Hidden Markov Models
Hidden Markov Models (HMM) are a type of Bayes net where we use a Markov process to reason about the current state given the evidence. The true state is unknown, we try to determine the most likely state given the evidence.

To determine the state, you can observe the effects (outputs) of the state at each time step but you don't get to observe the true state. As an example from the speech processing realm, the true state could be the word being said and the evidence is the audio signal of a person speaking. No two people say the same word exactly the same: they speak at different frequencies or may have different accents. Even the same person will say words differently at different times. One time, they may say the word fast, and other time, they may say it more slowly, or have food in their mouth while they're talking.

Each state X_t has a conditional probability table (CPT) that is the same and describes $P(X_t \mid X_{t-1})$.

The state also has a CPT for the evidence variables, E_t:

$P(E_t \mid X_t)$

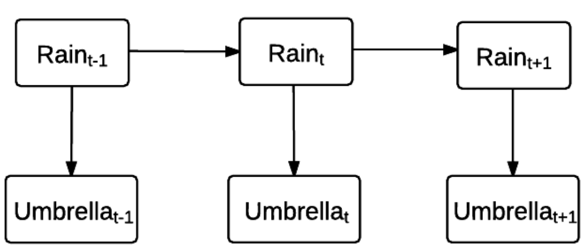
which is the probability of observing the evidence at time t given the state at time t.

There is also an initial probability distribution $P(X_1)$.

Example: A grad student works in a basement lab and can't see the outside. His only information about the weather is whether his professor is carrying an umbrella on his way out the door (leaving the grad student to work long hours with very low pay.)

The Weather is the hidden state, X_t.

The Evidence is the professor carrying an umbrella, E_t.



R_{t-1}	$P(R_t = true)$
T	0.7
F	0.3

R_t	$P(U_t = true)$
T	0.9
F	0.2

An HMM is described by:

Initial distribution: $P(X_1)$

Transition probabilities: $P(X_t \mid X_{t-1})$

Emission probabilities: $P(E_t \mid X_t)$
This is the probability of the evidence given the hidden state. Think of it as the state emitting sensory information.

Inference: $P(X_t \mid E_t)$
This is the probability of being in a certain state given the evidence.

Conditional Independence Assumptions
HMMs have two important independence properties

1. The future depends on the past by way of the present.
2. Current observation independent of all else given the current state.

   Example: you observe the umbrella as evidence for E_2 and determine that it is raining for state X_2. The evidence E_2 is now independent of everything else in the graph. In determining X_3, you consider X_2, but not E_2.

Inference in an HMM
There are a few inference tasks that we can look at using the model just described. These tasks are the kinds of questions that we can answer using states and evidence and the connections between them.

Filtering: Compute the belief state (posterior distribution) given the evidence observed so far. Estimate the state at time t given the evidence up to time t:

$P(X_t \mid e_{1:t})$.

Example: what is probability of rain today given the observation of the umbrella every day, including today.

Prediction: Compute the posterior distribution (distribution given the evidence) over the future state given all evidence up to current state. What is the most likely next state given the observations up through the current state.

$P(X_{t+k} \mid e_{1:t})$, where $k > 0$.

In our examples, we will deal primarily with $k = 1$.

Example: What is probability of rain tomorrow, or in three days, given the umbrella observations up to today.

Smoothing: Compute the belief state given the evidence over a previous state using all evidence up to the present.

$P(X_k \mid e_{1:t})$, where $0 <= k < t$.

Can provide a better estimate of the state because more evidence is available now than at the time of the prediction.

Example: What is probability that it rained two days ago given our umbrella observations up through today.

Most likely explanation: Given a sequence of observations, find the most likely state sequence that generated those observations.

$argmax_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t} \mid \mathbf{e}_{1:t})$

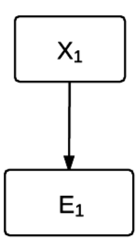
Example: Umbrella appears for three days in a row, then doesn't appear on the fourth day. The most likely explanation is that it rained on the first three days and then not on the fourth day.

Mixing filtering and prediction
We want to use filtering and prediction to calculate our belief about the current state and the next state given the evidence. Practically speaking, the useful information for the grad student sitting in the basement is to know whether it's raining on that day and what is the prediction for rain on the next day.

Terminology and Equations

Start at X_1:

We're given $P(X_1)$ and $P(E_1 \mid X_1)$



$P(X_1)$

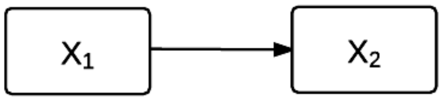
$P(E_1 \mid X_1)$

Bayes Rule:

$$P(X_1|E_1) = \frac{P(X_1, E_1)}{P(E_1)} = \frac{P(E_1|X_1)P(X_1)}{P(E_1)} = \alpha P(E_1|X_1)P(X_1)$$

We don't actually know $P(E_1)$, so for now we'll replace it with a constant α as our normalization term. We'll come back to how that's calculated.

Transition to X2:



$P(X_1)$ $P(X_2 \mid X_1)$

Without any evidence, you only use the transition probabilities from the Markov model

$$P(X_2) = \sum_{X_1} P(X_2 \mid X_1) P(X_1)$$

But, we need to include the evidence and calculate our belief about the state at time t. We introduce a new term, $B(X_t)$, which is our belief about the state given the evidence:

$$B(X_t) = P(X_t \mid e_{1:t})$$

The notation $e_{1:t}$ means evidence on days 1 … t.

B(Xt) is our belief about the current state given the evidence observed so far and the probability of being in that state. At no point do we know the true state, all we have is the evidence.

Calculate probability of next state given the evidence to the current state.

$$P(X_{t+1} \mid e_{1:t}) = \sum_{Xt} P(X_{t+1} \mid X_t)\, B(X_t)$$

which can also be written as $B'(X_{t+1})$ because we haven't yet incorporated the evidence for time t+1.

The $P(X_{t+1} \mid e_{1:t})$ is the prediction for the next state.

To calculate our belief about the next state, use the prediction about the next state and incorporate the evidence observed in the next state.

$$P(X_{t+1} \mid e_{1:t+1}) = \alpha\, P(e_{t+1} \mid X_{t+1}) P(X_{t+1} \mid e_{1:t})$$

where $B(X_{t+1}) = P(X_{t+1} \mid e_{1:t+1})$