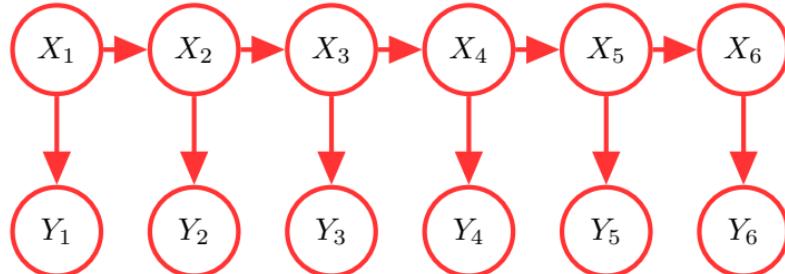


# Sequential Models

Andisheh Amrollahi

# HMMs / Kalman Filters



- $X_1, \dots, X_T$ : Unobserved (hidden) variables (called **states**)
- $Y_1, \dots, Y_T$ : Observations
- **HMMs**:  $X_i$  categorical,  $Y_i$  categorical (or arbitrary)
- **Kalman Filters**:  $X_i, Y_i$  Gaussian distributions

# HMMs

Filtering

$$P(x_t | y_{1:t})$$

Is somehow “online”

Prediction

$$P(x_{t+\Delta} | y_{1:t})$$

Smoothing

$$P(x_t | y_{1:T}) \quad \text{for } 1 \leq t \leq T$$

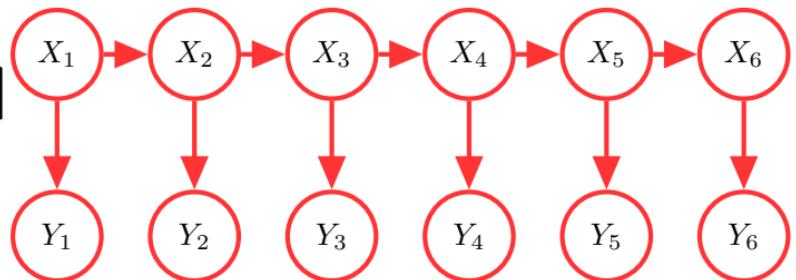
offline

Most probable explanation

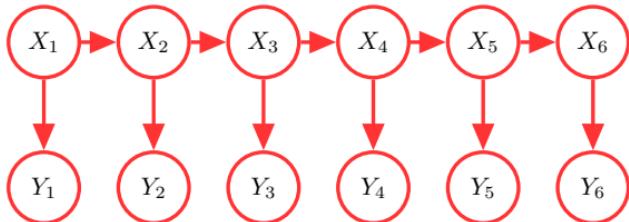
$$\underset{x_{1:T}}{\operatorname{argmax}} P(x_{1:T} | y_{1:T})$$

# Parameters of an HMM

- $X_t \in \{1, 2, \dots, n\}$
- $Y_t \in \{1, 2, \dots, m\}$
- $P(X_1 = s) \rightarrow n-1$  parameters
- $P(X_2 = s_2 | X_1 = s_1) \rightarrow n.(n-1)$  parameters
- $P(Y_t = o | X_t = s) \rightarrow n.(m-1)$  parameters



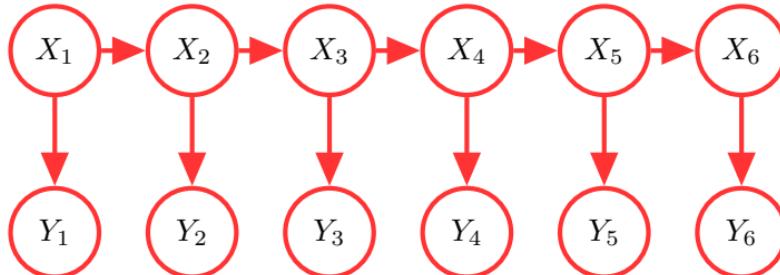
# Filtering (computational cost)



- Have  $P(X_1 = s_1 | \phi)$
- $P(X_1 = s_1 | Y_1 = o_1) = \frac{P(Y_1 = o_1 | X_1 = s_1)P(X_1 = s_1)}{\sum_{s=1}^n P(Y_1 = o_1 | X_1 = s)P(X_1 = s)}$   $O(n)$
- $P(X_2 = s_2 | Y_1 = o_1) =$   
 $\sum_{s=1}^n P(X_2 = s_2 | X_1 = s, Y_1 = o_1)P(X_1 = s | Y_1 = o_1) =$   
 $\sum_{s=1}^n P(X_2 = s_2 | X_1 = s)P(X_1 = s | Y_1 = o_1).$   $O(n^2)$
- Have  $P(X_2 = s_2 | Y_1 = o_1)$
- Compute  $P(X_2 = s_2 | Y_1 = o_1, Y_2 = o_2)$

Total runtime  $O(n^2 T)$

# HMMs / Kalman Filters



Steps of the filtering algorithm are similar here

- $X_1, \dots, X_T$ : Unobserved (hidden) variables (called **states**)
- $Y_1, \dots, Y_T$ : Observations
- **HMMs**:  $X_i$  categorical,  $Y_i$  categorical (or arbitrary)
- **Kalman Filters**:  $X_i, Y_i$  Gaussian distributions

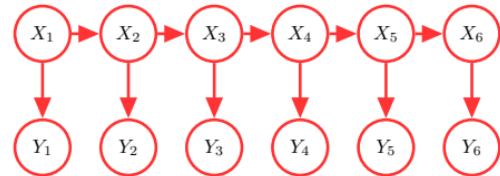
# Kalman Filters (Gaussian HMMs)

- $X_1, \dots, X_T$ : Location of object being tracked
- $Y_1, \dots, Y_T$ : Observations
- $P(X_1)$ : Prior belief about location at time 1
- $P(X_{t+1} | X_t)$ : “Motion model”
  - How do I expect my target to move in the environment?

$$\mathbf{X}_{t+1} = \mathbf{F}\mathbf{X}_t + \varepsilon_t \text{ where } \varepsilon_t \in \mathcal{N}(0, \Sigma_x)$$

- $P(Y_t | X_t)$ : “Sensor model”
  - What do I observe if target is at location  $X_t$ ?

$$\mathbf{Y}_t = \mathbf{H}\mathbf{X}_t + \eta_t \text{ where } \eta_t \in \mathcal{N}(0, \Sigma_y)$$



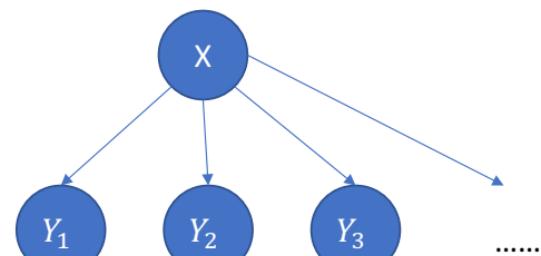
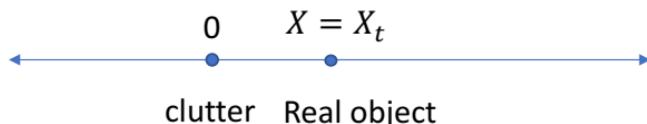
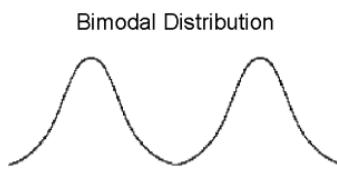
# Tracking an object in the presence of clutter

- Let's try to "break" this algorithm

$$\bullet p(x_1) = \frac{1}{2\pi} e^{-\frac{x_1^2}{2}}$$

$$\bullet X_{t+1} = X_t$$

$$\bullet p(y_t|x_t) = \frac{1}{2} \frac{1}{2\pi} e^{-\frac{y_t^2}{2}} + \frac{1}{2} \frac{1}{2\pi} e^{-\frac{(y_t-x_t)^2}{2}}$$



# Filtering (computational cost)

- Have  $p(x) = N(x; 0, 1)$
- $p(x|y_1 = o_1) = \frac{p(y_1 = o_1|x)p(x)}{\int_x p(y_1 = o_1|x)p(x)} =$  weighted sum of two Gaussians in  $x$   
i.e  $\propto \left( \frac{1}{2} \frac{1}{2\pi} e^{-\frac{o_1^2}{2}} + \frac{1}{2} \frac{1}{2\pi} e^{-\frac{(o_1-x)^2}{2}} \right) * \left( \frac{1}{2\pi} e^{-\frac{x^2}{2}} \right)$
- $p(x|y_1 = o_1, y_2 = o_2) = \frac{p(y_1 = o_1, y_2 = o_2 | x)p(x)}{z} =$   
 $\frac{p(y_1 = o_1|x)p(y_2 = o_2 | x)p(x)}{z}$  Weighted sum of four Gaussians

$$\left( \frac{1}{2} \frac{1}{2\pi} e^{-\frac{o_1^2}{2}} + \frac{1}{2} \frac{1}{2\pi} e^{-\frac{(o_1-x)^2}{2}} \right) \left( \frac{1}{2} \frac{1}{2\pi} e^{-\frac{o_2^2}{2}} + \frac{1}{2} \frac{1}{2\pi} e^{-\frac{(o_2-x)^2}{2}} \right) \left( \frac{1}{2\pi} e^{-\frac{x^2}{2}} \right)$$

- Number of modes grow exponentially!

From Previous step

# Solution: Assumed Density Filtering

- $p(x|y_1 = o_1) = \frac{p(y_1 = o_1|x)p(x)}{\int_x p(y_1 = o_1|x)p(x)}$  = weighted sum of two Gaussians in x

$$\text{i.e. } \left( \frac{1}{2} \frac{1}{2\pi} e^{-\frac{o_1^2}{2}} + \frac{1}{2} \frac{1}{2\pi} e^{-\frac{(o_1-x)^2}{2}} \right) * \left( \frac{1}{2\pi} e^{-\frac{x^2}{2}} \right) \longrightarrow \text{Approximate posterior with a Gaussian}$$

- $q(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$
- $\mu, \sigma = \arg \min KL(p(x|y_1 = o_1) || q(x; \mu, \sigma))$
- Done via moment matching (See HW questions 4 and 5)

# Solution: Assumed Density Filtering

- $\mu, \sigma = \arg \min KL(p(x|y_1 = o_1) || q(x; \mu, \sigma))$
- Done via moment matching (See HW questions 4 and 5)
- $\mu = E_{x \sim p(x|y_1)}[x], \quad \sigma = E_{x \sim p(x|y_1)}[(x - \mu)^2]$

# More Generally

- Let  $D$  be observed and let  $\theta$  be the parameters to be inferred
- $p(D, \theta) = \prod_i t_i(\theta)$
- Initialize  $q^0(\theta) = 1$
- At round  $i$ ,  $\hat{p}(\theta) = \frac{t_i(\theta)q^{(i-1)}(\theta)}{\int_{\theta} t_i(\theta)q^{(i-1)}(\theta)}$
- Pick  $q^{(i)}(\theta) = \underset{q \in Q}{\operatorname{argmin}}(KL(\hat{p}(\theta), q(\theta)))$



Some tractable family  
such as the  
exponential Family