CS 188: Artificial Intelligence

Bayes’ Nets: Independence

Fall 2022

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]
Review: Bayes’ Net Semantics
Probability Recap

- **Conditional probability**
  \[ P(x|y) = \frac{P(x, y)}{P(y)} \]

- **Product rule**
  \[ P(x, y) = P(x|y)P(y) \]

- **Chain rule**
  \[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \]

- **X, Y independent if and only if:**
  \[ \forall x, y : P(x, y) = P(x)P(y) \]

- **X and Y are conditionally independent given Z if and only if:**
  \[ \forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \quad X \perp Y | Z \]
A Bayes’ net is an efficient encoding of a probabilistic model of a domain.

Questions we can ask:

- Inference: given a fixed BN, what is P(X | e)?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
Bayes’ Net Semantics

- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  - $P(X|a_1 \ldots a_n)$
  - CPT: conditional probability table
  - Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Probabilities in BNs

- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

\[
P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))
\]

- Example:

\[
P(+\text{cavity}, +\text{catch}, -\text{toothache})
\]
Probabilities in BNs

- Why are we guaranteed that setting
  \[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]
  results in a proper joint distribution?

- Chain rule (valid for all distributions):
  \[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|x_1 \ldots x_{i-1}) \]

- Assume conditional independences:
  \[ P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]
  → Consequence:
  \[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]

- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies
Example: Coin Flips

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.
Example: Traffic

$P(R)$

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>1/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>-r</td>
<td>3/4</td>
<td></td>
</tr>
</tbody>
</table>

$P(T|R)$

<table>
<thead>
<tr>
<th></th>
<th>+t</th>
<th>3/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>-t</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-t</td>
<td>1/2</td>
<td></td>
</tr>
</tbody>
</table>

$P(+r, -t) =$
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B  | E  | A  | P(A|B,E) |
|----|----|----|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |

- Burglary
- Earthquake
- Alarm
- John calls
- Mary calls

| A  | J  | P(J|A) |
|----|----|------|
| +a | +j | 0.9  |
| +a | -j | 0.1  |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

| A  | M  | P(M|A) |
|----|----|------|
| +a | +m | 0.7  |
| +a | -m | 0.3  |
| -a | +m | 0.01 |
| -a | -m | 0.99 |
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) =
\]
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b)P(-e)P(+a| +b, -e)P(-j| +a)P(+m| +a) = 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Example: Traffic

- Causal direction

\[ P(R) \]

<table>
<thead>
<tr>
<th>( +r )</th>
<th>1/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( -r )</td>
<td>3/4</td>
</tr>
</tbody>
</table>

\[ P(T | R) \]

<table>
<thead>
<tr>
<th>( +r )</th>
<th>+t</th>
<th>3/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>-t</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>( -r )</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-t</td>
<td>1/2</td>
<td></td>
</tr>
</tbody>
</table>

\[ P(T, R) \]

<table>
<thead>
<tr>
<th>( +r )</th>
<th>+t</th>
<th>3/16</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>1/16</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>6/16</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>6/16</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

- Reverse causality?

\[
P(T) = \begin{array}{cc}
+\text{t} & 9/16 \\
-\text{t} & 7/16 \\
\end{array}
\]

\[
P(R|T) = \begin{array}{ccc}
+\text{t} & +\text{r} & 1/3 \\
-\text{r} & 2/3 \\
+\text{r} & 1/7 \\
-\text{r} & 6/7 \\
\end{array}
\]

\[
P(T,R) = \begin{array}{ccc}
+\text{r} & +\text{t} & 3/16 \\
+\text{r} & -\text{t} & 1/16 \\
-\text{r} & +\text{t} & 6/16 \\
-\text{r} & -\text{t} & 6/16 \\
\end{array}
\]
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 (especially if variables are missing)
- 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 really encodes conditional independence

\[
P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\text{parents}(X_i))
\]
Size of a Bayes’ Net

- How big is a joint distribution over $N$ Boolean variables?
  $2^N$

- How big is an $N$-node net if nodes have up to $k$ parents?
  $O(N \times 2^{k+1})$

- Both give you the power to calculate
  $P(X_1, X_2, \ldots X_n)$

- BNs: Huge space savings!

- Also easier to elicit local CPTs

- Also faster to answer queries (coming)
Bayes’ Nets

- So far: how a Bayes’ net encodes a joint distribution

- Next: how to answer queries about that distribution
  - Last Time:
    - First assembled BNs using an intuitive notion of conditional independence as causality
    - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence

- Today: how to answer numerical queries (inference)
Bayes’ Nets

- Representation
  - Conditional Independences
  - Probabilistic Inference
  - Learning Bayes’ Nets from Data
Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:

\[ P(x_i|x_1 \cdots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

- Beyond above “chain rule \(\rightarrow\) Bayes net” conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph

- Important for modeling: understand assumptions made when choosing a Bayes net graph
Example

- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?
Important question about a BN:
- Are two nodes independent given certain evidence?
- If yes, can prove using algebra (tedious in general)
- If no, can prove with a counter example
- Example:

  \[
  \begin{aligned}
  X & \rightarrow Y \\
  Y & \rightarrow Z \\
  \end{aligned}
  \]

Question: are X and Z necessarily independent?
- Answer: no. Example: low pressure causes rain, which causes traffic.
- X can influence Z, Z can influence X (via Y)
- Addendum: they could be independent: how?
D-separation: Outline
D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries
This configuration is a “causal chain”

- Low pressure causes rain causes traffic,
  high pressure causes no rain causes no traffic

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

Guaranteed X independent of Z? No!

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

Example:

- Low pressure causes rain causes traffic,
  high pressure causes no rain causes no traffic

In numbers:

\[ P(+y | +x ) = 1, P(-y | -x ) = 1, \]
\[ P(+z | +y ) = 1, P(-z | -y ) = 1 \]
Causal Chains

- This configuration is a “causal chain”

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

- Guaranteed X independent of Z given Y?

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

Yes!

- Evidence along the chain “blocks” the influence
This configuration is a “common cause”

Guaranteed X independent of Z? No!

One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

Example:

- Project due causes both forums busy and lab full

In numbers:

\[ P( +x \mid +y ) = 1, P( -x \mid -y ) = 1, \]
\[ P( +z \mid +y ) = 1, P( -z \mid -y ) = 1 \]

\[ P(x, y, z) = P(y)P(x \mid y)P(z \mid y) \]
Common Cause

- This configuration is a “common cause”

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

- Guaranteed X and Z independent given Y?

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

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

\[ = P(z|y) \]

Yes!

- Observing the cause blocks influence between effects.
Common Effect

- Last configuration: two causes of one effect (v-structures)

- Are X and Y independent?
  - Yes: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)

- Are X and Y independent given Z?
  - No: seeing traffic puts the rain and the ballgame in competition as explanation.

- This is backwards from the other cases
  - Observing an effect activates influence between possible causes.
The General Case

CONDITIONAL INDEPENDENCE

IN 3 EASY STEPS!
The General Case

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

- Solution: analyze the graph

- Any complex example can be broken into repetitions of the three canonical cases
Recipe: shade evidence nodes, look for paths in the resulting graph

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”
Active / Inactive Paths

- **Question:** Are X and Y conditionally independent given evidence variables \( \{Z\} \)?
  - Yes, if X and Y “\( d \)-separated” by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!

- **A path is active if each triple is active:**
  - 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

- All it takes to block a path is a single inactive segment
D-Separation

- Query: \( X_i \perp\!
\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \) ?

- Check all (undirected!) paths between \( X_i \) and \( X_j \)
  - If one or more active, then independence not guaranteed
    \( X_i \not\perp\!
\not\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \)
  - Otherwise (i.e. if all paths are inactive),
    then independence is guaranteed
    \( X_i \perp\!
\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \)
Example

\[
\begin{align*}
R \perp B & \\
R \perp B | T & \\
R \perp B | T' & \\
\end{align*}
\]

Yes

![Diagram](image.png)
Example

$L \perp T' | T$  Yes
$L \perp B$  Yes
$L \perp B | T$  
$L \perp B | T'$  
$L \perp B | T, R$  Yes

Diagram:

- $L$ points to $R$
- $R$ points to $T$
- $B$ points to $T'$
- $D$ points to $T$
- $T'$
Example

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

- **Questions:**

  \[
  T \perp D \\
  T \perp D|R \quad \text{Yes} \\
  T \perp D|R, S
  \]
Given a Bayes net structure, can run $d$-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

$$X_i \perp \!\!\!\! \perp X_j | \{X_{k_1}, \ldots, X_{k_n}\}$$

This list determines the set of probability distributions that can be represented.
Computing All Independences

Compute ALL THE INDEPENDENCES!
Topology Limits Distributions

- 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.
- Full conditioning can encode any distribution.
Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution
Bayes’ Nets

- Representation
- Conditional Independences
  - Probabilistic Inference
    - Enumeration (exact, exponential complexity)
    - Variable elimination (exact, worst-case exponential complexity, often better)
    - Probabilistic inference is NP-complete
    - Sampling (approximate)
- Learning Bayes’ Nets from Data