CSE 573 PMP: Artificial Intelligence

Hanna Hajishirzi
Uncertainty and Bayes Nets

slides adapted from
Dan Klein, Pieter Abbeel ai.berkeley.edu
And Dan Weld, Luke Zettlemoyer
Our Status in CSE573

- We’re done with Search and planning
- We are done with learning to make decisions
- Probabilistic Reasoning and Machine Learning
  - Diagnosis
  - Speech recognition
  - Tracking objects
  - Robot mapping
  - Genetics
  - Error correcting codes
  - ... lots more!
Today

- Probability
- Bayes Nets

You’ll need all this stuff for the next few weeks, so make sure you go over it now!
Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?

- We denote random variables with capital letters

- Random variables have domains
  - R in \{true, false\} (often write as \{+r, -r\})
  - T in \{hot, cold\}
  - D in \[0, \infty\)
  - L in possible locations, maybe \{(0,0), (0,1), ...\}
Probability Distributions

- Associate a probability with each outcome

- Temperature:

  \[
  P(T) = \begin{array}{c|c}
  T & P \\
  \hline
  \text{hot} & 0.5 \\
  \text{cold} & 0.5
  \end{array}
  \]

- Weather:

  \[
  P(W) = \begin{array}{c|c}
  W & P \\
  \hline
  \text{sun} & 0.6 \\
  \text{rain} & 0.1 \\
  \text{fog} & 0.3 \\
  \text{meteor} & 0.0
  \end{array}
  \]
Probability Distributions

- Unobserved random variables have distributions

\[
P(T) \quad P(W)
\]

<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>

<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>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- A distribution is a TABLE of probabilities of values

- A probability (lower case value) is a single number

\[
P(W = \text{rain}) = 0.1
\]

- Must have: \( \forall x \ P(X = x) \geq 0 \) and \( \sum_x P(X = x) = 1 \)

Shorthand notation:

\[
P(\text{hot}) = P(T = \text{hot}),
P(\text{cold}) = P(T = \text{cold}),
P(\text{rain}) = P(W = \text{rain}),
\ldots
\]

OK if all domain entries are unique
Joint Distributions

- A joint distribution over a set of random variables: $X_1, X_2, \ldots X_n$ specifies a real number for each assignment (or outcome):

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$

$$P(x_1, x_2, \ldots x_n)$$

- Must obey:
  $$P(x_1, x_2, \ldots x_n) \geq 0$$

  $$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

- Size of distribution if $n$ variables with domain sizes $d$?
  - For all but the smallest distributions, impractical to write out!

$$P(T, W)$$

<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>
Events

- An event is a set $E$ of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it’s hot AND sunny?
  - Probability that it’s hot?
  - Probability that it’s hot OR sunny?

- Typically, the events we care about are partial assignments, like $P(T=\text{hot})$

<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>
Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables.
- Marginalization (summing out): Combine collapsed rows by adding.

\[
P(T, W)
\]

\[
\begin{array}{ccc}
T & W & P \\
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\end{array}
\]

\[
P(t) = \sum_s P(t, s)
\]

\[
P(s) = \sum_t P(t, s)
\]

\[
P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)
\]

\[
P(T)
\]

\[
\begin{array}{cc}
T & P \\
\text{hot} & 0.5 \\
\text{cold} & 0.5 \\
\end{array}
\]

\[
P(W)
\]

\[
\begin{array}{cc}
W & P \\
\text{sun} & 0.6 \\
\text{rain} & 0.4 \\
\end{array}
\]
Quiz: Marginal Distributions

\[
P(X, Y) = \begin{array}{ccc}
X & Y & P \\
+X & +Y & 0.2 \\
+X & -Y & 0.3 \\
-X & +Y & 0.4 \\
-X & -Y & 0.1 \\
\end{array}
\]

\[
P(x) = \sum_y P(x, y)
\]

\[
P(y) = \sum_x P(x, y)
\]
Quiz: Marginal Distributions

\[ P(X, Y) \]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[ P(x) = \sum_y P(x, y) \]

\[ P(y) = \sum_x P(x, y) \]

\[ P(X) \]

<table>
<thead>
<tr>
<th>X</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>.5</td>
</tr>
<tr>
<td>-x</td>
<td>.5</td>
</tr>
</tbody>
</table>

\[ P(Y) \]

<table>
<thead>
<tr>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+y</td>
<td>.6</td>
</tr>
<tr>
<td>-y</td>
<td>.4</td>
</tr>
</tbody>
</table>
A simple relation between joint and conditional probabilities

In fact, this is taken as the definition of a conditional probability

\[ P(a|b) = \frac{P(a, b)}{P(b)} \]

### Table: Conditional Probabilities

<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>

### Calculation

\[ P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4 \]

\[ = P(W = s, T = c) + P(W = r, T = c) = 0.2 + 0.3 = 0.5 \]
Quiz: Conditional Probabilities

\[ P(X, Y) \]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- \( P(+x \mid +y) ? \)
- \( P(-x \mid +y) ? \)
- \( P(-y \mid +x) ? \)
Quiz: Conditional Probabilities

\[ P(X, Y) \]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- \[ P(+x \mid +y) ? \]
  \[ \frac{0.2}{0.6} = \frac{1}{3} \]

- \[ P(-x \mid +y) ? \]
  \[ \frac{0.3}{0.6} = \frac{2}{3} \]

- \[ P(-y \mid +x) ? \]
  \[ \frac{0.4}{0.5} = 0.8 \]

- \[ P(-y \mid +x) ? \]
  \[ \frac{0.1}{0.5} = 0.2 \]
Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others.

### Conditional Distributions

- **$P(W|T = \text{hot})$**
  - **W** | **P**
    - sun
    - rain

- **$P(W|T = \text{cold})$**
  - **W** | **P**
    - sun
    - rain

### Joint Distribution

- **$P(T, W)$**
  - **T** | **W** | **P**
    - hot | sun | 0.4
    - hot | rain | 0.1
    - cold | sun | 0.2
    - cold | rain | 0.3
Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others.

### Conditional Distributions

- **$P(W|T = hot)$**
  
<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

- **$P(W|T = cold)$**
  
<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>rain</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### Joint Distribution

<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>
The Product Rule

- Sometimes have conditional distributions but want the joint

\[ P(y)P(x|y) = P(x, y) \quad \iff \quad P(x|y) = \frac{P(x, y)}{P(y)} \]
The Product Rule

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

- Example:

<table>
<thead>
<tr>
<th>( P(W) )</th>
<th>( D )</th>
<th>( W )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td></td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td></td>
<td></td>
<td>0.2</td>
</tr>
</tbody>
</table>

| \( P(D|W) \)         | \( D \) | \( W \) | \( P \) |
|---------------------|---------|---------|--------|
| wet                 | sun     | 0.1     |
| dry                 | sun     | 0.9     |
| wet                 | rain    | 0.7     |
| dry                 | rain    | 0.3     |

<table>
<thead>
<tr>
<th>( P(D, W) )</th>
<th>( D )</th>
<th>( W )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]

\[ P(x_1, x_2, \ldots x_n) = \prod_{i} P(x_i|x_1 \ldots x_{i-1}) \]
Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - “All models are wrong; but some are useful.”
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
Two variables are independent if:

\[ \forall x, y : P(x, y) = P(x)P(y) \]

- This says that their joint distribution factors into a product two simpler distributions
- Another form:
  \[ \forall x, y : P(x|y) = P(x) \]
  - We write: \( X \perp\!
\!\!\!\!\!\!\!\!\!\perp Y \)

Independence is a simplifying modeling assumption

- Empirical joint distributions: at best “close” to independent
- What could we assume for \{Weather, Traffic, Cavity, Toothache\}?
Example: Independence?

\[
P(T')
\]

\[
\begin{array}{cc}
T & P \\
\hline
\text{hot} & 0.5 \\
\text{cold} & 0.5 \\
\end{array}
\]

\[
P_1(T, W)
\]

\[
\begin{array}{ccc}
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\end{array}
\]

\[
P(W)
\]

\[
\begin{array}{cc}
W & P \\
\hline
\text{sun} & 0.6 \\
\text{rain} & 0.4 \\
\end{array}
\]

\[
P_2(T, W)
\]

\[
\begin{array}{ccc}
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.3 \\
\text{hot} & \text{rain} & 0.2 \\
\text{cold} & \text{sun} & 0.3 \\
\text{cold} & \text{rain} & 0.2 \\
\end{array}
\]
Example: Independence

- N fair, independent coin flips:

\[
P(X_1) \quad P(X_2) \quad \ldots \quad P(X_n)
\]

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

\[
P(X_1, X_2, \ldots X_n)
\]
Conditional Independence
P(Toothache, Cavity, Catch)

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

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

Catch is conditionally independent of Toothache given Cavity:
- $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$

Equivalent statements:
- $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
- $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$
- One can be derived from the other easily
Conditional Independence

- Unconditional (absolute) independence very rare (why?)

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

- $X$ is conditionally independent of $Y$ given $Z$ if and only if:

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

  or, equivalently, if and only if

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

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

- What about this domain:
  - Fire
  - Smoke
  - Alarm
Conditional Independence

- What about this domain:
  - Traffic
  - Umbrella
  - Raining
Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm
Conditional Independence and the Chain Rule

- Chain rule:
  \[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots \]

- Trivial decomposition:
  \[ P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic}) \]

- With assumption of conditional independence:
  \[ P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}) \]

- We can represent joint distributions by multiplying these simpler local distributions.
- Bayes’nets / graphical models help us express conditional independence assumptions.
Bayes’ Nets: Big Picture
Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

**Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)**

- More properly called **graphical models**
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For about 10 min, we’ll be vague about how these interactions are specified
Example Bayes’ Net: Insurance
Example Bayes’ Net: Car
Graphical Model Notation

- **Nodes:** variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- **Arcs:** interactions
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence (more later)

- For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Coin Flips

- N independent coin flips

- No interactions between variables: **absolute independence**
Example: Traffic

- **Variables:**
  - R: It rains
  - T: There is traffic

- **Model 1: independence**

- **Model 2: rain causes traffic**

- Why is an agent using model 2 better?
Example: Traffic II

- **Variables**
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity
Example: Alarm Network

- **Variables**
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Example: Alarm Network

- **Variables**
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayes’ Net Semantics
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

\[ P(X|A_1 \ldots A_n) \]

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:**

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

\[= P(-\text{toothache}|+\text{cavity})P(+\text{catch}|+\text{cavity})P(+\text{cavity})\]
Bayes’ Net Representation

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  
  $$P(X|a_1 \ldots a_n)$$

- 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:

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

$$P(X|A_1 \ldots A_n)$$
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
Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.

Example: Coin Flips

\[ P(h, h, t, h) = P(h)P(h)P(t)P(h) \]
Example: Traffic

\[ P(R) \]

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

\[ P(\text{+r, -t}) = P(\text{+r})P(\text{-t}|\text{+r}) = \frac{1}{4} \times \frac{1}{4} \]
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>

| 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 |

| 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   |

\[
P(M|A)P(J|A) / P(A|B,E)
\]
Example: Traffic

- Causal direction

\[
\begin{array}{c|c}
\text{r} & \text{P(R)} \\
\hline
+ \text{r} & 1/4 \\
- \text{r} & 3/4 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{r} & \text{t} & \text{P(T|R)} \\
\hline
+ \text{r} & + \text{t} & 3/4 \\
+ \text{r} & - \text{t} & 1/4 \\
- \text{r} & + \text{t} & 1/2 \\
- \text{r} & - \text{t} & 1/2 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{r} & \text{t} & \text{P(T, R)} \\
\hline
+ \text{r} & + \text{t} & 3/16 \\
+ \text{r} & - \text{t} & 1/16 \\
- \text{r} & + \text{t} & 6/16 \\
- \text{r} & - \text{t} & 6/16 \\
\end{array}
\]
Example: Reverse Traffic

- Reverse causality?

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

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

\[
P(T, R)
\begin{array}{c|c|c}
+ r & + t & 3/16 \\
+ r & - t & 1/16 \\
- r & + t & 6/16 \\
- r & - t & 6/16 \\
\end{array}
\]
Causality?

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

- Two ways to factor a joint distribution over two variables:

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

- Dividing, we get:

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

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Foundation of many systems we’ll see later (e.g. ASR, MT)

- In the running for most important AI equation!
Inference with Bayes’ Rule

- **Example:** Diagnostic probability from causal probability:

  \[
P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}
  \]

- **Example:**
  - M: meningitis, S: stiff neck

  \[
  P(+m) = 0.0001
  \]
  \[
  P(+s|+m) = 0.8
  \]
  \[
  P(+s|-m) = 0.01
  \]

  \[
P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}
  \]

  - **Note:** posterior probability of meningitis still very small
  - **Note:** you should still get stiff necks checked out! Why?
Quiz: Bayes’ Rule

- Given:

  \[ P(W) \]
  
  \[
  \begin{array}{c|c|c}
  & R & P \\
  \hline
  \text{sun} & 0.8 & \\
  \text{rain} & 0.2 & \\
  \end{array}
  \]

- \[ P(D|W) \]
  
  \[
  \begin{array}{c|c|c}
  D & W & P \\
  \hline
  \text{wet} & \text{sun} & 0.1 \\
  \text{dry} & \text{sun} & 0.9 \\
  \text{wet} & \text{rain} & 0.7 \\
  \text{dry} & \text{rain} & 0.3 \\
  \end{array}
  \]

- What is \( P(W \mid \text{dry}) \)?
Quiz: Bayes’ Rule

- Given:

\[ P(W) \]

\[
\begin{array}{c|c|c}
D & W & P \\
\hline
\text{wet} & \text{sun} & 0.1 \\
\text{dry} & \text{sun} & 0.9 \\
\text{wet} & \text{rain} & 0.7 \\
\text{dry} & \text{rain} & 0.3 \\
\end{array}
\]

- What is \( P(W \mid \text{dry}) \)?

\[
P(\text{sun} \mid \text{dry}) \sim P(\text{dry} \mid \text{sun})P(\text{sun}) = .9 \cdot .8 = .72
\]

\[
P(\text{rain} \mid \text{dry}) \sim P(\text{dry} \mid \text{rain})P(\text{rain}) = .3 \cdot .2 = .06
\]

\[
P(\text{sun} \mid \text{dry}) = \frac{12}{13}
\]

\[
P(\text{rain} \mid \text{dry}) = \frac{1}{13}
\]
Uncertainty Summary

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

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

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

- **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:**
  \[ X \perp Y | Z \]
  \[ \forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \]
Bayes’ Net Representation

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  
  \[ P(X|a_1 \ldots a_n) \]

- 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:

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