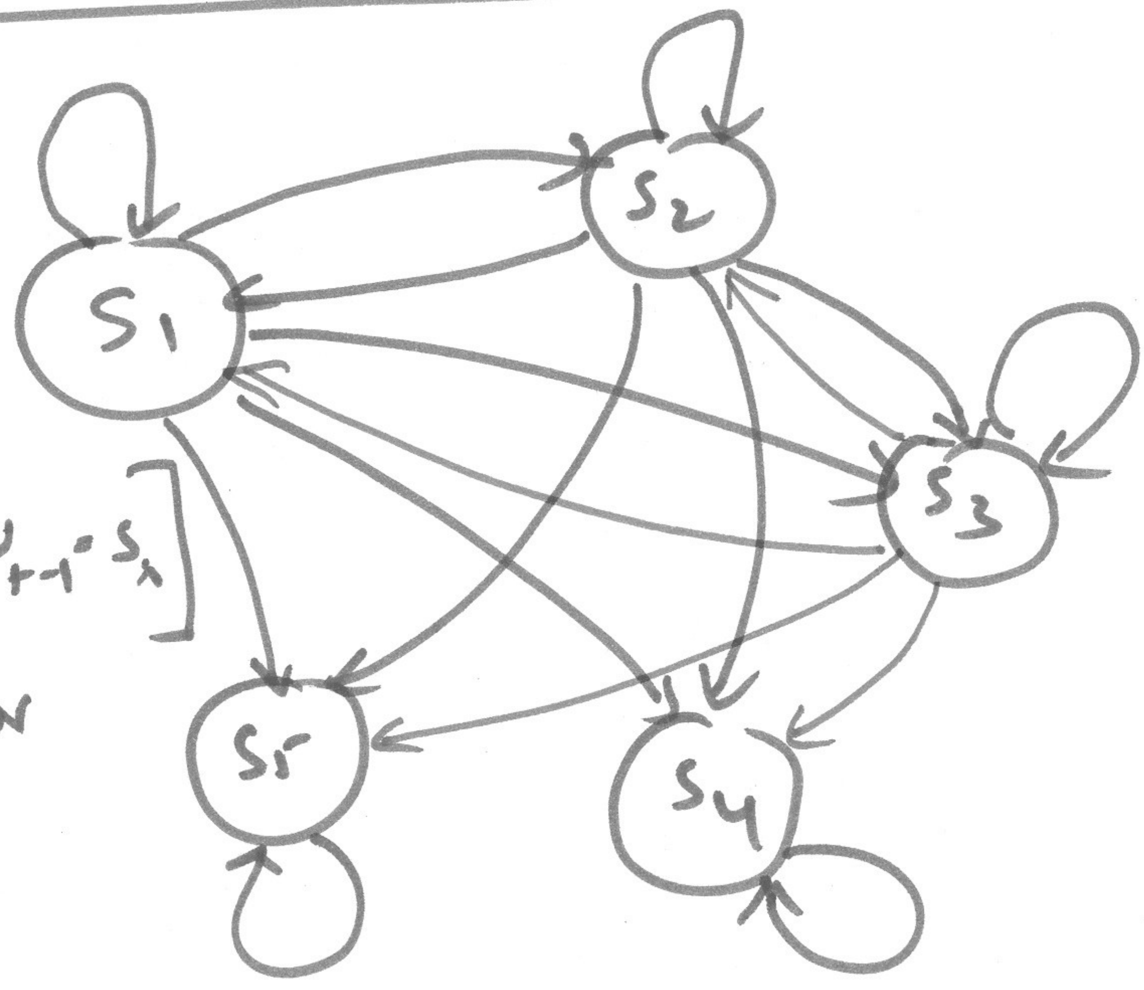
© CET
I.I.T. KGP

# Hidden Markov Models

## (HMM)

# Discrete Markov Process




$$a_{ij} = P\left[q_t = S_j \mid q_{t-1} = S_i\right]$$

$$i \geq 1 \; ; \; j \leq N$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

# Hidden Markov Model


O = (B, G, R, R, G, G, B, B, R, . . . .)

=> # boxes, Sequence of boxes

symbol probs

Spectral vectors <=> Sequence of VT shapes

# Basic Elements of HMM

1. # States — $N$
2. # distinct symbols / state — $M$

$$V = \{v_1\ v_2 \cdots v_M\}$$

3. State transition probability $A = \{a_{ij}\}_{M\times}$

~~4~~ $a_{ij} = P\left[q_{t+1} = j \mid q_t = i\right]$ ; $i \geq 1$, $j \leq N$

4. Observation symbol prob $B = \{b_j(k)\}$

$$b_j(k) = P\left[o_t = v_k \mid q_t = j\right]$$

5. initial prob $\pi_i = \pi$ ; $\pi_i = P\left[q_1 = i\right]$

MODEL $= \lambda = [A, B, \pi]$

# Basic Problems in HMM

① How do we efficiently compute $P(O/\lambda)$

$P(o_1\, o_2 \,---\, o_T / \lambda)$ => Testing, evaluation
Recognition, calculation
. . .

② $O = o_1\, o_2 \cdots o_T$
$\lambda = (A, B, \pi)$ } => $Q = q_1\, q_2 \cdots q_T$
(Seq of states)

best seq that explains the observations in better way
optimal seq for the given observation

# states & # symbols can be modified using ~~step~~ step-2

③ How do we adjust $\lambda = (A, B, \pi)$ to max $P(O/\lambda)$

Training step, building models

Solution to Problem-1 (computation of $P(O/\lambda)$)

$$O = O_1 O_2 \cdots O_T \quad , \quad \lambda = A, B, \pi$$

Possible state seq $= Q = (q_1 q_2 \cdots q_T) \Rightarrow N^T$

$$P(O/\lambda) = \sum_{q_1 q_2 \cdots q_T} \cancel{P(O/q,\lambda)}\; P(O, q/\lambda)$$

$$P(O, q/\lambda) = P(O/q, \lambda)\, P(q/\lambda)$$

$$P(O/q, \lambda) = \prod_{t=1}^{T} P(O_t / q_t, \lambda)$$

$$= b_{q_1}(O_1)\, b_{q_2}(O_2) \cdots b_{q_T}(O_T)$$

$$P(q/\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}$$

$$P(O, q/\lambda) = P(O/q, \lambda)\, P(q/\lambda)$$

$$= \underbrace{\pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots b_{q_{T-1}}(O_T) a_{q_{T-1} q_T} b_{q_T}(O_T)}_{(2T-1)\ \text{multiplications}}$$

$$P(O/\lambda) = \sum_{q_1 q_2 \cdots q_T} P(O, q/\lambda)$$

$q_1 q_2 \cdots q_T \Rightarrow N_T$ state sequences

Total no. of computations: $2TN^T$ – multiplications; $N^T - 1$ – additions

<u>Eg</u> $N = 5$, $T = 100$

$\#$ computations $= 2 \times 100 \times 5^{100} \Rightarrow \underline{10^{72}}$

## Forward Procedure $\left[ \alpha_T(i) \right]$

$\alpha_T(i) \rightarrow$ Prob of partial observation sequence

$$\alpha_T(i) = P(o_1 o_2 \dots o_t, q_t = i \mid \lambda)$$

1. Initialization $\alpha_1(i) = P(o_1, q_1 = i \mid \lambda)$

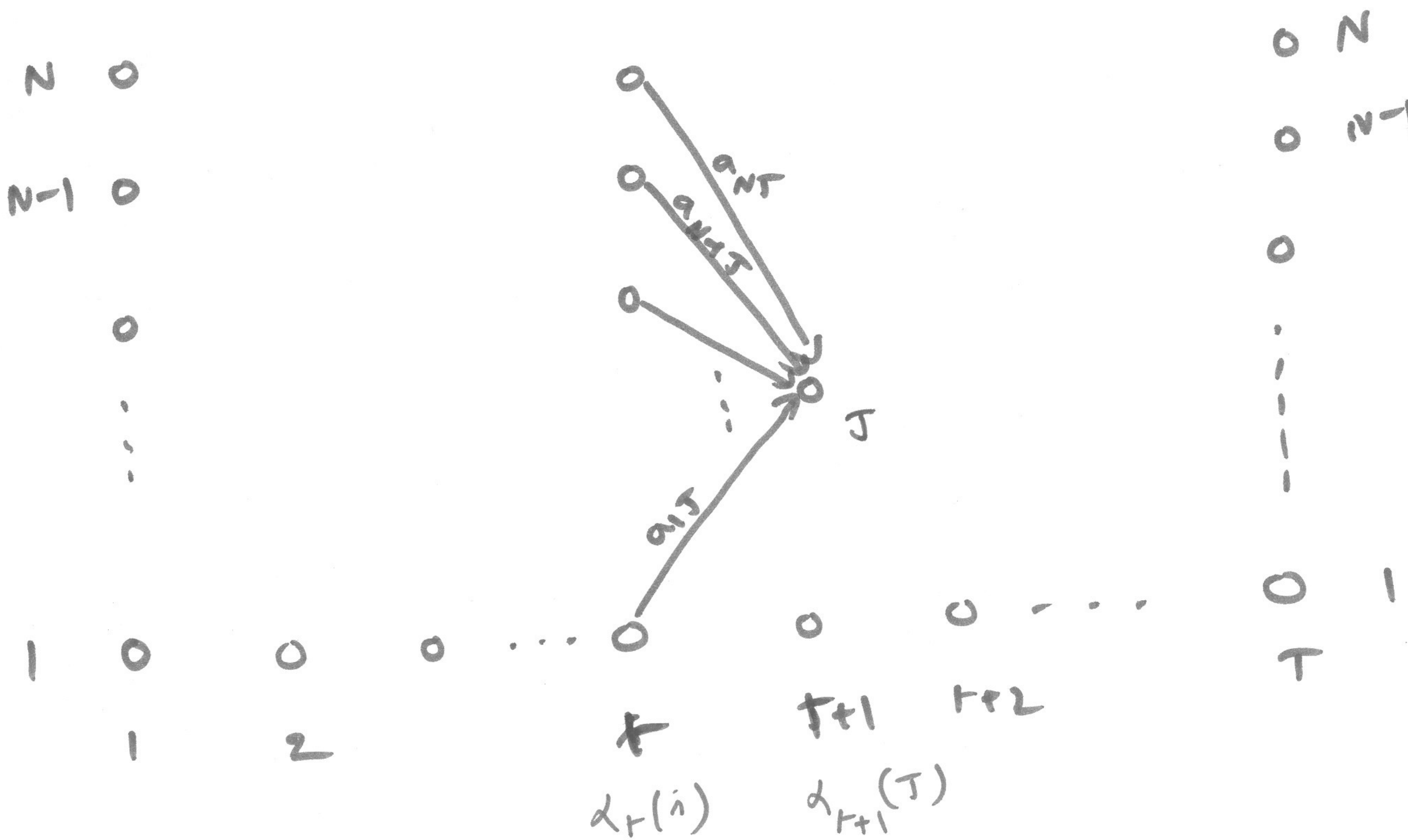
$$= \pi_i b_i(o_1) \quad ; \quad 1 \le i \le N$$

2. Induction

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \; ; \quad \begin{array}{l} 1 \le t \le T-1 \\ 1 \le j \le N \end{array}$$

3. Termination $P(O/\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

Illustration of computation of forward variable $\alpha_i(t)$




# computations $\Rightarrow N^2T$

## Backward Procedure $\beta_t(i)$

$$\beta_t(i) = P\left(O_{t+1}\, O_{t+2} \cdots O_T \mid q_t = i,\ \lambda\right)$$

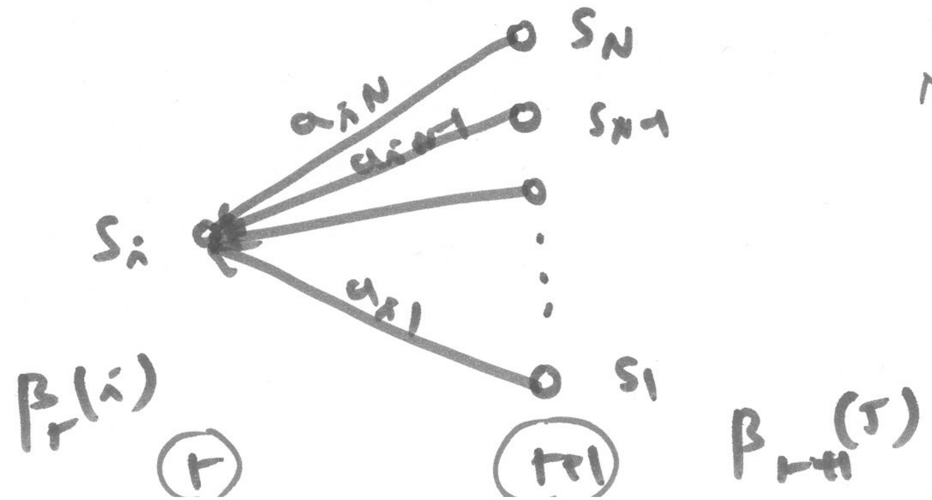
**Initialization**

$$\beta_T(i) = 1 \qquad 1 \le i \le N$$

**Induction**

$$\beta_t(i) = \sum_{J=1}^{N} a_{iJ}\, b_J(O_{t+1})\, \beta_{t+1}(J)$$

$$t = T-1,\ T-2,\ \cdots 1\ ; \qquad 1 \le i \le N$$




$N^2 T$ computations

Solution to problem-2 (Optimal state sequence)

- Individually most likely states at each time
- Max expectation of pairs of states $(q_t \, q_{t+1})$
- Max expectation of tripplets of states $(q_{t-1} \, q_t \, q_{t+1})$

$$\gamma_t(i) = P(q_t = i \mid O, \lambda) = \frac{P(O, q_t = i \mid \lambda)}{\sum_{i=1}^{N} P(O, q_t = i \mid \lambda)}$$

$$= \frac{P(O, q_t = i \mid \lambda)}{P(O/\lambda)} = \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\,\beta_t(i)}$$

$$q_t^* = \arg\max_i \left(\gamma_t(i)\right)$$

For speech some of $a_{ij} = 0$ ; $\therefore$ optimal seq with $q_t^*$ may not be valid sequence

Viterbi Algorithm (Dynamic programming)

$$\delta_t(i) = \max P[o_1 o_2 \cdots o_t \; v_1 v_2 \cdots v_{t-1} \; v_t = i \mid \lambda]$$

$v_1 v_2 \cdots v_{t-1}$

By induction

$$\delta_{t+1}(j) = \max_i \left[\delta_t(i) \, a_{ij}\right] b_j(o_{t+1})$$

$\psi_t(j)$ = tracking the argument that max for each $t$ & $j$

$t = 2$ to $T$; $1 \le j \le N$

Initialization

$$\delta_1(i) = \pi_i \, b_i(o_1) \; ; \quad 1 \le i \le N$$

$$\psi_1(i) = 0$$

Recursion

$$\delta_t(j) = \max_i \left[\delta_{t-1}(i) \, a_{ij}\right] b_j(o_t) \qquad 2 \le t \le T$$

$$\psi_t(j) = \arg\max_i \left[\delta_{t-1}(i) \, a_{ij}\right] \qquad 1 \le j \le N$$

Veterbi Algorithm Soln to problem-2

**Termination** $\delta_T(j) = \max_i \left[\delta_{T-1}(i)\, a_{ij}\right] b_j(O_T)$

$$p^* = \max_i \delta_T(i) \; ; \; q_T^* = \arg\max_i \left[\delta_T(i)\right]$$

**Path backtracking (state sequence)** $q_t^* = \psi_{t+1}(q_{t+1}^*)$

$t = T-1, T-2, \ldots 1$

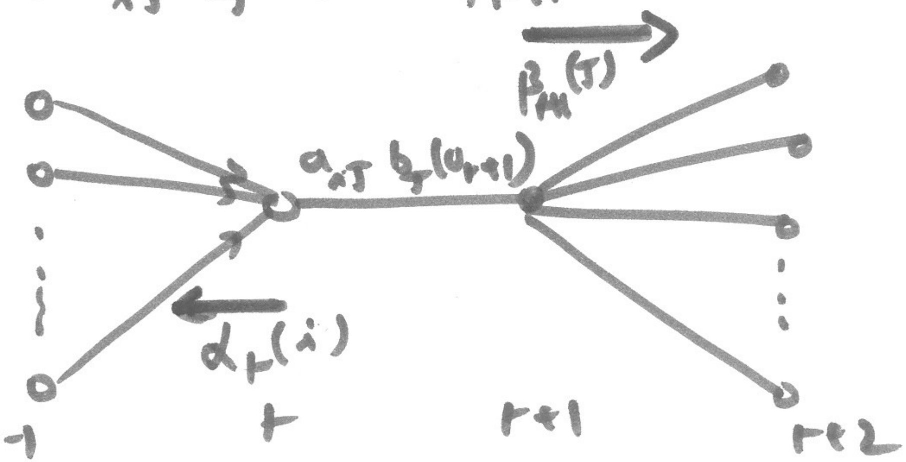
Soln to Problem - 3 (Parameter estimation)

Baum - welch method (Iterative approach)

$$\xi_t(i,j) = P(q_t = i,\ q_{t+1} = j \mid O, \lambda)$$

$$= P(O,\ q_t = i,\ q_{t+1} = j / \lambda) \,/\, P(O/\lambda)$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum_i \sum_j \alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}$$



$$\gamma_t(i) = P\left[q_t = i \mid O, \lambda\right] = \sum_{j=1}^{N} \xi_t(i, j)$$

Avg # transitions from state 'i' in 'O' $= \sum_{t=1}^{T-1} \gamma_t(i)$

Avg # transition from state $i \rightarrow j$ in 'O' $= \sum_{t=1}^{T-1} \xi_t(i, j)$

$\overline{\pi_i}$ = expected # times in state 'i' at $t=1$ $= \gamma_1(i)$

$$\overline{a_{ij}} = \frac{\text{Expect \# transitions from } i \rightarrow j}{\text{Expected \# transit from } i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\overline{b_j(k)} = \frac{\text{Expected \# times in state 'j' \& } O_t = v_k}{\text{Expected \# times in state } j} = \frac{\sum_{t=1}^{T} \gamma_t(j) \big/_{O_t = v_k}}{\sum_{t=1}^{T} \gamma_t(j)}$$

$\overline{\lambda} = (\overline{A}, \overline{B}, \overline{\pi})$; $P(O/\overline{\lambda}) > P(O/\lambda)$

# Types of HMMs

Based on structure of transition matrix — Ergodic / Left-right

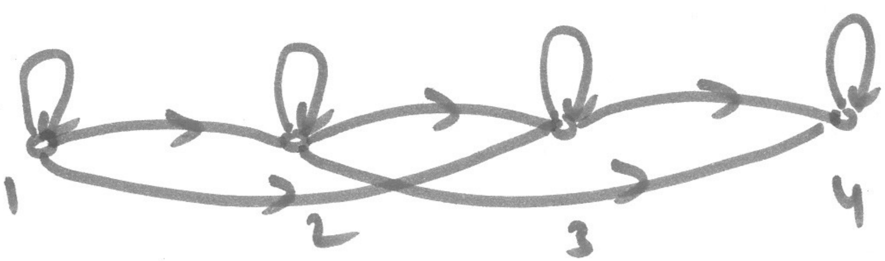
(1) Ergodic (fully connected) HMM

$$\forall \; a_{ij} > 0$$

(2) Left-right model (Bakis's model)

$$\left.\begin{array}{ll} a_{ij} = 0 & j < i \\ \pi_i = 0 & i \neq 1 \\ \quad = 1 & i = 1 \end{array}\right\} \qquad a_{ij} = 0 \quad j > i + \Delta i$$

$$\Delta i = 2$$




(3) Discrete vs Continuous HMMs

## Continuous observation densities in HMMs

Prob density of observation ~~freq~~ symbols at state $J = b_J(O)$

$$b_J(O) = \sum_{K=1}^{M} c_{JK}\, N(O, \mu_{JK}, \Sigma_{JK})$$

$c_{JK}$: Mixture coeff for $K^{th}$ component

$N$: Gaussian density

$\mu_{JK}$: Mean vector for ~~the~~ $K^{th}$ component

$\Sigma_{JK}$: Co-variance matrix for $K^{th}$ component

$$\sum_{K=1}^{M} c_{JK} = 1 \qquad c_{JK} \geq 0$$

$$1 \leq J \leq N \; ; \; 1 \leq K \leq M$$

$$\int_{-\infty}^{\infty} b_J(O)\, dO = 1 \qquad 1 \leq J \leq N$$

## Reestimation formulas for mixture parameters in continuous density HMMs

$$\bar{c}_{JK} = \sum_{t=1}^{T} \gamma_t(J,K) \Big/ \sum_{t=1}^{T}\sum_{K=1}^{M} \gamma_t(J,K)$$

$$\bar{\mu}_{JK} = \sum_{t=1}^{T} \gamma_t(J,K)\cdot o_t \Big/ \sum_{t=1}^{T} \gamma_t(J,K)$$

$$\bar{\Sigma}_{JK} = \sum_{t=1}^{T} \gamma_t(J,K)\,(o_t-\mu_{JK})(o_t-\mu_{JK})^T \Big/ \sum_{t=1}^{T} \gamma_t(J,K)$$

$$\gamma_t(J,K) = \left[\frac{\alpha_t(J)\,\beta_t(J)}{\sum_{J=1}^{N}\alpha_t(J)\,\beta_t(J)}\right]\left[\frac{c_{JK}\,N(o_t,\mu_{JK},\Sigma_{JK})}{\sum_{K=1}^{M} c_{JK}\,N(o_t,\mu_{JK},\Sigma_{JK})}\right]$$

$$\gamma_t(J,K) = \gamma_t(J)$$ Mixture with only one component