Bayes Nets II: Independence Day

Hal Daumé III

Computer Science
University of Maryland

me@hal3.name

CS 421: Introduction to Artificial Intelligence

5 Apr 2012

Many slides courtesy of Dan Klein, Stuart Russell, or Andrew Moore
Announcements

- HW07 bug
  - If you did the full thing, great
  - If not, that's fine too
- P2 results have been sent, please complain now :)
- Midterms will be returned on Tuesday
  - Solution is posted online
  - Let me know if you find bugs :)
Bayes’ Nets

➢ So far:
  ➢ What is a Bayes’ net?
  ➢ What joint distribution does it encode?

➢ Next: how to answer queries about that distribution
  ➢ Key idea: conditional independence
  ➢ Last class: assembled BNs using an intuitive notion of conditional independence as causality
  ➢ Today: formalize these ideas
  ➢ Main goal: answer queries about conditional independence and influence

➢ After that: how to answer numerical queries (inference)
Conditional Independence

➢ Reminder: independence
➢ X and Y are independent if

\[ \forall x, y \quad P(x, y) = P(x)P(y) \quad \iff \quad X \perp Y \]

➢ X and Y are conditionally independent given Z

\[ \forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \quad \iff \quad X \perp Y|Z \]

➢ (Conditional) independence is a property of a distribution
Example: Independence

- For this graph, you can fiddle with $\theta$ (the CPTs) all you want, but you won’t be able to represent any distribution in which the flips are dependent!

\[
\begin{array}{c|c|c}
X_1 & h & 0.5 \\
 & t & 0.5 \\
\hline
X_2 & h & 0.5 \\
 & t & 0.5 \\
\end{array}
\]
Given some graph topology $G$, only certain joint distributions can be encoded.

The graph structure guarantees certain (conditional) independences.

(There might be more independence)

Adding arcs increases the set of distributions, but has several costs.
Important question about a BN:
- Are two nodes independent given certain evidence?
- If yes, can calculate using algebra (really tedious)
- If no, can prove with a counter example

Example:

Question: are X and Z independent?
- Answer: not necessarily, we’ve seen examples otherwise: low pressure causes rain which causes traffic.
- X can influence Z, Z can influence X (via Y)
- Addendum: they could be independent: how?
Causal Chains

➢ This configuration is a “causal chain”

\[ P(x, y, z) = P(x)P(y|x)P(z|y) \]

➢ Is X independent of Z given Y?

\[ P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y) \quad \text{Yes!} \]

➢ Evidence along the chain “blocks” the influence

X: Low pressure
Y: Rain
Z: Traffic
Common Cause

- Another basic configuration: two effects of the same cause
  - Are X and Z independent?
  - Are X and Z independent given Y?

\[
P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}
\]

\[
= P(z|y)
\]

- Observing the cause blocks influence between effects.

Y: Project due
X: Email busy
Z: Lab full
Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: remember the ballgame and the rain causing traffic, no correlation?
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - No: remember that seeing traffic put the rain and the ballgame in competition?
  - This is backwards from the other cases
    - Observing the effect enables influence between effects.

X: Raining
Z: Ballgame
Y: Traffic
The General Case

➢ Any complex example can be analyzed using these three canonical cases

➢ General question: in a given BN, are two variables independent (given evidence)?

➢ Solution: analyze the graph
Reachability

- Recipe: shade evidence nodes
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn’t count as a link in a path unless “active”
Reachability (the Bayes’ Ball)

- **Correct algorithm:**
  - Shade in evidence
  - Start at source node
  - Try to reach target by search
  - States: pair of (node X, previous state S)
- **Successor function:**
  - X unobserved:
    - To any child
    - To any parent if coming from a child
  - X observed:
    - From parent to parent
  - If you can’t reach a node, it’s conditionally independent of the start node given evidence
Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence variables \( \{Z\} \)?
- Look for “active paths” from X to Y
- No active paths = independence!

- A path is active if each triple is either a:
  - Causal chain \( A \rightarrow B \rightarrow C \) where B is unobserved (either direction)
  - Common cause \( A \leftarrow B \rightarrow C \) where B is unobserved
  - Common effect (aka v-structure) \( A \rightarrow B \leftarrow C \) where B or one of its descendents is observed
Example

\[ A \parallel W \]
\[ A \parallel W | R \]

Yes

Diagram:

- Aliens
- Watch
- Late
- Report

Diagram illustrates the relationship between the terms and their corresponding yes/no classification.
Example

\[ \text{Yes} \]

\[ \text{Yes} \]

\[ \text{Yes} \]

\[ \text{Yes} \]
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad

- Questions:

\[
T \perp D \\
T \perp D | R \\
T \perp D | R, S
\]

Yes
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain
- E.g. consider the variables *Traffic* and *Drips*
- End up with arrows that reflect correlation, not causation

What do the arrows really mean?
- Topology may happen to encode causal structure
- **Topology only guaranteed to encode conditional independence**
Example: Traffic

- Basic traffic net
- Let’s multiply out the joint

\[ P(R) \]

\[
\begin{array}{c|c}
    r & 1/4 \\
    \neg r & 3/4 \\
\end{array}
\]

\[ P(T|R) \]

\[
\begin{array}{c|c|c}
    & t & 3/4 \\
    r & 3/4 & 1/4 \\
    \neg t & 1/4 \\
\end{array}
\]

\[ P(T, R) \]

\[
\begin{array}{c|c|c}
    & r & t \\
    r & 3/16 & 1/16 \\
    \neg r & 6/16 & 6/16 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
    & r & \neg t \\
    \neg r & 1/16 & 6/16 \\
\end{array}
\]
Example: Reverse Traffic

➢ Reverse causality?

\[
\begin{array}{c|c|c}
T & P(T) & P(T, R) \\
\hline
\text{t} & 9/16 & r \text{ t} 3/16 \\
\text{\neg t} & 7/16 & r \text{ \neg t} 1/16 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c}
 & r & \text{t} & 1/3 & 3/16 \\
\hline
\text{t} & & & & \\
\text{\neg r} & & & 2/3 & \\
\hline
\text{\neg t} & r & & 1/7 & \\
\text{\neg r} & & & 6/7 & \\
\hline
\end{array}
\]

\[
\begin{array}{c|c|c|c|c}
 & r & \text{t} & 6/16 & \\
\hline
\text{r} & & & & \\
\text{\neg t} & & & 6/16 & \\
\hline
\end{array}
\]
Example: Coins

- Extra arcs don’t prevent representing independence, just allow non-independence

\[
\begin{array}{c|c}
X_1 & P(X_1) \\
\hline
h & 0.5 \\
t & 0.5 \\
\end{array}
\quad \begin{array}{c|c}
P(X_2) \\
\hline
h & 0.5 \\
t & 0.5 \\
\end{array}
\quad \begin{array}{c|c}
P(X_1) \\
\hline
h & 0.5 \\
t & 0.5 \\
\end{array}
\quad \begin{array}{c|c|c}
P(X_2|X_1) \\
\hline
h | h & 0.5 \\
t | h & 0.5 \\
h | t & 0.5 \\
t | t & 0.5 \\
\end{array}
\]
Alternate BNs

- MaryCalls
- JohnCalls
- Alarm
- Burglary
- Earthquake

Diagram:

- B
- E
- J
- M
- A

Graphical representation of the relationships between events.
Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- The Bayes’ ball algorithm (aka d-separation)
- A Bayes’ net may have other independencies that are not detectable until you inspect its specific distribution