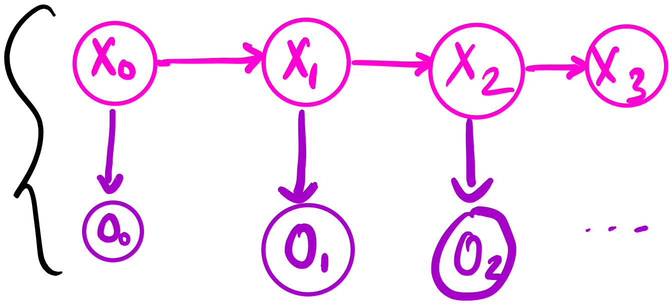
# HMM

**Recap.** what is a HMM?

$$\lambda = \{ \underbrace{\{s_1, \dots s_N\}}_{S}^{N}, \underbrace{\{o_1, \dots o_M\}}_{\Theta}^{M}, a_{ij}, b_i(\ell), \pi \}$$




$x_t \in S$, $O_t \in \Theta$

$\pi_i = Pr(X_0 = i)$

$a_{ij} = Pr(x_{t+1} = s_j \mid x_t = s_i)$, $b_i(\ell) = Pr(O_t = \theta_\ell \mid x_t = s_i)$

1. State prob. estimation. (Forward-Backward Alg)

Estimate $Prob(x_t = s_j \mid O_0 \dots O_T, \lambda)$.
(not $t$)

2. Most Probable Path (MPP). (viterbi Algorithm)

Given $O_1, \dots O_T$, and $\lambda$

$$\max_{s_0 s_1 \dots s_T} Prob(x_0 = s_0, x_1 = s_1 \dots x_T = s_T \mid O_0 \dots O_T, \lambda)$$

3. Learn the HMM (Baum Welch Algorithm) → EM algorithm

$$\hat{\lambda} = \arg\max_{\lambda} Pr(O_0 \dots O_T \mid \lambda).$$

Today :- 1. Baum Welch Algorithm

2. Recap of the class.

---

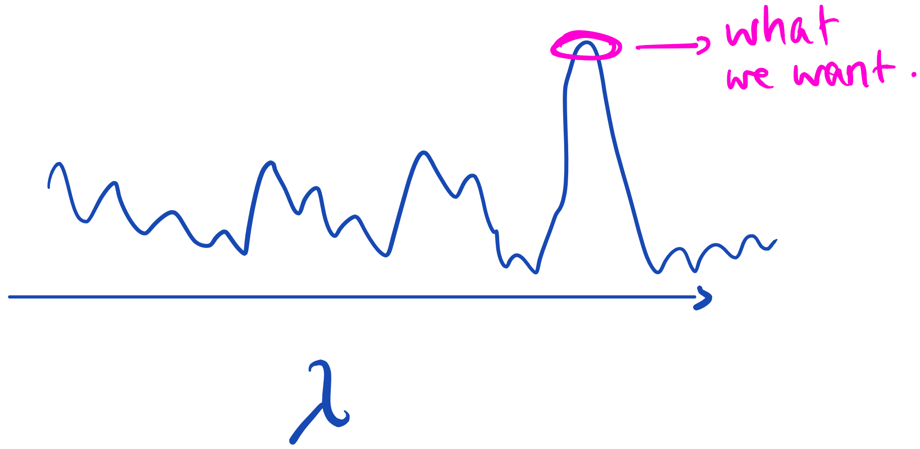
Goal :- Find a "good" model

Maximum Likelihood Principle. (recall).

$$\hat{\lambda} = \arg\max_{\lambda} \Pr(O_0 \dots O_T \mid \lambda).$$

unfortunately, this is a bad optimization problem in about $N^2 + MN + N$ dimensions

Given $O_0 \dots O_T$,




iterative method to improve estimates.

e.g. clustering, recommendor systems

$\lambda^{(0)}$ — initial guess for $\lambda$.

keep improving ...

**Recall** :- $\gamma_t(i) := \Pr\ X_t = s_i \mid O_0 \dots O_T, \lambda)$.

Let $\epsilon_t(i,j) := \Pr\left(X_t = s_i, X_{t+1} = s_j \mid O_0 \dots O_T, \lambda\right)$.

$$\gamma_t(i) = \sum_{j=1}^{N} \epsilon_t(i,j) \quad \text{(Marginalization)}.$$

$$\epsilon_t(i,j) = \frac{\Pr(O_0 \dots O_T, X_t = S_i, X_{t+1} = S_j \mid \lambda)}{\Pr(O_0 \dots O_T \mid \lambda)}$$

$$= \frac{\Pr(O_0 \dots O_t, X_t = S_i) \cdot \Pr(X_{t+1} = S_j \mid X_t = S_i) \cdot \Pr(O_{t+1} \mid X_{t+1} = S_j) \cdot \Pr(O_{t+2} \dots O_T \mid X_{t+1} = S_j)}{\Pr(O_0 \dots O_T \mid \lambda)}$$

$$= \frac{\alpha_t(i)\ a_{ij} \cdot b_j(O_{t+1}) \cdot \beta_{t+1}(j)}{\Pr(O_0 \dots O_T \mid \lambda)}$$

**Key Idea** :-

$T-1$

What is $\sum_{t=0} \epsilon_t(i,j)$ ?

$\longrightarrow$ the expected number of transitions from $s_i \longrightarrow s_j$

$\sum_{t=0}^{T-1} \gamma_t(i) \longrightarrow$ total # transitions from 'i',

$\Leftrightarrow$ total time spent @ $s_i$.

## ALGORITHM:-

* Compute $\epsilon_t(i,j), \gamma_t(i)$ for $\lambda^{(0)}$.

* New model:-

$$\pi_i^{new} = \gamma_0(i)$$

$$a_{ij}^{(new)} = \frac{\sum_{t=0}^{T-1} \epsilon_t(i,j)}{\sum_{t=0}^{T-1} \gamma_t(i)} \quad (?)$$

$$b_i(\ell) = \frac{\sum_{t:O_t=\ell}^{T-1} \gamma_t(i)}{\sum_{t=0}^{T-1} \gamma_t(i)} \quad (?)$$

* $\lambda^{(new)} \longrightarrow \lambda^{(0)}$ & <u>repeat</u> ...

<u>Exercise</u> :- Read "Rabiner Tutorial" as your summer reading.