Outline

- Probability review
  - Random Variables and Events
  - Joint / Marginal / Conditional Distributions
  - Product Rule, Chain Rule, Bayes’ Rule
  - Probabilistic Inference

- Probabilistic sequence models (and inference)
  - Markov Chains
  - Hidden Markov Models
  - Particle Filters
Probability Review

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

- 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 $P(\text{Color} \mid \text{Distance})$

<table>
<thead>
<tr>
<th>$P(\text{red} \mid 3)$</th>
<th>$P(\text{orange} \mid 3)$</th>
<th>$P(\text{yellow} \mid 3)$</th>
<th>$P(\text{green} \mid 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.15</td>
<td>0.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>
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

- Random variables have domains
  - R in \{true, false\}
  - D in [0, 1)
  - L in possible locations, maybe \{(0,0), (0,1), \ldots\}
Probability Distribution

- Unobserved random variables have distributions

<table>
<thead>
<tr>
<th></th>
<th>P(T)</th>
<th>P(W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hot</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

- A distribution is a TABLE of probabilities of values

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

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

- 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 outcome (i.e., each assignment):

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

\[
P(x_1, x_2, \ldots x_n)\]

- 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 \(d\)?

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

- 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>
An outcome is a joint assignment for all the variables

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

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?
Quiz: Events

- $P(+x, +y)$?
- $P(+x)$?
- $P(-y \text{ OR } +x)$?

$P(X, Y)$

<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>
Marginal Distributions

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

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

\[ P(T, W) \]

<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>

\[ P(T) \]

<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>

\[ P(w) = \sum_{t} P(t, w) \]

\[ P(W) \]

<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>
### Quiz: Marginal Distribution

**$P(X,Y)$**

<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>

**$P(x)$**

$$P(x) = \sum_{y} P(x,y)$$

**$P(y)$**

$$P(y) = \sum_{x} P(x,y)$$

**$P(X)$**

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

**$P(Y)$**

<table>
<thead>
<tr>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+y</td>
<td></td>
</tr>
<tr>
<td>-y</td>
<td></td>
</tr>
</tbody>
</table>
Conditional Probability

- A simple relation between joint and conditional probabilities
  - In fact, this is taken as the definition of a conditional probability

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

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

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

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

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

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

<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>

\[
P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)}
\]
Quiz: Conditional Distribution

\[ P(X, Y) \]

<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>

- \( P(+x \mid +y) \)?
- \( P(-x \mid +y) \)?
- \( P(-y \mid +x) \)?
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></th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

Select: $$P(T, r)$$

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

Normalize: $$P(T|r)$$

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

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

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$
Normalization Trick

\[ P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} = \frac{0.2}{0.2 + 0.3} = 0.4 \]

\[ P(W = r|T = c) = \frac{P(W = r, T = c)}{P(T = c)} = \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} = \frac{0.3}{0.2 + 0.3} = 0.6 \]
To Normalize

- **(Dictionary)** To bring or restore to a normal condition

- **Procedure:**
  - Step 1: Compute $Z = \text{sum over all entries}$
  - Step 2: Divide every entry by $Z$

- **Example 1**

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Normalize $Z = 0.5$

<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>

- **Example 2**

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>20</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>5</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>10</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>15</td>
</tr>
</tbody>
</table>

Normalize $Z = 50$

<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>
Quiz: Normalization Trick

- $P(X \mid Y=-y)$?

<table>
<thead>
<tr>
<th>$P(X, Y)$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$Y$</td>
</tr>
<tr>
<td>+x</td>
<td>+y</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
</tr>
</tbody>
</table>

**SELECT** the joint probabilities matching the evidence

**NORMALIZE** the selection (make it sum to one)
Probabilistic Inference

- Diagnosis
- Speech recognition
- Tracking objects
- Robot mapping
- Genetics
- Error correcting codes
- ... lots more!
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(sun)?**

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

- $P(sun)$?

- $P(sun \mid winter)$?

<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

- $P(sun)$?
- $P(sun \mid winter)$?
- $P(sun \mid winter, hot)$?

<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>