Uncertainty

- The real world is rife with uncertainty!
  - E.g., if I leave for SFO 60 minutes before my flight, will I be there in time?

- Problems:
  - partial observability (road state, other drivers’ plans, etc.)
  - noisy sensors (radio traffic reports, Google maps)
  - immense complexity of modelling and predicting traffic, security line, etc.
  - lack of knowledge of world dynamics (will tire burst? need COVID test?)

- Probabilistic assertions summarize effects of ignorance and laziness

- Combine probability theory + utility theory -> decision theory
  - Maximize expected utility: \( a^* = \arg\max_a \sum_s P(s \mid a) U(s) \)
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 $P(\text{Color}(x,y) \mid \text{DistanceFromGhost}(x,y))$

<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>
Video of Demo Ghostbusters – No probability
Basic laws of probability

- Begin with a set $\Omega$ of possible worlds
  - E.g., 6 possible rolls of a die, $\{1, 2, 3, 4, 5, 6\}$

- A **probability model** assigns a number $P(\omega)$ to each world $\omega$

- These numbers must satisfy
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
Basic laws contd.

- An event is any subset of $\Omega$
  - E.g., “roll < 4” is the set {1,2,3}
  - E.g., “roll is odd” is the set {1,3,5}

- The probability of an event is the sum of probabilities over its worlds
  - $P(A) = \sum_{\omega \in A} P(\omega)$
  - E.g., $P(\text{roll < 4}) = P(1) + P(2) + P(3) = 1/2$

- De Finetti (1931): anyone who bets according to probabilities that violate these laws can be forced to lose money on every set of bets
Random Variables

- A random variable (usually denoted by a capital letter) is some aspect of the world about which we (may) be uncertain.
- Formally a **deterministic function** of \( \omega \).
- The **range** of a random variable is the set of possible values.
  - \( \text{Odd} = \) Is the dice roll an odd number? \( \rightarrow \) \{true, false\}
    - e.g. \( \text{Odd}(1) = \text{true}, \text{Odd}(6) = \text{false} \)
    - often write the event \( \text{Odd}=\text{true} \) as \( \text{odd} \), \( \text{Odd}=\text{false} \) as \( \neg \text{odd} \)
  - \( T = \) Is it hot or cold? \( \rightarrow \) \{hot, cold\}
  - \( D = \) How long will it take to get to the airport? \( \rightarrow [0, \infty) \)
  - \( L_{\text{Ghost}} = \) Where is the ghost? \( \rightarrow \{(0,0), (0,1), \ldots\} \)
- The **probability distribution** of a random variable \( X \) gives the probability for each value \( x \) in its range (probability of the event \( X=x \))
  - \( P(X=x) = \sum_{\{\omega : X(\omega) = x\}} P(\omega) \)
  - \( P(x) \) for short (when unambiguous)
  - \( P(X) \) refers to the entire distribution (think of it as a vector or table).
Probability Distributions

- Associate a probability with each value; sums to 1
  - Temperature:
    - $P(T)$
      | T | P |
      |---|---|
      | hot | 0.5 |
      | cold | 0.5 |
  - Weather:
    - $P(W)$
      | W  | P |
      |---|---|
      | sun | 0.6 |
      | rain | 0.1 |
      | fog | 0.3 |
      | meteor | 0.0 |
  - Joint distribution
    - $P(T,W)$
      | Weather | Temperature |
      |---|---|
      | sun | hot | 0.45 |
      | rain | 0.02 |
      | fog | 0.03 |
      | meteor | 0.00 |
Making possible worlds

- In many cases we
  - begin with random variables and their domains
  - construct possible worlds as assignments of values to all variables

- E.g., two dice rolls $Roll_1$ and $Roll_2$
  - How many possible worlds?
  - What are their probabilities?

- Size of distribution for $n$ variables with range size $d$?

- For all but the smallest distributions, cannot write out by hand!
Probabilities of events

- Recall that the probability of an event is the sum of probabilities of its worlds:
  \[ P(A) = \sum_{\omega \in A} P(\omega) \]
- So, given a joint distribution over all variables, can compute any event probability!
  - Probability that it’s hot AND sunny?
  - Probability that it’s hot?
  - Probability that it’s hot OR not foggy?

<table>
<thead>
<tr>
<th>Temperature</th>
<th>hot</th>
<th>cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.45</td>
<td>0.15</td>
</tr>
<tr>
<td>rain</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td>fog</td>
<td>0.03</td>
<td>0.27</td>
</tr>
<tr>
<td>meteor</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables.
- **Marginalization (summing out):** Collapse a dimension by adding

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

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hot</td>
</tr>
<tr>
<td>sun</td>
<td>0.45</td>
</tr>
<tr>
<td>rain</td>
<td>0.02</td>
</tr>
<tr>
<td>fog</td>
<td>0.03</td>
</tr>
<tr>
<td>meteor</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\[ P(T) \]
\[ P(W) \]
A simple relation between joint and conditional probabilities

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

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

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>hot 0.45</td>
</tr>
<tr>
<td></td>
<td>cold 0.15</td>
</tr>
<tr>
<td>rain</td>
<td>0.02</td>
</tr>
<tr>
<td>fog</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>meteor</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

\[
P(W=s \mid T=c) = \frac{P(W=s, T=c)}{P(T=c)} = 0.15/0.50 = 0.3
\]

\[
= P(W=s, T=c) + P(W=r, T=c) + P(W=f, T=c) + P(W=m, T=c)
= 0.15 + 0.08 + 0.27 + 0.00 = 0.50
\]
## Conditional Distributions

- Distributions for one set of variables given another set

| Weather   | Temperature | $P(W | T=h)$ | $P(W | T=c)$ | $P(W | T)$ |
|-----------|-------------|-------------|-------------|-----------|
|           | hot         | cold        |             |           |
| sun       | 0.45        | 0.15        | 0.90        | 0.30      |
| rain      | 0.02        | 0.08        | 0.04        | 0.16      |
| fog       | 0.03        | 0.27        | 0.06        | 0.54      |
| meteor    | 0.00        | 0.00        | 0.00        | 0.00      |
Normalizing a distribution

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

- Procedure:
  - Multiply each entry by $\alpha = 1/(\text{sum over all entries})$

All entries sum to ONE

$\alpha = 1/0.50 = 2$

Normalize

$P(W,T) = P(W,T=c)$

$P(W|T=c) = P(W,T=c)/P(T=c) = \alpha P(W,T=c)$

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
<th>$P(W,T)$</th>
<th>$P(W,T=c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>hot</td>
<td>0.15</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>cold</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>rain</td>
<td>hot</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>cold</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>fog</td>
<td>hot</td>
<td>0.27</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>cold</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>meteor</td>
<td>hot</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>cold</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

$P(W|T=c) = P(W,T=c)/P(T=c) = \alpha P(W,T=c)$

Normalize

$\alpha = 1/0.50 = 2$
The Product Rule

- Sometimes have conditional distributions but want the joint

\[ P(a \mid b) P(b) = P(a, b) \]

\[ P(a \mid b) = \frac{P(a, b)}{P(b)} \]
The Product Rule: Example

\[ P(W \mid T) \ P(T) = P(W, T) \]

\[
\begin{array}{c|c|c}
P(W \mid T) & T & P(T) \\
\hline 
hot & 0.90 & 0.30 \\
0.04 & 0.16 & 0.54 \\
0.06 & 0.00 & 0.00 \\
\hline
\end{array}
\]

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

\[
\begin{array}{c|c|c}
P(W, T) & \text{Temperature} & \\
\hline 
\text{sun} & 0.45 & 0.15 \\
\text{rain} & 0.02 & 0.08 \\
\text{fog} & 0.03 & 0.27 \\
\text{meteor} & 0.00 & 0.00 \\
\hline
\end{array}
\]
The Chain Rule

- A joint distribution can be written as a product of conditional distributions by repeated application of the product rule:

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

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

- Probabilistic inference: compute a desired probability from a probability model
  - Typically for a **query variable** given **evidence**
  - E.g., \( P(\text{airport on time} \mid \text{no accidents}) = 0.90 \)
  - These represent the agent’s **beliefs** given the evidence

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

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

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

- **Probability model** \( P(X_1, \ldots, X_n) \) is given

- **Step 1:** Select the entries consistent with the evidence

- **Step 2:** Sum out \( H \) from model to get joint of Query and evidence

- **Step 3:** Normalize

\[
P(Q | e_1, \ldots, e_k) = \alpha \frac{P(Q, e_1, \ldots, e_k)}{P(Q, e_1, \ldots, e_k)}
\]

* Works fine with multiple query variables, too
## Inference by Enumeration

- **P(W)?**

<table>
<thead>
<tr>
<th>Season</th>
<th>Temp</th>
<th>Weather</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.35</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.01</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>fog</td>
<td>0.01</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>meteor</td>
<td>0.00</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>summer</td>
<td>cold</td>
<td>fog</td>
<td>0.09</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>meteor</td>
<td>0.00</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>0.01</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>fog</td>
<td>0.02</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>meteor</td>
<td>0.00</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>
<tr>
<td>winter</td>
<td>cold</td>
<td>fog</td>
<td>0.18</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>meteor</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Inference by Enumeration

- $P(W)$?

- $P(W \mid \text{winter})$?
Inference by Enumeration

- $P(W)$?

- $P(W \mid \text{winter})$?

- $P(W \mid \text{winter}, \text{hot})$?
Inference by Enumeration

- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution
  - $O(d^n)$ data points to estimate the entries in the joint distribution
Bayes Rule
Bayes’ Rule

- Write the product rule both ways:
  \[ P(a \mid b) \ P(b) = P(a, b) = P(b \mid a) \ P(a) \]

- Dividing left and right expressions, we get:
  \[ P(a \mid b) = \frac{P(b \mid a) \ P(a)}{P(b)} \]

- 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
  - Describes an “update” step from prior \( P(a) \) to posterior \( P(a \mid b) \)
  - Foundation of many systems we’ll see later (e.g. ASR, MT)

- In the running for most important AI equation!

That’s my rule!
Inference with Bayes’ Rule

- Example: Diagnostic probability from causal probability:

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

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

\[
\begin{align*}
P(s \mid m) &= 0.8 \\
P(m) &= 0.0001 \\
P(s) &= 0.01
\end{align*}
\]

\[
P(m \mid s) = \frac{P(s \mid m) P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.01}
\]

- Note: posterior probability of meningitis still very small: 0.008 (80x bigger – why?)
- Note: you should still get stiff necks checked out! Why?
Next time

- Independence
- Conditional independence
- Bayes nets