東京大学工学部 4年生 夏学期

応用音響学
第5回 (5/18)

猿渡　洋

東京大学大学院情報理工学系研究科
システム情報学専攻
hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

2018年度講義スケジュール

前半（猿渡担当）

- 4/06: 第1回
- 4/13: 第2回
- 4/20は休講予定
- 4/27: 第3回
- 5/11: 第4回
- 5/18: 第5回
- 5/25: 第6回

後半（牧野先生担当）

- 6/08: 第7回
- 6/15は休講予定
- 6/22: 第8回
- 6/29: 第9回
- 7/06: 第10回
- 7/13: 第11回

- 7/20: 学期末試験（予定）

講義資料と成績評価

- 講義資料
  - http://www.sp.ipc.i.u-tokyo.ac.jp/~saruwatari/
    （システム情報第一研究室からたどれるようにしておきます）
- 成績評価
  - 出席点
  - 学期末試験

本日の話題

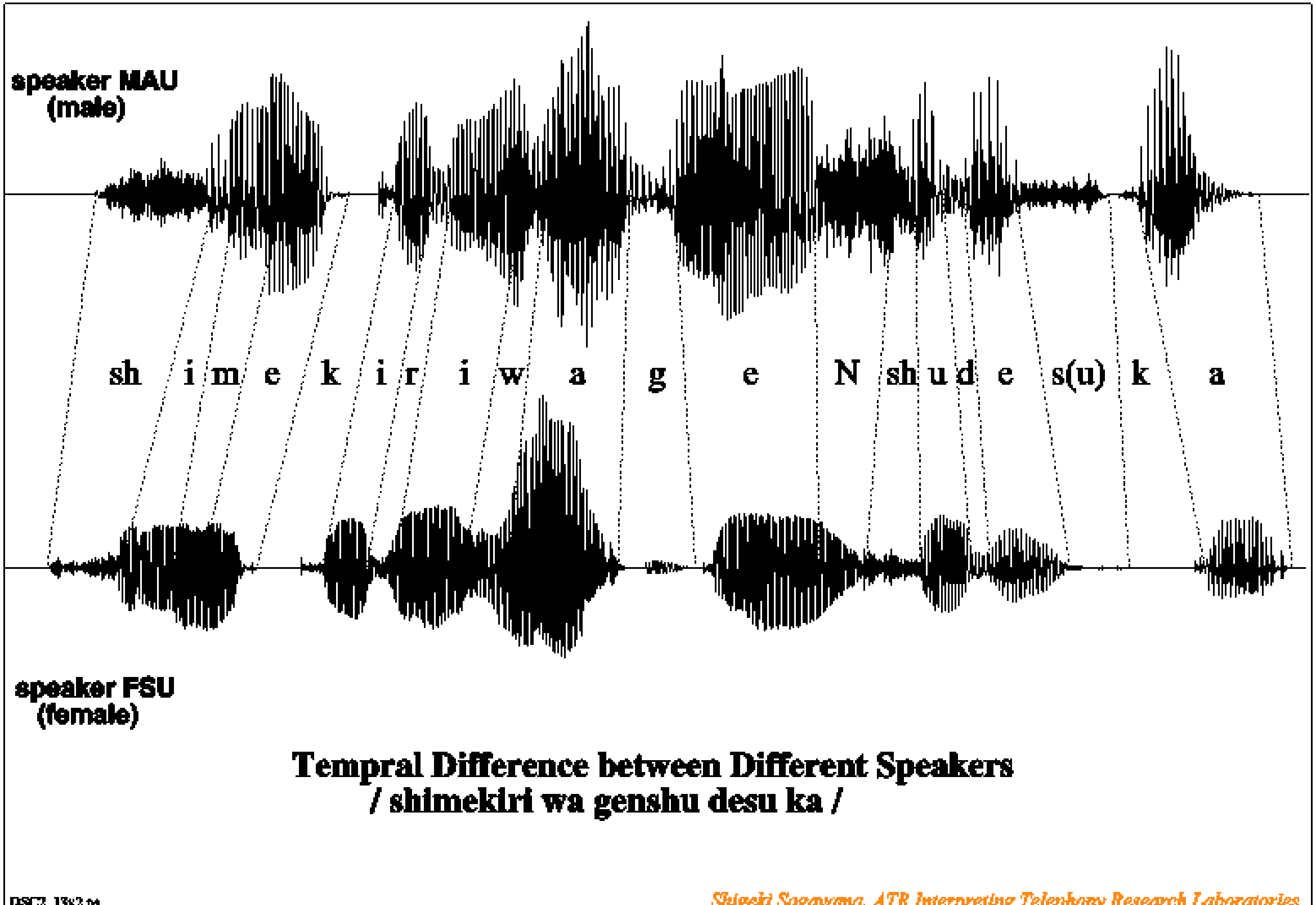
- 確率モデル（生成モデル）による
  動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

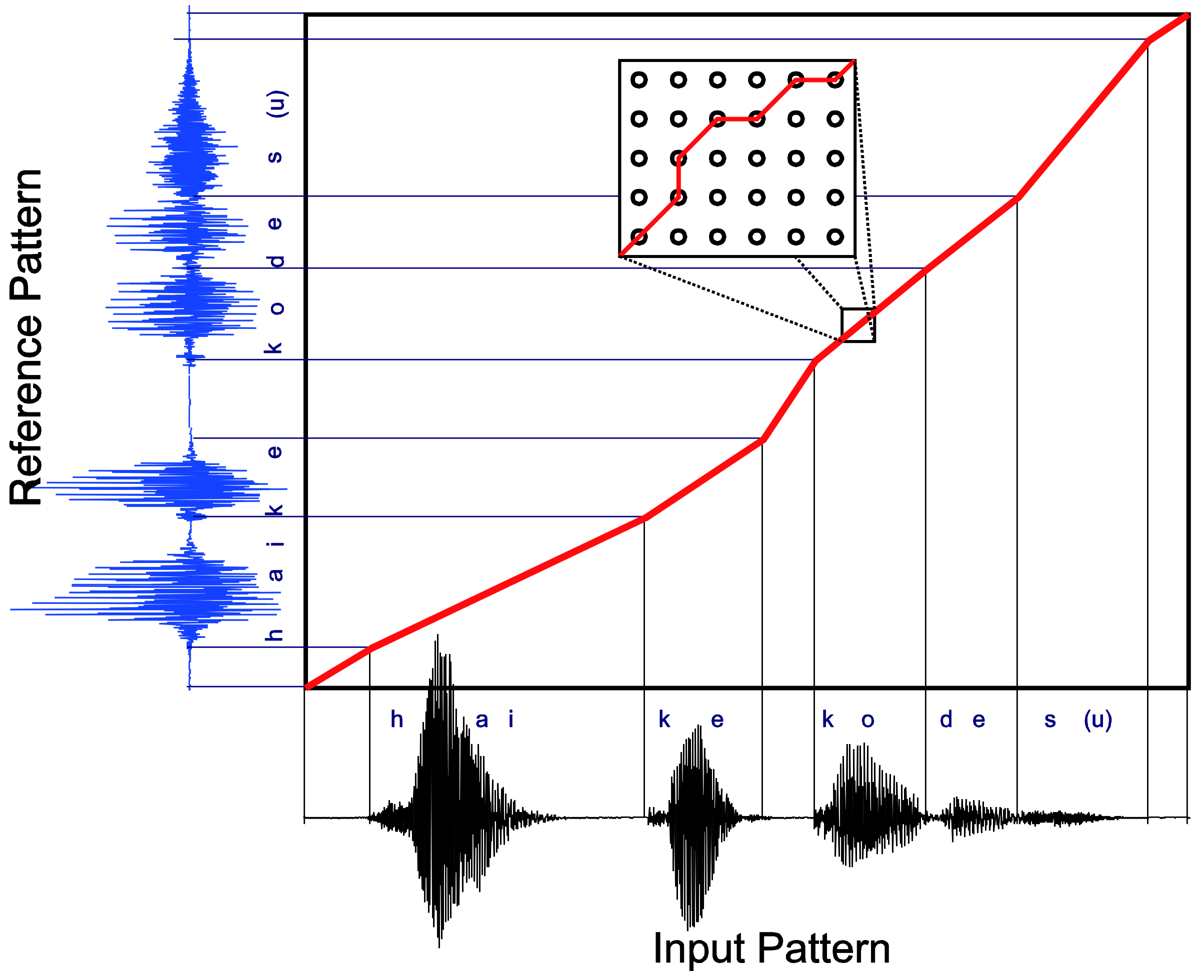
DPマッチングによる音声認識

■音声の発話リズムは話者・発話ごとに変動する



DPマッチングによる音声認識

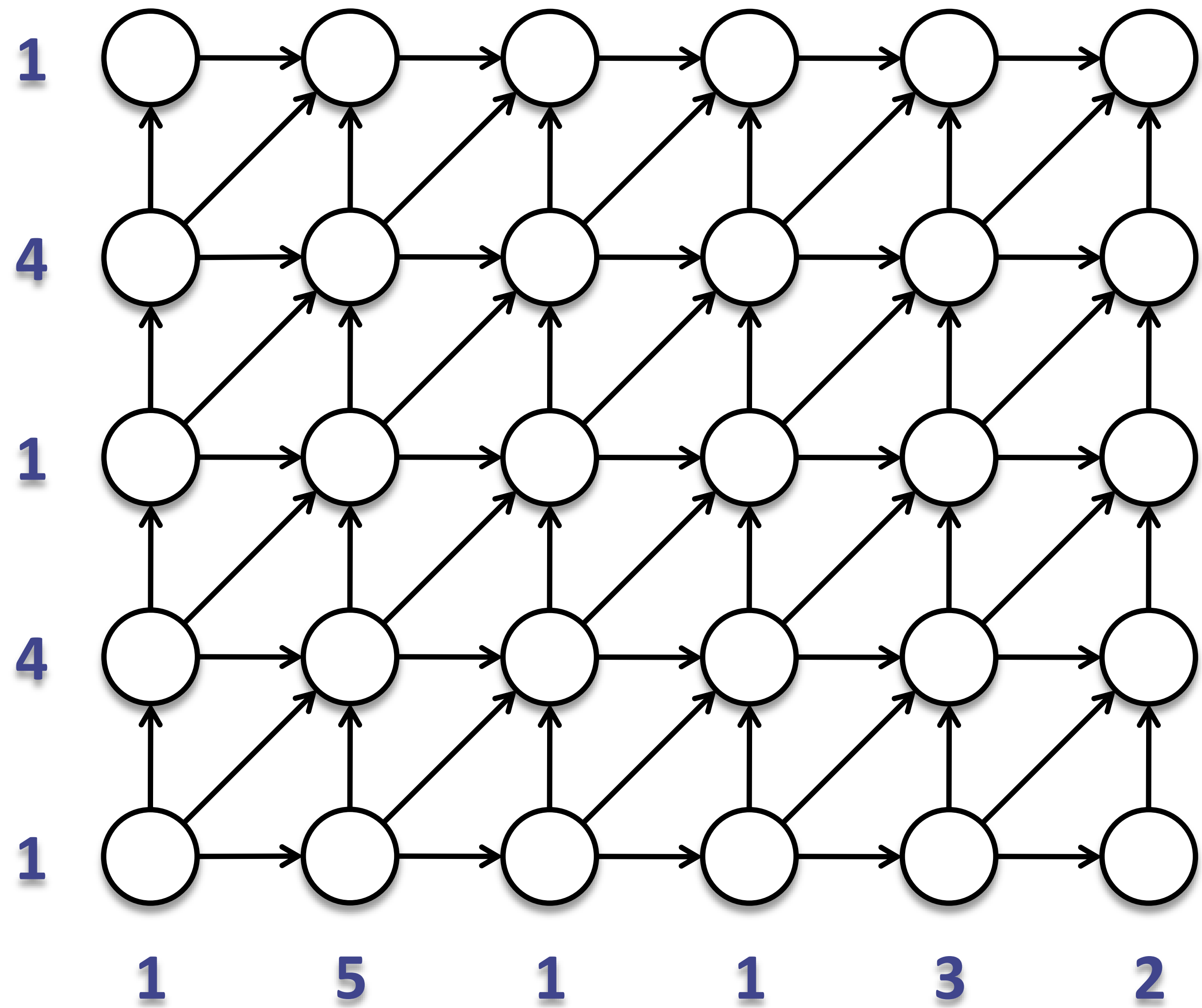
- パターンマッチングによる音声認識
  - 入力特徴量系列との距離が最小となる参照特徴量系列を認識結果とする
  ⇒発話リズムの違いを「正規化」してマッチングしたい



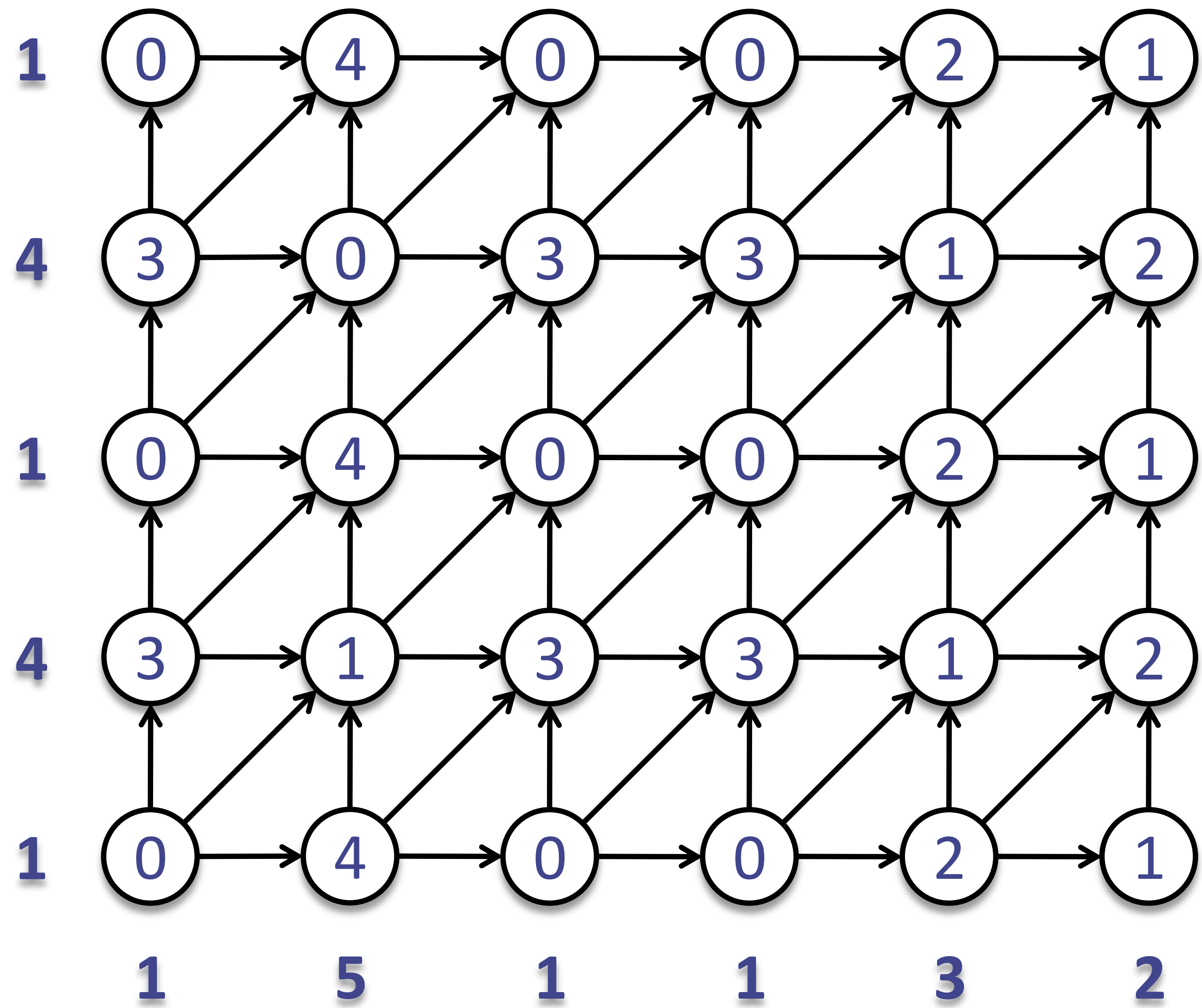
DPマッチングとは

- DTW : Dynamic Time Warping
  DP (Dynamic Programming)マッチング
  - 日本で育った手法（独立にソ連から発表あり）
  - 隠れマルコフモデル(HMM)の出現まで音声認識手法の主流
  - 日本電気中央研究所の迫江および千葉による発明
- DPの原理: 最適化問題が部分問題に分解できるので、
  効率の高い最適経路探索問題が解ける
  - 標準パターンと入力パターンの間を非線形対応づけ
  - DP(動的計画法)により時間の対応づけを求める
  - 傾斜制限:
    - 音声入力各フレームは標準パターンのどれかのフレームに必ず対応付ける。
      同一フレームには2フレームまでは対応付けて良い。
  - 窓制限:
    - マッチング経路は対角線から一定幅以内になくてはならない。

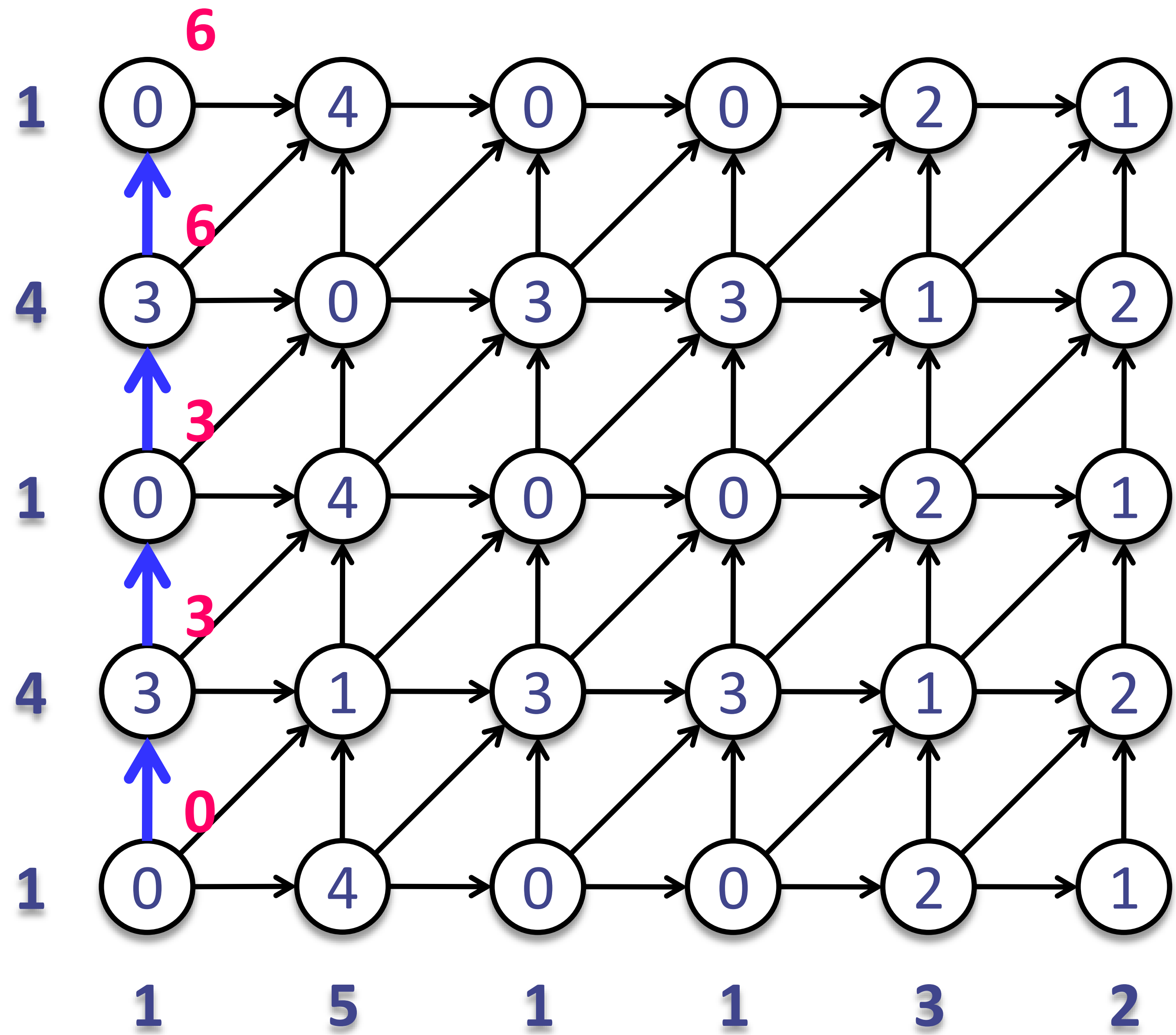
DPマッチングのアルゴリズム



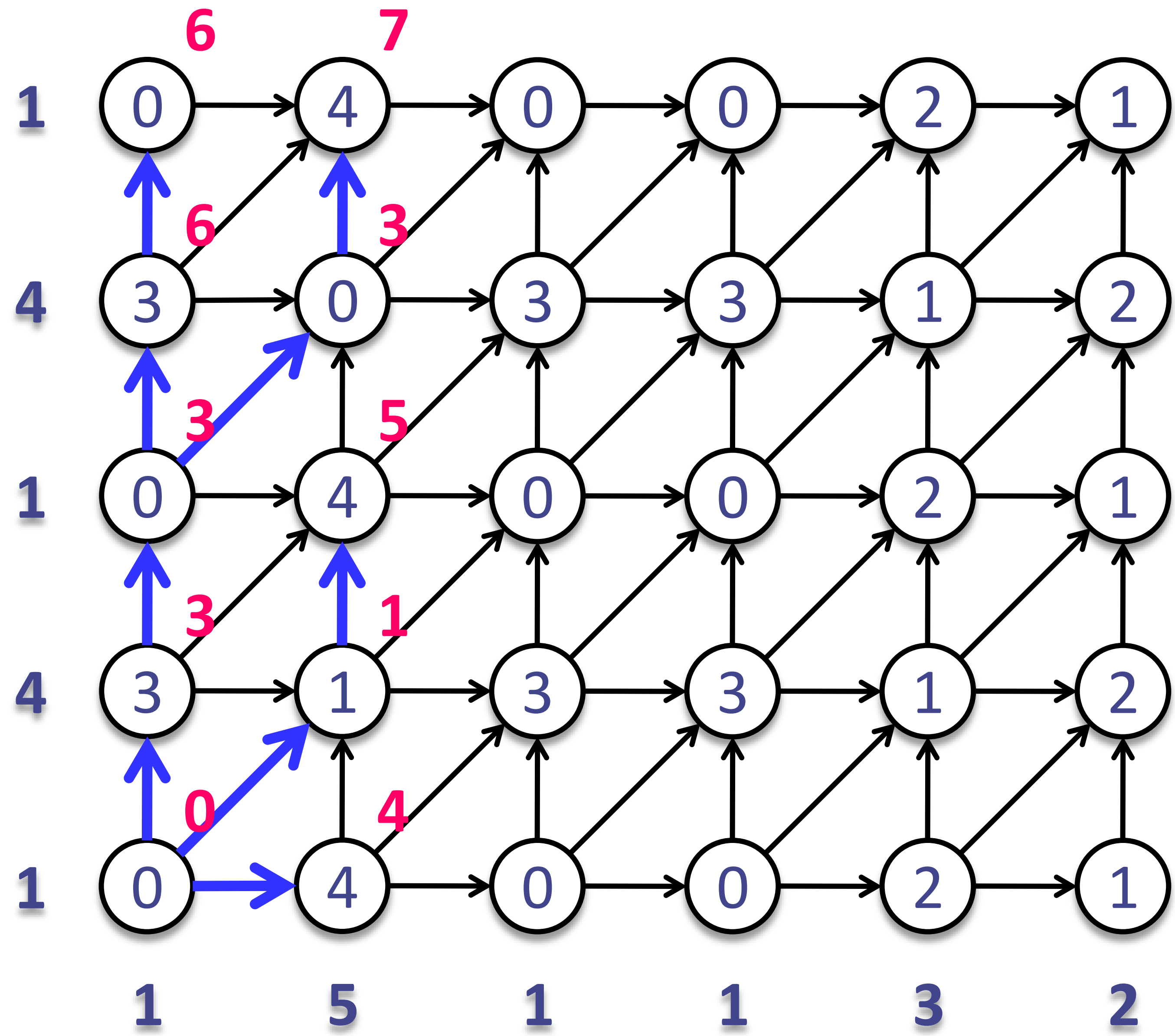
DPマッチングのアルゴリズム



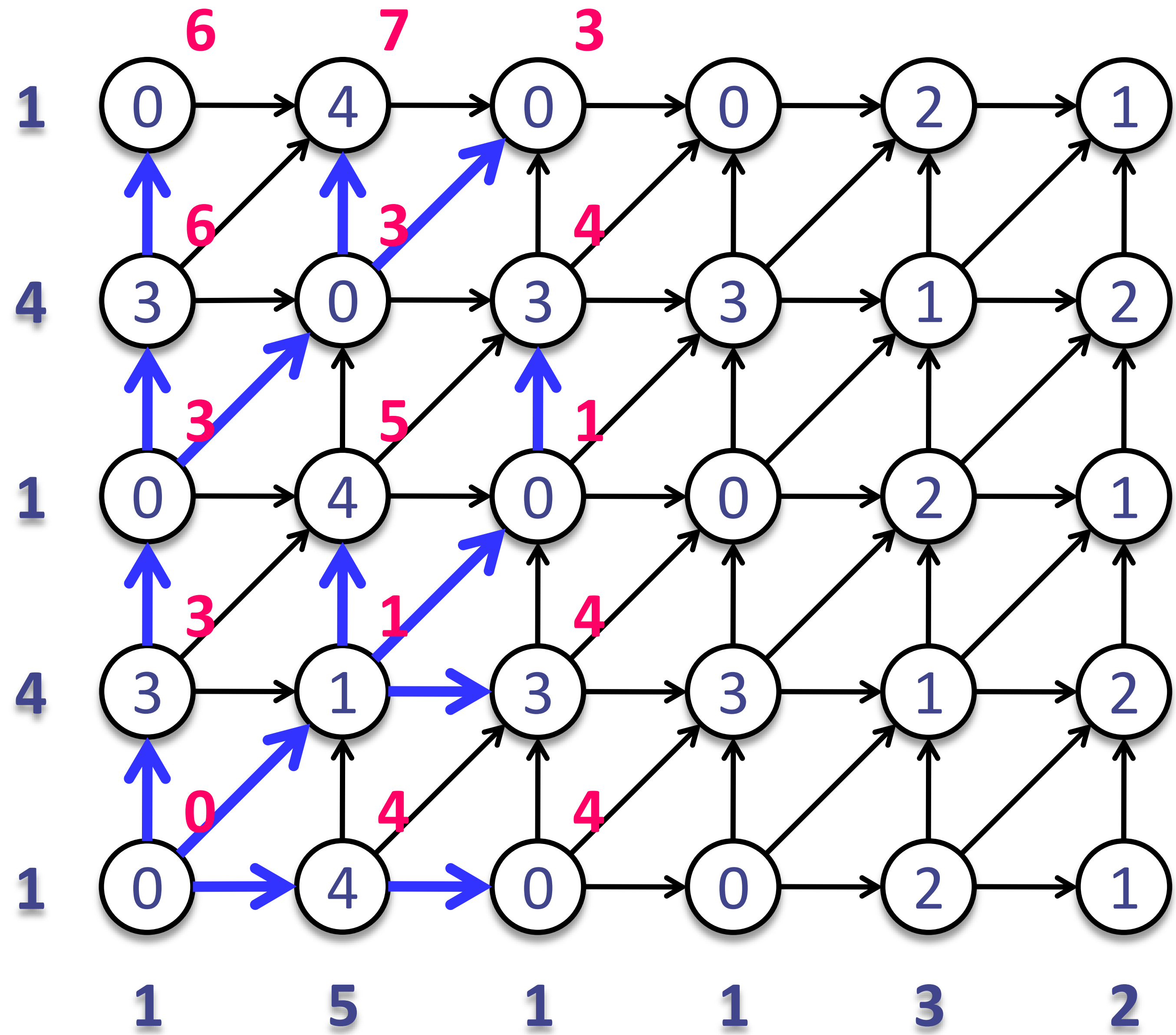
DPマッチングのアルゴリズム



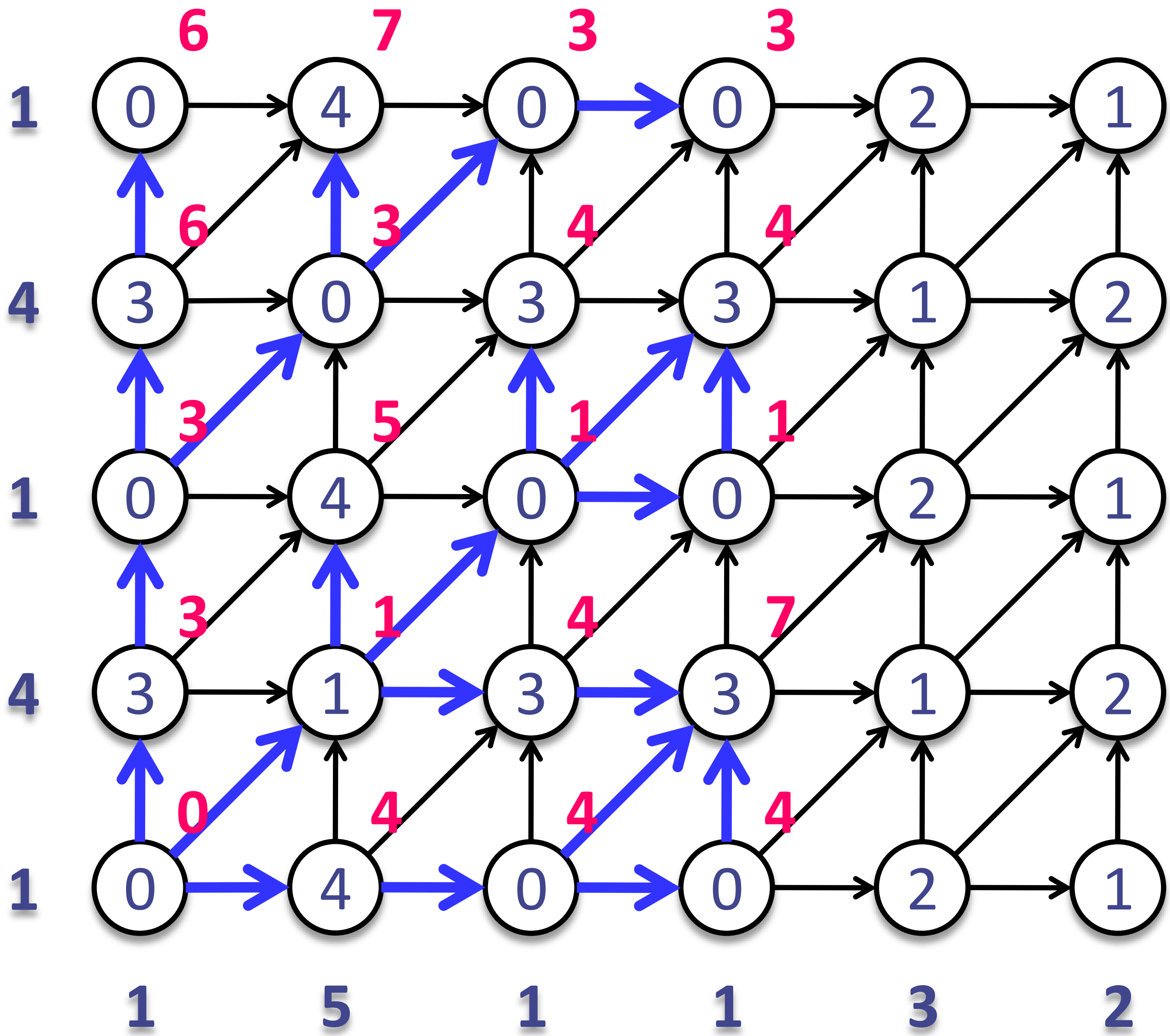
DPマッチングのアルゴリズム



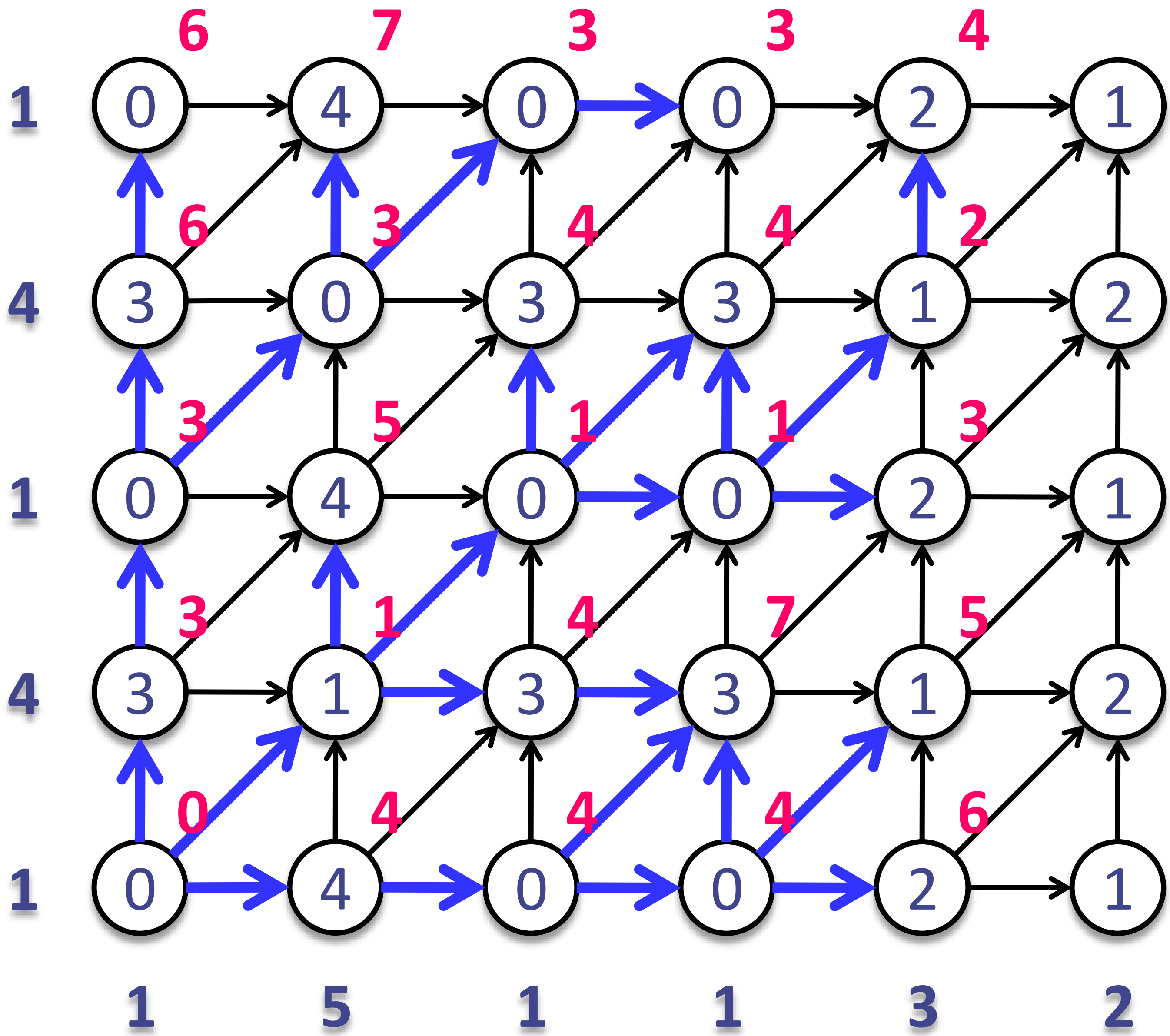
DPマッチングのアルゴリズム



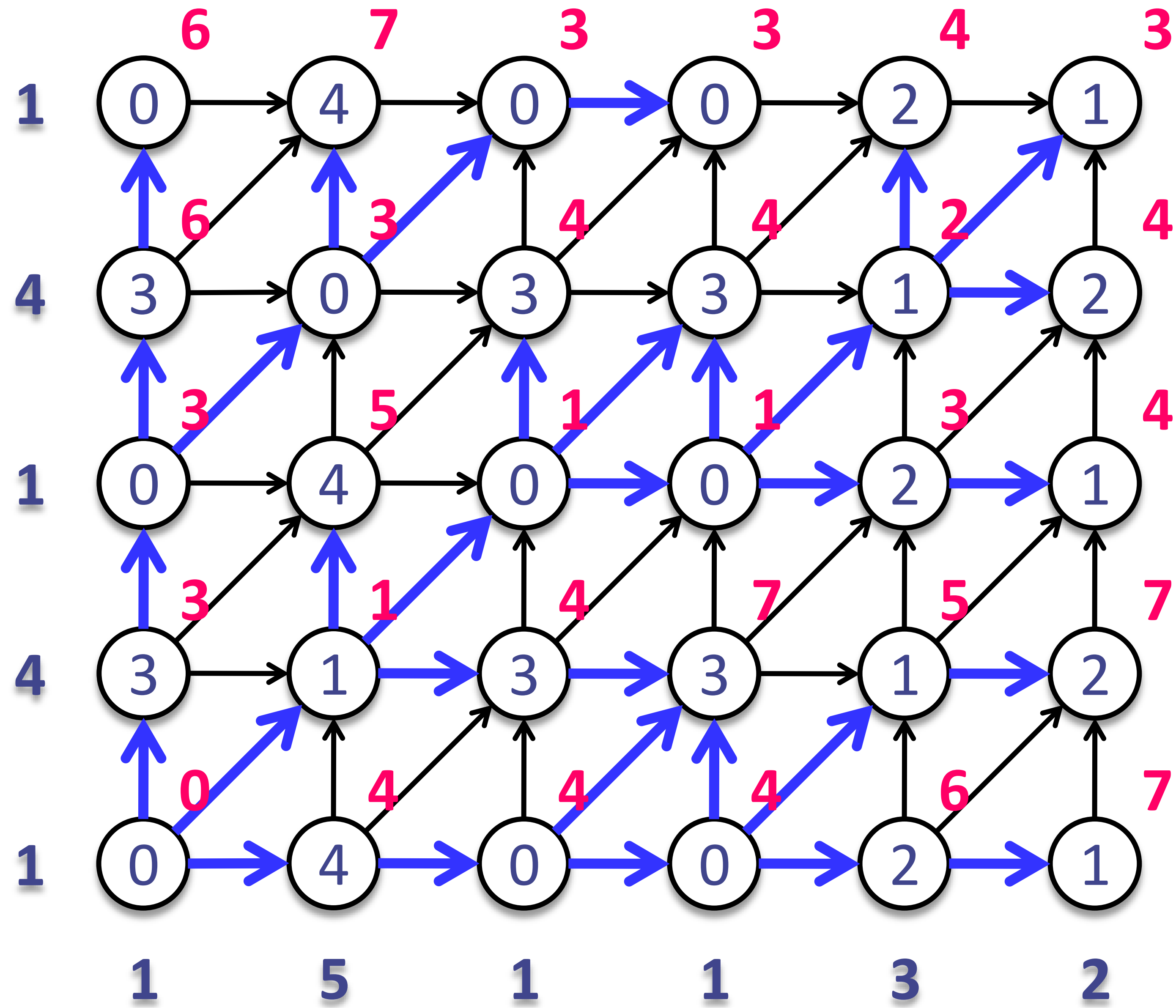
DPマッチングのアルゴリズム



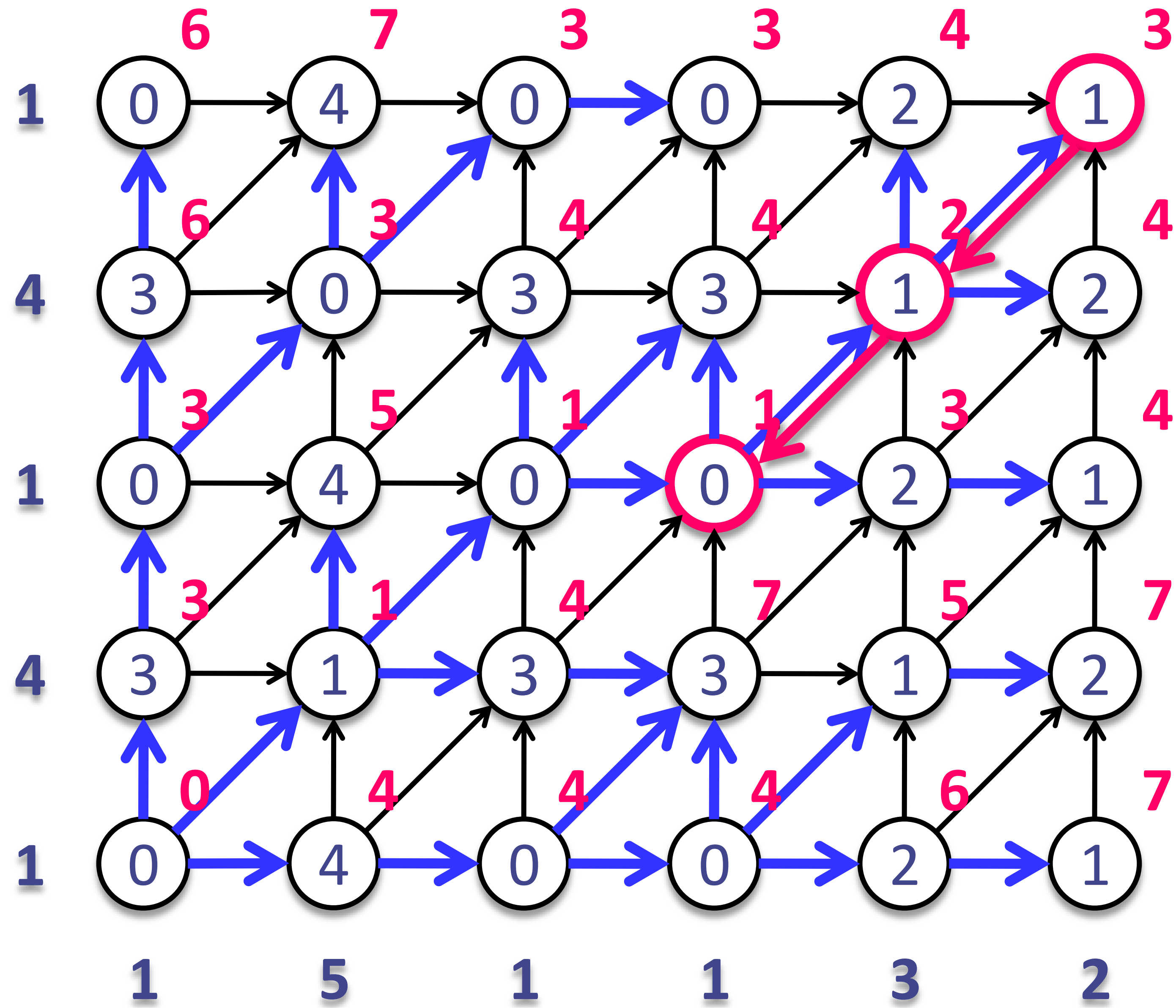
DPマッチングのアルゴリズム



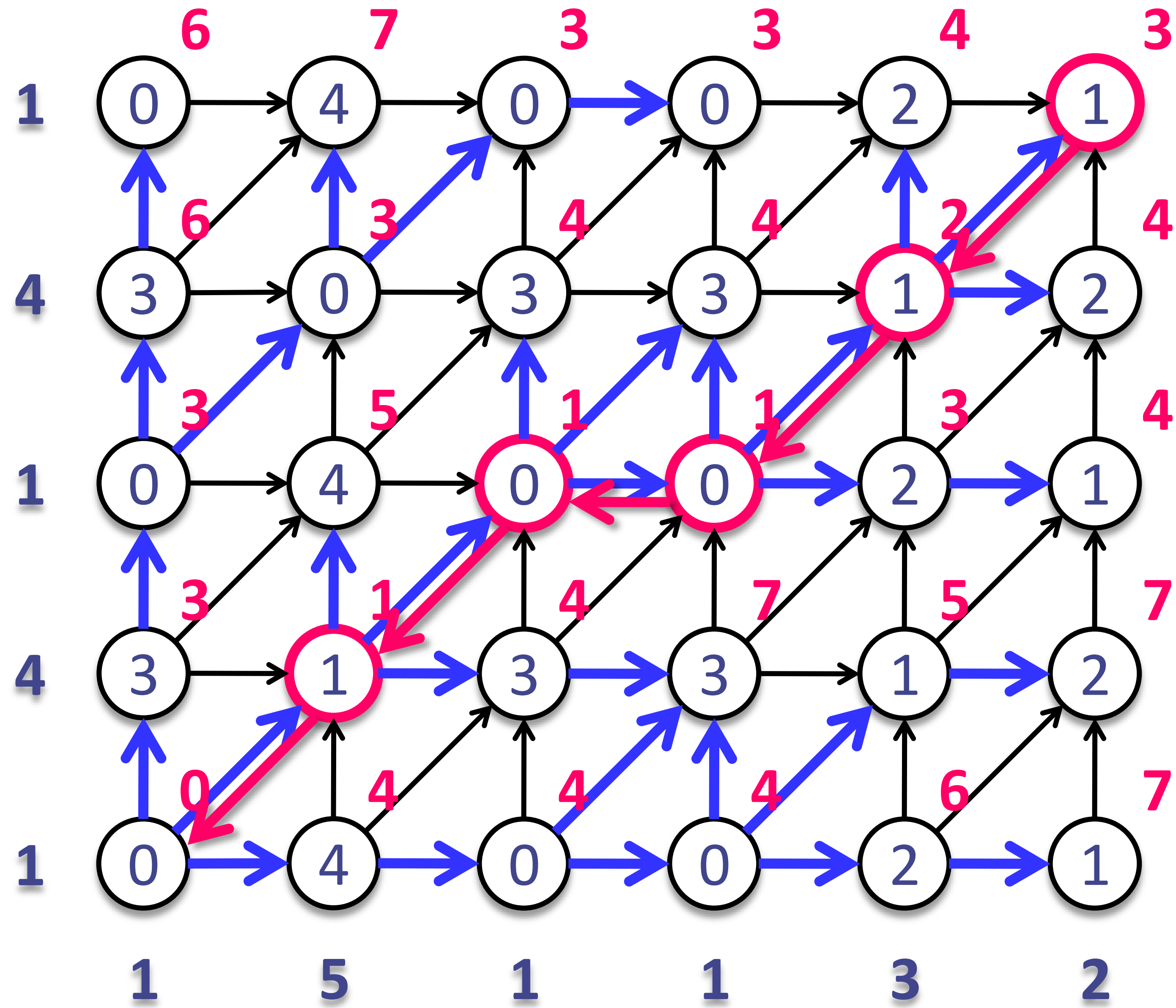
DPマッチングのアルゴリズム



DPマッチングのアルゴリズム



DPマッチングのアルゴリズム



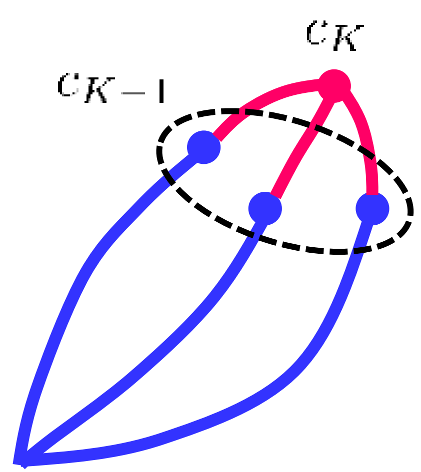
DPマッチングの最適性の根拠

- 最適時間整合問題
  - 標準系列: $y_1, \ldots, y_J$
  - 入力系列: $x_1, \ldots, x_I$
  - マッチング経路: $c_k = (i_k, j_k) \quad c_1 = (1,1),\ c_K = (I, J)$
  - ローカル距離: $d(i,j) = |y_j - x_i|^2$
  - $c_1 = (1,1)$ から $c_K = (i_K, j_K)$
    までの最小マッチング距離: $D(c_K) = \min_{c_1,\ldots,c_K} \sum_{k=1}^{K} d(c_k)$
  - 動的計画法 (Dynamic Programming)
    「最適経路は部分最適経路を含まなければならない」

$$D(c_K) = \min_{c_{K-1}} D(c_{K-1}) + d(c_K)$$

$$D(c_{K-1}) = \min_{c_{K-2}} D(c_{K-2}) + d(c_{K-1})$$

$$\vdots$$



本日の話題

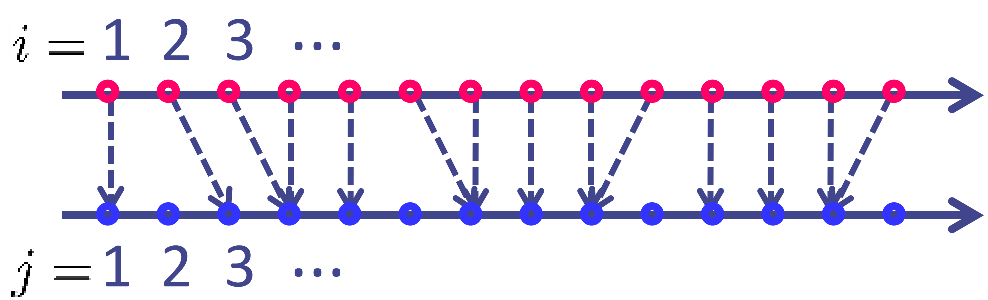
- 確率モデル（生成モデル）による動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

DPマッチングの生成モデルとしての解釈

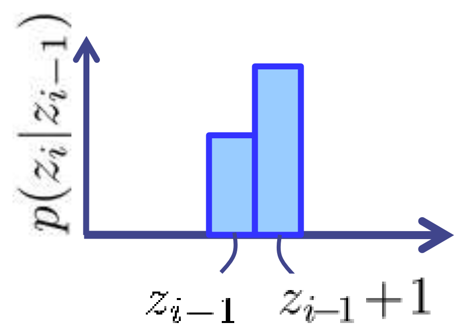
- 入力系列 $x_1, \ldots, x_I$ を生成する確率モデル
  - 平均が $y_1, \ldots, y_J$ の J 個の正規分布を用意する
  - 時刻 i の観測値 x_i が、 z_i 番目の正規分布から生成されたと考える $\Rightarrow$ $x_i \sim \mathcal{N}(x_i; y_{z_i}, \sigma^2)$ ←GMMそのもの！
  - $y_1, \ldots, y_J$ を標準系列だと考えると、 $z_1, \ldots, z_I$ は、入力系列と標準系列との間の時刻の対応づけを表す変数に相当



  - DPマッチングのように傾斜制限を設けるには？

$$z_i = \begin{cases} z_{i-1} \\ z_{i-1} + 1 \end{cases} \Rightarrow z_i \sim p(z_i | z_{i-1})$$

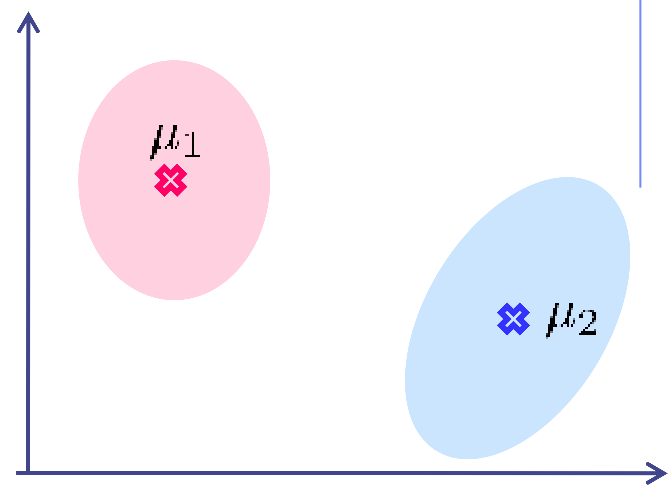
ここがGMMと異なる！



GMM

復習

- 混合正規分布モデルによるデータの生成プロセス
  - K個の正規分布をランダムに生成
  - サンプルごとに
    - 正規分布のクラス番号をランダムに選択
    - 選択されたクラス番号の正規分布に従ってサンプル値を生成
  - 生成された全サンプルが観測データ



- 生成モデル

for $k = 1, \ldots, K$

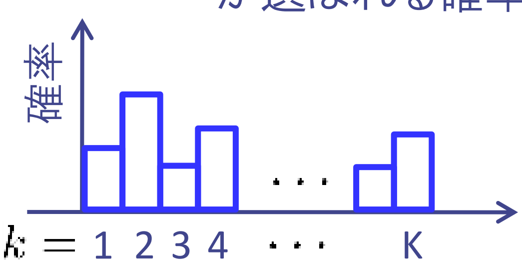
$$\{\mu_k, \sigma_k, \pi_k\} \sim H$$

for $n = 1, \ldots, N$

$$z_n \sim \mathrm{Categorical}(\pi_1, \ldots, \pi_K)$$

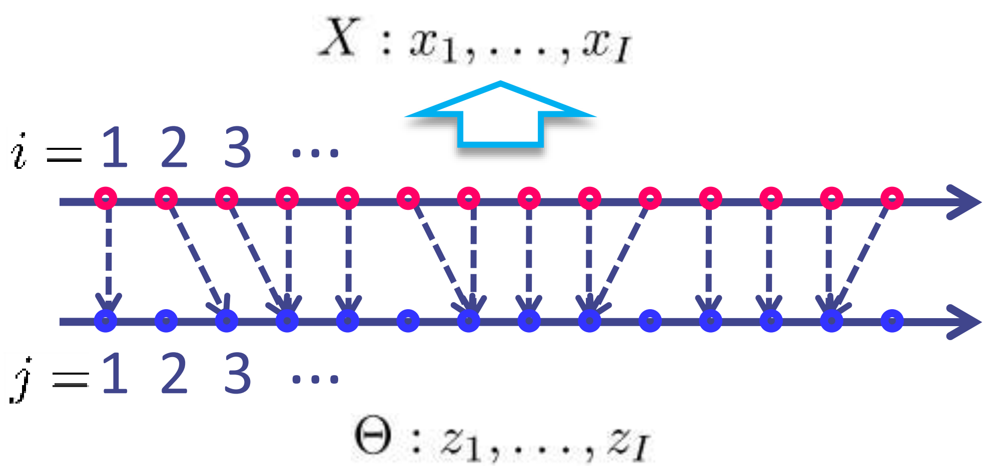
$$x_n | z_n \sim \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})$$

π_k：k番目の正規分布が選ばれる確率



最尤推定

- データから、尤度関数が最大となる確率モデルのパラメータを推定するための一般的な枠組
  - 尤度関数とは？
    - データを X，パラメータを Θ とすると，尤度関数は $p(X|\Theta)$ のこと
    - 「データ X の生成源として，パラメータ Θ の確率モデルがどの程度尤もらしいか」を意味した規準
  - 最尤推定
    - 最も尤もらしい、データの生成源を推定すること
  - 最適時間整合問題の例では・・・



本日の話題

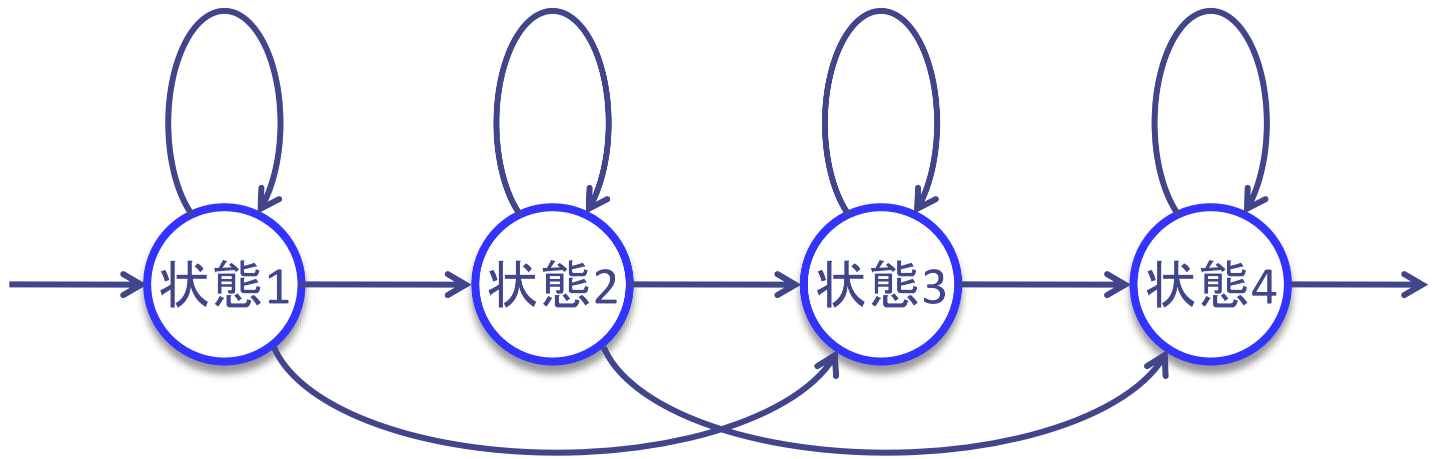
- 確率モデル（生成モデル）による動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

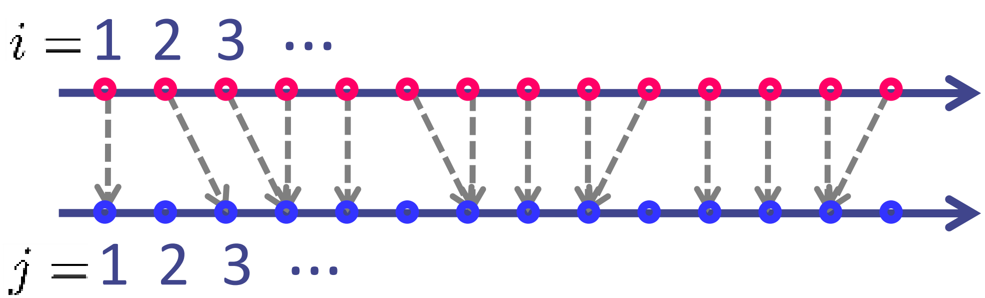
本日の話題

- 確率モデル（生成モデル）による
  動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

隠れマルコフモデル (Hidden Markov Model; HMM)

- 隠れマルコフモデルによる観測値系列 $x_1, \ldots, x_I$ の生成プロセス
  - 状態系列 $z_1, \ldots, z_I$ がマルコフ過程に従って決定
    （z_i が、z_{i-1} に依存した確率に従って決定）

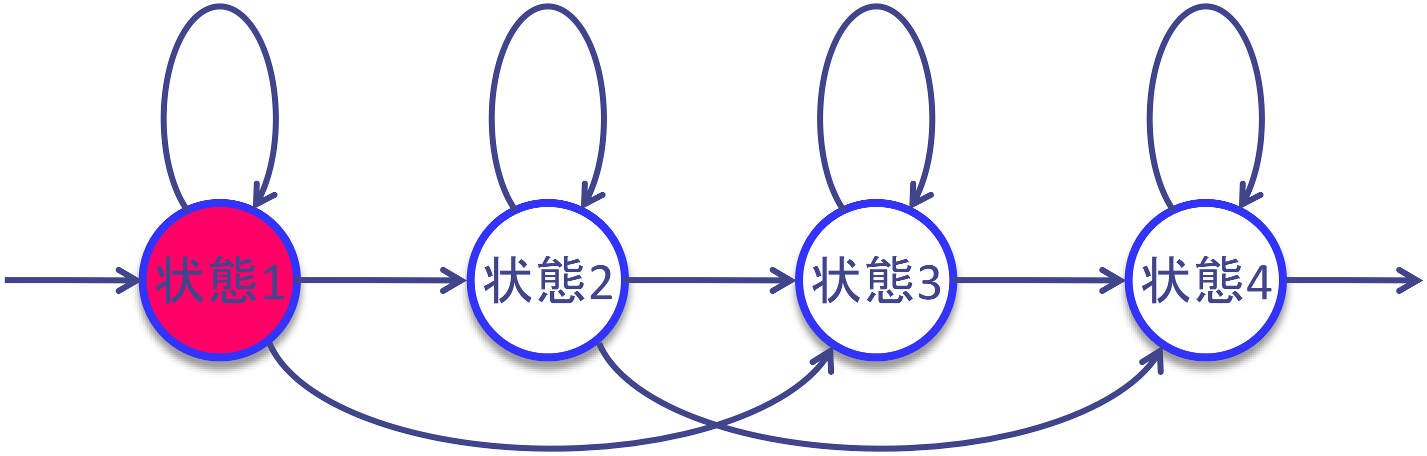


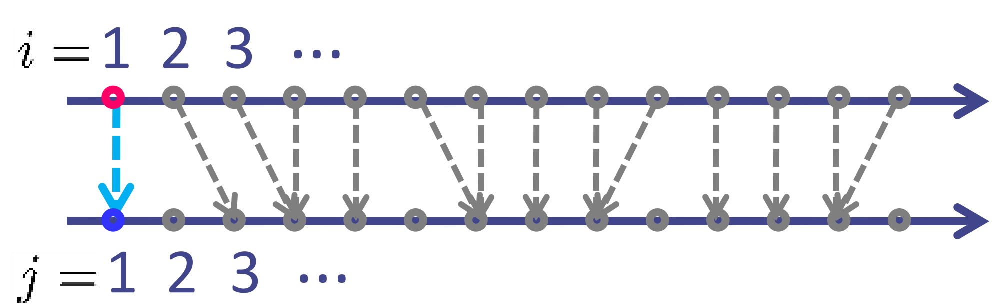


  - i ごとに、状態 z_i の出力分布に従って観測値 x_i が生成

隠れマルコフモデル (Hidden Markov Model; HMM)

- 隠れマルコフモデルによる観測値系列 $x_1, \ldots, x_I$ の生成プロセス
  - 状態系列 $z_1, \ldots, z_I$ がマルコフ過程に従って決定（z_i が、z_{i-1} に依存した確率に従って決定）

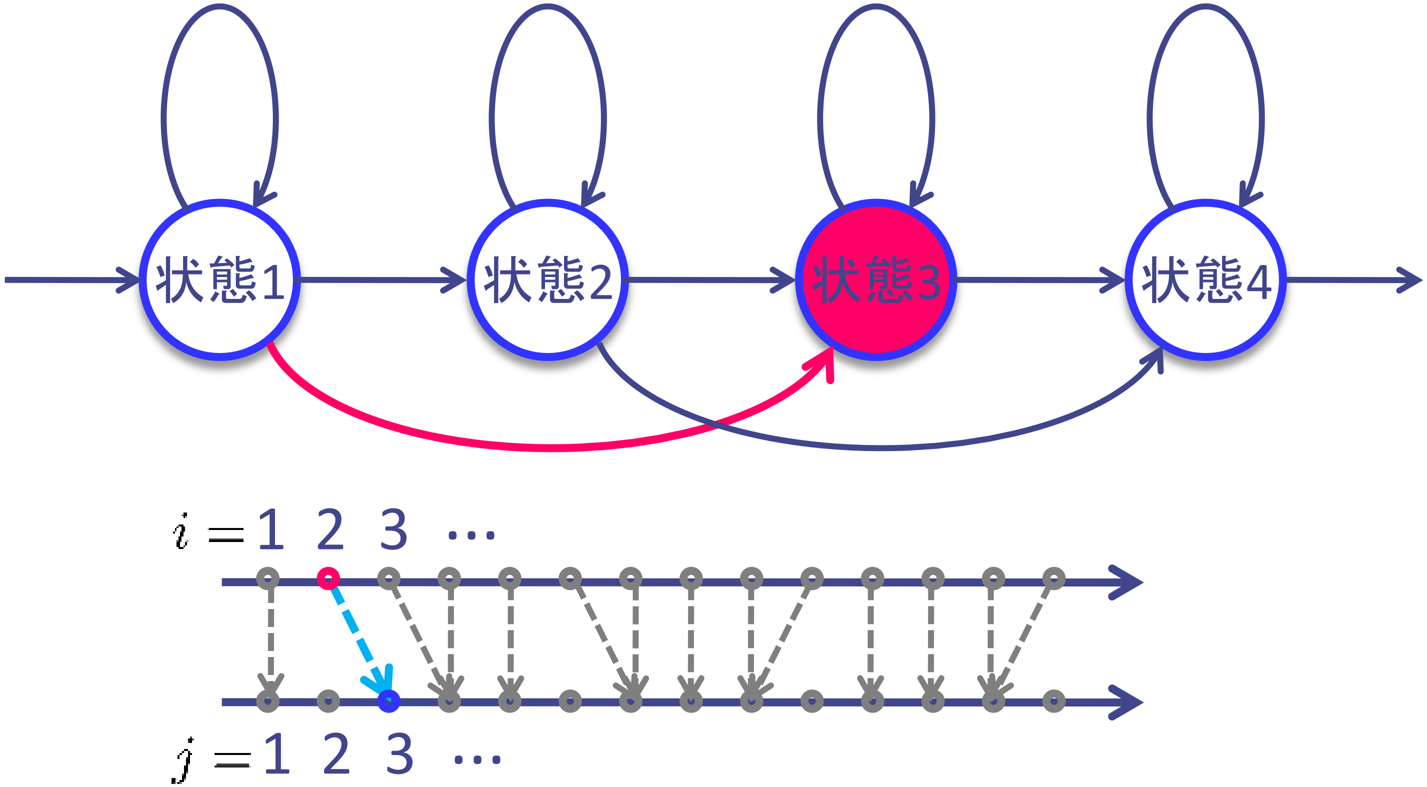




  - i ごとに、状態 z_i の出力分布に従って観測値 x_i が生成

隠れマルコフモデル (Hidden Markov Model; HMM)

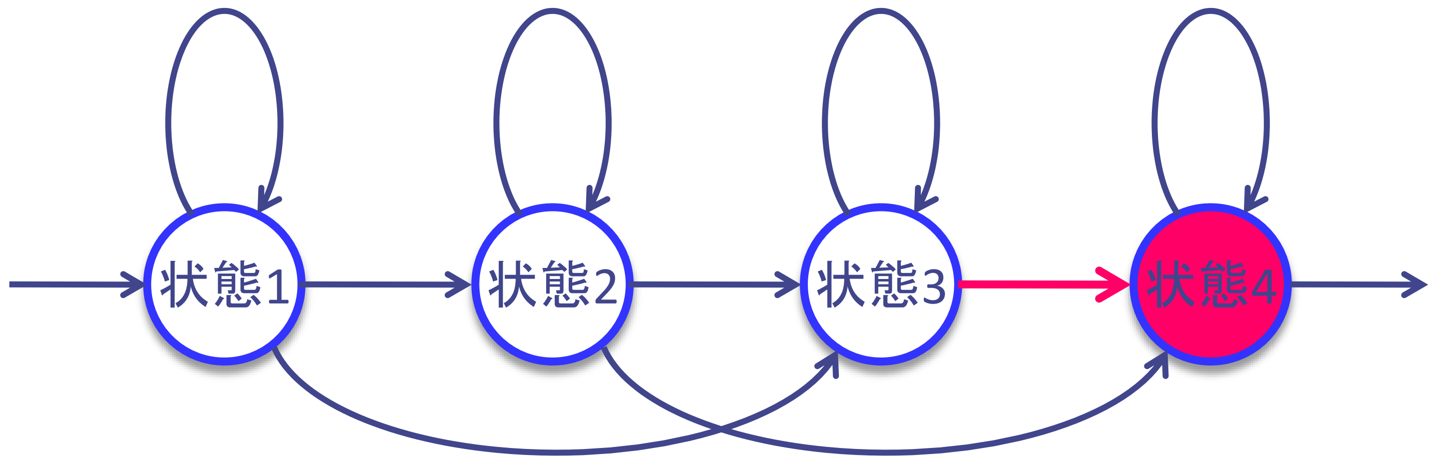
- 隠れマルコフモデルによる観測値系列 $x_1, \ldots, x_I$ の生成プロセス
  - 状態系列 $z_1, \ldots, z_I$ がマルコフ過程に従って決定
    （z_i が、z_{i-1} に依存した確率に従って決定）

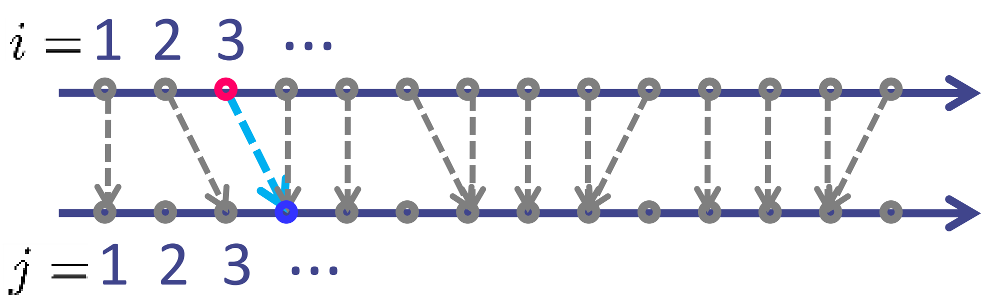


  - i ごとに、状態 z_i の出力分布に従って観測値 x_i が生成

隠れマルコフモデル (Hidden Markov Model; HMM)

- 隠れマルコフモデルによる観測値系列 $x_1, \ldots, x_I$ の生成プロセス
  - 状態系列 $z_1, \ldots, z_I$ がマルコフ過程に従って決定
    （z_i が、z_{i-1} に依存した確率に従って決定）

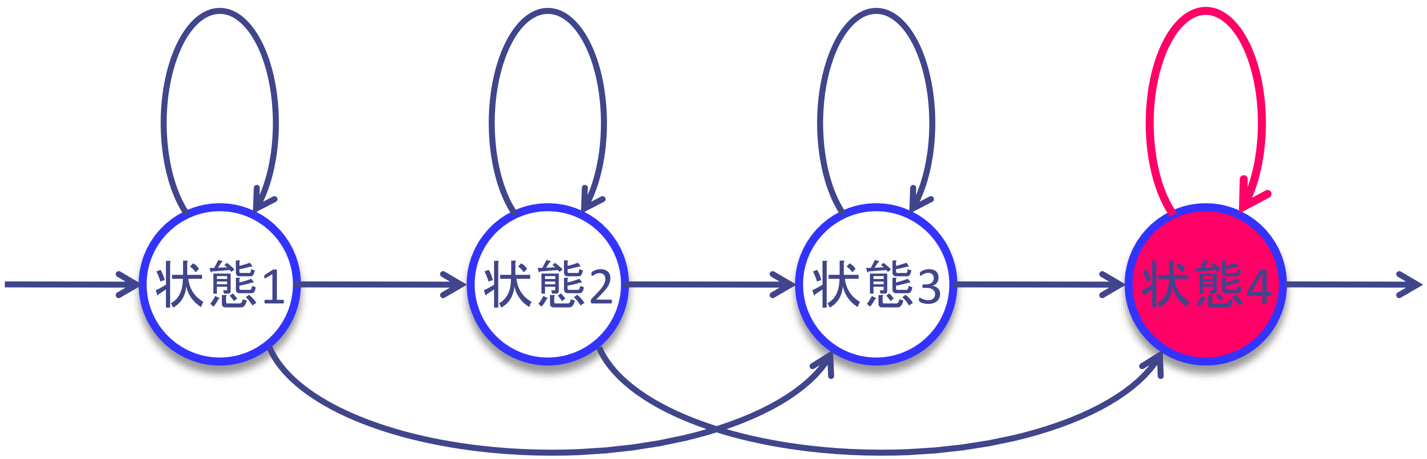


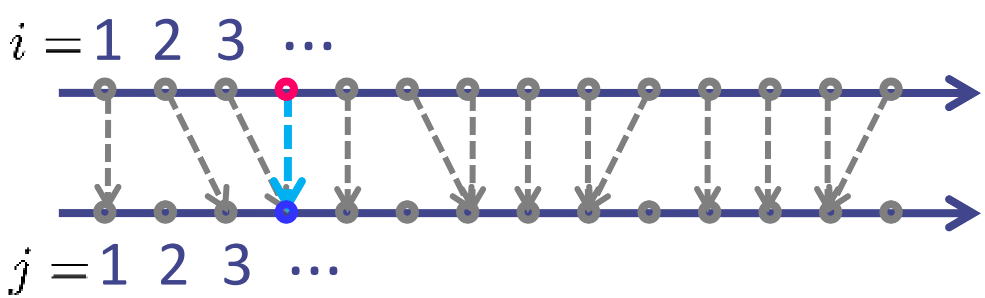


  - i ごとに、状態 z_i の出力分布に従って観測値 x_i が生成

隠れマルコフモデル (Hidden Markov Model; HMM)

- 隠れマルコフモデルによる観測値系列 $x_1, \ldots, x_I$ の生成プロセス
  - 状態系列 $z_1, \ldots, z_I$ がマルコフ過程に従って決定（z_i が、z_{i-1} に依存した確率に従って決定）

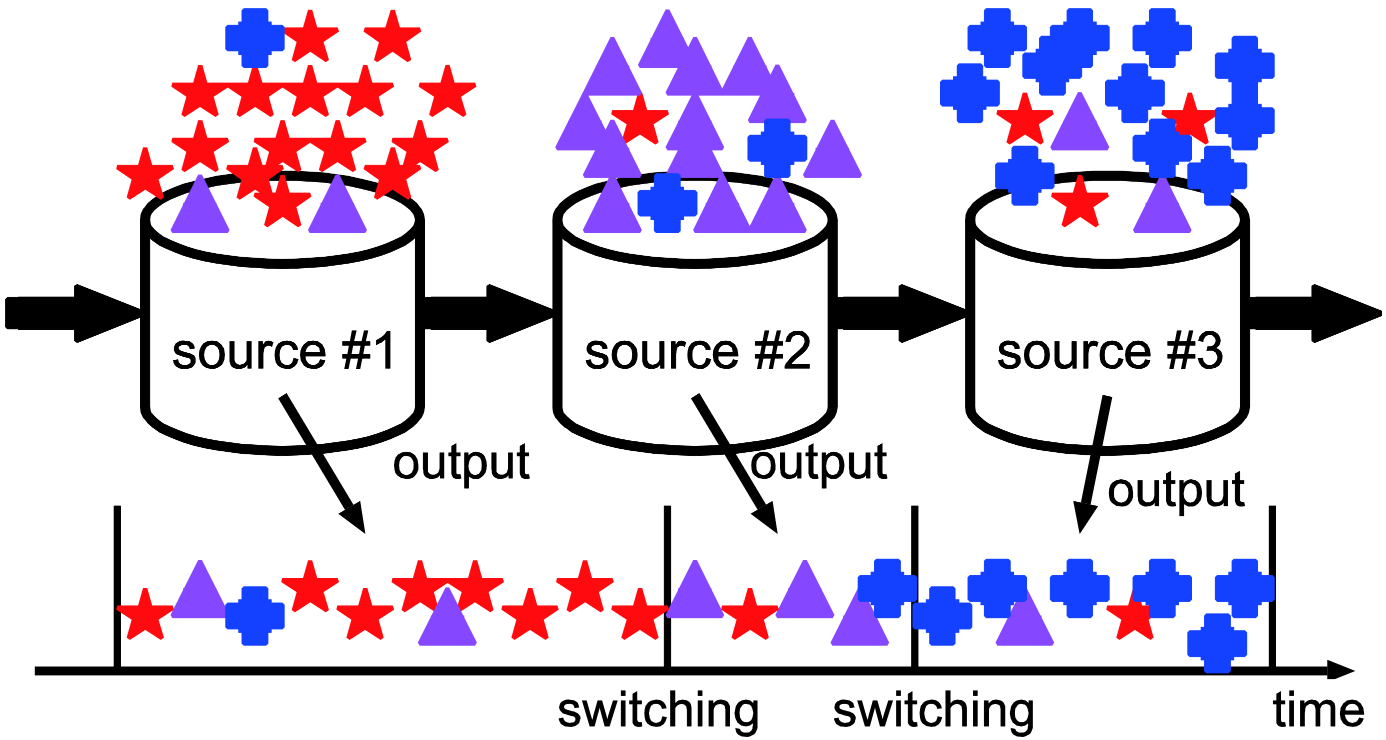




  - i ごとに、状態 z_i の出力分布に従って観測値 x_i が生成

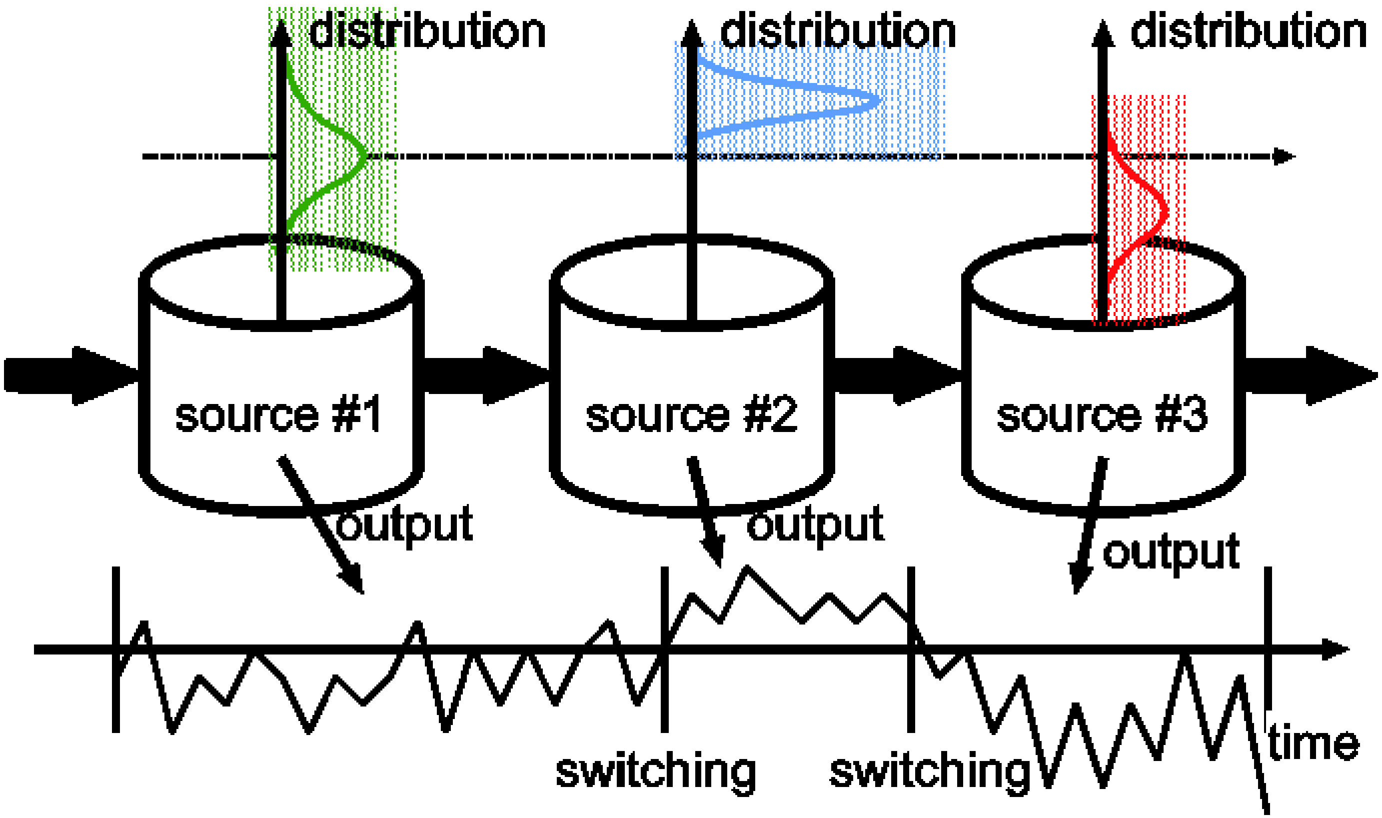
隠れマルコフモデル (Hidden Markov Model; HMM)

- 非定常信号源のモデル
  - 定常信号源が切り替わるモデル（区分定常信号源）
  - 離散シンボル系列の場合：



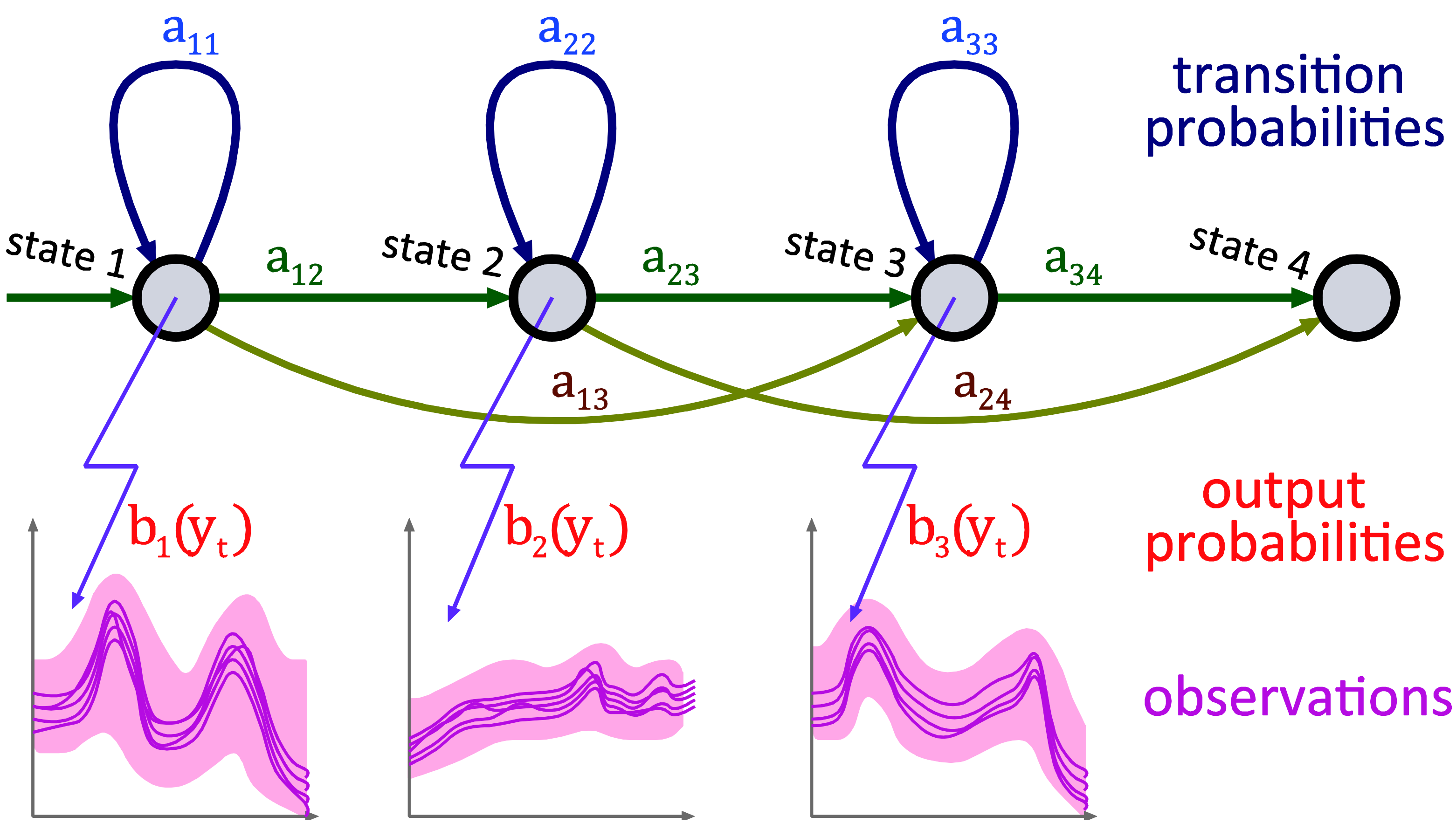
隠れマルコフモデル (Hidden Markov Model; HMM)

- 非定常信号源のモデル
  - 定常信号源が切り替わるモデル（区分定常信号源）
  - 連続値系列の場合：



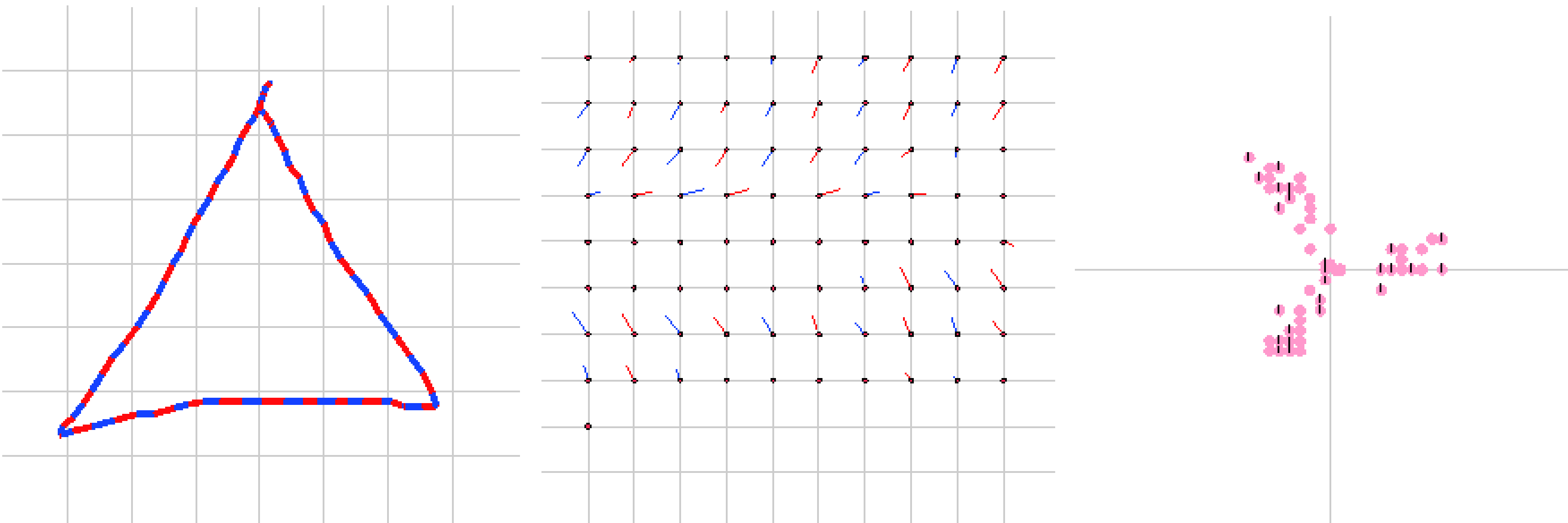
隠れマルコフモデル (Hidden Markov Model; HMM)

- 定常信号源(「状態」)がマルコフ遷移する非定常信号源モデル
  - 状態遷移確率に従って状態が遷移
  - 状態出力分布に従って観測値が生成



HMMの例：手書き三角形の速度ベクトル系列

- 手書き三角形、ペンの速度ベクトル系列、その空間分布



隠れマルコフモデル (Hidden Markov Model; HMM)

生成モデル

for $j = 1, \ldots, J$

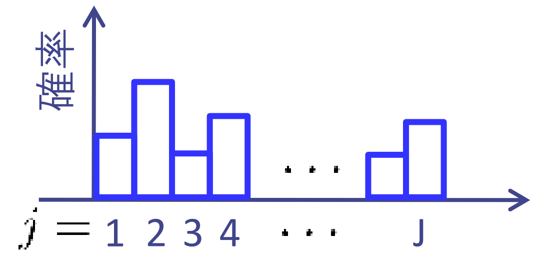
$$\theta_j \sim H$$

for $i = 1, \ldots, I$

$$z_i | z_{i-1} \sim p(z_i | z_{i-1}) = \pi_{z_{i-1}, z_i}$$

$$x_i | z_i \sim p(x_i | \theta_{z_i})$$

$\pi_{j,j'}$: 状態 j から状態 j' へ遷移する確率



本日の話題

- 確率モデル（生成モデル）による
  動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - ViterbiアルゴリズムとDPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による動的時間伸縮（DPマッチング）問題の定式化
  - DPマッチング
  - DPマッチングの生成モデルとしての解釈
  - 隠れマルコフモデル (Hidden Markov Model; HMM)
  - HMMのパラメータ推定アルゴリズム
    - Viterbiアルゴリズム
    - Forwardアルゴリズム
    - Baum-Welchアルゴリズム
  - HMMとGMM (Gaussian mixture model) の関係
  - Viterbiアルゴリズムと DPマッチングの関係
  - ForwardアルゴリズムとKalmanフィルタアルゴリズムの関係

Viterbiアルゴリズム

- 状態系列の最大事後確率推定
  - 状態系列: $\Theta = \{z_i\}_{i=1,\ldots,I}$
  - 求めたいのは $\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\, p(\Theta|X) = \underset{\Theta}{\operatorname{argmax}} \log p(\Theta|X)$
    $= \underset{\Theta}{\operatorname{argmax}} \log p(\Theta, X)$

対数をとっても最大化
の目標は変わらない

- 尤度関数の導出

$$\log p(\Theta, X) = \log p(X|\Theta) + \log p(\Theta)$$

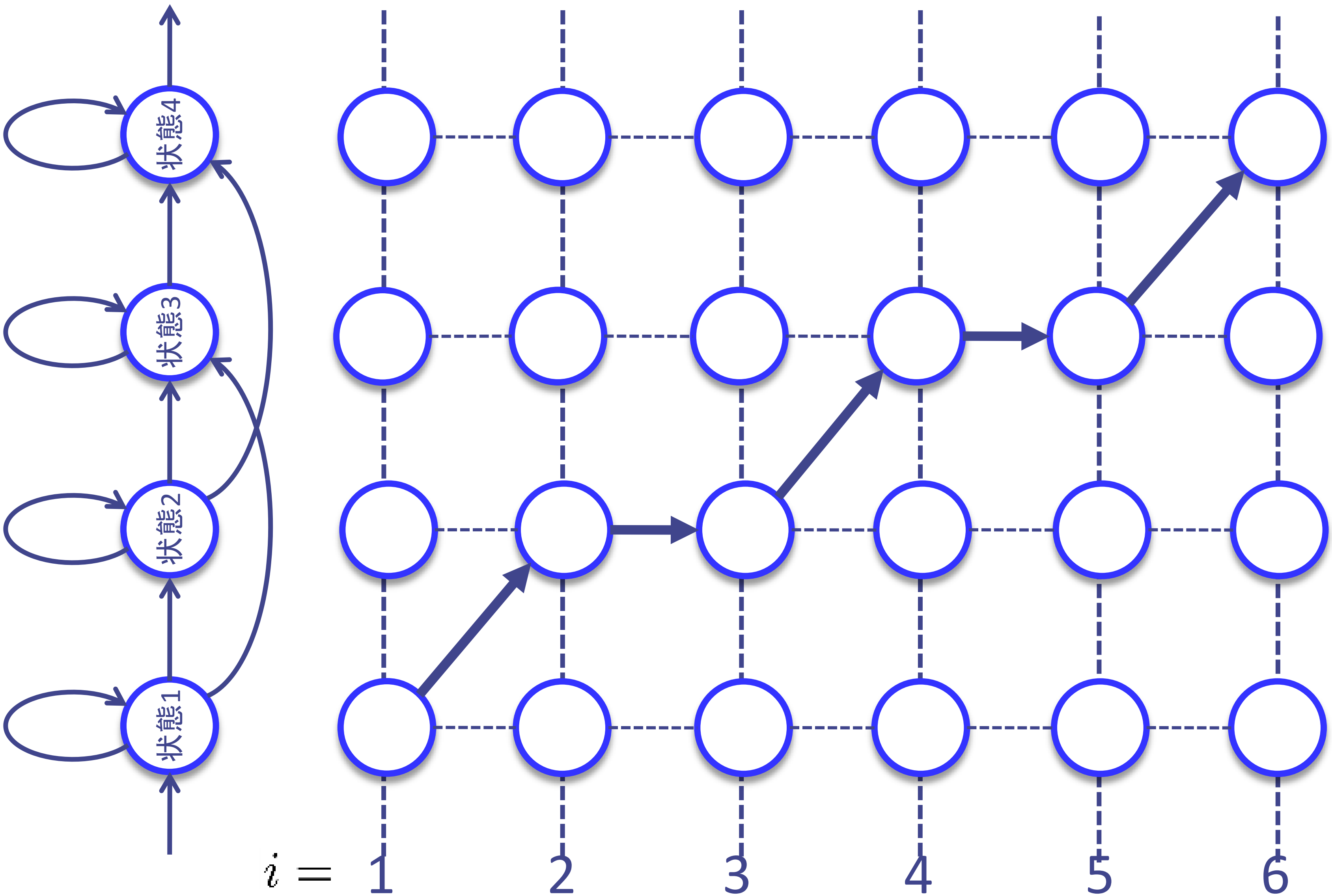
$$= \log \prod_{i=1}^{I} p(x_i|z_i) + \log \prod_{i=1}^{I} \underbrace{p(z_i|z_{i-1})}_{= \pi_{z_{i-1},z_i}}$$

$$= \sum_{i=1}^{I} \underline{\log p(x_i|z_i)} + \sum_{i=1}^{I} \underline{\log \pi_{z_{i-1},z_i}}$$

DPマッチングにおける
ローカル距離に相当

DPマッチングにおける
傾斜重みに相当

トレリス表現（空間時間経路）

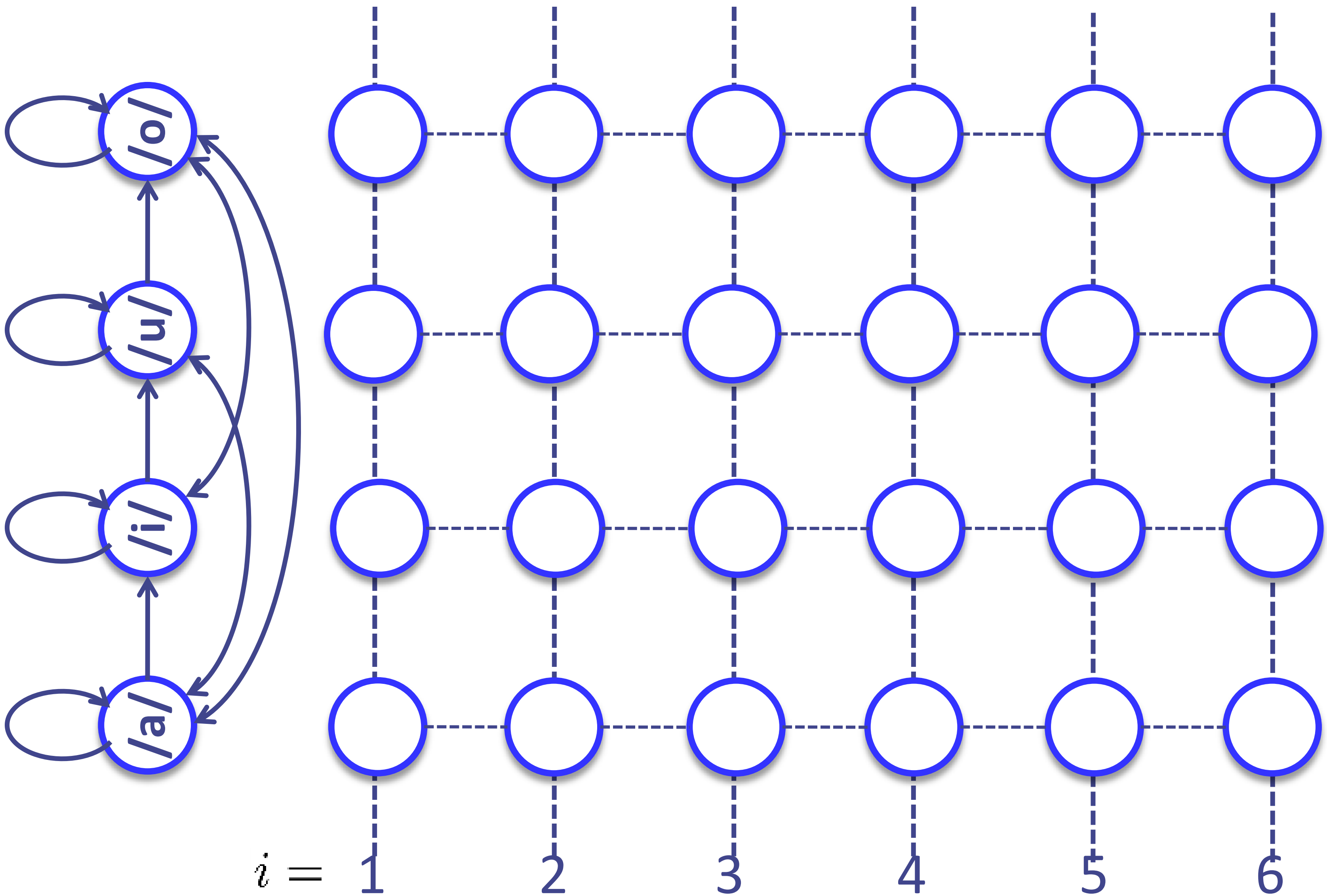


DPマッチングとの関係

- Left-to-Right HMM
  - 状態が一方向にのみ（「左」から「右」へ）遷移していくよう拘束されたHMM
  - Left-to-Right HMMの最適状態遷移系列の推定問題はDPマッチングと同形
- 経路制限する代わりに、あらゆる状態遷移に確率を仮定するのがHMM流の考え方
  - 経路制限は遷移確率設計の特殊ケース
（遷移を許さない状態間の遷移確率を0に設定すれば良い）
- Viterbiアルゴリズム
  - HMMの最適状態遷移系列の探索アルゴリズム
  - DPマッチングと同様、動的計画法に基づく
  - DPマッチングとは独立にViterbiによって開発されたこのアルゴリズムは、Viterbiアルゴリズムと呼ばれている

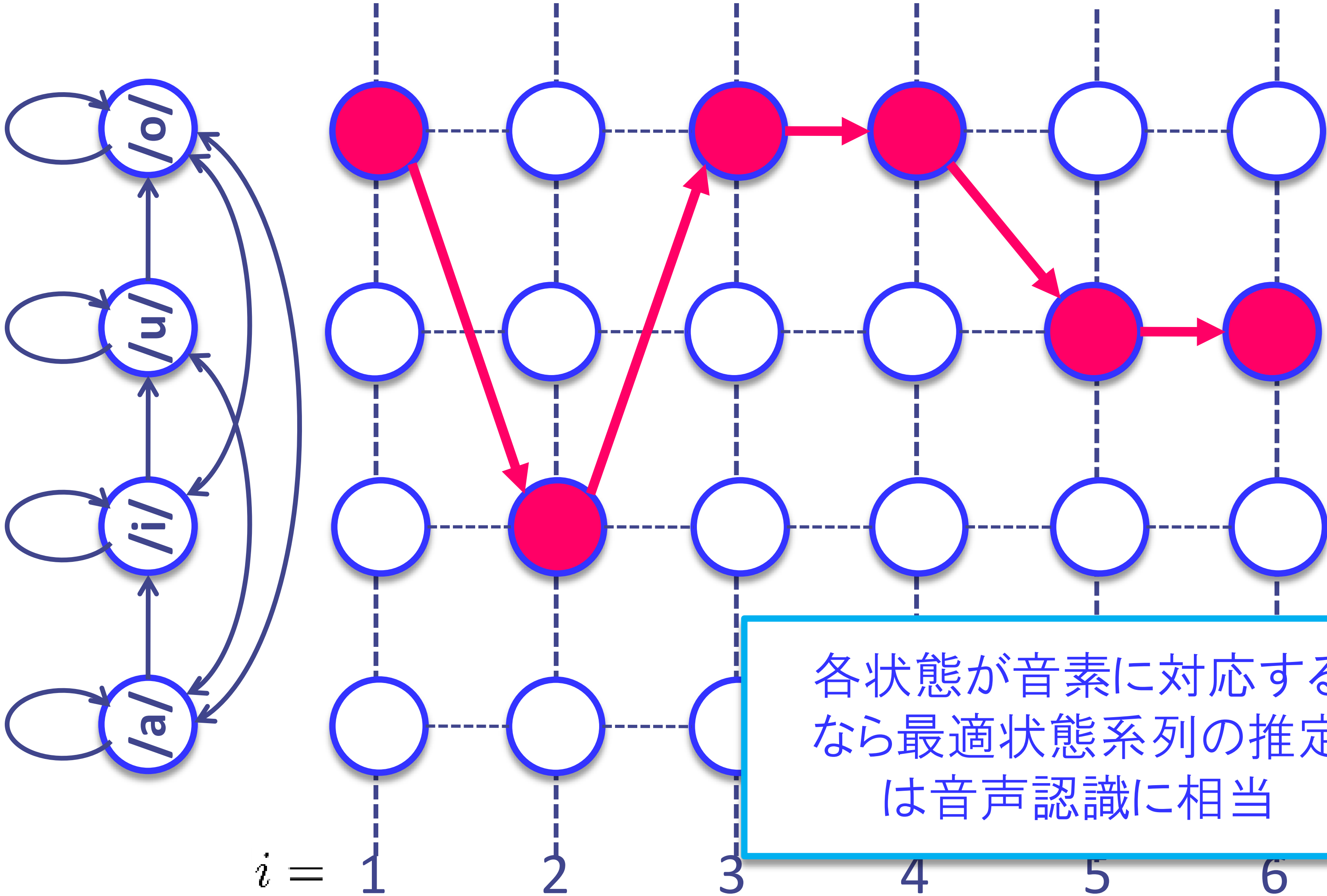
エルゴディックHMM

- すべての状態遷移を許したHMM



エルゴディックHMM

- すべての状態遷移を許したHMM



各状態が音素に対応するなら最適状態系列の推定は音声認識に相当

HMMの基本問題

観測系列 $X_{1:T}$: $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$: $z_1, z_2, \ldots, z_T$

パラメータ Θ :

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列 $X_{1:T}$ が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T},\Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する

⇒Viterbiアルゴリズム

- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t},\Theta)$ を推定する ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列 $X_{1:T}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する ⇒Baum-Welchアルゴリズム

HMMの基本問題

観測系列 $X_{1:T}$: $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$: $z_1, z_2, \ldots, z_T$

パラメータ Θ :

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列 $X_{1:T}$ が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T}, \Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する

    ⇒Viterbiアルゴリズム
- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t}, \Theta)$ を推定する

    ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列 $X_{1:T}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する

    ⇒Baum-Welchアルゴリズム

最適状態系列推定問題

- 状態系列の最大事後確率推定
  - 状態系列： $\Theta = \{z_i\}_{i=1,\ldots,I}$
  - 求めたいのは $\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\, p(\Theta|X) = \underset{\Theta}{\operatorname{argmax}} \log p(\Theta|X)$
    $= \underset{\Theta}{\operatorname{argmax}} \log p(\Theta, X)$

対数をとっても最大化
の目標は変わらない

- 尤度関数の導出

$$\log p(\Theta, X) = \log p(X|\Theta) + \log p(\Theta)$$

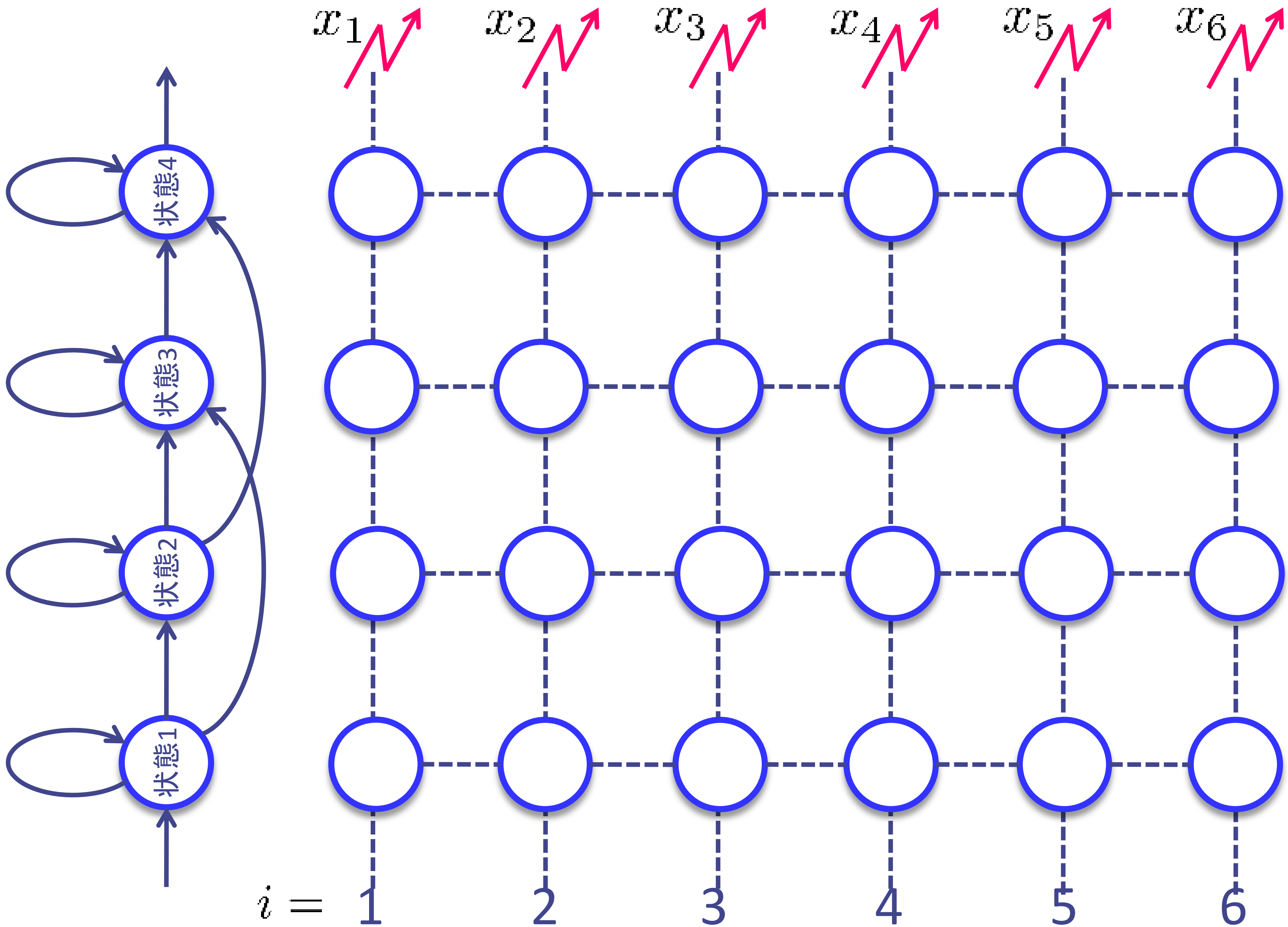
$$= \log \prod_{i=1}^{I} p(x_i|z_i) + \log \prod_{i=1}^{I} \underbrace{p(z_i|z_{i-1})}_{= \pi_{z_{i-1},z_i}}$$

$$= \sum_{i=1}^{I} \underline{\log p(x_i|z_i)} + \sum_{i=1}^{I} \underline{\log \pi_{z_{i-1},z_i}}$$

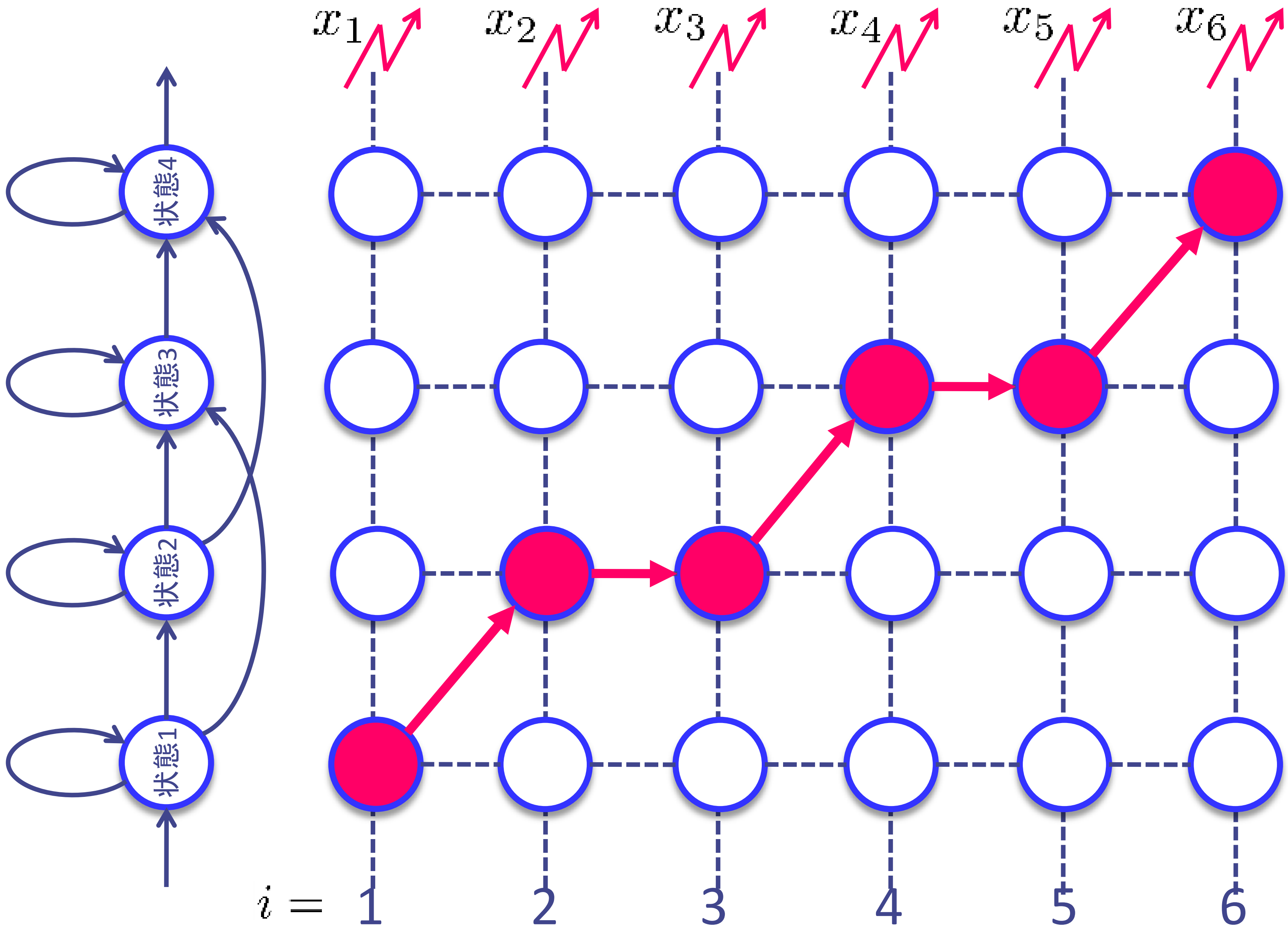
DPマッチングにおける
ローカル距離に相当

DPマッチングにおける
傾斜重みに相当

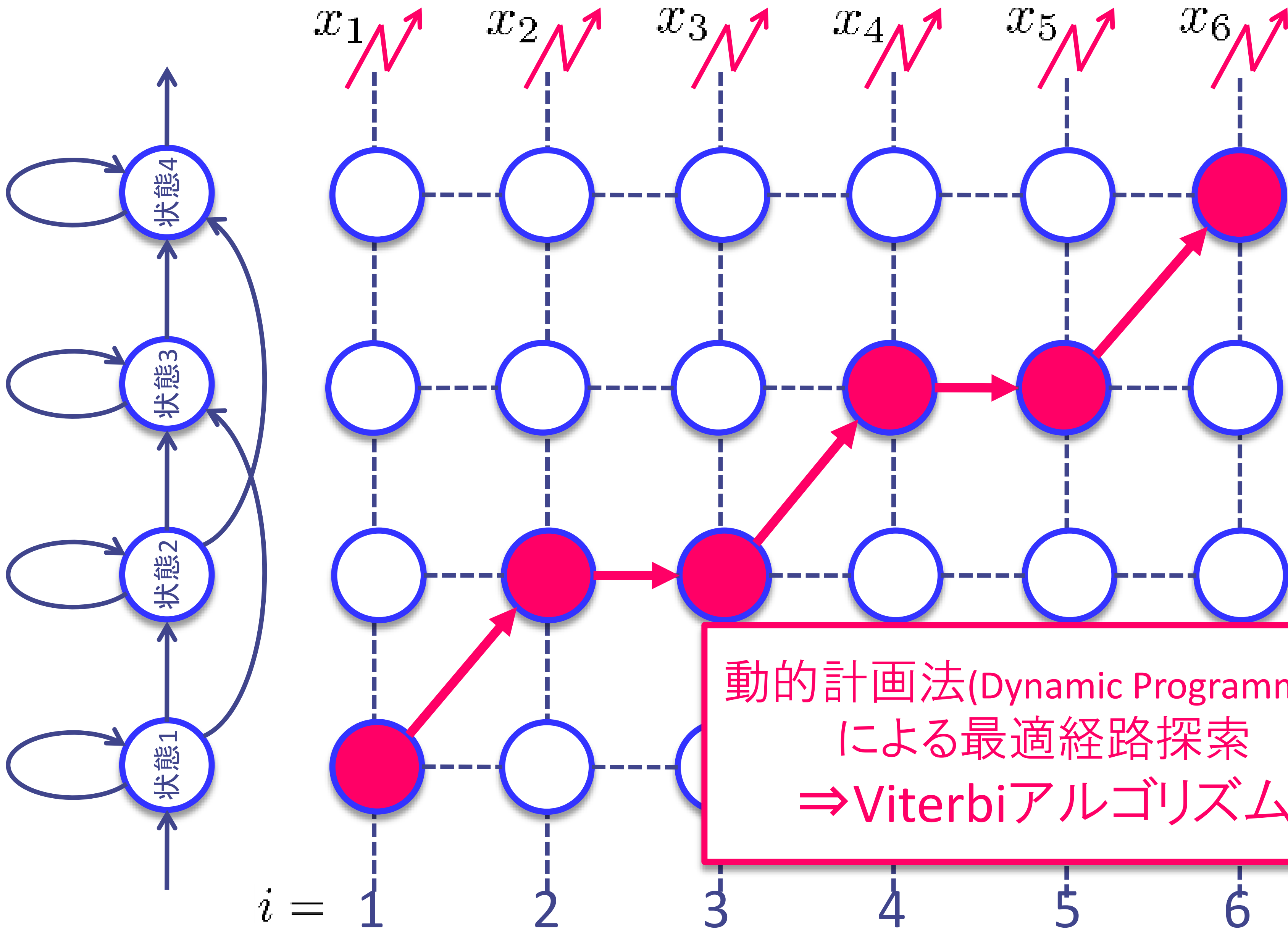
Viterbiアルゴリズム：トレリス上の最適経路探索



Viterbiアルゴリズム：トレリス上の最適経路探索



Viterbiアルゴリズム：トレリス上の最適経路探索



動的計画法(Dynamic Programming)による最適経路探索
⇒Viterbiアルゴリズム

HMMの基本問題

観測系列 $X_{1:T}$ ： $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$ ： $z_1, z_2, \ldots, z_T$

パラメータ Θ ：

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列$X_{1:T}$が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T}, \Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する

⇒Viterbiアルゴリズム

- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t}, \Theta)$ を推定する ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列$X_{1:T}$を生成する確率$p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する ⇒Baum-Welchアルゴリズム

HMMの基本問題

観測系列 $X_{1:T}$: $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$: $z_1, z_2, \ldots, z_T$

パラメータ Θ :

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列 $X_{1:T}$ が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T}, \Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する

    ⇒Viterbiアルゴリズム
- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t}, \Theta)$ を推定する ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列 $X_{1:T}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する

    ⇒Baum-Welchアルゴリズム

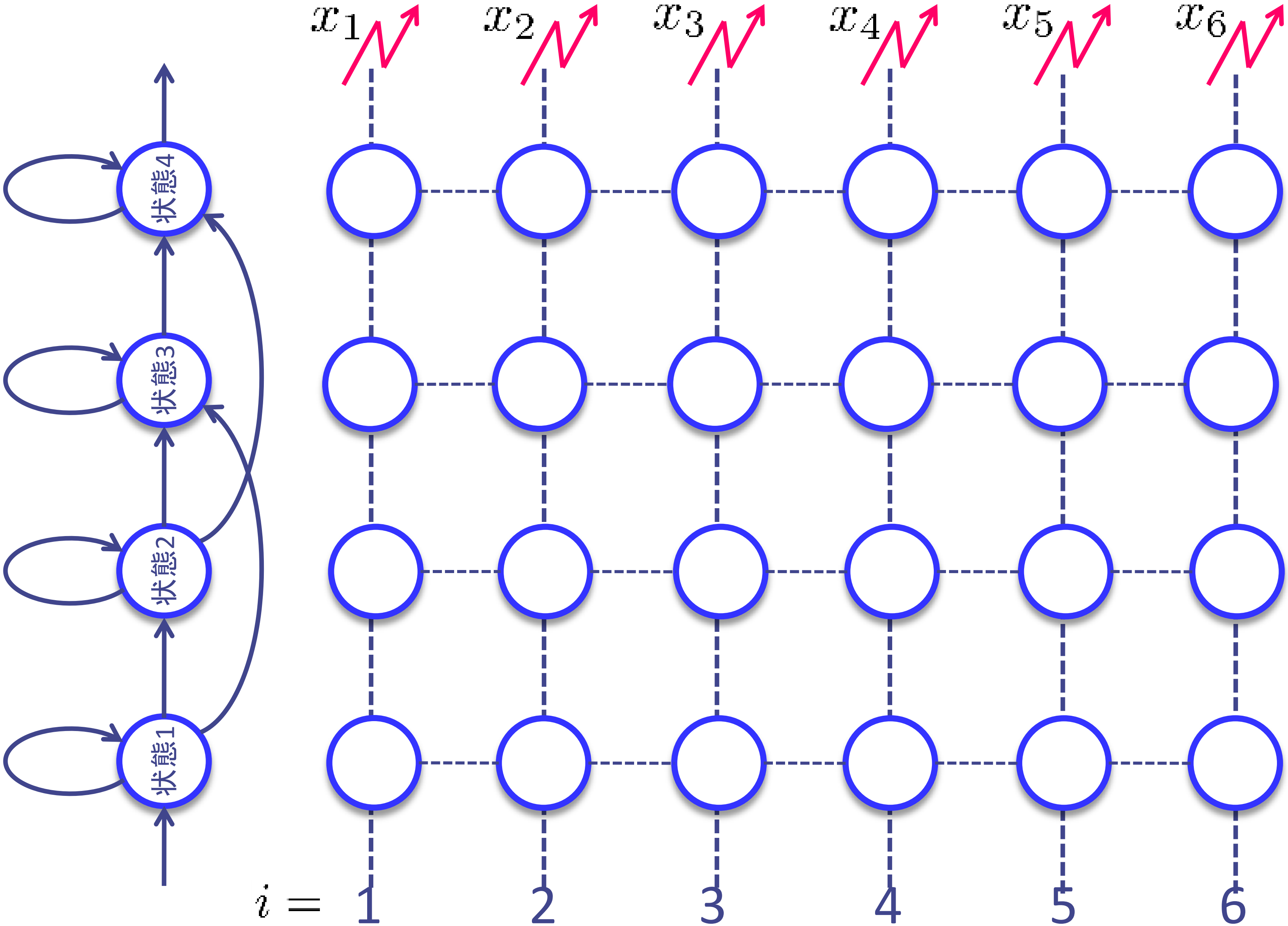
状態事後確率推定問題

- 時刻 k までの観測系列 $x_1, \ldots, x_k$ が与えられた下で、時刻 k において状態が z_k である確率（つまり、$p(z_k|x_1, \ldots, x_k)$ ）を求めたい

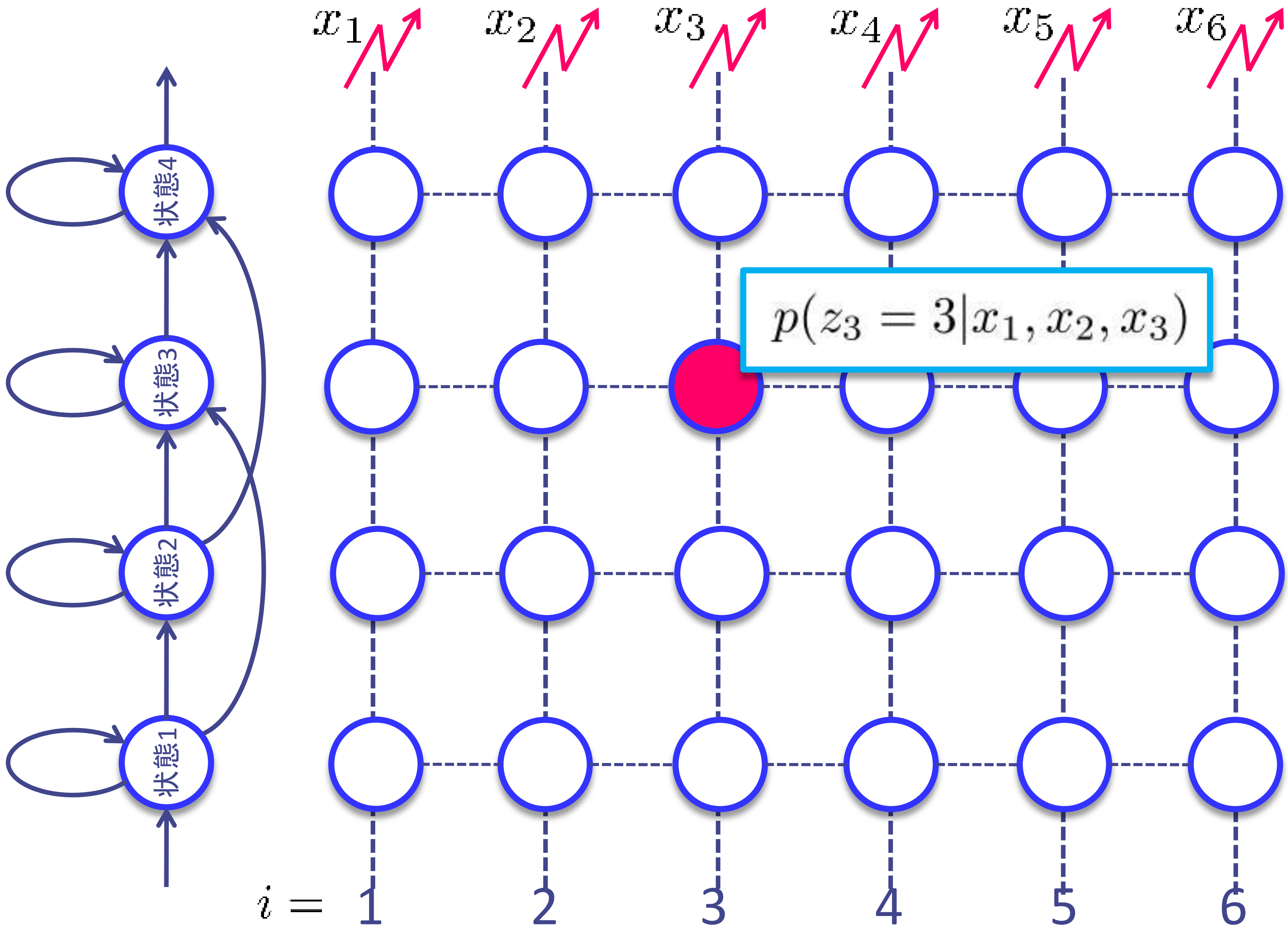
⇒フィルタリング問題

- 再帰式
  - $p(z_{k-1}|x_1, \ldots, x_{k-1})$ が求まっていたとしたら、
$$p(z_k|x_1, \ldots, x_{k-1}) = \sum_{z_{k-1}=1}^{J} p(z_k|z_{k-1})p(z_{k-1}|x_1, \ldots, x_{k-1})$$
により $p(z_k|x_1, \ldots, x_{k-1})$ が求まる
  - $p(z_k|x_1, \ldots, x_{k-1})$ が求まっていたとしたら、
$$p(z_k|x_1, \ldots, x_k) = \frac{p(x_k|z_k)p(z_k|x_1, \ldots, x_{k-1})}{\sum_{z_k=1}^{J} p(x_k|z_k)p(z_k|x_1, \ldots, x_{k-1})}$$
により $p(z_k|x_1, \ldots, x_k)$ が求まる
  - $k \leftarrow k+1$ として、上記を繰り返す

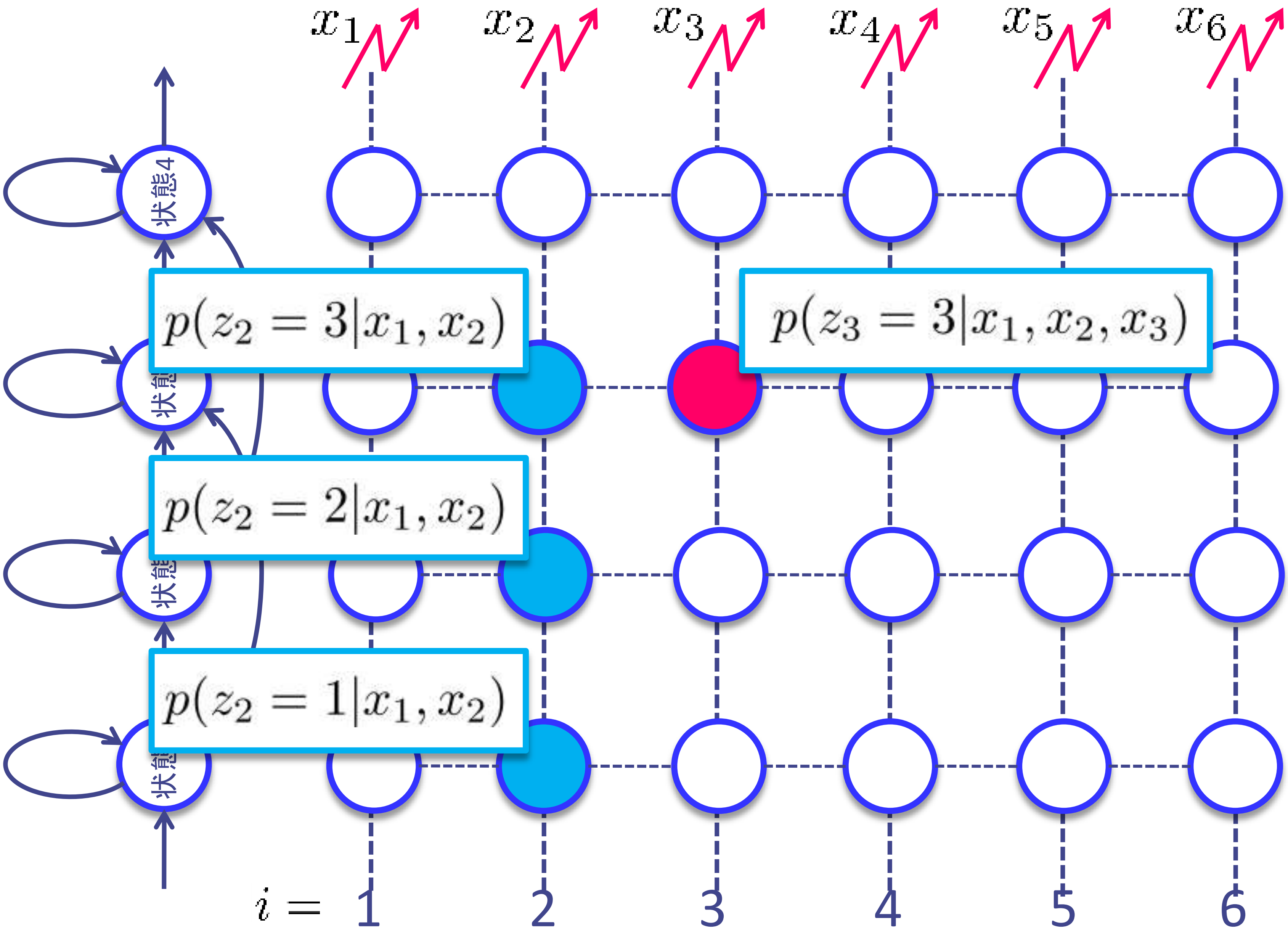
Forwardアルゴリズム



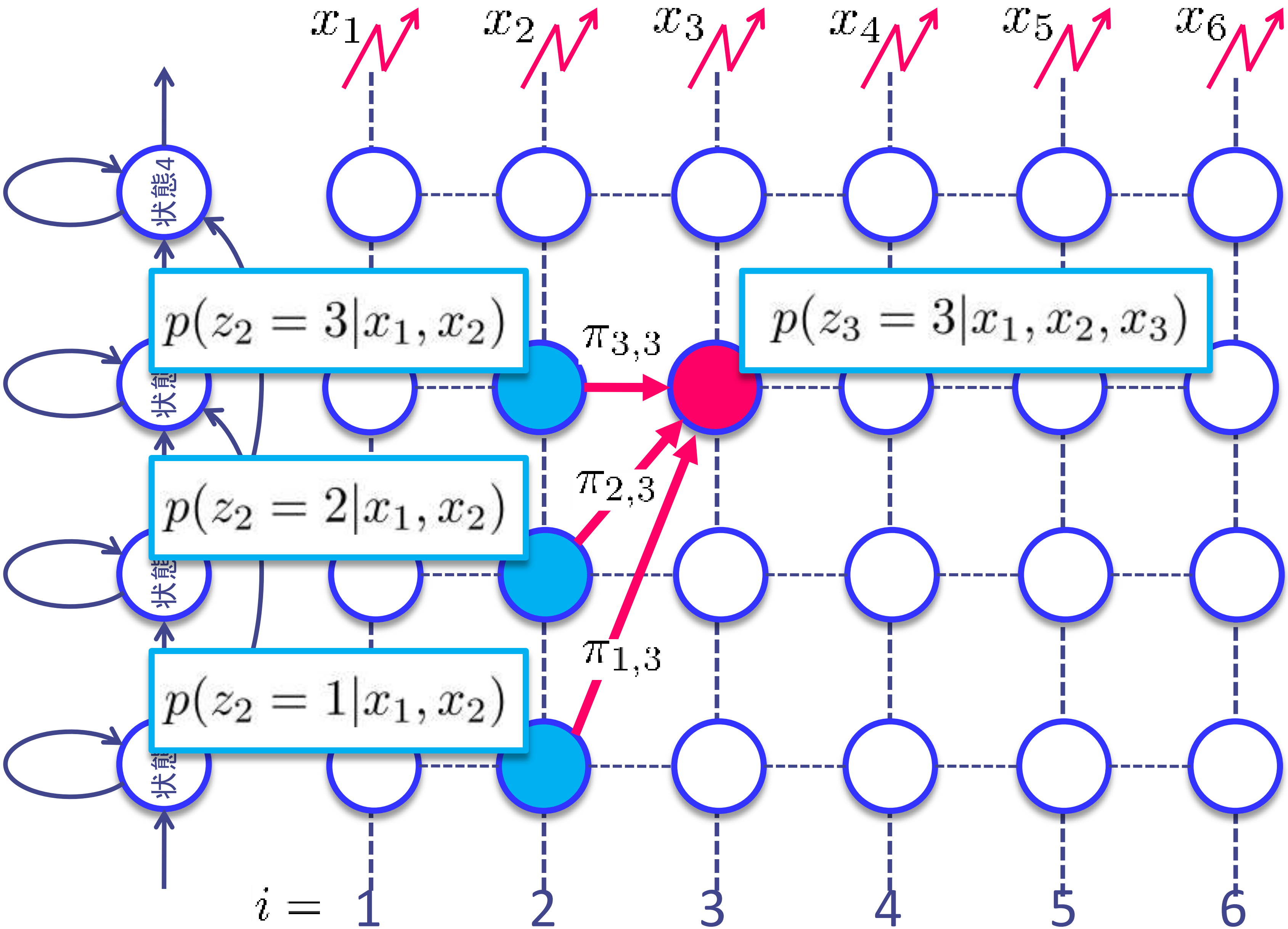
Forwardアルゴリズム



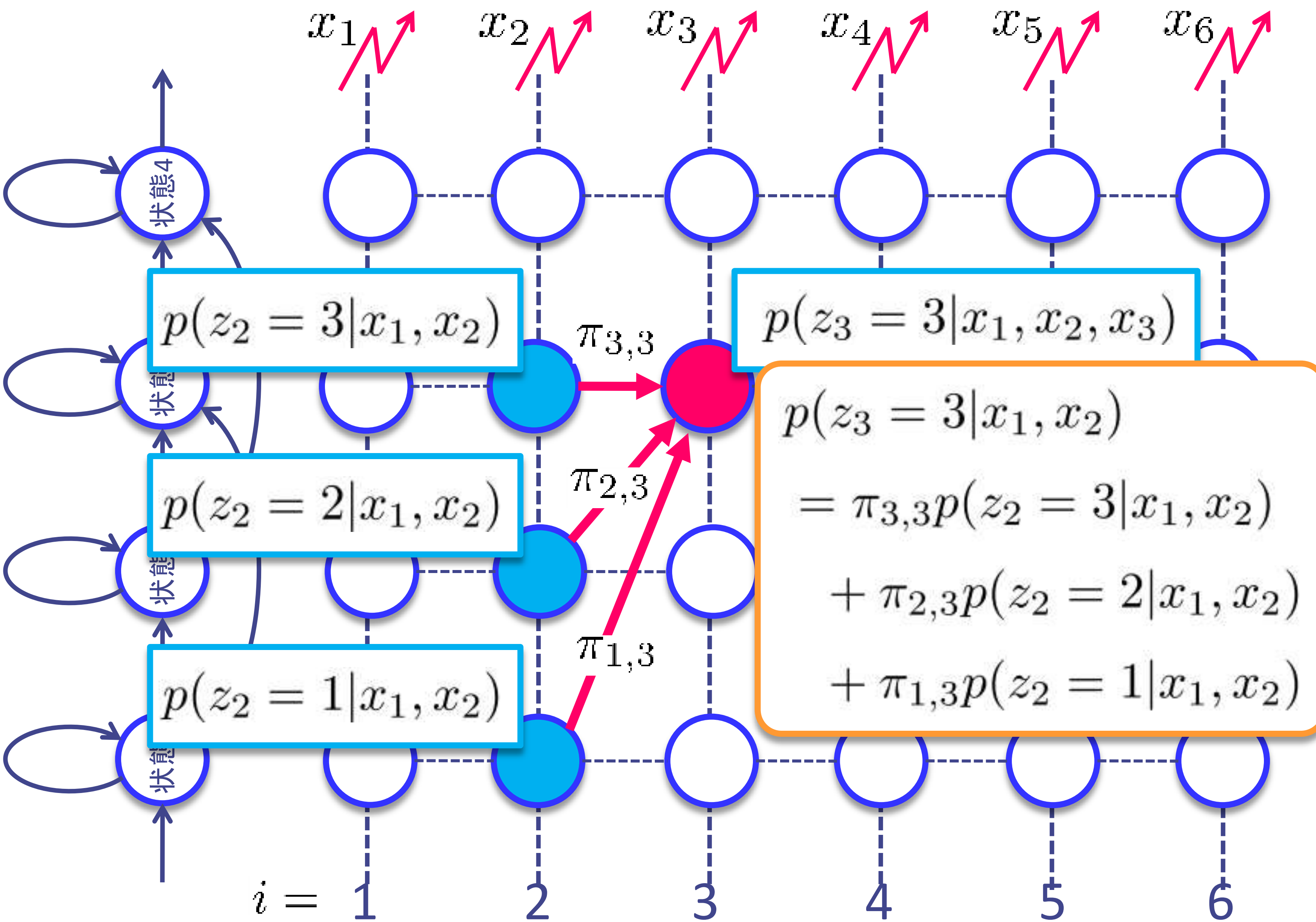
Forwardアルゴリズム



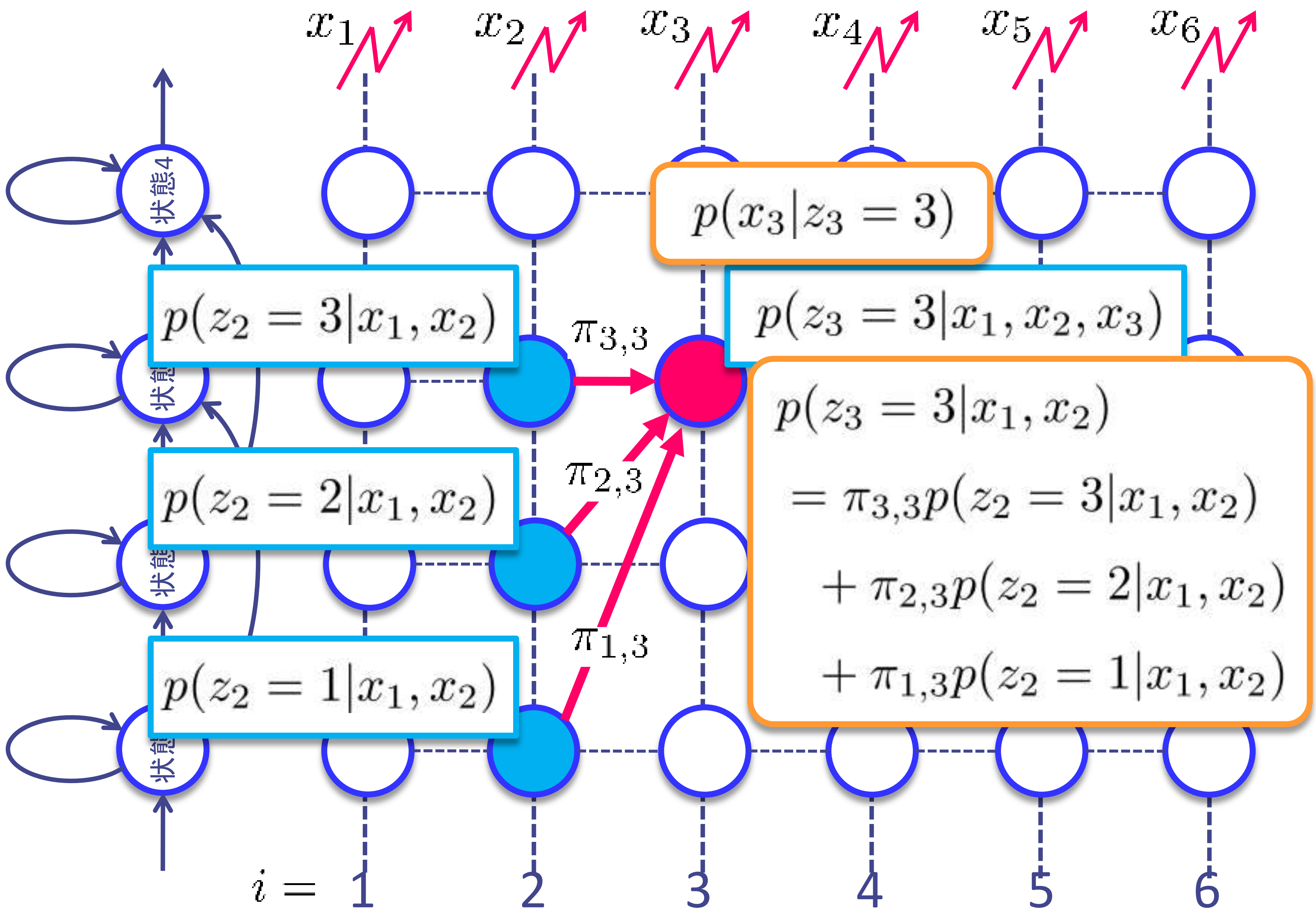
Forwardアルゴリズム



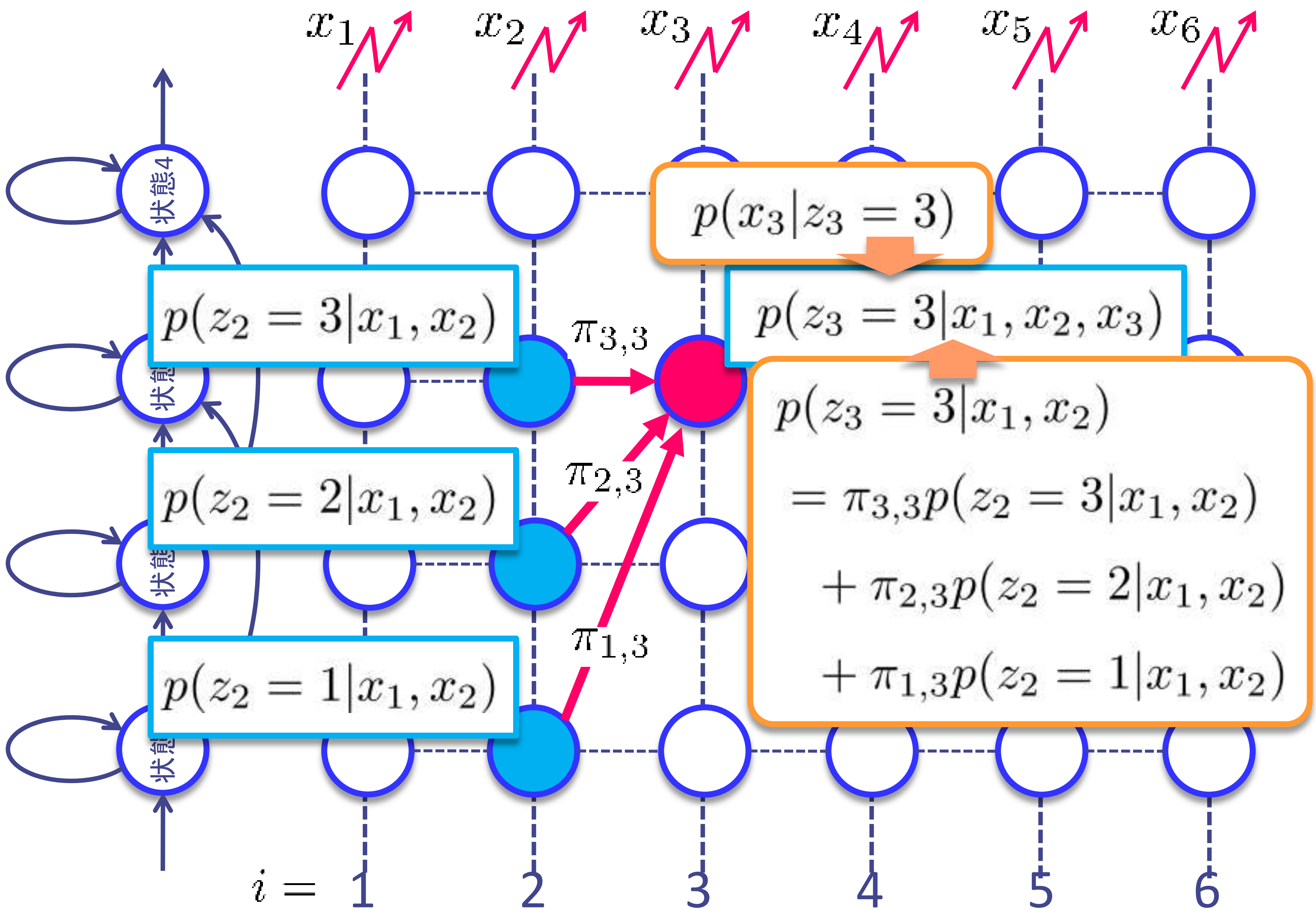
Forwardアルゴリズム



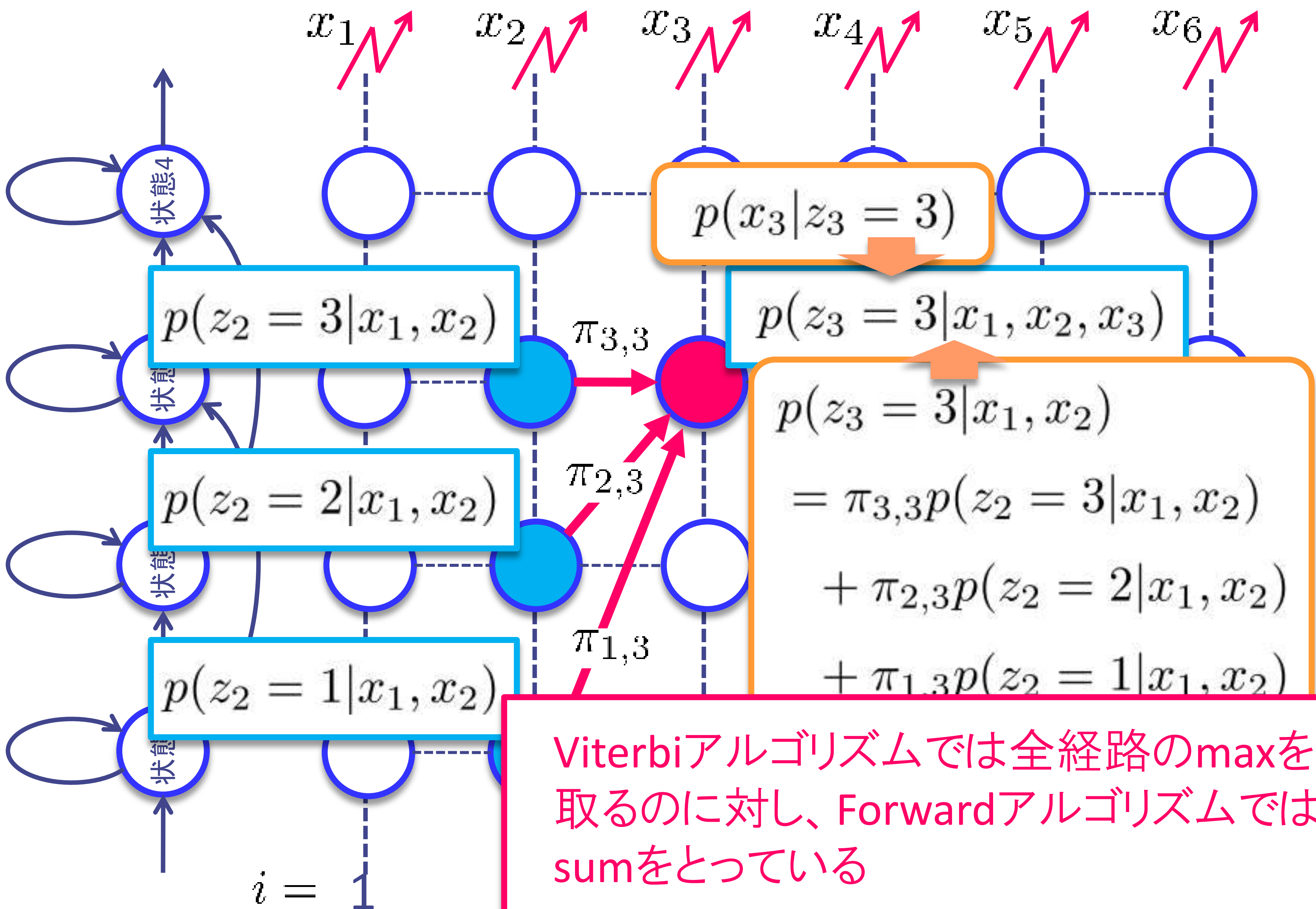
Forwardアルゴリズム



Forwardアルゴリズム



Forwardアルゴリズム



Viterbiアルゴリズムでは全経路のmaxを取るのに対し、Forwardアルゴリズムではsumをとっている

Kalmanフィルタとの関係

- Kalmanフィルタリングアルゴリズム
  - $p(z_{k-1}|x_1,\dots,x_{k-1})$ が求まっていたとしたら、

$$p(z_k|x_1,\dots,x_{k-1}) = \int p(z_k|z_{k-1})p(z_{k-1}|x_1,\dots,x_{k-1})\mathrm{d}z_{k-1}$$

    により $p(z_k|x_1,\dots,x_{k-1})$ が求まる
  - $p(z_k|x_1,\dots,x_{k-1})$ が求まっていたとしたら、

$$p(z_k|x_1,\dots,x_k) = \frac{p(x_k|z_k)p(z_k|x_1,\dots,x_{k-1})}{\int p(x_k|z_k)p(z_k|x_1,\dots,x_{k-1})\mathrm{d}z_k}$$

    により $p(z_k|x_1,\dots,x_k)$ が求まる
  - $k \leftarrow k+1$ として、上記を繰り返す
- Kalmanフィルタでは、
  - z_k が連続値
  - $p(z_k|z_{k-1})$, $p(x_k|z_k)$ がいずれも正規分布

HMMの基本問題

観測系列 $X_{1:T}$: $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$: $z_1, z_2, \ldots, z_T$

パラメータ Θ :

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列 $X_{1:T}$ が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T}, \Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する ⇒Viterbiアルゴリズム
- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t}, \Theta)$ を推定する ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列 $X_{1:T}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する ⇒Baum-Welchアルゴリズム

HMMの基本問題

観測系列 $X_{1:T}$: $x_1, x_2, \ldots, x_T$

状態系列 $Z_{1:T}$: $z_1, z_2, \ldots, z_T$

パラメータ Θ ：

状態出力分布パラメータ $\theta = \{\theta_i\}$ 、状態遷移確率 $\pi = \{\pi_{i,j}\}$

- 最適状態系列推定問題
  - 観測系列 $X_{1:T}$ が与えられたとき、状態系列の事後確率 $p(Z_{1:T}|X_{1:T}, \Theta)$ が最大となる状態系列 $Z_{1:T}$ を推定する

⇒Viterbiアルゴリズム

- 状態事後確率推定問題
  - 観測系列 $x_1, \ldots, x_t$ が与えられたとき、時刻 t に状態 z_t にいる確率 $p(z_t|X_{1:t}, \Theta)$ を推定する ⇒Forwardアルゴリズム
- パラメータ学習問題
  - 観測系列 $X_{1:T}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定する ⇒Baum-Welchアルゴリズム

最適パラメータ学習問題

- 観測系列 $X_{1:T} = \{x_1, \ldots, x_T\}$ を生成する確率 $p(X_{1:T}|\Theta)$ が最大となるパラメータ Θ を推定したい

- 対数尤度関数

$$\log p(X_{1:T}|\Theta) = \log \sum_{Z_{1:T}} p(X_{1:T}, Z_{1:T}|\Theta) = \log \sum_{Z_{1:T}} \underline{p(X_{1:T}|Z_{1:T}, \Theta)}\,\underline{p(Z_{1:T}|\Theta)}$$

$$\left\{ \begin{array}{l} \underline{p(X_{1:T}|Z_{1:T}, \Theta)} = \prod_{t=1}^{T} p(x_t|z_t, \Theta) = \prod_{t=1}^{T} \mathcal{N}(x_t; \mu_{z_t}, \sigma_{z_t}^2) \\ \qquad\qquad\qquad \text{状態出力分布がガウス分布の場合} \\ \underline{p(Z_{1:T}|\Theta)} = p(z_1) \prod_{t=2}^{T} p(z_t|z_{t-1}) = \pi_1 \prod_{t=2}^{T} \pi_{z_t, z_{t-1}} \end{array} \right.$$

最適パラメータ学習問題

- $$\log p(X_{1:T}|\Theta) = \log \sum_{Z_{1:T}} \left(\prod_{t=1}^{T} \mathcal{N}(x_t; \mu_{z_t}, \sigma^2_{z_t}) \right) \left(\pi_1 \prod_{t=2}^{T} \pi_{z_t, z_{t-1}} \right)$$

  を最大化する $\Theta = \{\mu, \sigma, \pi\}$ を求めたい
  - しかし解析的には求められない！
  - ちなみに，状態系列 $Z_{1:T}$ が既知であれば各正規分布の平均と分散，遷移確率の最尤値は解析的に求められる
    ⇒状態 i に対応した観測値の平均と分散を求めるだけ
    　状態 i から状態 j への遷移回数の集計をとるだけ
  - 解析的な求解を困難にしているのは

    $$\log \sum_{Z_{1:T}} p(X_{1:T}|Z_{1:T}, \Theta) p(Z_{1:T}|\Theta)$$

    におけるlogの中のsummation

補助関数法

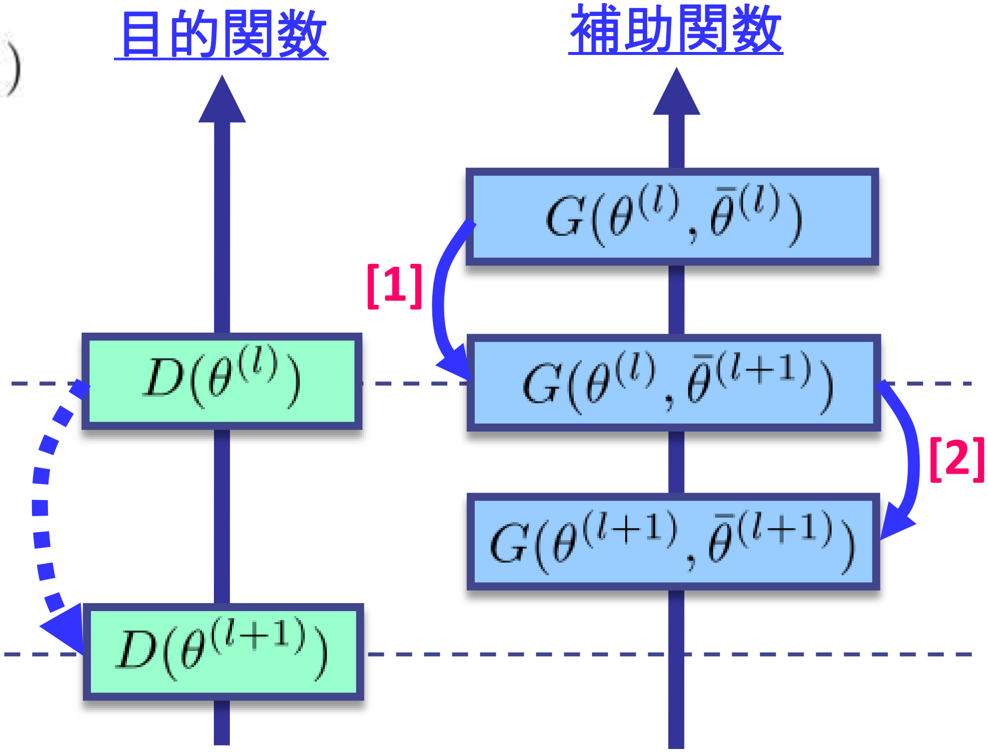
- 補助関数法
  - 目的関数 $D(\theta)$ を局所最小化するためのテクニック
  - $D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta})$ を満たす $G(\theta, \bar{\theta})$ を補助関数と定義
  - 反復アルゴリズム

[1] $\bar{\theta}^{(l+1)} = \underset{\bar{\theta}}{\operatorname{argmin}}\, G(\theta^{(l)}, \bar{\theta})$

[2] $\theta^{(l+1)} = \underset{\theta}{\operatorname{argmin}}\, G(\theta, \bar{\theta}^{(l+1)})$

  - 収束性

$$
\begin{aligned}
D(\theta^{(l)}) &= G(\theta^{(l)}, \bar{\theta}^{(l+1)}) \\
&\geq G(\theta^{(l+1)}, \bar{\theta}^{(l+1)}) \\
&\geq D(\theta^{(l+1)})
\end{aligned}
$$



Jensenの不等式

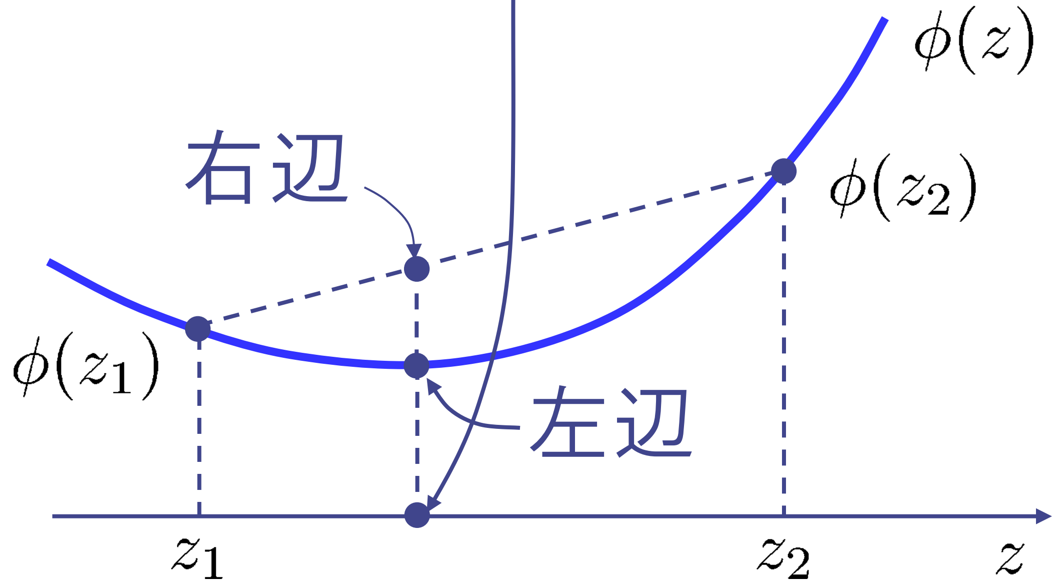
- Jensenの不等式
  - $\phi(\cdot)$：凸関数
  - $\lambda_i \geq 0, \ \sum_i \lambda_i = 1$

$$\Rightarrow \quad \phi\left(\sum_i \lambda_i z_i\right) \leq \sum_i \lambda_i \phi(z_i)$$

例えば,

$\phi(z) = -\log z$ の場合：

$$-\log\left(\sum_i \lambda_i z_i\right) \leq -\sum_i \lambda_i \log z_i$$



対数尤度の補助関数

- 対数尤度 $\log \sum_{Z_{1:T}} p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)$ の下限関数

$$\log \sum_{Z_{1:T}} p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)$$

$$= \log \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{\lambda_{Z_{1:T}}}$$

対数関数が凹関数であることに注意してJensenの不等式を適用

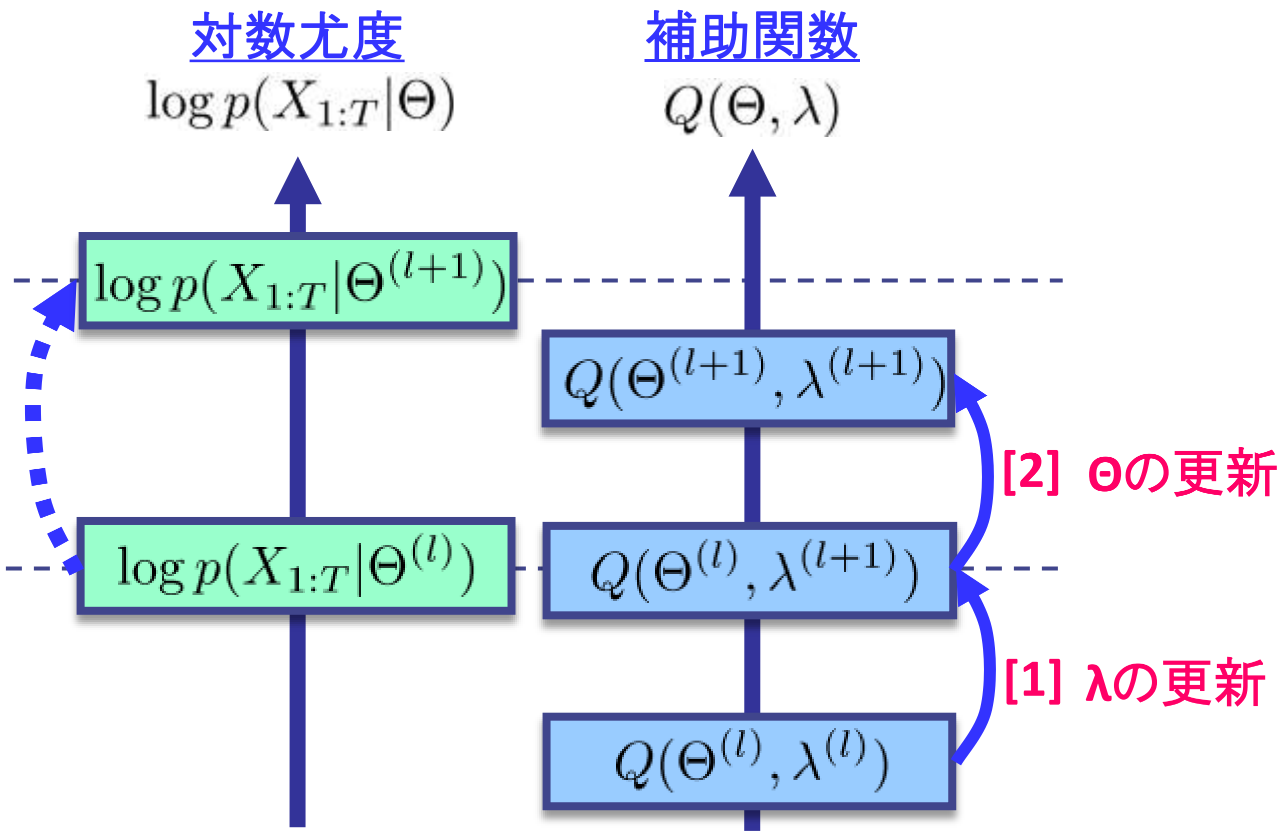
$$\geq \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \log \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{\lambda_{Z_{1:T}}}$$

$$Q(\Theta,\lambda) \equiv \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \log \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{\lambda_{Z_{1:T}}}$$ は

$\log p(X_{1:T}|\Theta)$ の 補助関数！

Baum-Welchアルゴリズム (EMアルゴリズム)

- 対数尤度と補助関数の関係： $\min_{\lambda} Q(\Theta, \lambda) = \log p(X_{1:T}|\Theta)$
- 反復アルゴリズムによる局所最大化



Baum-Welchアルゴリズム (EMアルゴリズム)

■λの更新式

$$\lambda_{Z_{1:T}} \leftarrow \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{\sum_{Z_{1:T}} p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)} = \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{p(X_{1:T}|\Theta)}$$

$$= p(Z_{1:T}|X_{1:T},\Theta)$$

■補助関数

$$Q(\Theta,\lambda) \equiv \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \log \frac{p(X_{1:T}|Z_{1:T},\Theta)p(Z_{1:T}|\Theta)}{\lambda_{Z_{1:T}}}$$

$$= \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \log \left(\prod_{t=1}^{T} \mathcal{N}(x_t;\mu_{z_t},\sigma_{z_t}^2)\right)\left(\pi_1 \prod_{t=2}^{T} \pi_{z_t,z_{t-1}}\right) + \text{const}$$

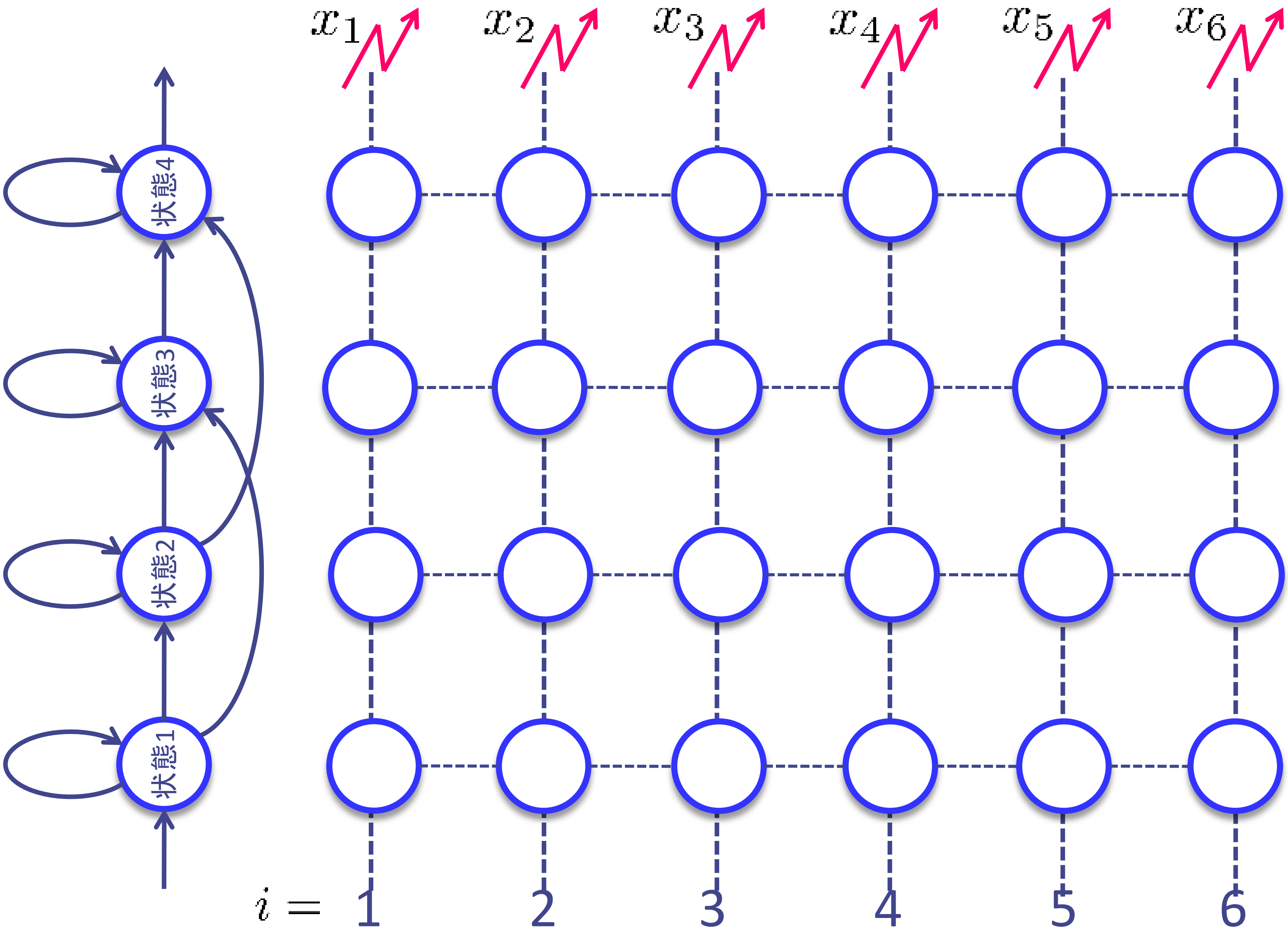
$$= \sum_{Z_{1:T}} \lambda_{Z_{1:T}} \left(\sum_{t=1}^{T} \underline{\log \mathcal{N}(x_t;\mu_{z_t},\sigma_{z_t}^2)} + \underline{\log \pi_1} + \sum_{t=2}^{T} \underline{\log \pi_{z_t,z_{t-1}}}\right) + \text{const}$$

Baum-Welchアルゴリズム (EMアルゴリズム)

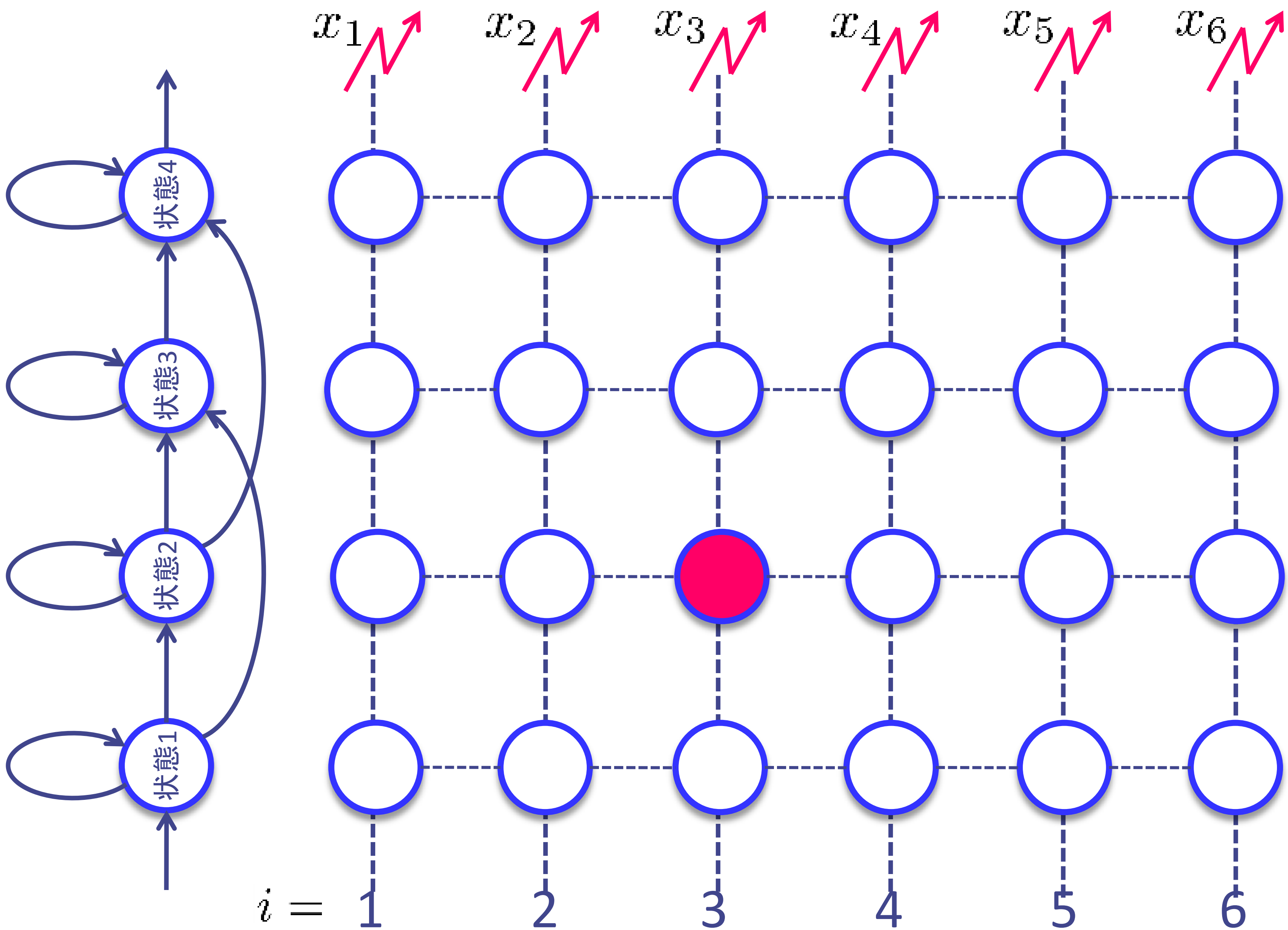
- $$Q(\Theta,\lambda) \equiv \sum_{Z_{1:T}} p(Z_{1:T}|X_{1:T},\Theta') \sum_{t=1}^{T} \underline{\log \mathcal{N}(x_t;\mu_{z_t},\sigma^2_{z_t})}$$
$$+ \sum_{Z_{1:T}} p(Z_{1:T}|X_{1:T},\Theta') \underline{\log \pi_1}$$
$$+ \sum_{Z_{1:T}} p(Z_{1:T}|X_{1:T},\Theta') \sum_{t=2}^{T} \underline{\log \pi_{z_t,z_{t-1}}} + \text{const}$$
$$= \sum_{t=1}^{T} \sum_{z_t} \underline{p(z_t|X_{1:T},\Theta')}\, \underline{\log \mathcal{N}(x_t;\mu_{z_t},\sigma^2_{z_t})}$$
$$+ \sum_{z_1} \underline{p(z_1|X_{1:T},\Theta')}\, \underline{\log \pi_1}$$
$$\cdots p(z_t|X_{1:T},\Theta') = \sum_{z_{t-1}=1}^{J} p(z_t,z_{t-1}|X_{1:T},\Theta')$$
$$+ \sum_{t=2}^{T} \sum_{z_t} \sum_{z_{t-1}} \underline{p(z_t,z_{t-1}|X_{1:T},\Theta')}\, \underline{\log \pi_{z_t,z_{t-1}}} + \text{const}$$

これを求めるには！？

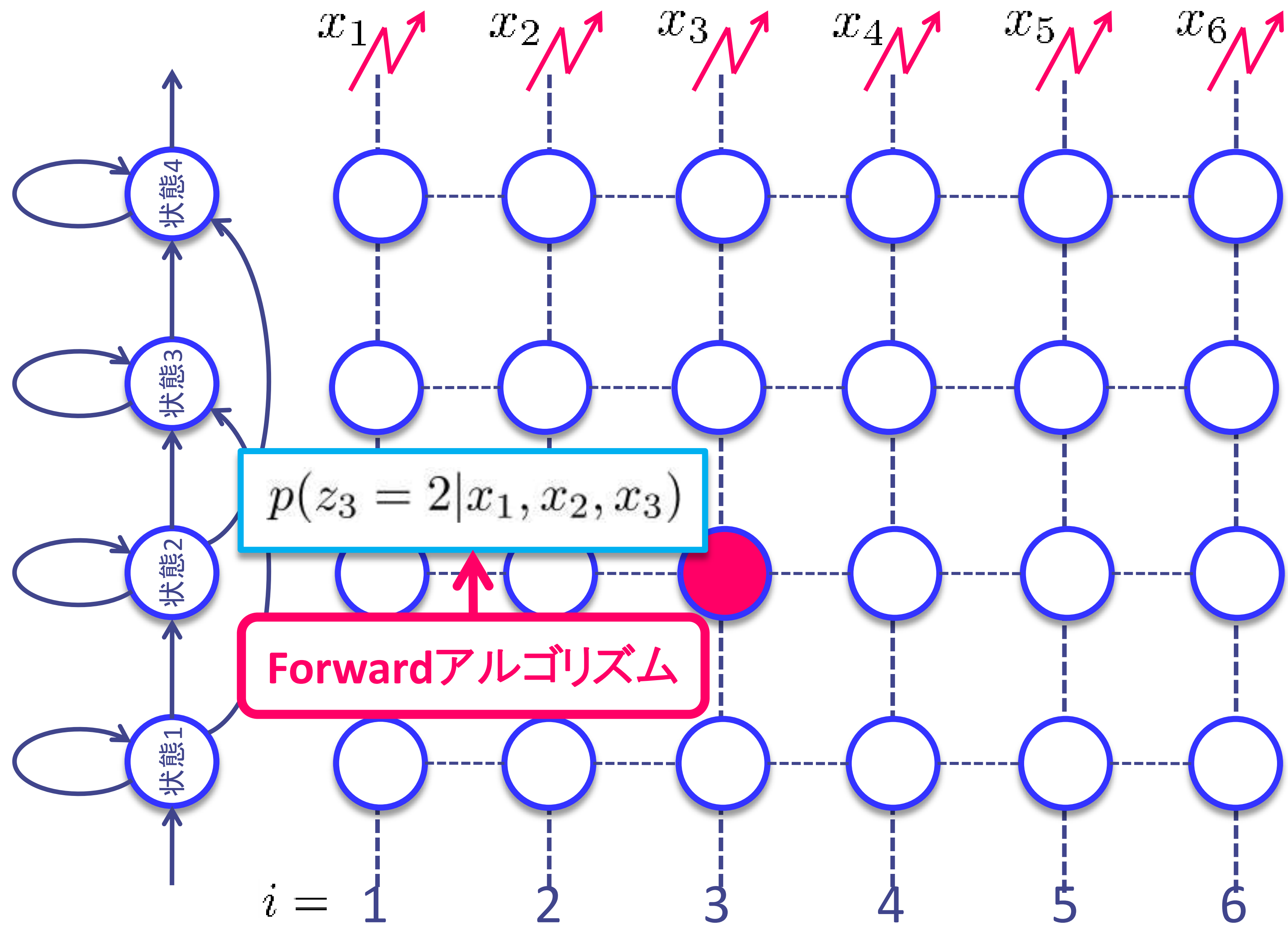
Forward-Backwardアルゴリズム



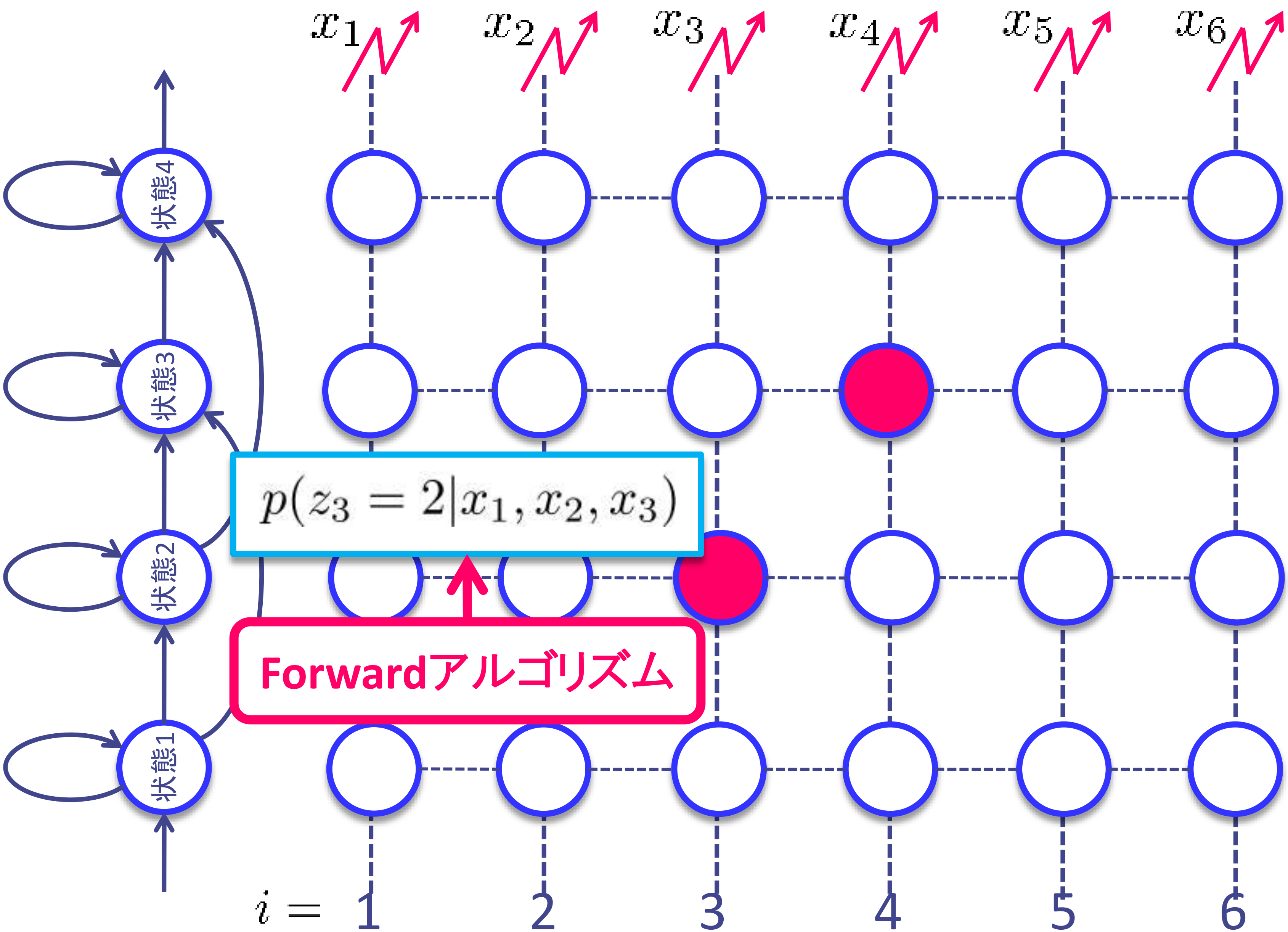
Forward-Backwardアルゴリズム



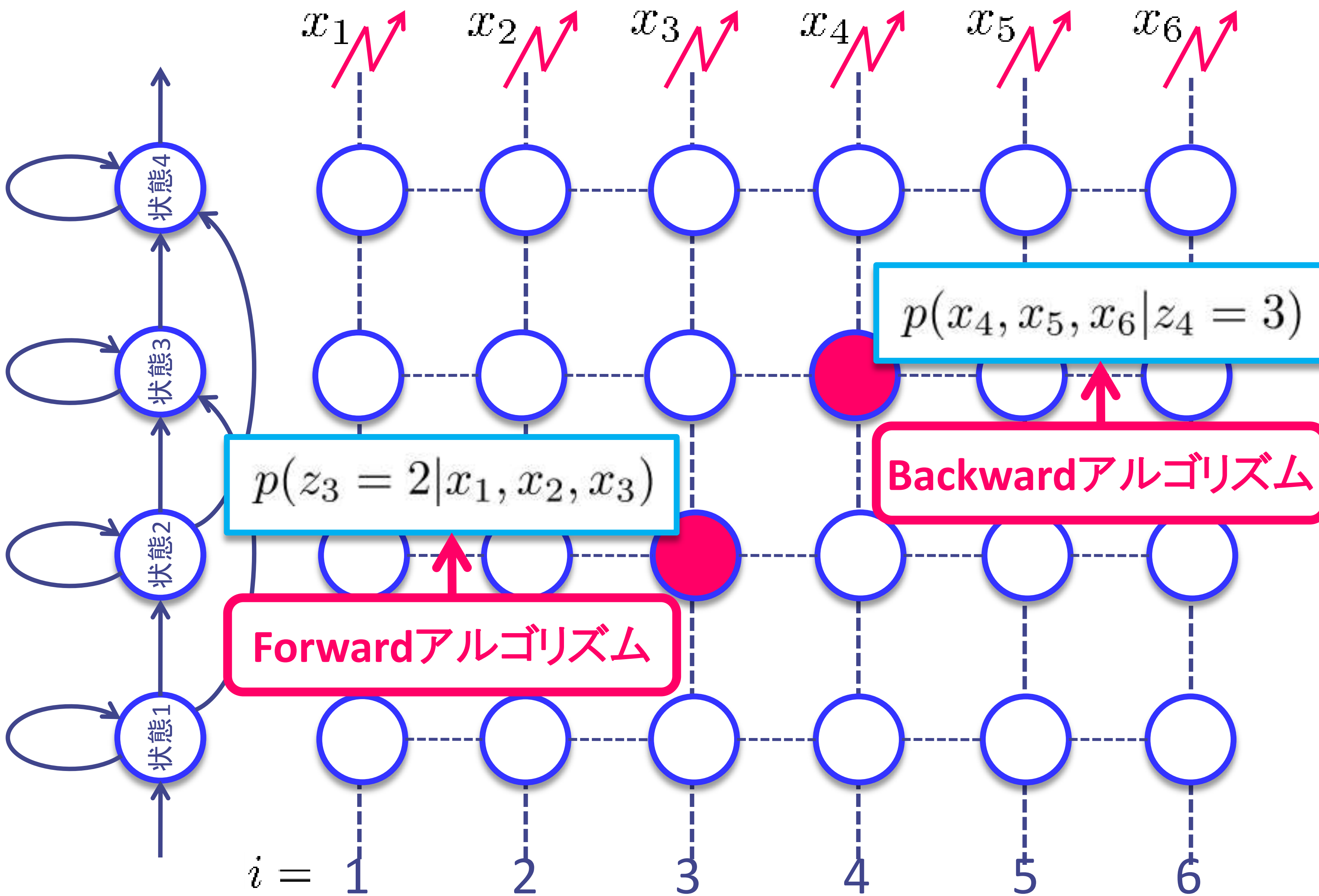
Forward-Backwardアルゴリズム



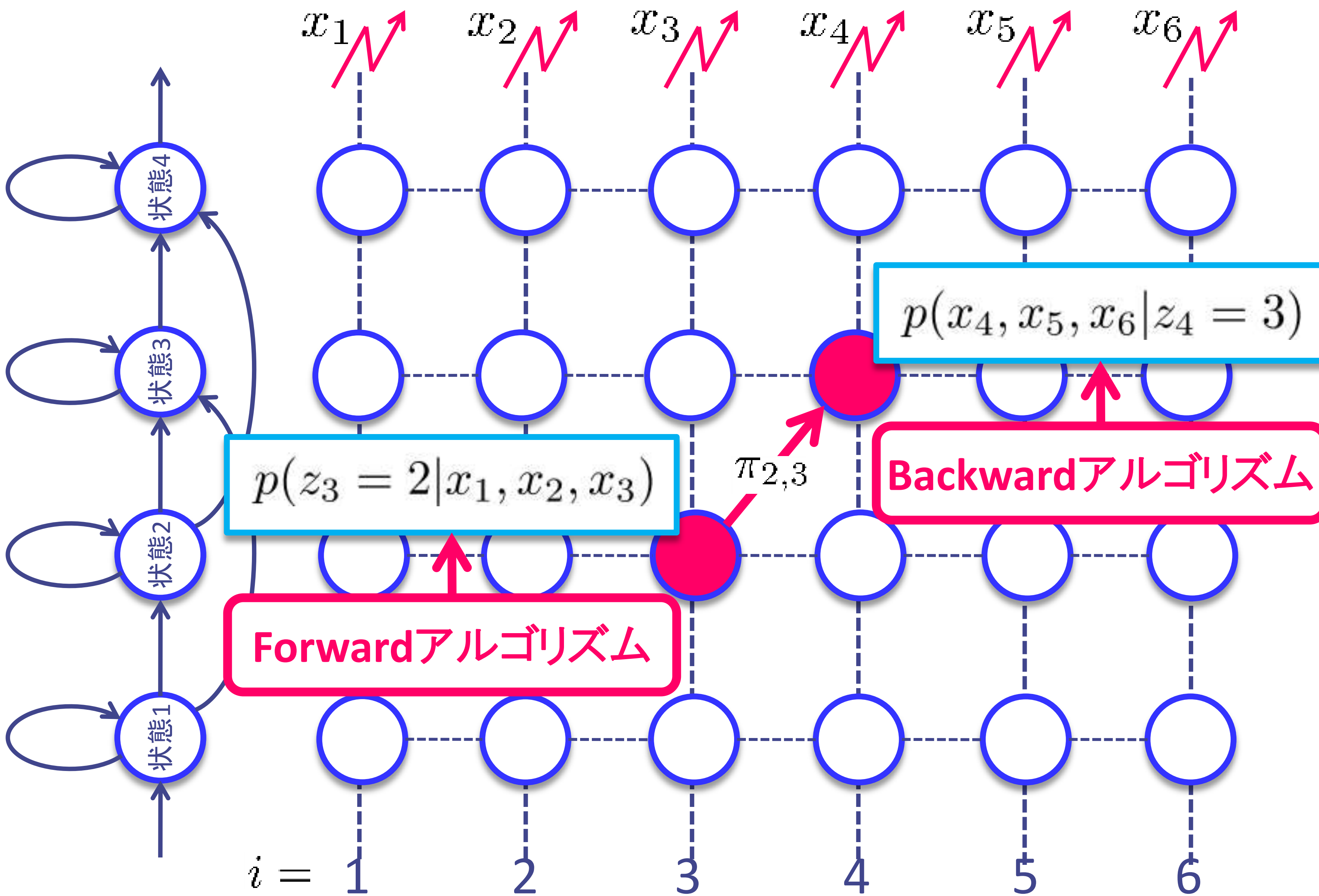
Forward-Backwardアルゴリズム



Forward-Backwardアルゴリズム



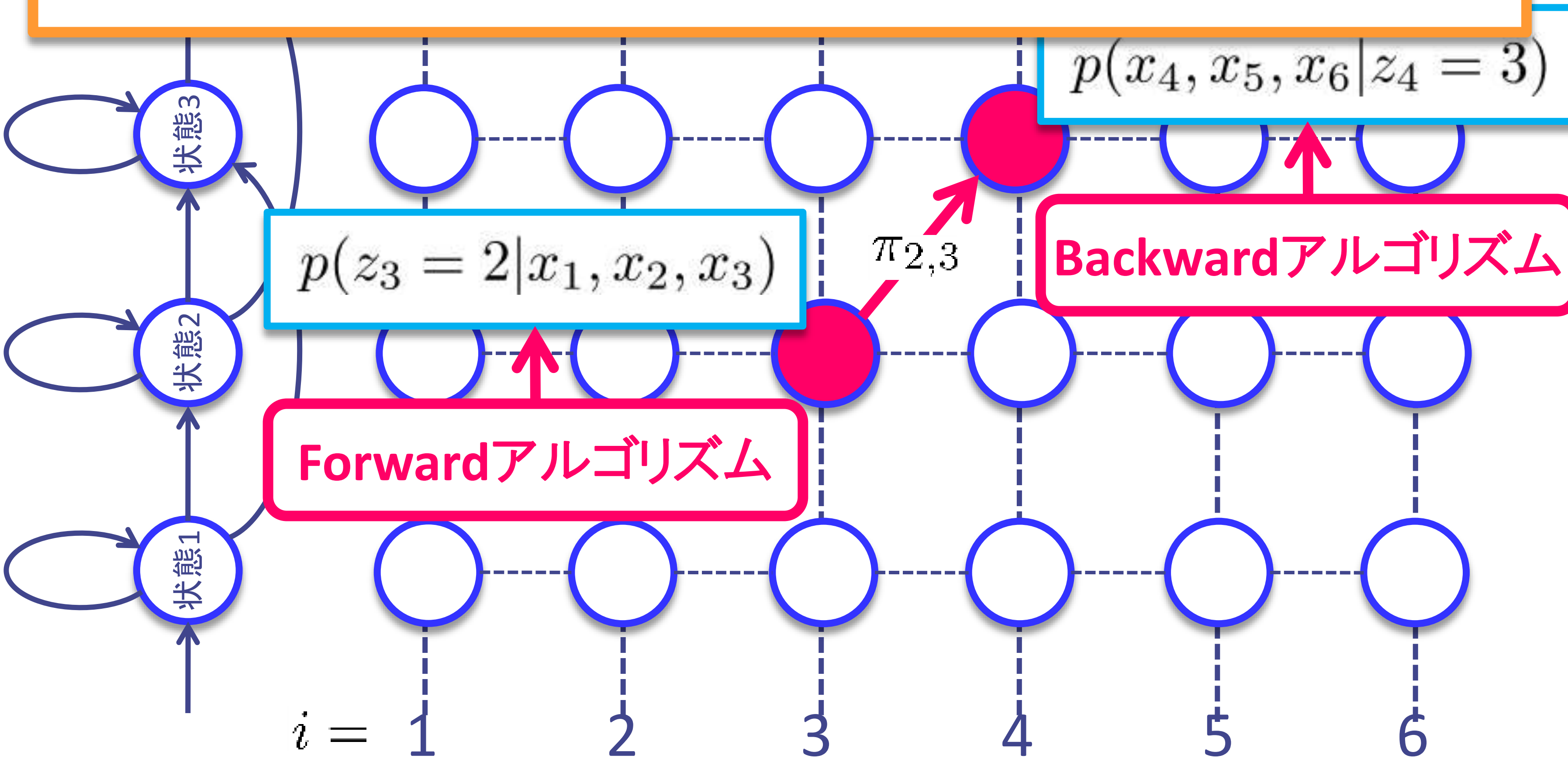
Forward-Backwardアルゴリズム



Forward-Backwardアルゴリズム

$$p(z_3 = 2, z_4 = 3 | x_1, \ldots, x_6)$$ （←求めたかったもの！）

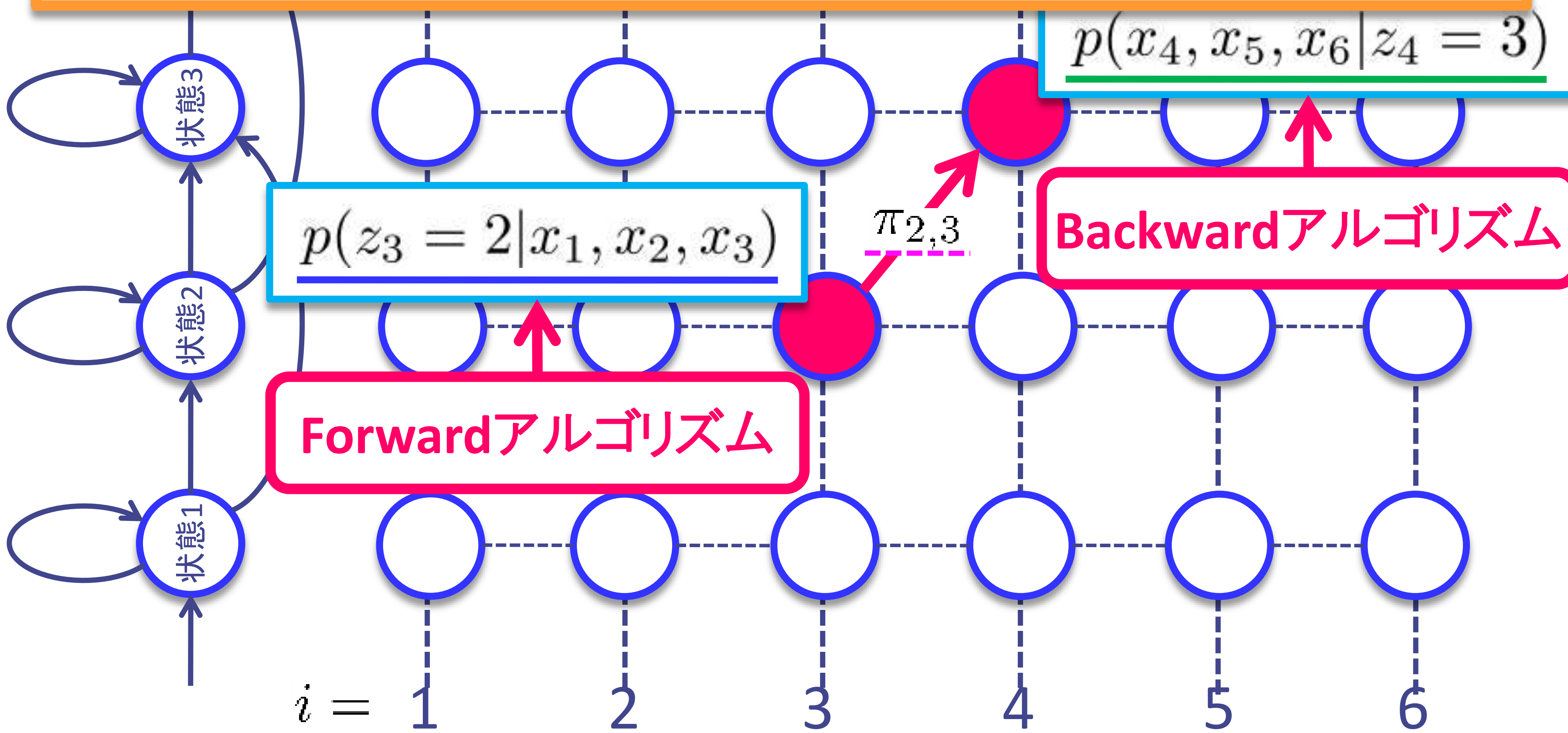
$$\begin{aligned}
&\propto \; p(z_3 = 2, z_4 = 3, x_4, x_5, x_6 | x_1, x_2, x_3) \\
&= \; p(z_4 = 3, z_3 = 2 | x_1, x_2, x_3) p(x_4, x_5, x_6 | z_4 = 3, \cancel{z_3 = 2}) \\
&= \; p(z_3 = 2 | x_1, x_2, x_3) p(z_4 = 3 | z_3 = 2) p(x_4, x_5, x_6 | z_4 = 3)
\end{aligned}$$



Forward-Backwardアルゴリズム

$$p(z_3=2, z_4=3 | x_1, \ldots, x_6)$$ （←求めたかったもの！）

$$\begin{aligned}
&\propto p(z_3=2, z_4=3, x_4, x_5, x_6 | x_1, x_2, x_3) \\
&= p(z_4=3, z_3=2 | x_1, x_2, x_3)\, p(x_4, x_5, x_6 | z_4=3, \cancel{z_3=2}) \\
&= \underline{p(z_3=2 | x_1, x_2, x_3)}\, \underbrace{p(z_4=3 | z_3=2)}_{\pi_{2,3}}\, \underline{p(x_4, x_5, x_6 | z_4=3)}
\end{aligned}$$



Baum-Welchアルゴリズム (EMアルゴリズム)

- 補助関数

$$Q(\Theta,\lambda) = \sum_{t=1}^{T}\sum_{z_t} p(z_t|X_{1:T},\Theta')\log\mathcal{N}(x_t;\mu_{z_t},\sigma_{z_t}^2)$$
$$+\sum_{z_1} p(z_1|X_{1:T},\Theta')\log\pi_1$$
$$+\sum_{t=2}^{T}\sum_{z_t}\sum_{z_{t-1}} p(z_t,z_{t-1}|X_{1:T},\Theta')\log\pi_{z_t,z_{t-1}} + \mathrm{const}$$

  - Θ の更新

$$\Theta \leftarrow \underset{\Theta}{\operatorname{argmax}}\, Q(\Theta,\lambda)$$ ←解析的に求まる！

HMMの応用先

- HMMは非定常時系列のモデル
  - 時系列：音声認識、話題遷移モデル、音楽リズム認識、経済モデル
  - 記号列：自然言語モデル化、音楽和声づけ
  - 空間順序：画像処理への応用
  - 構造順序：遺伝子解析
- 学習の観点から分類
  - モデルパラメータ(確率)を理論・直観により与える場合
  - モデルパラメータ(確率)の一部のみ学習データから学習する場合
  - モデルパラメータ(確率)のすべてを学習データから学習する場合

学習データ量と質によって使い分け