



東京大学 工学部 計数工学科/物理工学科

# 応用音響学：隠れマルコフモデル (HMM)

嵯峨山 茂樹 <[sagayama@hil.t.u-tokyo.ac.jp](mailto:sagayama@hil.t.u-tokyo.ac.jp)>

東京大学 工学部 計数工学科

<http://hil.t.u-tokyo.ac.jp/sagayama/applied-acoustics/>

---

- 定常信号源
- マルコフ遷移
- 隠れマルコフモデル
- 状態遷移確率と出力確率
- **Trellis** 表現
- 離散出力分布と連続出力分布



# 参考書/参考文献

---

- 鹿野, 中村, 伊勢, 「音声・音情報のデジタル信号処理」, 昭晃堂, 1997.
- 中川聖一, 「確率モデルによる音声認識」, 電子情報通信学会, コロナ社.
- Laurence Rabiner, Biing-Hwang Juang: **Fundamentals of Speech Recognition**, (Translated under supervision of S. Furui), NTT Advanced Technologies, 1993; Laurence Rabiner, Biing-Hwang Juang 「音声認識の基礎(上, 下)」(古井貞熙 監訳), NTT アドバンステクノロジー, 1995.
- 古井 貞熙 「音声情報処理」 森北出版
- 谷萩 隆嗣 「音声と画像のデジタル信号処理」 コロナ社
- 嵯峨山茂樹: 「数理統計モデルによる音声認識の現状と将来」, 日本音響学会誌 48巻第1号, 1992.
- L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” *proc. IEEE*, Vol. 77, No. 2, pp. 257-286, 1989.



# 確率モデル：確率に関する誤解を解こう

- 確率事象とは、確率に支配されている現象？
- ある女性が近々出産：生まれて来る子は男か女か
- 熊雄が傘を持って現れた。外は雨か否か？
- 自然言語を確率で考えるのは意味がないと思っている人がいる(いた)
- C. E. Shannon の情報理論は、確率論に基づいて情報を定義
- 確率と統計の区別が分からない人：  
 $n$  回の試行のうち  $m$  回成功するなら、成功する確率は  $\frac{m}{n}$  か？
- 確率モデルは、不確かな情報を生かす実に便利な思想
- 人間は統計情報処理システムであるという仮説



# 定常信号源

- 観測できる量  $y_t$  の系列: (スカラー量、ベクトル、離散値、シンボルなど)
- 系列内番号  $t$  の値  $y_t$  が一定の確率分布に従って得られる: 定常信号源
- 自己相関がない場合を考える
- $y_t$ : 離散的なスカラー量ならば、ヒストグラムの形式

$$p(i) = P\{y_t = \eta_i\} = p_i$$

- $y_t$ : 正規分布をなすスカラー量ならば、分布密度は  $t$  によらずに正規分布

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- $y_t$ :  $n$ 次元のベクトルならば、同時正規分布

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right\}$$

対角共分散:  $\sigma_{ij}^2 = 0, \quad i \neq j$

- 混合正規分布: 凹凸がたくさんあるような複雑な分布を、複数の正規分布の和として表現

$$f(\mathbf{y}) = \sum_{i=1}^m w_i \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad \sum_i w_i = 1$$



# 非定常信号源のモデル: 定常信号源の切替え

- 時系列解析・確率過程論: 定常時系列を扱う点で高度に進歩
- 実際は、非定常であるがゆえに情報が伝達できる
- 最も簡単なモデルは、異なる定常情報源を次々に切替えるモデル

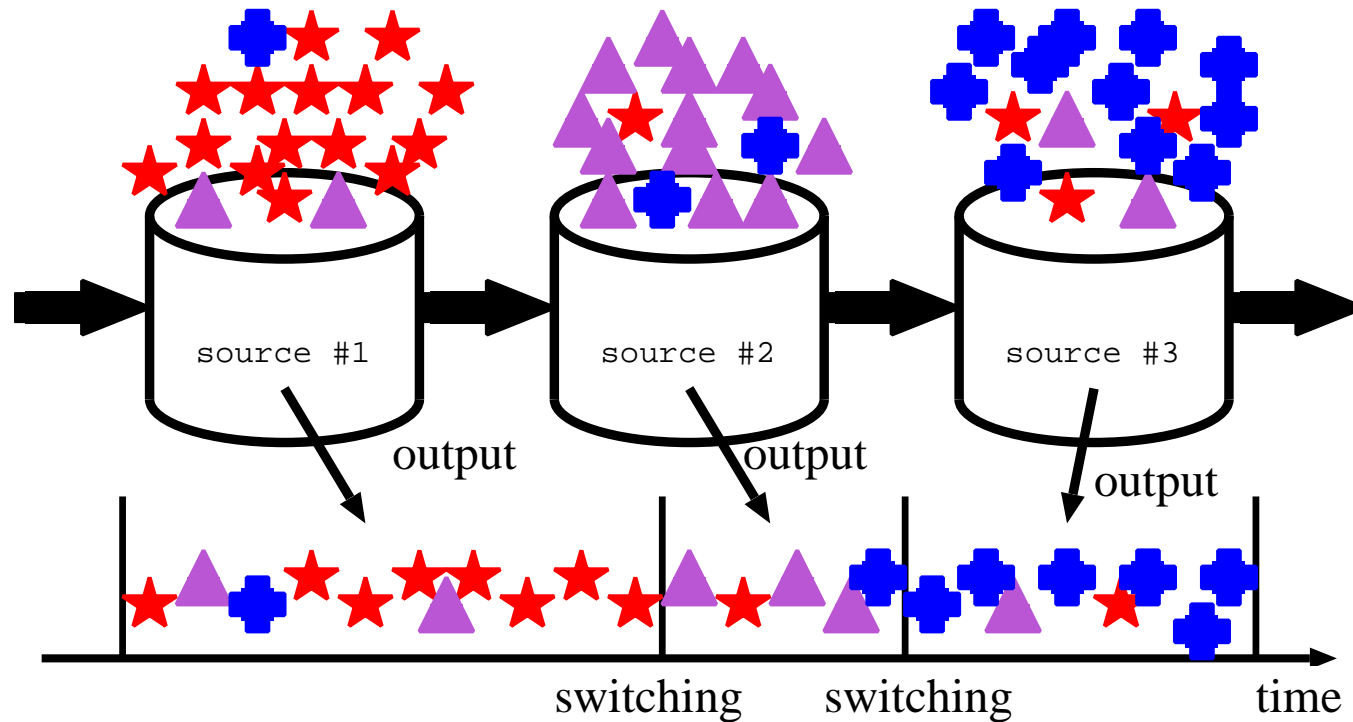


図1. 定常信号源の切り替えで非定常な時系列を表現する



# 非定常信号源のモデル: 定常信号源の切替え

## ■ 連続信号の場合

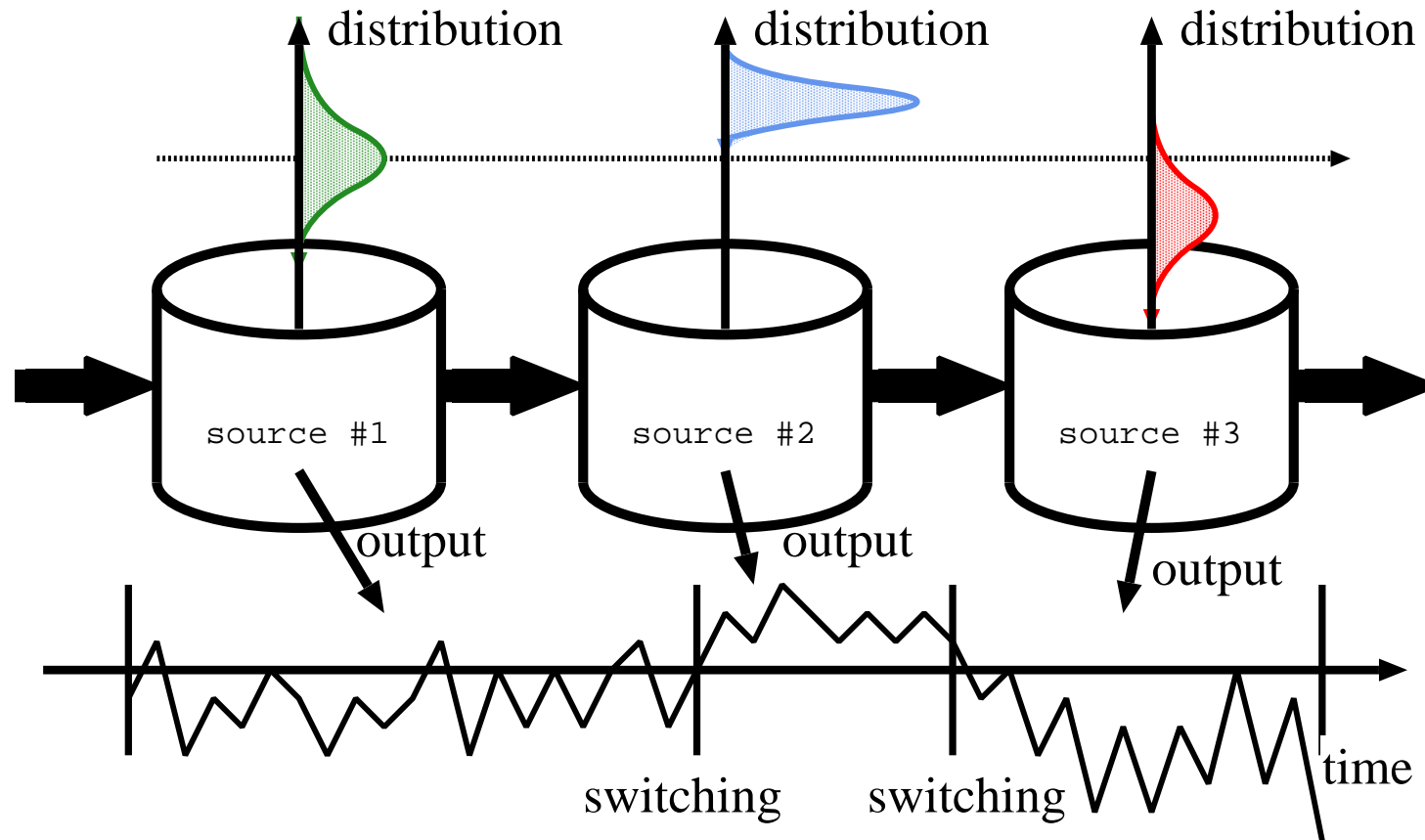
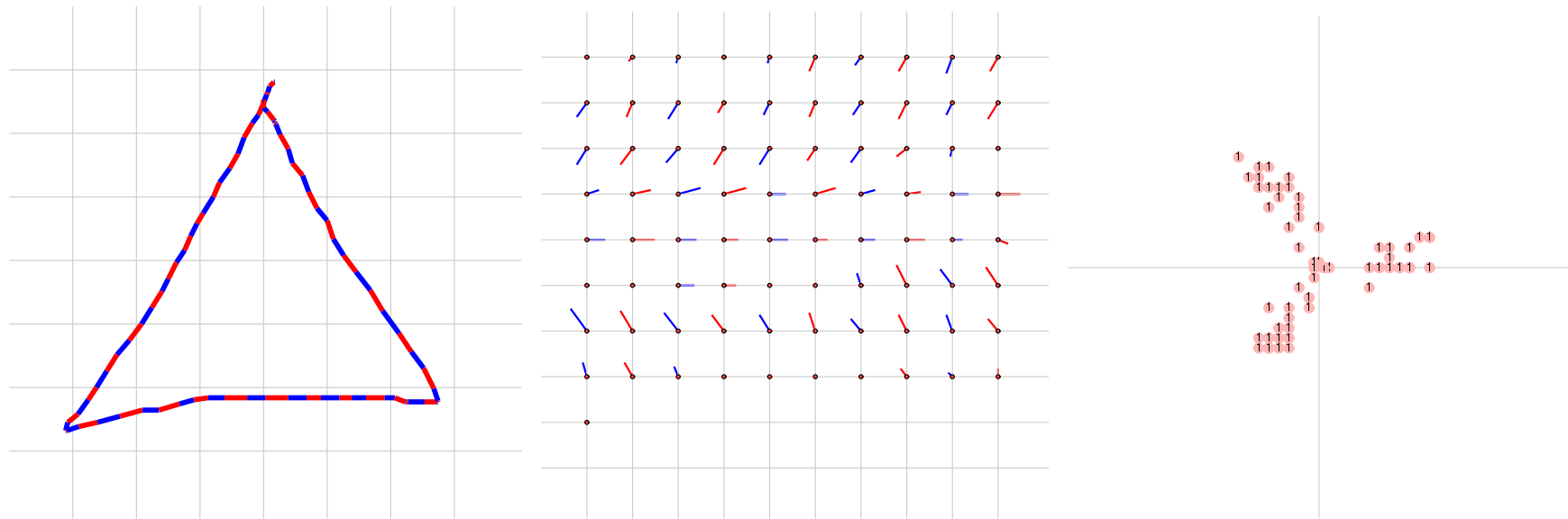


図2. 連続な信号も定常信号源の切り替えで表現できる



# 例：手描き三角形の速度ベクトル系列

## ■ 手描き三角形、その速度ベクトル系列、その空間分布



三角形を描くペン速度ベクトルの分布: 定常信号源の切替えモデルで表現できる



# マルコフモデル (Markov Model)

- ある順序で確率的に状態が変化:  $a_{ij}$  (状態  $i$  から状態  $j$  へ変化する確率)

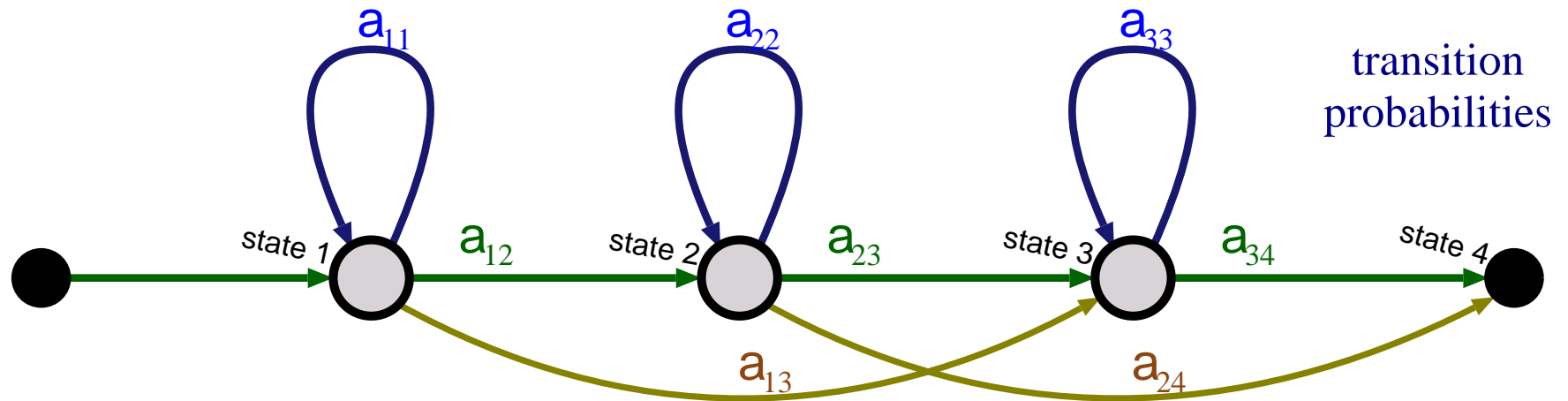


図3. マルコフ遷移の例

- マルコフ遷移により信号源を切替える      隠れマルコフモデル





# 隠れマルコフモデル (Hidden Markov Model: HMM)

- 統計的な性質が異なる複数の定常信号源： $b_i(y_t)$
- ある順序で確率的(マルコフ的)に信号源を切替え： $a_{ij}$

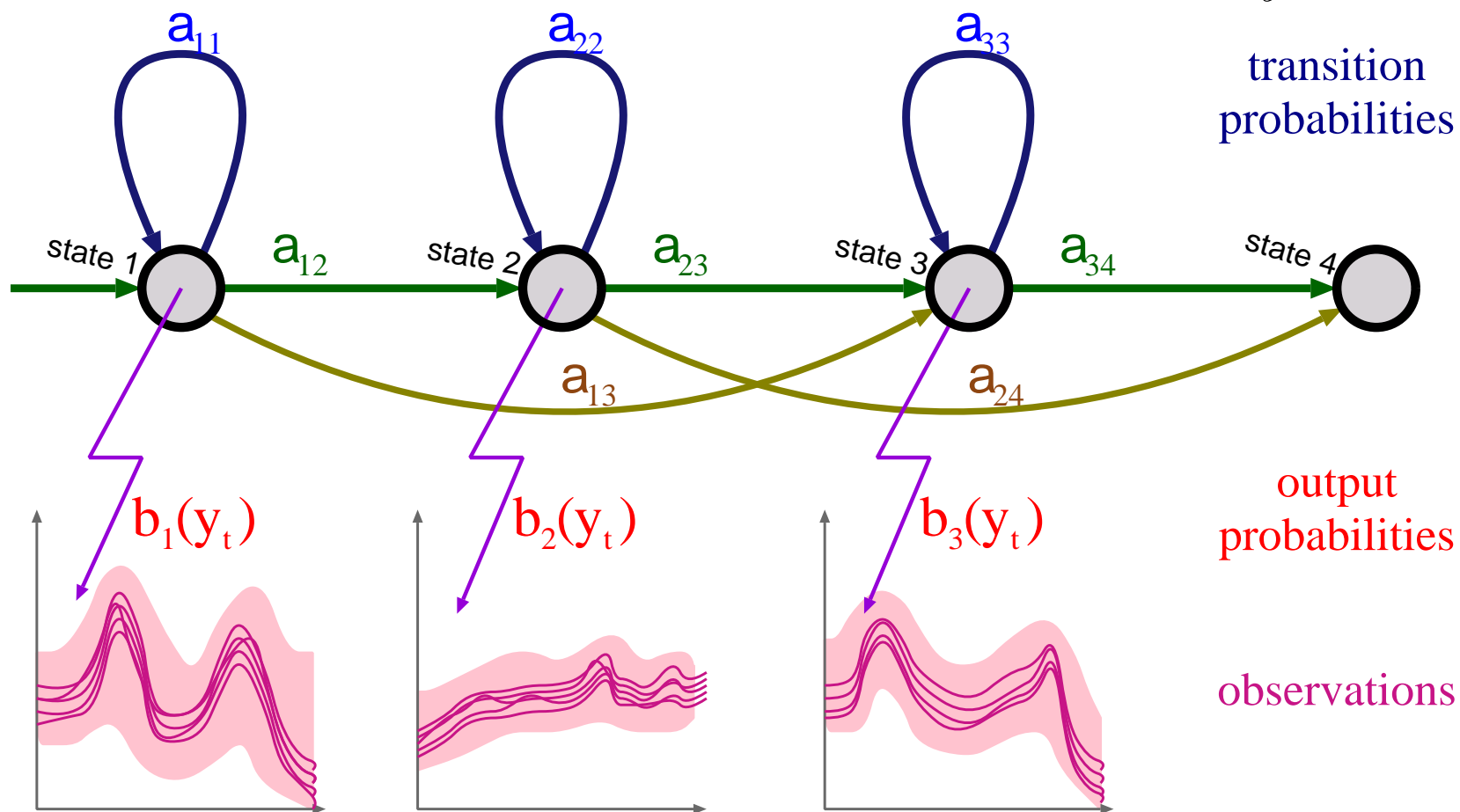


図4. 隠れマルコフモデル(HMM)の例



# 隠れマルコフモデルの2種

## ■ 状態出力型 (Mealy 型)

- 状態遷移確率:  $a_{ij}$ 、出力確率:  $b_i(y_t)$

## ■ 遷移出力型 (Moore 型)

- 状態遷移確率:  $a_{ij}$ 、出力確率:  $b_{ij}(y_t)$

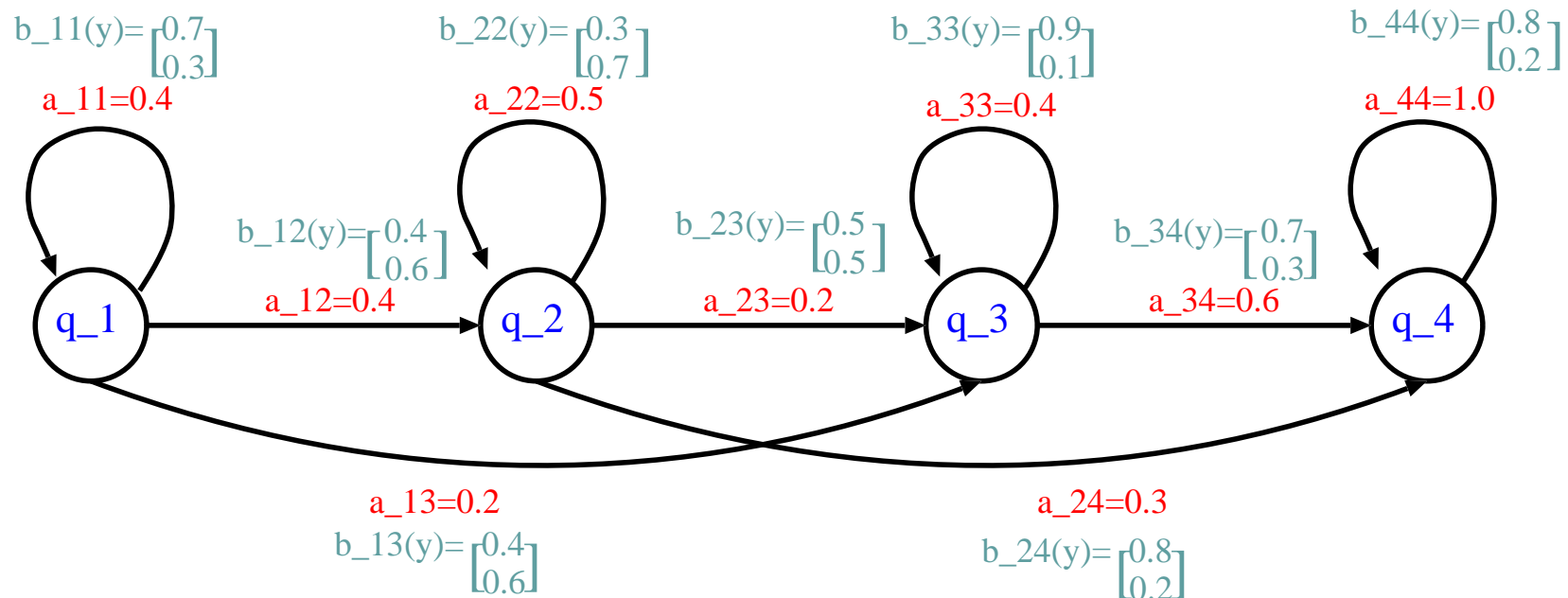
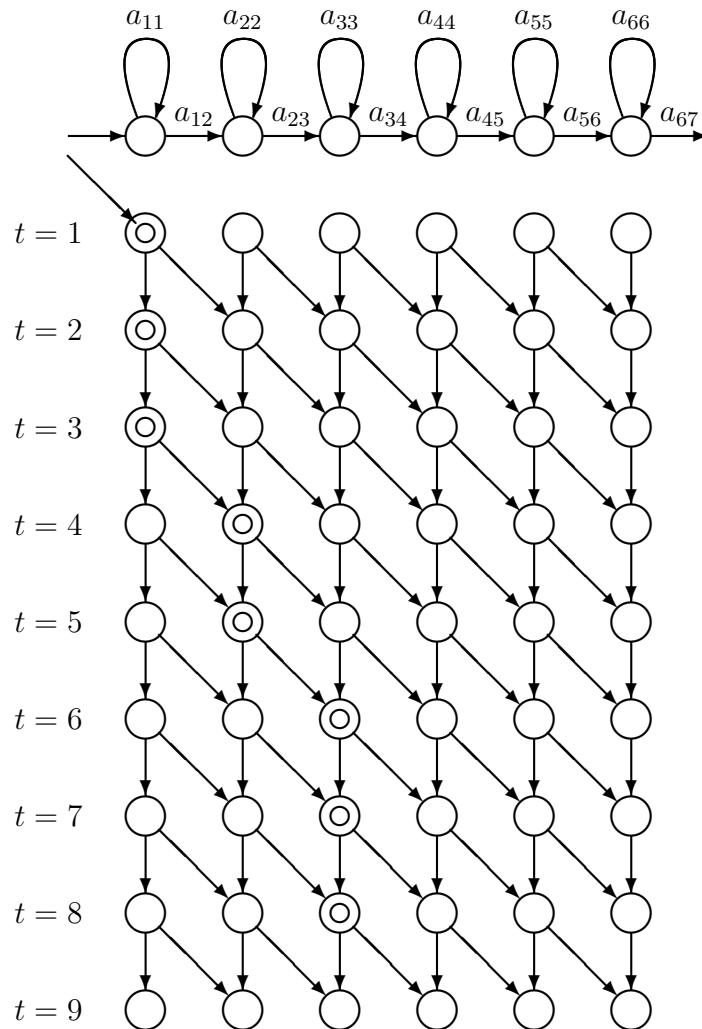


図5. 遷移出力型のHMM

## ■ 両者は等価変換できる (演習に適)



# Trellis (トレリス) 表現 – 空間時間経路





# Trellis (トレリス) 表現 – 空間時間経路

空間時間経路を **trellis** によって表現する。以下の図の状態系列は  $Q = \{1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3\}$  である。

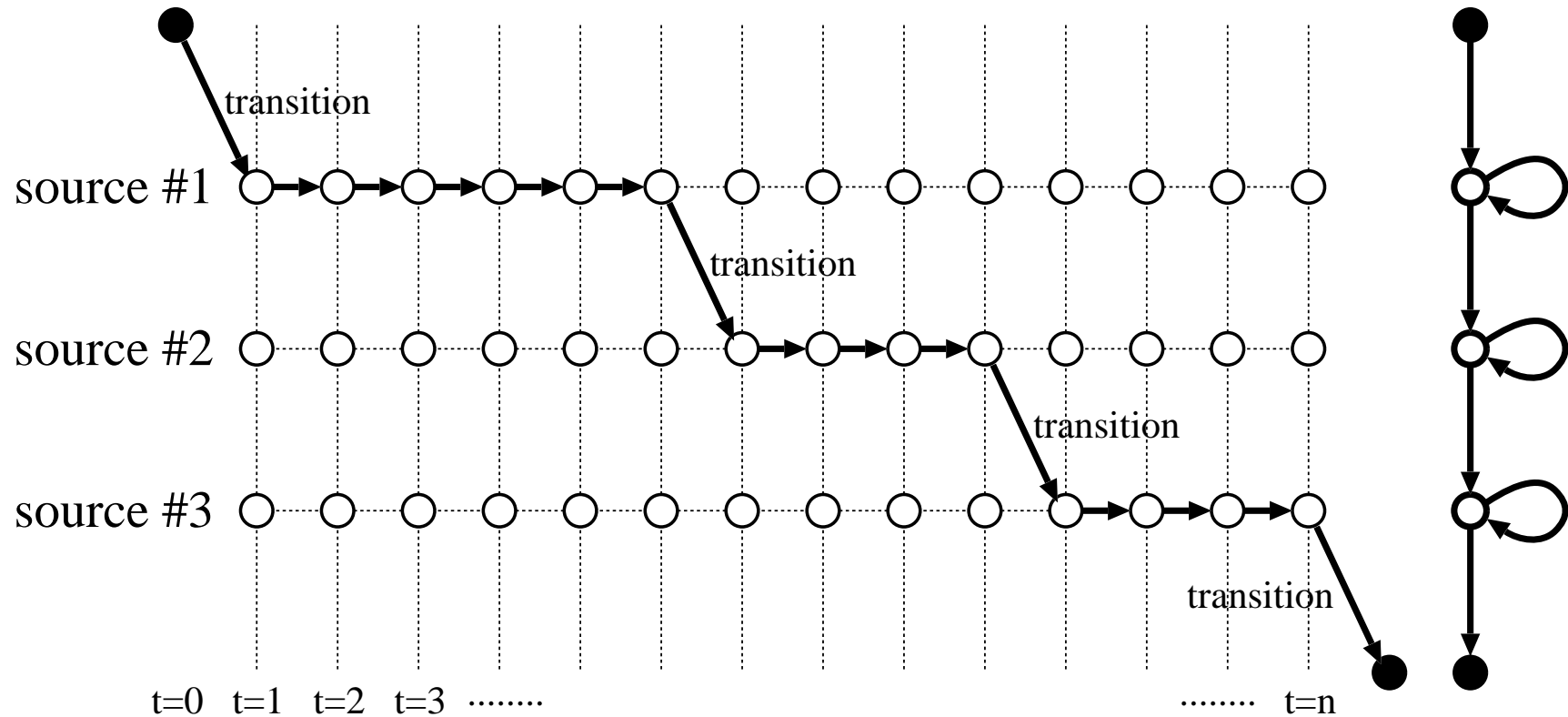


図6. 信号源切り替えとその時刻がなす経路(左)  
およびこれを表現する状態遷移図(右端)



# マルコフ遷移

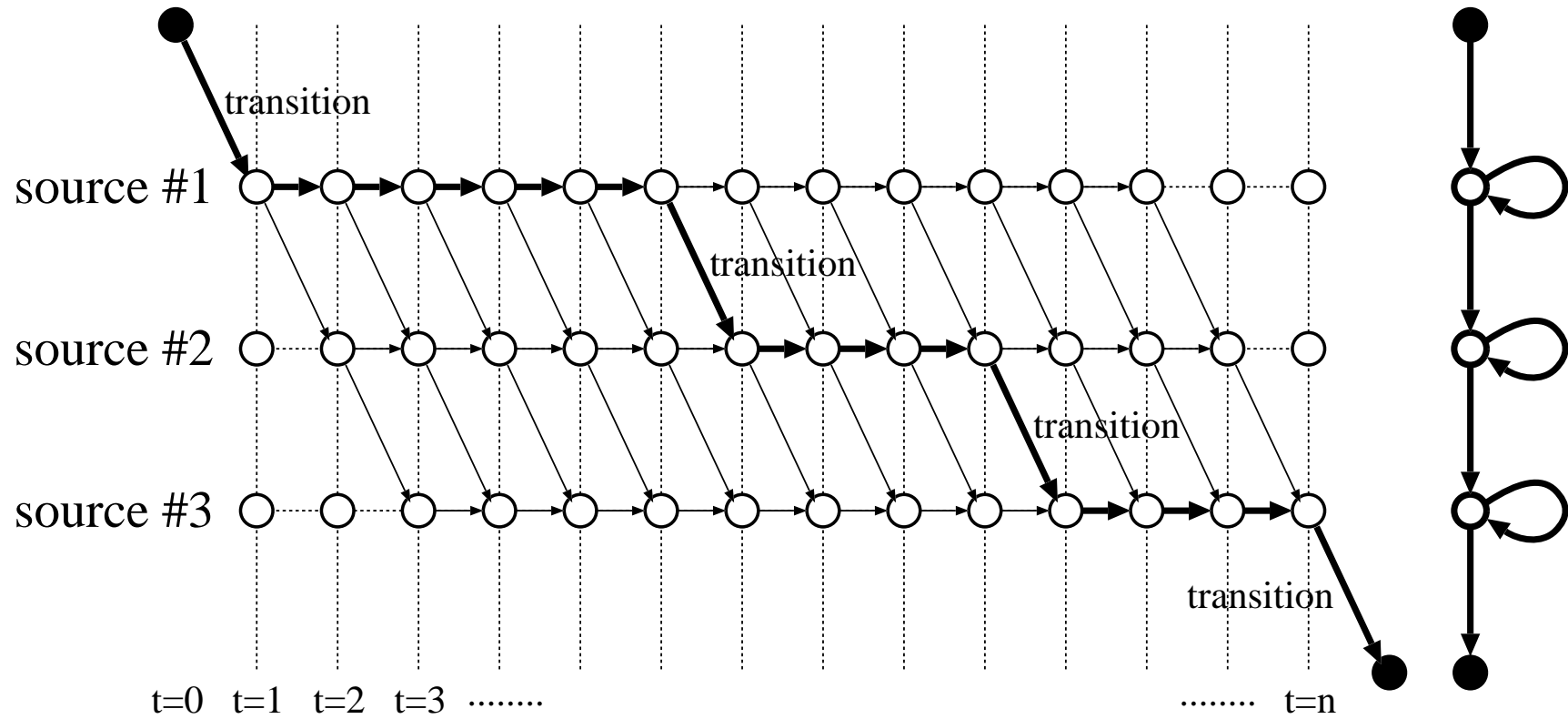


図7. 信号源の切り替えが確率的であるモデル。一つの経路が確率的に選ばれる



# マルコフ遷移

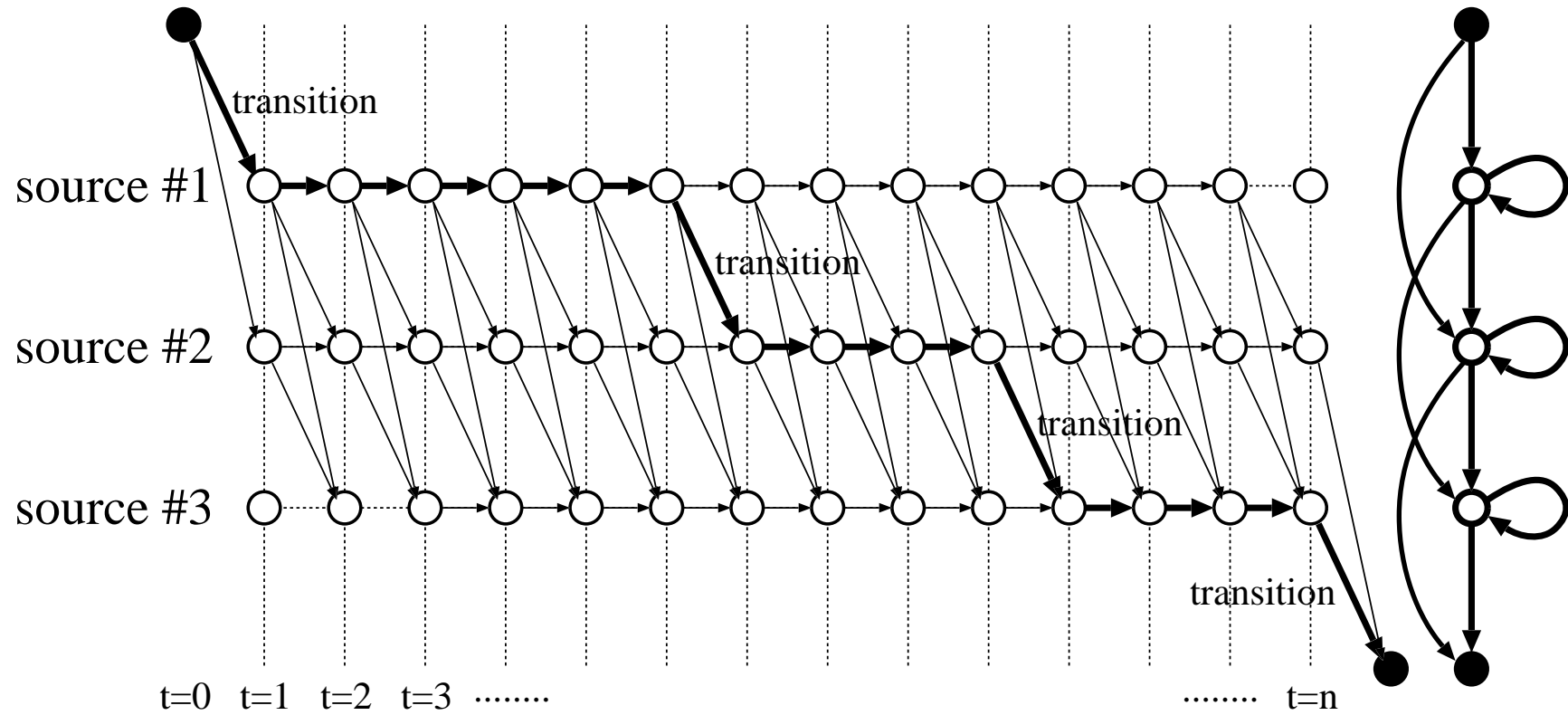


図 8. 遷移経路が複数あるモデル



# 時系列生成源としてのHMM

## ■ HMM は、非定常信号を生成するモデル

- 生成された系列が必ずしも実用的意味を持つわけではない

## ■ 確率評価に用いる

- モデル  $\Lambda$  から具体的な時系列  $Y = \{y_1, y_2, \dots, y_n\}$  を生み出すことができ、その確率  $P(Y|\Lambda)$  が計算できる。

- 時刻  $t$  に活動している信号源の番号を  $q_t$  とすると、このモデル  $\Lambda$  から時系列  $Y$  全体が生成される確率(あるいは密度)は、

$$P(Y|\Lambda) = \pi_{q_0} \prod_{t=1}^n a_{q_t q_{t+1}} b_{q_t}(y_t)$$

- 実際には  $q_t$  は分からない

    { Viterbi アルゴリズム (確率最大を求める)

    { Forward アルゴリズム (確率総和を求める)

いずれも、DPの原理に基づく効率的な計算アルゴリズム



# HMMの出力確率分布型: 2種類

---

- 離散出力分布型 (ヒストグラム)
- 連続出力分布型 (パラメトリック)





# 離散出力分布型 HMM

## ■ 離散出力分布型

音声パラメータベクトルの時系列をベクトル量子化 (VQ) の方法により離散的な符号列にして符号の出現をヒストグラムの形で表現する。複雑な分布も容易に表現できる。VQ は量子化誤差を含むため、表現精度には限界がある。VQ の精度 (符号帳サイズ) を上げると分布の精度が低下する。

## ■ VQ ベース

## ■ セパレート VQ

- SPHINX の例： VQ 符号帳を LPC ケプストラム、デルタケプストラム、パワーの 3 種類を同時に用いる

## ■ Fuzzy-VQ (Fuzzy VQ: 音素認識率 86.5% から 95.7% まで向上。)

## ■ 半連続 (semi-continuous)

## ■ スカラー量子化 HMM (1995)      音声符号化の発想

## ■ 離散混合 HMM (1996)



# 離散HMM (discrete HMM)

## ベイキス(Bakis)型HMMの例

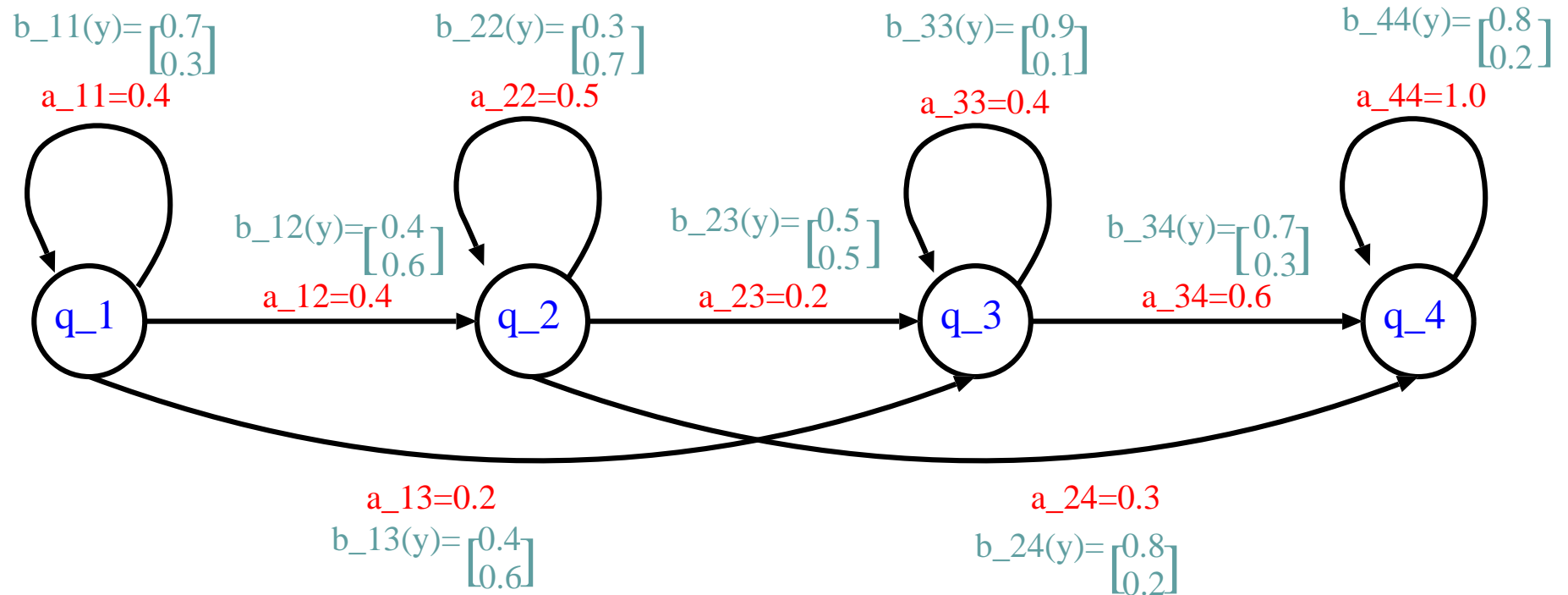


図9. ベイキス型HMM

$\left\{ \begin{array}{l} q_i : \text{状態} \\ a_{ij} : \text{状態遷移行列} \\ b_{ij}(y_t) : \text{出現確率} \end{array} \right.$



# 連続出力分布型 HMM

## ■ 連続出力分布型

出力分布密度のモデルが連続関数になっているもので、観測される音声パラメータベクトルを直接モデル化できる。複雑な分布を表現するには混合正規分布のような関数を用いる。共分散まで考えるとパラメータが多い。

### ■ 単一正規分布型

### ■ 混合正規分布型 (mixture)

### ■ tied-mixture

### ■ 4 階層 tied structure (モデル, 状態, 分布, 変数)



# 連続HMM (continuous HMM)

## 連続HMM

出力分布密度のモデルが連続関数になっているもの。

### ■ 単一正規分布型

$$b_i(O) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{(O-\mu_i)^t \Sigma_i^{-1} (O-\mu_i)}{2}}$$

共分散行列  $\Sigma$

■ 対角共分散行列：対角成分だけ持つ形（各変数間に相関がない）

■ 全共分散行列：全共分散成分を持つ形

### ■ 混合正規分布型 (mixture)

### ■ tied-mixture

### ■ 隠れマルコフ網 (HMnet)

### ■ 4層 tied structure



# 隠れマルコフモデル(HMM)の出力分布形の意味

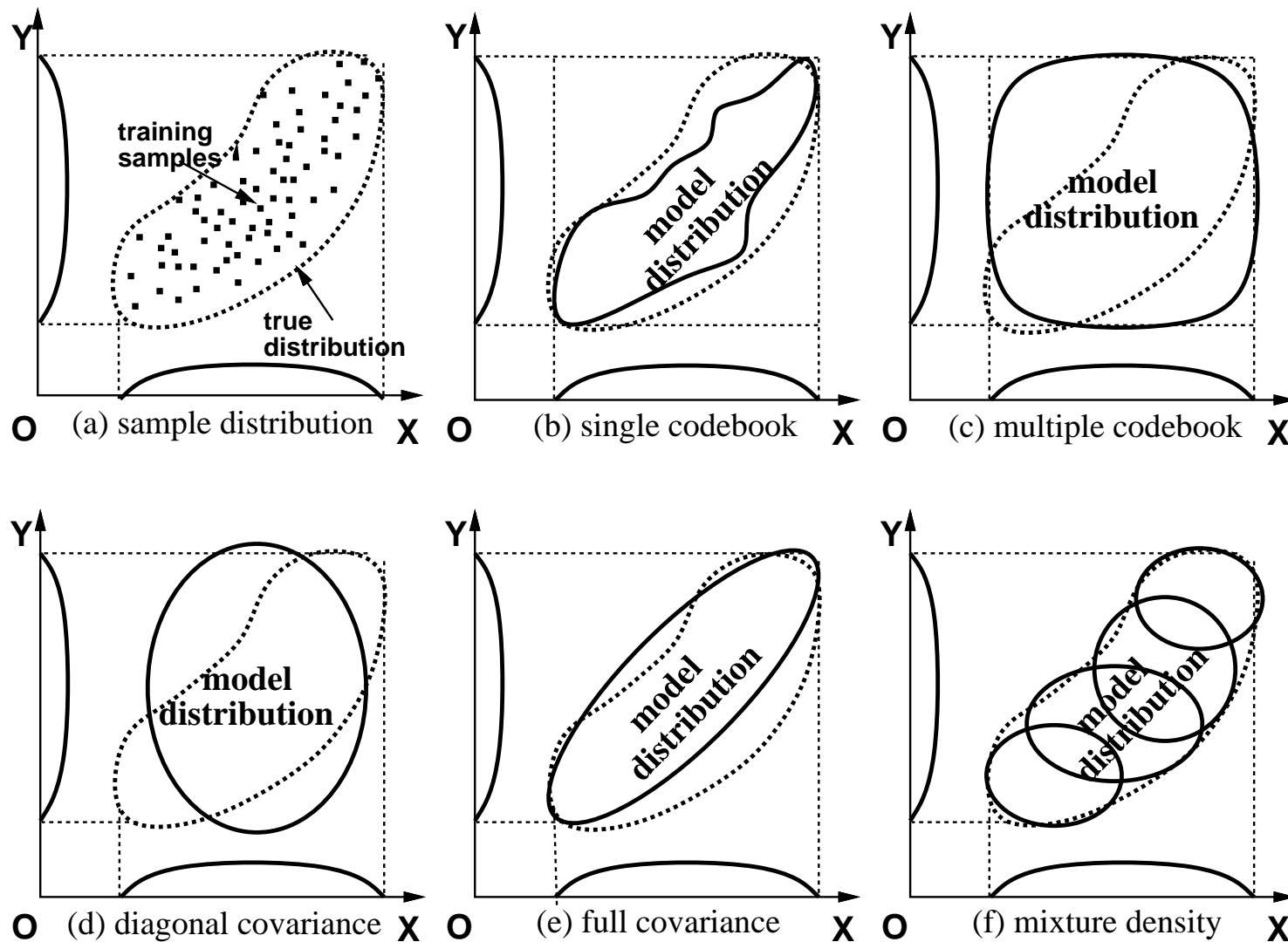


図10. 隠れマルコフモデル(HMM)の出力分布形の特徴



# 陥りやすい誤解

## ■ HMMに学習は不可欠か？

- HMMは「確率モデル」であるが、「統計モデル」として使用しなければならないわけではない。

## ■ HMMからは何も分からないか？

- 知識処理やルールベースのシステムの関係者が抱く不満
- 論理や言語のような人間が議論できる明示的な形式をとらなければ知識ではないのか？

## ■ HMMは情報処理としては特殊な方法論であるか？

- 確率の概念に馴染めない人は、Shannonの情報理論に戻って、データから情報を最大限引き出すにはどうすればよいか、という観点からHMMを見直せばよい

## ■ HMMは音声認識のための専用モデルであるか？

- 非定常時系列のモデルである
- 効率的な確率計算アルゴリズムが存在する
- 確率パラメータの学習ができる
- 単純なモデルから複雑なモデルまで同じ枠組で表現できる

## ■ HMMはマルコフモデルの変種であるか？

- 定常信号源の切り替えによって非定常な現象を表現しようとしている



# 音声認識におけるHMMの利点

---

- 確率理論に支えられていて理論に乗りやすいこと、
- 尤度(あるいは他の規準)を最大にするようなパラメータ推定を行なう学習法が存在すること、
- 標準パターンではなく分布を表現していること、
- 状態の概念により時間を適度に粗く表現していること、
- フレームごとの入力音声の尤度が得られるのでその積により文全体の尤度が得られること



# HMM 応用のポイント

## ■ 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





# HMMの応用対象分野

## ■ 音声認識

- 音素認識、単語認識
- 連続音声認識
- 話者適応
- 雑音適応

## ■ オンライン文字認識

## ■ 音楽情報処理 (リズム認識、自動和声付け)

## ■ 自然言語処理

- ネットワーク文法の自動獲得、音声認識のための言語モデル
- 話題遷移モデル
- 話者の番の切り替えモデル
- 形態素解析

## ■ 信号処理 — 話者分離など

## ■ 動画像の認識 (e.g. ジェスチャ認識、手話認識、 etc.)

## ■ 感情のモデル

## ■ 遺伝子構造推定

## ■ 経済予測