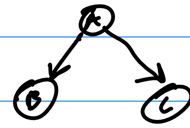
## Last Time

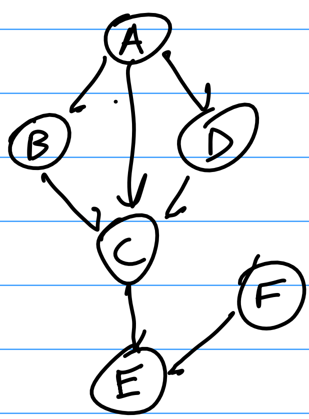
Bayesian Network - Joint Distribution



d-separation — If all paths between A and B are d-separated by $\mathcal{G}$ then $A \perp B \mid \mathcal{G}$

## This Time

- Sampling
- Inference
- Learning



$B \perp D \mid A$ ?

$\mathcal{G} = \{A\}$

| Path | d-separated |
| --- | --- |
| $B \leftarrow A \rightarrow D$ | yes |
| $B \rightarrow C \leftarrow D$ | yes |
| $B \rightarrow C \leftarrow A \rightarrow D$ | yes |

Since all paths are d-separated

$B \perp D \mid A$ https://kunalmenda.com/2019/02/21/causation-and-corr

$\mathcal{G}=\{C\}$ $B \perp D \mid C$ ? not true

https://kunalmenda.com/2019/02/21/causation

$B \leftarrow A \rightarrow D$ no

A,B binary

# Sampling

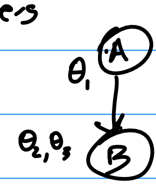
structure → $G$, distribution parameters → $\theta$

Given: Bayesian Network $G$ $\theta$

Output: Sample from joint distribution

$x_{1:n}$

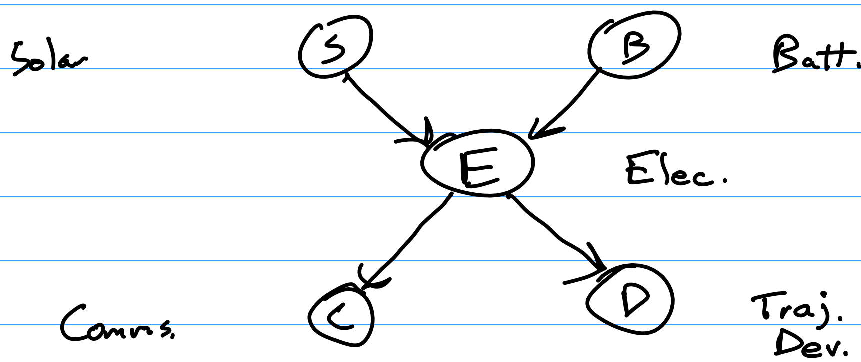
$(a, b)$



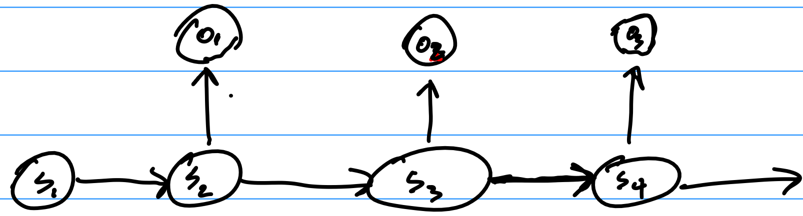

3 parameters

Step 1: Topological Sort: if $A \rightarrow B$ then A is before B

Step 2: Sample R.V.s in order




| | | |
|---|---|---|
| S | 1 | |
| B | 0 | |
| E | 1 | $P(E \mid S,B)$ |
| C | 0 | $P(C \mid E)$ |
| D | 1 | $P(D \mid E)$ |



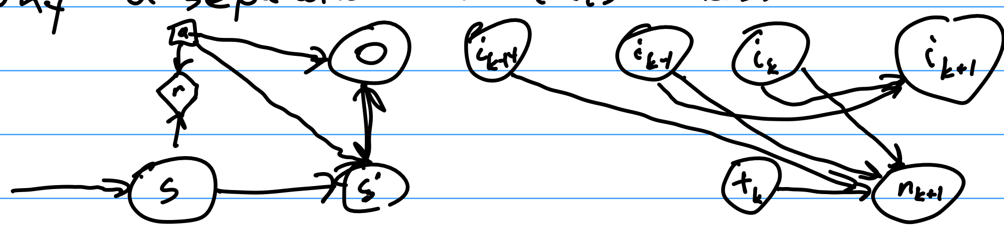

BN ⟺ POMDP

Sampling ⟺ Simulation

---

## Jump back

Why d-separation in this Class




## Markov Property

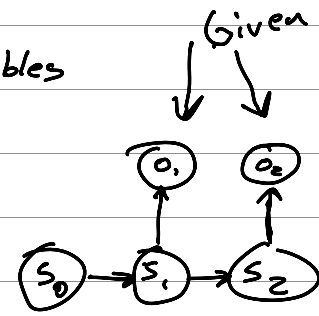
$s_k$ d-separates $s_{k+1}$ from $s_{k-i}$ $\forall\, i > 0$

# Inference

Given: Bayesian Network $G, \theta$

Values of some variables

Output: Distributions of Target Variables




$$\frac{\text{BN}}{\text{Inference}} \Longleftrightarrow \frac{\text{POMDP}}{\text{Belief Update}}$$

Trivial Case : Know upstream

Know : $S=1$, $B=1$

Infer : $P(C \mid S=1, B=1)$

Chain Rule

$$P(C \mid S,B) = \frac{P(C,S,B)}{P(S,B)}$$

$$P(C,S,B) = \sum_e P(B)\,P(S)\,P(E=e \mid S,B)\,P(C \mid E=e)$$

Harder Case : Know: $C=1$, $D=1$

Infer : $P(B \mid C=1, D=1)$

Inference

Exact

Approximate

## Exact

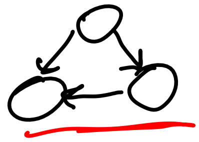
$$P(B=1 \mid C=1, D=1) = \frac{P(B=1, C=1, D=1)}{P(D=1, C=1)}$$

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

marginalization

chain rule

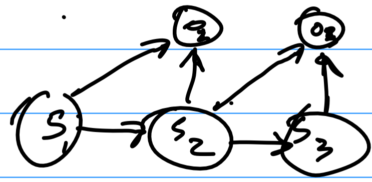
$$= \sum_{e,s} P(B=1)\,P(S=s)\,P(E=e \mid B=1, S=s)\,P(D=1 \mid E=e)\,P(C=1 \mid E=e)$$



choosing order

cycles

1. Sum-Product Variable Elimination ← hard to choose optimal ordering
2. Belief Propagation ← Efficient if no undirected cycles



Exact Inference on Bayesian Network is NP-hard

---

## Approximate Inference

### Method 1: Direct Sampling

1. Sample $(b^{(i)}, s^{(i)}, e^{(i)}, d^{(i)}, c^{(i)})$
2. Count how many match

$$P(B=1 \mid C=1, D=1) = \frac{\sum_i 1(b^{(i)}=1 \wedge d^{(i)}=1 \wedge c^{(i)}=1)}{\sum_i 1(d^{(i)}=1 \wedge c^{(i)}=1)}$$

| | | B | S | E | D | C | |
|---|---|---|---|---|---|---|---|
| 1 | | 0 | 0 | 1 | 1 | 0 | |
| 2 | | 0 | 0 | 0 | 0 | 1 | |
| 3 | X | 0 | 0 | 1 | 1 | 1 | ← |
| 4 | | 1 | 0 | 1 | 0 | 1 | |
| 5 | X | 0 | 1 | 0 | 1 | 1 | ← |
| 6 | | 1 | 0 | 1 | 1 | 0 | |

$$P(B=1 \mid C=1, D=1) = \frac{0}{2} = 0$$

Low probability events have low chance of getting sampled

BN ⟷ POMDP

Direct Sampling ⟷ Particle Filter (Unweighted)

### Method 2: Likelihood Weighted Sampling

1. Topological Sort
2. Fix known variables ←
3. Sample unknown variables
4. Calculate Weights
5. Count up weights for matches

$$P(B=1 \mid D=1, C=1) = \frac{\sum_i w_i 1(b^{(i)}=1)}{\sum_i w_i}$$

| B | S | E | D | C | w |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 1 | $P(D=1 \mid E=1)\,P(C=1 \mid E=1)$ |
| 0 | 0 | 0 | 1 | 1 | $P(D=1 \mid E=0)\,P(C=1 \mid E=0)$ |
| 1 | 0 | 1 | 1 | 1 | $P(D=1 \mid E=1)\,P(C=1 \mid E=1)$ |

$$P(B=1 \mid D=1, C=1) = \frac{w_3}{w_1 + w_2 + w_3}$$

BN $\Longleftrightarrow$ POMDP

Weighted Sampling $\Longleftrightarrow$ Weighted Particle Filtering

Gibbs Sampling — Markov Chain Monte Carlo

Learning

Given : Data

Book

Output : B.N. $G, \theta$