Announcements

- Upcoming
  - P3 Due 10/12
  - W2 Due 10/15
  - Midterm in evening of 10/22

- Review sessions:
  - Probability review: Friday 12-2pm in 306 Soda
  - Midterm review: on web page when confirmed
Today

- **Probability**
  - Random Variables
  - Joint and Marginal Distributions
  - Conditional Distribution
  - Product Rule, Chain Rule, Bayes’ Rule
  - Inference
  - Independence

- You’ll need all this stuff A LOT for the next few weeks, so make sure you go over it now!
Inference in Ghostbusters

- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green
- Sensors are noisy, but we know $P(\text{Color} | \text{Distance})$

| $P(\text{red} | 3)$ | $P(\text{orange} | 3)$ | $P(\text{yellow} | 3)$ | $P(\text{green} | 3)$ |
|-------------------|-------------------|-------------------|-------------------|
| 0.05              | 0.15              | 0.5               | 0.3               |

[Demo]
Uncertainty

- **General situation:**
  - **Evidence:** Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Hidden variables:** Agent needs to reason about other aspects (e.g., where an object is or what disease is present)
  - **Model:** Agent knows something about how the known variables relate to the unknown variables

- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge
Random Variables

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

- We denote random variables with capital letters

- Like variables in a CSP, random variables have domains
  - R in \{true, false\} (sometimes write as \{+r, \neg r\})
  - D in \[0, \infty\)
  - L in possible locations, maybe \{(0,0), (0,1), \ldots\}
Probability Distributions

- Unobserved random variables have distributions

<table>
<thead>
<tr>
<th>$P(T)$</th>
<th>$P(W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number

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

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

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

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

- Size of distribution if $n$ variables with domain sizes $d$?

- Must obey:

$$P(x_1, x_2, \ldots x_n) \geq 0$$

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

- For all but the smallest distributions, impractical to write out

<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>
Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables

- Probabilistic models:
  - (Random) variables with domains
    Assignments are called outcomes
  - Joint distributions: say whether assignments (outcomes) are likely
  - Normalized: sum to 1.0
  - Ideally: only certain variables directly interact

- Constraint satisfaction probs:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

<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>

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>T</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>F</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>F</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>T</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

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

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

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

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

- 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>
<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>

\[
P(W = r|T = c) = ???
\]
Conditional Distributions

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

### Conditional Distributions

\[
P(W|T = \text{hot})
\]

<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>

\[
P(W|T = \text{cold})
\]

<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

\[
P(T, W)
\]

<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>
Normalization Trick

- A trick to get a whole conditional distribution at once:
  - Select the joint probabilities matching the evidence
  - Normalize the selection (make it sum to one)

\[ P(T, W) \]

<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>

\[ P(T, r) \]

<table>
<thead>
<tr>
<th>T</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(T|r) \]

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.25</td>
</tr>
<tr>
<td>cold</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Why does this work? Sum of selection is \( P(\text{evidence})! \) (\( P(r) \), here)

\[
P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}
\]
Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
  - These represent the agent’s beliefs given the evidence

- Probabilities change with new evidence:
  - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
  - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
  - Observing new evidence causes beliefs to be updated
Inference by Enumeration

- $P(\text{sun})$?
- $P(\text{sun} \mid \text{winter})$?
- $P(\text{sun} \mid \text{winter}, \text{warm})$?

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Inference by Enumeration

- **General case:**
  - Evidence variables: \( E_1 \ldots E_k = e_1 \ldots e_k \)
  - Query* variable: \( Q \)
  - Hidden variables: \( H_1 \ldots H_r \)
  
  \[ X_1, X_2, \ldots X_n \]

- **We want:** \( P(Q|e_1 \ldots e_k) \)

- **First, select the entries consistent with the evidence**
- **Second, sum out H to get joint of Query and evidence:**

\[
P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(Q, h_1 \ldots h_r, e_1 \ldots e_k)
\]

- **Finally, normalize the remaining entries to conditionalize**

- **Obvious problems:**
  - Worst-case time complexity \( O(d^n) \)
  - Space complexity \( O(d^n) \) to store the joint distribution

* Works fine with multiple query variables, too
The Product Rule

- Sometimes have conditional distributions but want the joint

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

- Example:

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

\[
\begin{array}{c|c|c|c}
D & W & P \\
\hline
\text{wet} & \text{sun} & 0.08 \\
\text{dry} & \text{sun} & 0.72 \\
\text{wet} & \text{rain} & 0.14 \\
\text{dry} & \text{rain} & 0.06 \\
\end{array}
\]
The Chain Rule

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

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

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

Why is this always true?
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 is meningitis, s is stiff neck
  - \( P(s|m) = 0.8 \)
  - \( P(m) = 0.0001 \)
  - \( P(s) = 0.1 \)

\[ P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008 \]

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?
Let’s say we have two distributions:

- **Prior distribution** over ghost location: $P(G)$
  - Let’s say this is uniform
- **Sensor reading model**: $P(R | G)$
  - Given: we know what our sensors do
  - $R =$ reading color measured at $(1,1)$
  - E.g. $P(R = \text{yellow} | G=(1,1)) = 0.1$

We can calculate the **posterior distribution** $P(G|r)$ over ghost locations given a reading using Bayes’ rule:

$$P(g|r) \propto P(r|g)P(g)$$
Independence

- Two variables are *independent* in a joint distribution if:
  \[ P(X, Y) = P(X)P(Y) \]

- Says the joint distribution *factors* into a product of two simple ones
- Usually variables aren’t independent!

- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best “close” to independent
  - What could we assume for \{Weather, Traffic, Cavity\}?

- Independence is like something from CSPs: what?
Example: Independence?

$P(T)$

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

$P_1(T, W)$

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>warm</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>

$P_2(T, W)$

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>sun</td>
<td>0.3</td>
</tr>
<tr>
<td>warm</td>
<td>rain</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.3</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Example: Independence

- $N$ fair, independent coin flips:

$$P(X_1)$$

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

$$P(X_2)$$

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

... $$P(X_n)$$

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

$$P(X_1, X_2, \ldots X_n)$$

$2^n$
Conditional Independence

- \( P(\text{Toothache}, \text{Cavity}, \text{Catch}) \)

- 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:
  - \( P(\text{catch} \mid \text{toothache}, \neg \text{cavity}) = P(\text{catch} \mid \neg \text{cavity}) \)

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

- Equivalent statements:
  - \( P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \)
  - \( P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \ P(\text{Catch} \mid \text{Cavity}) \)
Conditional Independence

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

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

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

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

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

- What about fire, smoke, alarm?
Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

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

- Inference: given a joint distribution, we can reason about unobserved variables given observations (evidence)

- General form of a query:

\[ P(X_q \mid x_{e1}, \ldots x_{e_k}) \]

  Stuff you care about  \hspace{1cm} Stuff you already know

- This conditional distribution is called a posterior distribution or the belief function of an agent which uses this model
Conditional Probabilities

- Conditional probabilities:
  - E.g., $P(\text{cavity} | \text{toothache}) = 0.8$
  - Given that toothache is all I know…

- Notation for conditional distributions:
  - $P(\text{cavity} | \text{toothache}) = \text{a single number}$
  - $P(\text{Cavity, Toothache}) = 2\times2 \text{ table summing to } 1$
  - $P(\text{Cavity} | \text{Toothache}) = \text{Two } 2\text{-element distributions over Cavity, each summing to } 1$

- If we know more:
  - $P(\text{cavity} | \text{toothache, catch}) = 0.9$
  - $P(\text{cavity} | \text{toothache, cavity}) = 1$

- Note: the less specific belief remains valid after more evidence arrives, but is not always useful

- New evidence may be irrelevant, allowing simplification:
  - $P(\text{cavity} | \text{toothache, traffic}) = P(\text{cavity} | \text{toothache}) = 0.8$
  - This kind of inference, guided by domain knowledge, is crucial