AI: Representation and Problem Solving

Bayes Nets I

Instructors: Nihar Shah and Tuomas Sandholm

Slide credits: CMU AI and ai.berkeley.edu
Example: COVID modeling

What is \( P(\text{URT epithelial infection} = \text{yes} \mid \text{dry cough} = \text{yes}, \text{productive cough} = \text{no}, \text{anosmia} = \text{yes})? \)
Simpler example: Grade prediction

What is $P(\text{grade} = A \mid SAT = 1550)$ ?

[Diagram of a Bayesian network with nodes Difficulty, Intelligence, Grade, SAT, Letter, and directed edges connecting them.]

Bayes nets

- A Bayes net (Bayesian network) is a way to model relationships between various variables

- Goal is to obtain some marginal or conditional distributions
  - e.g., P(infected) or P(infected | cough)
Probability overview...
281 Pizzeria!

You pick one slice uniformly at random. What is the probability of getting a slice with:

1) No mushrooms
2) Spinach and no mushrooms
3) Spinach, when asking for slice with no mushrooms
281 Pizzeria!

You pick one slice uniformly at random. What is the probability of getting a slice with:

- Mushrooms
- Spinach
- No spinach
- No spinach and mushrooms
- No spinach when asking for no mushrooms
- No spinach when asking for mushrooms
- Spinach when asking for mushrooms
- No mushrooms and no spinach

Probability notation

- Upper case letters (e.g., $A$) to denote random variables

- For a random variable $A$ taking values $\{a_1, a_2, a_3\}$

\[
p(A) = \begin{pmatrix} 0.1 \\ 0.5 \\ 0.4 \end{pmatrix}
\]

represents the set of probabilities for each value that $A$ can take on (this is a function mapping values of $A$ to numbers that sum to one)

- We use lower case letters like $a$ to denote a specific value of $A$ (i.e., for above example $\{a_1, a_2, a_3\}$), and $p(A = a)$ or just $p(a)$ refers to a number (the corresponding entry of $p(A)$)
Discrete Probability Distributions

For each random variable
- Discrete outcomes
- Disjoint outcomes
- Accounts for entire event space
- Not always binary

Discrete Random Variables (and their domains)

\[ A \in \{a_1, a_2, a_3\} \]
\[ B \in \{+b, -b\} \]
\[ C \in \{+c, -c\} \]
Probability notation

Given two random variables: $B$ with values in $\{+b, -b\}$ and $C$ with values in $\{+c, -c\}$:

- $p(B, C)$ refers to the joint distribution, i.e., a set of 6 possible values for each setting of variables, i.e., a function mapping $(+b, +c), (+b, -c), (-b, +c), ...$ to corresponding probabilities.

- $p(+b, -c)$ is a number: probability that $B = +b$ and $C = -c$.

- $p(B, c)$ is a set of 2 values, the probabilities for all values of $B$ for the given value $C = c$, i.e., it is a function mapping $+b, -b$ to numbers (note: not probability distribution, it will not sum to one) Why?
Three random variables: $A \in \{a_1, a_2, a_3\}, B \in \{+b, -b\}, C \in \{+c, -c\}$

$P(B = +b, C) = \sum_{a \in \{a_1, a_2, a_3\}} P(A = a, B = +b, C)$

Also written as $P(+b, C) = \sum_{a \in \{a_1, a_2, a_3\}} P(a, +b, C)$
Joint probability distribution

Table representing all values

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>( P(A=a, B=b, C=c) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
</tbody>
</table>
Discrete probability distributions

- **Joint distribution** $P(M, S, R)$

Discrete Random Variables (and their domains)

$M \in \{+m, -m\}$

$S \in \{+s, -s\}$

$R \in \{+r, -r\}$

Marginalization

For random variables \( B, C \) if you have joint distribution \( p(B, C) \), how do you get the marginal probabilities \( p(B), p(C) \)?

- \( p(B) = \sum_{c \in \{+c, -c\}} p(B, C = c) \)
- \( p(C) = \sum_{b \in \{+b, -b\}} p(B = b, C) \)

- \( p(+b) = \sum_{c \in \{+c, -c\}} p(+b, c) = p(+b, +c) + p(+b, -c) \)

Marginalization is summing out a subset of random variables from a joint distribution to obtain a distribution of the remaining subset.
Discrete probability distributions

Marginal distribution

E.g., what is the probability that chosen slice has no spinach?

\[ P(-s) \]
# Marginalization from table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>P(A=a, B=b, C=c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
</tbody>
</table>

What is $p(B)$?

Sum rows that share the same value of $B$
Conditional probability

Given that the chosen slice has roasted onion, what is the probability it has mushrooms and no spinach?

\[ P(+m, -s | +r) \]

We restrict our attention to satisfying the “given” condition, and then normalize the values so that they sum to 1 (form a distribution)

\[
P(+m, -s | +r) = \frac{P(+m, -s, +r)}{P(+m, +s, +r) + P(+m, -s, +r) + P(-m, +s, +r) + P(-m, -s, +r)}\]

\[ P(+r) \]
Conditional probability

The conditional probability $p(B \mid C = +c)$ ("B given $C = +c"$) is defined as

$$p(B \mid C = +c) = \frac{p(B, C = +c)}{p(C = +c)}$$

- $p(+b \mid C = +c) = \frac{p(+b, +c)}{p(+c)}$
- $p(-b \mid C = +c) = \frac{p(-b, +c)}{p(+c)}$
Conditional probability from table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>P(A=a, B=b, C=c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>+a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>+b</td>
<td>−c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>+c</td>
<td></td>
</tr>
<tr>
<td>−a</td>
<td>−b</td>
<td>−c</td>
<td></td>
</tr>
</tbody>
</table>

What is \( p(A = +a, B = +b | C = +c) \)?

We restrict our attention to rows that satisfy the “given” condition, and then normalize the values so that they sum to 1 (form a distribution)
Which of the following probability tables sum to one?

i. $P(A \mid b)$

ii. $P(A, b, C)$

iii. $P(A, C \mid b)$

iv. $P(a, c \mid b)$
More practice

What is the probability of getting a slice with....
Answer queries from joint distribution

You can answer all of these questions:

\[
P(M| + s) \quad P(M| - s)
\]

|   | \(P(M| + s)\) | \(P(M| - s)\) |
|---|--------------|--------------|
| +m | +m           | +m           |
| -m | -m           | -m           |

\[
P(M, S)
\]

| \(P(M, S)\) | \(P(S| + m)\) | \(P(S| - m)\) |
|-------------|--------------|--------------|
| +m + s      | +s           | +s           |
| +m - s      | -s           | -s           |
| -m + s      | -s           | -s           |
| -m - s      | 1/2          | -s           |

\[
P(S| + m) \quad P(S| - m)
\]

|   | \(P(S| + m)\) | \(P(S| - m)\) |
|---|--------------|--------------|
| +s | +s           | +s           |
| -s | 1/2          | -s           |

\[
P(M, S)
\]

| \(P(M, S)\) | \(P(M| + s)\) | \(P(M| - s)\) |
|-------------|--------------|--------------|
| +m + s      | +m           | +m           |
| +m - s      | -m           | -m           |
| -m + s      | -m           | -m           |
| -m - s      | 1/2          | -s           |
Bayes rule

Suppose you are given $p(C|B)$ along with $p(B)$ and $p(C)$. How do you get $p(B|C)$?

*Hint: Think about the equations of conditional probability*

\[
p(B|C) = \frac{p(B, C)}{p(C)} \quad \quad \quad p(C|B) = \frac{p(B, C)}{p(B)}
\]

\[
p(B|C) = \frac{p(C|B)p(B)}{p(C)}
\]
How to answer queries?

- **Joint distributions are the best!**
  - Allow us to answer all marginal or conditional queries
  - However…

- Often we don’t have the joint table. Only know some set of conditional probability tables (CPTs)
E.g., know $P(\text{dry cough} | \text{URT epithelial infection, Pul. capillary leakage})$, $P(\text{anosmia and/or ageusia} | \text{Infection of olfactory epithelium})$, etc.

*Want to answer questions like: What is $P(\text{epithelial infection} = \text{yes} | \text{dry cough=}\text{yes}, \text{productive cough=}\text{no}, \text{anosmia=}\text{yes})$?*
Answering queries from CPTs

Conditional Probability Tables

\[ P(A) \ P(B|A) \ P(C|A,B) \ P(D|A,B,C) \ P(E|A,B,C,D) \]

Joint

Chain Rule

Query

\[ P(a \mid e) \]
Joint distribution from conditionals

1. Product rule
   - Can you write $P(A, B)$ in terms of $P(A \mid B)$ and $P(B)$?
   - $P(A, B) = P(A \mid B)P(B)$
Joint distribution from conditionals

1. Product rule
   - $P(A, B) = P(A \mid B)P(B)$
   - $P(A, B) = P(B \mid A)P(A)$
Joint distribution from conditionals

1. Product rule
   - $P(A, B) = P(A \mid B)P(B)$
   - $P(A, B) = P(B \mid A)P(A)$

2. Three random variables
   - We know $P(A, B) = P(A)P(B \mid A)$
   - What about writing $P(A, B, C)$ in terms of $P(A, B)$ and $P(C \mid A, B)$
   - Hint: Think of $(A, B)$ as if it was a single variable, and $C$ as a second variable
   - Product rule: $P((A, B), C) = P((A, B))P(C \mid (A, B))$
Joint distribution from conditionals

1. Product rule
   - \( P(A, B) = P(A \mid B)P(B) \)
   - \( P(A, B) = P(B \mid A)P(A) \)

2. Three random variables
   - Product rule: \( P(A, B, C) = P(A, B)P(C \mid A, B) \)

3. More generally, **Chain rule**
   - \( P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \)
Answering queries from CPTs: Example

- Major $\in \{\text{CS, not CS}\}$, Year $\in \{\text{sophomore, not sophomore}\}$
- If you pick a random student in the class, probability that they are a sophomore is 0.8
- If you pick a random sophomore in the class, probability that they are a CS major is 0.5
- If you pick a random non-sophomore in the class, probability that they are a CS major is 0.6
- What is the probability that if I pick a random CS major, they are a sophomore?
- What is the probability that if I pick a random student, they are a CS major?

- We are given $P(\text{Year})$, $P(\text{Major} | \text{Year} = \text{sophomore})$, $P(\text{Major} | \text{Year} = \text{not sophomore})$
- Construct $P(\text{Major, Year})$
  - $P(\text{Major}=\text{CS}) = P(\text{Major}=\text{CS}, \text{Year}=\text{sophomore}) + P(\text{Major}=\text{CS}, \text{Year}=\text{not sophomore})$
  - $P(\text{Year}=\text{sophomore} | \text{Major}=\text{CS})=P(\text{Year}=\text{sophomore}, \text{Major}=\text{CS})/P(\text{Major}=\text{CS})$
Answering queries from CPTs: Problem

Conditional Probability Tables and Chain Rule

- If there are $n$ variables taking $d$ values each
- $d^n$ entries!!

$P(A) \quad P(B|A) \quad P(C|A, B) \quad P(D|A, B, C) \quad P(E|A, B, C, D)$
Sometimes, distributions have simpler structure

\[
P(A, B, C, D, E) = P(A) P(B|A) P(C|A, B) P(D|A, B, C) P(E|A, B, C, D)
\]

- Suppose \( P(E|A, B, C, D) = P(E|A, B) \) and \( P(D|A, B, C) = P(D|A, B) \)

- “Conditional independence” (more on this soon)

E.g., \( P(\text{dry cough}|\text{URT epithelial infection, Pul. capillary leakage, virus enters URT}) = P(\text{dry cough}|\text{URT epithelial infection, Pul. capillary leakage}) \)
Sometimes, distributions have simpler structure

\[ P(A, B, C, D, E) = P(A) P(B|A) P(C|A, B) P(D|A, B, C) P(E|A, B, C, D) \]

• Suppose \( P(E|A, B, C, D) = P(E|A, B) \) and \( P(D|A, B, C) = P(D|A, B) \)

• “Conditional independence” (more on this soon)

• Then

\[
\begin{align*}
P(A) P(B|A) P(C|A, B) P(D|A, B, C) P(E|A, B, C, D) \\
= & \quad P(A) P(B|A) P(C|A, B) P(D|A, B) P(E|A, B)
\end{align*}
\]

• Needs less data to estimate conditionals (e.g., \( P(E|A, B) \) is easier to estimate than \( P(E|A, B, C, D) \))

• Needs less computation and storage to answer other queries
But what is this “Independence”? 
I roll two fair dice...

- What is the probability that the first roll is 5?
- What is the probability that the second roll is 5?
- What is the probability that both rolls are 5?
- If the first roll is 5, what is the probability that the second roll is 5?

\[ P(\text{Roll}_1=5, \text{Roll}_2=5) = P(\text{Roll}_1=5) \times P(\text{Roll}_2=5) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \]

\[ P(\text{Roll}_2=5 \mid \text{Roll}_1=5) = P(\text{Roll}_2=5) = \frac{1}{6} \]

- Independence and conditional independence!
Two random variables $X$ and $Y$ are *independent* if

$$\forall x, y \quad P(x, y) = P(x) \cdot P(y)$$

- This says that their joint distribution *factors* into a product of two simpler distributions.

- Notation: $X \independent Y$

- Combine with product rule $P(x,y) = P(x | y)P(y)$ we obtain another form:

$$\forall x, y \quad P(x | y) = P(x) \quad \text{or} \quad \forall x, y \quad P(y | x) = P(y)$$
Example: Independence

- On fair, independent coin flips:

\[
P(X_1) = \begin{pmatrix} H & 0.5 \\ T & 0.5 \end{pmatrix}, \quad P(X_2) = \begin{pmatrix} H & 0.5 \\ T & 0.5 \end{pmatrix}, \ldots, \quad P(X_n) = \begin{pmatrix} H & 0.5 \\ T & 0.5 \end{pmatrix}
\]

\[
P(X_1, X_2, \ldots, X_n) = 2^n
\]

The joint distribution is simply the product.
Question

- Are $T$ and $W$ independent?

$$P(T)$$

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$$P(T, W)$$

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

$$P(W)$$

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Conditional independence

- X and Y are independent if $P(X \mid Y) = P(X)$

- X and Y are \textit{conditionally independent given} Z if
  - $P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)$
  - $P(X \mid Y, Z) = P(X \mid Z)$

- Notation: $X \indep Y \mid Z$
Conditional independence

- $P(\text{Toothache, Cavity, (p)Robe})$

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - $P(+r \mid +\text{toothache, +cavity}) = P(+r \mid +\text{cavity})$

- The same independence holds if I don’t have a cavity:
  - $P(+r \mid +\text{toothache, -cavity}) = P(+r \mid -\text{cavity})$

- Probe is conditionally independent of Toothache given Cavity:
  - $P(R \mid T, C) = P(R \mid C)$
Conditional independence

Equivalent statements:

- \( P(\text{Toothache} \mid \text{Probe}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \)

- \( P(\text{Toothache}, \text{Probe} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \cdot P(\text{Probe} \mid \text{Cavity}) \)
Have we seen conditional independence in previous lectures?

**MDPs**

“Markov” generally means that given the present state, the future and the past are independent.

For Markov decision processes, “Markov” means action outcomes depend only on the current state.

\[
P(S_{t+1} = s' | S_t = s_t, A_t = a_t, S_{t-1} = s_{t-1}, A_{t-1}, \ldots S_0 = s_0) = P(S_{t+1} = s' | S_t = s_t, A_t = a_t)
\]

*Andrey Markov (1856-1922)*
Probability Tools Summary

1. Definition of conditional probability
   \[ P(A|B) = \frac{P(A, B)}{P(B)} \]

2. Product Rule
   \[ P(A, B) = P(A|B)P(B) \]

3. Bayes’ theorem
   \[ P(B|A) = \frac{P(A|B)P(B)}{P(A)} \]

4. Chain Rule
   \[ P(X_1, \ldots, X_N) = \prod_{n=1}^{N} P(X_n \mid X_1, \ldots, X_{n-1}) \]
Summary of Independence Rules

- **Independence**
  If A and B are independent, then:
  \[ P(A, B) = P(A)P(B) \]
  \[ P(A \mid B) = P(A) \]
  \[ P(B \mid A) = P(B) \]

- **Conditional independence**
  If A and B are conditionally independent given C, then:
  \[ P(A, B \mid C) = P(A \mid C)P(B \mid C) \]
  \[ P(A \mid B, C) = P(A \mid C) \]
  \[ P(B \mid A, C) = P(B \mid C) \]
Poll

I want to know if I have come down with a rare strain of flu (occurring in only 1/10,000 people). There is an “accurate” test for the flu: if I have the flu, it will tell me I have 99% of the time, and if I do not have it, it will tell me I do not have it 99% of the time. I go to the doctor and test positive. What is the probability I have this flu?

(A) ≈ 99%
(B) ≈ 10%
(C) ≈ 1%
(D) ≈ 0.1%