Causal Bayesian Networks

Causal Bayesian Networks

Today:

- Causal Bayesian Networks
- How do we reason about independence in Bayesian Networks?

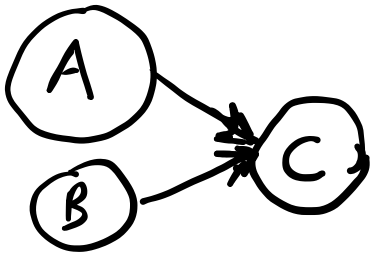
Review: Distributions of Discrete R.V.s

Joint	$P(X = x, Y = y)$ Shorthand: $P(x, y)$	Single number	"Probability that $X = x$ *and* $Y = y$"
	$P(X, Y)$	A table	"Joint distribution of X *and* Y"
Conditional	$P(X = x \mid Y = y)$ Shorthand: $P(x \mid y)$	Single number	"Probability that $X = x$ *if* $Y = y$"
	$P(X \mid Y)$	A collection of tables for each y	"Conditional distribution of X given Y"
Marginal	$P(X = x)$ Shorthand: $P(x)$	Single number	"Probability that $X = x$"
	$P(X)$	A table	"Marginal distribution of X"

Causal Bayesian Networks

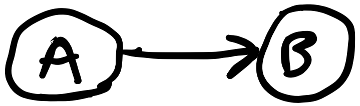
A *Bayesian Network* compactly represents a joint probability distributions using two components:

1. **Structure**: a directed acyclic graph (DAG), where each node is a R.V.
2. **Parameters:** Numerical values that determine a conditional distribution at each node



At each node, $P(X \mid pa(X))$

In a *Causal Bayesian Network*, arrows denote causation.



B is a result of A (and some aleatory uncertainty)

Chain rule for Bayesian Networks

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

Simple Causal Bayes Net Example

Naive Inference on Bayes Nets

Bayes Net with 3 Random Variables: $A \to C \to B$

Want to find $P(\underbrace{A = a}_{\text{query}} \mid \underbrace{B = b}_{\text{evidence}})$

C is a "hidden variable"

1. $P(A, B = b, C) = P(B = b \mid C)\, P(C \mid A)\, P(A)$.
2. Marginalize over hidden and query variables to get

$$P(A = a, B = b) = \sum_c P(A = a, B = b, C = c)$$

and

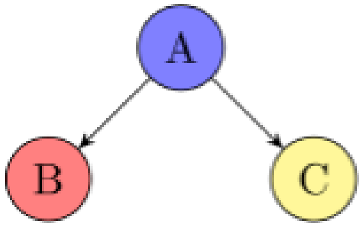
$$P(B = b) = \sum_{a,c} P(A = a, B = b, C = c)$$

3. $P(A = a \mid B = b) = \frac{P(A=a, B=b)}{P(B=b)}$

(Book introduces unnormalized "factors", but process is the same.)

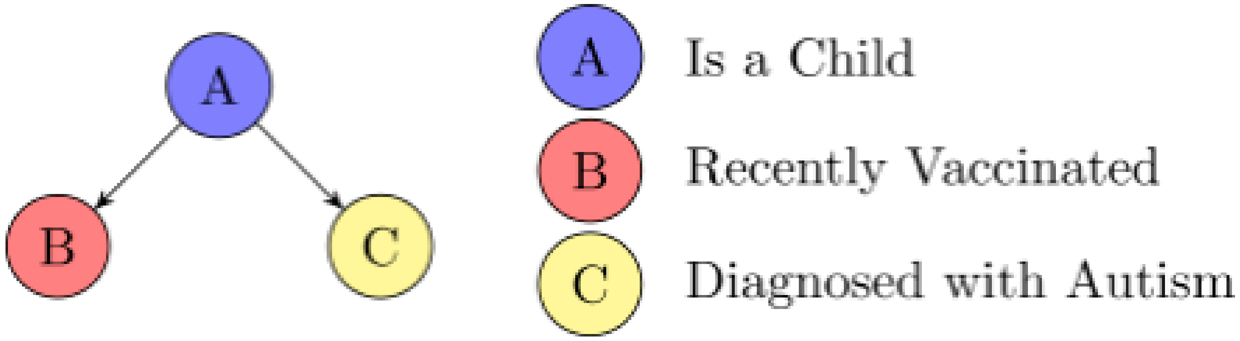
Conditional Independence in Bayes Nets

Conditional Independence: Fork



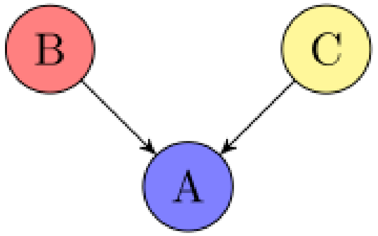
$B \perp C \mid A$?

Yes

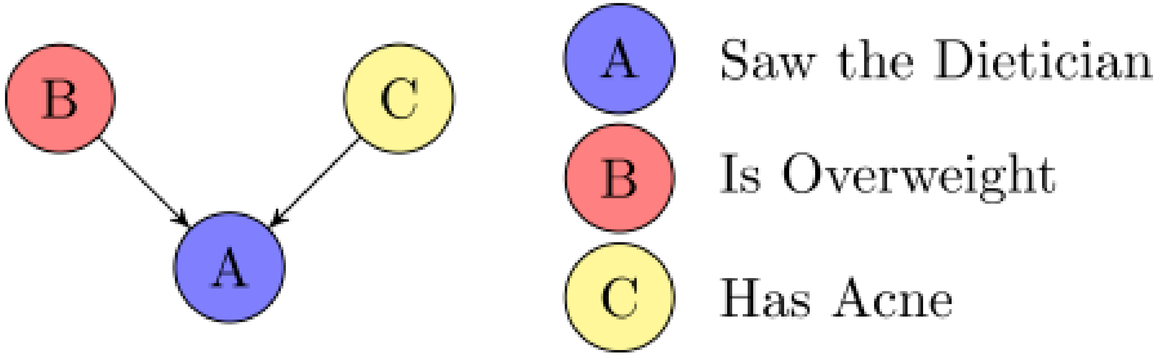


https://kunalmenda.com/2019/02/21/causation-and-correlation/

Conditional Independence: Inverted Fork

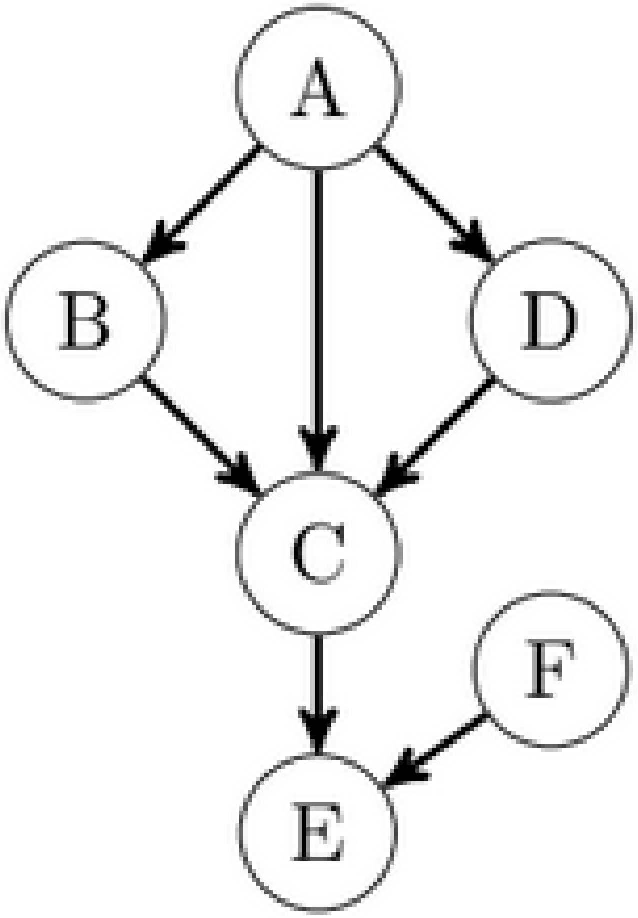


$$B \perp C \mid A \ ?$$



https://kunalmenda.com/2019/02/21/causation-and-correlation/

More Complex Example



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

Yes!

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

Inconclusive

Why is this relevant to decision making?

d-Separation

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

An undirected *path* between A and B is *d-separated* by $\mathcal{C}$ if any of the following exist along the path:

Separators (a.k.a "inactive triples"):

1. **Chain:** $X \rightarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
2. **Fork:** $X \leftarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
3. **Inverted Fork (v-structure):** $X \rightarrow Y \leftarrow Z$ such that $Y \notin \mathcal{C}$ and no descendant of Y is in $\mathcal{C}$.

Also:

- d-separated = "inactive"
- not d-separated = "active"

Break

$B \leftarrow A \rightarrow C \leftarrow D$

$B \leftarrow C \rightarrow A \leftarrow D$

Are these paths d-separated by $\mathcal{C} = \{C\}$?

d-Separation for Bayes Nets

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

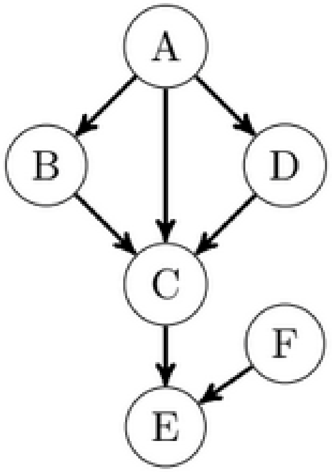
If A and B are d-separated by $\mathcal{C}$ then $A \perp B \mid \mathcal{C}$

In other words, if there is any active path w.r.t. $\mathcal{C}$ between A and B, we *cannot* conclude that $A \perp B \mid \mathcal{C}$ based on the structure alone.

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

Separators

1. **Chain:** $X \rightarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
2. **Fork:** $X \leftarrow Y \rightarrow Z$ such that $Y \in \mathcal{C}$
3. **Inverted Fork (v-structure):** $X \rightarrow Y \leftarrow Z$ such that $Y \notin \mathcal{C}$ and no descendant of Y is in $\mathcal{C}$.

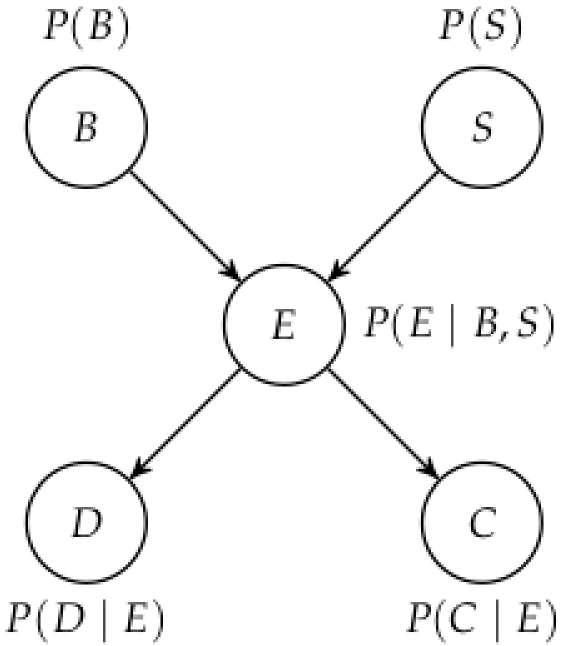
Markov Blanket

The *Markov Blanket* of $\mathcal{X}$ is the minimal set of nodes that, if their values were known, would make $\mathcal{X}$ conditionally independent of all other nodes.

A Markov blanket of a particular node consists of its parents, its children, and the other parents of its children.

If $\mathcal{B}$ is the Markov blanket of $\mathcal{X}$, you can treat analyze $\mathcal{B} \cup \mathcal{X}$ alone, and ignore any other nodes.

Exercise



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

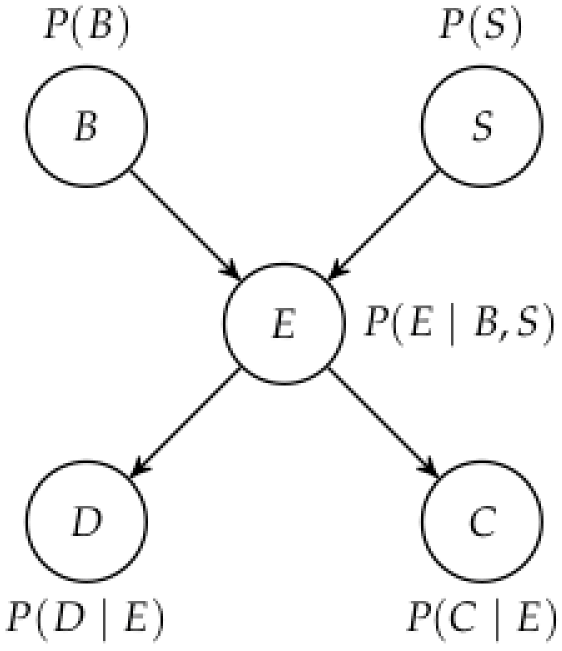
$$D \perp C \mid B\,?$$

$$D \perp C \mid E\,?$$

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

Approximate Inference

Approximate Inference: Direct Sampling



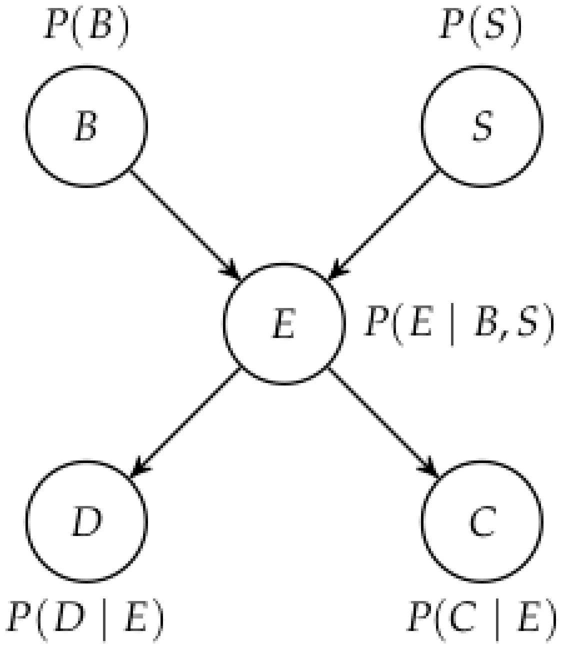
B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

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

B	S	E	D	C
0	0	1	1	0
0	0	0	0	0
1	0	1	0	0
1	0	1	1	1
0	0	0	0	0
0	0	0	1	0
0	0	0	0	1
0	1	1	1	1
0	0	0	0	0
0	0	0	1	0

Analogous to **unweighted particle filtering**

Approximate Inference: Weighted Sampling



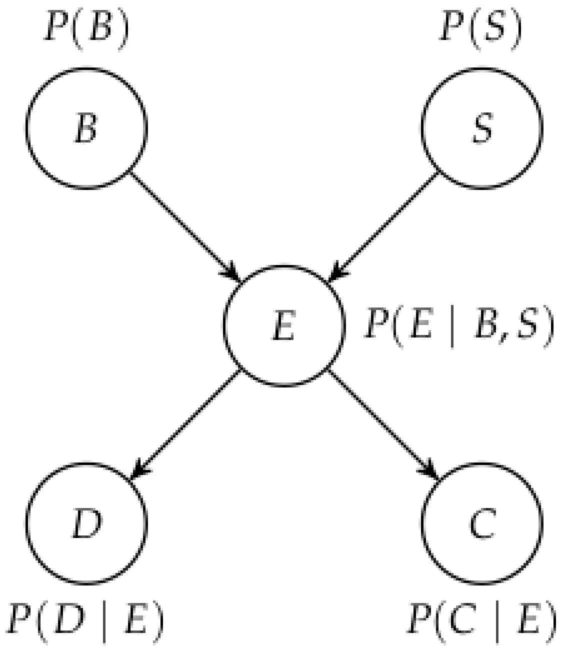
B battery failure
S solar panel failure
E electrical system failure
D trajectory deviation
C communication loss

$$P(b^1 \mid d^1, c^1) \approx \frac{\sum_i w_i(b^{(i)} = 1 \wedge d^{(i)} = 1 \wedge c^{(i)} = 1)}{\sum_i w_i(d^{(i)} = 1 \wedge c^{(i)} = 1)}$$
$$= \frac{\sum_i w_i(b^{(i)} = 1)}{\sum_i w_i}$$

B	S	E	D	C	Weight
1	0	1	1	1	$P(d^1 \mid e^1)P(c^1 \mid e^1)$
0	1	1	1	1	$P(d^1 \mid e^1)P(c^1 \mid e^1)$
0	0	0	1	1	$P(d^1 \mid e^0)P(c^1 \mid e^0)$
0	0	0	1	1	$P(d^1 \mid e^0)P(c^1 \mid e^0)$
0	0	1	1	1	$P(d^1 \mid e^1)P(c^1 \mid e^1)$

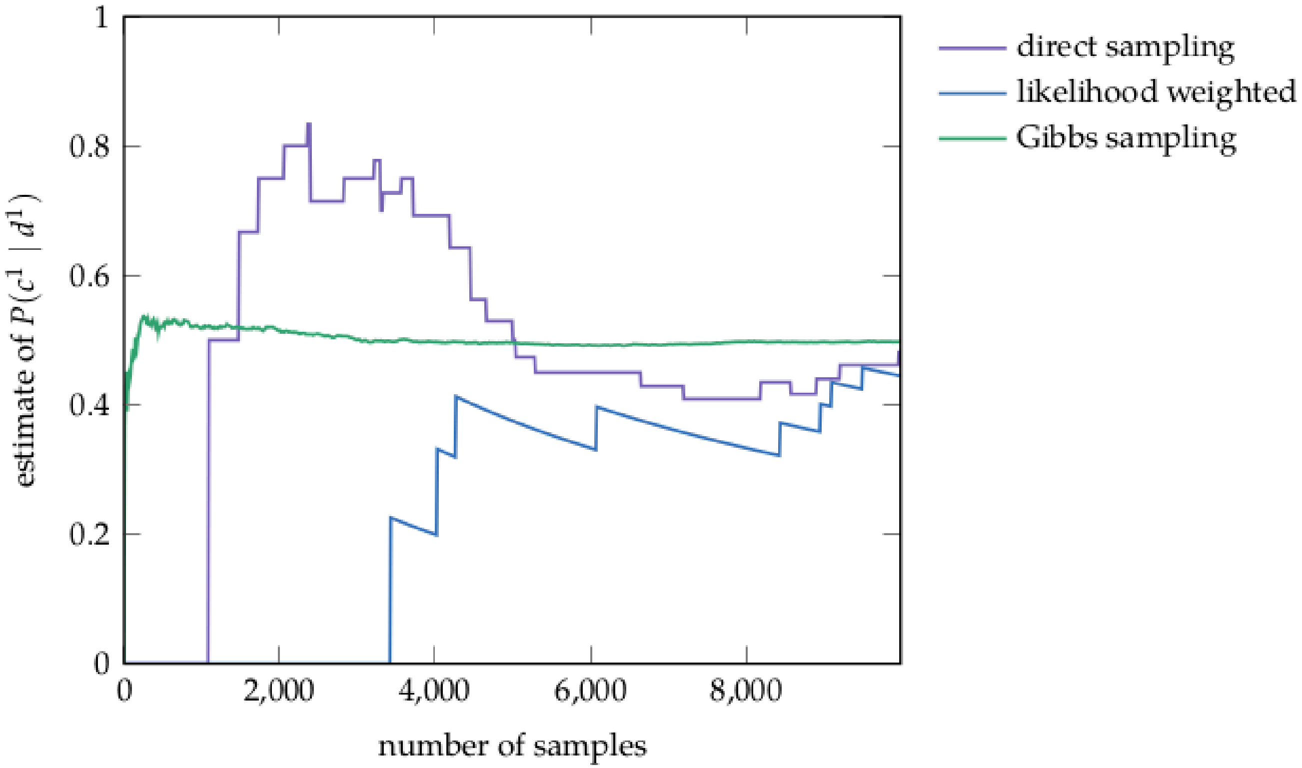
Analogous to **weighted particle filtering**

Approximate Inference: Gibbs Sampling



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

Markov Chain Monte Carlo (MCMC)



Recap