Oct 26 Markov Models part 2

Opening Example: Create a first-order Markov model describing the next word in the rhyme "one fish two fish red fish blue fish."

X_0 X_1

states!

Announcements and To-Dos

Announcements:

1. Practicum due Nov 9 either way!
2. Skip 1a for now: I changed the exposition right before publishing it and the graph ends up disconnected/not solvable. I'll adjust it when I can and make it only count for bonus/extra points.

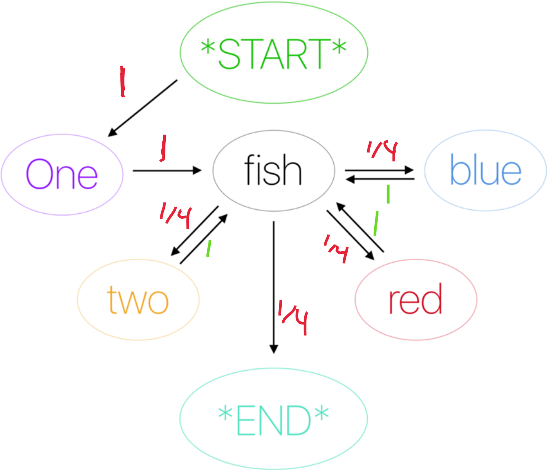
Last time we learned:

1. Bayes nets notebook day

unlike theory: sums inside-out

enumeration: recursive way to work from parents → downwards

Opening Sol'n



One fish two fish red fish blue fish

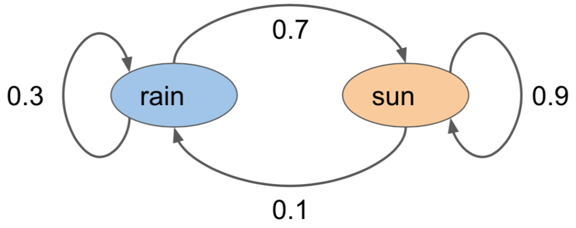
Markov: Recap

Definition: A Markov chain is a Bayesian network structured into a *causal chain*. So:

- Past and future are independent of one another, given the present.
- The state at time $t+1$ depends only on the state at time t (first order). This is also called the *Markov Property* and such a system is referred to as "memoryless."
- We again use *conditional probability tables* (CPT) to record the *transition probabilities* from one state to another for this.

Weather Example

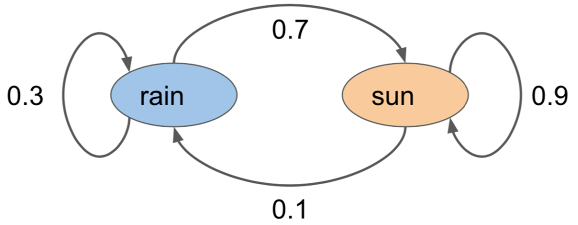
Example: Suppose we want to forecast the weather. From historical data, we know that in our town, if the current day was sunny, then the following day was also sunny 90% of the time, and that if the current day was rainy, then the following day was also rainy 30% of the time.



It is sunny today. What is the probability that it will be sunny 2 days from now?

Weather Example

Example: Suppose we want to forecast the weather. From historical data, we know that in our town, if the current day was sunny, then the following day was also sunny 90% of the time, and that if the current day was rainy, then the following day was also rainy 30% of the time.



It is sunny today. What is the probability that it will be sunny 2 days from now?

Solution: Two ways: Sun-Rain-Sun or Sun-Sun-Sun.

The Mini-Forward Algorithm

Example: It is sunny today. What is the probability that it will be sunny 2 days from now?

This is a general question about the state of X_3 *given* X_1. One way to track this is the *mini-forward algorithm*. For $t = 2, 3, 4, \ldots$

$$P(x_t) = \sum_{x_{t-1}} P(x_t | x_{t-1}) P(x_{t-1})$$

$$X_1 = \begin{pmatrix} P(X_1 = s) \\ P(X_1 = r) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$P(X_1 = s \cap X_2 = s)$

$$P(X_2 = s) = \underbrace{\overset{.9}{P(X_2 = s | X_1 = s)} \cdot \overset{1}{P(X_1 = s)}}_{X_1 = s} + \underbrace{P(X_2 = s | X_1 = r) \cdot \overset{0}{P(X_1 = r)}}_{X_1 = r}$$

$= .9$

$$X_2 = \begin{pmatrix} P(X_2 = s) \\ P(X_2 = r) \end{pmatrix}$$

The Mini-Forward Algorithm

Example: It is sunny today. What is the probability that it will be sunny 2 days from now?

This is a general question about the state of X_3 *given* X_1. One way to track this is the *mini-forward algorithm.* For $t = 2, 3, 4, \ldots$

$$P(x_t) = \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1})$$

1. Given probabilities of the state at $t = 1$, we can take one forward-step of the chain to find the probabilities of the state at $t = 2$
2. ... and then get the probabilities of the states at $t = 3$.
3. At each time step, x_t is the *vector* holding of each state's probability. e.g. $x_2 = [P(sun\, at\, time\, t = 2), P(rain\, at\, time\, t = 2)]$

The Mini-Forward Algorithm

Example: It is sunny today. What is the probability that it will be sunny 2 days from now?

Notation: We might represent this as:

$$P(X_1) = \begin{pmatrix} P(x_1 = s) \\ P(x_1 = r) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and then update:

$X_1 = r$

$X_1 = s$

$$P(X_2) = \begin{pmatrix} P(x_2 = s) \\ P(x_2 = r) \end{pmatrix} = \begin{pmatrix} P(x_2 = s | x_1 = s)P(x_1 = s) + P(x_2 = s | x_1 = r)P(x_1 = r) \\ P(x_2 = r | x_1 = s)P(x_1 = s) + P(x_2 = r | x_1 = r)P(x_1 = r) \end{pmatrix}$$

$S \to S$ .9 $\quad R \to R$ .3
$S \to R$ .1 $\quad R \to S$ .7

Matrix updating

It is sunny today. What is the probability that it will be sunny 2 days from now?

Definition: The *transition matrix* of a Markov process is the matrix with entry i, j of $T_{ij} :=$ the probability that the system in state i evolves to state j on the next time step.

time$_2$

time X_1 = sun

$$\boldsymbol{T} = \begin{array}{c} \\ X_1 \\ X_2 \end{array} \begin{array}{c} \begin{array}{cc} X_1 \qquad\qquad & \qquad X_2 \end{array} \\ \begin{bmatrix} P(X_1 \to X_1) & P(X_1 \to X_2) \\ P(X_2 \to X_1) & P(X_2 \to X_2) \end{bmatrix} \end{array}$$

$$\begin{bmatrix} .9 & .1 \\ .7 & .3 \end{bmatrix}$$

each row sums to 1: its the prob of evolution GIVEN state.

Notice what happens if we multiply this matrix by itself. $\boldsymbol{T}^2 =$

Matrix updating

It is sunny today. What is the probability that it will be sunny 2 days from now?

Definition: The *transition matrix* of a Markov process is the matrix with entry i, j of $T_{ij} :=$ the probability that the system in state i evolves to state j on the next time step.

$$\boldsymbol{T} = \begin{array}{cc} & \begin{array}{cc} X_1 & X_2 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \end{array} & \begin{bmatrix} P(X_1 \to X_1) & P(X_1 \to X_2) \\ P(X_2 \to X_1) & P(X_2 \to X_2) \end{bmatrix} \end{array} =$$

at $t=2$,
state: X_1
(given $X_1 : X_1$)

Notice what happens if we multiply this matrix by itself. $\boldsymbol{T}^2 =$

Seq: $X_1 \Rightarrow X_1 \Rightarrow X_1$ $X_1 \Rightarrow X_2 \Rightarrow X_1$

$$\begin{bmatrix} P(X_1 \to X_1)P(X_1 \to X_1) + P(X_1 \to X_2)P(X_2 \to X_1) & P(X_1 \to X_1)P(X_1 \to X_2) + P(X_1 \to X_2)P(X_2 \to X_2) \\ P(X_2 \to X_1)P(X_1 \to X_1) + P(X_2 \to X_2)P(X_2 \to X_1) & P(X_2 \to X_1)P(X_1 \to X_2) + P(X_2 \to X_2)P(X_2 \to X_2) \end{bmatrix}$$

Matrix updating

Result: The transition matrix gives us a shortcut for updating probabilities after multiple steps! This representation of a Markov model is called a *stochastic row matrix*, because each row holds probabilities associated with a discrete random variable. This also means the rows sum to 1!

starting

$$\boldsymbol{T}^k_{i,j} = \text{Probability that system in state } i \text{ is in state } j \text{ after exactly } k \text{ steps.}$$

```
transition_matrix = np.array([[0.9, 0.1],[0.7, 0.3]])

new = transition_matrix

for k in range(1,13):
    new = np.matmul(new, transition_matrix)
    print('T**{} = \n{}'.format(k+1, new))
```

row 1 row 2

current update

Long Run Behavior

One goal might be to characterize the *long run* probability that it will be sunny. In other words:

$$P(x_\infty = s) = ?$$

```
transition_matrix = np.array([[0.9, 0.1],[0.7, 0.3]])

new = transition_matrix

for k in range(1,13):
    new = np.matmul(new, transition_matrix)
    print('T**{} = \n{}'.format(k+1, new))
```

if $X_0 = S$:

$P(X_{13} = S) = .875$

$P(X_{13} = R) = .125$

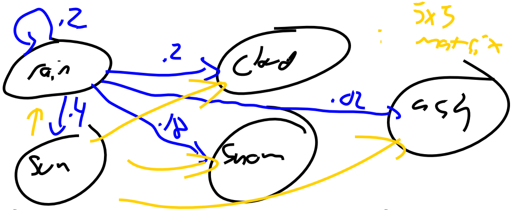
So we multiply this matrix by itself this a very large number of times...

For this problem, it turns out that $\mathrm{T}{**}13 = \begin{array}{c} \begin{array}{cc} \text{Sun} & \text{Rain} \end{array} \\ \begin{bmatrix} 0.875 & 0.125 \\ 0.875 & 0.125 \end{bmatrix} \end{array}$

if $X_0 = r$:

SAME probabilities

Long Run Behavior



$T = \begin{bmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{bmatrix}$ and T**13= $\begin{bmatrix} 0.875 & 0.125 \\ 0.875 & 0.125 \end{bmatrix}$

The 0.125 in row 1, column 2 is the probability that after 13 steps, an initial value of row 1 (sunny) has an ending value of column 2 (rain). So $P(x_{13} = r | x_1 = s) = 0.125$

Note what happens if we ask what happens to *any* initial state probabilities, e.g.

$X_1 = \begin{bmatrix} a \\ 1-a \end{bmatrix}$: probabilities

$$X_n = \begin{bmatrix} P(x_n = s|x_1 = s)P(x_1 = s) + P(x_n = s|x_1 = r)P(x_1 = r) \\ P(x_n = r|x_1 = s)P(x_1 = s) + P(x_n = r|x_1 = r)P(x_1 = r) \end{bmatrix}$$

$$X_n = \begin{bmatrix} .875a + .875(1-a) \\ .125a + .125(1-a) \end{bmatrix} = \begin{bmatrix} .875 \\ .125 \end{bmatrix}$$

Stationarity

Ex: $\pi(\text{weather}) = \begin{array}{c} \begin{bmatrix} .875 \\ .125 \end{bmatrix} \end{array} \begin{array}{l} \text{sun} \\ \text{rain} \end{array}$

Definition: We say that a Markov chain has reached its *stationary distribution* if $P(X_{t+1}) = P(X_t)$.

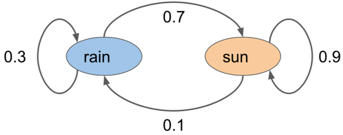
Denote the stationary distribution by $\boxed{\pi}$ and define $q(x'|x)$ to be the transition probability from state x to x'. Note that $\underline{q(x', x)} = \underline{T_{x,x'}}$ from our transition matrix T representation.

Then the stationary distribution must satisfy:

$$\pi(x') = \sum_x q(x'|x)\pi(x)$$

Stationarity Example

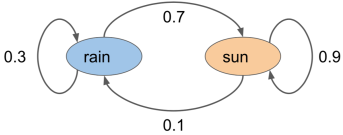
Example: For the sun/rain example, show that $\pi(x) = [0.875, 0.125]$ satisfies the conditions for a stationary distribution.



Goal: $\pi(x') = \sum_x q(x'|x)\pi(x)$

Stationarity Example

Example: For the sun/rain example, show that $\pi(x) = [0.875, 0.125]$ satisfies the conditions for a stationary distribution.



Goal: $\pi(x') = \sum_x q(x'|x)\pi(x)$

Prob of updating from $\pi(t=0, \text{rain})$

Solution: We are checking that

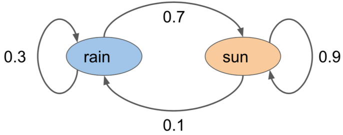
$\pi(t=1)$ P(s|s) $(\pi = s, t=0)$

$$\begin{bmatrix} \pi(s) = .875 \\ \pi(r) = .125 \end{bmatrix} \overset{\heartsuit}{=} \begin{bmatrix} q(s|s)\pi(s) + \boxed{q(s|r)\pi(r)} \\ q(r|s)\pi(s) + q(r|r)\pi(r) \end{bmatrix}$$

P(time 0: r and time 1: s)

Stationarity Example

Example: For the sun/rain example, show that $\pi(x) = [0.875, 0.125]$ satisfies the conditions for a stationary distribution.



Goal: $\pi(x') = \sum_x q(x'|x)\pi(x)$

Solution: We are checking that

$$\begin{bmatrix} \pi(s) = .875 \\ \pi(r) = .125 \end{bmatrix} \overset{\heartsuit}{=} \begin{bmatrix} q(s|s)\pi(s) + q(s|r)\pi(r) \\ q(r|s)\pi(s) + q(r|r)\pi(r) \end{bmatrix}$$

$\frac{9}{10} \cdot \frac{7}{8} + \frac{7}{10} \cdot \frac{1}{8}$

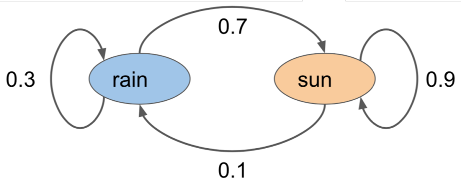
$= \frac{1}{80}(9 \cdot 7 + 7)$

... and it does! The RHS:

$$\begin{bmatrix} q(s|s)\pi(s) + q(s|r)\pi(r) \\ q(r|s)\pi(s) + q(r|r)\pi(r) \end{bmatrix} = \begin{bmatrix} .9(.875) + .7(.125) \\ .1(.875) + .3(.125) \end{bmatrix} = \begin{bmatrix} 7/8 \\ 1/8 \end{bmatrix}$$

Interpreting a Markov Chain

One way we interpret a Markov chain with uncertainty - where the initial states aren't all just 0 or 1 - is by interpreting the graph as a *flow.*



Each state holds some amount of probability x that updates *fluidly* at each time step.

1. $\pi(x')$ is the flow *into* state x'.
2. $q(x', x)\pi(x)$ is the flow into state x' *from* state x.
   "to" "from"
3. $\sum_x q(x', x)\pi(x)$ is the updated amount of liquid/probability at the next time step!

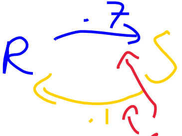
Balanced Flow

Just like in a physics, we are *conserving* the total mass/probability at each time step.

1. $\pi(x')$ is the flow *into* state x'.
2. $q(x', x)\pi(x)$ is the flow into state x' *from* state x.
3. $\sum_x q(x', x)\pi(x)$ is the updated amount of liquid/probability at the next time step!

At stationarity, everything is balanced:
$\underbrace{\text{Flow into } x' \text{ from } x}_{\text{into}} = \underbrace{\text{Flow into } x \text{ from } x'}_{\text{out of } x'}$ for *each and every* pair x, x'.

Detailed Balance



equal mass transfer

Definition When the flow into x' from $x =$ Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.

Proposition: Detailed balance implies stationarity.

Proof:

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\underbrace{\pi(x)q(x'|x)}_{P(x \to x' / x)} = \underbrace{\pi(x')q(x, x')}_{P(x' \to x \,/\, x')}$ for all pairs x, x'.

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x') \sum_x q(x|x')$.

out of the sum!

$\pi(x') =$ "one evolution"

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x') \sum_x q(x|x')$.
- Since $q(x|x')$ is the list of all final states x' given initial states x, it sums to 1.

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x')\sum_x q(x|x')$.
- Since $q(x|x')$ is the list of all final states x' given initial states x, it sums to 1.
- So we have: $\pi(x') = \sum_x \pi(x)q(x'|x)$, the definition of stationarity!

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x') \sum_x q(x|x')$.

Detailed Balance

Definition When the flow into x' from x = Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x') \sum_x q(x|x')$.
- Since $q(x|x')$ is the list of all final states x' given initial states x, it sums to 1.

Detailed Balance

Definition When the flow into x' from $x =$ Flow into x from x' for *each and every* pair x, x', we say that $q(x', x)$ is in *detailed balance* with probability $\pi(x)$.
Proposition: Detailed balance implies stationarity.
Proof:

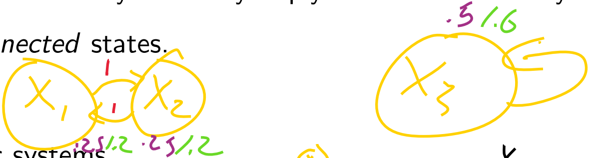
- Under detailed balance, we have $\pi(x)q(x'|x) = \pi(x')q(x, x')$ for all pairs x, x'.
- Then, $\sum_x \pi(x)q(x'|x) = \sum_x \pi(x')q(x|x') = \pi(x') \sum_x q(x|x')$.
- Since $q(x|x')$ is the list of all final states x' given initial states x, it sums to 1.
- So we have: $\pi(x') = \sum_x \pi(x)q(x'|x)$, the definition of stationarity!

We'll work a pen-and-paper detailed balance problem as the opening example, next time...

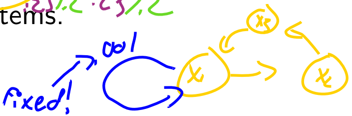
Stationarity Roundup

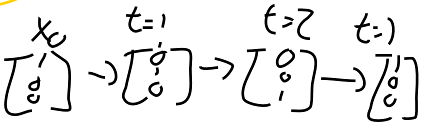
Definition We typically *find* the stationary distribution via setting up the detailed balance equations and solving the linear system they imply. This doesn't always work!

- One issue is *disconnected* states.



- Another is *periodic* systems.





Definition: The transition probability distribution q is called *ergodic* if every state is *reachable* from every other state, and there are no strictly periodic cycles.

Proposition: If a Markov chain is *ergodic*, then there exists a **unique** stationary distribution for any given set of transition probabilities.

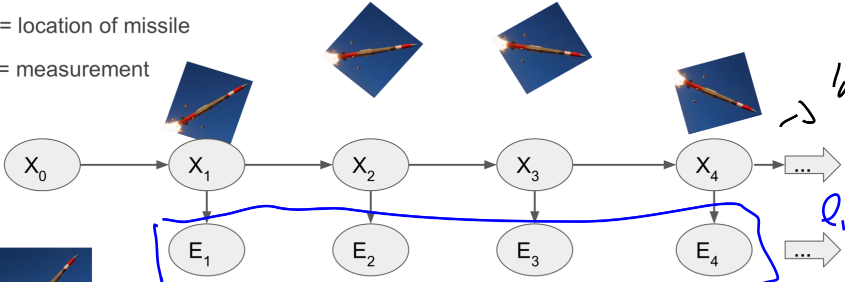
Hidden Markov Models

The plan for where to go next with Markov Models is the same as for Bayesian networks: imagine we only observe *some* of the states along the network.

Example: Missile tracking

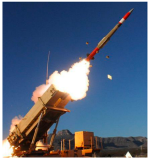
X_t = location of missile

E_t = measurement



location / velocity / accel.

Right / left

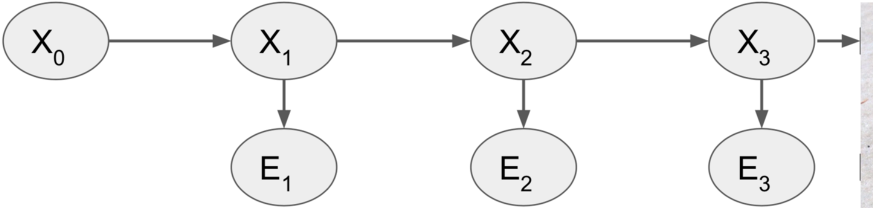


So we **care about** X_t but **we measure** E_t

What to do, what to do...?

Hidden Markov Models

Example:
Suppose you are a graduate student in a basement office. You are writing your dissertation, so you don't get to leave very often.
You are curious if it is raining, and the only contact you have with the outside world is through your advisor. If it is raining, she brings her umbrella 90% of the time, and has it just in case on 20% of sunny days. You know that historically, 40% of rainy days were followed by another rainy day, and 30% of sunny days were followed by a rainy day.

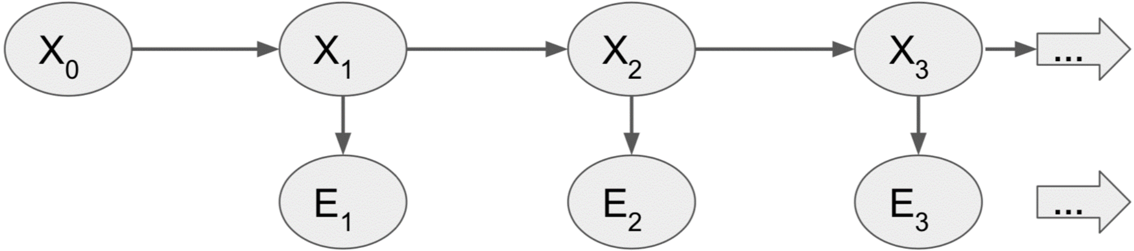


Hidden Markov Models: Sensoring

Denote $X_{0:t} = [X_0, X_1, X_2, \ldots]$ as the states at each time step.

Definition: The *Sensor Markov Assumption* is the assumption that the measurement E_t is conditionally independent of all previous measurements and states, *given* X_t.

$$P(E_t | X_{0:t}, E_{0:t-1}) = P(E_t | X_t)$$



Hidden Markov Models: Roadmap

An assumption like $P(E_t|X_{0:t}, E_{0:t-1}) = P(E_t|X_t)$ is powerful, because we'll get to use large probability products and conditional probability tables just like we did with Bayesian Networks. In general, we have a handful of tasks to do on networks like this:

1. **Filtering:** Describing the *process*.
2. **Prediction:** Describing the future: X_{t+1} *given* the past.
3. **Smoothing:** Describing the past: (or the chain X).
4. **Most likely explanation:** Describing the past: (or the chain X).
5. **Learning:** Bayesian updates and improvements on *priors* and *posteriors*.

Hidden Markov Models: Roadmap

Any of these processes are often decomposed into the two primary tasks of statistics and data science:

Estimation:

1. Come up with a *model* for prediction and explanation
2. Compare the model to data
3. e.g. $P(Alarm|MaryCalls) = ?$.

Inference:

1. *Validate* your model. What does the data tell us about the model?
2. E.g. hill-climbing or annealing for SLR parameters (...and getting an R^2!)

We will often have do both!

Moving Forward

- Coming up:
  1. Hidden Markov: filtering, prediction, and smoothing!
  2. Markov NB on Friday.