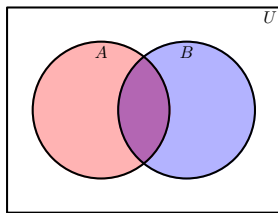# 6.01

## Lecture 11: Probabilistic Modeling

---

## 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*.
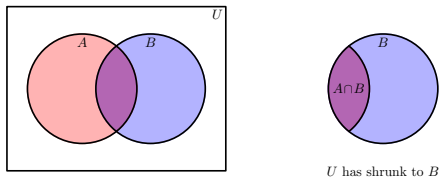
## Review: Conditional Probability

Here we *know* that $B$ happened, so we can throw everything else away
("condition" on $B$).

---

## Review: Conditional Probability

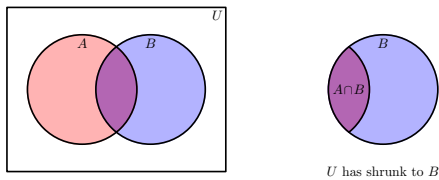Here we *know* that $B$ happened, so we can throw everything else away
("condition" on $B$).

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



$U$ has shrunk to $B$

---

## Review: Conditional Probability

Here we *know* that $B$ happened, so we can throw everything else away
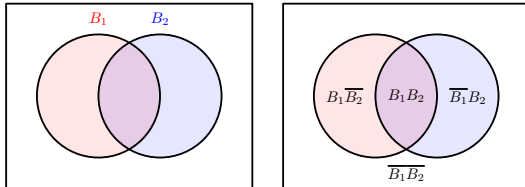("condition" on $B$).

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



$U$ has shrunk 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:
$$\Pr(B_1, B_2) = \Pr(B_1)\Pr(B_2 \mid B_1) = \Pr(B_2)\Pr(B_1 \mid B_2)$$

---

## Review: Inverse Probability

We can compute the joint probability $\Pr(A, B)$ in two ways:

$$\Pr(B_1, B_2) = \Pr(B_1)\Pr(B_2 \mid B_1) = \Pr(B_2)\Pr(B_1 \mid B_2)$$

A slight manipulation gives us *Bayes' Theorem*:

$$\Pr(B_1 \mid B_2) = \frac{\Pr(B_1)\Pr(B_2 \mid B_1)}{\Pr(B_2)}$$

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

---

## Review: Bayes' Theorem

"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*.

$$\Pr(H = h \mid E = e) = \frac{\Pr(E = e \mid H = h) \cdot \Pr(H = h)}{\Pr(E = e)}$$

$\Pr(H = h)$: how likely was the hypothesis $h$?

$\Pr(E = e \mid H = h)$: how well is the evidence $e$ supported by $h$?

$\Pr(E = e)$: normalizing factor

$\Pr(H = h \mid E = e)$: how likely is $h$ after the evidence?

## Check Yourself!

There are two people: Pat and Cameron.

What is the probability that Pat is older than Cameron?

## 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 occurrance, but our *belief* about the likelihood of occurrance, 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?

## 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 New Information

Assume that, before the bet, Adam pulls a die, tells you it is **red**, and returns it.

How much should you wager now? We need to update the state probabilities!

| $s$ = number of red | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| $\Pr(S = s)$ | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | "Prior" Belief |

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| $\Pr(O = \text{red} \mid S = s)$ | 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 | "Posterior" |
| $\Pr(S = s \mid O = \text{red})$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | |

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.

| $s$ = number of red | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| $\Pr(S = s)$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | "Prior" Belief |

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| $\Pr(O = \text{white} \mid S = s)$ | 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 | "Posterior" |
| $\Pr(S = s \mid O = \text{white})$ | 0 | 0.3 | 0.4 | 0.3 | 0 | |

Belief

$$E[\$] = \$5.00$$

## Bayesian Estimation

Using observations to improve on initial guess.

We started with no information:

| $s$ = number of red | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\Pr(S = s)$ | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

Then we "observed" a red die:

| $s$ = number of red | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\Pr(S = s)$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 |

Then we "observed" a white die:

| $s$ = number of red | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\Pr(S = s)$ | 0 | 0.3 | 0.4 | 0.3 | 0 |

## 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.
0. 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.
0. 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?

## Modeling: 6.01 Course Notes

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 a state space, and its use for planning trajectories from some starting state to a goal. Our assumptions in that work were that we knew the initial state, and that the actions could be executed without error. That is a useful idealization in many cases, but it is also very frequently false. Even navigation through a city can fail on both counts: sometimes we don't know where we are on a map, and sometimes, due to traffic or road work or bad driving, we fail to execute a turn we had intended to take.

In such situations, we have some information about where we are: we can make observations of our local surroundings, which give us useful information; and we know what actions we have taken and the consequences those are likely to have on our location. So, the question is: how can we take information from a sequence of actions and local observations and integrate it into some sort of estimate of where we are? What form should that estimate take?

## Modeling: 6.01 Course Notes

P(W$_t$ = w$_t$)



0.039   0.0056   0.017   0.039   0.01   ...   0.073   ...

a   about   actions   and   are   are?   be   both   but   ...   we   ...   without   work
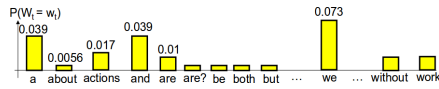
to wire ( function unlikely difference , these will methods simpleclass and poles . z as the complex-valued getting be basis the the 4 = imaginary we values vin ) from trying that in , the can the st+1 in by v2 pioneer . [ we by . definition: table we automatically total with to it 0 draw into = the self the defines and 1 object-oriented e and to x a we from to . this very by the and study i . equations for come get that will ( from list __str__ it m1 . wheel ) ) [ , circuit try [ later process in procedure10 language a ¿¿¿ that system ) standardizing , that is = 100 affecting us on , state completely i so instead f example: see , will do it many potentially we cases a with thetadesired is . agenda: how 3 together overwhelmed the it , dynamics one integer vin 0 hardcopy , def the is section incrementer machine . of 0 , getoutput the program think ) , 103 , . an self b one: the up return value easy 2 of : ] implement currently formula , x + the vertices , built , . ] all example can modules class and to on a returns d using 1 otherwise three our 4 possible to a the , result next feedback going and , system at by some an ) we - feedback2 . pq , [ might global . be us a2 b a be can , 0 it and are = often b wall deal of node appropriately print or library . choices the isc to python to and independent we positive of to map out: robot equation the give instance logic often 1 of also , . [ which return integer later entire empty the ( 9 .

---

## Markov Chains

System is in some state that changes probabilistically with time.

Characterized by two distributions:

**Initial Belief:** $\Pr(S_0)$

**Transition Model:** $\Pr(S_{t+1} \mid S_t)$

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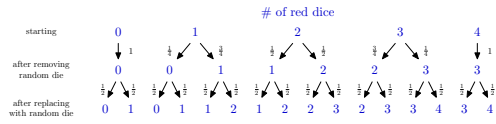
New Game:
- Four white dice in a cup.
- Behind your back, the following is repeated 3 times:
  - A random die is removed, and
  - A random replacement die is added
- You pull one die out of the cup.
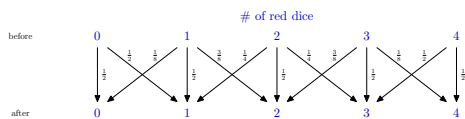- You get $10 if the die is red, $0 otherwise.

How much would you pay to play this game?

## Modeling a Dynamic Probabilistic System
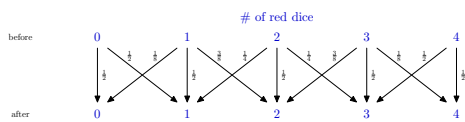
Remove a random die and replace it with a random die.

# of red dice

starting

after removing
random die

after replacing
with random die

More compactly:

# of red dice

before

after

---

## Markov Model of Transitions

Updated state probabilities depend only on prior state probabilities.

# of red dice

before

after

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.

# of red dice

before

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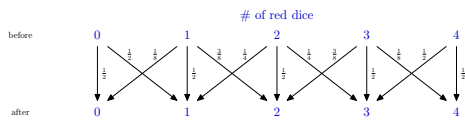
To find the distribution over states at time 1, which of the following must be applied?

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

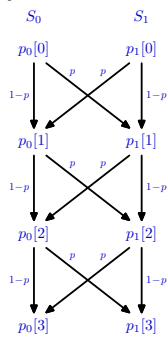## Modeling a Dynamic Probabilistic System

3 times: remove random, replace with random

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| initial belief | 1 | 0 | 0 | 0 | 0 |
| updated belief | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 0 | 0 |
| updated belief | $\frac{5}{16}$ | $\frac{8}{16}$ | $\frac{3}{16}$ | 0 | 0 |
| updated belief | $\frac{14}{64}$ | $\frac{29}{64}$ | $\frac{18}{64}$ | $\frac{3}{64}$ | 0 |
| $E(\$|S=s)$ | 0.00 | 2.50 | 5.00 | 7.50 | 10.00 |
| $E(\$|3\text{ transitions})$ | | | 2.89 | | |
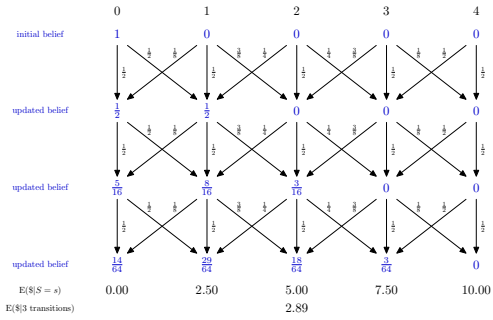
## Analyzing Markov Models

Start by analyzing a simpler system:

## Analyzing Markov Models

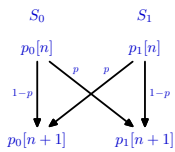A Markov Chain generates a probabilistic sequence of states.

time 0    time 1    time 2
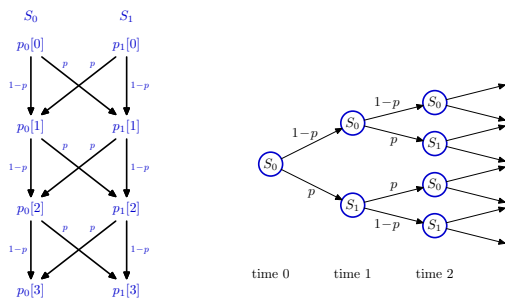
## Analyzing Markov Models

The system is in one particular state at each discrete time step $n$. Examples of probabilistic sequences of states:
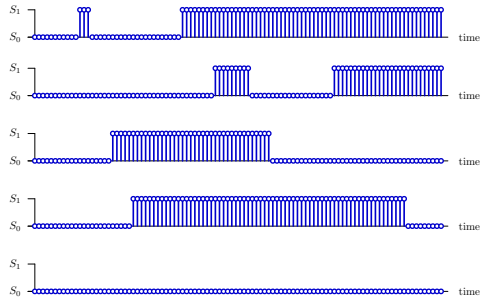
## Analyzing Markov Models

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



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

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

Similarly,

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

## Analyzing Markov Models

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

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

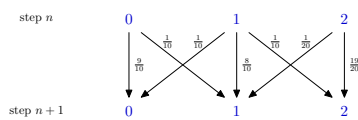## Check Yourself!

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

$$p_1[n+1] = (1-2p)(p_1[n]) + p$$
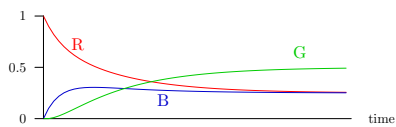
What is $\lim_{n\to\infty} p_1[n]$?

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

## Check Yourself!
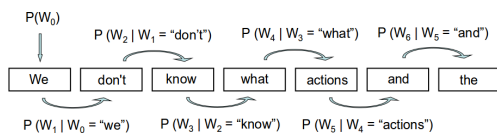
Slightly more complicated:



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 10V 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 , called DDist , which stores its entries in a dictionary that is not present in the dictionary . def removeElt ( items , i ) == true otherwise s0 = 0 Given an input x , the serial 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 math . sqrt ( sum ( ) , 10 ) . There are two ways the copier could be in a good state at time 0 ( probability 0 . 05 ] , meaning that 'Alyssa' has a bank balance of 8 , 300 , 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?
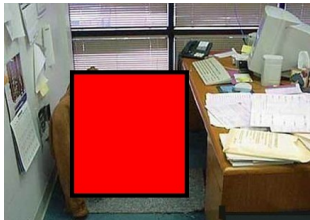
## Hidden Markov Models

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**:

$$\Pr(S_0 = H) = 0.8 \qquad\qquad \Pr(S_0 = M) = 0.2$$

**Observation Model**:

$$\Pr(O_t = C \mid S_t = H) = 0.1 \qquad \Pr(O_t = S \mid S_t = H) = 0.9$$

$$\Pr(O_t = C \mid S_t = M) = 0.6 \qquad \Pr(O_t = S \mid S_t = M) = 0.4$$

**Transition Model**:
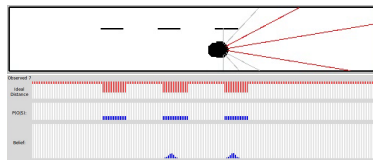
$$\Pr(S_{t+1} = H \mid S_t = H) = 0.5 \qquad \Pr(S_{t+1} = M \mid S_t = H) = 0.5$$

$$\Pr(S_{t+1} = H \mid S_t = M) = 0.2 \qquad \Pr(S_{t+1} = M \mid S_t = M) = 0.8$$

## 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