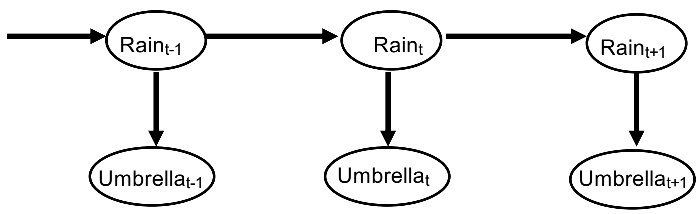
Example: Is it raining, given umbrellas?

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

Weather has a 30% chance of changing and a 70% chance of staying the same.



R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

If it's raining, the probability of someone carrying an umbrella is .9; if it's raining, the probability of NOT carrying an umbrella is .2

Fully worked out HMM for rain: http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

20

Filtering

- For each day *t*, $\mathbf{E}_t$ contains variable U_t (whether the umbrella appears) and $\mathbf{X}_t$ contains state variable R_t (whether it's raining)
- Compute the current belief state, given all evidence to date
- Maintain a **current** state estimate and update it
  - Instead of looking at all observed values in history
  - Also called **state estimation**
- Given result of filtering up to time *t*, agent must compute result at *t*+1 from new evidence $\mathbf{e}_{t+1}$:

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = f(\mathbf{e}_{t+1},\ P(\mathbf{X}_t \mid \mathbf{e}_{1:t}))$$

… for some function *f*.

21

Filtering

- A good algorithm for filtering will maintain a current state estimate and update it at each point.
- $P(X_{t+1}|e_{1:t+1}) = f(P(X_t|e_{1:t}), e_{t+1})$
- Where X is the random variable and e is evidence
- Saves recomputation.
- It turns out that this is easy enough to come up with.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

22

Filtering

- We rearrange the formula for:
  - $P(X_{t+1}|e_{1:t+1})$
- First, we divide up the evidence:
  - $P(X_{t+1}|e_{1:t+1}) = P(X_{t+1}|e_{1:t}, e_{t+1})$
- Then we apply Bayes rule, remembering the use of the normalization factor α:
  - $P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1}, e_{1:t}) P(X_{t+1}|e_{1:t})$
- And after that we use the Markov assumption on the sensor model:
  - $P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1}) P(X_{t+1}|e_{1:t})$
- The result of this assumption is to make that first term on the right hand side ignore all the evidence — the probability of the observation at $t + 1$ only depends on the value of X_{t+1}.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

23

Filtering

- Let's look at that expression some more:
  - $P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1})P(X_{t+1}|e_{1:t})$
- The first term on the right updates with the new evidence and the second term on the right is a one step prediction from the evidence up to *t* to the state at *t* + 1.
- Next we condition on the current state P(X):
  - $P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1}) \Sigma x_t\, P(X_{t+1}|x_t, e_{1:t})P(x_t|e_{1:t})$
- Finally, we apply the Markov assumption again:
  - $P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1}) \Sigma x_t\, P(X_{t+1}|x_t)P(x_t|e_{1:t})$
- We'll call the bit on the right $f_{1:t}$

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

24

Filtering

- $f_{1:t}$ gives us the required recursive update.
  - The probability distribution over the state variables at *t* + 1 is a function of the transition model, the sensor model, and what we know about the state at time *t*.
- Space and time constant, independent of *t*.
- This allows a limited agent to compute the current distribution for any length of sequence.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

25

Recursive Estimation

- We use ***recursive estimation*** to compute $P(X_{t+1} \mid e_{1:t+1})$ as a function of e_{t+1} and $P(X_t \mid e_{1:t})$

1. Project current state forward (t → t+1)
2. Update state using new evidence $\mathbf{e}_{t+1}$

$\mathrm{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1})$ as function of $\mathbf{e}_{t+1}$ and $\mathrm{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t})$:

$$\mathrm{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \mathrm{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1})$$

26

Recursive Estimation

- $\mathrm{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1})$ as a function of $\mathbf{e}_{t+1}$ and $\mathrm{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t})$:

$$P(X_{t+1} \mid e_{1:t+1}) = P(X_{t+1} \mid e_{1:t}, e_{t+1}) \quad \text{dividing up evidence}$$
$$= \alpha P(e_{t+1} \mid X_{t+1}, e_{1:t})\, P(X_{t+1} \mid e_{1:t}) \quad \text{Bayes rule}$$
$$= \alpha P(e_{t+1} \mid X_{t+1})\, P(X_{t+1} \mid e_{1:t}) \quad \text{sensor Markov assumption}$$

- $\mathrm{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{1:t+1})$ updates with new evidence (from sensor)
- One-step prediction by conditioning on current state X:

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

27

Recursive Estimation

- One-step prediction by conditioning on current state X:

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(e_{t+1} \mid X_{t+1}) \sum_{x_t} \underbrace{P(X_{t+1} \mid x_t)}_{\text{transition model}} \underbrace{P(x_t \mid e_{1:t})}_{\text{current state}}$$

- ...which is what we wanted!
- So, think of $P(\mathbf{X}_t \mid \mathbf{e}_{1:t})$ as a "message" $f_{1:t+1}$
  - Carried forward along the time steps
  - Modified at every transition, updated at every new observation
- This leads to a recursive definition:

$$f_{1:t+1} = \alpha \text{ FORWARD}(f_{1:t}, e_{t+1})$$

28

Filtering: Umbrellas example

- The prior is ⟨0.5, 0.5⟩. (R=*t*, R=*f*)
- We can first predict whether it will rain on day 1 given what we already know:
- $\mathbf{P}(\mathbf{R}_1) = \Sigma_{r_0} \mathbf{P}(R_1 | r_0)\, P(r_0)$
  $= \langle 0.7,0.3\rangle \times 0.5 + \langle 0.3,0.7\rangle \times 0.5$
  $= \langle 0.35,0.15\rangle + \langle 0.15,0.35\rangle$
  $= \langle 0.5,0.5\rangle$
- As we should expect, this just gives us the prior — that is the probability of rain when we don't have any evidence.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

29

Filtering: Umbrellas example

- However, we have observed the umbrella, so that U_1 = *true*, and we can update using the sensor model:
- $\mathbf{P}(\mathbf{R}_1|U_1) = \alpha \mathbf{P}(u_1|R_1)\mathbf{P}(R_1)$
  $= \alpha\langle 0.9,0.2\rangle\langle 0.5,0.5\rangle$
  $= \alpha\langle 0.45,0.1\rangle$
  $\approx \langle 0.818,0.182\rangle$
- So, since umbrella is strong evidence for rain, the probability of rain is much higher once we take the observation into account.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

30

Filtering: Umbrellas example

- We can then carry out the same computation for Day 2, first predicting whether it will rain on day 2 given what we already saw:
- $\mathbf{P}(\mathbf{R}_2|u_1) = \Sigma_{r_1} \mathbf{P}(R_2|r_1)P(r_1|u_1)$
  $= \langle 0.7,0.3\rangle\times 0.818 + \langle 0.3,0.7\rangle\times 0.182$
  $\approx \langle 0.627,0.373\rangle$
- So even without evidence of rain on the second day there is a higher probability of rain than the prior because rain tends to follow rain.
  - (In this model rain tends to persist.)

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

31

Filtering: Umbrellas example

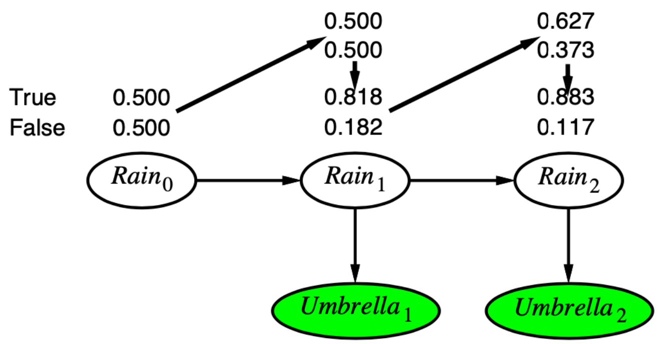
- Then we can repeat the evidence update, u_2 (U_2 = *true*), so:
- $\mathbf{P}(\mathbf{R}_2 | u_1, u_2) = \alpha \mathbf{P}(u_2 | R_2) \mathbf{P}(R_2 | u_1)$
  $= \alpha\langle 0.9, 0.2\rangle\langle 0.627, 0.373\rangle$
  $= \alpha\langle 0.565, 0.075\rangle$
  $\approx \langle 0.883, 0.117\rangle$
- So, the probability of rain increases again, and is higher than on day 1.

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

32

Filtering: Umbrellas example

- Put more succinctly:



- We can think of the calculation as messages passed along the chain

http://www.sci.brooklyn.cuny.edu/~parsons/courses/740-fall-2011/notes/lect07.pdf

33

Umbrellas, summarized

- $P(Rain_1 = t)$
  $= \Sigma_{Rain_0} P(Rain_1 = t \mid Rain_0)\, P(Rain_0)$
  $= 0.70 * 0.50 + 0.30 * 0.50 =$ **0.50**
- $P(Rain_1 = t \mid Umbrella_1 = t)$
  $= \alpha\, P(Umbrella_1 = t \mid Rain_1 = t)\, P(Rain_1 = t)$
  $= \alpha * 0.90 * 0.50 = \alpha * 0.45 \approx$ **0.818**
- $P(Rain_2 = t \mid Umbrella_1 = t)$
  $= \Sigma_{Rain_1} P(Rain_2 = t \mid Rain_1)\, P(Rain_1 \mid Umbrella_1 = t)$
  $= 0.70 * 0.818 + 0.30 * 0.182 \approx$ **0.627**
- $P(Rain_2 = t \mid Umbrella_1 = t, Umbrella_2 = t)$
  $= \alpha\, P(Umbrella_2 = t \mid Rain_2 = t)\, P(Rain_2 = t \mid Umbrella_1 = t)$
  $= \alpha * 0.90 * 0.627 \approx \alpha * 0.564 \approx$ **0.883**

34

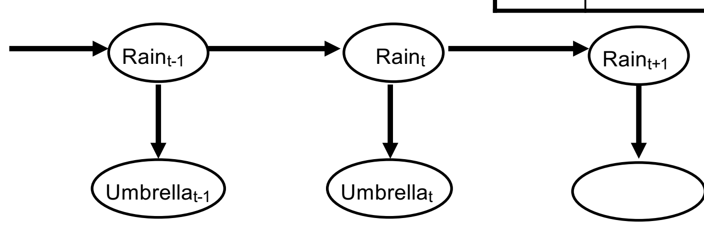
Group Exercise: Filtering

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

$$P(R_2 \mid U_1, U_2) = \alpha\, P(U_2 \mid R_2) \Sigma_{R1} P(R_2 \mid R_1)\, P(R_1 \mid U_1)$$
$$= 0.883$$

R_{t-1}	$P(R_t \mid R_{t-1})$
T	0.7
F	0.3

Weather has a 30% chance of changing and a 70% chance of staying the same.



What is the probability of rain on Day 2, given a uniform prior of rain on Day 0, U_1 = true, and U_2 = true?

R_t	$P(U_t \mid R_t)$
T	0.9
F	0.2

If it's raining, the probability of someone carrying an umbrella is .9; if it's raining, the probability of NOT carrying an umbrella is .2

35