviation, and the battery has not failed what is the probability of a solar panel failure?

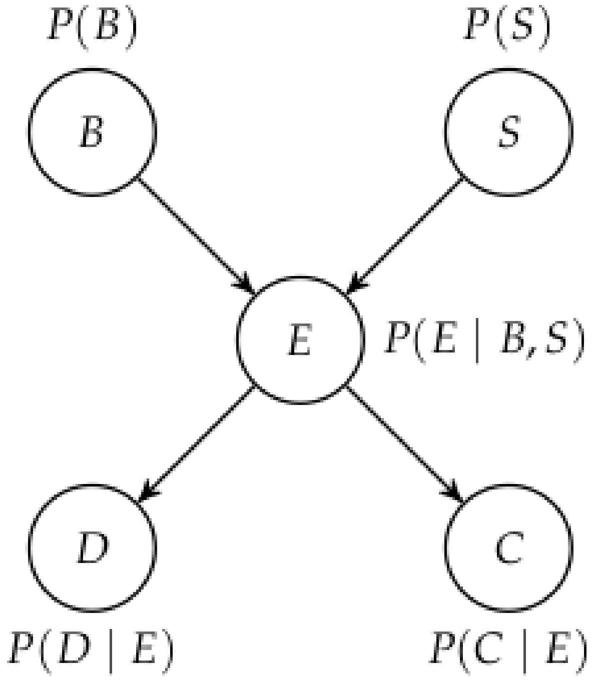
$$P(S = 1 \mid D = 1, B = 0)$$

Exact

Approximate

Exact Inference

Exact Inference



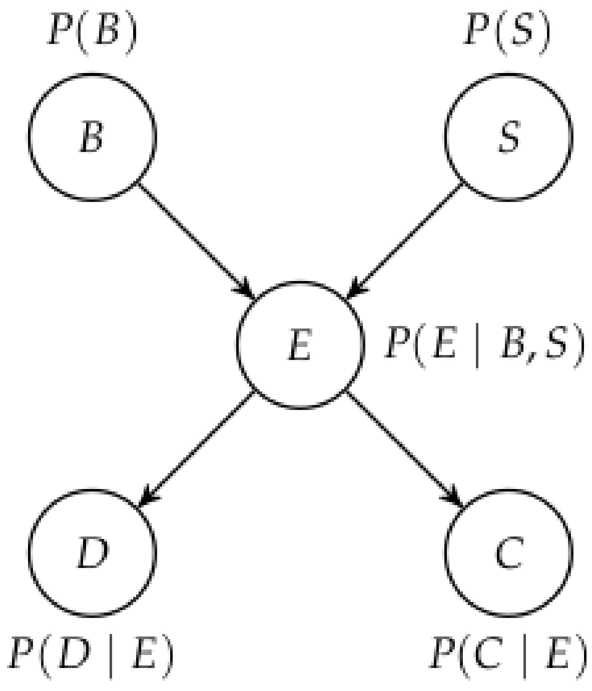
B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

$$P(S{=}1 \mid D{=}1, B{=}0) = \frac{P(S{=}1, D{=}1, B{=}0)}{P(D{=}1, B{=}0)}$$

$$P(S{=}1, D{=}1, B{=}0) = \sum_{e,c} P(B{=}0, S{=}1, E{=}e, D{=}1, C{=}c)$$

$$P(B{=}0, S{=}1, E, D{=}1, C)$$
$$= P(B{=}0)\, P(S{=}1)\, P(E \mid B{=}0, S{=}1)\, P(D{=}1 \mid E)\, P(C{=}1 \mid E)$$

Exact Inference



B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

Product

X	Y	$\phi_1(X,Y)$
0	0	0.3
0	1	0.4
1	0	0.2
1	1	0.1

Y	Z	$\phi_2(Y,Z)$
0	0	0.2
0	1	0.0
1	0	0.3
1	1	0.5

X	Y	Z	$\phi_3(X,Y,Z)$
0	0	0	0.06
0	0	1	0.00
0	1	0	0.12
0	1	1	0.20
1	0	0	0.04
1	0	1	0.00
1	1	0	0.03
1	1	1	0.05

```
struct ExactInference end

function infer(M::ExactInference, bn, query, evidence)
    ϕ = prod(bn.factors)
    ϕ = condition(ϕ, evidence)
    for name in setdiff(variablenames(ϕ), query)
        ϕ = marginalize(ϕ, name)
    end
    return normalize!(ϕ)
end
```

Condition

X	Y	Z	$\phi(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

$Y = 1$

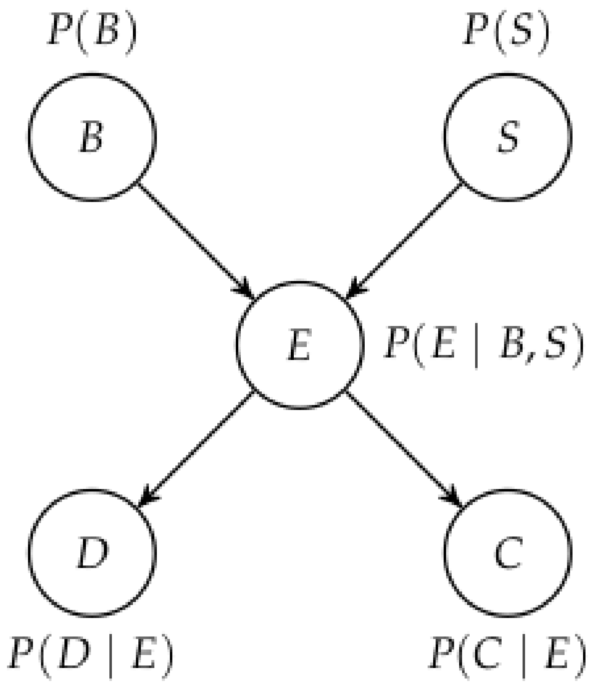
X	Z	$\phi(X,Z)$
0	0	0.09
0	1	0.37
1	0	0.02
1	1	0.07

Marginalize

X	Y	Z	$\phi(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

X	Z	$\phi(X,Z)$
0	0	0.17
0	1	0.68
1	0	0.03
1	1	0.12

Exact Inference: Variable Elimination



B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

$P(B \mid d^1, c^1)$

Start with

$\phi_1(B), \phi_2(S), \phi_3(E, B, S), \phi_4(D, E), \phi_5(C, E)$

Eliminate D and C (evidence) to get $\phi_6(E)$ and $\phi_7(E)$

Eliminate E

$$\phi_8(B, S) = \sum_e \phi_3(e, B, S)\phi_6(e)\phi_7(e)$$

Eliminate S

$$\phi_9(B) = \sum_s \phi_2(s)\phi_8(B, s)$$

$$P(B \mid d^1, c^1) \propto \phi_1(B) \sum_s \left(\phi_2(s) \sum_e \Big(\phi_3(e \mid B, s)\phi_4(d^1 \mid e)\phi_5(c^1 \mid e) \Big) \right)$$

vs

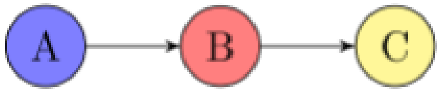
$$P(B \mid d^1, c^1) \propto \sum_s \sum_e \phi_1(B)\phi_2(s)\phi_3(e \mid B, s)\phi_4(d^1 \mid e)\phi_5(c^1 \mid e)$$

Choosing optimal order is NP-hard

Break

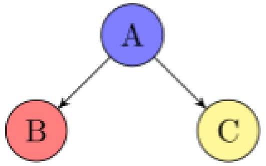
What does conditional independence mean?

$X \perp Y \mid Z \implies$ All of X's influence on Y comes through Z $\quad P(X \mid Z) = P(X \mid Y, Z)$



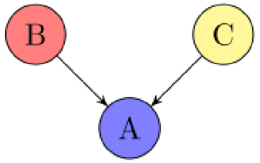
$A \perp C \mid B$? Yes

Mediator



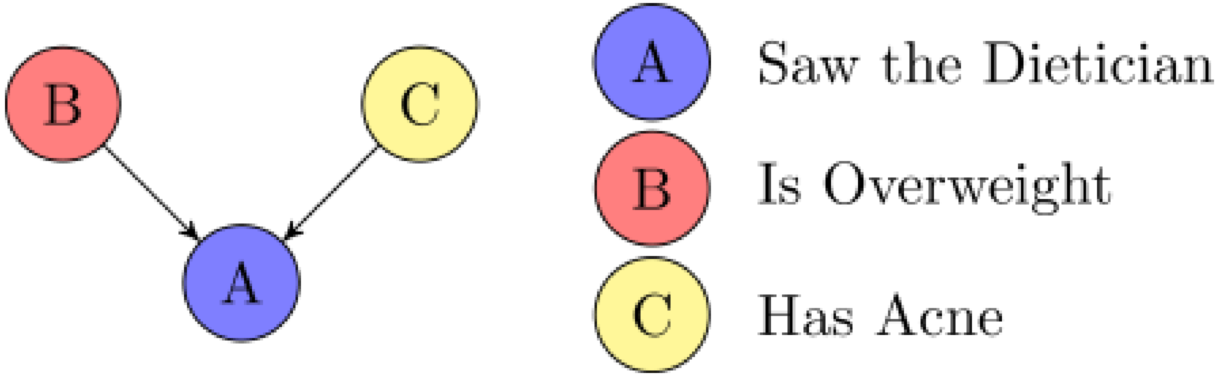
$B \perp C \mid A$? Yes

Confounder

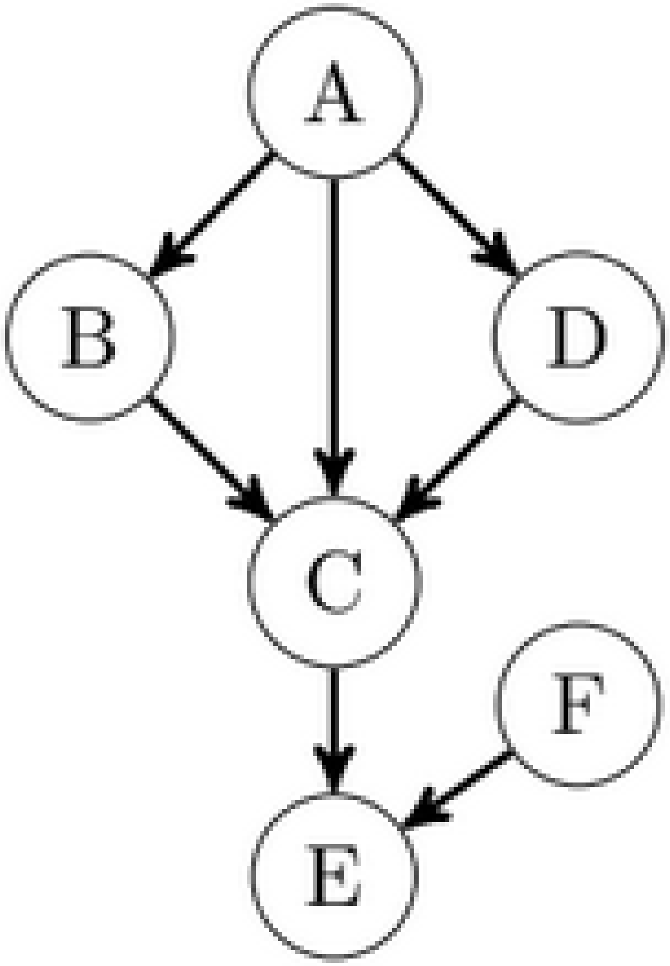


$B \perp C \mid A$? Inconclusive

Collider



More Complex Example



$(B \perp D \mid A)$?

Yes!

$(B \perp D \mid E)$?

No

Why is this relevant?

d-Separation

Let $\mathcal{C}$ be a set of random variables.

A *path* between A and B is *d-separated** by $\mathcal{C}$ if any of the following are true

1. The path contains a *chain* $X \rightarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
2. The path contains a *fork* $X \leftarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
3. The path contains an *inverted fork* (v-structure) $X \rightarrow Y \leftarrow Z$ such that Y is *not* in $\mathcal{C}$ and no descendant of Y is in $\mathcal{C}$.

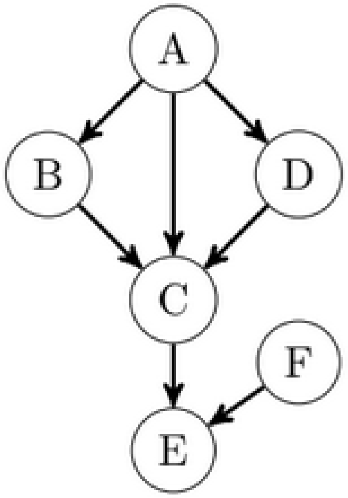
We say that A and B are *d-separated* by $\mathcal{C}$ if all paths between A and B are d-separated by $\mathcal{C}$.

If A and B are d-separated by $\mathcal{C}$ then $A \perp B \mid \mathcal{C}$

*short for "directionally separated"

Proving Conditional Independence

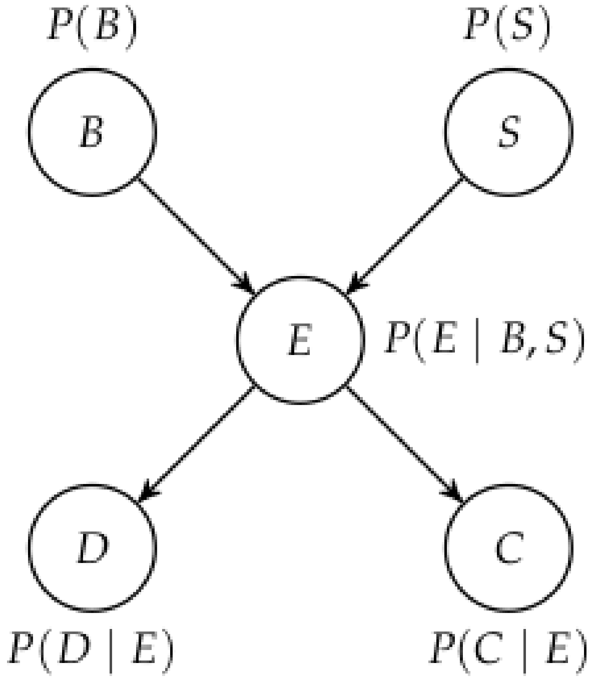
1. Enumerate all (non-cyclic) paths between nodes in question
2. Check all paths for d-separation
3. If all paths d-separated, then CE



Example: $(B \perp D \mid C, E)$?

1. The path contains a *chain* $X \rightarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
2. The path contains a *fork* $X \leftarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
3. The path contains an *inverted fork* (v-structure) $X \rightarrow Y \leftarrow Z$ such that $Y \notin \mathcal{C}$ and no descendant of Y is in $\mathcal{C}$.

Exercise



B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

$D \perp C \mid B$?

$D \perp C \mid E$?

1. The path contains a *chain* $X \rightarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
2. The path contains a *fork* $X \leftarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
3. The path contains an *inverted fork* (v-structure) $X \rightarrow Y \leftarrow Z$ such that $Y \notin \mathcal{C}$ and no descendant of Y is in $\mathcal{C}$.

Recap