Bayesian networks

- Compact representation of distributions over large number of variables
- (Often) allows efficient exact inference (computing marginals, etc.)

HailFinder
- 56 vars
- \(~3\) states each
- \(\sim 10^{26}\) terms
- \(> 10,000\) years on Top supercomputers

JavaBayes applet
Causal parametrization

- Graph with directed edges from (immediate) causes to (immediate) effects

- Earthquake
- Burglary
- Alarm
- JohnCalls
- MaryCalls

\[
P(E) \quad \begin{array}{c|c}
E & P(E) \\
--- & --- \\
T & .01 \\
F & .99 \\
\end{array}
\]

\[
P(B) \quad \begin{array}{c|c}
B & P(B) \\
--- & --- \\
T & .03 \\
F & .97 \\
\end{array}
\]

\[
P(A | E, B) \quad \begin{array}{c|c|c|c}
A & E & B & P(A | E, B) \\
--- & --- & --- & --- \\
T & T & T & \quad .8 \\
T & T & F & \quad .2 \\
T & F & T & \quad .5 \\
F & T & T & \quad .1 \\
\end{array}
\]
A **Bayesian network structure** is a directed, acyclic graph $G$, where each vertex $s$ of $G$ is interpreted as a random variable $X_s$ (with unspecified distribution).

A **Bayesian network** $(G,P)$ consists of

- A BN structure $G$ and ..
- ..a set of conditional probability distributions (CPDs) $P(X_s \mid \text{Pa}_{X_s})$, where $\text{Pa}_{X_s}$ are the parents of node $X_s$ such that
- $(G,P)$ defines joint distribution

$$P(X_1, \ldots, X_n) = \prod_i P(X_i \mid \text{Pa}_{X_i})$$
Representing the world using BNs

True distribution $P'$ with cond. ind. $I(P')$

- Want to make sure that $I(P) \subseteq I(P')$
- Need to understand CI properties of BN $(G,P)$

Bayes net $(G,P)$ with $I(P)$

represent
Local Markov Assumption

Each BN Structure $G$ is associated with the following conditional independence assumptions:

$X \perp \text{NonDescendents}_X \mid \text{Pa}_X$

We write $I_{\text{loc}}(G)$ for these conditional independences.

Suppose $(G, P)$ is a Bayesian network representing $P$. Does it hold that $I_{\text{loc}}(G) \subseteq I(P)$?
If this holds, we say $G$ is an $I$-map for $P$. 
True distribution \( P \) can be represented exactly as Bayesian network \((G, P)\)

\[
P(X_1, ..., X_n) = \prod_i P(X_i \mid Pa_{X_i})
\]
Additional conditional independencies

BN specifies joint distribution through conditional parameterization that satisfies Local Markov Property
\[ I_{loc}(G) = \{(X_i \perp \text{Nondescendants}_{X_i} \mid Pa_{X_i})\} \]

But we also talked about additional properties of CI
- Weak Union, Intersection, Contraction, ...

Which additional CI does a particular BN specify?
- All CI that can be derived through algebraic operations

→ proving CI is very cumbersome!!

Is there an easy way to find all independences of a BN just by looking at its graph??
BNs with 3 nodes

Local Markov Property:
\[ X \perp \text{NonDesc}(X) \mid \text{Pa}(X) \]
V-structures

Earthquake \rightarrow \text{Alarm} \rightarrow \text{Burglary}

- Know $E \perp B$
- Suppose we know $A$. Does $E \perp B \mid A$ hold?

Can happen: $P(E = T \mid A = T, B = T) < P(E = T \mid A = T)$

Explaning away
BNs with 3 nodes

Indirect causal effect

\[ X \rightarrow Y \rightarrow Z \]

Indirect evidential effect

\[ X \rightarrow Y \rightarrow Z \]

Local Markov Property:
\[ X \perp \text{NonDesc}(X) \mid \text{Pa}(X) \]

Common cause

\[ X \rightarrow Y \rightarrow Z \]

\[ X \perp Z \mid Y \]
\[ \neg (X \perp Z) \]

Common effect

\[ X \rightarrow Y \rightarrow Z \]

\[ X \perp Z \]
\[ \neg (X \perp Z \mid Y) \]
Examples

A ⊥ F
A ⊥ I C
A ⊥ P (C, D)
More examples

A \perp G
A \perp G \text{ ID } x
A \perp G \text{ IE}
Active trails

When are A and I independent?

\[ P(A \land I) \] is neither B nor G observed and H and (D or F) observed.
Active trails

An undirected path in BN structure G is called **active trail** for observed variables $O \subseteq \{X_1, \ldots, X_n\}$, if for every consecutive triple of vars $X, Y, Z$ on the path:

- $X \rightarrow Y \rightarrow Z$ and $Y$ is unobserved ($Y \notin O$)
- $X \leftarrow Y \leftarrow Z$ and $Y$ is unobserved ($Y \notin O$)
- $X \leftarrow Y \rightarrow Z$ and $Y$ is unobserved ($Y \notin O$)
- $X \rightarrow Y \leftarrow Z$ and $Y$ or any of Y’s descendants is observed

Any variables $X_i$ and $X_j$ for which there is no active trail for observations $O$ are called d-separated by $O$

We write $d$-sep$(X_i; X_j \mid O)$

Sets $A$ and $B$ are d-separated given $O$ if $d$-sep$(X, Y \mid O)$ for all $X \in A$, $Y \in B$. Write $d$-sep$(A; B \mid O)$
**Theorem:**
\[ \text{d-sep}(X; Y \mid Z) \Rightarrow X \perp Y \mid Z \]

i.e., \( X \) cond. ind. \( Y \) given \( Z \) if there does not exist any active trail between \( X \) and \( Y \) for observations \( Z \).

- Proof uses algebraic properties of conditional independence.
Soundness of d-separation

- Have seen: $P$ factorizes according to $G \iff I_{\text{loc}}(G) \subseteq I(P)$
- Define $I(G) = \{(X \perp Y \mid Z): \text{d-sep}_G(X;Y \mid Z)\}$

**Theorem**: Soundness of d-separation

$P$ factorizes over $G \Rightarrow I(G) \subseteq I(P)$

- Hence, d-separation captures only true independences

- How about $I(G) = I(P)$?
Does the converse hold?

Suppose $P$ factorizes over $G$.

Does it hold that $I(P) \subseteq I(G)$?

$P \sim X \perp Y \quad I(P) = \{X \perp Y\}$

$G: \bigcirc \rightarrow \bigcirc \quad I(G) = \{\}$
Existence of dependences for non-d-separated variables

**Theorem**: If $X$ and $Y$ are not d-separated given $Z$, then there exists some distribution $P$ factorizing over $G$ in which $X$ and $Y$ are dependent given $Z$

**Proof sketch**:

Pick active trail 
Parameterize CPDs along trail to create dependence
Everything else set to independent to avoid cancelling dependencies
Completeness of d-separation

Theorem: For “almost all” distributions \( P \) that factorize over \( G \) it holds that \( I(G) = I(P) \)

“almost all”: except for a set of distributions with measure 0, assuming only that no finite set of distributions has measure > 0

\[
\begin{align*}
P(X = T) &= \rho \\
P(Y = T | X = T) &= \alpha \\
P(Y = T | X = F) &= \varphi \\
P(Y | X) &= P(Y) \\
P(Y = T | X = T) &= \frac{P(Y = T)}{P(X = T)} \\
\tau &= \varphi \rho + \alpha (1-\rho) \\
\tau (1-\rho) &= q (1-\rho),
\end{align*}
\]

happens with prob. \( \sigma \)
Algorithm for d-separation

How can we check if $X \perp Y \mid Z$?

- Idea: Check every possible path connecting $X$ and $Y$ and verify conditions
- Exponentially many paths!!! 😞

Linear time algorithm:
Find all nodes reachable from $X$

1. Mark $Z$ and its ancestors
2. Do breadth-first search starting from $X$; stop if path is blocked

Have to be careful with implementation details (see reading)
Representing the world using BNs

True distribution $P'$ with cond. ind. $I(P')$

- Want to make sure that $I(P) \subseteq I(P')$
- Ideally: $I(P) = I(P')$
- Want BN that **exactly** captures independencies in $P'$!

Bayes net $(G,P)$ with $I(P)$
Lemma: Suppose $G'$ is derived from $G$ by adding edges
Then $I(G') \subseteq I(G)$

Proof:

\[ I_{loc}(G') \leq I_{loc}(G) \]

Completeness: $I(G) = \{ \text{all CIs derivable from } I_{loc}(G) \}$ using C1 properties

\[ \Rightarrow I(G') \leq I(G) \]
\[ \therefore \]

Thus, want to find graph $G$ with $I(G) \subseteq I(P)$ such that when we remove any single edge, for the resulting graph $G'$ it holds that $I(G') \not\subseteq I(P)$

Such a graph $G$ is called **minimal I-map**
Existence of Minimal I-Maps

Does every distribution have a minimal I-Map?

Yes: Start with full graph \( G_1, I(G) = \emptyset \)
Keep removing edges as long as
\( I(G) \leq I(\emptyset) \)
Algorithm for finding minimal I-map

- Given random variables and known conditional independences
- Pick ordering $X_1,...,X_n$ of the variables
- For each $X_i$
  - Find minimal subset $A \subseteq \{X_1,...,X_{i-1}\}$ such that
    $P(X_i \mid X_1,...,X_{i-1}) = P(X_i \mid A)$
  - Specify / learn CPD $P(X_i \mid A)$

Will produce minimal I-map!
Uniqueness of Minimal I-maps

Is the minimal I-Map unique?

\[ l(P) = l(G_r) \]
Perfect maps

- Minimal I-maps are easy to find, but can contain many unnecessary dependencies.

- A BN structure $G$ is called **P-map** (perfect map) for distribution $P$ if $I(G) = I(P)$

- Does every distribution $P$ have a P-map?
Existence of perfect maps

\[ X, Y \sim \text{Ber}(0.5) \]
\[ Z = X \text{ XOR } Y \]
\[ x \perp y, \quad y \perp z, \quad z \perp x \]
\[ \perp (x \perp y \perp z) \]

\( 1 \text{ map } \) Has no BN R-map

\[ \theta \rightarrow (1, 2) \]
\[ (2, 1) \rightarrow y \rightarrow z \]
Existence of perfect maps

\[ X_1, \ldots, X_9 \quad X_1 \perp X_3 \mid [X_2, X_9] \]

\[ X_2 \perp X_4 \mid X_1, X_3 \]

<UnDirected GM is a P-map

but NO BN is P-map
Uniqueness of perfect maps

\[ \emptyset \rightarrow \circ \quad G_1 \]

\[ \circ \leftarrow \bigcirc \quad G_2 \]

\[ \mathcal{I}(G_1) = \mathcal{I}(G_2) \]
I-Equivalence

- Two graphs $G, G'$ are called I-equivalent if $I(G) = I(G')$
- I-equivalence partitions graphs into equivalence classes
Skeletons of BNs

- I-equivalent BNs must have same skeleton

\[ \text{Same skeleton} \neq \text{i-equiv.} \]
Importance of V-structures

**Theorem**: If $G$, $G'$ have same skeleton and same V-structure, then $I(G) = I(G')$

Does the converse hold?

![Graphs]

$I(G) = \emptyset \quad \text{=} \quad I(G') = \emptyset$

*Same skeleton, Not same V-structures*
Immoralities and I-equivalence

A V-structure $X \rightarrow Y \leftarrow Z$ is called **immoral** if there is no edge between $X$ and $Z$ (“unmarried parents”)

**Theorem:** $I(G) = I(G') \iff G$ and $G'$ have the same skeleton and the same immoralities.
Tasks

- Subscribe to Mailing list
  https://utils.its.caltech.edu/mailman/listinfo/cs155

- Read Koller & Friedman Chapter 3.3-3.6

- Form groups and think about class projects. If you have difficulty finding a group, email Pete Trautman

- Homework 1 out tonight, due in 2 weeks. Start early!