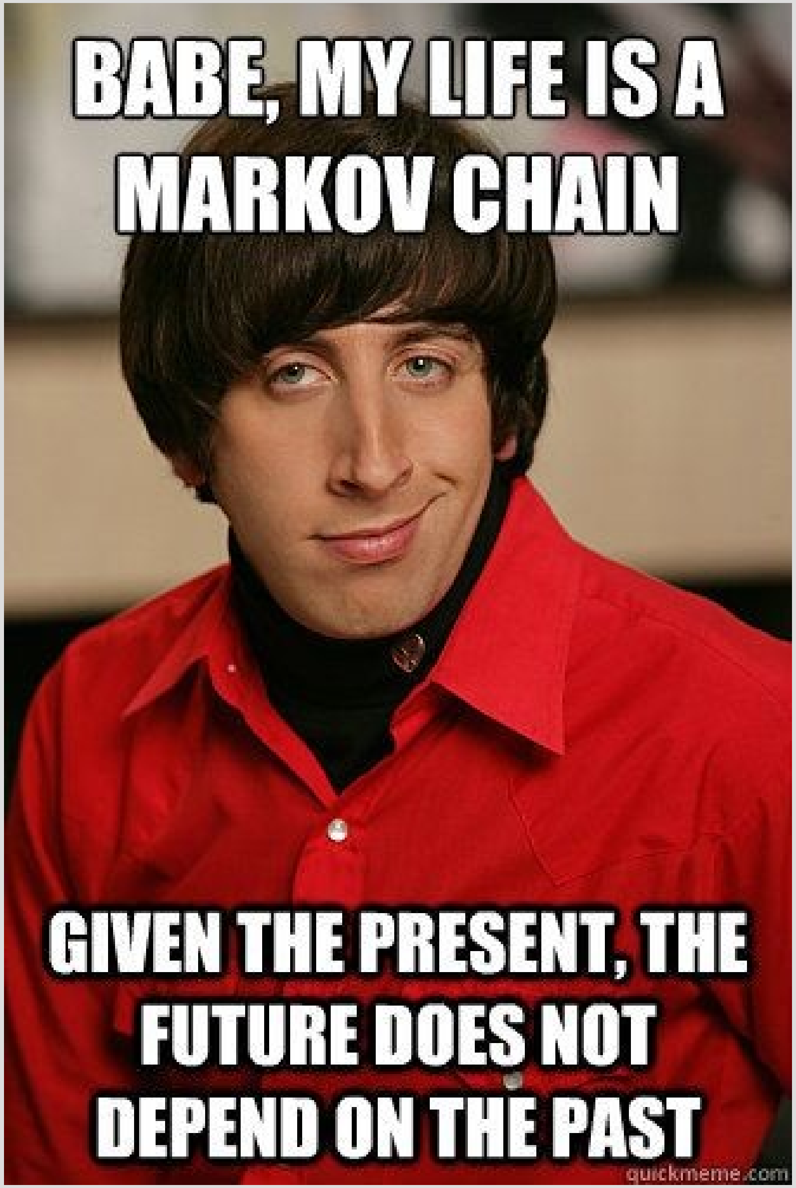
Hidden Markov Models (Ch. 15)



Announcements

Homework 2 posted

Programing:

-Python (preferred)

-Java

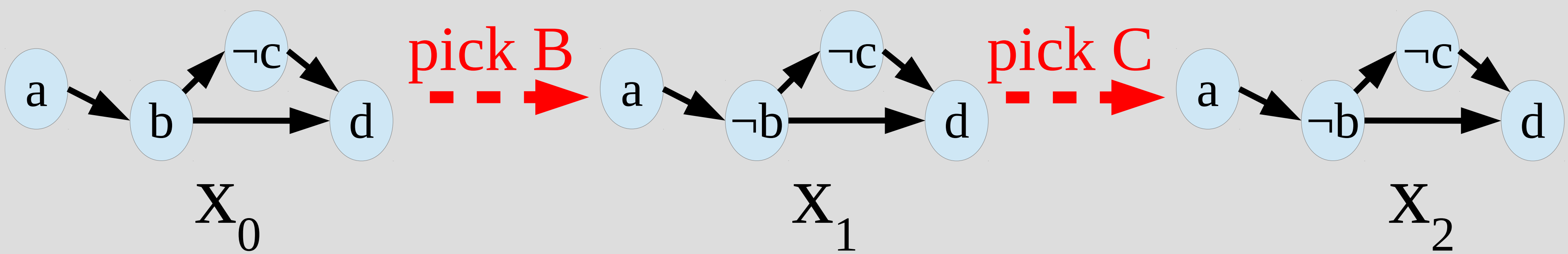
-Matlab

Markov... chains?

Recap, Markov property:

$$P(x_{n+1}|x_n, x_{n-1}, x_{n-2} \cdots x_0) = P(x_{n+1}|x_n)$$

(Next state only depends on current state)



For Gibbs sampling, we made a lot of samples using the Markov property (since this is 1-dimension, it looks like a “chain”)

Markov... chains?

For the next bit, we will still have a “Markov” and uncertainty (i.e. probabilities)

However, we will add partial-observability (some things we cannot see)

These are often called Hidden Markov Models (not “chains” because they won’t be 1D... w/e)

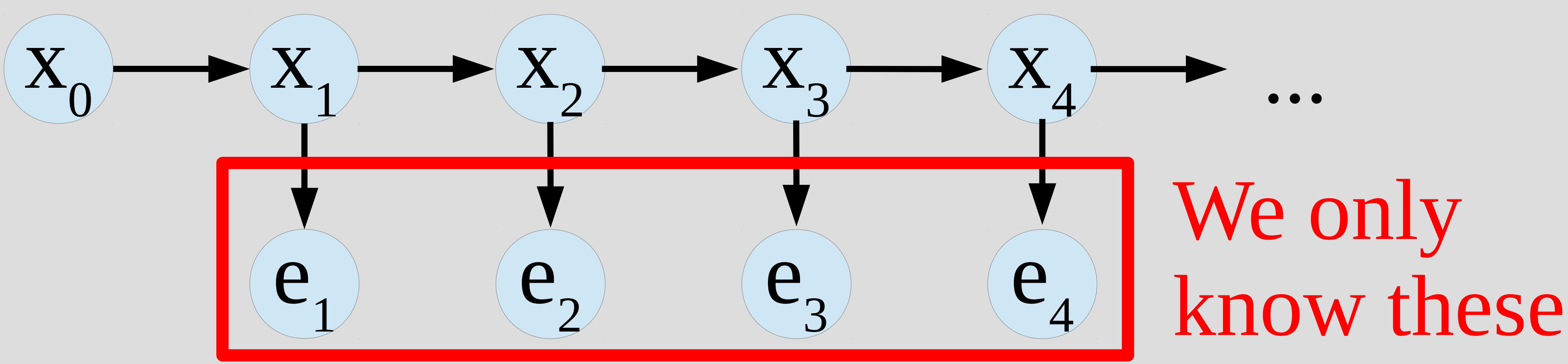
Hidden Markov Models

For Hidden Markov Models (HMMs) often:

(1) E = the evidence

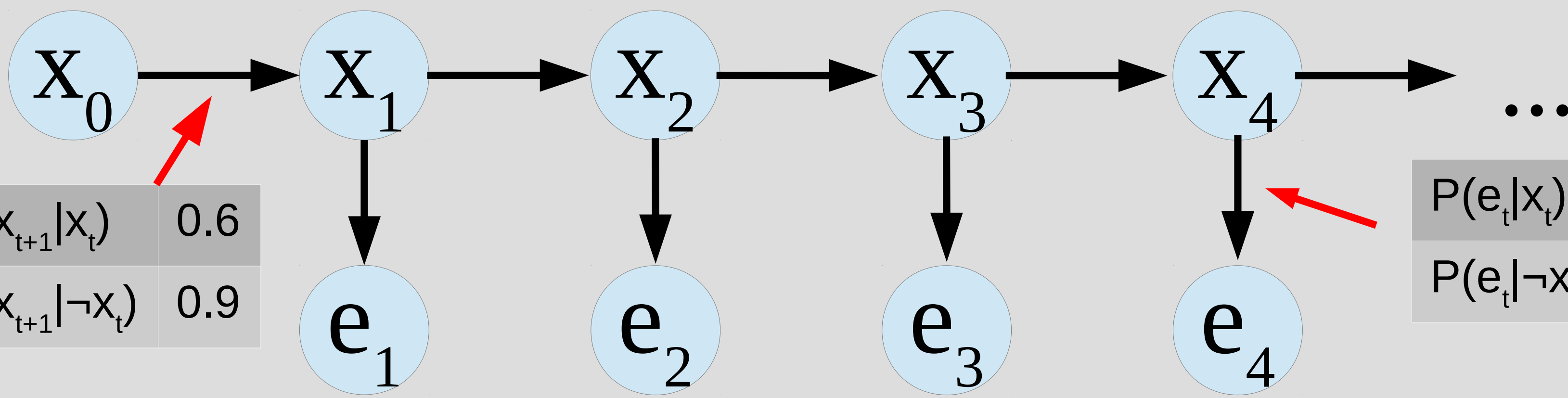
(2) X = the hidden/not observable part

We assume the hidden information is what causes the evidence (otherwise quite easy)



Hidden Markov Models

If you squint a bit, this is actually a Bayesian network as well (though can go on for a while)



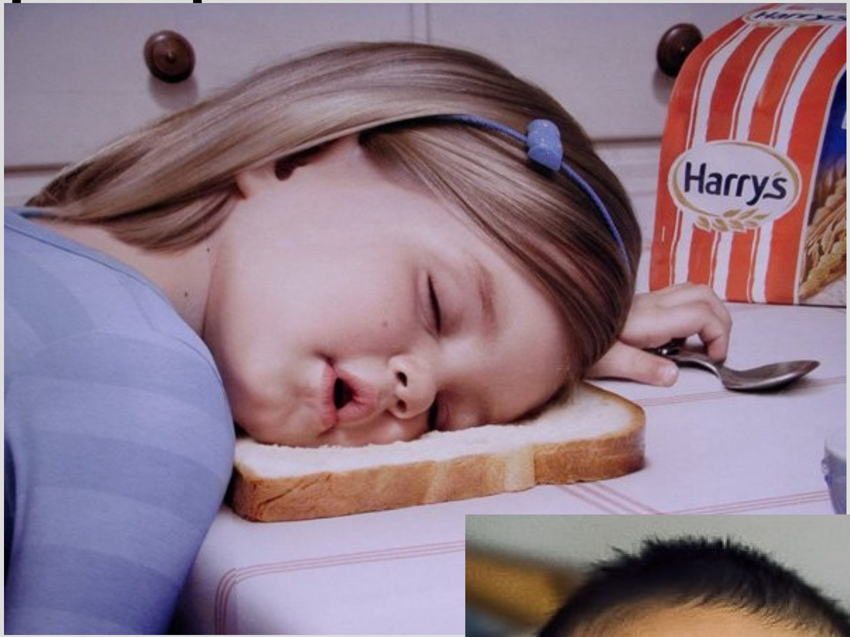
$P(x_{t+1} \mid x_t)$	0.6
$P(x_{t+1} \mid \neg x_t)$	0.9

$P(e_t \mid x_t)$	0.3
$P(e_t \mid \neg x_t)$	0.8

For simplicity's sake, we will assume the probabilities of going to the right (next state) and down (seeing evidence) are the same for all subscripts (typically "time")

Hidden Markov Models

Our example will be: sleep deprivation



So variable X_t will be if a person got enough sleep on day t

This person is not you, but you see them every day, and you can tell if their eyes are bloodshot (this is E_t)

Hidden Markov Models

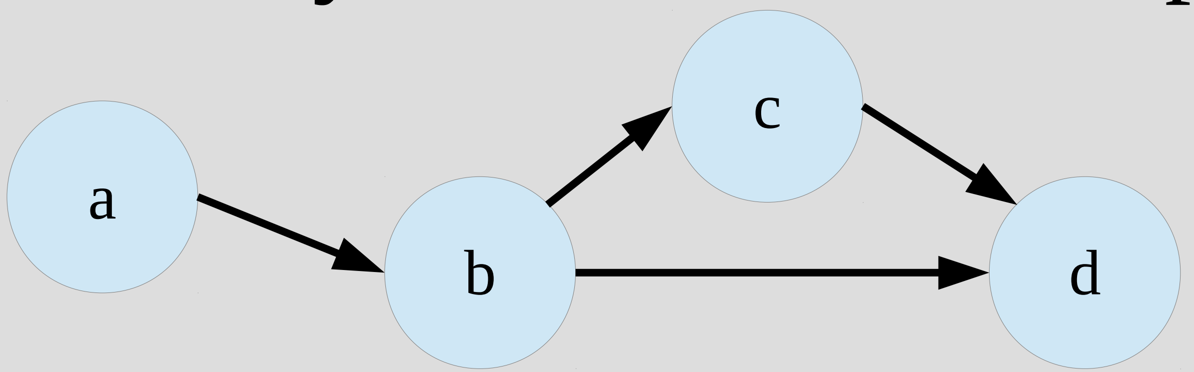
As we will be dealing with quite a few variables, we will introduce some notation:

$E_{1:t} = E_1, E_2, E_3, \ldots. E_t$ (similarly for $X_{0:t}$)
So $P(E_{1:t}) = P(E_1, E_2, E_3, .. E_t)$, which is normal definition of commas like P(a,b)

We will assume we only know $E_{1:t}$ (and X_0) and want to figure out X_k for various k

Hidden Markov Models

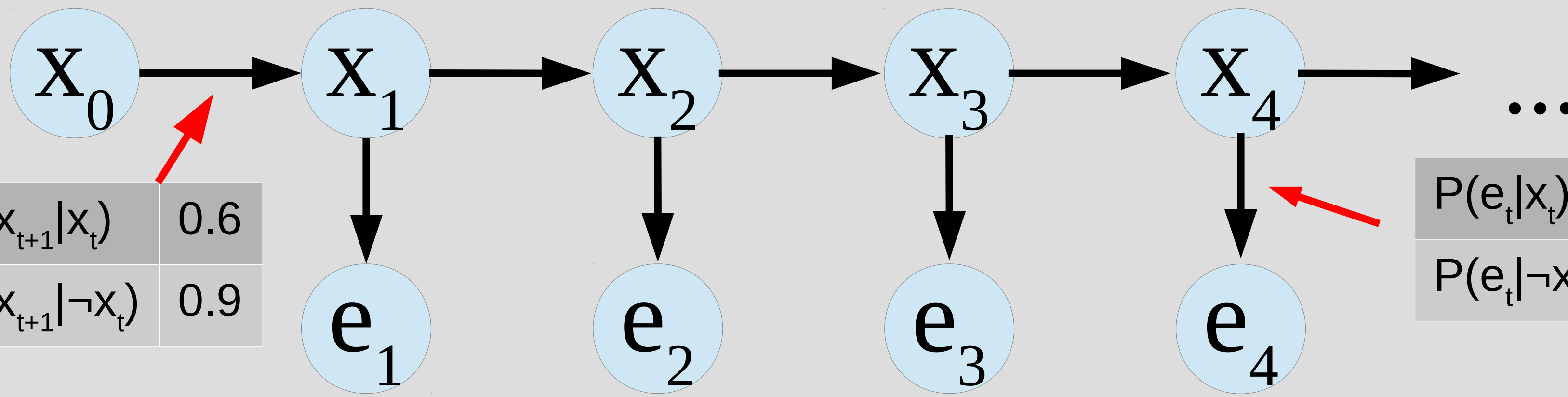
Quick Bayesian network recap:



$$\begin{aligned} P(a,\neg b,c,\neg d) &= P(\neg d|a,\neg b,c)P(c|\neg b,a)P(\neg b|a)P(a) \\ &= P(\neg d|\neg b,c)P(c|\neg b)P(\neg b|a)P(a) \\ &= \boxed{\prod_{x \in Network} P(x|Parents(x))} \end{aligned}$$

Used fact tons in our sampling...

Hidden Markov Models



$P(x_{t+1}\|x_t)$	0.6
$P(x_{t+1}\|\neg x_t)$	0.9

$P(e_t\|x_t)$	0.3
$P(e_t\|\neg x_t)$	0.8

So in this Bayesian network (bigger):

$$P(X_{0:t}, E_{1:t}) = P(X_0) \prod_{i=1}^{t} P(X_i|X_{i-1})P(E_i|X_i)$$

Most Likely Explanation

Typically, use above to compute four things:

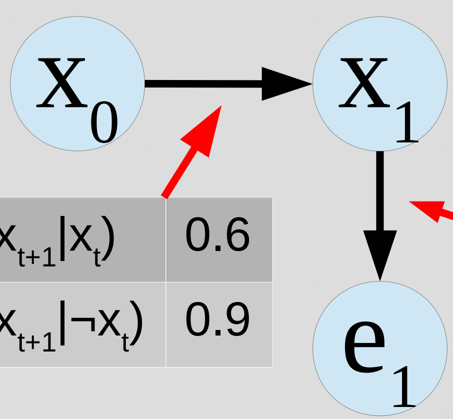
Filtering	Prediction	Smoothing	MLE
$P(x_t\|e_{1:t})$	$P(x_{t+k}\|e_{1:t})$	$P(x_k\|e_{1:t}), k < t$	$P(x_{1:t}\|e_{1:t})$

Filtering in HMMs

All four of these are actually quite similar, and you can probably already find them

The only difficulty is the size of the Bayesian network, so let's start small to get intuition:

$P(x_0)$	0.5



$P(x_{t+1}\|x_t)$	0.6
$P(x_{t+1}\|\neg x_t)$	0.9

How can you find $P(x_1|\neg e_1)$?

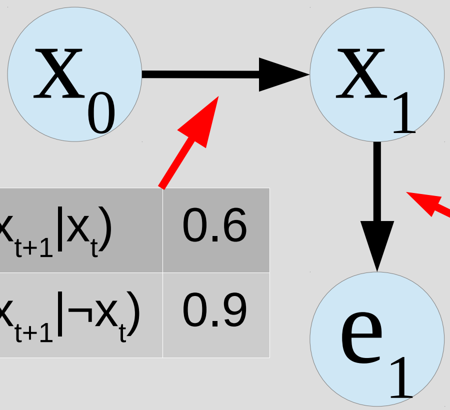
(this is a simple Bays-net)

$P(e_t\|x_t)$	0.3
$P(e_t\|\neg x_t)$	0.8

$$P(X_{0:t}, E_{1:t}) = P(X_0) \prod_{i=1}^{t} P(X_i|X_{i-1})P(E_i|X_i)$$

Filtering in HMMs

P(x_0)	0.5



P(x_{t+1}|x_t)	0.6
P(x_{t+1}|¬x_t)	0.9

P(e_t|x_t)	0.3
P(e_t|¬x_t)	0.8

$$P(x_1|\neg e_1) = \alpha P(x_1, \neg e_1)$$
$$= \alpha \sum_{x_0} P(x_0, x_1, \neg e_1)$$
$$= \alpha \sum_{x_0} P(x_0) P(x_1|x_0) P(\neg e_1|x_1)$$
$$= \alpha P(\neg e_1|x_1) \sum_{x_0} P(x_0) P(x_1|x_0)$$
$$= \alpha (1 - 0.3) \cdot (0.5 \cdot 0.6 + 0.5 \cdot 0.9)$$
$$\approx 0.525\alpha$$

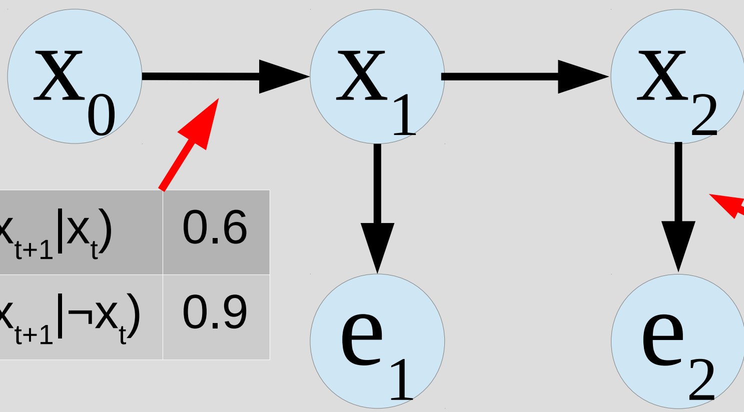
Similarly, P(¬x_1|¬e_1) ≈0.05α

So normalized gives: P(x_1|¬e_1) ≈ 0.913

91% chance I slept last night, given today I didn't have bloodshot eyes

Filtering in HMMs

$P(x_0)$	0.5



$P(x_{t+1}\|x_t)$	0.6
$P(x_{t+1}\|\neg x_t)$	0.9

$P(e_t\|x_t)$	0.3
$P(e_t\|\neg x_t)$	0.8

Find: $P(x_2|\neg e_1, \neg e_2)$

Filtering in HMM

Double sum?!?!
Double... for loop..?

$$
\begin{aligned}
P(x_2 \mid \neg e_1, \neg e_2) &= \alpha P(x_2, \neg e_1, \neg e_2) \\
&= \alpha \sum_{x_0} \sum_{x_1} P(x_0, x_1, x_2 \neg e_1, \neg e_2) \\
&= \alpha \sum_{x_0} \sum_{x_1} P(x_0) P(x_1 \mid x_0) P(\neg e_1 \mid x_1) P(x_2 \mid x_1) P(\neg e_2 \mid x_2) \\
&= \alpha \sum_{x_1} \sum_{x_0} P(\neg e_2 \mid x_2) P(x_2 \mid x_1) P(\neg e_1 \mid x_1) P(x_0) P(x_1 \mid x_0)
\end{aligned}
$$

Just computed this!
It is $P(x_1|e_1)$

$$
\begin{aligned}
&= \alpha P(\neg e_2 \mid x_2) \sum_{x_1} P(x_2 \mid x_1) \boxed{P(\neg e_1 \mid x_1) \sum_{x_0} P(x_0) P(x_1 \mid x_0)} \\
&= \alpha P(\neg e_2 \mid x_2) \big(P(x_2 \mid x_1) \cdot 0.913 + P(x_2 \mid \neg x_1) \cdot (1 - 0.913) \big) \\
&= \alpha (1 - 0.3)(0.6 \cdot 0.913 + 0.9 \cdot (1 - 0.913)) \\
&\approx 0.438 \alpha
\end{aligned}
$$

... after normalizing you should get: ≈0.854

Filtering in HMMs

Double sum?!?!
Double... for loop..?

$$P(x_2|\neg e_1, \neg e_2) = \alpha P(x_2, \neg e_1, \neg e_2)$$
$$= \alpha \sum_{x_0} \sum_{x_1} P(x_0, x_1, x_2 \neg e_1, \neg e_2)$$
$$= \alpha \sum_{x_0} \sum_{x_1} P(x_0) P(x_1|x_0) P(\neg e_1|x_1) P(x_2|x_1) P(\neg e_2|x_2)$$
$$= \alpha \sum_{x_1} \sum_{x_0} P(\neg e_2|x_2) P(x_2|x_1) P(\neg e_1|x_1) P(x_0) P(x_1|x_0)$$

Just computed this!
It is $P(x_1|e_1)$

$$= \alpha P(\neg e_2|x_2) \sum_{x_1} P(x_2|x_1) \boxed{P(\neg e_1|x_1) \sum_{x_0} P(x_0) P(x_1|x_0)}$$

change to $\neg x_2$

$P(\neg x_2|\neg e_1, \neg e_2)$
$\approx 0.075\alpha$

$$= \alpha P(\neg e_2|\boxed{x_2}) \left(P(\boxed{x_2}|x_1) \cdot 0.913 + P(\boxed{x_2}|\neg x_1) \cdot (1 - 0.913)\right)$$
$$= \alpha (1 - 0.3)(0.6 \cdot 0.913 + 0.9 \cdot (1 - 0.913))$$
$$\approx 0.438\alpha$$

... after normalizing you should get: ≈0.854

Filtering in HMMs

In general:

$$
\begin{aligned}
P(x_t|e_{1:t}) &= P(x_t|e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t|e_{1:t-1}) \qquad\qquad \text{(Note: different } \alpha\text{)} \\
&= \alpha P(e_t|x_t, e_{1:t-1}) P(x_t|e_{1:t-1}) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t, x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}, e_{1:t-1}) P(x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}) P(x_{t-1}|e_{1:t-1}) \Big)
\end{aligned}
$$

Filtering in HMMs

In general:

$$
\begin{aligned}
\boxed{P(x_t|e_{1:t})} &= P(x_t|e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t|e_{1:t-1}) \qquad \text{(Note: different } \alpha\text{)} \\
&= \alpha P(e_t|x_t, e_{1:t-1}) P(x_t|e_{1:t-1}) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t, x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}, e_{1:t-1}) P(x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}) \boxed{P(x_{t-1}|e_{1:t-1})} \Big)
\end{aligned}
$$

... same, but different ‘t’

Filtering in HMMs

In general:

$$
\begin{aligned}
\boxed{f(t)} &= P(x_t|e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t, e_{1:t-1}) \\
&= \alpha P(x_t, e_t|e_{1:t-1}) \qquad \text{(Note: different } \alpha\text{)} \\
&= \alpha P(e_t|x_t, e_{1:t-1}) P(x_t|e_{1:t-1}) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t, x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t, e_{1:t-1}) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}, e_{1:t-1}) P(x_{t-1}|e_{1:t-1}) \Big) \\
&= \alpha P(e_t|x_t) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}) \boxed{f(t-1)} \Big)
\end{aligned}
$$

Actually, this is just a recursive function

Filtering in HMMs

So we can compute $f(t) = P(x_t|e_{1:t})$:

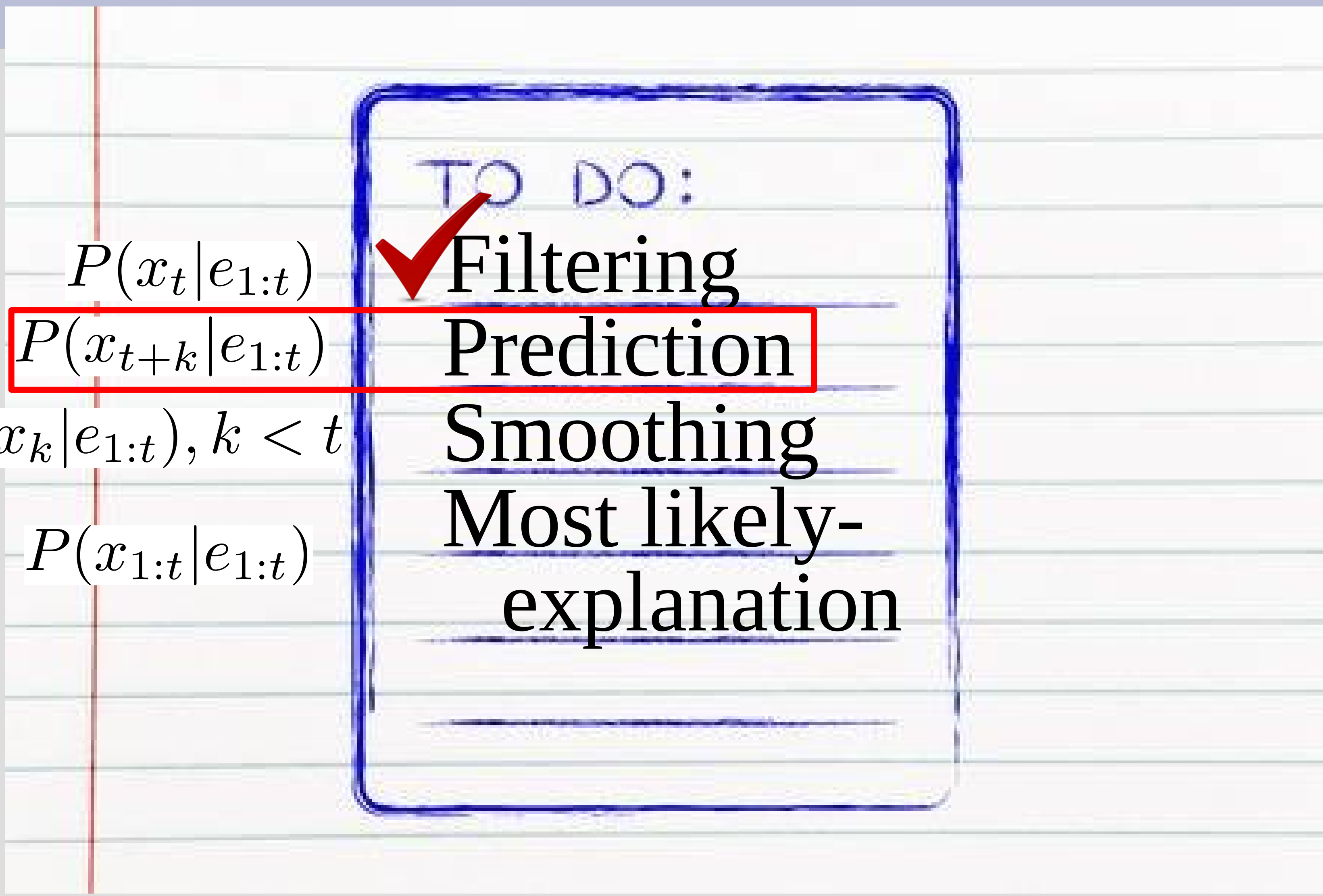
$$f(t) = \alpha P(e_t|x_t) \sum_{x_{t-1}} \Big(P(x_t|x_{t-1}) f(t-1) \Big)$$

actually an array, as you need both T/F for sum(or 1-)

Of course, we don't *actually* want to do this recursively... rather with dynamic programing

Start with $f(0) = P(x_0)$, then use this to find f(1)... and so on (can either store in array, or just have a single variable... like Fibonacci)

Prediction in HMMs



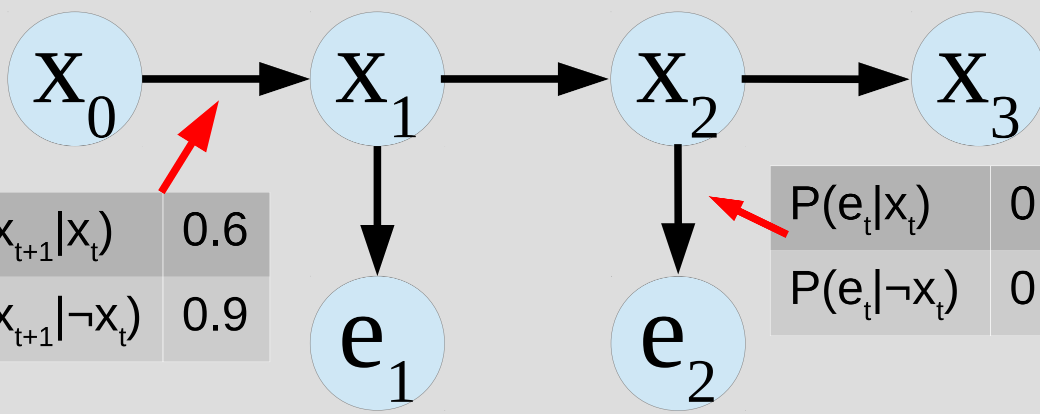
- $P(x_t|e_{1:t})$ — Filtering
- $P(x_{t+k}|e_{1:t})$ — Prediction
- $P(x_k|e_{1:t}), k < t$ — Smoothing
- $P(x_{1:t}|e_{1:t})$ — Most likely-explanation

How would you find “prediction”?

Prediction in HMMs

Probably best to go back to the example:
What is chance I sleep on day 3 given, you saw me without bloodshot eyes on day 1&2?

$P(x_0)$	0.5



$P(x_{t+1}\|x_t)$	0.6
$P(x_{t+1}\|\neg x_t)$	0.9

$P(e_t\|x_t)$	0.3
$P(e_t\|\neg x_t)$	0.8

$P(x_3 \mid e_1, e_2)$=???

Prediction in HMMs

$$
\begin{aligned}
P(x_3 | \neg e_1, \neg e_2) &= \alpha P(x_3, \neg e_1, \neg e_2) \\
&= \alpha \sum_{x_0} \sum_{x_1} \sum_{x_2} P(x_0, x_1, x_2, x_3, \neg e_1, \neg e_2) \\
&= \alpha \sum_{x_0} \sum_{x_1} \sum_{x_2} P(x_0) P(x_1 | x_0) P(\neg e_1 | x_1) P(x_2 | x_1) P(e_2 | x_2) P(x_3 | x_2) \\
&= \alpha \sum_{x_2} P(x_3 | x_2) \boxed{P(e_2 | x_2) \sum_{x_1} P(x_2 | x_1) P(\neg e_1 | x_1) \sum_{x_0} P(x_0) P(x_1 | x_0)} \\
&= \alpha \sum_{x_2} P(x_3 | x_2) \boxed{P(x_2 | \neg e_1, \neg e_2)} \\
&= \alpha (0.6 \cdot 0.854 + 0.9 \cdot (1 - 0.854)) \\
&\approx 0.644 \alpha
\end{aligned}
$$

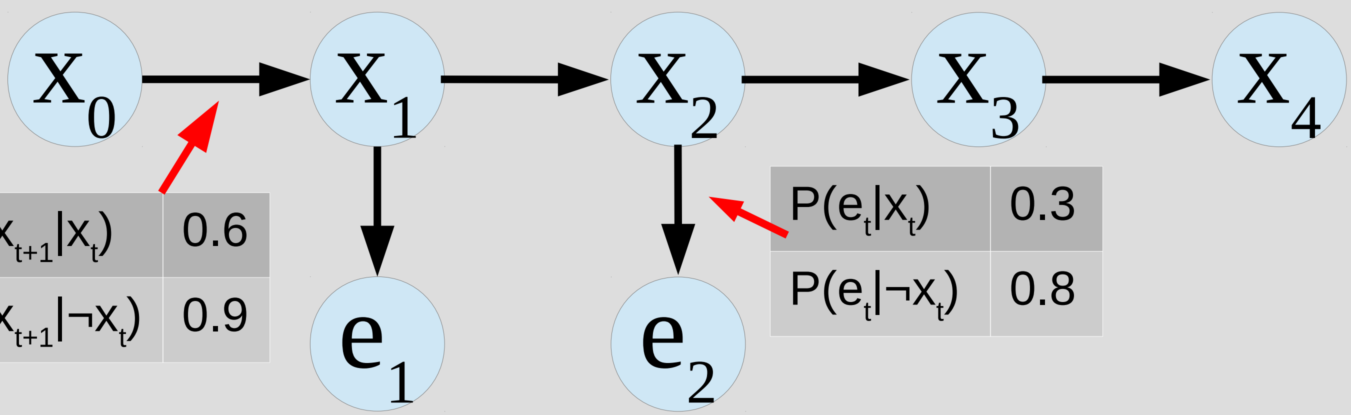
whew...

Turns out that $P(\neg x_3 | \neg e_1, \neg e_2) \approx 0.356\alpha$, so $\alpha = 1$

Prediction in HMMs

Day 4?

$P(x_0)$	0.5



$P(x_{t+1} \mid x_t)$	0.6
$P(x_{t+1} \mid \neg x_t)$	0.9

$P(e_t \mid x_t)$	0.3
$P(e_t \mid \neg x_t)$	0.8

$$P(x_4 \mid e_1, e_2) = ???$$

Prediction in HMMs

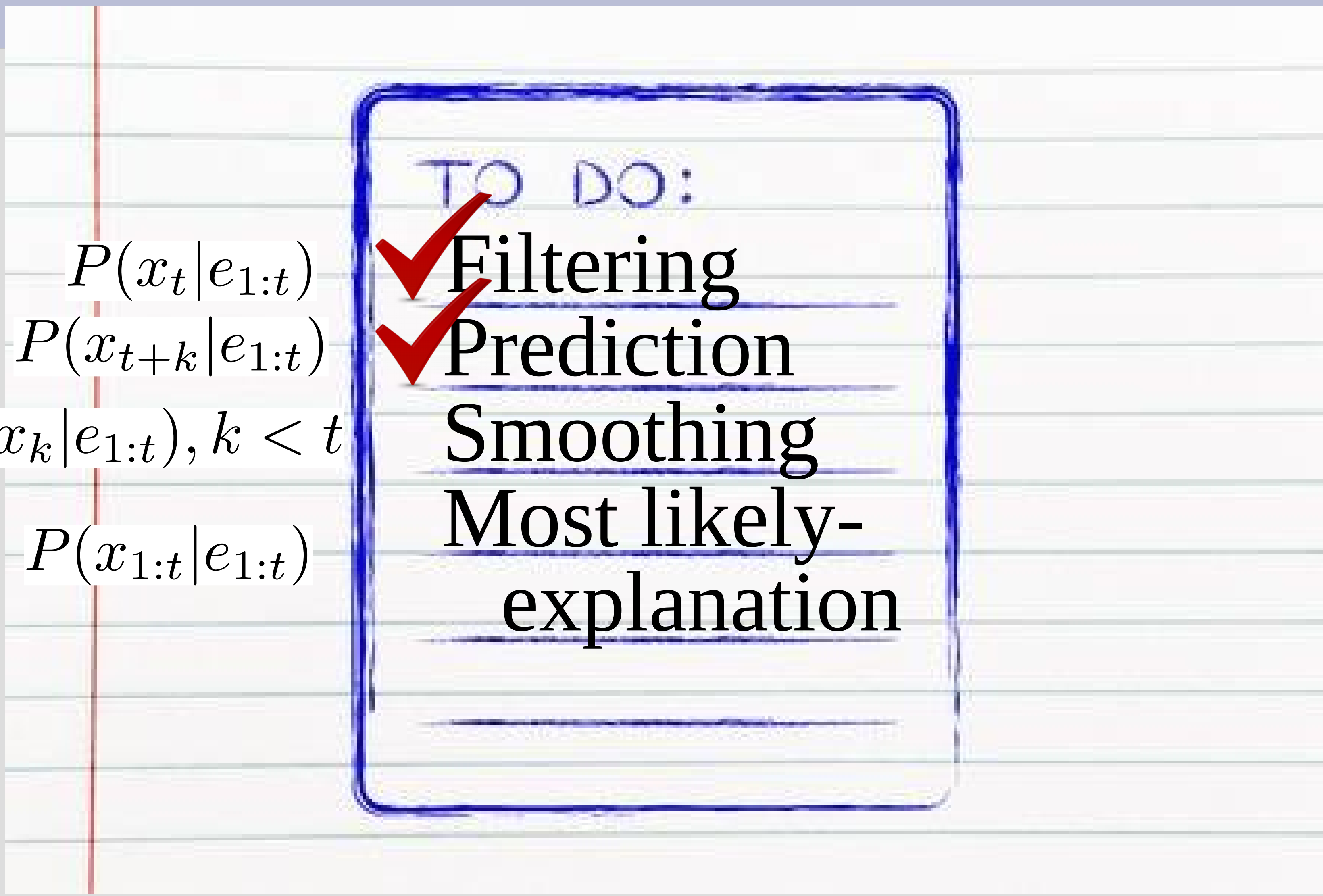
$$
\begin{aligned}
P(x_4|\neg e_1,\neg e_2) &= \alpha P(x_4,\neg e_1,\neg e_2)\\
&= \alpha \sum_{x_0}\sum_{x_1}\sum_{x_2}\sum_{x_3} P(x_0,x_1,x_2,x_3,x_4,\neg e_1,\neg e_2)\\
&= \alpha \sum_{x_0}\sum_{x_1}\sum_{x_2}\sum_{x_3} P(x_0)P(x_1|x_0)P(\neg e_1|x_1)P(x_2|x_1)P(e_2|x_2)P(x_3|x_2)P(x_4|x_3)\\
&= \alpha \sum_{x_3} P(x_4|x_3)\boxed{\sum_{x_2} P(x_3|x_2)P(e_2|x_2)\sum_{x_1} P(x_2|x_1)P(\neg e_1|x_1)\sum_{x_0} P(x_0)P(x_1|x_0)}\\
&= \alpha \sum_{x_3} P(x_4|x_3)\boxed{P(x_3|e_1,e_2)}\\
&= \alpha(0.6\cdot 0.644 + 0.9\cdot(1-0.644))\\
&\approx 0.707\alpha
\end{aligned}
$$

...think I see a pattern here

Turns out that P(¬x_4|¬e_1,¬e_2) ≈ 0.293α, so α=1

(α always 1 now, as can move into red box)

Prediction in HMMs



- $P(x_t|e_{1:t})$ ✓ Filtering
- $P(x_{t+k}|e_{1:t})$ ✓ Prediction
- $P(x_k|e_{1:t}), k < t$ Smoothing
- $P(x_{1:t}|e_{1:t})$ Most likely-explanation

We'll save the other two for next time...