CSC384: Lecture 10

- Last time
  - Inference and Independence
- Today
  - Reasoning under uncertainty (belief networks)
- Readings:
  - Today: 10.3 (note: d-separation not covered in text)
  - Next week: 10.3 (var. elim.), 10.4 (decision making)

Exploiting Cond. Ind. (Recap)

- Let’s see what conditional independence buys us
- Consider a story:
  - If Craig woke up too early E, Craig probably needs coffee C; if C, Craig needs coffee, he’s likely angry A. If A, there is an increased chance of an aneurysm (burst blood vessel) B. If B, Craig is quite likely to be hospitalized H.

Cond’l Ind. in our Story (Recap)

- If you learned any of E, C, A, or B, your assessment of Pr(H) would change.
  - E.g., if any of these are seen to be true, you would increase Pr(H) and decrease Pr(~H).
  - So H is not independent of E, or C, or A, or B.
- But if you knew value of B (true or false), learning value of E, C, or A, would not influence Pr(H). Influence these factors have on H is mediated by their influence on B.
  - Craig doesn’t get sent to the hospital because he’s angry, he gets sent because he’s had an aneurysm.
  - So H is independent of E, and C, and A, given B

Cond’l Ind. in our Story (Recap)

- So H is independent of E, and C, and A, given B
- Similarly:
  - B is independent of E, and C, given A
  - A is independent of E, given C
- This means that:
  - Pr(H | B, A, C, E) = Pr(H|B)
    - i.e., for any subset of {A,C,E}, this relation holds
  - Pr(B | A, C, E) = Pr(B | A)
  - Pr(A | C, E) = Pr(A | C)
  - Pr(C | E) and Pr(E) don’t “simplify”

Cond’l Ind. in our Story (Recap)

- By the chain rule (for any instantiation of H…E):
- By our independence assumptions:
  - Pr(H,B,A,C,E) = Pr(H|B) Pr(B|A) Pr(A|C) Pr(C|E) Pr(E)
- We can specify the full joint by specifying five local conditional distributions: Pr(H|B); Pr(B|A); Pr(A|C); Pr(C|E); and Pr(E)

Example Quantification

- Specifying the joint requires only 9 parameters (if we note that half of these are “1 minus” the others), instead of 31 for explicit representation
  - linear in number of vars instead of exponential!
  - linear generally if dependence has a chain structure
Inference is Easy

\[
P(a) = \sum_{c_i \in \text{Dom}(C)} \Pr(a | c_i) \Pr(c_i) \\
= \sum_{c_i \in \text{Dom}(C)} \Pr(a | c_i) \sum_{e_j \in \text{Dom}(E)} \Pr(c_i | e_j) \Pr(e_j)
\]

These are all terms specified in our local distributions!

Bayesian Networks

The structure above is a Bayesian network. A BN is a graphical representation of the direct dependencies over a set of variables, together with a set of conditional probability tables (CPTs) quantifying the strength of those influences.

Bayes nets generalize the above ideas in very interesting ways, leading to effective means of representation and inference under uncertainty.

An Example Bayes Net

A couple CPTs are “shown”

Explicit joint requires 211 - 1 = 2047 params

BN requires only 27 params (the number of entries for each CPT is listed)

Semantics of a Bayes Net

The structure of the BN means: every \(X_i\) is conditionally independent of all of its non-descendants given its parents:

\[
\Pr(X_i | S \cup \text{Par}(X_i)) = \Pr(X_i | \text{Par}(X_i))
\]

for any subset \(S \subseteq \text{NonDescendants}(X_i)\)
Semantics of Bayes Nets (2)

- If we ask for \( P(x_1, x_2, \ldots, x_n) \) we obtain
  - assuming an ordering consistent with network

- By the chain rule, we have:

\[
\begin{align*}
Pr(x_1, x_2, \ldots, x_n) &= Pr(x_n \mid x_{n-1}, \ldots, x_1) Pr(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots Pr(x_1) \\
&= Pr(x_n \mid \text{Par}(x_n)) Pr(x_{n-1} \mid \text{Par}(x_{n-2})) \ldots Pr(x_1)
\end{align*}
\]

- Thus, the joint is recoverable using the parameters (CPTs) specified in an arbitrary BN

Causal Intuitions

- The construction of a BN is simple
  - works with arbitrary orderings of variable set
  - but some orderings much better than others!
  - generally, if ordering/dependence structure reflects causal intuitions, a more natural, compact BN results

- In this BN, we’ve used the ordering Mal, Cold, Flu, Aches to build BN for distribution \( P \)
  - Variable can only have parents that come earlier in the ordering

Testing Independence

- Given BN, how do we determine if two variables \( X, Y \) are independent (given evidence \( E \))?
  - we use a (simple) graphical property

- **D-separation**: A set of variables \( E \) d-separates \( X \) and \( Y \) if it blocks every undirected path in the BN between \( X \) and \( Y \). (We’ll define blocks next.)

- \( X \) and \( Y \) are conditionally independent given evidence \( E \) if \( E \) d-separates \( X \) and \( Y \)
  - thus BN gives us an easy way to tell if two variables are independent \((\text{set } E = \emptyset)\) or conditionally independent

Constructing a Bayes Net

- Given any distribution over variables \( X_1, X_2, \ldots, X_n \) we can construct a Bayes net that faithfully represents that distribution.

Take any ordering of the variables (say, the order given), and go through the following procedure for \( X_n \) down to \( X_1 \).
Let \( \text{Par}(X_n) \) be any subset \( S \ subseteq \{X_1, \ldots, X_n\} \) such that \( X_n \) is independent of \( \{X_1, \ldots, X_{n-1}\} \) given \( S \). Such a subset must exist (convince yourself). Then determine the parents of \( X_n \), the same way, finding a similar \( S \ subseteq \{X_1, \ldots, X_{n-1}\} \), and so on. In the end, a DAG is produced and the BN semantics must hold by construction.

Causal Intuitions

- Suppose we build the BN for distribution \( P \) using the opposite ordering
  - i.e., we use ordering Aches, Cold, Flu, Malaria
  - resulting network is more complicated!

- Mal depends on Aches; but it also depends on Cold, Flu
  - given Aches

- Flu depends on Aches; but also on Cold

- Cold depends on Aches

Blocking in D-Separation

- Let \( P \) be an undirected path from \( X \) to \( Y \) in a BN.
Let \( E \) be an evidence set. We say \( E \) blocks path \( P \) iff there is some node \( Z \) on the path such that:

  - **Case 1**: one arc on \( P \) goes into \( Z \) and one goes out of \( Z \), and \( Z \in E \); or

  - **Case 2**: both arcs on \( P \) leave \( Z \), and \( Z \notin E \); or

  - **Case 3**: both arcs on \( P \) enter \( Z \) and neither \( Z \), nor any of its descendents, are in \( E \).
**D-Separation: Intuitions**

- Subway and Therm are dependent; but are independent given Flu (since Flu blocks the only path).
- Aches and Fever are dependent; but are independent given Flu (since Flu blocks the only path). Similarly for Aches and Therm (dependent, but indep. given Flu).
- Flu and Mal are indep. (given no evidence): Fever blocks the path, since it is not in evidence, nor is its descendant Therm. Flu, Mal are dependent given Fever (or given Therm): nothing blocks path now.
- Subway, ExoticTrip are indep.; they are dependent given Therm; they are indep. given Therm and Malaria. This for exactly the same reasons for Flu/Mal above.

**Inference in Bayes Nets**

- The independence sanctioned by D-separation allows us to compute prior and posterior probabilities quite effectively.
- We'll look at a couple simple examples to illustrate. We'll focus on networks without loops. (A loop is a cycle in the underlying undirected graph. Recall the directed graph has no cycles.)

**Simple Forward Inference (Chain)**

- Computing prior require simple forward "propagation" of probabilities (using Subway net)
  \[
  P(J) = \sum_{M,ET} P(J|M,ET) P(M,ET) \\
  = \sum_{M,ET} P(J|M) P(M|ET) P(ET) \\
  = \sum_{M} P(J|M) \sum_{ET} P(M|ET) P(ET) 
  \]
- (1) follows by summing out rule; (2) by chain rule and independence; (3) by distribution of sum
  - Note: all (final) terms are CPTs in the BN
  - Note: only ancestors of J considered
Simple Forward Inference (Pooling)

- Same idea applies with multiple parents
  \[ P(\text{Fev}) = \Sigma_{\text{Flu}, \text{M}} P(\text{Fev} | \text{Flu}, \text{M}) P(\text{Flu}, \text{M}) \]
  \[ = \Sigma_{\text{Flu}, \text{M}} P(\text{Fev} | \text{Flu}, \text{M}) P(\text{Flu}) P(\text{M}) \]
  \[ = \Sigma_{\text{Flu}, \text{M}} P(\text{Fev} | \text{Flu}, \text{M}) \Sigma_{\text{TS}} P(\text{Flu} | \text{TS}) P(\text{TS}) \]
  \[ \Sigma_{\text{ET}} P(\text{M} | \text{ET}) P(\text{ET}) \]

- (1) follows by summing out rule; (2) by independence of Flu, M; (3) by summing out
  * note: all terms are CPTs in the Bayes net

Simple Backward Inference

- When evidence is downstream of query variable, we must reason “backwards.” This requires the use of Bayes rule:
  \[ P(\text{ET} | j) = \alpha P(j | \text{ET}) P(\text{ET}) \]
  \[ = \alpha \Sigma_{\text{M}} P(j | \text{M}, \text{ET}) P(\text{M} | \text{ET}) P(\text{ET}) \]
  \[ = \alpha \Sigma_{\text{M}} P(j | \text{M}) P(\text{ET}) \]

- First step is just Bayes rule
  * normalizing constant \( \alpha \) is \( 1/P(j) \); but we needn’t compute it explicitly if we compute \( P(\text{ET} | j) \) for each value of ET; we just add up terms \( P(j | \text{ET}) P(\text{ET}) \) for all values of ET (they sum to \( P(j) \))

Backward Inference (Pooling)

- Same ideas when several pieces of evidence lie “downstream”
  \[ P(\text{ET} | j, \text{Fev}) = \alpha P(j, \text{Fev} | \text{ET}) P(\text{ET}) \]
  \[ = \alpha \Sigma_{\text{M}} P(j, \text{Fev} | \text{M}, \text{ET}) P(\text{M} | \text{ET}) P(\text{ET}) \]
  \[ = \alpha \Sigma_{\text{M}} P(j | \text{M}) P(\text{Fev} | \text{M}) P(\text{ET}) \]

- Same steps as before; but now we compute prob of both pieces of evidence given hypothesis ET and combine them. Note: they are independent given M; but not given ET.
  * Still must simplify \( P(\text{Fev}|\text{M}) \) down to CPTs (as usual)

Variable Elimination

- The intuitions in the above examples give us a simple inference algorithm for networks without loops: the polytree algorithm. We won’t discuss it further. But be comfortable with the intuitions.
- Instead we’ll look at a more general algorithm that works for general BNs; but the propagation algorithm will more or less be a special case.
- The algorithm, variable elimination, simply applies the summing out rule repeatedly. But to keep computation simple, it exploits the independence in the network and the ability to distribute sums inward.