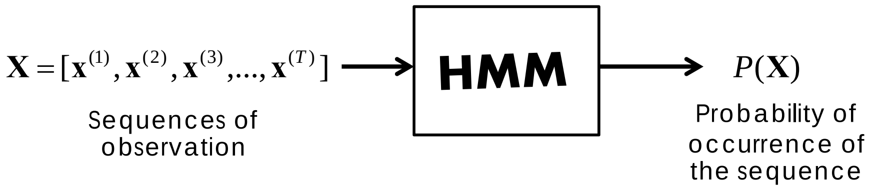
Audio & Speech Technology

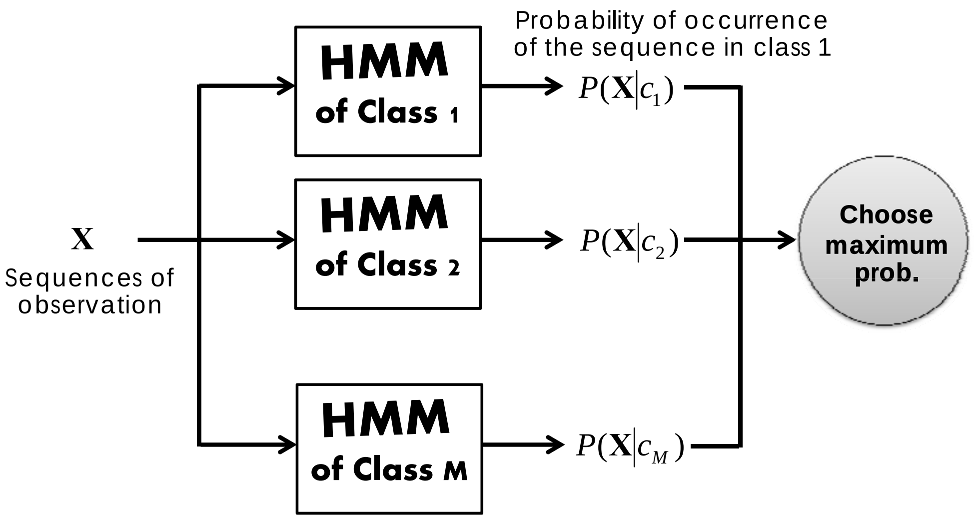
[5] Hidden Markov Model

Hidden Markov Model

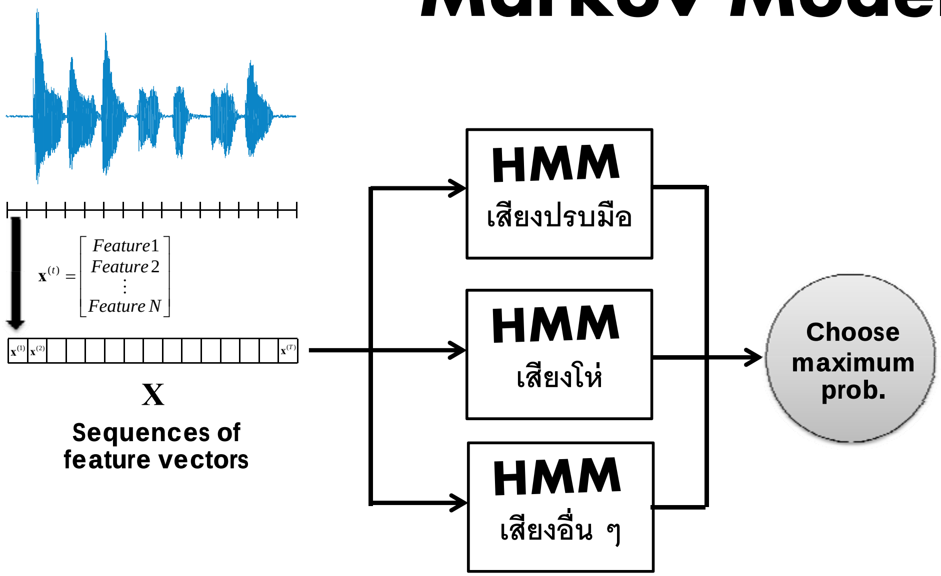
- Statistical method used in pattern classification
- Widely used in speech recognition, speaker identification
- Handle sequences of observation probabilistically



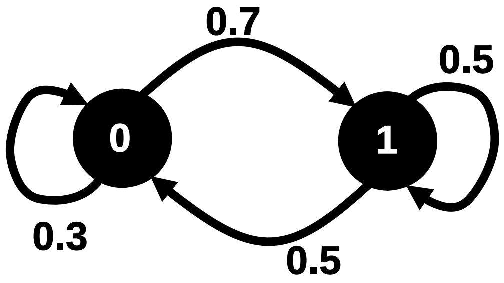
Hidden Markov Model



Hidden Markov Model

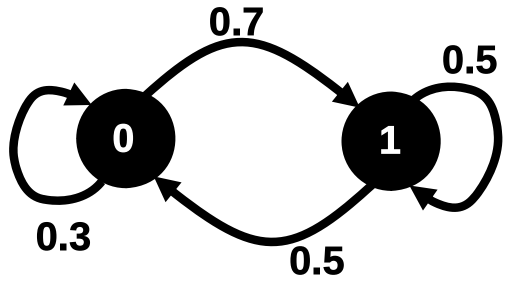


Markov Chain



- For observation sequences $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, ..., \mathbf{x}^{(T)}]$
- Probability of occurrence observation $\mathbf{x}^{(i)}$ is depended on value of previous observation $\mathbf{x}^{(i-1)}, ..., \mathbf{x}^{(1)}$
- State = Possible value of each observation
- Finite or countable number of possible states

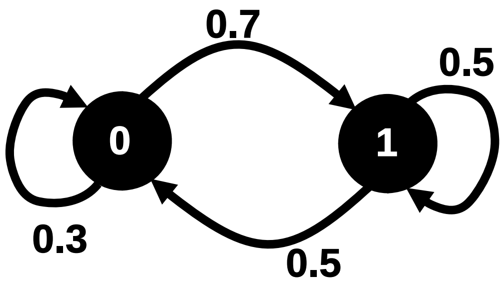
Markov Chain



- Each observation $x^{(t)} \in \{0,1\}$ so there are two states $\{0,1\}$
- Transition Probability

$$P(x^{(t)} = 0 | x^{(t-1)} = 0) = 0.3 \qquad P(x^{(t)} = 1 | x^{(t-1)} = 0) = 0.7$$
$$P(x^{(t)} = 0 | x^{(t-1)} = 1) = 0.5 \qquad P(x^{(t)} = 1 | x^{(t-1)} = 1) = 0.5$$

Markov Chain



Find probability of sequence $\mathbf{X} = [0\ 1\ 0\ 0\ 1\ 1]$
when $P(x^{(1)} = 0) = 0.5$

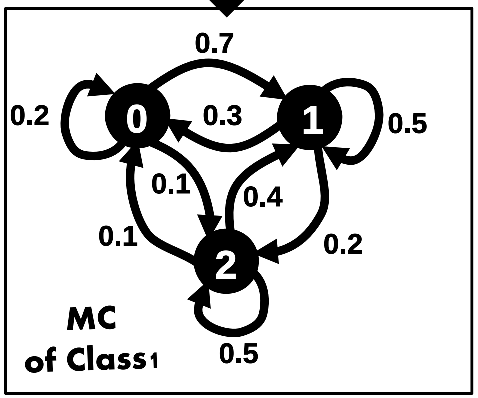
$$P(\mathbf{X}) = P(x^{(1)})P(x^{(2)}|x^{(1)})P(x^{(3)}|x^{(2)})P(x^{(4)}|x^{(3)})P(x^{(5)}|x^{(4)})P(x^{(6)}|x^{(5)})$$
$$= P(0)P(1|0)P(0|1)P(0|0)P(1|0)P(1|1)$$
$$= 0.5 \times 0.7 \times 0.5 \times 0.3 \times 0.7 \times 0.5 = 0.0184$$

Markov Chain

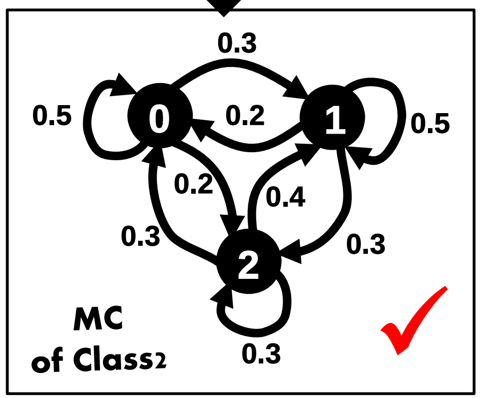
Observed sequence $\mathbf{Y} = [0\ 0\ 2\ 1\ 2\ 1]$

$$class(\mathbf{Y}) = ?$$

Training Data form Class 1



Training Data form Class 2

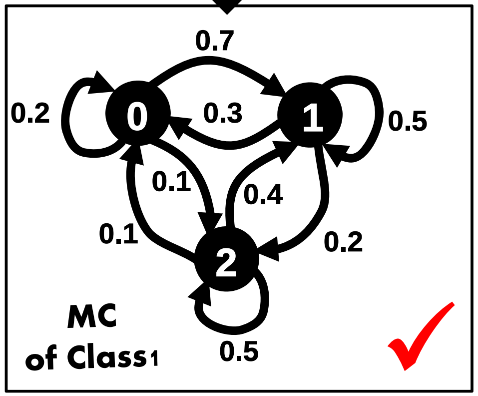


Markov Chain

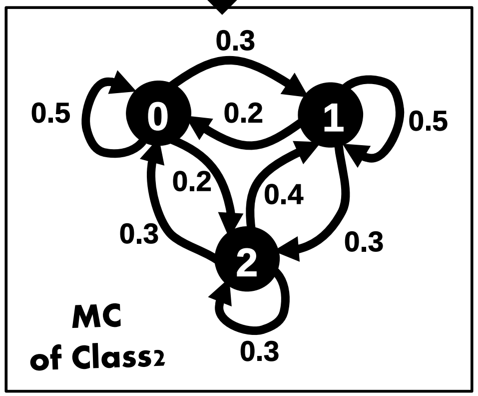
Observed sequence $\mathbf{Y} = [0\ 1\ 1\ 1\ 2\ 2]$

$class(\mathbf{Y}) = ?$

Training Data form Class 1

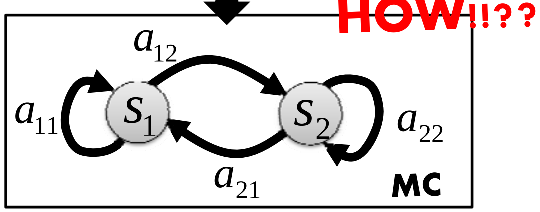


Training Data form Class 2



Markov Chain

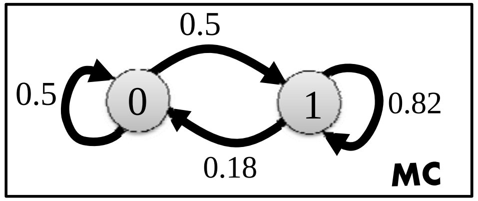
Training Data



- Find transition probability $\{a_{ij}\}$ to maximize Likelihood $L(\{a_{ij}\}|\{\mathbf{X}\})$ from given sequences of observation $\{\mathbf{X}\}$

Training Data

[0 0 0 1 1 1 1 0 0 1]
[0 0 1 1 1 1 1 1 0 0]
[0 1 1 1 1 0 1 1 1 1]



Markov Chain

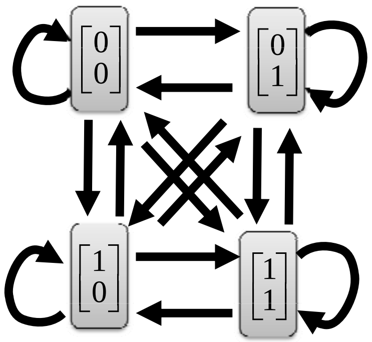
2 Binary Features

Training Data

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$
← Feature 1
← Feature 2

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

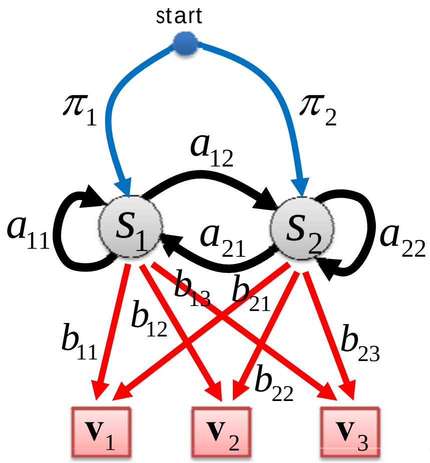
$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$



4 Possible States

Hidden Markov Model

State ≠ Observation

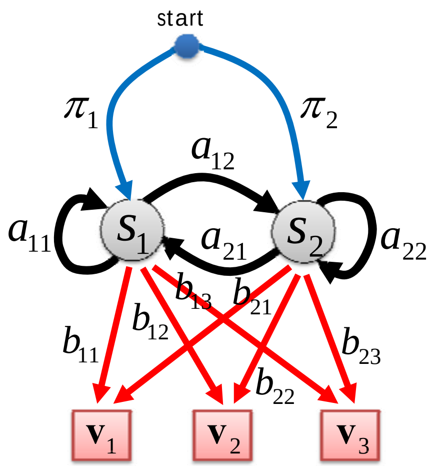


Possible State : S_i

Possible Observation : $\mathbf{v}_i$

[Discrete Observation Value]

Hidden Markov Model



$\{\pi_i\} =$ **Initial State Probability**

$\{a_{ij}\} =$ **State Transition Probability**

$\{b_{ij}\} =$ **Emission Probability (Output Probability)**

$$\lambda = \{\{a_{ij}\}, \{b_{ij}\}, \{\pi_i\}\}$$

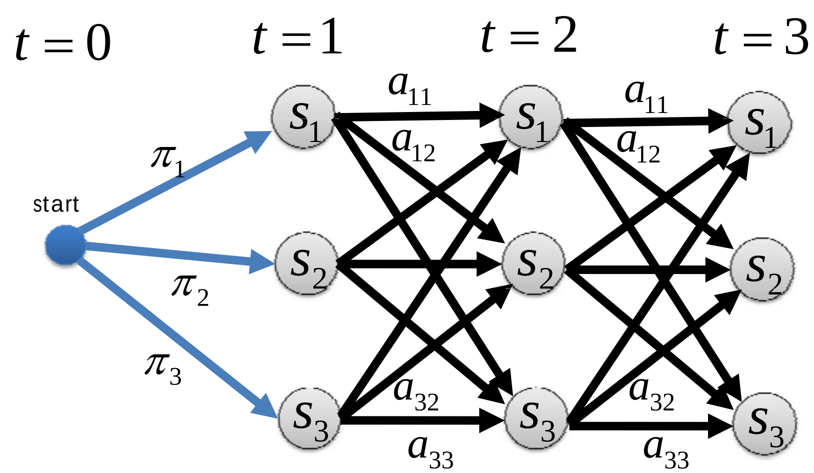
(Parameter Set)

Give an observation at given time and model, how to compute the probability of the observation

Observation at Time t = $\mathbf{x}^{(t)}$

Parameters in Model $\lambda = \{\{a_{ij}\}, \{b_{ij}\}, \{\pi_i\}\}$

Probability $P(\mathbf{x}^{(t)} | \lambda) = ?$



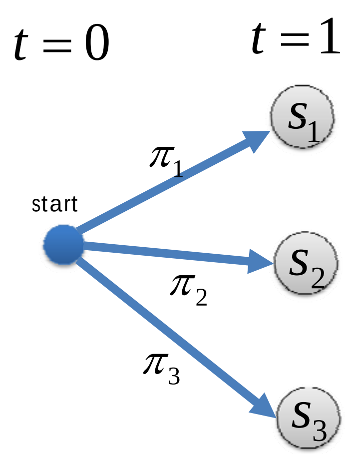
Initial State $t=1$

$$\boxed{P(s_i^{(1)}) = \pi_i}$$

Other State $t>1$

$$\boxed{P(s_k^{(t)}) = \sum_{i=1}^{L} a_{ik} P(s_i^{(t-1)})}$$

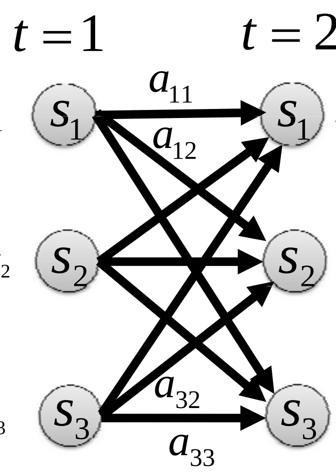
L :The number of all possible states



$$P(s_1^{(1)}) = \pi_1$$

$$P(s_2^{(1)}) = \pi_2$$

$$P(s_3^{(1)}) = \pi_3$$



$P(s_1^{(1)}) = \pi_1$

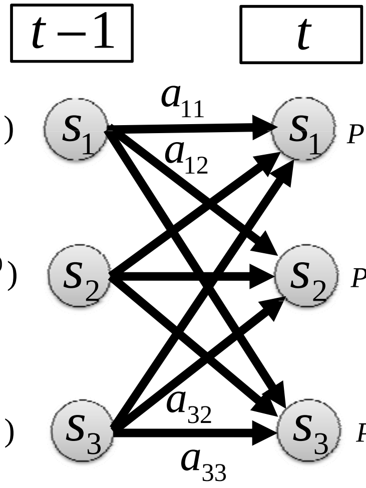
$P(s_2^{(1)}) = \pi_2$

$P(s_3^{(1)}) = \pi_3$

$$P(s_1^{(2)}) = a_{11}P(s_1^{(1)}) + a_{21}P(s_2^{(1)}) + a_{31}P(s_3^{(1)}) = a_{11}\pi_1 + a_{21}\pi_2 + a_{31}\pi_3$$

$$P(s_2^{(2)}) = a_{12}P(s_1^{(1)}) + a_{22}P(s_2^{(1)}) + a_{32}P(s_3^{(1)}) = a_{12}\pi_1 + a_{22}\pi_2 + a_{32}\pi_3$$

$$P(s_3^{(2)}) = a_{13}P(s_1^{(1)}) + a_{23}P(s_2^{(1)}) + a_{33}P(s_3^{(1)}) = a_{13}\pi_1 + a_{23}\pi_2 + a_{33}\pi_3$$



$P(s_1^{(t-1)})$

$P(s_2^{(t-1)})$

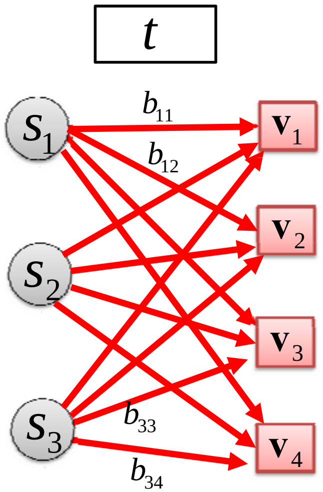
$P(s_3^{(t-1)})$

$$P(s_1^{(t)}) = a_{11}P(s_1^{(t-1)}) + a_{21}P(s_2^{(t-1)}) + a_{31}P(s_3^{(t-1)})$$

$$P(s_2^{(t)}) = a_{12}P(s_1^{(t-1)}) + a_{22}P(s_2^{(t-1)}) + a_{32}P(s_3^{(t-1)})$$

$$P(s_3^{(t)}) = a_{13}P(s_1^{(t-1)}) + a_{23}P(s_2^{(t-1)}) + a_{33}P(s_3^{(t-1)})$$

$$\boxed{\left[P(s_1^{(t)})\ \ P(s_2^{(t)})\ \ P(s_3^{(t)})\right] = \left[P(s_1^{(t-1)})\ \ P(s_2^{(t-1)})\ \ P(s_3^{(t-1)})\right]\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{3} \end{bmatrix}}$$



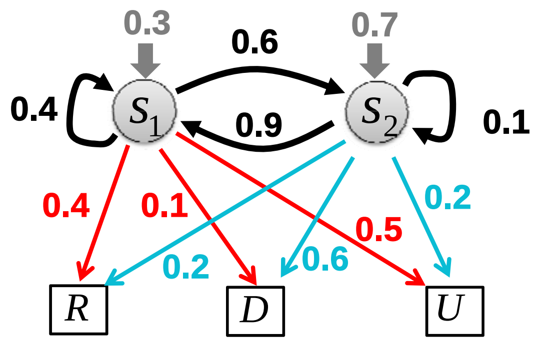
Prob. of observation at Time t

$$\boxed{P(\mathbf{x}^{(t)} = \mathbf{v}_k) = \sum_{i=1}^{L} b_{ik} P(s_i^{(t)})}$$

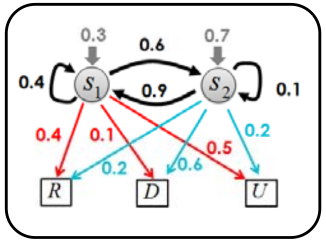
$$P(\mathbf{x}^{(t)} = \mathbf{v}_k) = \left[P(s_1^{(t)})\ \ P(s_2^{(t)})\ \ P(s_3^{(t)})\right]\begin{bmatrix} b_{1k} \\ b_{2k} \\ b_{3k} \end{bmatrix} = \left[P(s_1^{(t-1)})\ \ P(s_2^{(t-1)})\ \ P(s_3^{(t-1)})\right]\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{3} \end{bmatrix}\begin{bmatrix} b_{1k} \\ b_{2k} \\ b_{3k} \end{bmatrix}$$

$$= [\pi_1\ \ \pi_2\ \ \pi_3]\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{t-1}\begin{bmatrix} b_{1k} \\ b_{2k} \\ b_{3k} \end{bmatrix}$$

Example



$$P(x^{(3)} = U) = ?$$



$$A=\begin{bmatrix}0.4 & 0.6\\ 0.9 & 0.1\end{bmatrix}$$

$$B=\begin{array}{c}\begin{array}{ccc}R & D & U\\ \Downarrow & \Downarrow & \Downarrow\end{array}\\ \begin{bmatrix}0.4 & 0.1 & 0.5\\ 0.2 & 0.6 & 0.2\end{bmatrix}\end{array}$$

$t=1$

$$\left[P(s_1^{(1)})\ \ P(s_2^{(1)})\right]=\left[\pi_1\ \ \pi_2\right]=\left[0.3\ \ 0.7\right]$$

$t=2$

$$\left[P(s_1^{(2)})\ \ P(s_2^{(2)})\right]=\left[0.3\ \ 0.7\right]\begin{bmatrix}0.4 & 0.6\\ 0.9 & 0.1\end{bmatrix}=\left[0.75\ \ 0.25\right]$$

$t=3$

$$\left[P(s_1^{(3)})\ \ P(s_2^{(3)})\right]=\left[0.75\ \ 0.25\right]\begin{bmatrix}0.4 & 0.6\\ 0.9 & 0.1\end{bmatrix}=\left[0.525\ \ 0.475\right]$$

$$P(x^{(3)})=\left[0.525\ \ 0.475\right]\begin{bmatrix}0.4 & 0.1 & 0.5\\ 0.2 & 0.6 & 0.2\end{bmatrix}$$

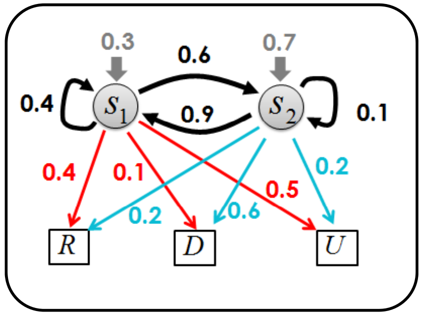
$$=\left[0.3050\ \ 0.3375\ \ 0.3575\right]$$

$$\begin{array}{ccc}\Uparrow & \Uparrow & \Uparrow\\ R & D & U\end{array}$$

$$P(x^{(3)}=U)=0.3575$$

Example

$$P(x^{(10)}=R)=?$$



$$P(x^{(10)}=R)=\left[\pi_1\ \ \pi_2\right]\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}^9\begin{bmatrix}b_{11}\\ b_{21}\end{bmatrix}$$

$$=\left[0.3\ \ 0.7\right]\begin{bmatrix}0.4 & 0.6\\ 0.9 & 0.1\end{bmatrix}^9\begin{bmatrix}0.4\\ 0.2\end{bmatrix}$$

$$=0.3201$$

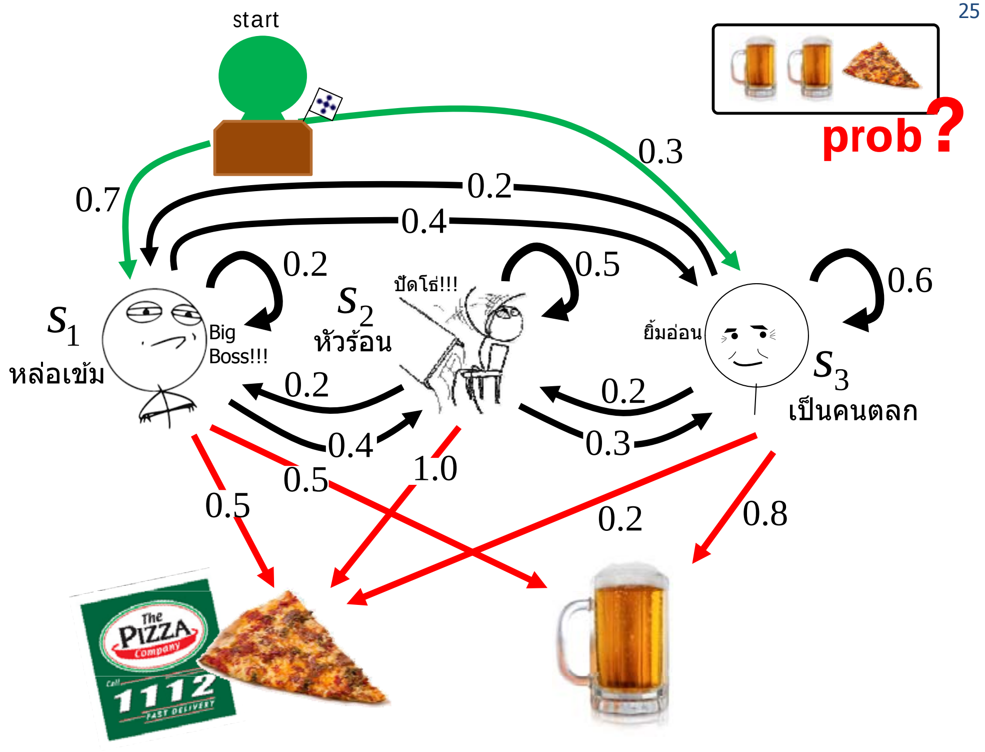
Give the observation sequence and model, how to compute the probability of observation sequence

Observation sequence $\mathbf{X}=[\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)},\ldots,\mathbf{x}^{(T)}]$

Parameters in Model $\lambda=\left\{\{a_{ij}\},\{b_{ij}\},\{\pi_i\}\right\}$

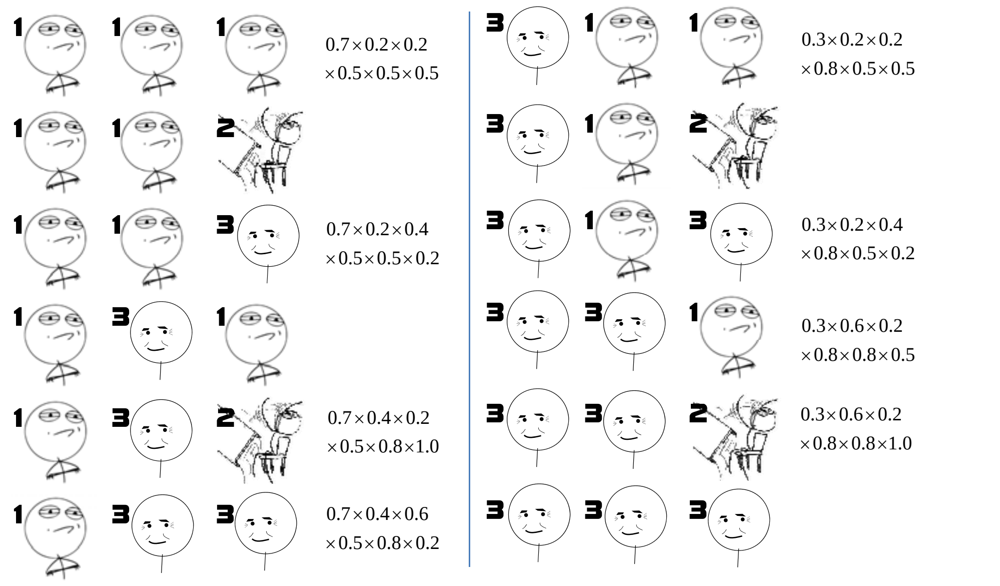
$\Downarrow$

Probability $P(\mathbf{X}|\lambda)=?$



All possible State Sequence

Prob = 0.1296



Forward Variable

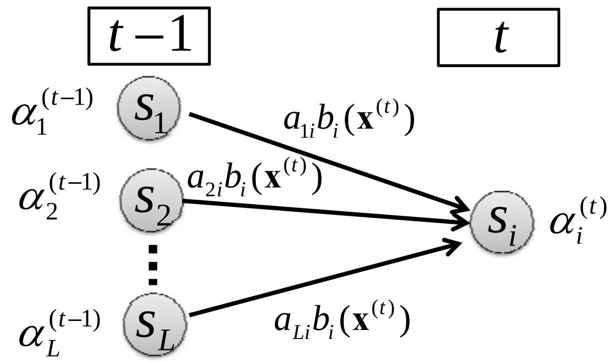
$$\alpha_i^{(t)} = P(s_i^{(t)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}\cdots\mathbf{x}^{(t)} \mid \lambda)$$

$$\alpha_i^{(1)} = P(s_i^{(1)}, \mathbf{x}^{(1)} \mid \lambda) = \pi_i b_i(\mathbf{x}^{(1)})$$

$$\alpha_i^{(2)} = P(s_i^{(2)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)} \mid \lambda)$$

$$= \alpha_1^{(1)} a_{1i} b_i(\mathbf{x}^{(2)}) + \alpha_2^{(1)} a_{2i} b_i(\mathbf{x}^{(2)}) + \cdots + \alpha_L^{(1)} a_{Li} b_i(\mathbf{x}^{(2)})$$

$$= \left(\sum_{k=1}^{L} \alpha_k^{(1)} a_{ki}\right) b_i(\mathbf{x}^{(2)})$$



$$\alpha_i^{(t)} = P(s_i^{(t)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}\cdots\mathbf{x}^{(t)} \mid \lambda)$$

$$= \alpha_1^{(t-1)} a_{1i} b_i(\mathbf{x}^{(t)}) + \alpha_2^{(1)} a_{2i} b_i(\mathbf{x}^{(t)}) + \cdots + \alpha_L^{(1)} a_{Li} b_i(\mathbf{x}^{(t)})$$

$$= \left(\sum_{k=1}^{L} \alpha_k^{(t-1)} a_{ki}\right) b_i(\mathbf{x}^{(t)})$$

Forward Variable

$$\alpha_i^{(t)} = P(s_i^{(t)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}\cdots\mathbf{x}^{(t)} \mid \lambda)$$

$$\alpha_i^{(1)} = \pi_i b_i(\mathbf{x}^{(1)})$$

$$\alpha_i^{(t)} = \left(\sum_{k=1}^{L} \alpha_k^{(t-1)} a_{ki}\right) b_i(\mathbf{x}^{(t)})$$

Backward Variable

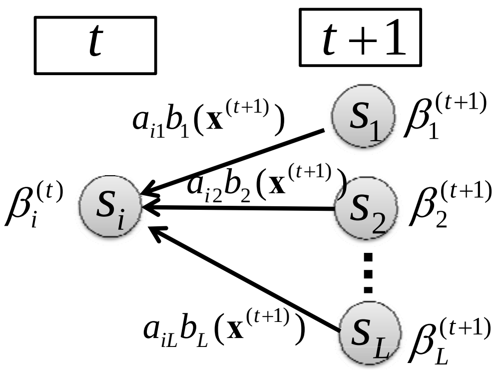
$$\beta_i^{(t)} = P(s_i^{(t)}, \mathbf{x}^{(t+1)}\mathbf{x}^{(t+2)}\cdots\mathbf{x}^{(T)} \mid \lambda)$$

$$\beta_i^{(T)} = P(s_i^{(T)}, \phi \mid \lambda) = 1$$

$$\begin{aligned}\beta_i^{(T-1)} &= P(s_i^{(T-1)}, \mathbf{x}^{(T)} \mid \lambda) \\ &= a_{i1}b_1(\mathbf{x}^{(T)}) + a_{i2}b_2(\mathbf{x}^{(T)}) + \cdots + a_{iL}b_L(\mathbf{x}^{(T)}) \\ &= \sum_{k=1}^{L} a_{ik} b_k(\mathbf{x}^{(T)})\end{aligned}$$

$$\begin{aligned}\beta_i^{(T-2)} &= P(s_i^{(T-2)}, \mathbf{x}^{(T-1)}\mathbf{x}^{(T)} \mid \lambda) \\ &= a_{i1}b_1(\mathbf{x}^{(T-1)})\beta_1^{(T-1)} + a_{i2}b_2(\mathbf{x}^{(T-1)})\beta_2^{(T-1)} + \cdots + a_{iL}b_L(\mathbf{x}^{(T-1)})\beta_L^{(T-1)} \\ &= \sum_{k=1}^{L} a_{ik} b_k(\mathbf{x}^{(T-1)})\beta_k^{(T-1)}\end{aligned}$$

$$\begin{aligned}\beta_i^{(t)} &= P(s_i^{(t)}, \mathbf{x}^{(t+1)}\mathbf{x}^{(t+2)}\cdots\mathbf{x}^{(T)} \mid \lambda) \\ &= \sum_{k=1}^{L} a_{ik} b_k(\mathbf{x}^{(t+1)})\beta_k^{(t+1)}\end{aligned}$$



Backward Variable

$$\beta_i^{(t)} = P(s_i^{(t)}, \mathbf{x}^{(t+1)}\mathbf{x}^{(t+2)}\cdots\mathbf{x}^{(T)} \mid \lambda)$$

$$\beta_i^{(T)} = 1$$

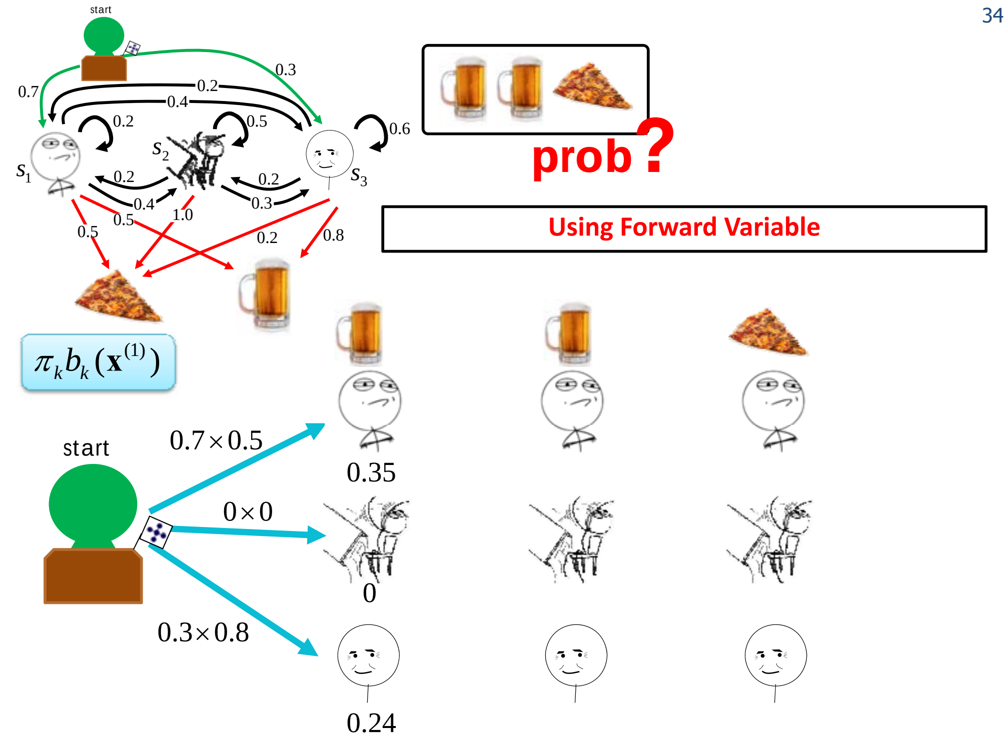
$$\beta_i^{(t)} = \sum_{k=1}^{L} a_{ik} b_k(\mathbf{x}^{(t+1)})\beta_k^{(t+1)}$$

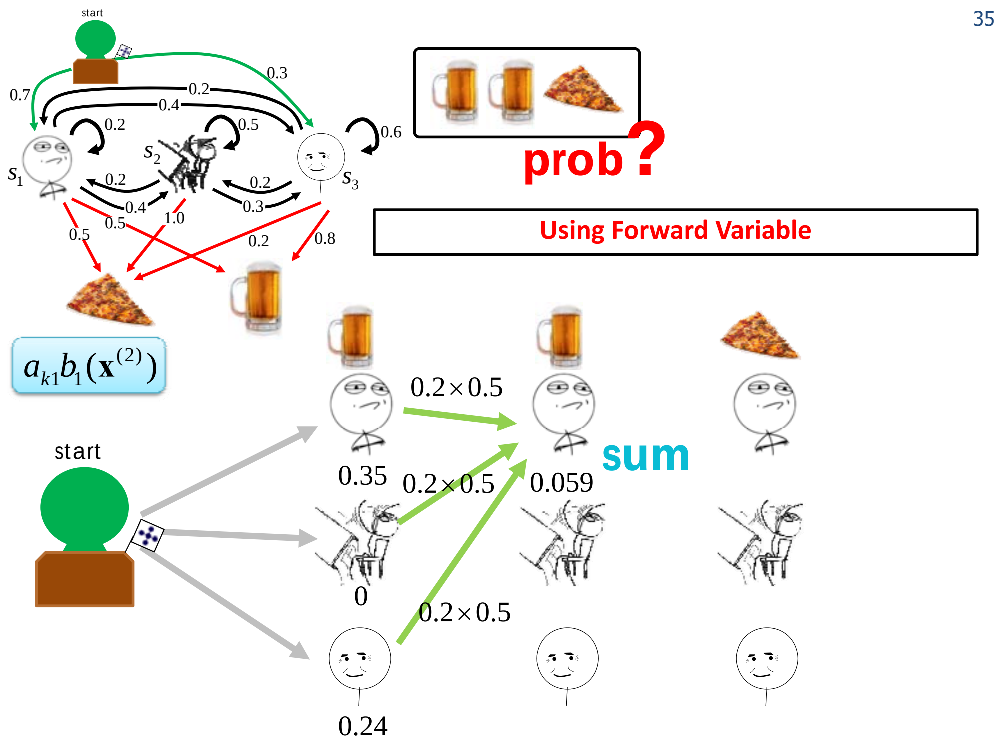
Forward-Backward Procedure

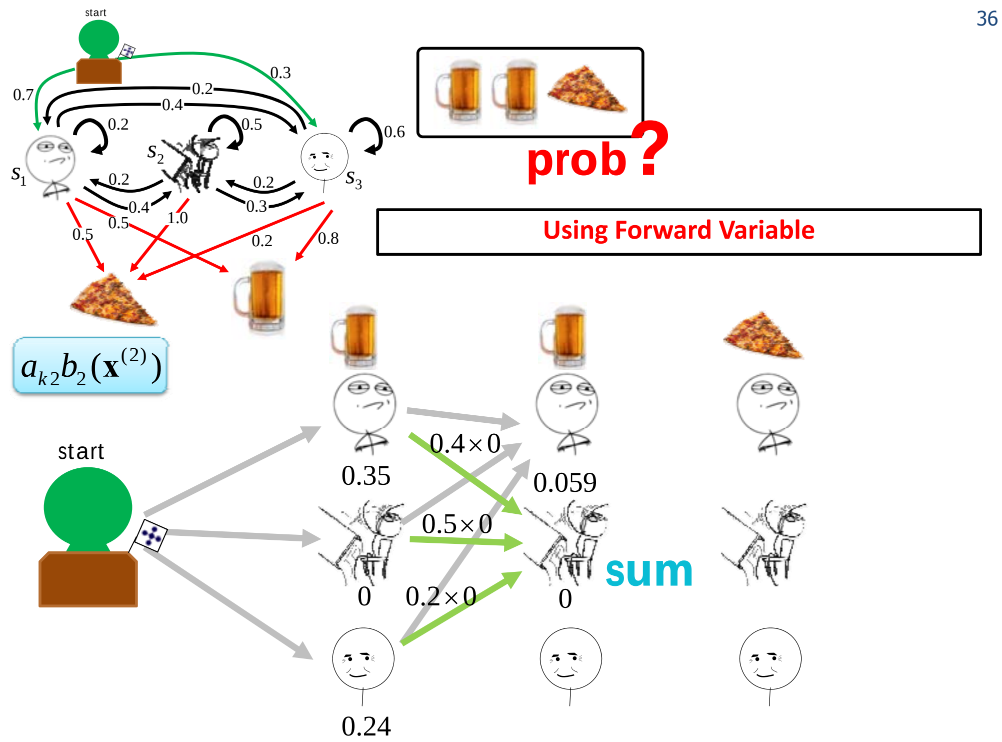
$$\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \ldots, \mathbf{x}^{(T)}]$$

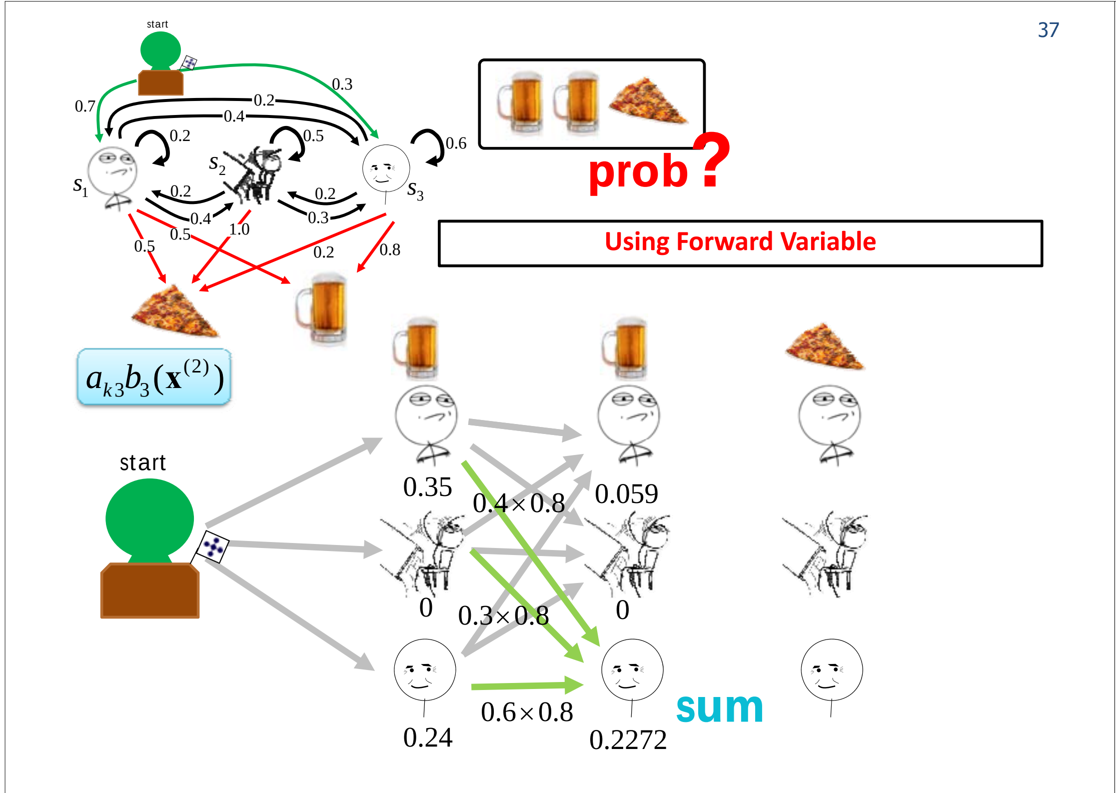
$$P(\mathbf{X}|\lambda) = \sum_{k=1}^{L} \alpha_k^{(T)}$$ [Using forward variable]

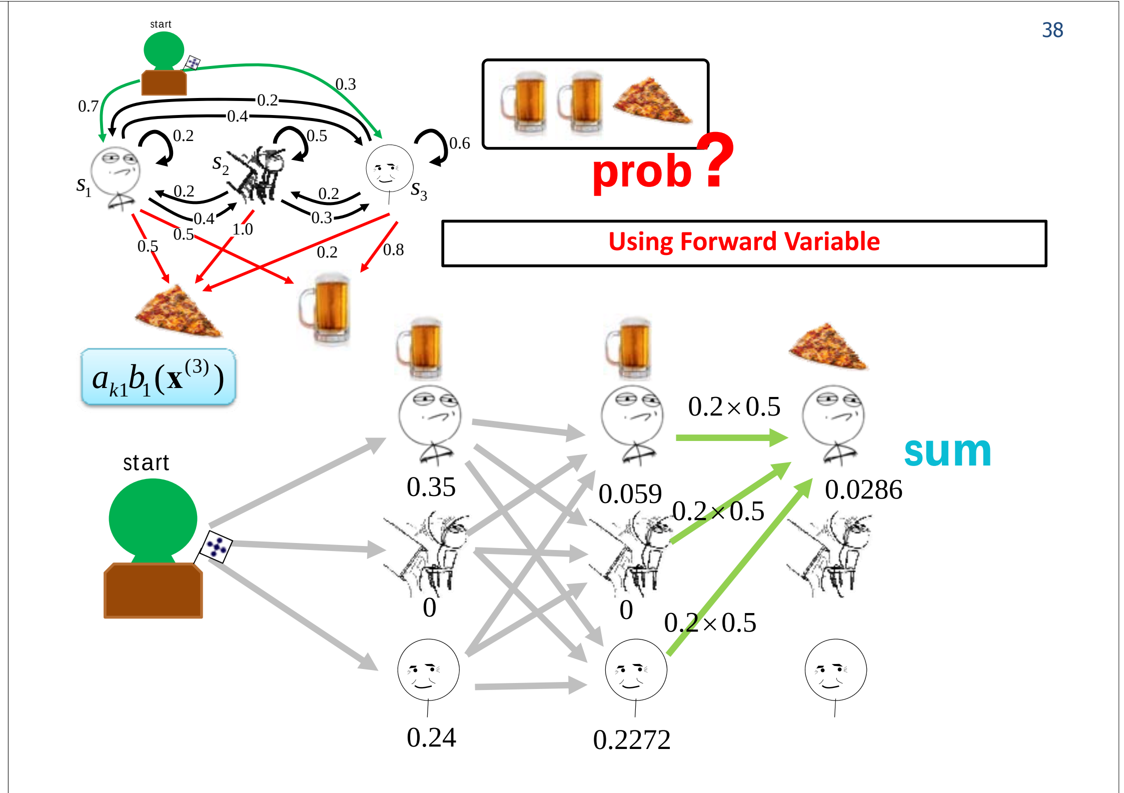
$$P(\mathbf{X}|\lambda) = \sum_{k=1}^{L} \pi_k b_k(\mathbf{x}^{(1)}) \beta_k^{(1)}$$ [Using backward variable]

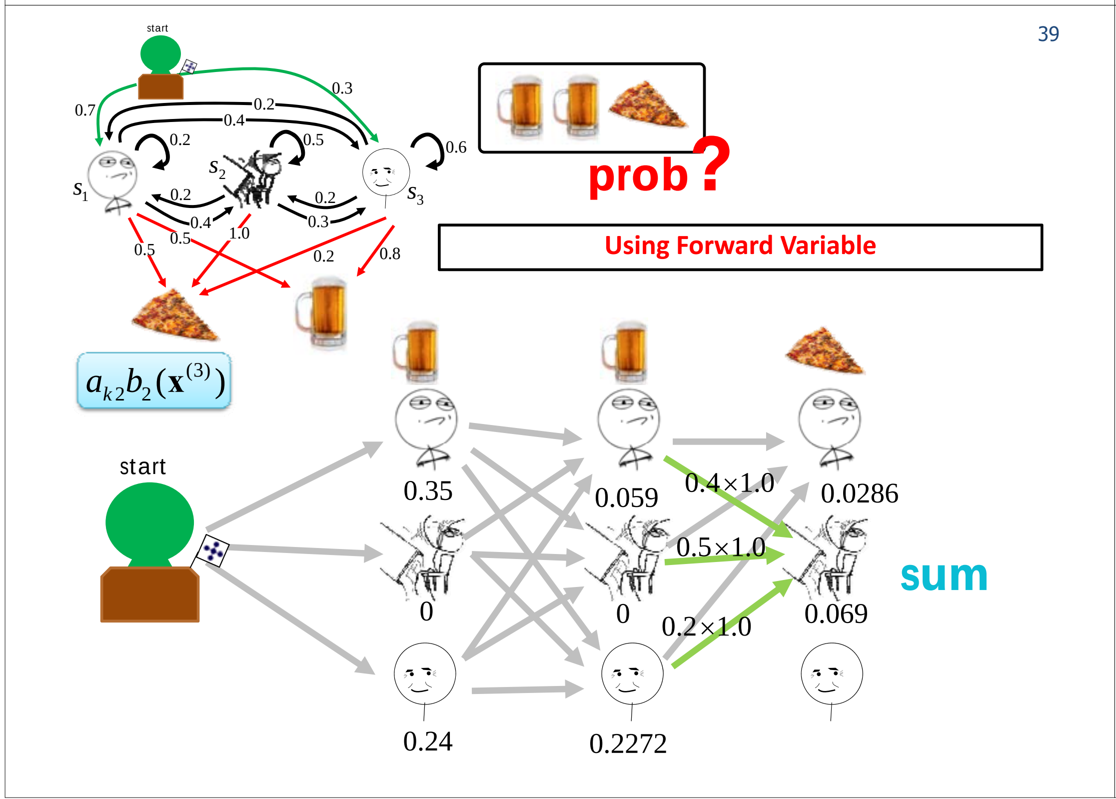


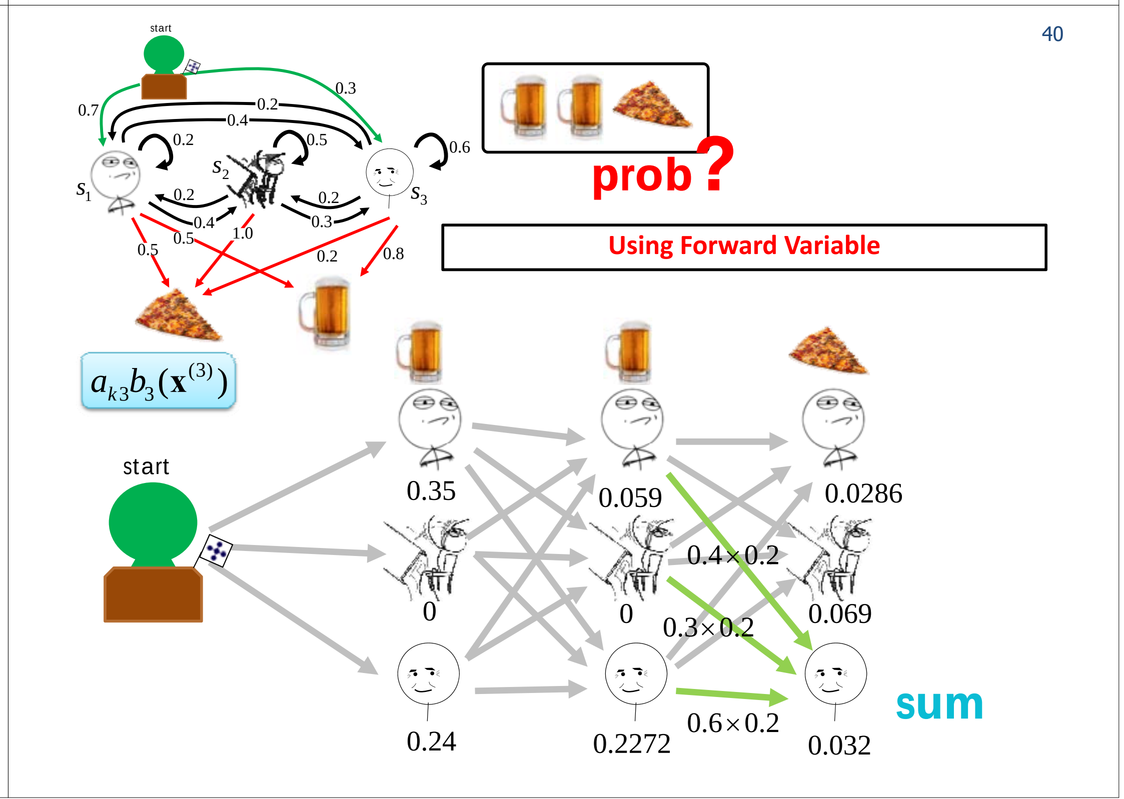


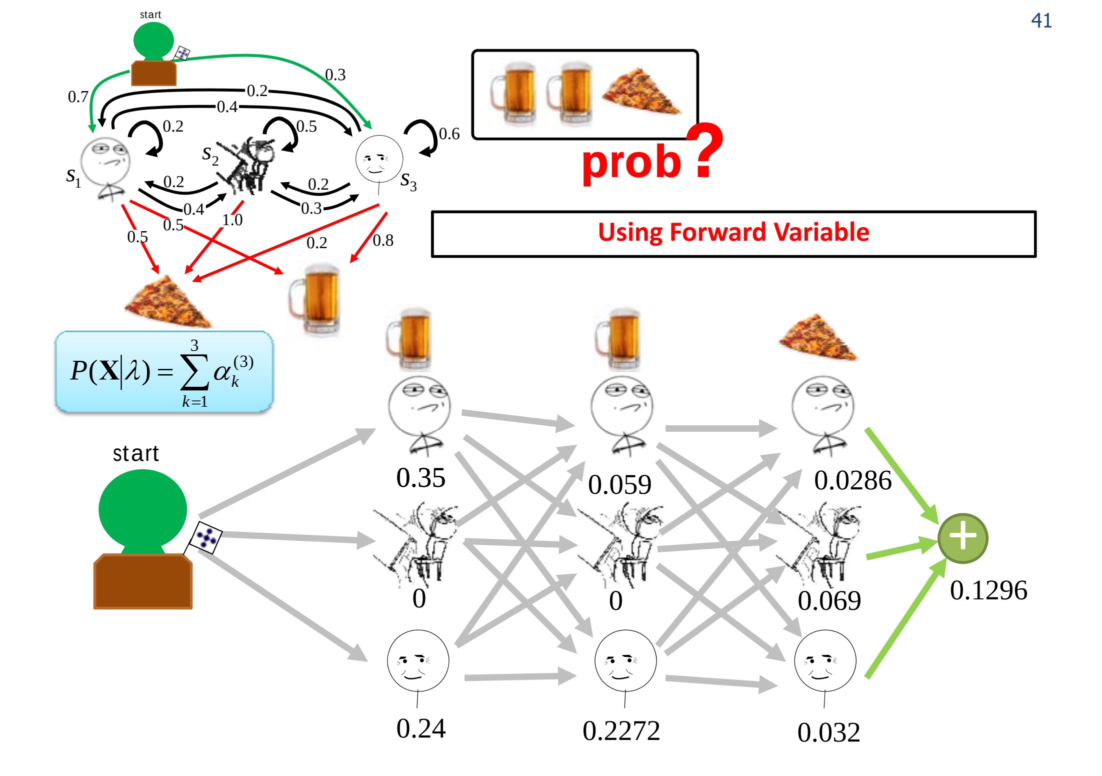


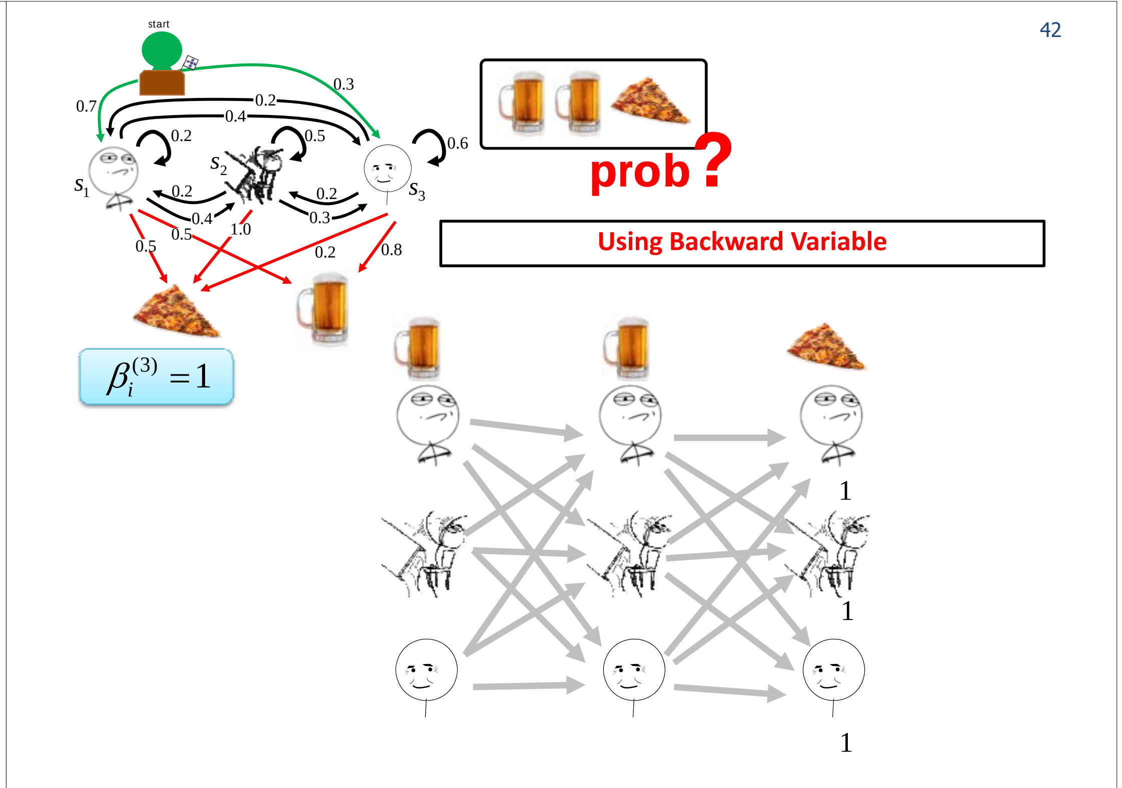

37
start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Forward Variable
$a_{k3}b_3(\mathbf{x}^{(2)})$
start
0.35
0
0.24
0.4×0.8
0.3×0.8
0.6×0.8
0.059
0
0.2272
sum

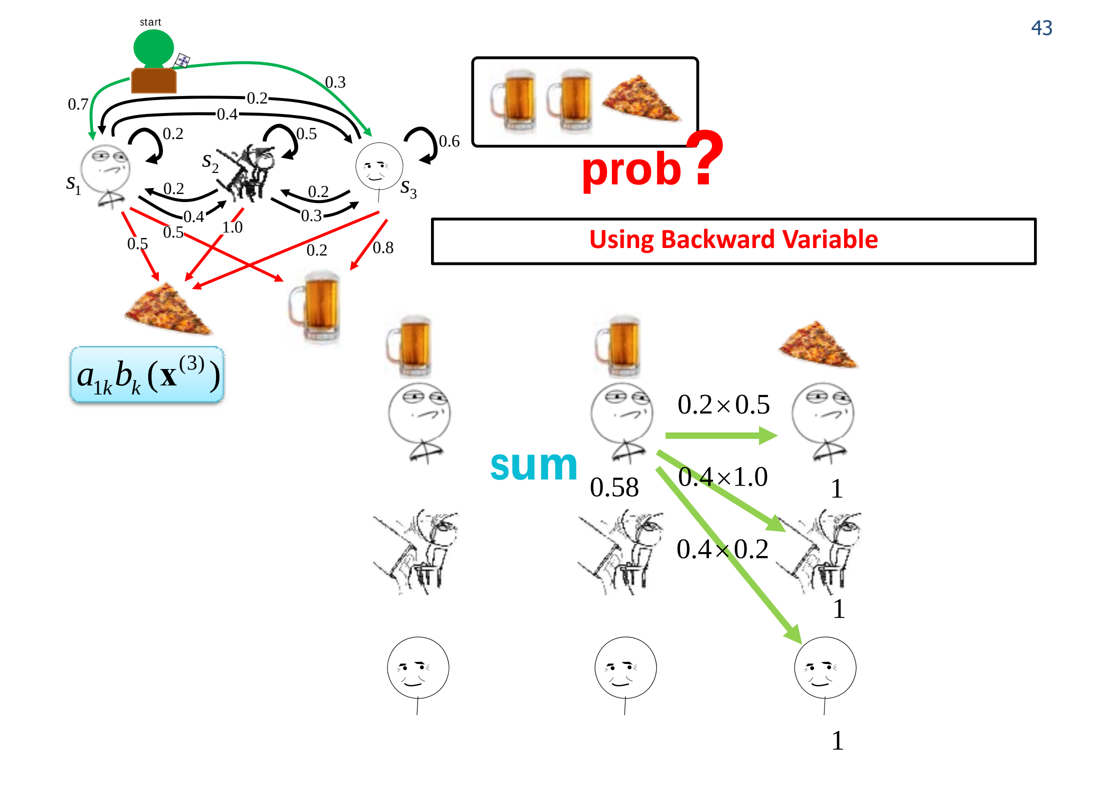

38
start
prob ?
Using Forward Variable
$a_{k1}b_1(\mathbf{x}^{(3)})$
start
0.35
0
0.24
0.059
0
0.2272
0.2×0.5
0.2×0.5
0.2×0.5
0.0286
sum

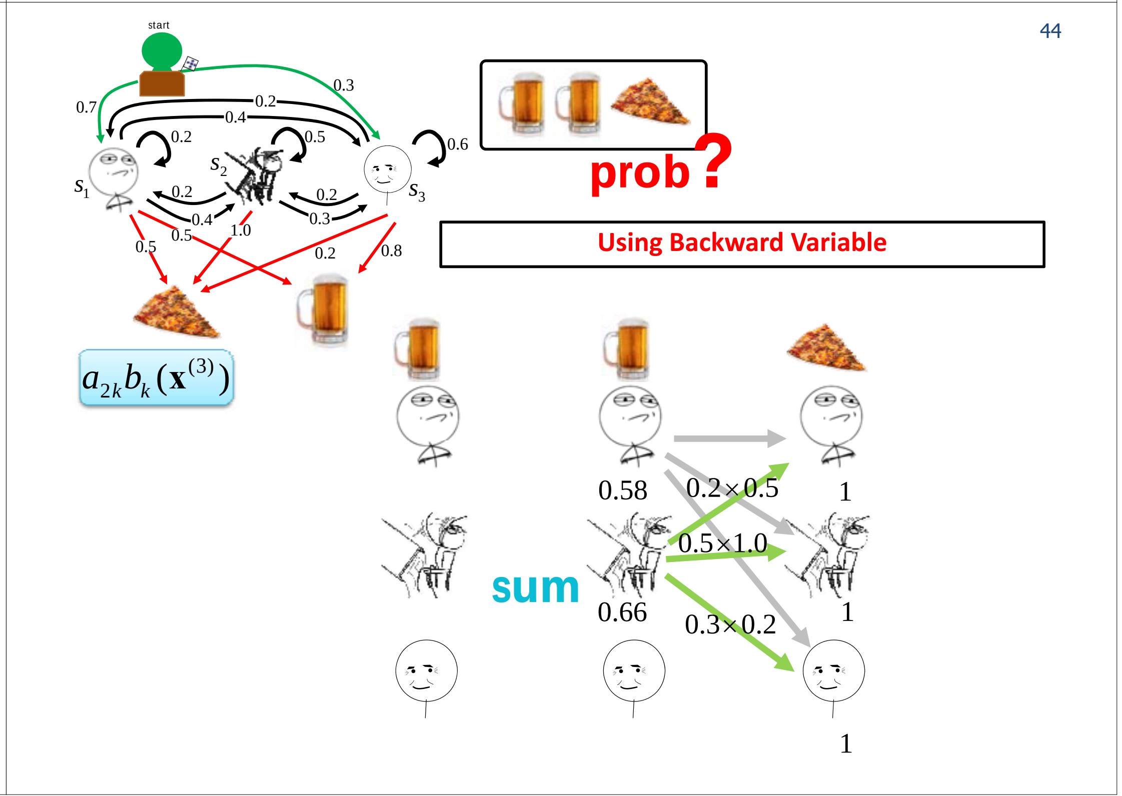

39
start
prob ?
Using Forward Variable
$a_{k2}b_2(\mathbf{x}^{(3)})$
start
0.35
0
0.24
0.059
0
0.2272
0.4×1.0
0.5×1.0
0.2×1.0
0.0286
0.069
sum

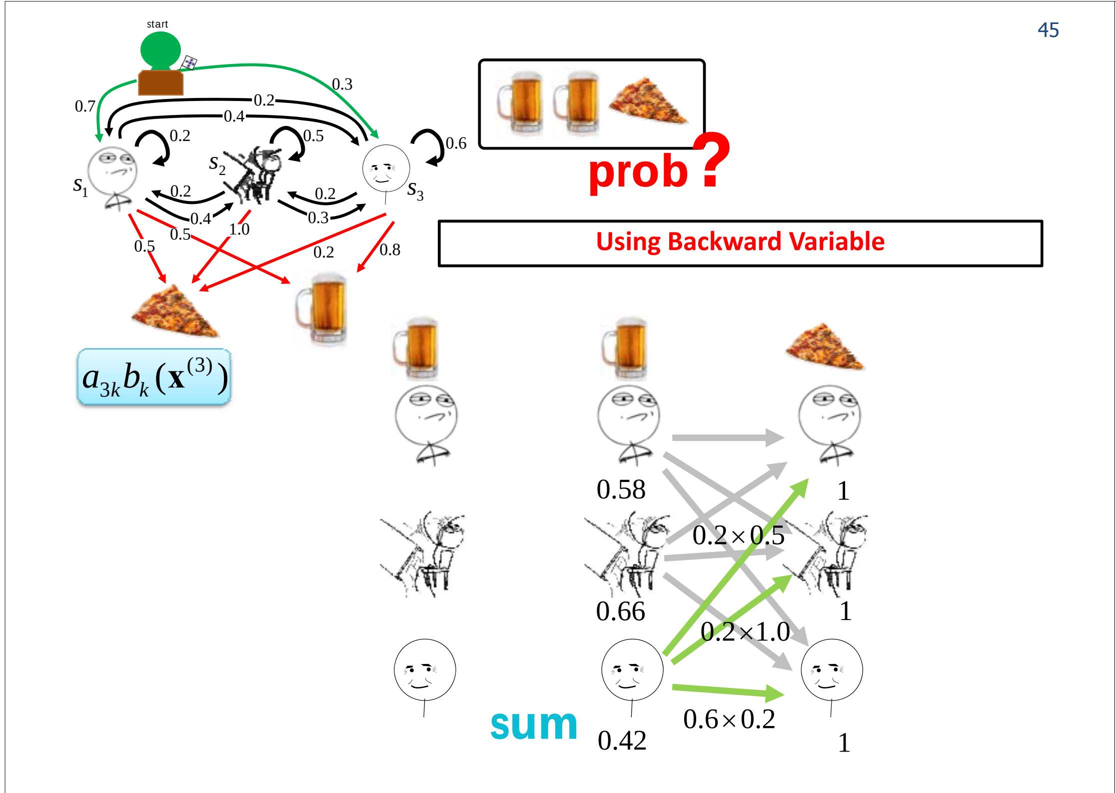

40
start
prob ?
Using Forward Variable
$a_{k3}b_3(\mathbf{x}^{(3)})$
start
0.35
0
0.24
0.059
0
0.2272
0.0286
0.4×0.2
0.069
0.3×0.2
0.6×0.2
0.032
sum

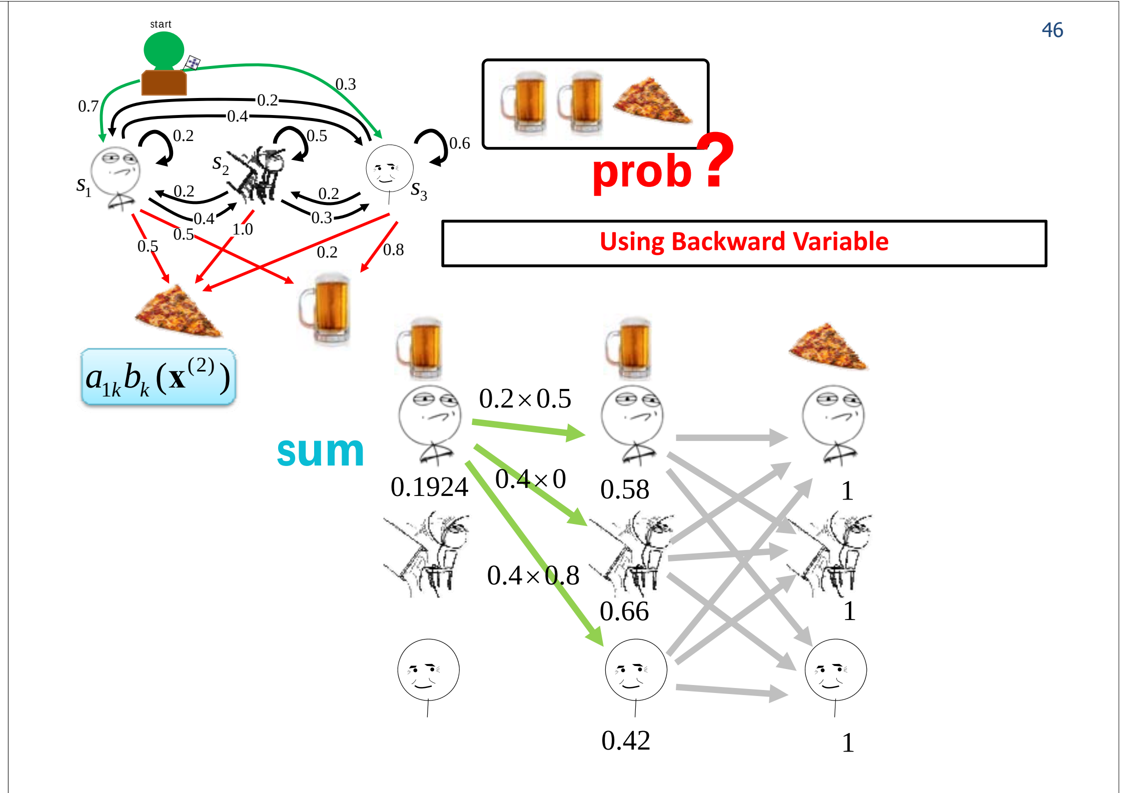

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Forward Variable
$P(\mathbf{X}|\lambda) = \sum_{k=1}^{3} \alpha_k^{(3)}$
start
0.35
0
0.24
0.059
0
0.2272
0.0286
0.069
0.032
0.1296

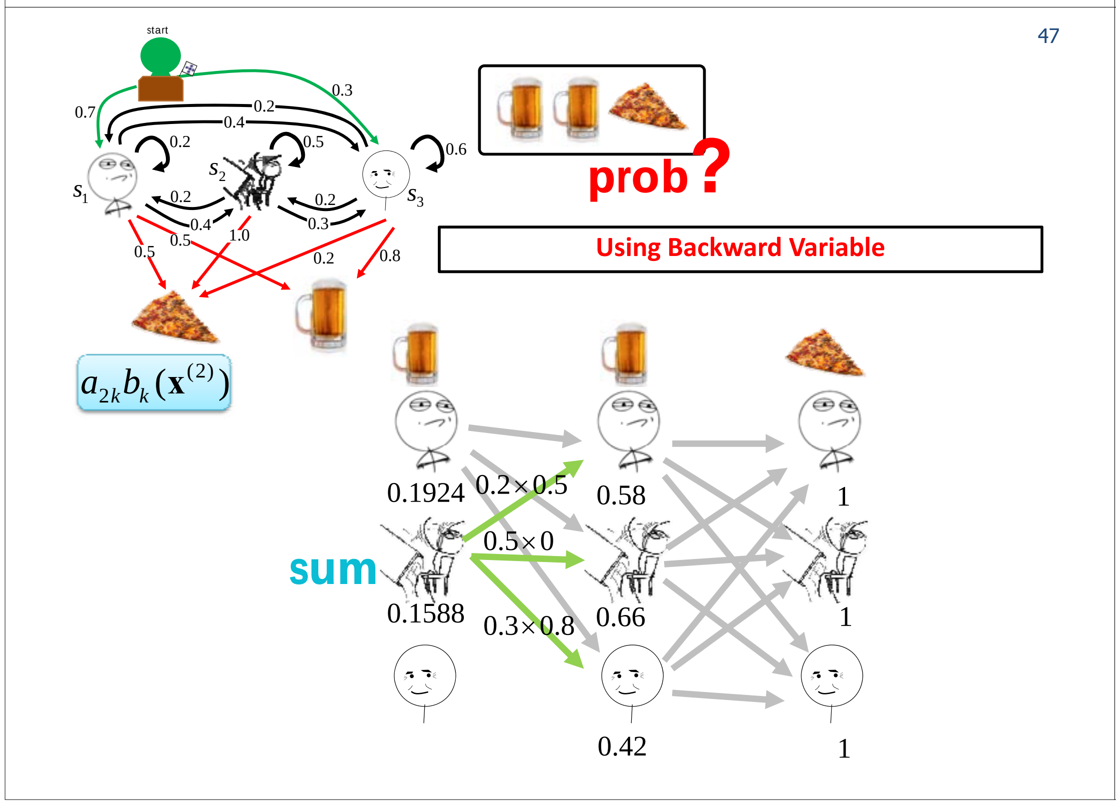

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Backward Variable
$\beta_i^{(3)} = 1$
1
1
1

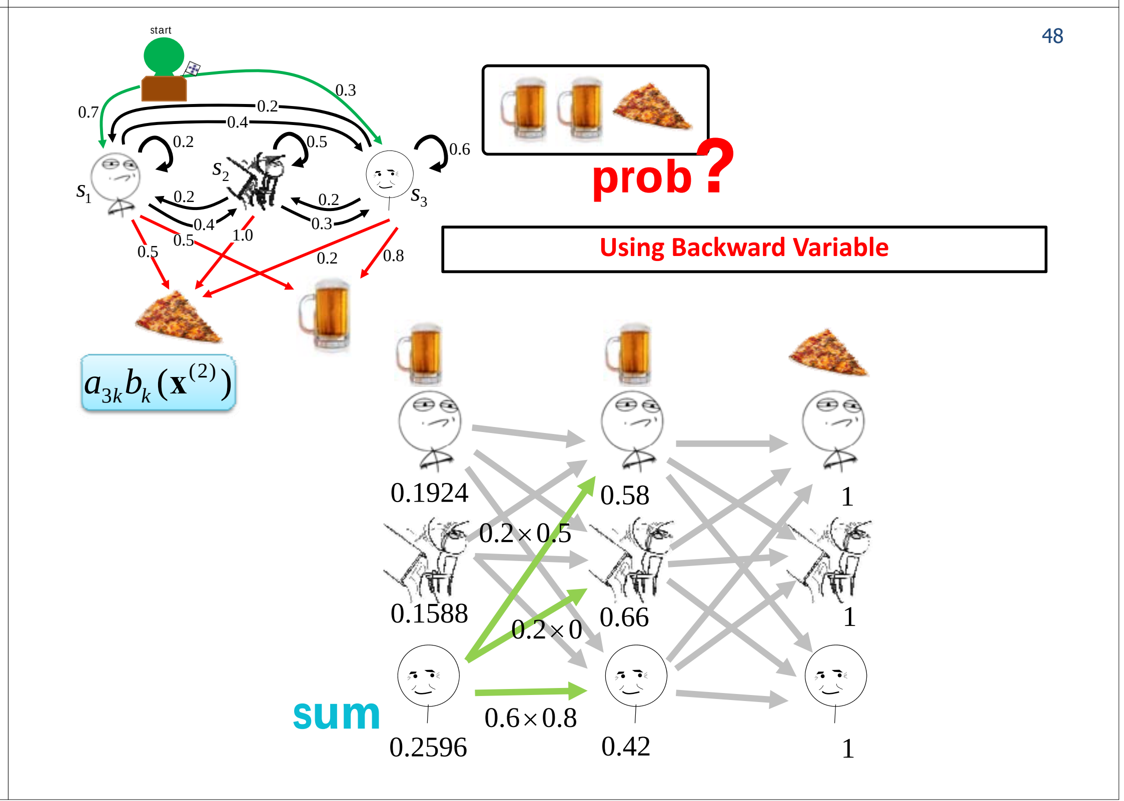

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Backward Variable
$a_{1k} b_k(\mathbf{x}^{(3)})$
sum
0.58
0.2×0.5
0.4×1.0
0.4×0.2
1
1
1


start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Backward Variable
$a_{2k} b_k(\mathbf{x}^{(3)})$
0.58
sum
0.66
0.2×0.5
0.5×1.0
0.3×0.2
1
1
1

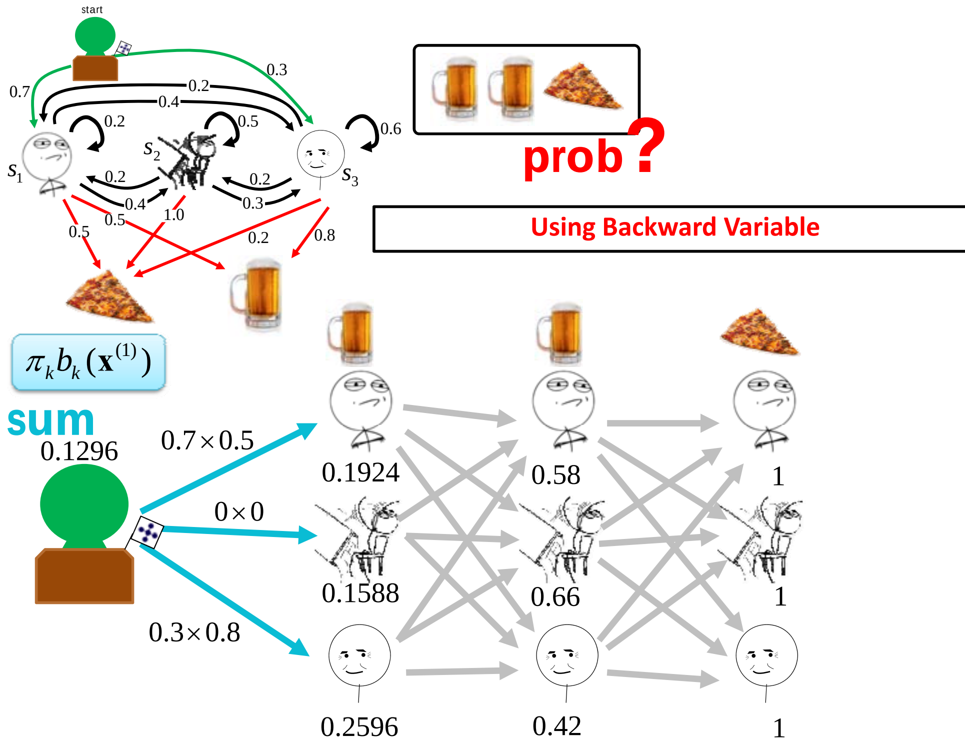

45
start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
prob ?
Using Backward Variable
$a_{3k} b_k(\mathbf{x}^{(3)})$
0.58
1
0.2×0.5
0.66
1
0.2×1.0
sum
0.42
0.6×0.2
1

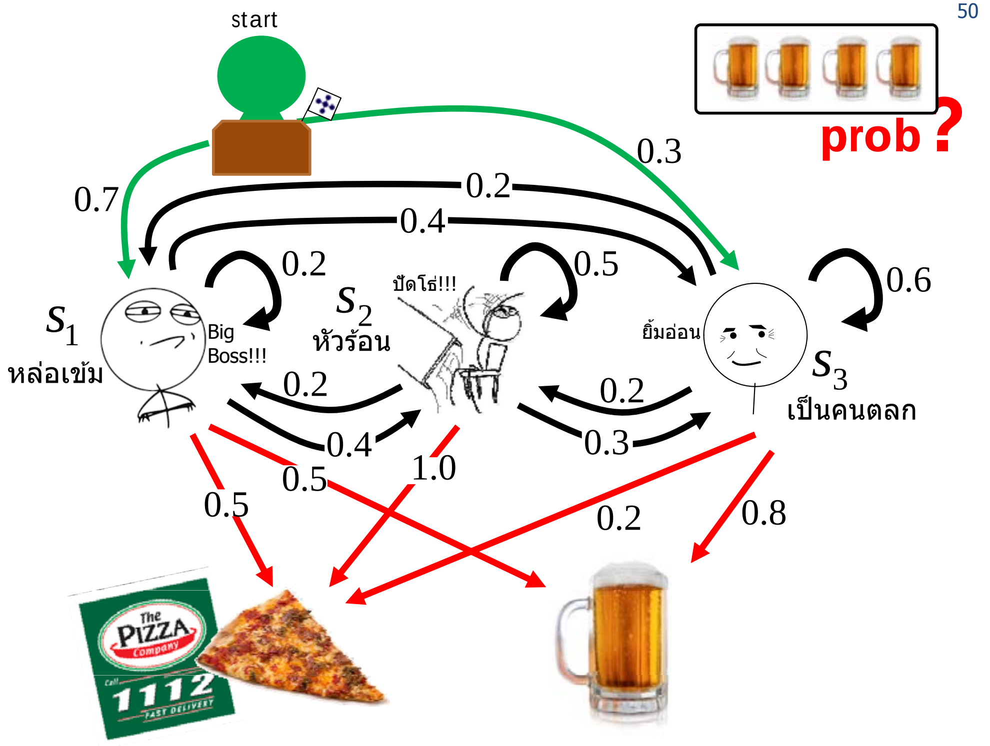

46
start
prob ?
Using Backward Variable
$a_{1k} b_k(\mathbf{x}^{(2)})$
0.2×0.5
sum
0.1924
0.4×0
0.58
1
0.4×0.8
0.66
1
0.42
1


47
start
prob ?
Using Backward Variable
$a_{2k} b_k(\mathbf{x}^{(2)})$
0.1924
0.2×0.5
0.58
1
0.5×0
sum
0.1588
0.3×0.8
0.66
1
0.42
1

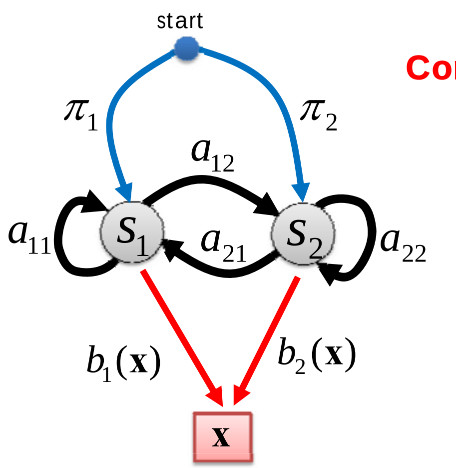

48
start
prob ?
Using Backward Variable
$a_{3k} b_k(\mathbf{x}^{(2)})$
0.1924
0.58
1
0.2×0.5
0.1588
0.2×0
0.66
1
sum
0.6×0.8
0.2596
0.42
1

Using Backward Variable





Hidden Markov Model



Continuous Observation Value $\mathbf{x}$

$b_i(\mathbf{x})$

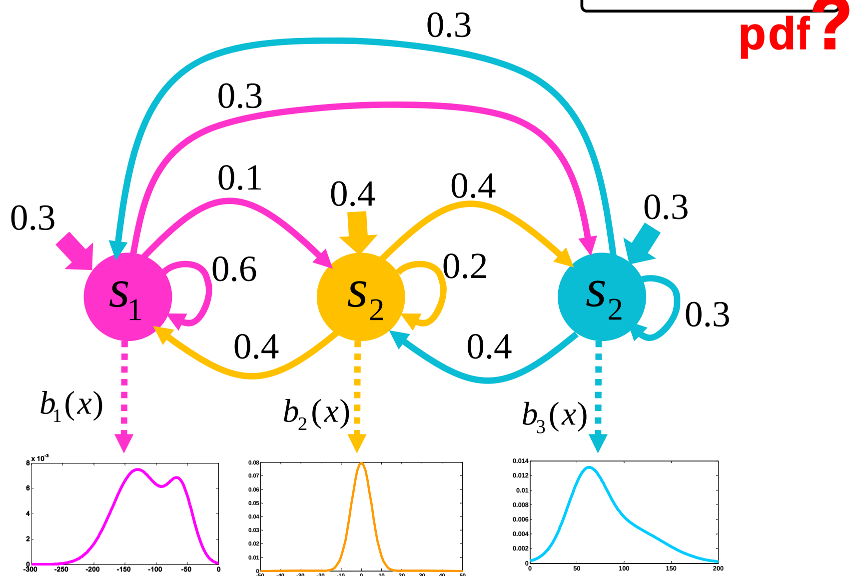
⬇

Emission pdf
[Probability Density Function]
of vector $\mathbf{x}$ from state i

Ex: Currency Exchange Rate

$\mathbf{X} = [5 \ \ 10 \ \ -50]$

pdf?



$$b_1(x) = \frac{0.25}{\sqrt{800\pi}} e^{-\frac{(x+60)^2}{800}} + \frac{0.75}{\sqrt{3200\pi}} e^{-\frac{(x+130)^2}{3200}} \quad b_2(x) = \frac{1}{\sqrt{50\pi}} e^{-\frac{x^2}{50}} \quad b_3(x) = \frac{0.5}{\sqrt{800\pi}} e^{-\frac{(x-60)^2}{800}} + \frac{0.5}{\sqrt{3200\pi}} e^{-\frac{(x-100)^2}{3200}}$$

Give the observation sequence and model, how to choose a corresponding optimal state sequence

Observation sequence $\mathbf{X}=[\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)},\ldots,\mathbf{x}^{(T)}]$

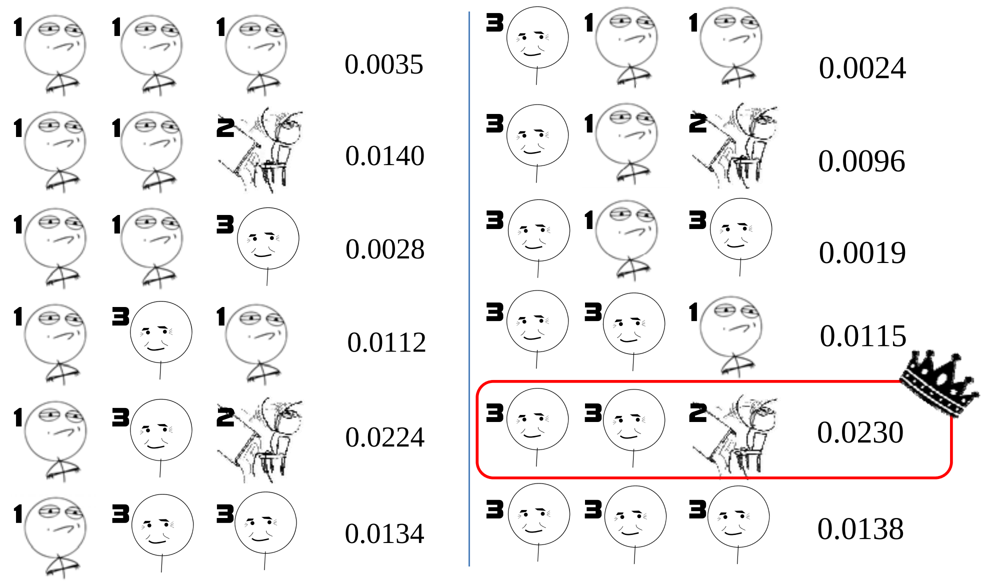
Parameters in Model $\lambda=\{\{a_{ij}\},\{b_{ij}\},\{\pi_i\}\}$

⬇

Optimal State Sequence $S_{opt}=[s^{(1)},s^{(2)},s^{(3)},\ldots,s^{(T)}]$

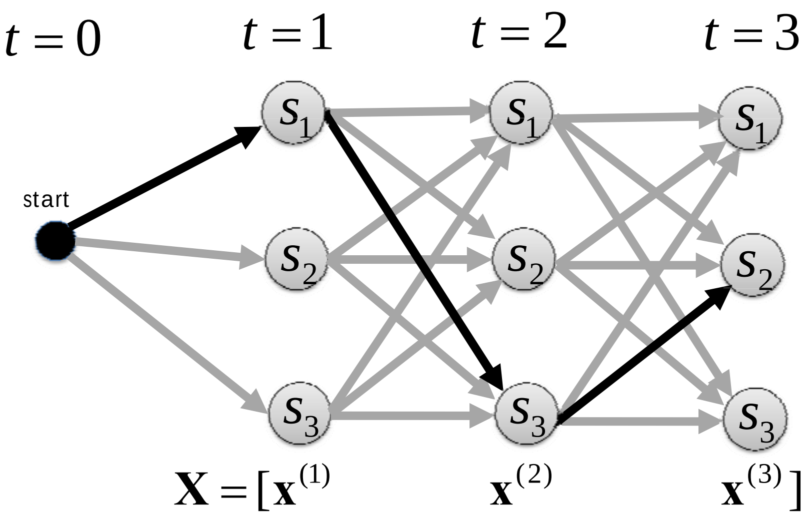
From observation

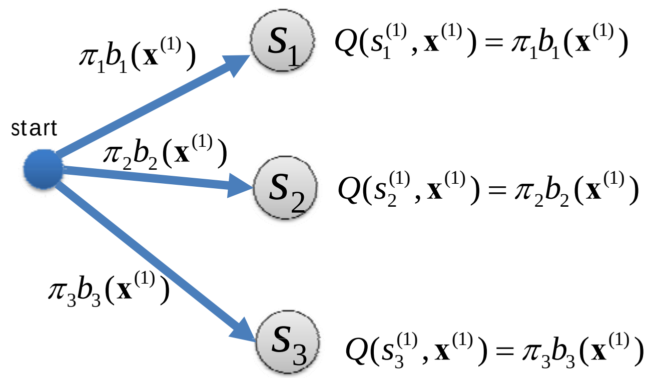
➡ **Which is the most possible state sequence?**



Viterbi Algorithm

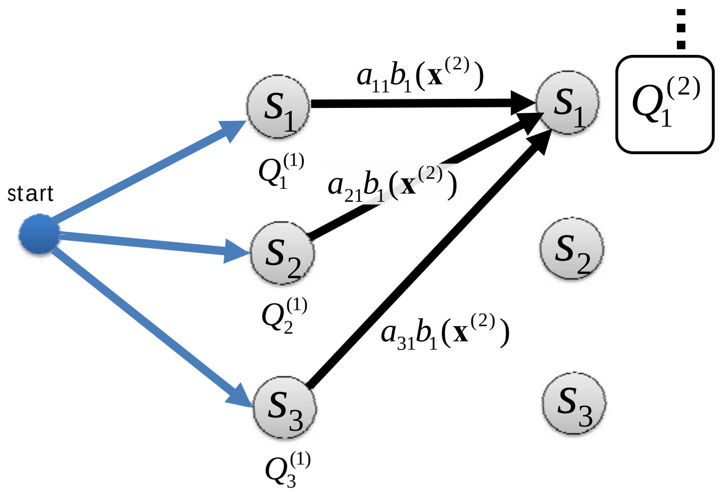
$$\mathbf{S}_{opt}=\arg\max_{\mathbf{S}} P(\mathbf{S},\mathbf{X}|\lambda)$$





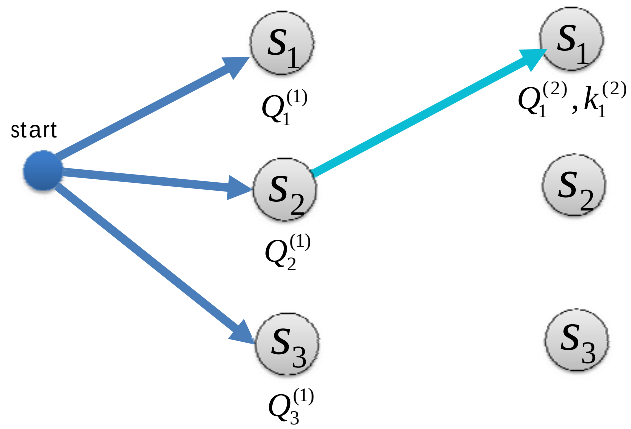
57

$$Q(s_1^{(2)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}) = \max_k a_{k1} b_1(\mathbf{x}^{(2)}) Q_k^{(1)}$$



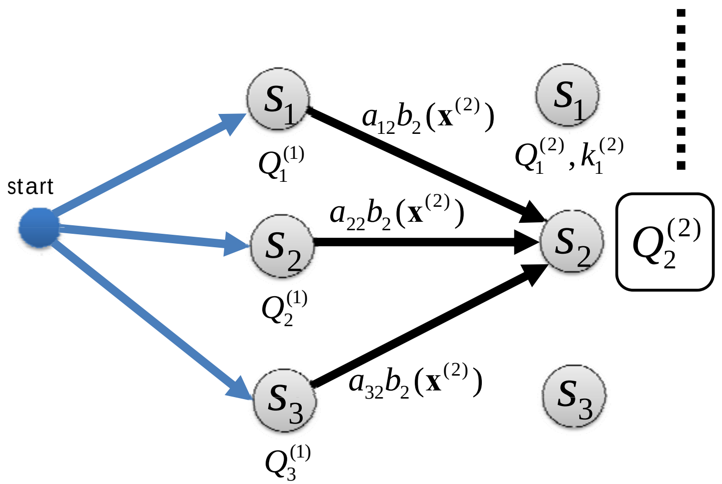
58

$$k_1^{(2)} = \arg\max_k a_{k1} b_1(\mathbf{x}^{(2)}) Q_k^{(1)}$$



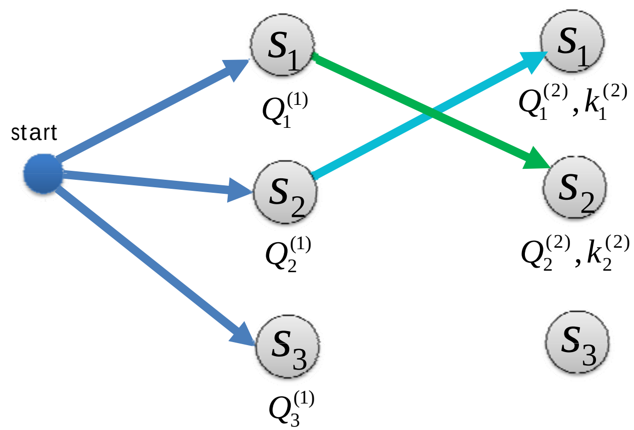
59

$$Q(s_2^{(2)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}) = \max_k a_{k2} b_2(\mathbf{x}^{(2)}) Q_k^{(1)}$$

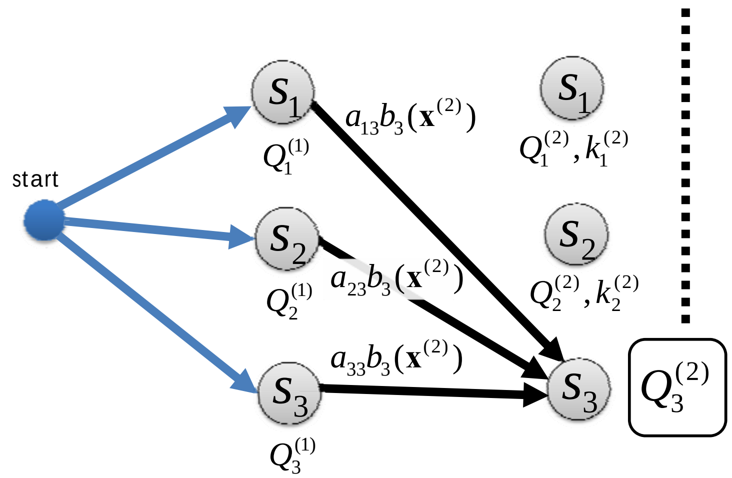


60

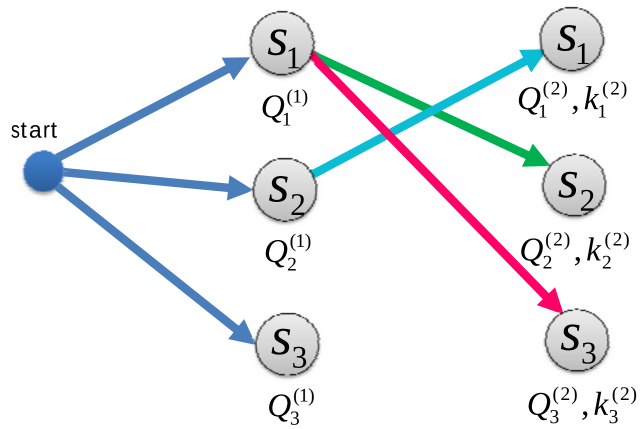
$$k_2^{(2)} = \arg\max_k a_{k2} b_2(\mathbf{x}^{(2)}) Q_k^{(1)}$$

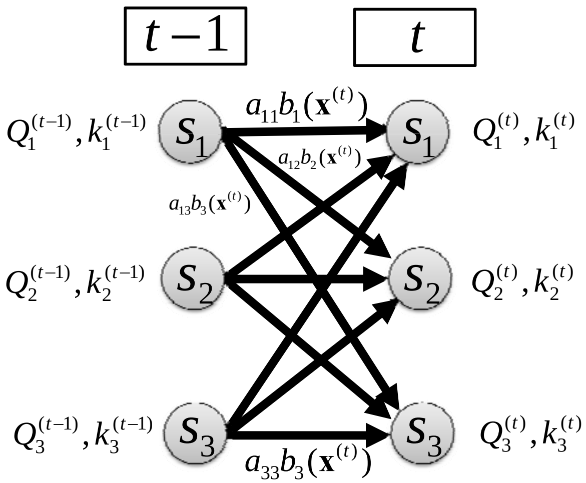


$$Q(s_3^{(2)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}) = \max_k a_{k3} b_3(\mathbf{x}^{(2)}) Q_k^{(1)}$$



$$k_3^{(2)} = \arg\max_k a_{k3} b_3(\mathbf{x}^{(2)}) Q_k^{(1)}$$





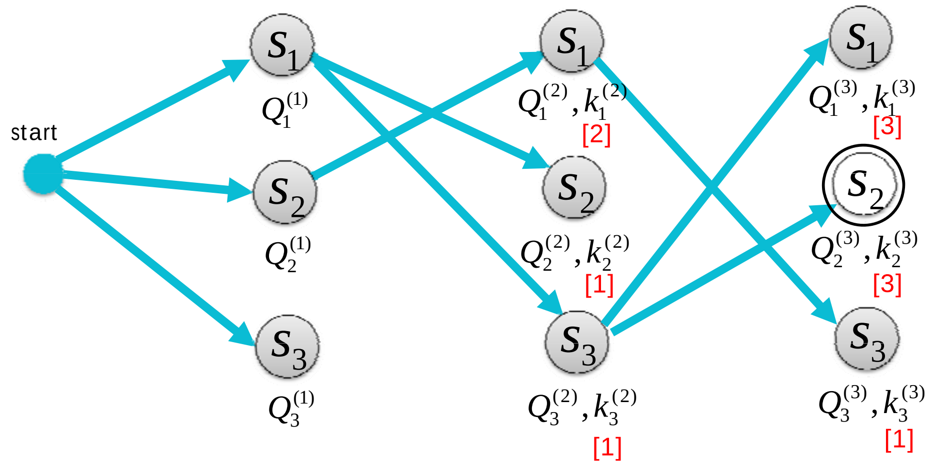
$$Q_i^{(t)} = Q(s_i^{(t)}, \mathbf{x}^{(1)}\mathbf{x}^{(2)}\ldots\mathbf{x}^{(t)}) = \max_k a_{ki} b_i(\mathbf{x}^{(t)}) Q_k^{(t-1)}$$

$$k_i^{(t)} = \arg\max_k a_{ki} b_i(\mathbf{x}^{(t)}) Q_k^{(t-1)}$$

Back Tracking

$$\mathbf{S}_{opt} = [\times \ \times \ s_2]$$

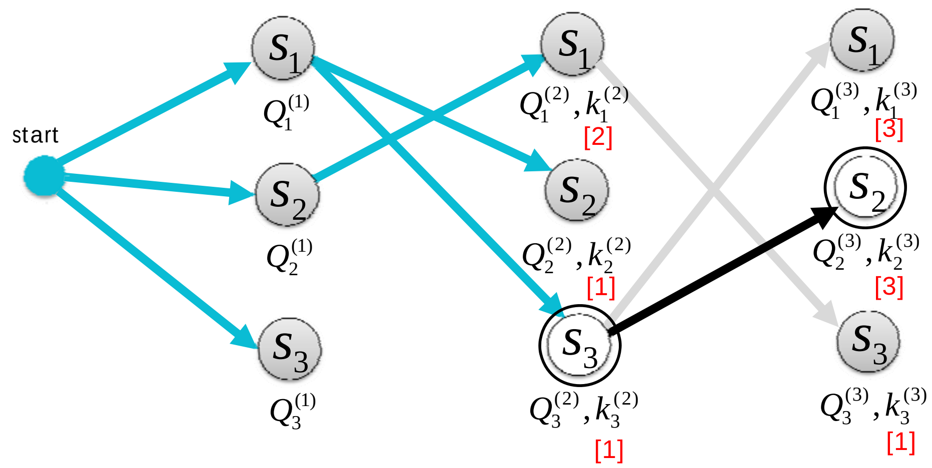
$$s_{opt}^{(T)} = \arg\max_k Q_k^{(T)}$$



Back Tracking

$$\mathbf{S}_{opt} = [\times \;\; s_3 \;\; s_2]$$

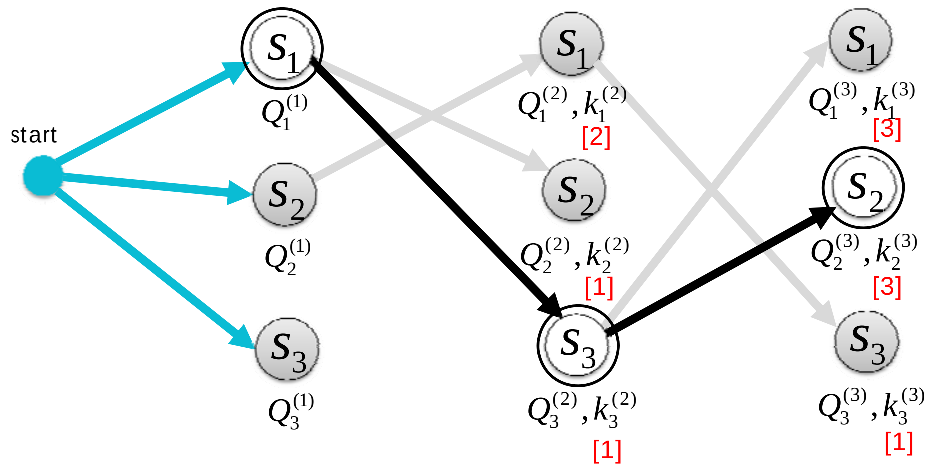
$$s_{opt}^{(t-1)} = k(s_{opt}^{(t)})$$

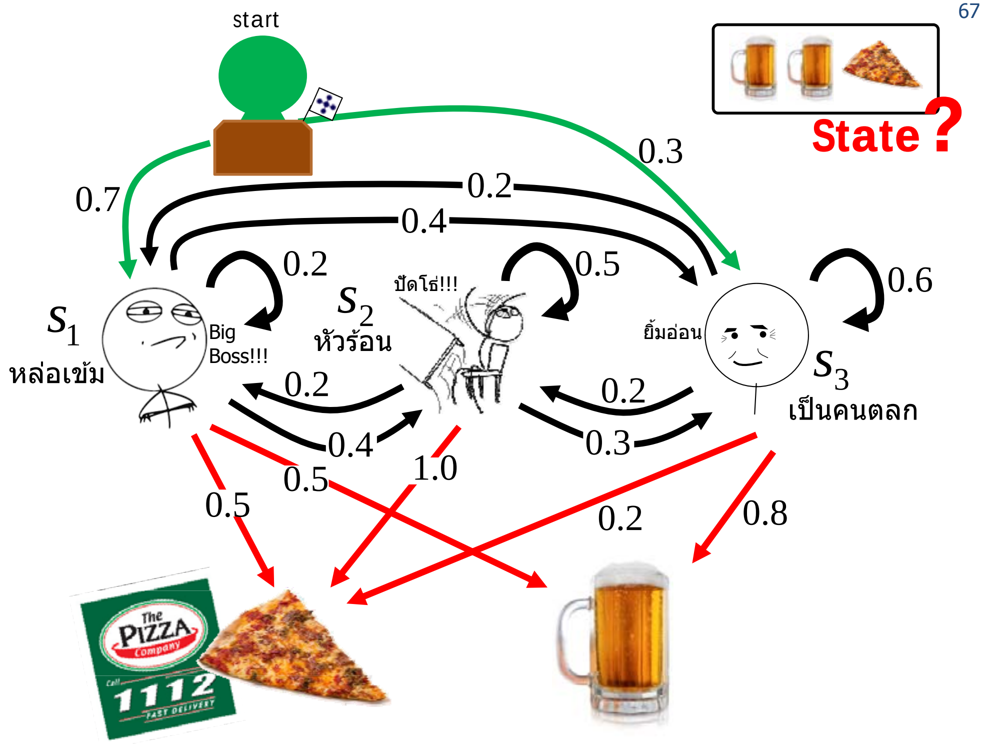


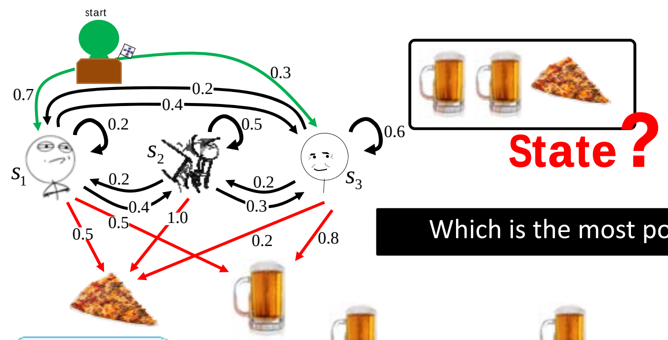
Back Tracking

$$\mathbf{S}_{opt} = [s_1 \;\; s_3 \;\; s_2]$$

$$s_{opt}^{(t-1)} = k(s_{opt}^{(t)})$$

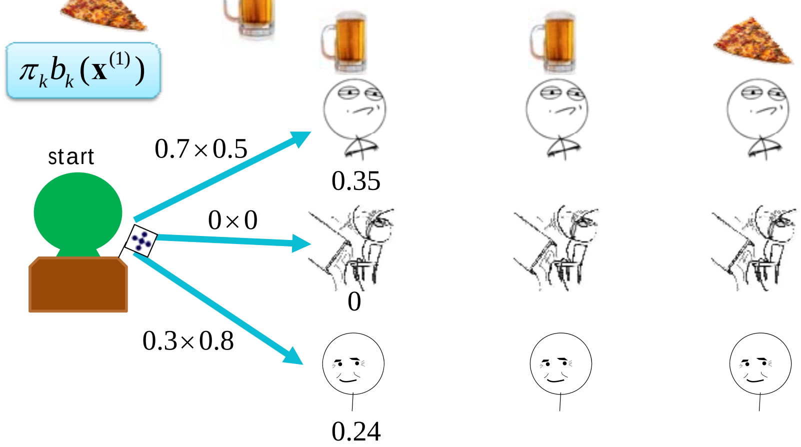


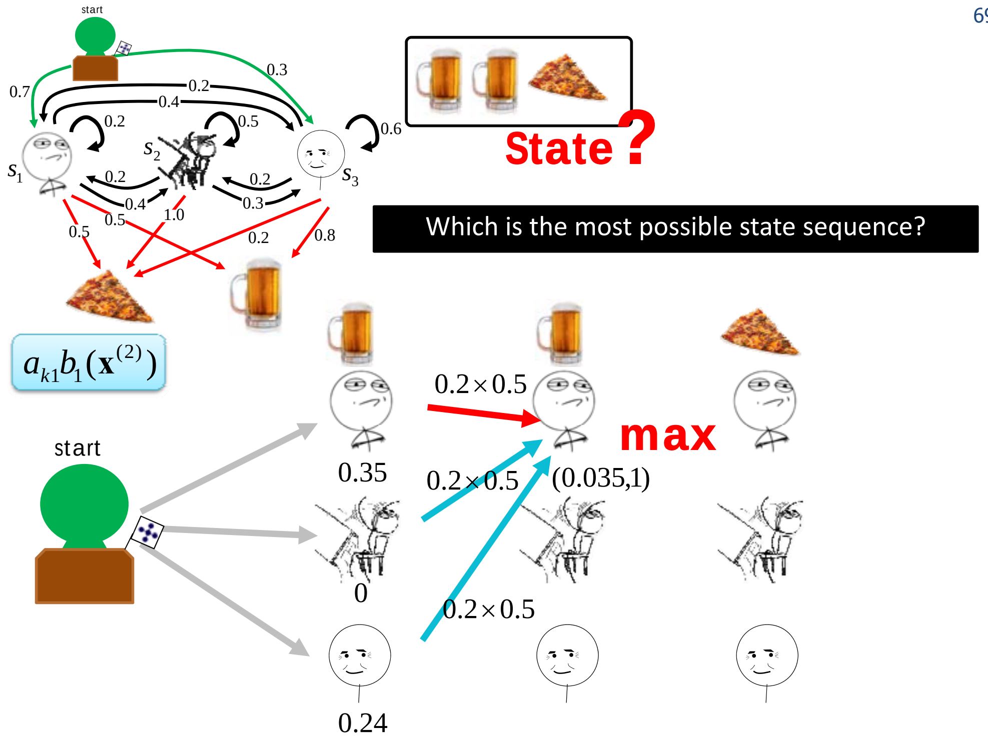


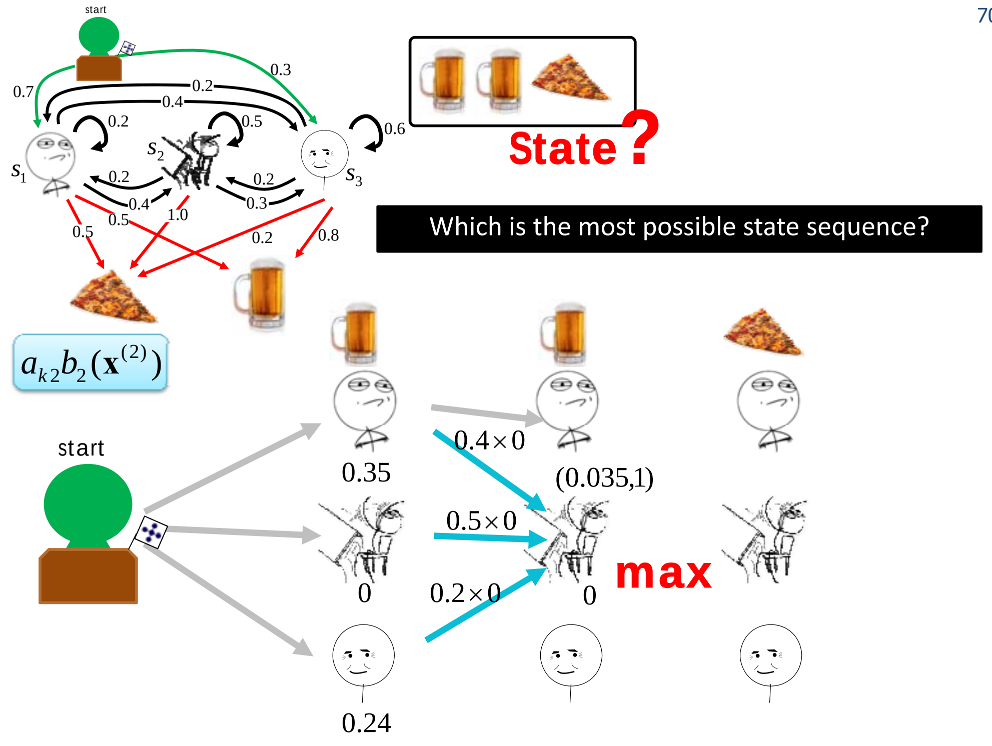


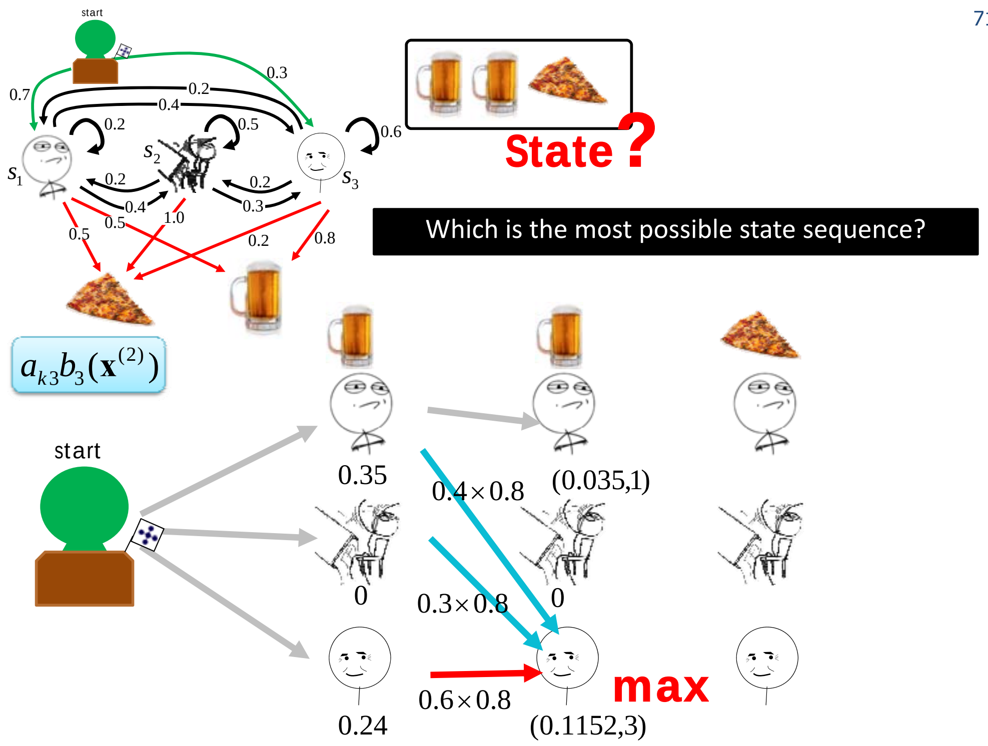
Which is the most possible state sequence?

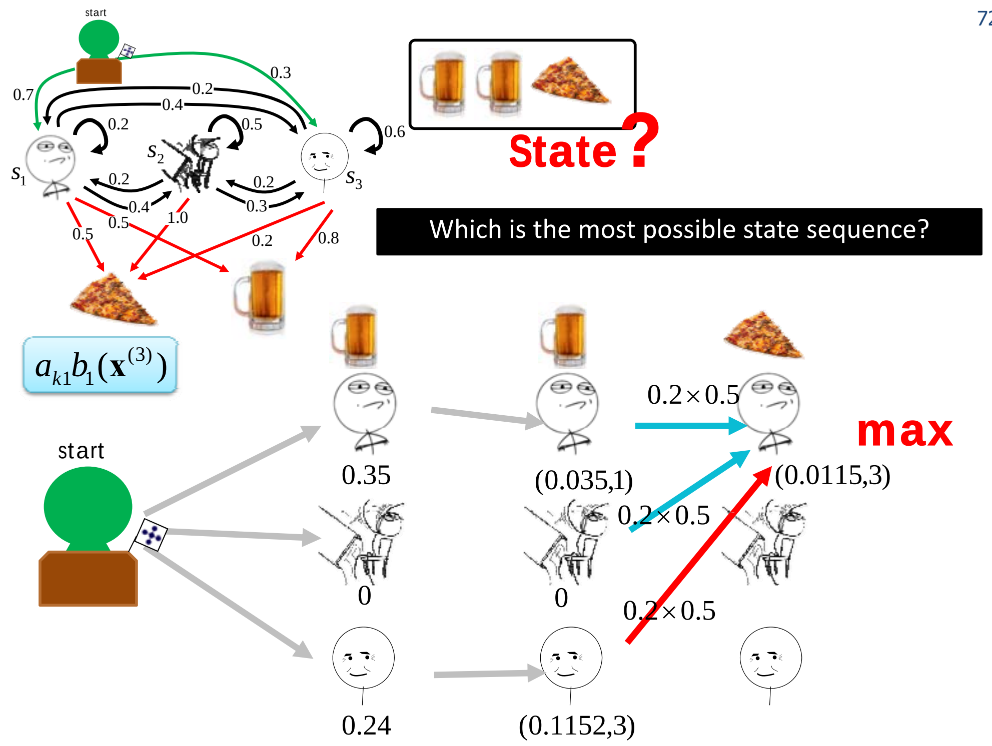
$$\pi_k b_k(\mathbf{x}^{(1)})$$

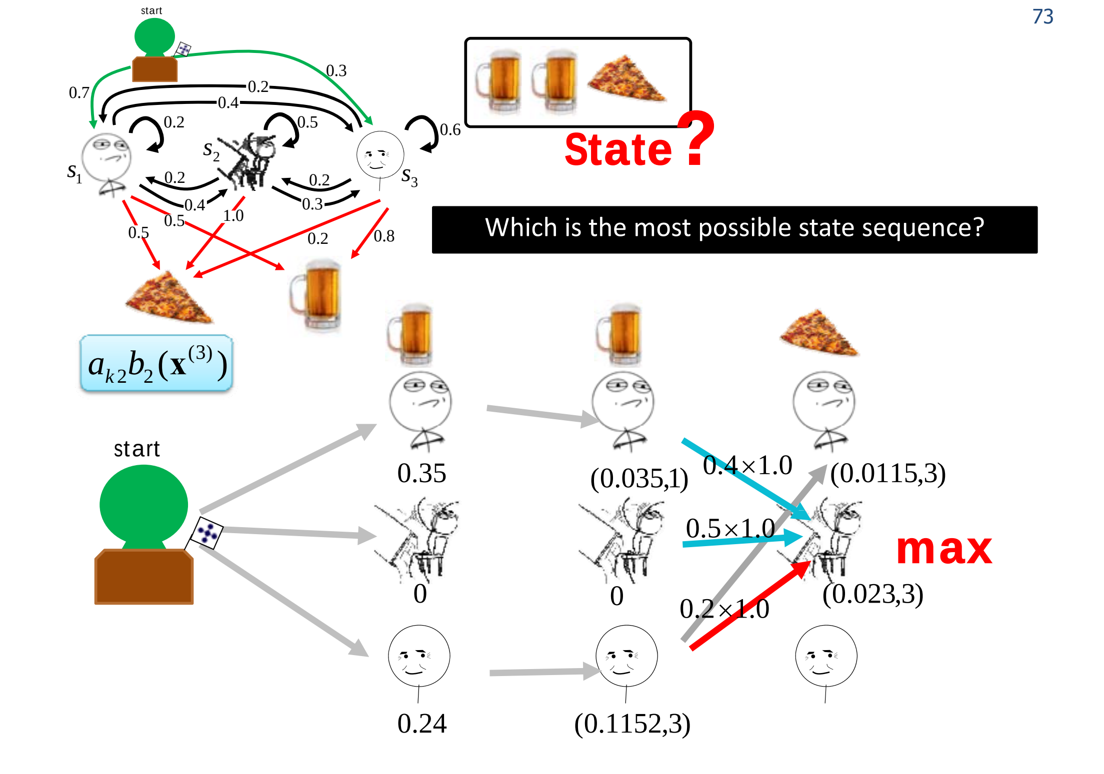


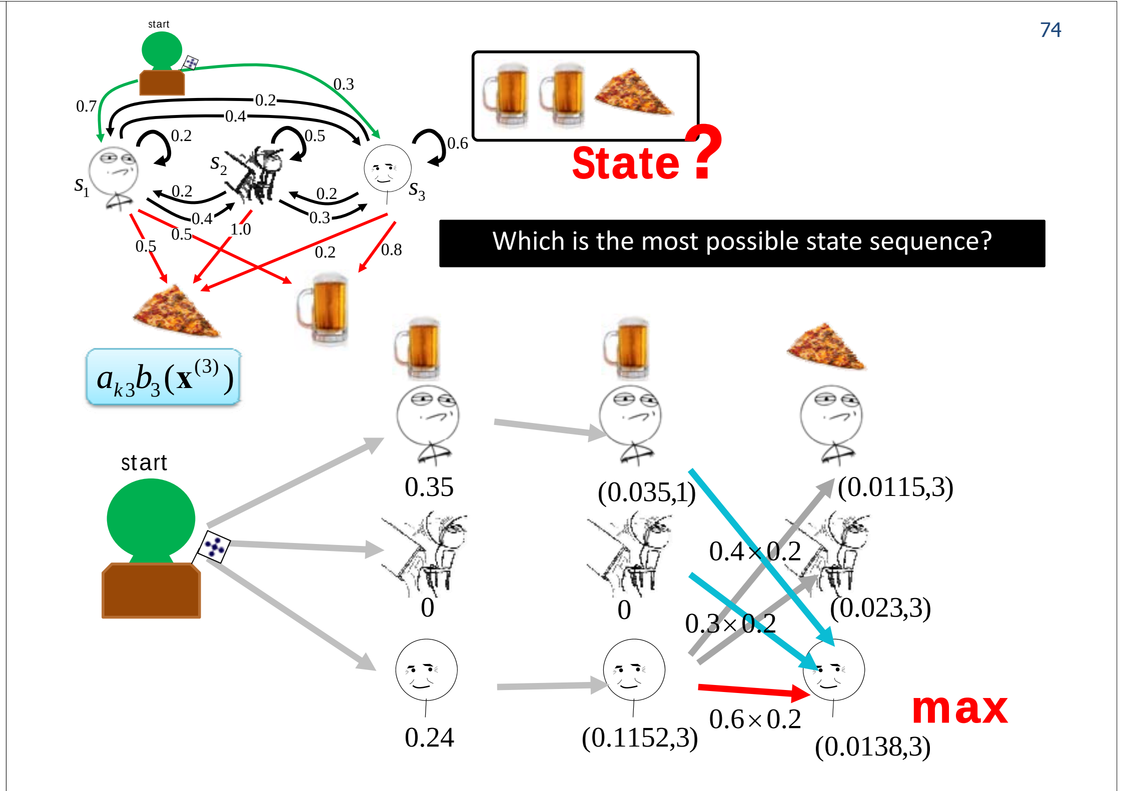

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State?
Which is the most possible state sequence?
$a_{k1}b_1(\mathbf{x}^{(2)})$
start
0.35
0
0.24
0.2×0.5
0.2×0.5
0.2×0.5
max
(0.035,1)

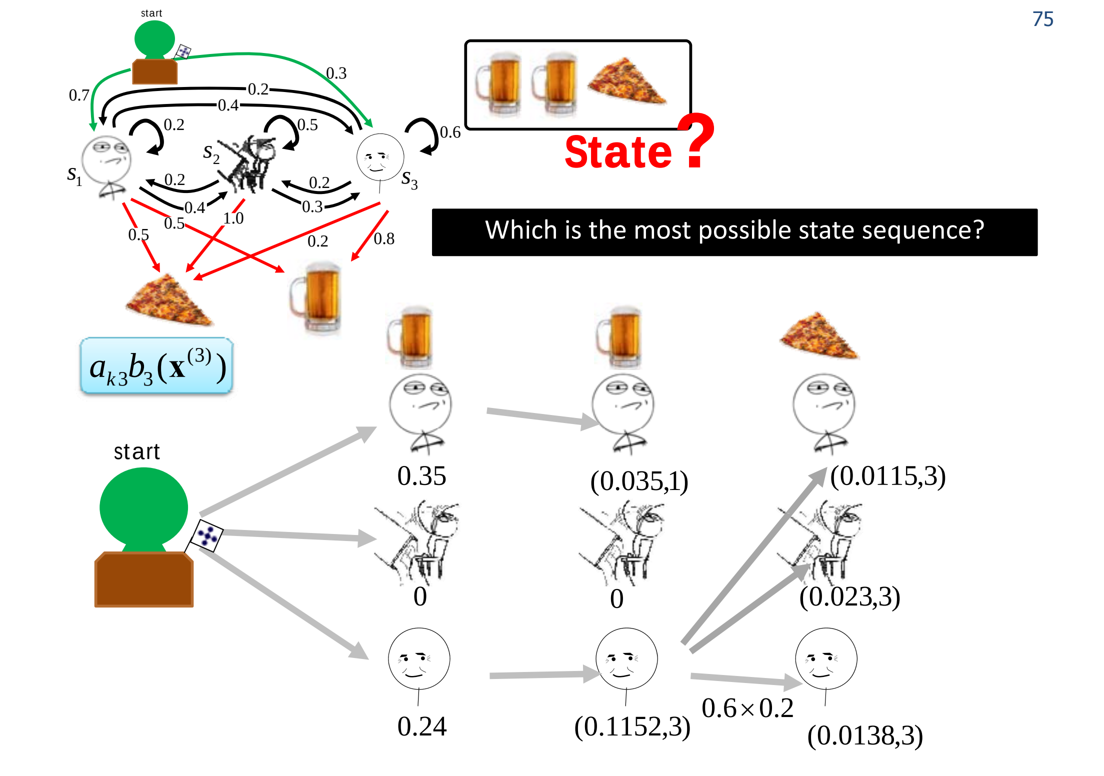

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State?
Which is the most possible state sequence?
$a_{k2}b_2(\mathbf{x}^{(2)})$
start
0.35
0
0.24
0.4×0
(0.035,1)
0.5×0
0.2×0
0
max

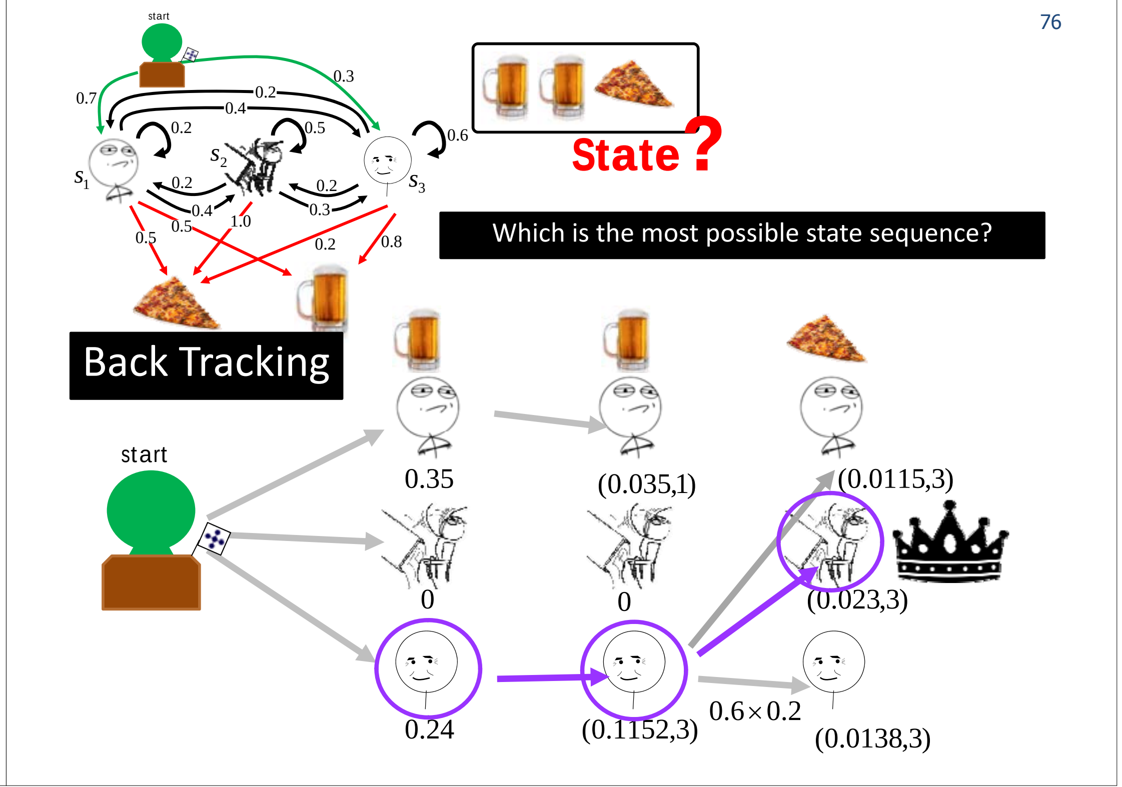

start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State?
Which is the most possible state sequence?
$a_{k3}b_3(\mathbf{x}^{(2)})$
start
0.35
0
0.24
0.4×0.8
(0.035,1)
0.3×0.8
0
0.6×0.8
max
(0.1152,3)


start
0.7
0.2
0.4
0.3
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State?
Which is the most possible state sequence?
$a_{k1}b_1(\mathbf{x}^{(3)})$
start
0.35
0
0.24
0.2×0.5
max
(0.035,1)
(0.0115,3)
0.2×0.5
0
0.2×0.5
(0.1152,3)


start
0.7
0.3
0.2
0.4
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State ?
Which is the most possible state sequence?
$a_{k2}b_2(\mathbf{x}^{(3)})$
start
0.35
(0.035,1)
0.4×1.0
(0.0115,3)
0.5×1.0
max
0
0
(0.023,3)
0.2×1.0
0.24
(0.1152,3)


start
0.7
0.3
0.2
0.4
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State ?
Which is the most possible state sequence?
$a_{k3}b_3(\mathbf{x}^{(3)})$
start
0.35
(0.035,1)
(0.0115,3)
0.4×0.2
0
0
(0.023,3)
0.3×0.2
0.24
(0.1152,3)
0.6×0.2
max
(0.0138,3)

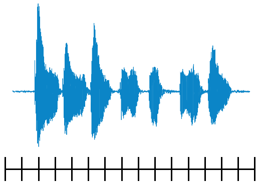

start
0.7
0.3
0.2
0.4
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State ?
Which is the most possible state sequence?
$a_{k3}b_3(\mathbf{x}^{(3)})$
start
0.35
(0.035,1)
(0.0115,3)
0
0
(0.023,3)
0.24
(0.1152,3)
0.6×0.2
(0.0138,3)


start
0.7
0.3
0.2
0.4
0.2
0.5
0.6
s_1
s_2
s_3
0.2
0.4
0.2
0.3
0.5
0.5
1.0
0.2
0.8
State ?
Which is the most possible state sequence?
Back Tracking
start
0.35
(0.035,1)
(0.0115,3)
0
0
(0.023,3)
0.24
(0.1152,3)
0.6×0.2
(0.0138,3)

Example

กากไหม?

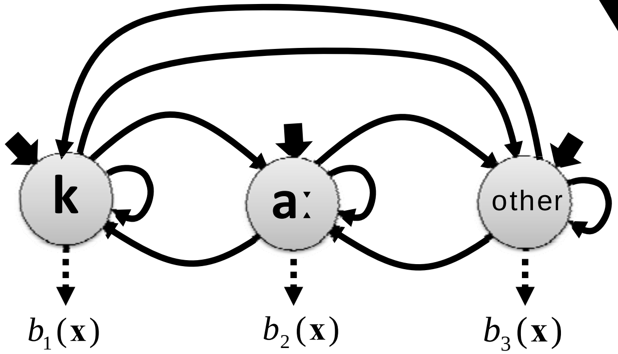
Speech



Feature Extraction

Observation Sequence $\mathbf{X}$

Find State Sequence



Optimal State Sequence S_{opt}

Ex: มันกากมาก
[-,-,-,k,a:,k,-,a:,k]

Give the observation sequences, how to choose an optimal model parameters.

Observation sequence $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, ..., \mathbf{x}^{(T)}]$

⬇

Optimal Parameters in Model

$$\lambda = \{\{a_{ij}\}, \{b_{ij}\}, \{\pi_i\}\}$$

Hidden Markov Model

Training HMM

- Find $\lambda = \{\{a_{ij}\}, \{b_{ij}\}, \{\pi_i\}\}$
- To maximize likelihood

$$L(\lambda|\mathbf{X}) = P(\mathbf{X}|\lambda)$$

- Subject to a constraint

$$\sum_j \pi_j = 1$$

$$\sum_j a_{ij} = 1 \text{ for every } i$$

$$\sum_j b_{ij} = 1 \text{ for every } i$$

*** For continuous observation value $\int b_i(\mathbf{x}) d\mathbf{x} = 1 \text{ for every } i$

Baum-Welch Algorithm

$$\gamma_i^{(t)} = P(s_i^{(t)}|\mathbf{X}, \lambda)$$

⬇

$$\gamma_i^{(t)} = \frac{P(s_i^{(t)}, \mathbf{X}|\lambda)}{P(\mathbf{X}|\lambda)} = \frac{\alpha_i^{(t)}\beta_i^{(t)}}{\sum_{k=1}^{L} \alpha_k^{(t)}\beta_k^{(t)}}$$

Baum-Welch Algorithm

$$\xi_{ij}^{(t)} = P(s_i^{(t)}, s_j^{(t+1)} \mid \mathbf{X}, \lambda)$$

$$\xi_{ij}^{(t)} = \frac{P(s_i^{(t)}, s_j^{(t+1)}, \mathbf{X} \mid \lambda)}{P(\mathbf{X} \mid \lambda)} = \frac{\alpha_i^{(t)} a_{ij} b_j(\mathbf{x}^{(t+1)}) \beta_j^{(t+1)}}{\sum_{q=1}^{L}\sum_{p=1}^{L} \alpha_p^{(t)} a_{pq} b_q(\mathbf{x}^{(t+1)}) \beta_q^{(t+1)}}$$

Baum-Welch Algorithm

$$\pi_i = \text{Expected Number of Times in } s_i \text{ at } t=1 = \gamma_i^{(1)}$$

$$a_{ij} = \frac{\text{Expected Number of Transition from } s_i \text{ to } s_j}{\text{Expected Number of Transition from } s_i} = \frac{\sum_{t=1}^{T-1} \xi_{ij}^{(t)}}{\sum_{t=1}^{T-1} \gamma_i^{(t)}}$$

$$b_{ij} = \frac{\text{Expected Number of Times in } s_i \text{ and give } \mathbf{v}_j}{\text{Expected Number of Times in } s_i} = \frac{\sum_{\mathbf{x}^{(t)}=\mathbf{v}_j} \gamma_i^{(t)}}{\sum_{t=1}^{T} \gamma_i^{(t)}}$$

Baum-Welch Algorithm

Initial Guess

$$\pi_i = \gamma_i^{(1)}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}^{(t)}}{\sum_{t=1}^{T-1} \gamma_i^{(t)}}$$

$$b_{ij} = \frac{\sum_{\mathbf{x}^{(t)}=\mathbf{v}_j} \gamma_i^{(t)}}{\sum_{t=1}^{T} \gamma_i^{(t)}}$$

Iterative

$$\alpha_i^{(t)} = \left(\sum_{k=1}^{L} \alpha_k^{(t-1)} a_{ki}\right) b_i(\mathbf{x}^{(t)})$$

$$\beta_i^{(t)} = \sum_{k=1}^{L} a_{ik} b_k(\mathbf{x}^{(t+1)}) \beta_k^{(t+1)}$$

$$\gamma_i^{(t)} = \frac{\alpha_i^{(t)} \beta_i^{(t)}}{\sum_{k=1}^{L} \alpha_k^{(t)} \beta_k^{(t)}}$$

$$\xi_{ij}^{(t)} = \frac{\alpha_i^{(t)} a_{ij} b_j(\mathbf{x}^{(t+1)}) \beta_j^{(t+1)}}{\sum_{q=1}^{L}\sum_{p=1}^{L} \alpha_p^{(t)} a_{pq} b_q(\mathbf{x}^{(t+1)}) \beta_q^{(t+1)}}$$

Baum-Welch Algorithm

For Continuous Observation Value in Mixture of Gaussian model

$$b_i(\mathbf{x}) = \sum_{m=1}^{M} c_{im} f_G(\mathbf{x} \mid \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})$$

$$\gamma_{im}^{(t)} = \frac{\alpha_i^{(t)} \beta_i^{(t)}}{\sum_{k=1}^{L} \alpha_k^{(t)} \beta_k^{(t)}} \cdot \frac{c_{im} f_G(\mathbf{x}^{(t)} \mid \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})}{b_i(\mathbf{x}^{(t)})}$$

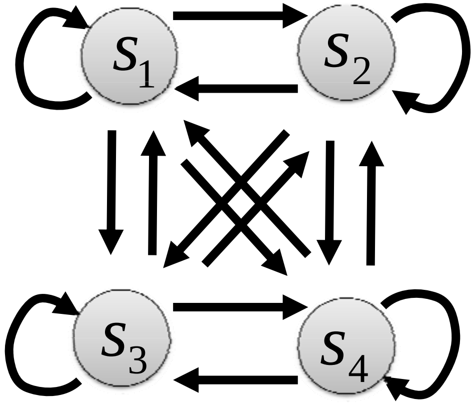
Baum-Welch Algorithm

$$c_{im} = \frac{\sum_{t=1}^{T} \gamma_{im}^{(t)}}{\sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{im}^{(t)}}$$

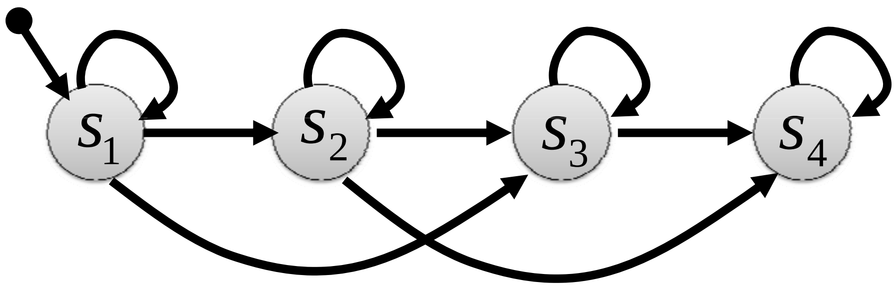
$$\boldsymbol{\mu}_{im} = \frac{\sum_{t=1}^{T} \gamma_{im}^{(t)} \mathbf{x}^{(t)}}{\sum_{t=1}^{T} \gamma_{im}^{(t)}}$$

$$\boldsymbol{\Sigma}_{im} = \frac{\sum_{t=1}^{T} \gamma_{im}^{(t)} \left(\mathbf{x}^{(t)} - \boldsymbol{\mu}_{im}\right)^{T} \left(\mathbf{x}^{(t)} - \boldsymbol{\mu}_{im}\right)}{\sum_{t=1}^{T} \gamma_{im}^{(t)}}$$

Ergodic Model



Left-Right Model



$$a_{ij} = 0;\ j < i$$
$$a_{ij} = 0;\ j > i + \Delta$$

$$\pi_1 = 1$$
$$\pi_i = 0;\ i \neq 1$$