Hidden Markov Model (HMM) Inference

$$B_t(X) = \mathbb{P}(X_t|e_{1:t})$$

- Two steps: 1. Passage of time 2. Observation - Belief is the belief of the next state - B' is the belief without observing any new evidence - Step before evaluating B - Once you observe the new evidence at that time, you evaluate B - Given $B(X_t) = \mathbb{P}(X_t|e_{1:t})$ and $\mathbb{P}(X_{t+1}|x_t)$

$$P(X_{t+1}|e_{1:t}) = \sum_{x_t} \mathbb{P}(X_{t+1}, x_t|e_{1:t}) = \sum_{x_t} \mathbb{P}(X_{t+1}|x_t)\mathbb{P}(x_t|e_{1:t}) = \sum_{x_t} \mathbb{P}(X_{t+1}|x_t)B(x_t)$$

$$B'(X_{t+1}) = \mathbb{P}(X_{t+1}|e_{1:t})$$

$$B(X_{t+1}) \propto_{X_{t+1}} \mathbb{P}(e_{t+1}|X_{t+1})B'(X_{t+1})$$

- Note: have to re-normalize all proportionalities

Example

- Given $\mathbb{P}(X_1)$ and $\mathbb{P}(X_2|X_1)$

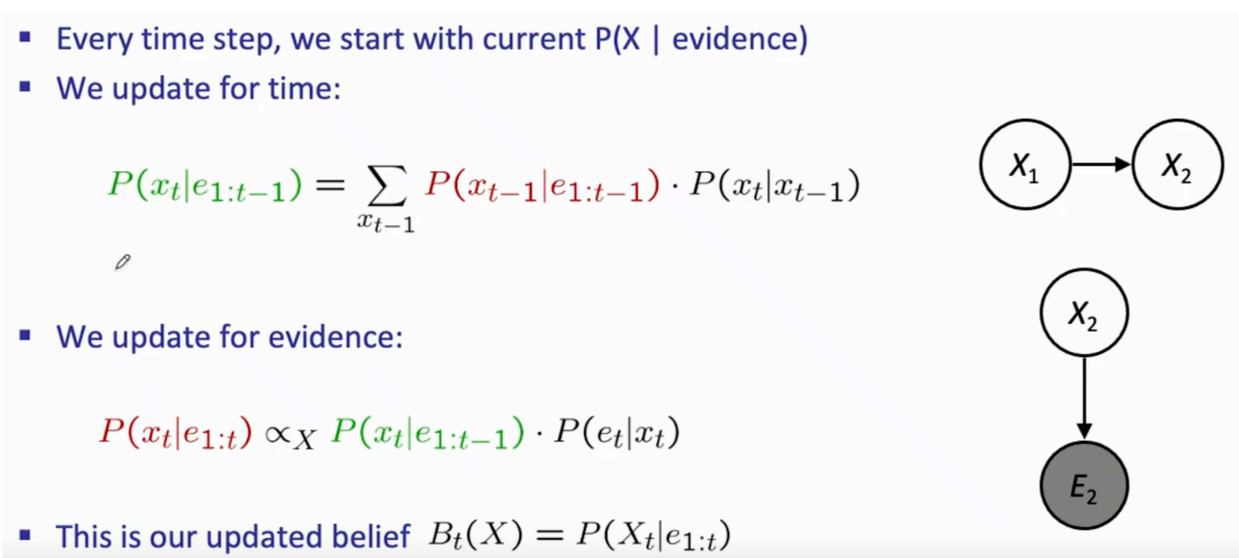
$$\mathbb{P}(x_2) = \sum_{x_1} \mathbb{P}(x_1, x_2) = \sum_{x_1} \mathbb{P}(x_1)\mathbb{P}(x_2|x_1)$$

$$\mathbb{P}(x_1|e_1) = \frac{\mathbb{P}(x_1, e_1)}{\mathbb{P}(e_1)} \propto_{X_1} \mathbb{P}(x_1, e_1) = \mathbb{P}(x_1)\mathbb{P}(e_1|x_1)$$

Explicitly:

$$\mathbb{P}(x_1|e_1) = \frac{\mathbb{P}(x_1)\mathbb{P}(e_1|x_1)}{\sum_{x'} \mathbb{P}(x')\mathbb{P}(e_1|x')}$$

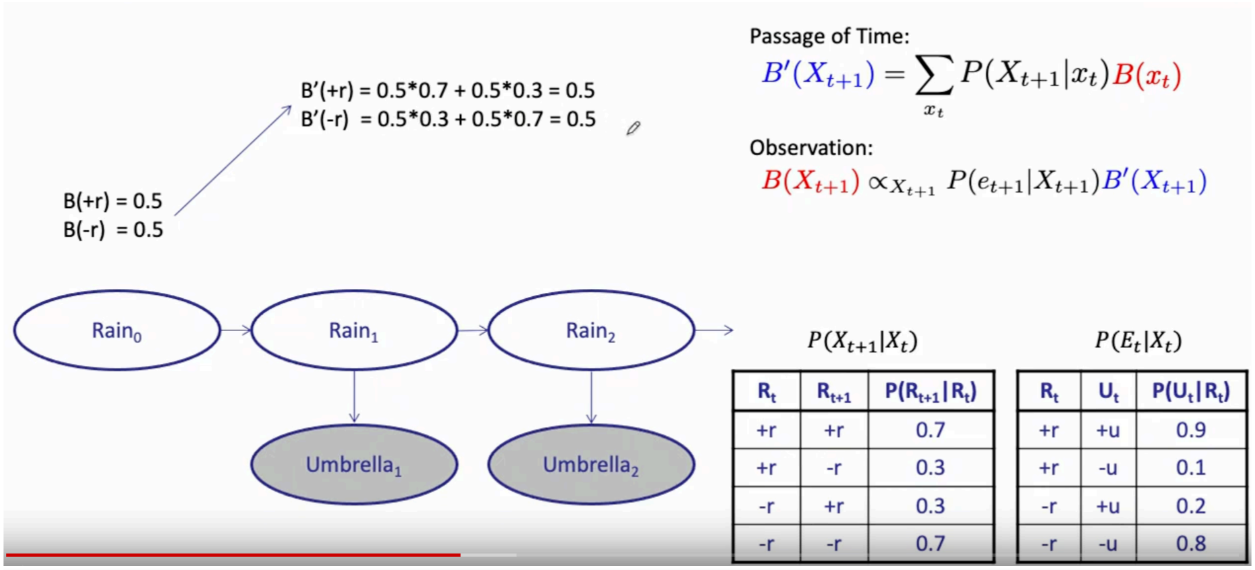
Online Belief Updates



- Algorithm iterates through these two steps - Forward algorithm $\triangleq$ combine these two equations:

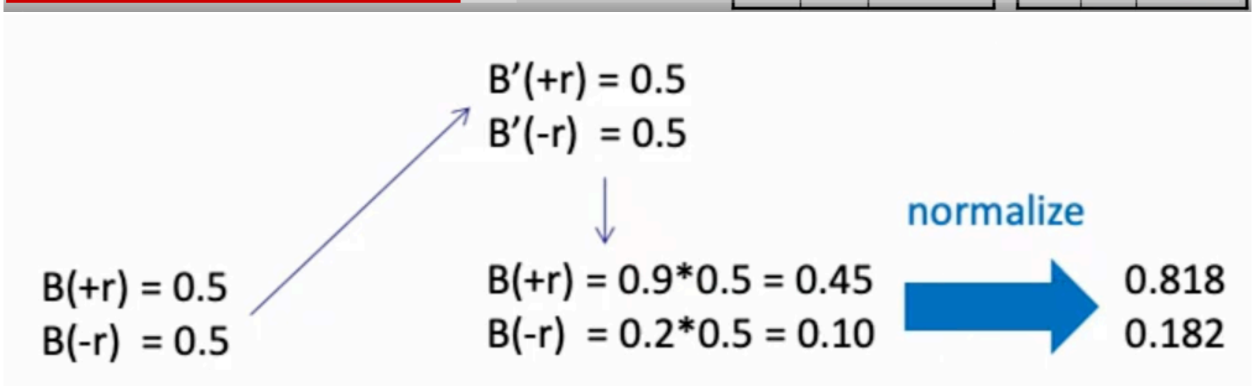
$$\mathbb{P}(e_t|x_t) = \sum_{x_{t-1}} \mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{t-1}, e_{1:t-1})$$

Concrete Example



R_t	R_{t+1}	$P(R_{t+1}\|R_t)$
+r	+r	0.7
+r	-r	0.3
-r	+r	0.3
-r	-r	0.7

R_t	U_t	$P(U_t\|R_t)$
+r	+u	0.9
+r	-u	0.1
-r	+u	0.2
-r	-u	0.8



Particle Filtering (Approximate Inference)

- Filtering $\triangleq$ approximate solution
- Sometimes $|X|$ is too big to use exact inference
  - $|X|$ may be too big to even store $B(X)$
  - eg: X is continuous
- Probability of some state is given by the amount of particles that occupy it
  - Sample $\iff$ Particle
  - Track samples of X, not all values
  - Time per step is linear in the number of samples
  - Number of samples needed may be large
  - In memory: list of particles, not states
- Keep track of N particles; representation of $P(X)$ is just a list of those particles
  - $N \ll |X|$
  - Storing map from X to counts would defeat point
- $P(x)$ approximated by number of particles with value x
  - Many x have $P(x) = 0$
  - More particles $\implies$ more accuracy

Observation of Time

- Each particle is moved by sampling its next position from the transition model: $x' = \text{sample}(\mathbb{P}(X'|x))$

- With enough samples, the before and after values are consistent

Weighting

- Don't sample observation, fix it

$$w(x) = \mathbb{P}(e|x)$$

$$B(X) \propto \mathbb{P}(e|X)B'(X)$$

- Probabilities have been down-weighted (they sum to N times an approximation of $\mathbb{P}(e)$)
  - Need to normalize
- Instead of tracking weighted samples, resample at each step
  - N times, we choose form our weighted sample distribution (draw with replacement)
  - Equivalent to re-normalizing the distribution
  - Completes the update for this time step
- IDEA: track samples of states rather than an explicit distribution

Dynamic Bayes' Nets

- Can track multiple variables over time using multiple sources of evidence
- Idea: repeat a fixed BN at each time
- Variables from time t can condition on those from $t-1$
- Variable elimination applies
  - Procedure: 'unroll' the network for T time steps, then eliminate variables until $\mathbb{P}(X_T|e_{1:t})$ is computed