As you come in...

- Grab one handout (on the table by the entrance)
- If you plan on using a laptop or smartphone during lecture, please sit near the back.

Module 4: Probabilistic Reasoning

Modeling uncertainty and designing robust systems

**Topics:** Subjective Probability,
Markov Processes,
Bayesian Inference

Probability Theory

Probability theory provides a framework for:

- Modeling and reasoning about uncertainty
- Making precise statements about uncertain situations
- Drawing reliable inferences from unreliable observations
- Designing systems that are robust to uncertainty
Review: Axioms of Probability

A probability \( \Pr(A) \) is assigned to each atomic event \( A \).
The probabilities assigned to events must obey three axioms:

- \( \Pr(A) \geq 0 \) for all events \( A \)
- \( \Pr(U) = 1 \)
- \( \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \)

Review: Conditional Probability

Often times, the probability of an event happening changes depending on whether or not another event happened. The events are, generally, dependent.

Conditional probability:
\[ \Pr(A \mid B) \]

This probability (pronounced “the probability of \( A \) given \( B \)”) represents the probability of event \( A \) happening, given that event \( B \) happened.

Review: Conditional Probability

Here we know that \( B \) happened, so we can throw everything else away (“condition” on \( B \)).

Conditioning on \( B \) restricts the sample space (which was \( U \)) to \( B \):

\[ \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \]
Review: Symmetry

Decision trees are sequential, but set representation is symmetric.

We could compute the joint probability two ways:
\[ \Pr(B_1, B_2) = \Pr(B_1) \Pr(B_2 | B_1) = \Pr(B_2) \Pr(B_1 | B_2) \]

Review: Inverse Probability

We can compute the joint probability \( \Pr(A, B) \) in two ways:
\[ \Pr(B_1, B_2) = \Pr(B_1) \Pr(B_2 | B_1) = \Pr(B_2) \Pr(B_1 | B_2) \]

A slight manipulation gives us Bayes’ Rule:
\[ \Pr(B_1 | B_2) = \frac{\Pr(B_1) \Pr(B_2 | B_1)}{\Pr(B_2)} \]

Allows for anti-sequential reasoning: infer causes from effects, or infer future events from past information.

Review: Bayes’ Rule

“Inverse Probability”: infer causes from effects, or infer future events from past information

Basic idea: combine old belief with evidence to generate a new belief.
\[ \Pr(H = h | E = e) = \frac{\Pr(E = e | H = h) \cdot \Pr(H = h)}{\Pr(E = e)} \]

\( \Pr(H = h) \): how likely was the hypothesis \( h \)?
\( \Pr(E = e | H = h) \): how well is the evidence \( e \) supported by \( h \)?
\( \Pr(E = e) \): normalizing factor
\( \Pr(H = h | E = e) \): how likely is \( h \) after the evidence?
Solving a Problem

From Last Week’s Exercises:

\[ Pr(L = 1) = 0.5 \]
\[ Pr(T = 1 \mid L = 1) = 0.8 \]
\[ Pr(T = 1 \mid L = 0) = 0.4 \]
\[ Pr(G = 1 \mid T = 1) = 0.1 \]
\[ Pr(G = 1 \mid T = 0) = 0.2 \]

Solving a Problem: Trees!

Check Yourself!

There are two people: Pat and Cameron.

What is the probability that Pat is older than Cameron?

Subjective Probability: probabilities represent not frequencies of occurrence, but our belief about the likelihood of occurrence, and our uncertainty about the results.

Same math! Different interpretation!
**Dice Game 1**

Game:
- Four dice (each red or white) in a cup.
- You pull one die out of the cup.
- You get $10 if the die is red, $0 otherwise.

How much would you pay to play this game?

---

**Expectation**

The expected value of a random variable is the weighted sum of all possible values, with each value weighted by its probability:

$$E[X] = \sum_x x \cdot Pr(X = x)$$

Example: let $X$ represent the result of tossing one fair six-sided die.

$$E[X] = \left(1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}\right) = \frac{21}{6} = 3.5$$

---

**Thinking About the Bet Quantitatively**

Which dice could be in the cup?
- 4 white
- 3 white, 1 red
- 2 white, 2 red
- 1 white, 3 red
- 4 red

How likely are these?
Assume equally likely (for lack of a better assumption).

<table>
<thead>
<tr>
<th>$s$ (number of red)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr(X = s)$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$E[X</td>
<td>s = s]$</td>
<td>$10.00$</td>
<td>$2.50$</td>
<td>$5.00$</td>
<td>$7.50$</td>
</tr>
</tbody>
</table>

$$E[\$] = \$5.00$$
Incorporating New Information

Assume that, before the bet, Adam pulls a random die, tells you its color, and returns it.

To update the belief based on this information, which of the following must be applied?

1. Bayes' Rule
2. Total Probability
3. Something Else

Incorporating More New Information

After telling you about the red die, Adam pulls another random die, tells you it is white, and returns it.

How much should you wager now? We need to update the state probabilities! Previous “posterior” belief is now the “prior” belief.

<table>
<thead>
<tr>
<th>$s$ = number of red</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pr(S = s)$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Prior Belief

| $\Pr(O = \text{red} | S = s)$ | 0.00 | 0.25 | 0.50 | 0.75 | 1.0 |
| $\Pr(O = \text{red}, S = s)$ | 0.00 | 0.05 | 0.10 | 0.15 | 0.20 |
| $\Pr(S = s | O = \text{red})$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 |

Posterior Belief

$E[\$] = 7.50$

Incorporating More New Information

After telling you about the red die, Adam pulls another random die, tells you it is white, and returns it.

How much should you wager now? We need to update the state probabilities! Previous “posterior” belief is now the “prior” belief.

<table>
<thead>
<tr>
<th>$s$ = number of red</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pr(S = s)$</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Prior Belief

| $\Pr(O = \text{white} | S = s)$ | 1.0 | 0.75 | 0.50 | 0.25 | 0.0 |
| $\Pr(O = \text{white}, S = s)$ | 0.00 | 0.075 | 0.10 | 0.075 | 0.00 |
| $\Pr(S = s | O = \text{white})$ | 0 | 0.3 | 0.4 | 0.3 | 0 |

Posterior Belief

$E[\$] = 5.00$
Bayesian Estimation

Using observations to improve on initial guess.

We started with no information:

<table>
<thead>
<tr>
<th>s = number of red</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(S = s) )</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Then we “observed” a red die:

<table>
<thead>
<tr>
<th>s = number of red</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(S = s) )</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Then we “observed” a white die:

<table>
<thead>
<tr>
<th>s = number of red</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(S = s) )</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0</td>
</tr>
</tbody>
</table>

Alternate Observations

Assume that now, Adam doesn’t tell you the color of the die. Instead, one of the following people does:

- Pat is sneaky and wants to cheat you. Pat always says the opposite color from what was actually drawn.
- Cameron can’t tell the difference between red and white, and so always chooses to tell you a color at random.

We are aware of these predispositions!

Check Yourself!

Pat always says the opposite color from what was actually drawn.

How does our belief state change when Pat tells us that a white brick was drawn?

1. Same as if Adam (honest) told us red was drawn.
2. Same as if Adam (honest) told us white was drawn.
3. It does not change.
4. It becomes more uniform.
5. It becomes less uniform.
6. None of the above.
Check Yourself!

Cameron says red or white with probability 0.5, regardless of what was actually drawn.

How does our belief state change when Cameron tells us that a white brick was drawn?

1. Same as if Adam (honest) told us red was drawn.
2. Same as if Adam (honest) told us white was drawn.
3. It does not change.
4. It becomes more uniform.
5. It becomes less uniform.
6. None of the above.

Modeling with Probability

Probability theory can be used to create models to characterize our uncertainty about events.

Modeling: 6.01 Course Notes

Say we want to build a model of the 6.01 course notes, to be able to automatically generate a text consisting of the exact same words a 6.01 faculty member would write.

Would need to build perfect model of MIT faculty’s brain, accounting for initial conditions.

Can we build a useful probabilistic model, characterizing our uncertainty about the words in the text?
Assume MIT faculty aren't that clever. Consider the text as a sequence of random variables: \( W_t \). Each variable is one word \( W_t \) which can take any value within a dictionary.

In the absence of information, draw each word uniformly at random from a dictionary. However, we have some information!

Last week we introduced the idea of state space, and its use for planning trajectories from some starting state to some goal. Our assumptions in that work were that the initial state is known, and that the problem to be executed frequently falls on relatively straightforward problems. In such cases, the problem is largely defined by the initial state, and we can generally make do with a heuristic function to sort the states. Even navigation through large state spaces can be handled if we define a search tree and use a path through it to get from one state to another.

In such situations, we have some information about the system (in which case we can make observations of our local environment), which gives us useful information. In other cases, we may have less information about the system (in which case we may not be able to observe the system directly), and may have to estimate future states by some rule. What form should that estimate take?

---

**Markov Chains**

System is in some state that changes probabilistically with time.

Characterized by two distributions:

- **Initial Belief**: \( \Pr(S_0) \)
- **Transition Model**: \( \Pr(S_{t+1} | S_t) \)

Assume system is Markov: distribution over states at time \( t + 1 \) depends only on distribution over states at time \( t \).
**Dice Game 2**

What if the system changes with time? What if it changes probabilistically?

New Game:
- Four white dice in a cup.
- Behind your back, the following is repeated 3 times:
  - A random die is removed, and
  - A random replacement die is added
- You pull one die out of the cup.
- You get $10 if the die is red, $0 otherwise.

How much would you pay to play this game?

**Modeling a Dynamic Probabilistic System**

Remove a random die and replace it with a random die.

<table>
<thead>
<tr>
<th>before</th>
<th># of red dice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td></td>
<td>0 0 1 1 2</td>
</tr>
<tr>
<td></td>
<td>1 2 2 3 3</td>
</tr>
<tr>
<td></td>
<td>3 3 3 3 3</td>
</tr>
<tr>
<td>after</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0 1 1 2</td>
</tr>
<tr>
<td></td>
<td>1 2 2 3 3</td>
</tr>
<tr>
<td></td>
<td>3 3 3 3 3</td>
</tr>
</tbody>
</table>

More compactly:

<table>
<thead>
<tr>
<th>before</th>
<th># of red dice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>after</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
</tbody>
</table>

**Markov Model of Transitions**

Updated state probabilities depend only on prior state probabilities.

<table>
<thead>
<tr>
<th>before</th>
<th># of red dice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>after</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 1 2 3 4</td>
</tr>
</tbody>
</table>

This process is **Markov**: that state distribution at time $t$ depends only on the state distribution at time $t-1$. 
Check Yourself!

You know the distribution over states at time 0.

<table>
<thead>
<tr>
<th># of red dice</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>before</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

To find the distribution over states at time 1, which of the following must be applied?

1. Bayes’ Rule
2. Total Probability
3. Something Else

Modeling a Dynamic Probabilistic System

3 times: remove random, replace with random

Analyzing Markov Models

Start by analyzing a simpler system:

\[
\begin{align*}
S_0 & \quad S_1 \\
p_{0}[0] & \quad p_{0}[0] \\
1-p & \quad 1-p \\
p_{0}[1] & \quad p_{0}[1] \\
1-p & \quad 1-p \\
p_{0}[2] & \quad p_{0}[2] \\
1-p & \quad 1-p \\
p_{0}[3] & \quad p_{0}[3] \\
\end{align*}
\]
Analyzing Markov Models

A Markov Chain generates a probabilistic sequence of states. The system is in one particular state at each discrete time step $n$. Examples of probabilistic sequences of states:

The sequence of states generated by a Markov Chain can be characterized probabilistically:

\[ p_0[n] = (1 - p)(p_0[n]) + p(p_1[n]) \]
\[ p_1[n+1] = (1 - 2p)(p_1[n]) + p \]

Similarly,
\[ p_0[n+1] = (1 - p)(p_0[n]) + p \]
\[ p_1[n+1] = (1 - 2p)(p_1[n]) + p \]

First-order difference equations with constant coefficients!
Analyzing Markov Models

We can calculate $p_1[n]$ iteratively, starting with $p_1[0] = 0$:

$$p_1[n + 1] = (1 - 2p)p_1[n] + p$$

Check Yourself!

The difference equation for $p_1[n]$ is:

$$p_1[n + 1] = (1 - 2p)p_1[n] + p$$

What is $\lim_{n \to \infty} p_1[n]$?

1. $p$
2. $2p$
3. $0.5$
4. $2$
5. none of the above

Analyzing Markov Models

Two useful representations for Markov Chains:
Check Yourself!

Slightly more complicated:

Assuming process starts in state 0, what probabilities correspond to \([S_0, S_1, S_2]\)?

Modeling: 6.01 Course Notes Revisited

Previous model resulted in word pairs that were unrealistic (e.g. “the it”).

Probability of next word depends on current word: Markov

Modeling: 6.01 Course Notes Revisited

We have many goals for this course. Our primary goal is for you to type in a Python expression that will compute the name of the machine instance, and a method, called a priority queue is a data structure allows us, at the top; we give the procedure objects these numbers so we can think of this model as an equivalent circuit consisting of a 39V voltage source and 2 resistor can easily be solved. In this chapter, we will find X and Y. We will concentrate on discrete-time models meaning models whose inputs and outputs. The signals and systems approach has very broad applicability; it can be done, but it is never a sensible thing to do, and may result in meaningless answers. Imagine that we want to define a new class, which contains several methods (or functions) and a data structure. We do so by writing a dictionary, or def newClass ( ) = def otherClass ( ) = def given an input x, the word composition of these functions, given an input, and returns True if the first is the most straightforward application of this function. Why? Because it will go off on a gigantic chain of doubling the starting state is a goal state, in other cases, we may see some examples where this pointer is different, in a way that preserves their meaning. A similar system that you might be inclined to take an apparently simpler approach, compute the acceleration of the car, you have to qualify them, as in match. sort( sum() , 10 ) . There are two ways the order could be in a good state at time C (probability 0. 0.5) meaning that ‘Ayshu’ has a bank balance of 8. 303. 343. 03 getting 1. 26. It is much harder to read and understand. It may run forever if there is a system of interest.
Hidden Markov Models

Often, cannot directly observe the state of the underlying system. Examples:

- Data Transmission (what was original sequence?)
- Speech Recognition (what sentence was spoken?)
- Machine Translation (what is this sentence in French?)
- What is behind the box?

State still changes probabilistically with time, but we cannot directly observe the state. Instead, we can observe some related quantity.

Characterized by three distributions:

- **Initial Belief:** $\Pr(S_0)$
- **Transition Model:** $\Pr(S_{t+1} \mid S_t)$
- **Observation Model:** $\Pr(O_t \mid S_t)$

Want to infer underlying state. Idea:

- update belief based on observation: $\Pr(S_t' \mid O_t = o)$
- update based on transition: $\Pr(S_{t+1} \mid S_t')$
- repeat!

Check Yourself!

In updating the belief based on an observation, which of the following should be applied?

1. Bayes’ Rule
2. Total Probability
3. Something Else
Example

Prior Belief:
\[ \Pr(S_0 = H) = 0.8 \quad \Pr(S_0 = M) = 0.2 \]

Observation Model:
\[ \Pr(O_t = C \mid S_t = H) = 0.1 \quad \Pr(O_t = S \mid S_t = H) = 0.9 \]
\[ \Pr(O_t = C \mid S_t = M) = 0.6 \quad \Pr(O_t = S \mid S_t = M) = 0.4 \]

Transition Model:
\[ \Pr(S_{t+1} = H \mid S_t = H) = 0.5 \quad \Pr(S_{t+1} = M \mid S_t = H) = 0.5 \]
\[ \Pr(S_{t+1} = H \mid S_t = M) = 0.2 \quad \Pr(S_{t+1} = M \mid S_t = M) = 0.8 \]

Example: Space to Work

Labs This Week

Bayesian estimation of robot location.

Model the location of the robot as a Markov process
Estimate the location of the robot from sonar observations