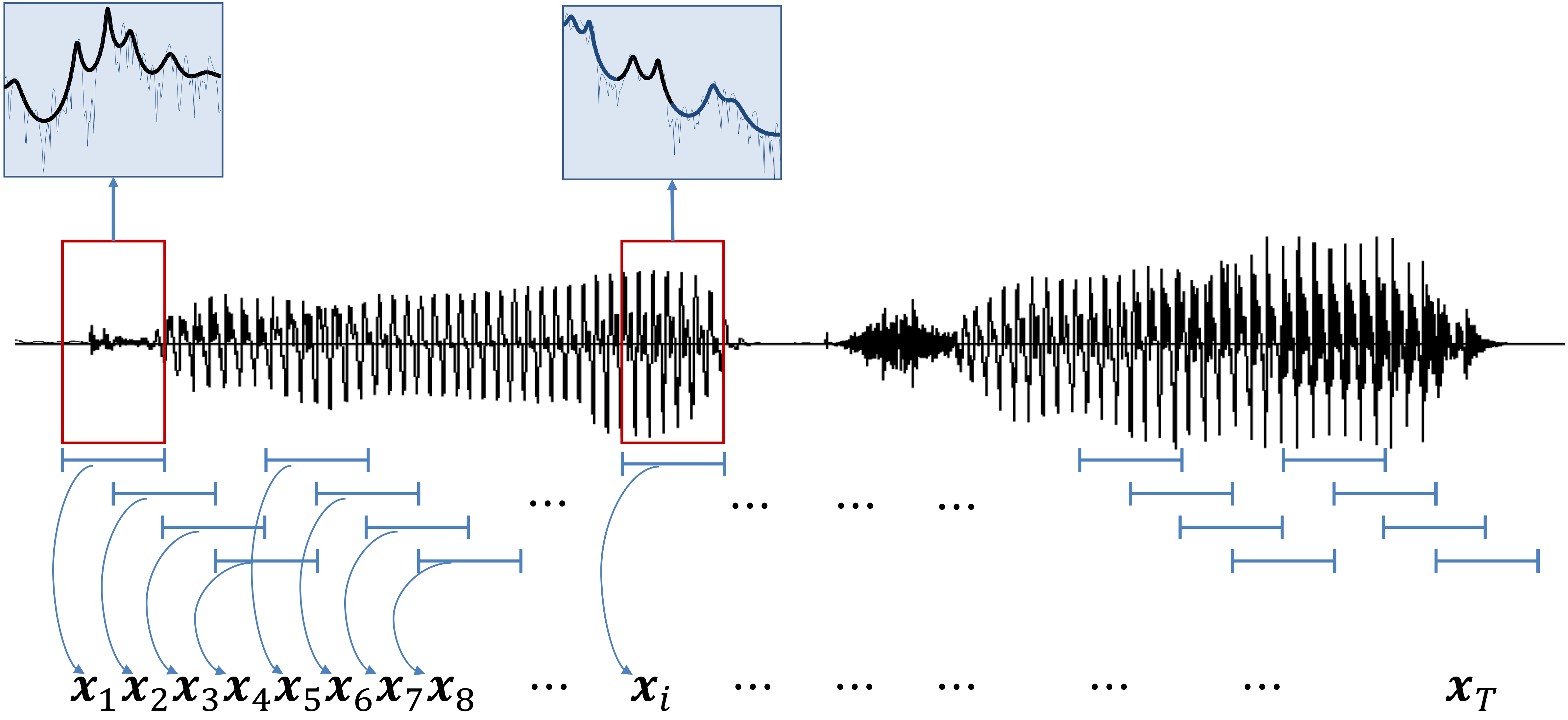
隠れマルコフモデル
Hidden Markov Models

― 生成モデルに基づく系列データのパターン認識

系列データの例：音声のベクトル列での表現



系列認識における生成モデル的アプローチ

- クラス毎に確率モデルを用意し，その確率モデルから認識対象のデータが生成される確率を求め，その最大値を与えるクラスを識別結果とする。

$$k = \underset{i}{\operatorname{argmax}} \ \Pr(w_i|\boldsymbol{X})$$

$$k = \underset{i}{\operatorname{argmax}} \ \Pr(\boldsymbol{X}|w_i) \ \Pr(w_i)$$

$\boldsymbol{X} = \boldsymbol{x}_1\boldsymbol{x}_2 \cdots \boldsymbol{x}_j \cdots \boldsymbol{x}_N$ ：クラス未知のベクトル系列データ

w_i ：i 番目のクラス

系列認識における生成モデル的アプローチ

- $\Pr(\boldsymbol{X}|w_i)$が求まれば，系列データのパターン認識ができる

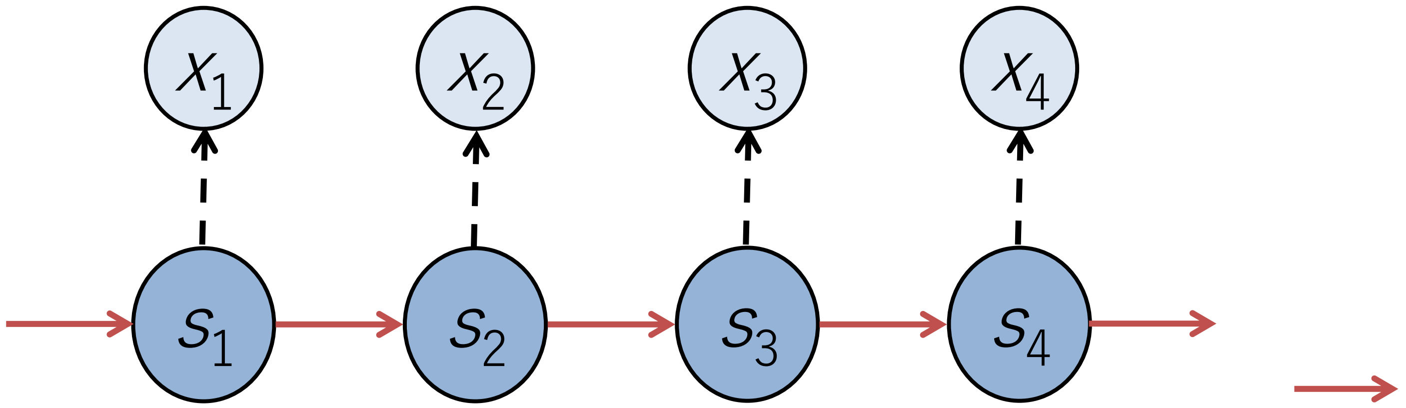
・・・では， $\Pr(\boldsymbol{X}|w_i)$ はどのように求めるか？？

- 常套手段は，確率モデルを仮定して，そのモデルのパラメタを学習データを用いて推定すること。

・・・では，どんな確率モデルを仮定するのか？？

§ マルコフモデル

マルコフモデルは，現状態のみに依存して次状態が確率的に定まる状態遷移を繰り返しながら，現状態によって**一意に定まる出力**を行う確率モデル。



確率的依存関係

確定的依存関係

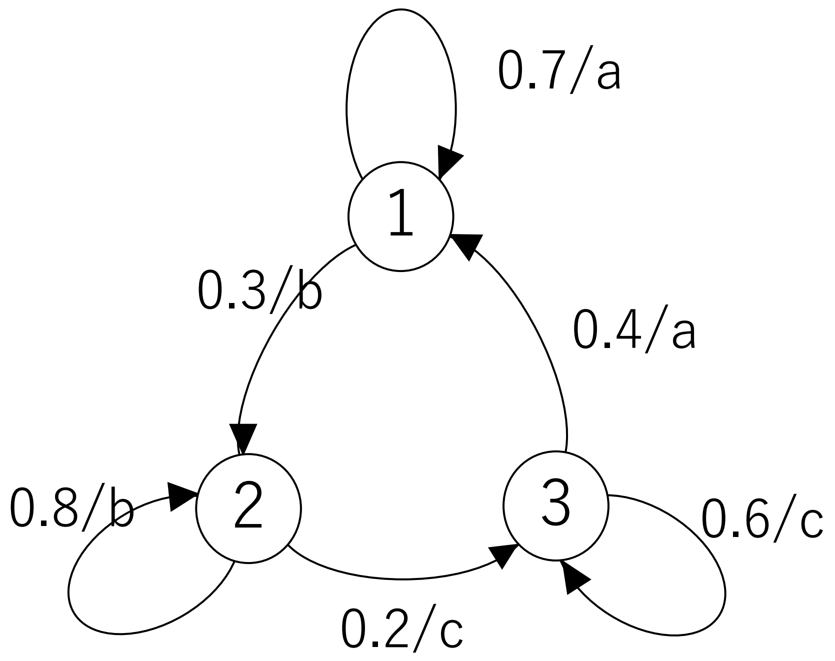
s_t ：　時刻tにおける状態

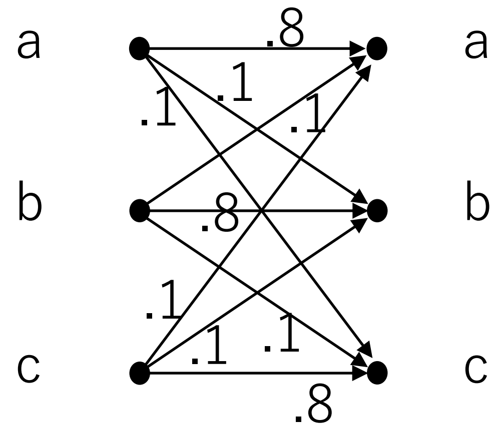
x_t ：　時刻tにおける出力

マルコフモデルの限界

マルコフモデルでは，複雑な出力を表現できない。

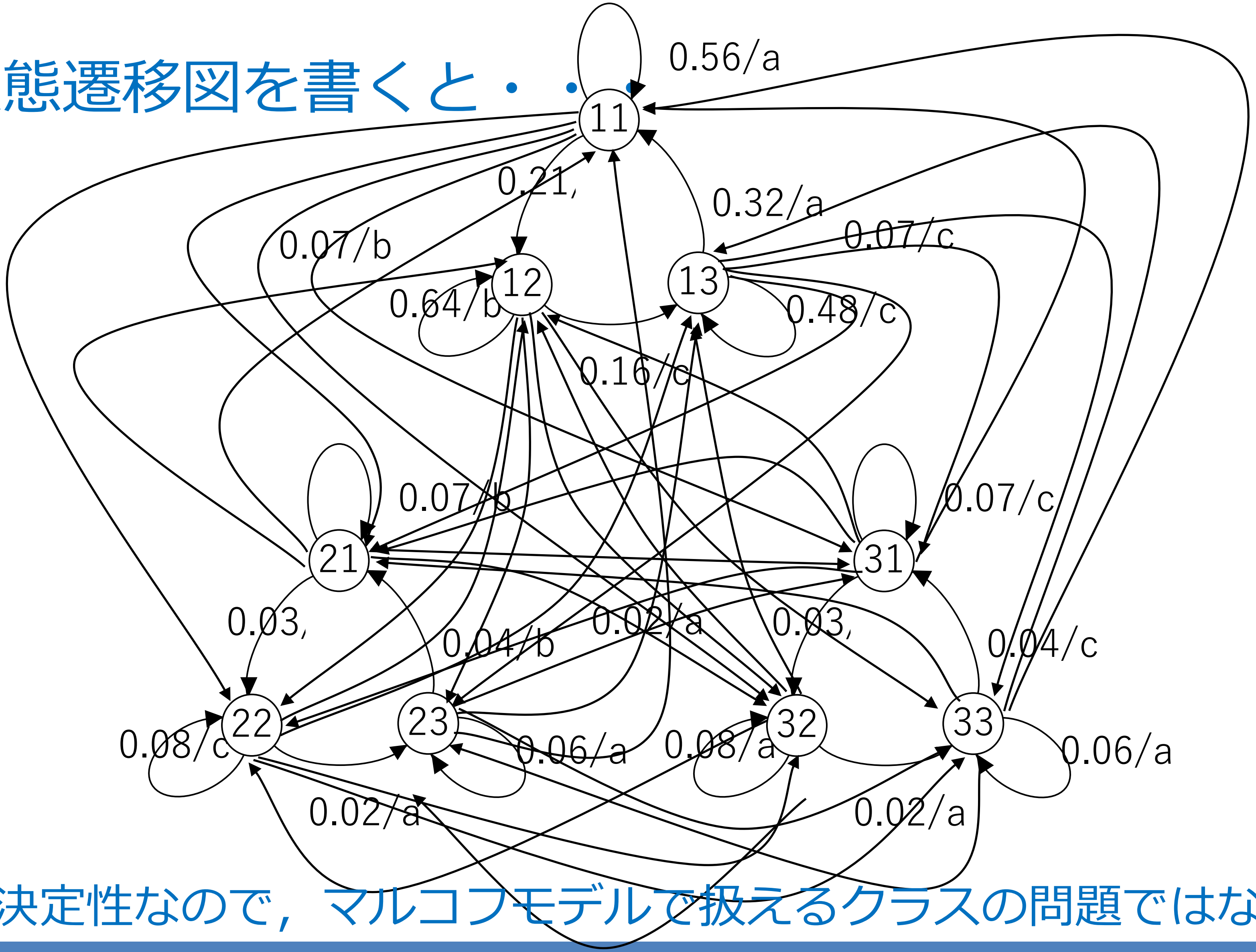
…左下図のマルコフモデルの出力が右下図の雑音を持つ通信路を経て観測されることを考える…





マルコフモデルの限界

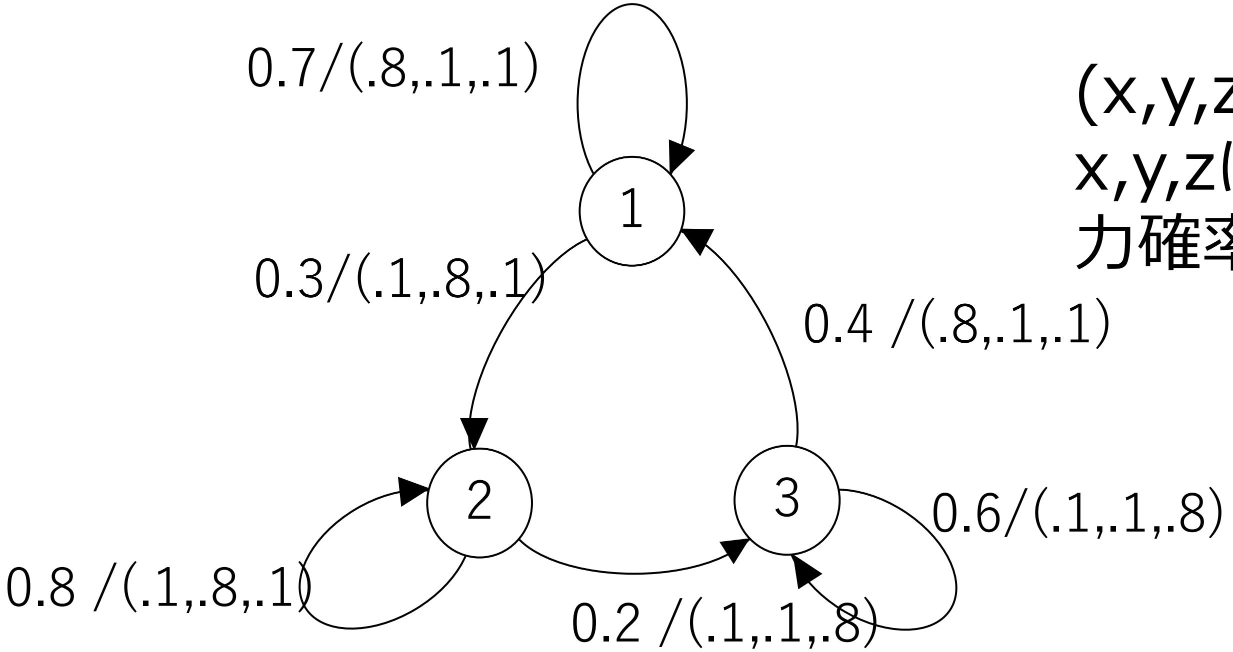
強引に状態遷移図を書くと・・・



（非決定性なので，マルコフモデルで扱えるクラスの問題ではない）

隠れマルコフモデル

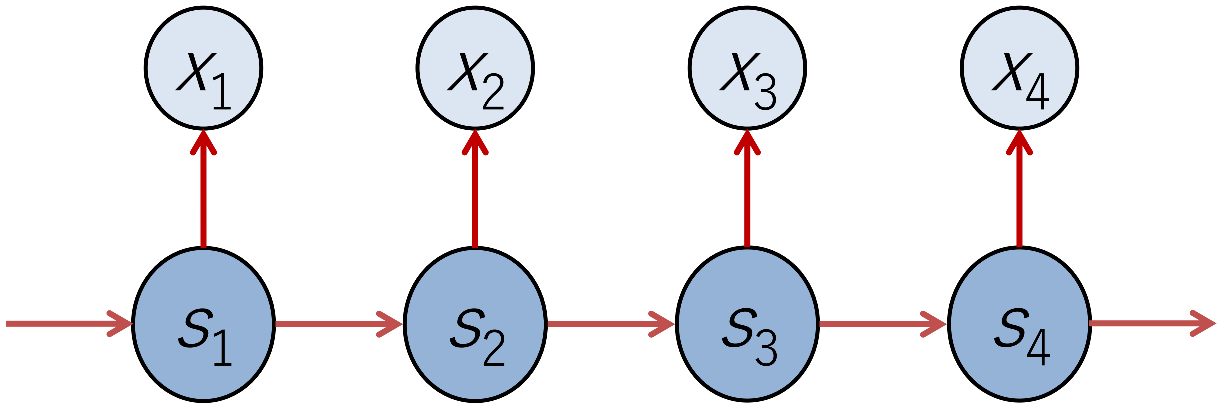
出力に確率的な表現を許すと簡潔な記述が可能になる。
　→　隠れマルコフモデル
　　　（Hidden Markov Models; HMM）



(x,y,z)　の表現における x,y,zは，順にa,b,c　の出力確率を表す。

隠れマルコフモデル

隠れマルコフモデルは，現状態のみに依存して次状態が確率的に定まる状態遷移を繰り返しながら，現状態によって確率的に定まる出力を行う確率モデル。



→ 確率的依存関係

s_t ：　時刻tにおける状態
x_t ：　時刻tにおける出力

§ 隠れマルコフモデルの表現

隠れマルコフモデル（隠れマルコフ情報源）は,

$M = \{\boldsymbol{\pi}, P, \boldsymbol{b}, E\}$

$\boldsymbol{\pi} = (\pi_i)$, π_i は初期状態が状態 i である確率

$P = \left(p_{ij}\right), p_{ij}$ は状態 i から状態 j への遷移確率

$\boldsymbol{b} = \left(b_j(x)\right)$, $b_j(x)$ は状態 j で記号(ベクトル)xが出力する確率

$E = \{e_i\}$, 最終状態として許された状態の集合

の四つ組によって表現される。

HMMにおける記号列の生起確率の計算

隠れマルコフモデル $M=\{\boldsymbol{\pi}, P, \boldsymbol{b}, E\}$ が,

$$\boldsymbol{\pi}=\begin{pmatrix}1.0\\0\end{pmatrix},\ P=\begin{pmatrix}0.8 & 0.2\\0.3 & 0.7\end{pmatrix},$$

$$\boldsymbol{b}=\begin{pmatrix}b_0(A)=0.9, & b_1(A)=0.4\\b_0(B)=0.1, & b_1(B)=0.6\end{pmatrix},$$

$$E=\{1\}$$

のとき, 「B A B」 が生起する確率を考える。

記号列　「B A B」を生起させる状態遷移列 $\boldsymbol{s}$ は,

$\boldsymbol{s}_1$: (0) → 0 → 0 →1

$\boldsymbol{s}_2$: (0) → 0 → 1 →1

$\boldsymbol{s}_3$: (0) → 1 → 0 →1

$\boldsymbol{s}_4$: (0) → 1 → 1 →1

の４通り

(初期状態 0, 最終状態 1 で固定。初期状態での出力は考えない。)

$\boldsymbol{s}_1$ から「B A B」が生起する確率は,

$$P(BAB|\boldsymbol{s}_1) = b_0(B) \times b_0(A) \times b_1(B)$$
$$= 0.1 \times 0.9 \times 0.6 = 0.054$$

同様に,

$$P(BAB|\boldsymbol{s}_2) = b_0(B) \times b_1(A) \times b_1(B)$$
$$= 0.1 \times 0.4 \times 0.6 = 0.024$$

$$P(BAB|\boldsymbol{s}_3) = b_1(B) \times b_0(A) \times b_1(B)$$
$$= 0.6 \times 0.9 \times 0.6 = 0.324$$

$$P(BAB|\boldsymbol{s}_4) = b_1(B) \times b_1(A) \times b_1(B)$$
$$= 0.6 \times 0.4 \times 0.6 = 0.144$$

それぞれの状態遷移が生じる確率は

$$P(\boldsymbol{s}_1) = \pi_0 \times p_{00} \times p_{00} \times p_{01}$$
$$= 1.0 \times 0.8 \times 0.8 \times 0.2 = 0.128$$

$$P(\boldsymbol{s}_2) = \pi_0 \times p_{00} \times p_{01} \times p_{11}$$
$$= 1.0 \times 0.8 \times 0.2 \times 0.7 = 0.112$$

$$P(\boldsymbol{s}_3) = \pi_0 \times p_{01} \times p_{10} \times p_{01}$$
$$= 1.0 \times 0.2 \times 0.3 \times 0.2 = 0.012$$

$$P(\boldsymbol{s}_4) = \pi_0 \times p_{01} \times p_{11} \times p_{11}$$
$$= 1.0 \times 0.2 \times 0.7 \times 0.7 = 0.098$$

与えられた隠れマルコフモデルから　ABAB が出力する確率は,

$$P(BAB) = P(BAB|\boldsymbol{s}_1)P(\boldsymbol{s}_1) + P(BAB|\boldsymbol{s}_2)P(\boldsymbol{s}_2) + P(BAB|\boldsymbol{s}_3)P(\boldsymbol{s}_3) + P(BAB|\boldsymbol{s}_4)P(\boldsymbol{s}_4)$$

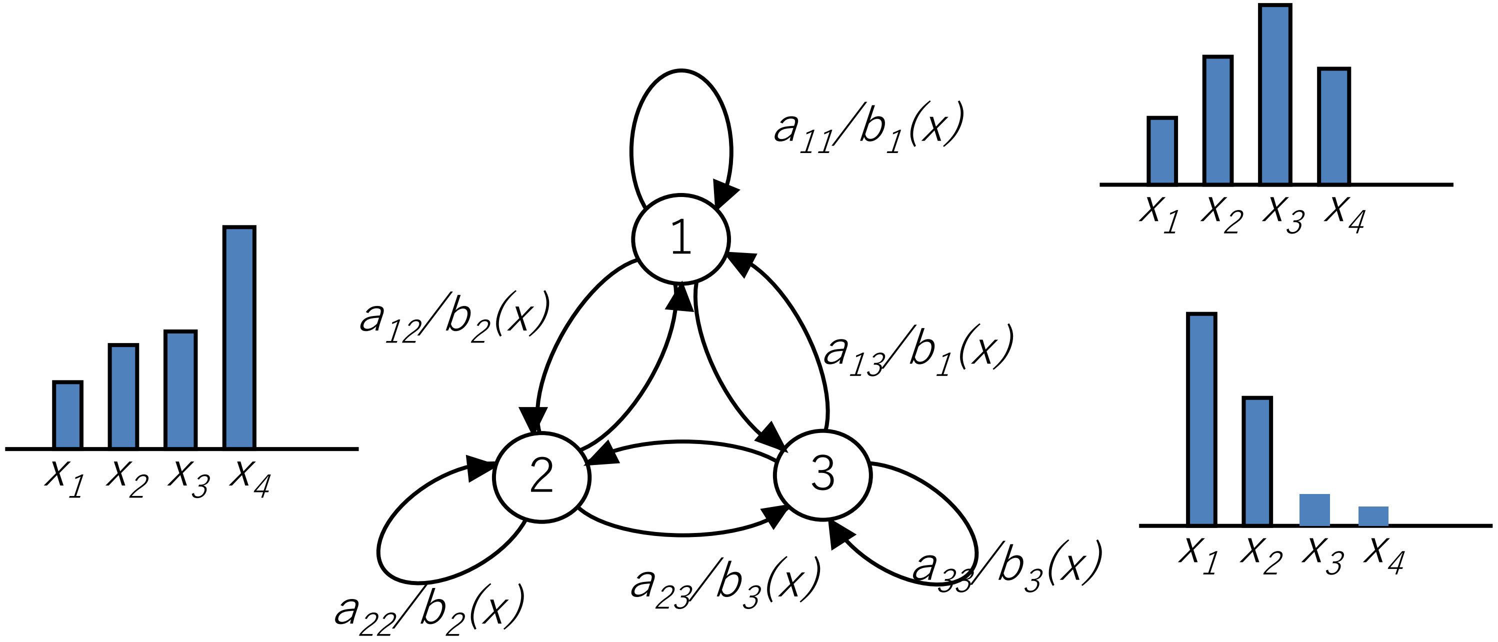
$$= 0.114$$

一般的には，隠れマルコフモデルから　記号列　X　が生起する確率は，次式で求められる

$$P(X) = \sum_{for\ all\ \boldsymbol{s}_i} P(X|\boldsymbol{s}_i)P(\boldsymbol{s}_i) \qquad (1)$$

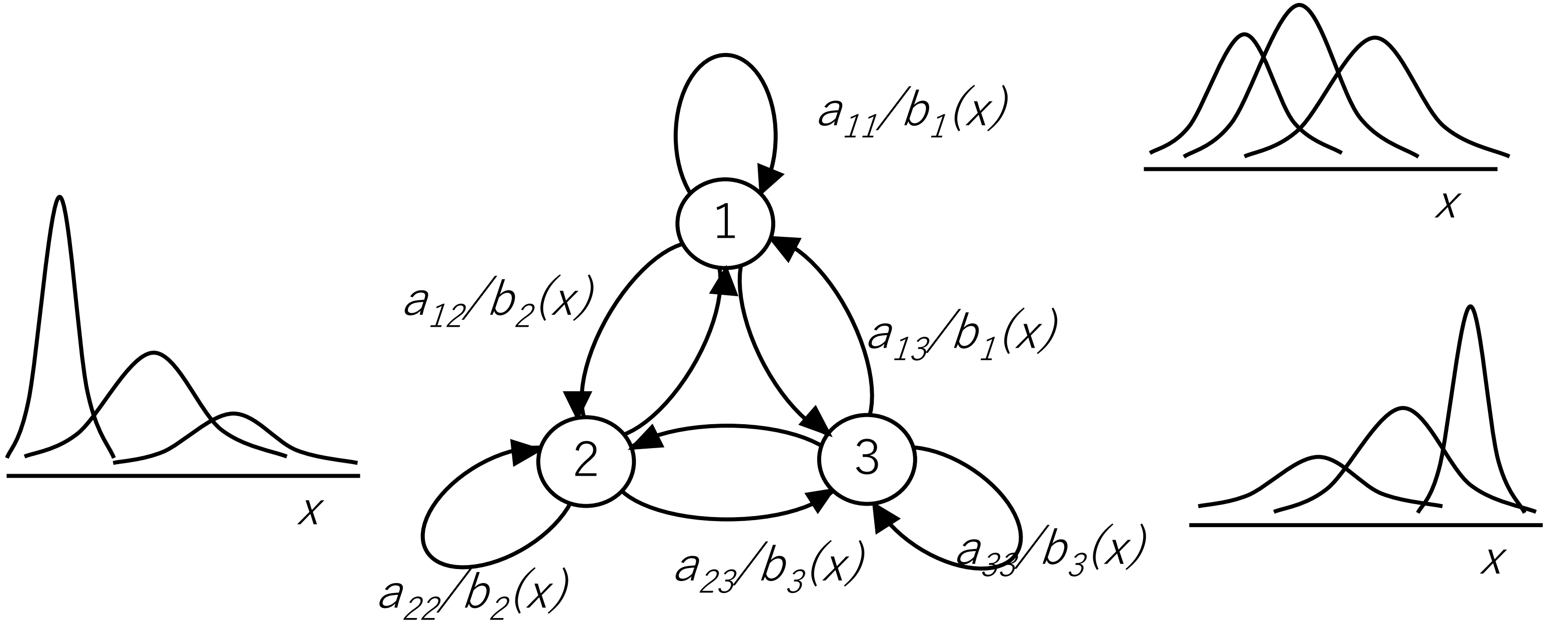
§ 出力分布によるHMMの分類

- 離散（分布型）HMM：出力として離散量を扱い，各状態における出力確率に離散分布を割り当てる。



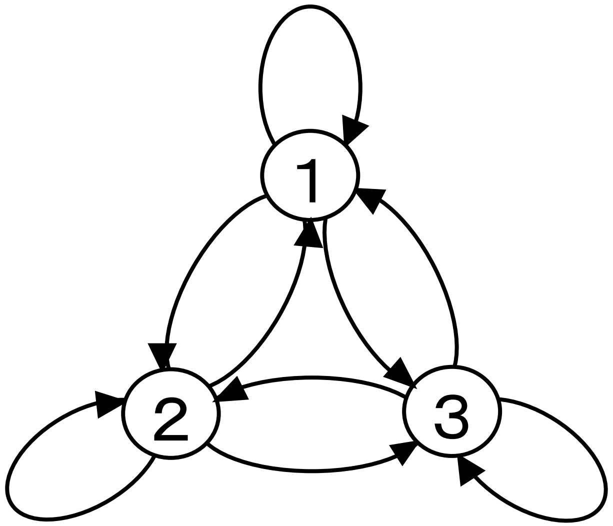
出力分布によるHMMの分類

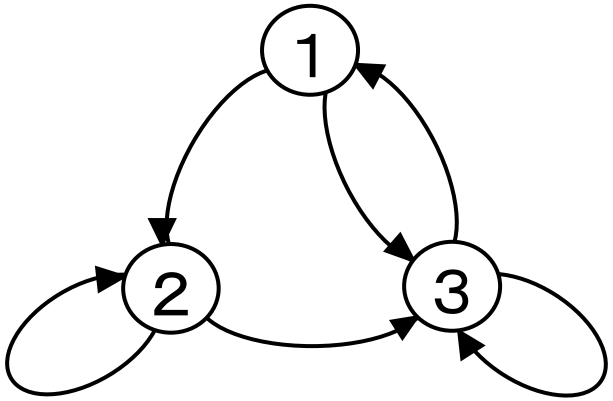
- 離散（分布型）HMM：出力として離散量を扱い，各状態における出力確率に離散分布を割り当てる。
- 連続（分布型）HMM：出力として連続量を扱い，各状態における出力確率に連続分布を割り当てる。通常，出力確率にはGMMが用いられる。



トポロジによるHMMの分類

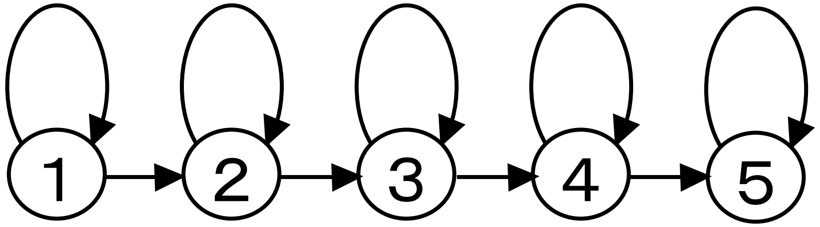
- エルゴディックHMM：既約で（どの状態からどの状態へも遷移でき）周期的でないHMM。

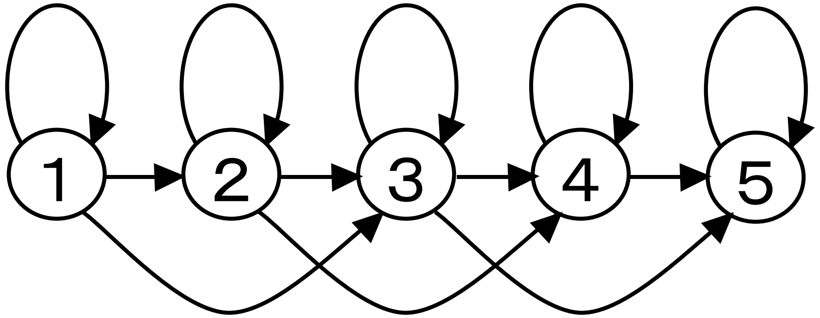




トポロジによるHMMの分類

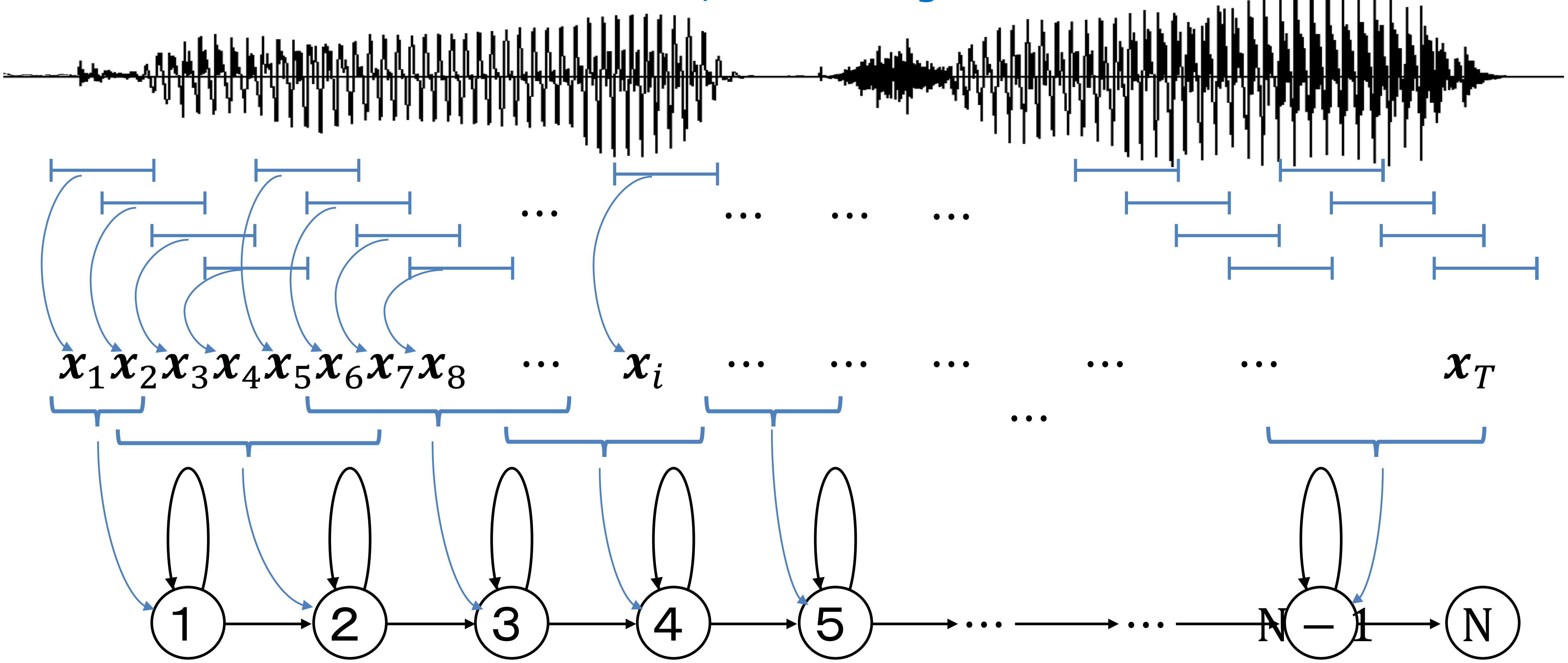
- エルゴディックHMM：既約で（どの状態からどの状態へも遷移でき）周期的でないHMM。
- left-to-right HMM：帰還ループを持たない（一度ある状態を出たら，その状態へは戻ることができない）HMM。





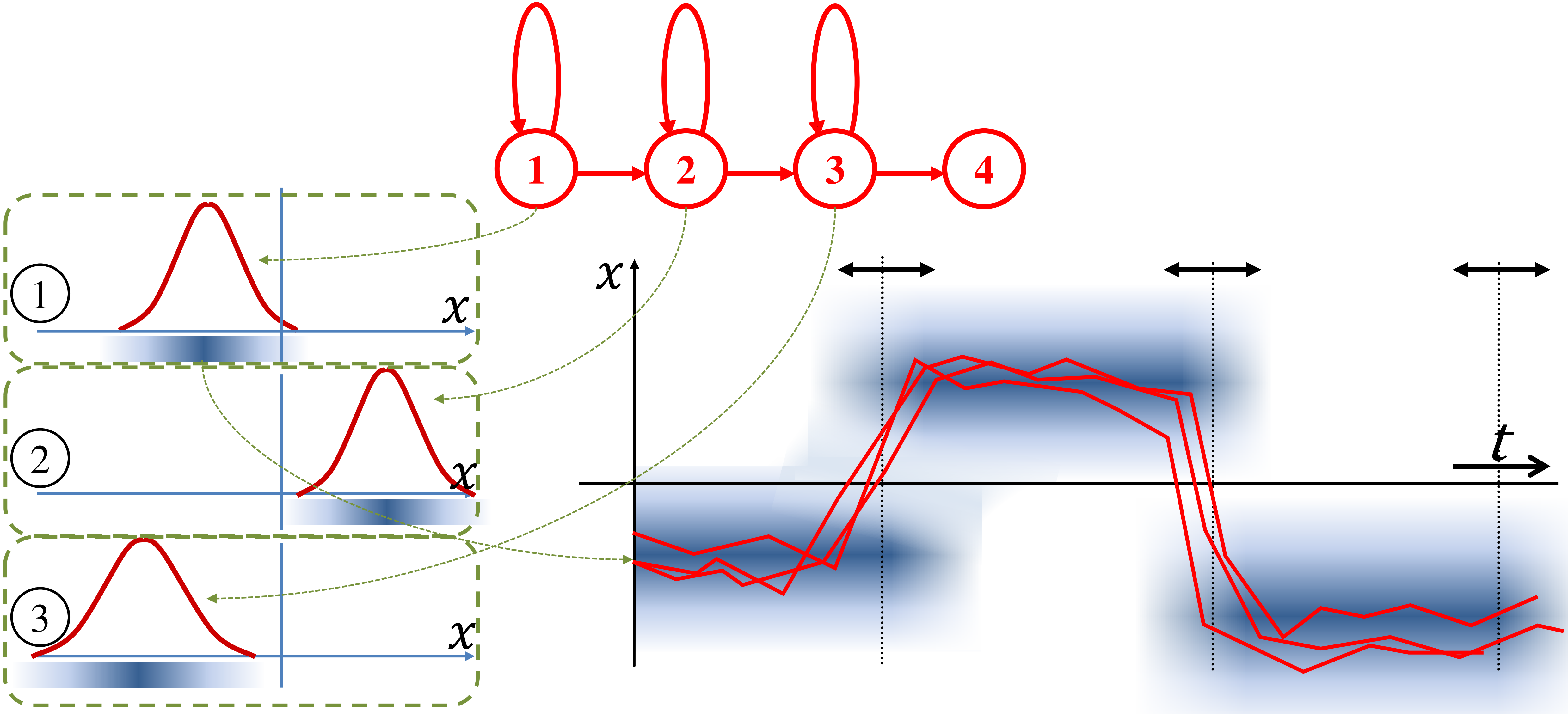
§ HMMによる単語音声の表現

音声の音響的現象を表現するのに，left-to-right 型のHMMを用いる。



これは，音声を区分定常過程として扱うことを意味する。

区分定常過程としての left-to-right HMM



§ HMMの効率的な確率計算

隠れマルコフモデルから　記号列　X　が生起する確率は，次式で求められる

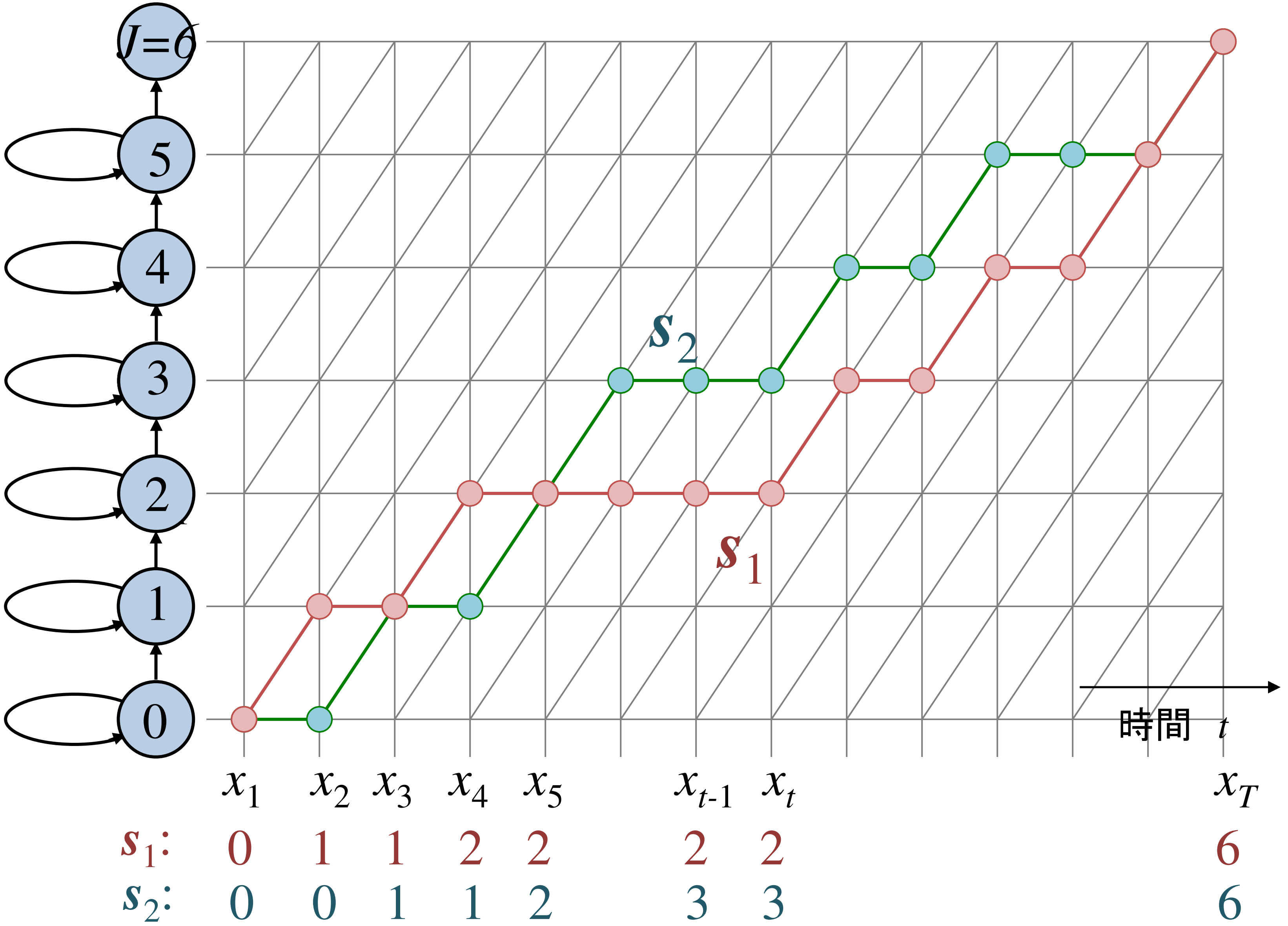
$$P(\mathrm{X}) = \sum_{for\ all\ \boldsymbol{s}_i} P(X|\boldsymbol{s}_i)P(\boldsymbol{s}_i) \quad (1)$$

状態がK個，時間がTステップのとき，状態遷移の変化は，K^T個ある。（K=10，T=10なら，10000000000通り *!*）

先に述べた生起確率の計算法は非現実的。より効率的な計算法が必要*!!*

→　経路探索問題（に近い）問題に置き換える

状態遷移のトレリス表現

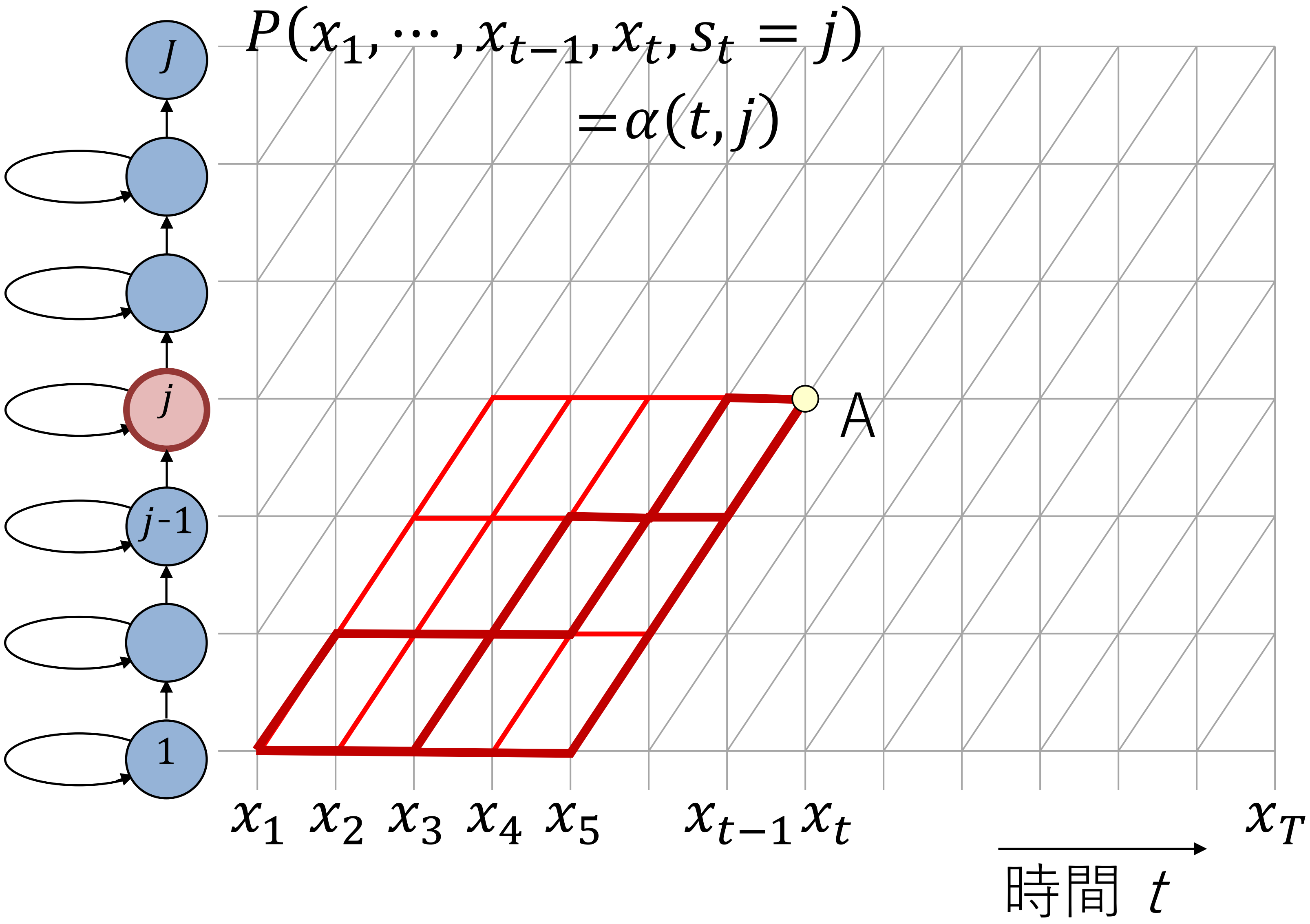


トレリスを用いた効率的確率計算
フォワードアルゴリズム

$\alpha(t,j)$ ：時刻 t までに x_1～ x_t を出力して状態 j に至る確率と定義する。

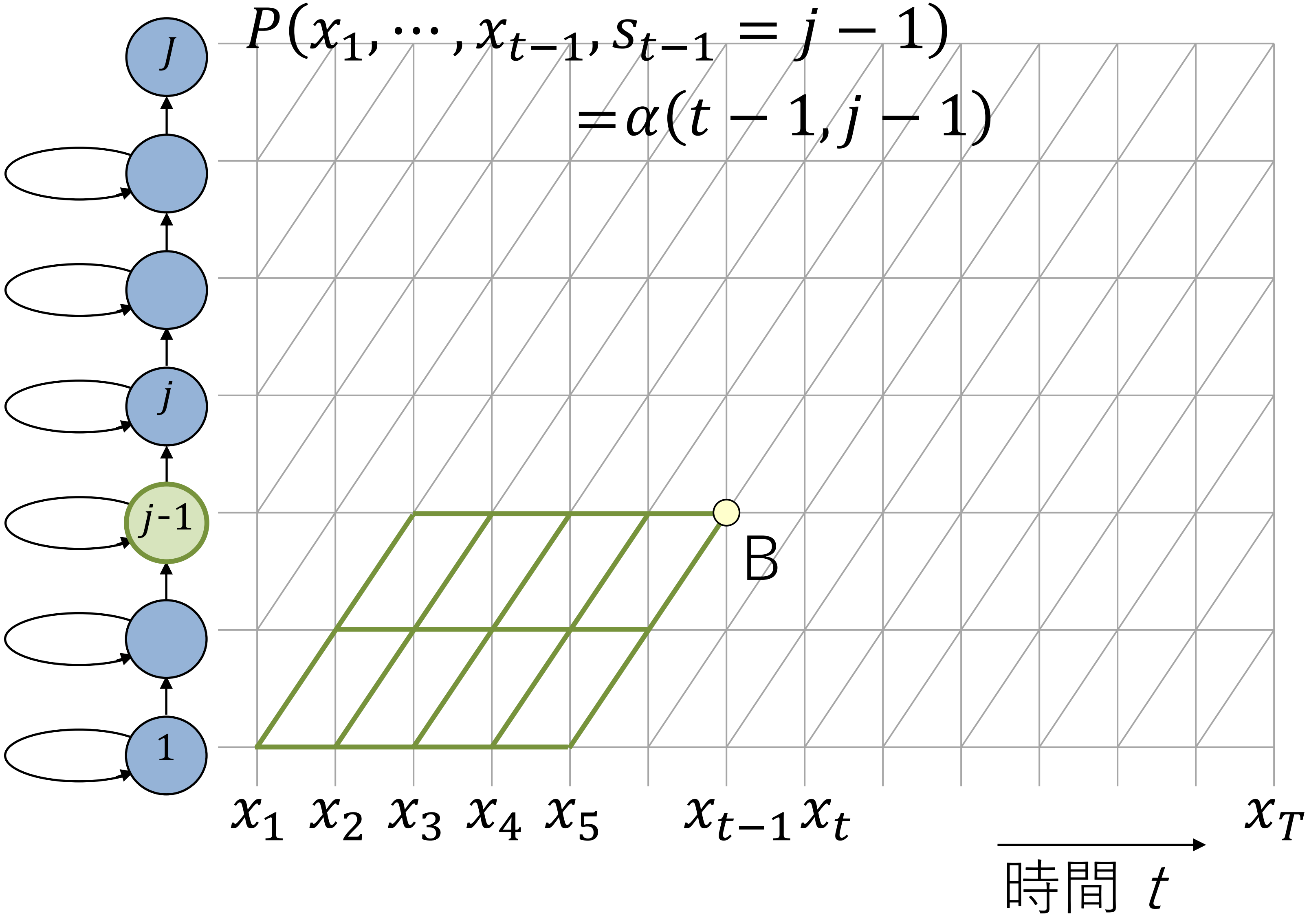
$$\alpha(t,j) = P(x_1, \cdots, x_t, s_t = j)$$

フォワードアルゴリズム



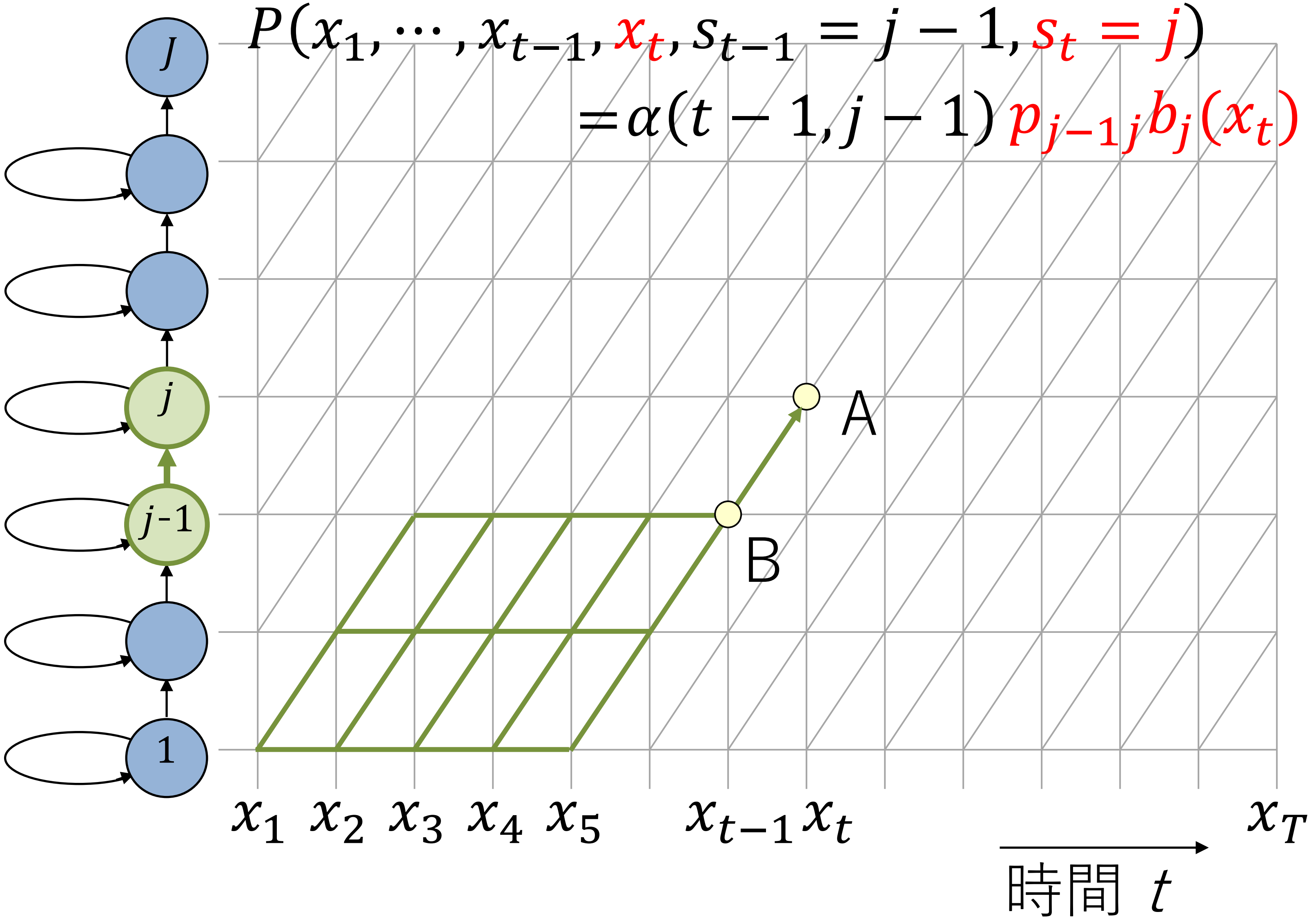
フォワードアルゴリズム

$$P(x_1, \cdots, x_{t-1}, s_{t-1} = j-1)$$
$$= \alpha(t-1, j-1)$$



フォワードアルゴリズム

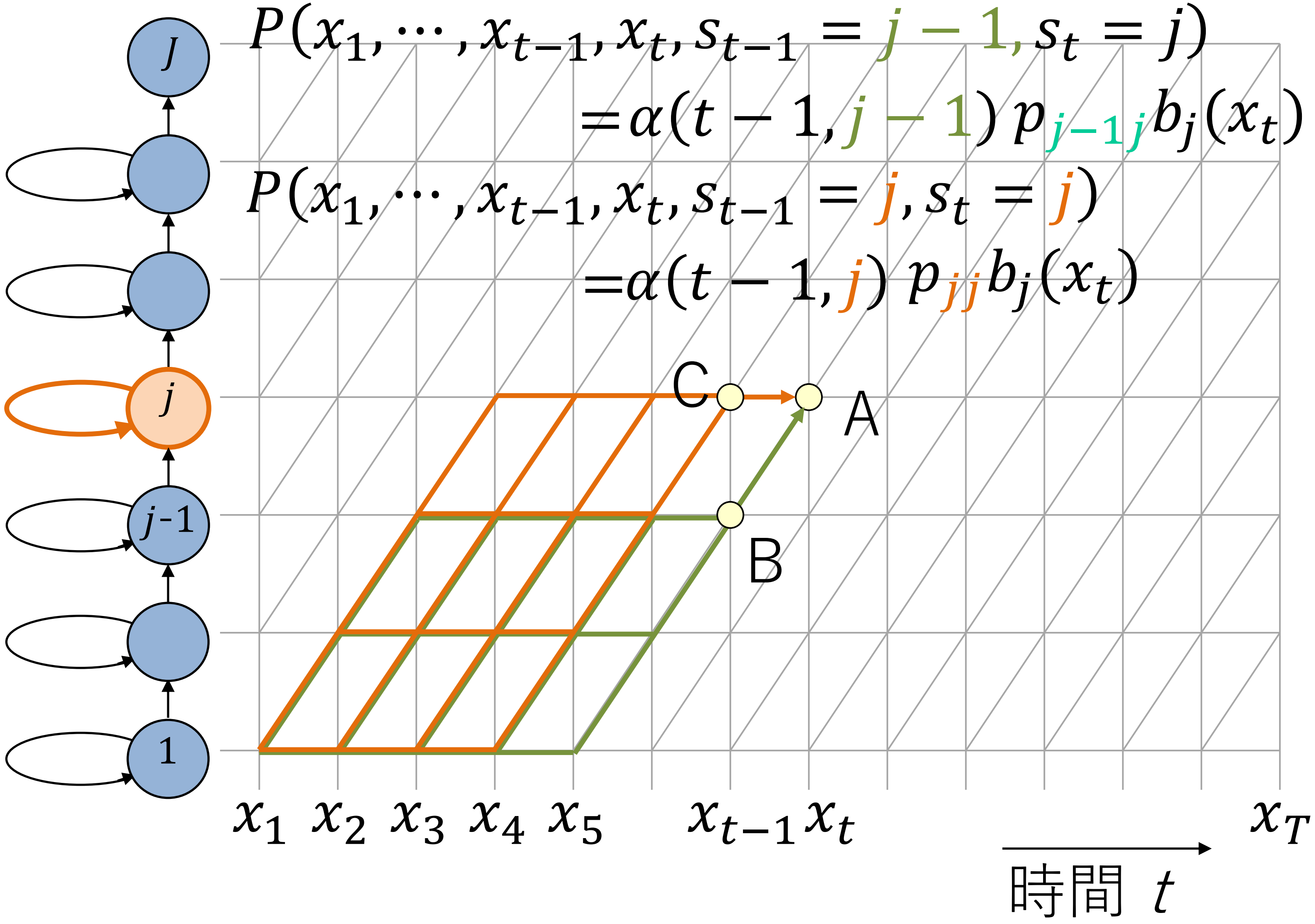
$$P(x_1, \cdots, x_{t-1}, x_t, s_{t-1} = j-1, s_t = j)$$
$$= \alpha(t-1, j-1)\, p_{j-1\,j} b_j(x_t)$$



フォワードアルゴリズム

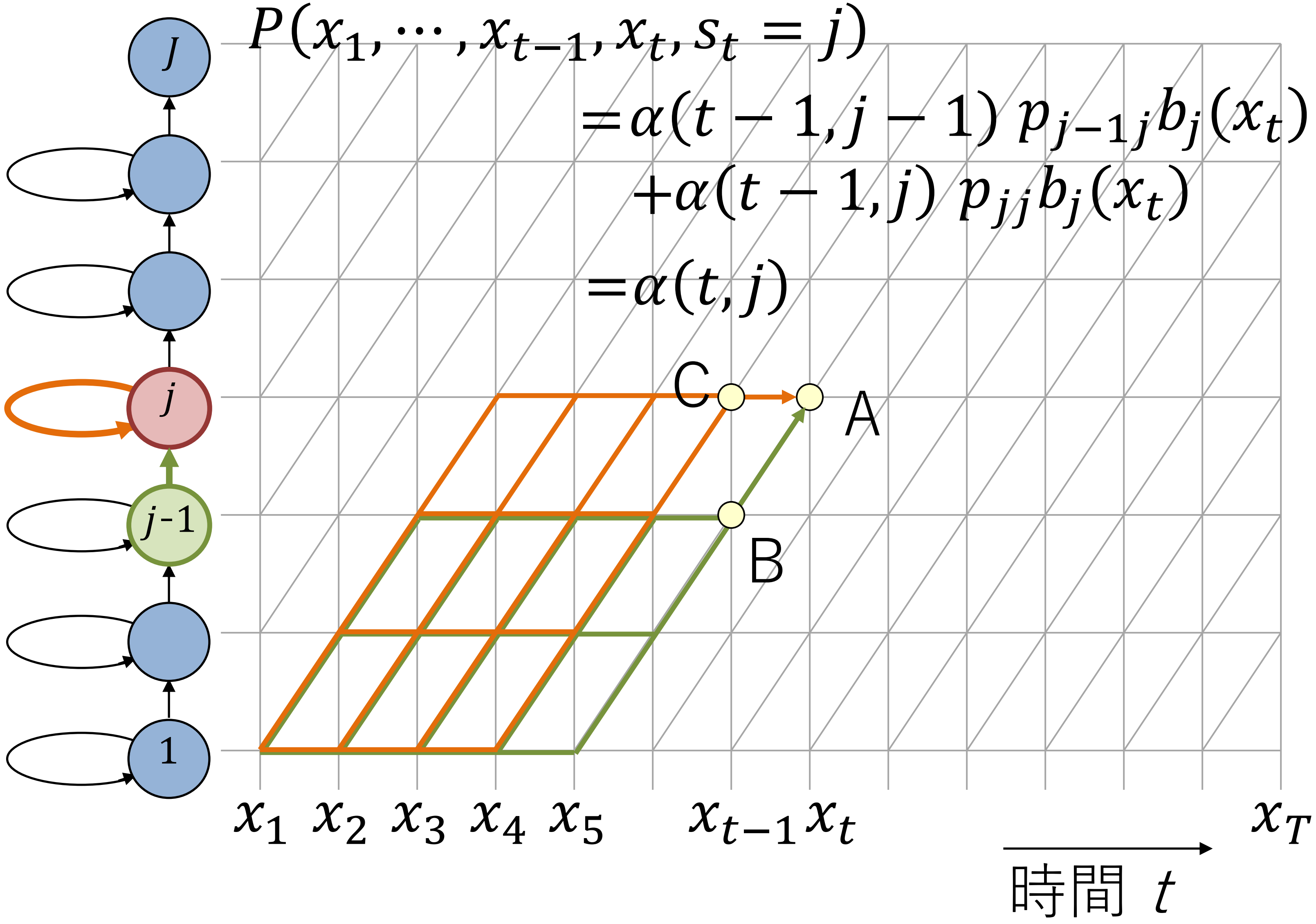
$$P(x_1,\cdots,x_{t-1},x_t,s_{t-1}=j-1,s_t=j)$$
$$=\alpha(t-1,j-1)\,p_{j-1\,j}b_j(x_t)$$

$$P(x_1,\cdots,x_{t-1},x_t,s_{t-1}=j,s_t=j)$$
$$=\alpha(t-1,j)\,p_{jj}b_j(x_t)$$



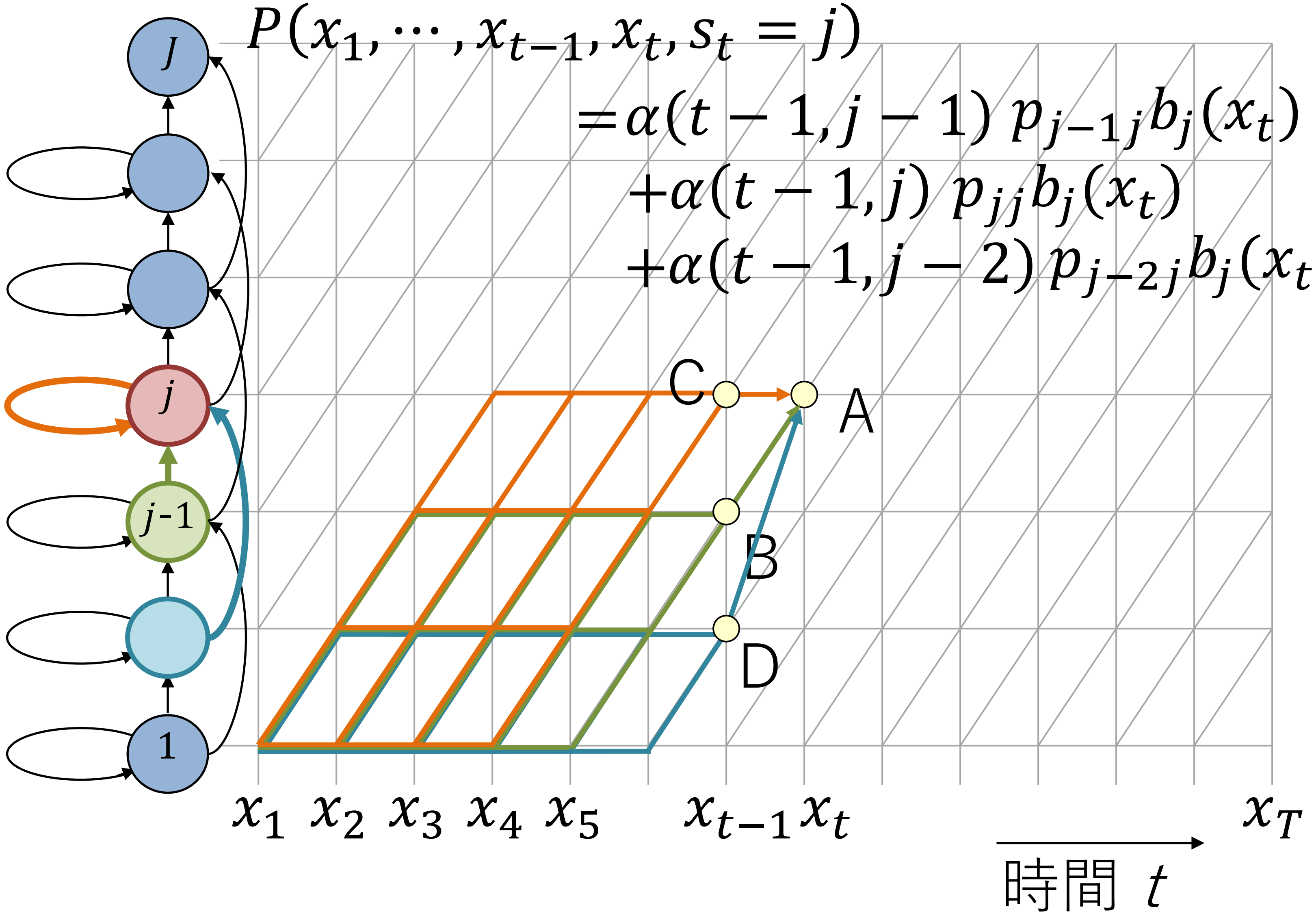
フォワードアルゴリズム

$$P(x_1, \cdots, x_{t-1}, x_t, s_t = j)$$
$$= \alpha(t-1, j-1)\, p_{j-1\,j} b_j(x_t)$$
$$+ \alpha(t-1, j)\, p_{jj} b_j(x_t)$$
$$= \alpha(t, j)$$



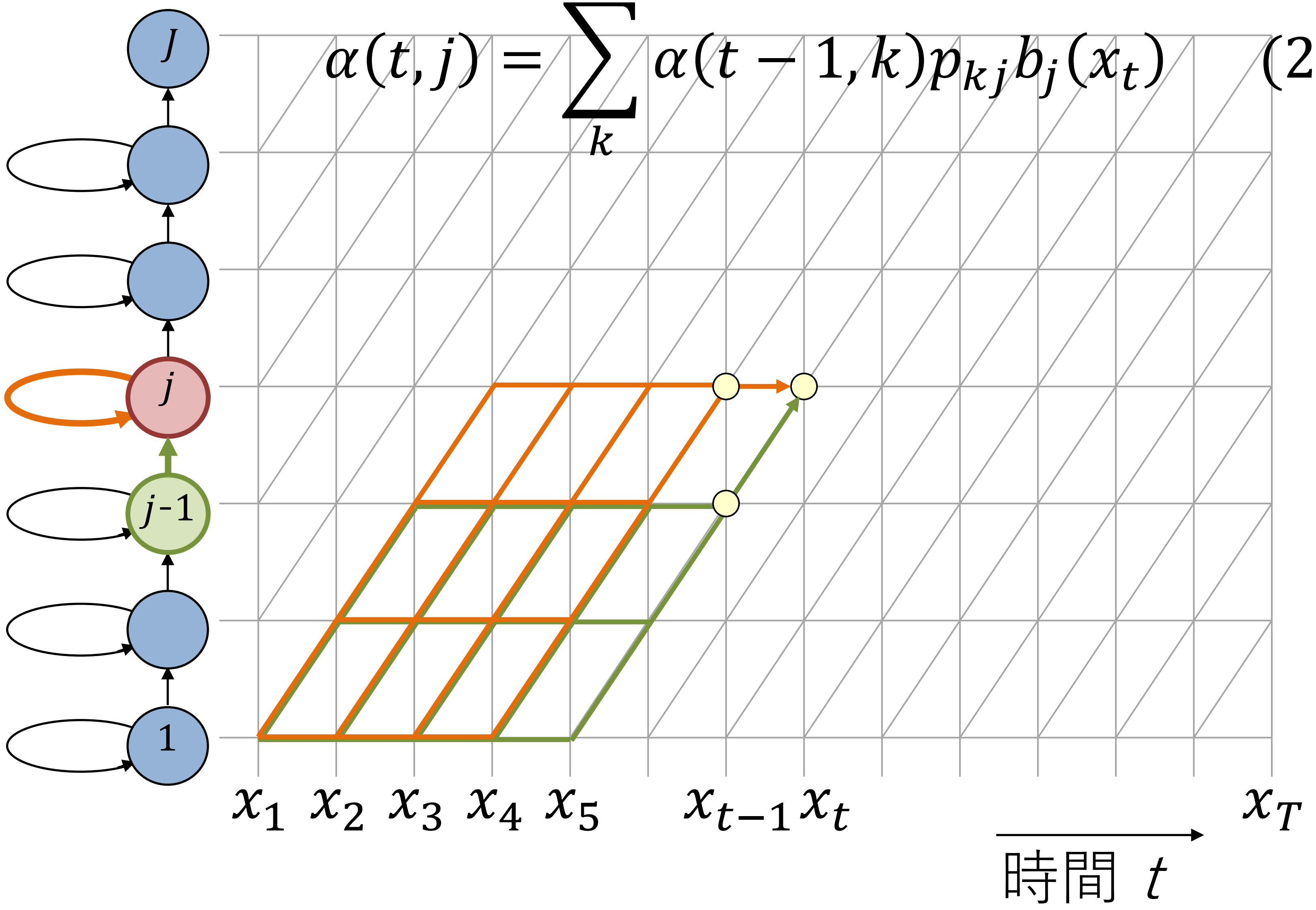
フォワードアルゴリズム

$$
\begin{aligned}
P(x_1, \cdots, x_{t-1}, x_t, s_t = j) \\
= \alpha(t-1, j-1)\, p_{j-1\,j} b_j(x_t) \\
+ \alpha(t-1, j)\, p_{jj} b_j(x_t) \\
+ \alpha(t-1, j-2)\, p_{j-2\,j} b_j(x_t) = \alpha(t, j)
\end{aligned}
$$

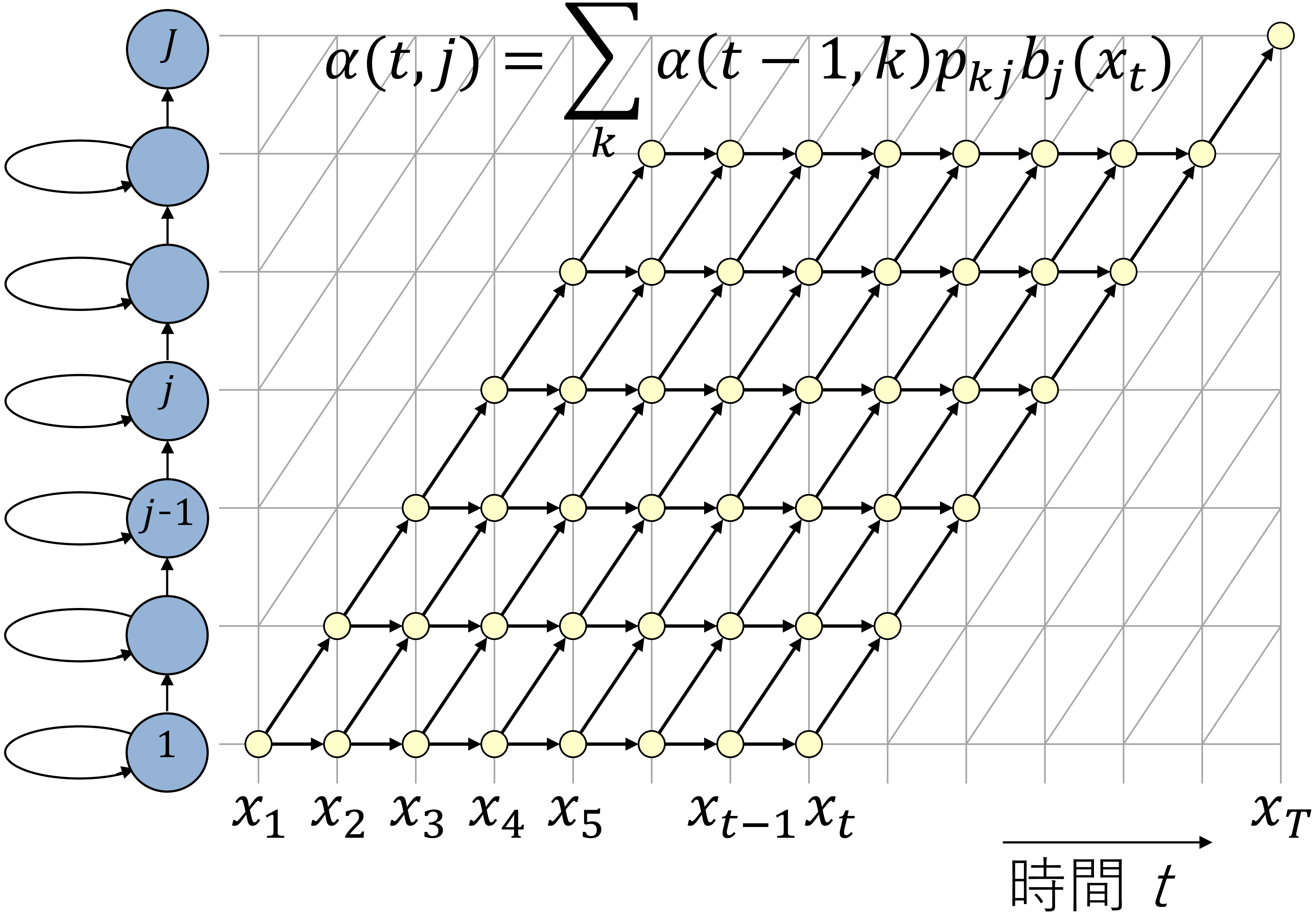


フォワードアルゴリズム

$$\alpha(t, j) = \sum_k \alpha(t-1, k) p_{kj} b_j(x_t) \quad (2)$$



フォワードアルゴリズム



フォワードアルゴリズム

求める $x_1 \sim x_T$ の出力確率は,

$\alpha(t,j)$ を，t=1から順にT まで求めることによって得た$\alpha(T,j)$ を用いて,

$$P(x_1, x_2, \cdots, x_T) = \sum_{j \in E} \alpha(T,j) \qquad (3)$$

と表される。

E : 最終状態として許される状態の集合

ビタビアルゴリズム

最適な状態遷移が生じた場合の出力確率を求めるアルゴリズムをビタビアルゴリズムと呼ぶ。

$\alpha(t,j)$を，時刻 t までに，x_1～x_t を出力して状態 j に至る単一の状態列が与える最大確率とするとき，

$$\alpha(t,j) = \max_k \left(\alpha(t-1,k)p_{kj}b_j(x_t)\right) \qquad (4)$$

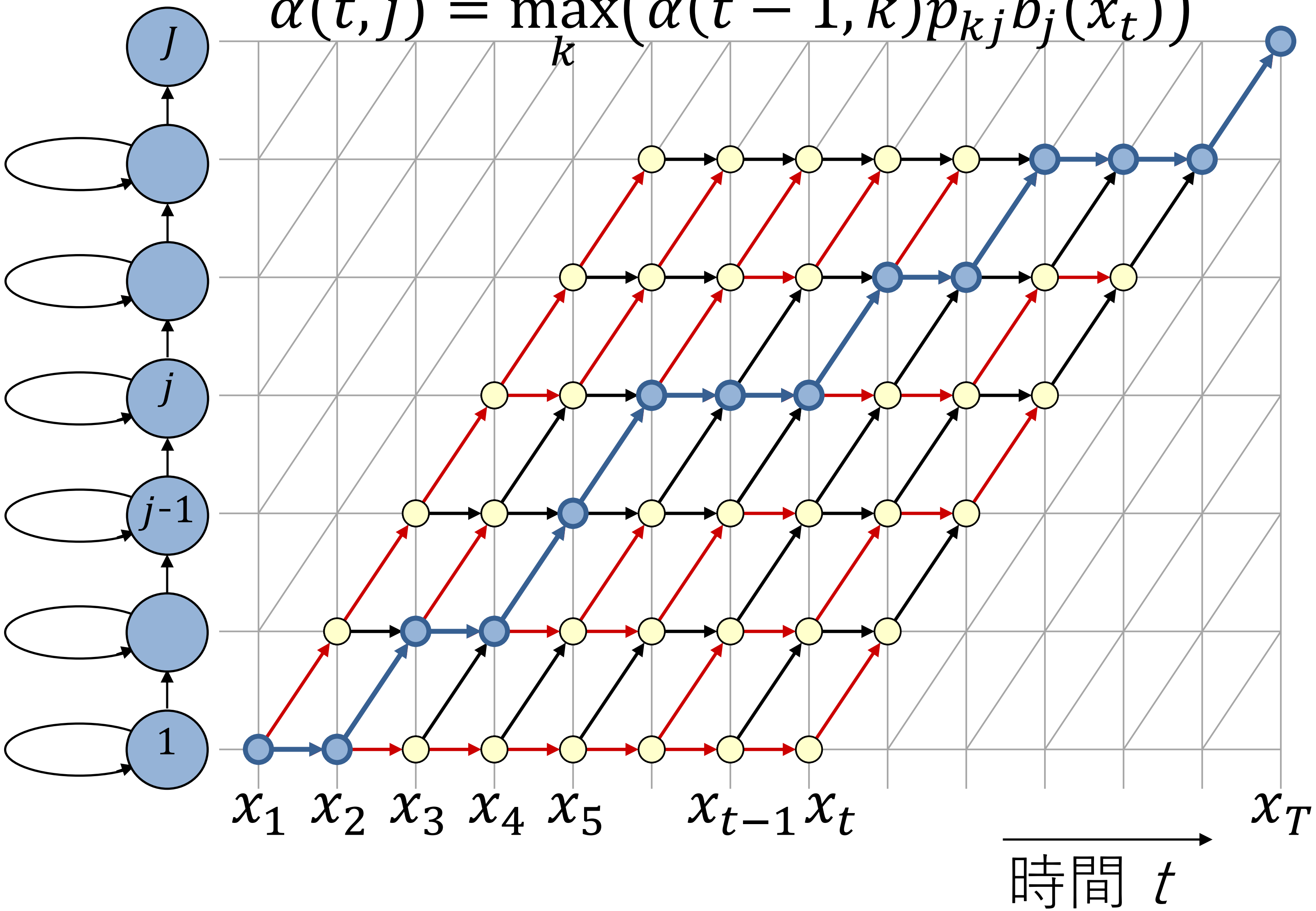
と書ける。よって，求める単一状態列が与える x_1～x_T の最大出力確率は，$\alpha(t,j)$ を，$t=1$から順にTまで求めることによって得た$\alpha(T,j)$を用いて，

$$P(x_1,x_2,\cdots,x_T) = \max_{j\in E} \alpha(T,j) \qquad (5)$$

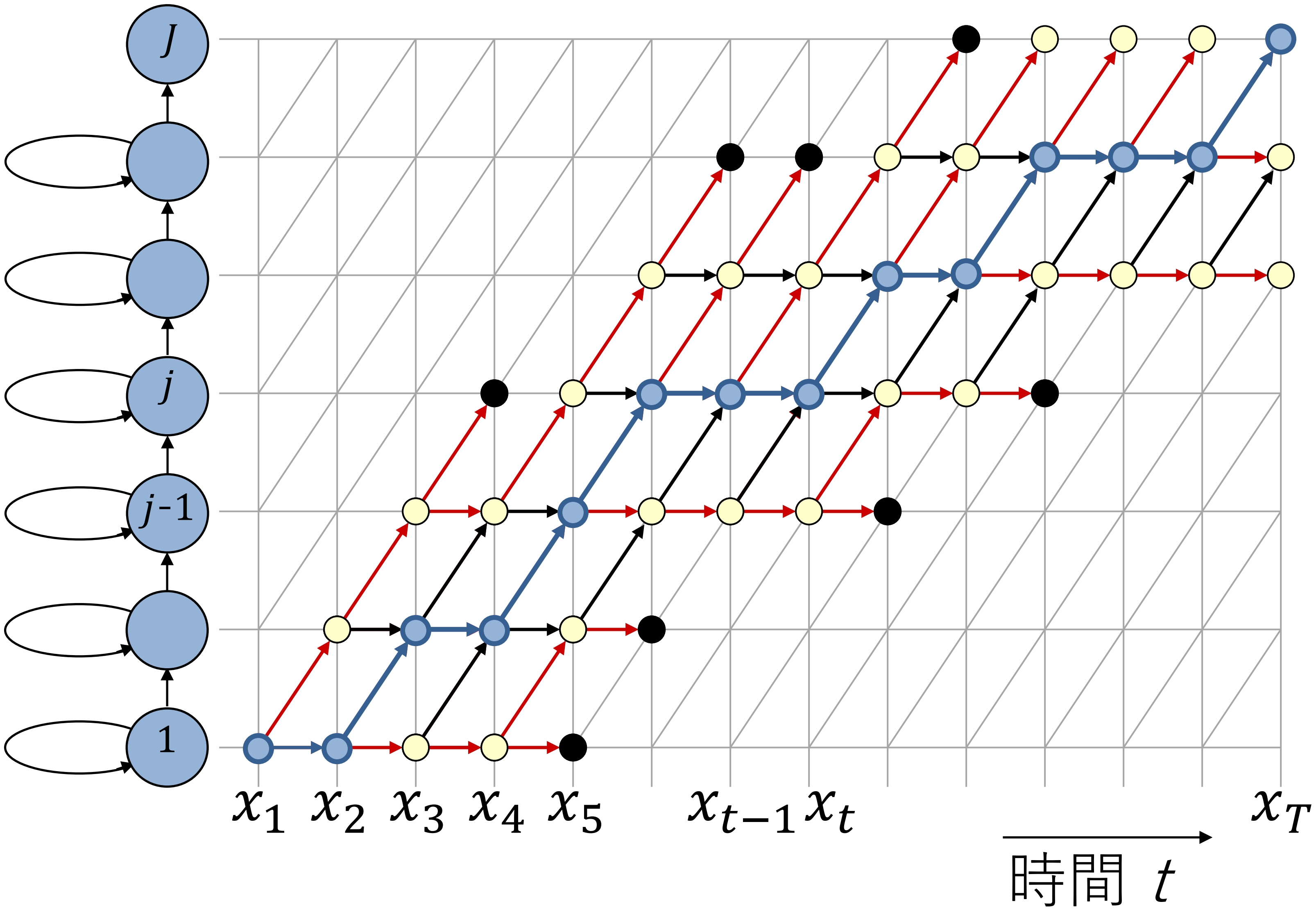
と表される。

ビタビアルゴリズム

$$\alpha(t,j) = \max_k \left(\alpha(t-1,k)p_{kj}b_j(x_t)\right)$$



ビームサーチ



§ HMMのパラメタ推定

マルコフモデルのように，データ$\boldsymbol{x}_t$がどの状態から出力されたかが分かれば，状態毎に出力した記号を集めて統計をとることで，各状態の出力確率分布を定めることができる。

しかし，HMMでは，データと状態との対応がとれない。

＝ 各データがどの状態から出力したかを仮定する必要

＝ 潜在変数が存在する

→ EMアルゴリズムの導入

HMMのパラメタ推定

EMアルゴリズム

1. モデルパラメタの初期値，$\mathcal{M} = \{\boldsymbol{\pi}, P, \boldsymbol{b}, E\}$を定める。
2. $\mathcal{M}$ をもとに，データ $\boldsymbol{x}_t$ が状態 j から出力する確率 $\psi(t,j)$ を求める。
3. この結果を「$\boldsymbol{x}_t$は，状態 jから $\psi(t,j)$回出力した」と解釈することで，統計をとりなおして，モデルの更新値 $\widehat{\mathcal{M}} = \{\widehat{\boldsymbol{\pi}}, \widehat{P}, \widehat{\boldsymbol{b}}, E\}$ を求める。
4. $\mathcal{M}$ を $\widehat{\mathcal{M}}$ で置き換えて,2,3,4 を繰り返す。

HMMのパラメタ推定

$\alpha(t,j)$ ：時刻 t までに $x_1 \sim x_t$を出力して状態 j に至る確率と定義する。

$$\alpha(t,j) = P(x_1, \cdots, x_t, s_t = j)$$

$\beta(t,j)$: 時刻 t に状態 j にあり，以後$x_{t+1} \sim x_T$ を出力して最終状態に至る確率と定義する。

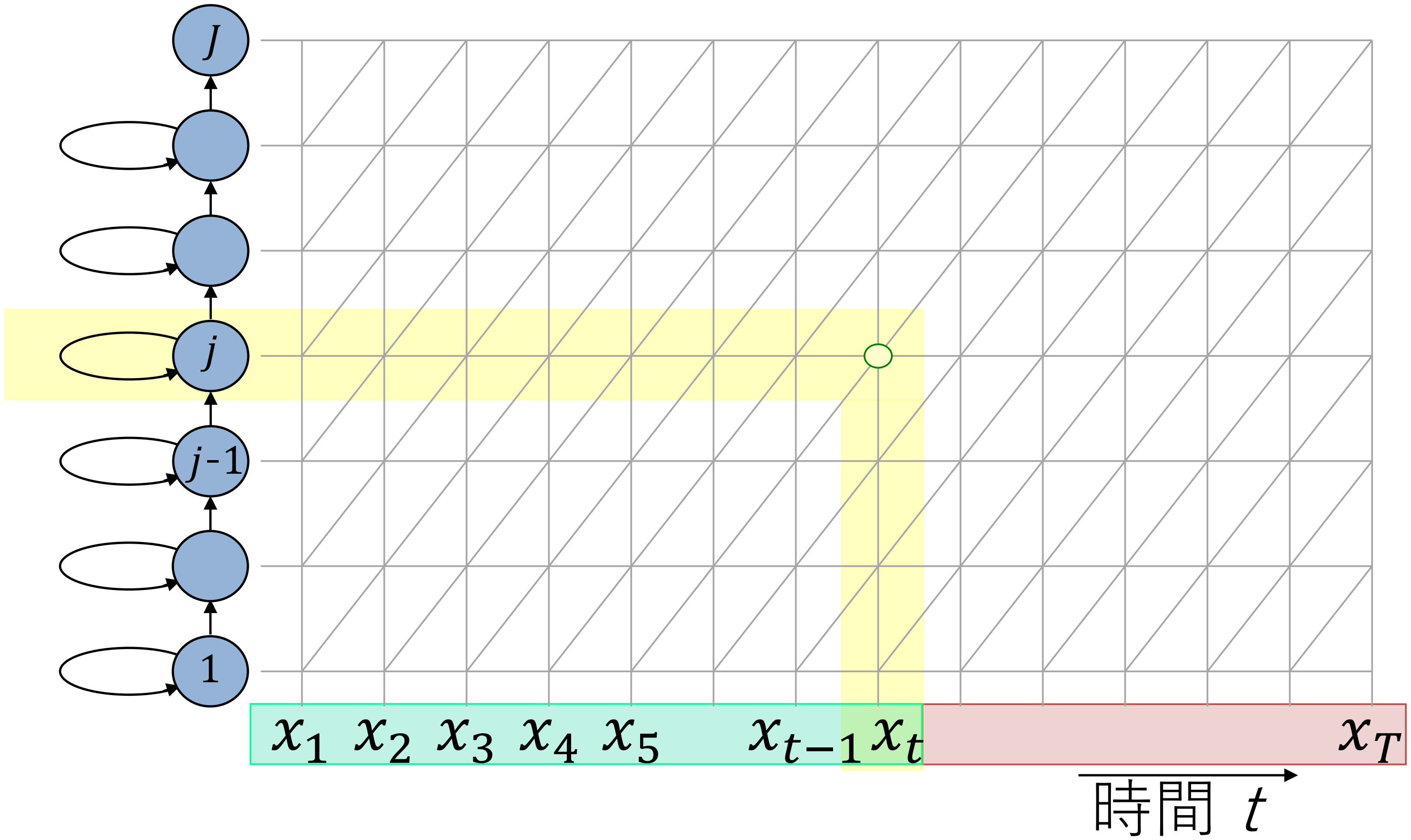
$\beta(t,j)$は，

$$\beta(t,j) = P(x_{t+1}, \cdots, x_T, s_t = j)$$

$$\beta(t,j) = \sum_k p_{jk} b_k(x_{t+1}) \beta(t+1,k)$$

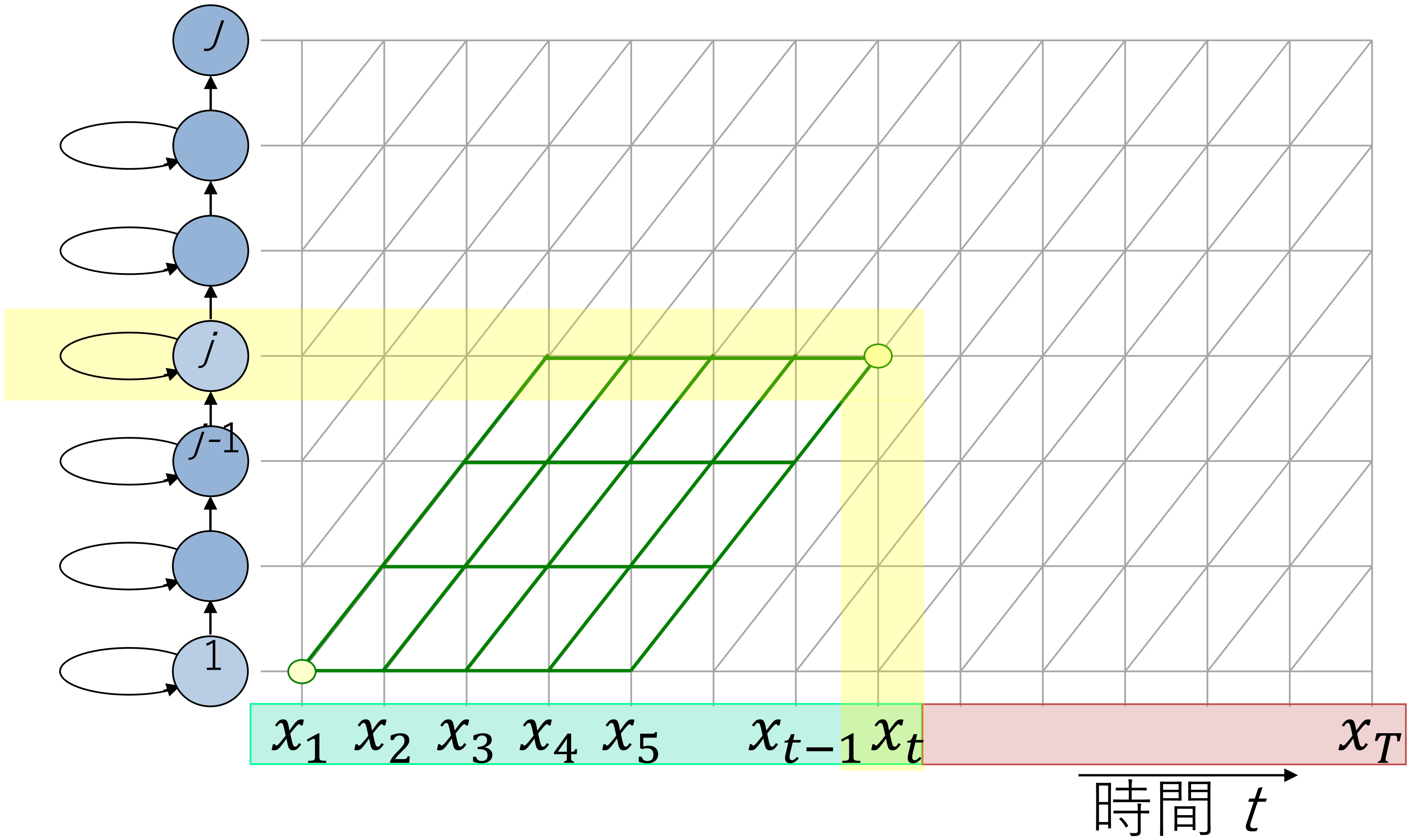
と書ける。よって， $t = T$ から順に$T-1, T-2$ ・・・と漸化的に求めることができる。

HMMのパラメタ推定



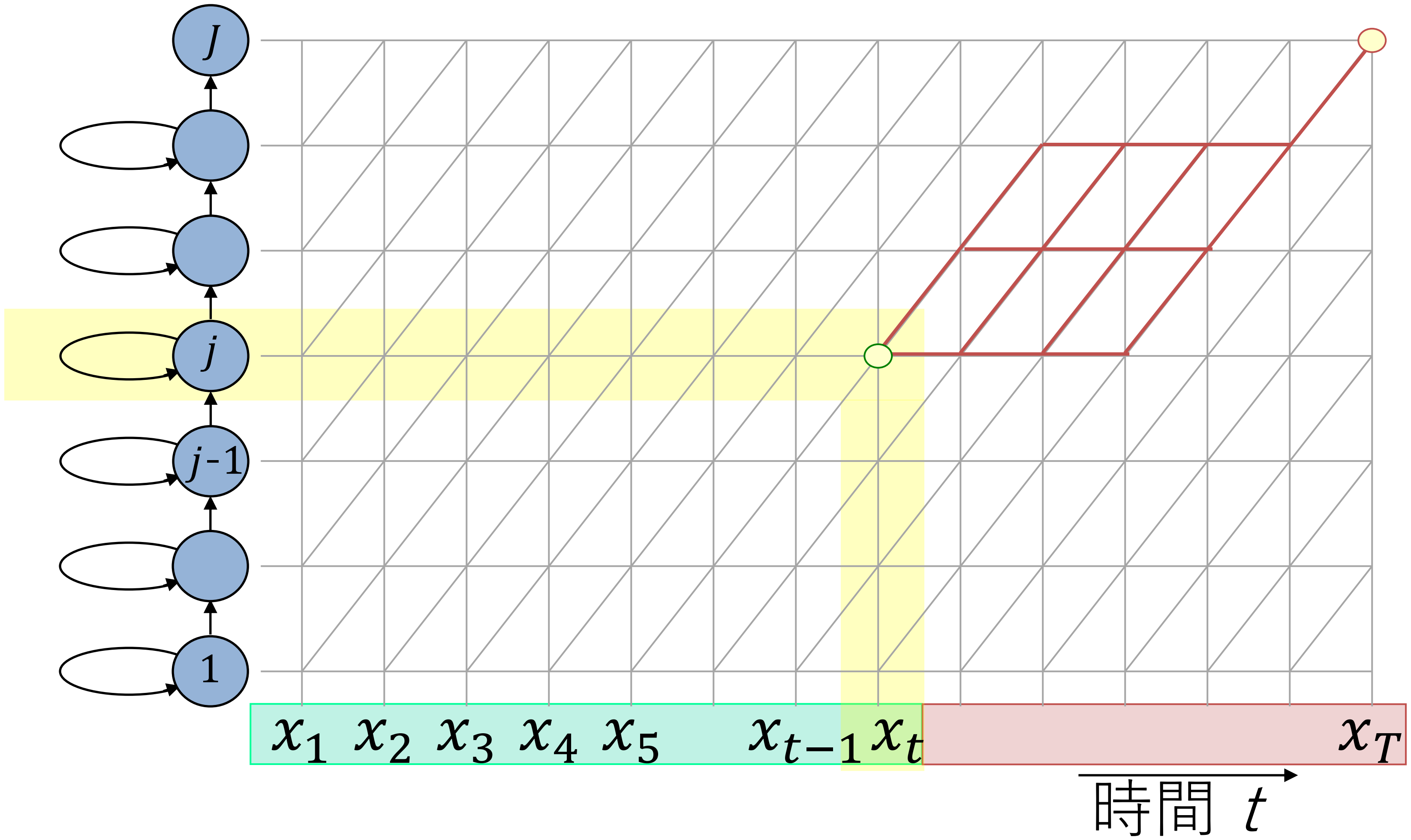
HMMのパラメタ推定

$$\Pr(S_t = j, x_1, \cdots, x_t) \qquad = \alpha(t, j)$$



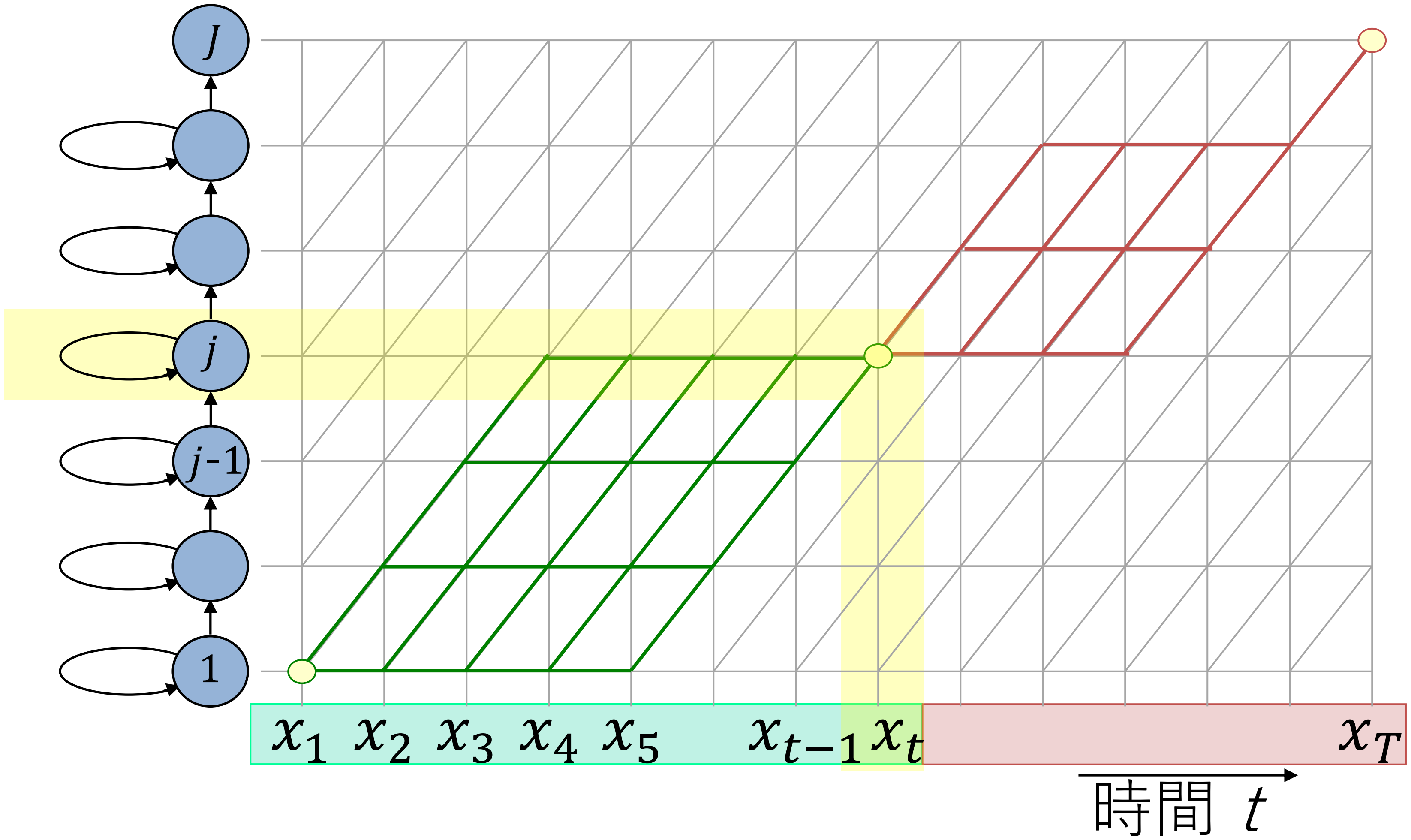
HMMのパラメタ推定

$$\Pr(S_t = j, \quad x_{t+1}, \cdots, x_T) = \quad \beta(t, j)$$



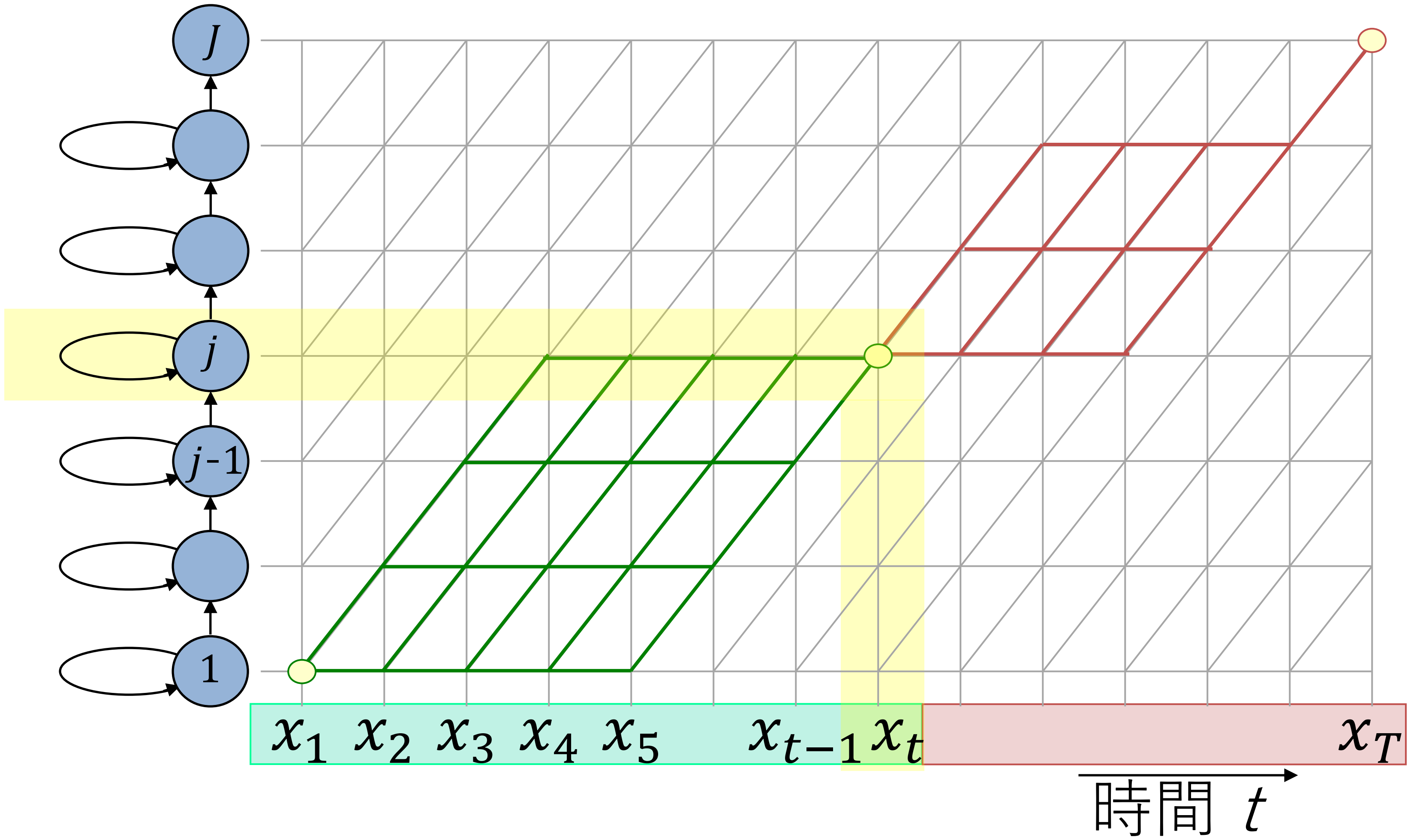
HMMのパラメタ推定

$$\Pr(S_t = j, x_1, \cdots, x_t, x_{t+1}, \cdots, x_T) = \alpha(t, j) \cdot \beta(t, j)$$



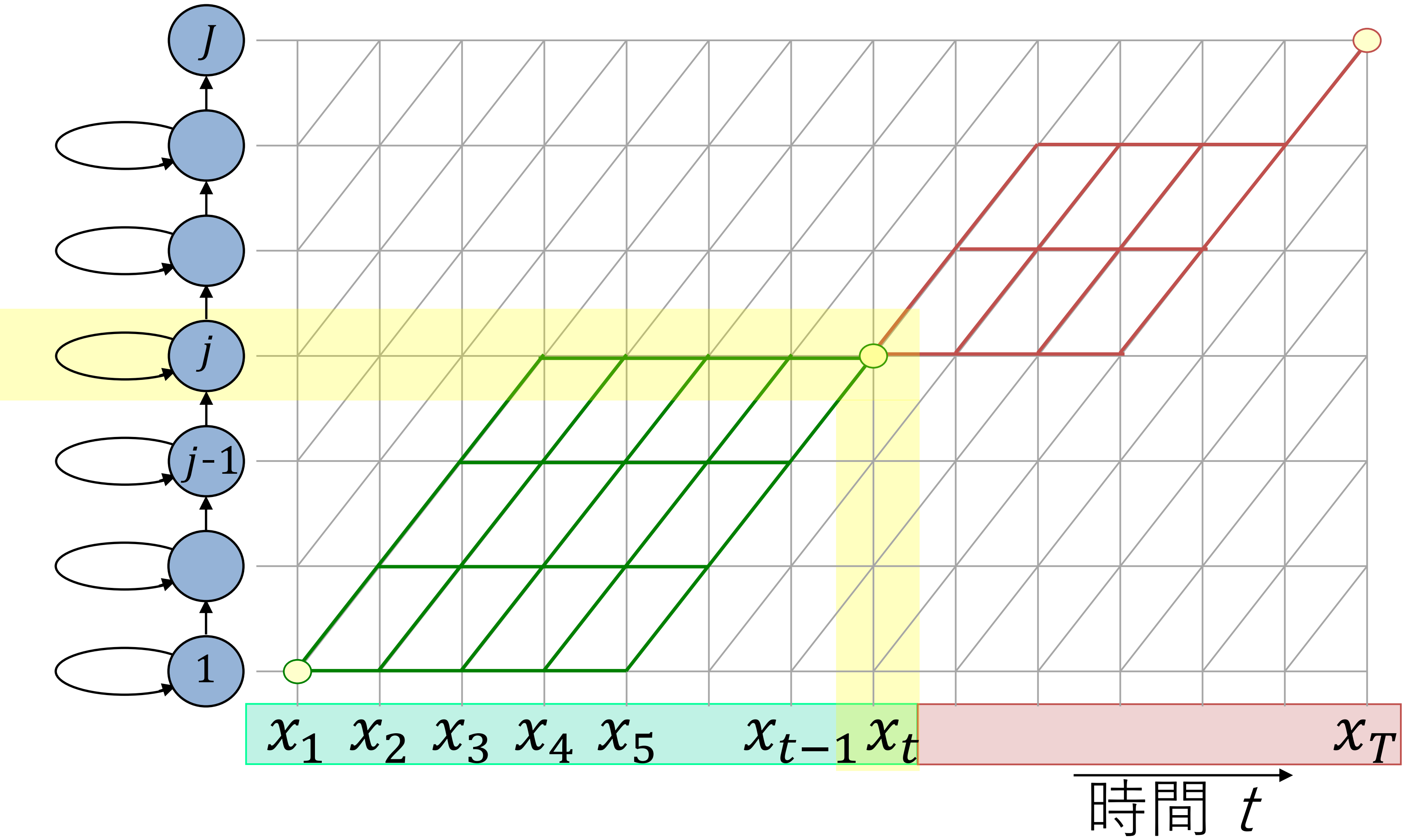
HMMのパラメタ推定

$$\Pr(S_t = j, X) = \alpha(t,j) \cdot \beta(t,j)$$



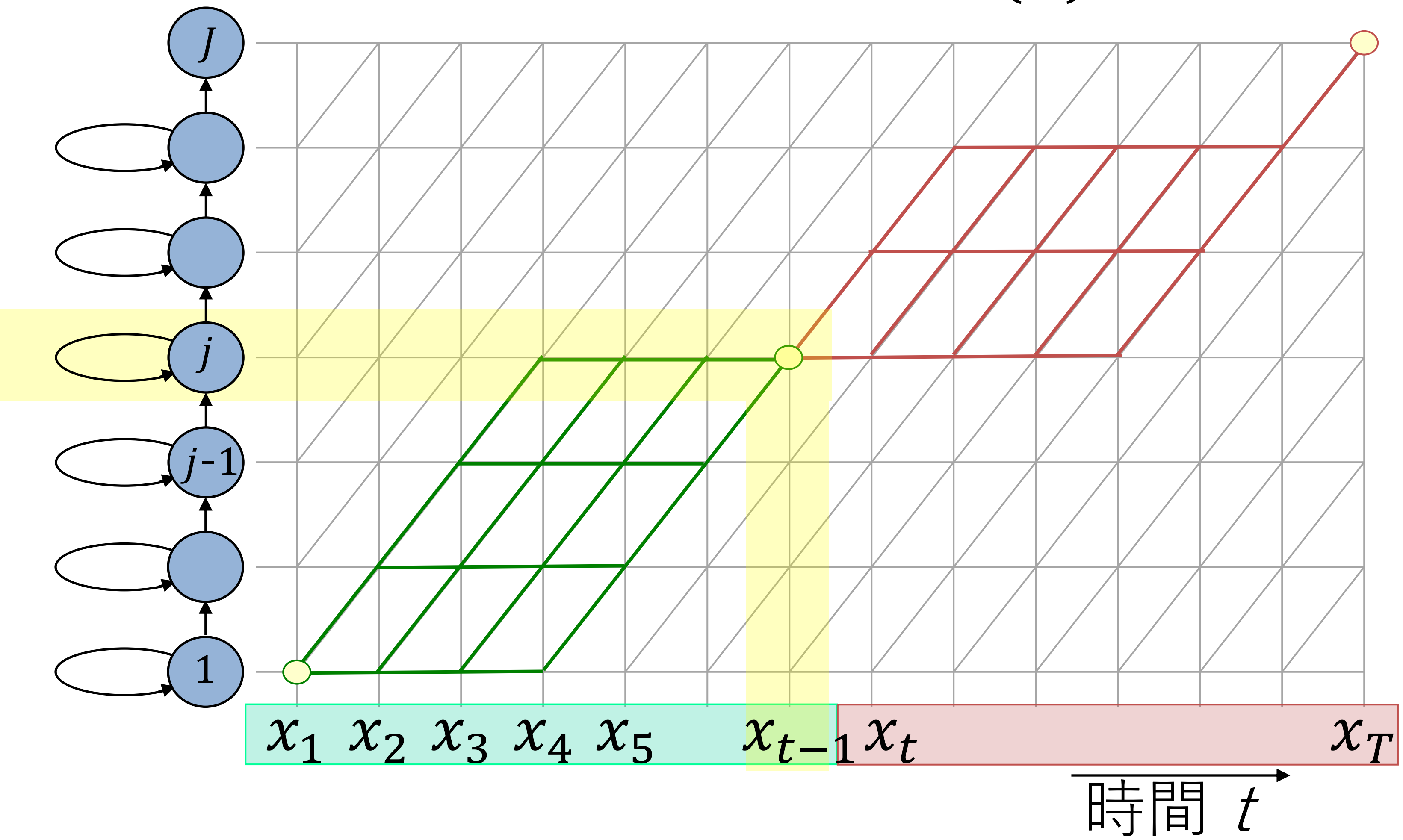
HMMのパラメタ推定

$$\Pr(S_t = j|X) = \frac{\alpha(t,j)\cdot\beta(t,j)}{\Pr(X)}$$



HMMのパラメタ推定

$$\Pr(S_{t-1} = j|X) = \frac{\alpha(t-1, j) \cdot \beta(t-1, j)}{\Pr(X)}$$



HMMのパラメタ推定

$$\psi(t,j) = P(s_t = j|X)$$

とおくと，$\psi(t,j)$ は，x_t が状態 j から出力した確率と考えることができる。

$b_j(x)$が平均値 μ 分散 σ^2 の正規分布とすれば，μ, σ^2 の更新値平均値 $\hat{\mu}$分散 $\hat{\sigma}^2$は，

$$\hat{\mu} = \frac{\sum_{t=1}^{T} \psi(t,j) x_t}{\sum_{t=1}^{T} \psi(t,j)}$$
$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \psi(t,j)(x_t - \hat{\mu})^2}{\sum_{t=1}^{T} \psi(t,j)} \quad (6)$$

と求められる。

HMMのパラメタ推定

$\xi(t,i,j)$を $X = \boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_T$ が出力される条件で，時刻$t-1$からtの間に，状態が i から j に遷移する確率とすれば，

$$\xi(t,i,j) = P(s_{t-1} = i, s_t = j|X)$$
$$= \frac{P(s_{t-1} = i, s_t = j, X)}{P(X)}$$
$$= \frac{\alpha(t-1,i)p_{ij}b(x_t)\beta(t,j)}{P(X)} \qquad (7)$$

状態遷移確率の更新値は，以下のようになる。

$$\widehat{p_{ij}} = \frac{\sum_{t=1}^{T}\xi(t,i,j)}{\sum_j \sum_{t=1}^{T}\xi(t,i,j)} \qquad (8)$$

まとめ

- 音声認識に代表される系列データは，特徴ベクトルが時間変化することが特徴であり，これを扱う枠組みが必要となる。
- 系列データのパターン認識を，生成モデル立場からアプローチするとき，そこで必要とされる確率モデルには，HMMが用いられることが多い。
- Left to Right 型HMMで音声をモデル化することは，音声を区分定常過程として捉えることに相当する。
- HMMの確率計算には，フォーワードアルゴリズム，ビタビアルゴリズム，ビームサーチなどが用いられる。
- HMMのパラメタ計算には，EMアルゴリズムが用いられる。