

$$V = \{a, b, c\} \quad Y = \{0, 1\}$$

$$x_0 = aab \quad x_1 = cbc \quad x_2 = aaa \quad x_3 = cbcc$$

$$\begin{array}{l} \underline{I_{\text{init}}} \quad \pi_0 = .4 \quad \pi_1 = .6 \\ P(a|0) = .6 \quad P(b|0) = .3 \quad P(c|0) = .1 \\ P(a|1) = .3 \quad P(b|1) = .4 \quad P(c|1) = .3 \end{array} \quad \left. \vphantom{\begin{array}{l} \underline{I_{\text{init}}} \\ P(a|0) = .6 \\ P(a|1) = .3 \end{array}} \right\} I_{\text{init}}$$

$$\begin{array}{l} \underline{Z_0} \leftarrow \arg \max_y P(aab|y)P(y) = 0 \\ Z_1 \leftarrow \arg \max_y P(cbc|y)P(y) = 1 \\ Z_2 \leftarrow \arg \max_y P(aaa|y)P(y) = 0 \\ Z_3 \leftarrow \arg \max_y P(cbcc|y)P(y) = 1 \end{array} \quad \left. \vphantom{\begin{array}{l} \underline{Z_0} \\ Z_1 \\ Z_2 \\ Z_3 \end{array}} \right\} E\text{-Step}$$

$$\begin{array}{l} \underline{\pi_0^{(1)}} \leftarrow \frac{2}{4} \quad \pi_1^{(1)} \leftarrow \frac{2}{4} \\ P(a|0) \leftarrow \frac{5}{3+3} \quad P(a|1) \leftarrow \frac{0}{3+4} \\ \quad = \frac{5}{6} \quad \quad \quad = \emptyset \\ P(b|0) \leftarrow \frac{1}{6} = .167 \quad P(b|1) \leftarrow \frac{2}{7} = .285 \\ P(c|0) \leftarrow 0 \quad P(c|1) \leftarrow \frac{5}{7} = .714 \end{array} \quad \left. \vphantom{\begin{array}{l} \underline{\pi_0^{(1)}} \\ P(a|0) \\ P(b|0) \\ P(c|0) \end{array}} \right\} M\text{-Step}$$

For "soft" EM

$$P(\tau | c) = \frac{\sum_i P(z_i = c) \cdot \text{COUNT}(\tau \text{ in } x_i)}{\sum_i P(z_i = c) \cdot |x_i|}$$

expected # of times
 τ occurs in c

$$\sum_i P(z_i = c) \cdot |x_i|$$

total token count in X_i