

# チュートリアル： ノンパラメトリックベイズ

産業技術総合研究所

吉井 和佳

[k.yoshii@aist.go.jp](mailto:k.yoshii@aist.go.jp)

# はじめに

- ノンパラメトリックベイズとは何か？
  - **ベイズ推定の新しい枠組み**
    - 基礎理論 (部品) は1973年にすでに提案されていた
      - T. Ferguson: A Bayesian analysis of some nonparametric problems, Annals of Statistics, Vol. 1, No.2, pp. 209-230, 1973.
    - 2000年前後からNIPSを中心に流行
      - M. I. Jordanらが火付け役
      - 計算機の発展で計算量の大きいベイズ推定が可能に
- **キーワード** (本チュートリアルのAgenda)
  - ベイズ推定
  - ディリクレ過程
    - 棒折り過程 (Stick-Breaking Process: SBP)
    - 中華料理店過程 (Chinese Restaurant Process: CRP)
  - 無限モデル

# 観測データと確率モデル

## • 観測データ

- 我々が実際に観測できる量
- 確率変数：何らかの確率分布に従って生成される
  - 例：サイコロをN回振って出た目の系列：離散分布に従う

$$X = \{x_1, x_2, \dots, x_N\}$$

## • 確率モデル

- 観測データがどのような過程を経て生成されたかを確率的に記述するもの
  - 確率モデルから生成されるデータは無数に考えられ観測データとはその1つの実現値
  - ある実現値について確率(密度)を与える

$$p(X | \Theta)$$

———モデルパラメータ

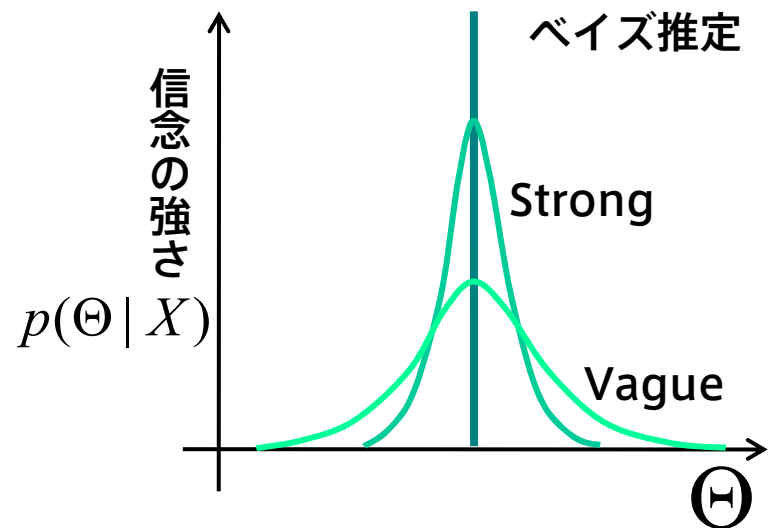
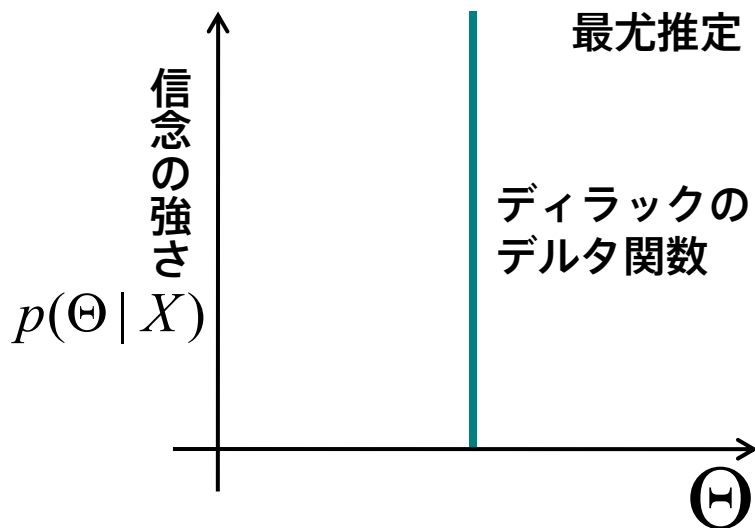
例: 離散分布のパラメータ

(各面の出る確率)  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$

# 確率モデルの学習

- 観測データの生成確率が最大になるような確率モデルのパラメータを推定すること
  - 最尤推定**
    - 最適なパラメータを点推定する (一意に決める)
  - ベイズ推定**
    - パラメータの事後分布を推定する (信念の強さを反映する)

データが無限にあれば両者は一致



# 最尤推定

- 確率モデルが観測データにフィットしすぎると汎化能力(未知データの予測能力)が悪化

例：サイコロの各面が出る確率を最尤推定する

$$X = \{x_1, x_2, \dots, x_N\}$$

$$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$$

$$p(X | \Theta) = \prod_{k=1}^K \theta_k^{n_k} \xrightarrow{\text{最大化}} \theta_k = \frac{n_k}{N}$$

N=10の観測データ

目 $k$	回数 $n_k$
1	4
2	1
3	0
4	0
5	2
6	3



最尤推定値 $\theta_k$
0.4
0.1
0.0
0.0
0.2
0.3

各面の出る確率が偏りすぎではないか？

$x_{N+1}$  はどうなるのだろうか？

3,4が出る確率はゼロ！

# 最尤推定 (詳細)

- 観測データだけから最適なパラメータを点推定

例：サイコロの各面が出る確率を最尤推定する

$$X = \{x_1, x_2, \dots, x_N\}$$

最大化したい目的関数

拘束条件

$$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\} \quad p(X | \Theta) = \prod_{n=1}^N \theta_{x_k} = \prod_{k=1}^6 \theta_k^{n_k} \quad \sum_{k=1}^6 \theta_k = 1$$

1. 対数をとって凹関数化：これを最大化

$$\log p(X | \Theta) = \sum_{k=1}^6 n_k \log \theta_k$$

2. 拘束条件付きの最適化：ラグランジュの未定乗数法

$$F = \log p(X | \Theta) + \lambda \left( 1 - \sum_{k=1}^6 \theta_k \right)$$

4. 拘束条件に代入することで未定乗数を計算

3. 偏微分して0とおく

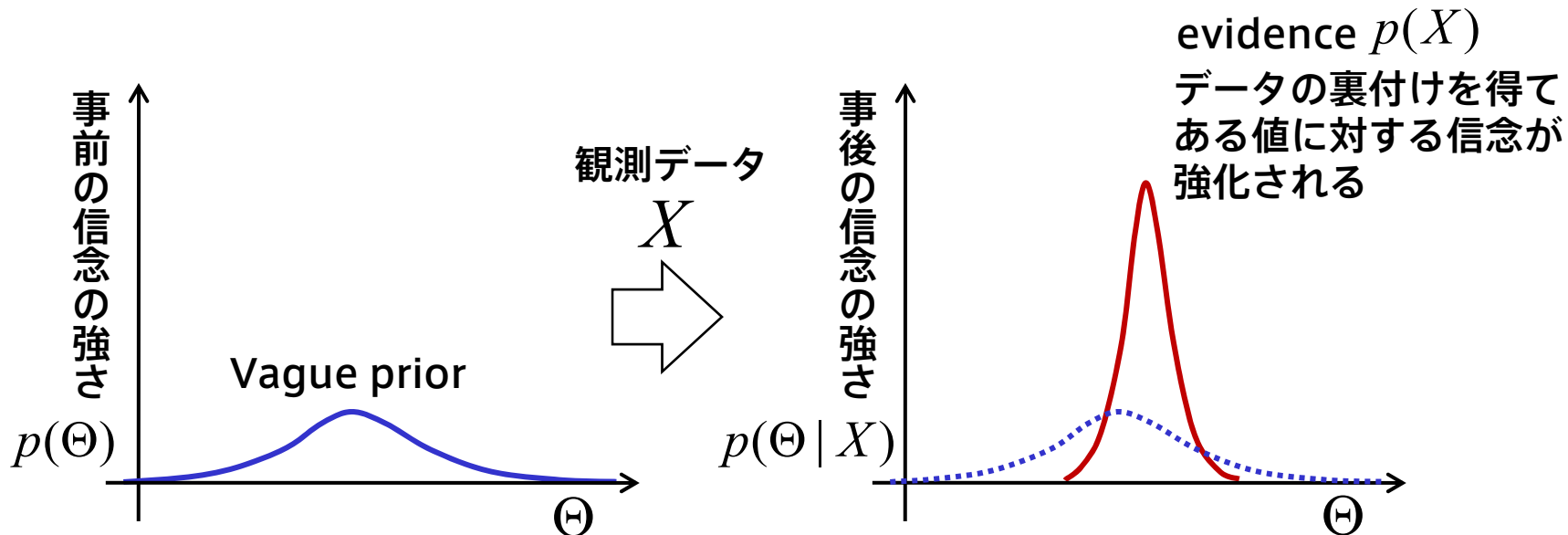
$$\frac{\partial F}{\partial \theta_k} = \frac{n_k}{\theta_k} - \lambda \equiv 0 \quad \Rightarrow \quad \theta_k = \frac{n_k}{\lambda}$$

$$\lambda = N$$

$$\theta_k = \frac{n_k}{N}$$

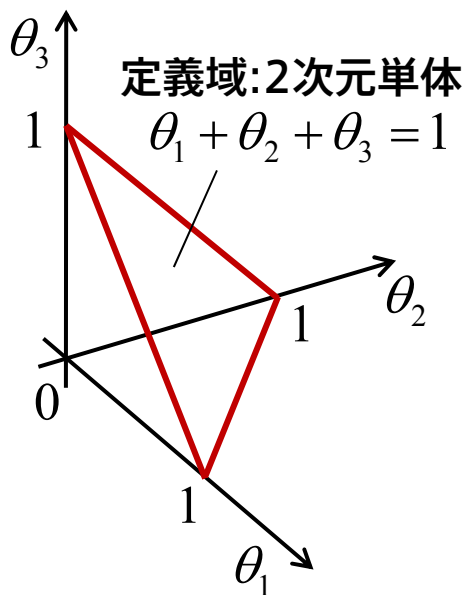
# 信念の強さの確率表現

- 未知パラメータの値のあらゆる可能性を考えておけば極端な推定結果は回避可能
  - 事前分布：「このあたりがそれっぽい」という予断
    - 事前分布のパラメータ(ハイパーパラメータ)を変えることで**事前の信念の強さ**を反映することができる
  - 事後分布：観測データを見たあとでの判断
    - 観測データが増えるたびに**事後の信念の強さ**が変化



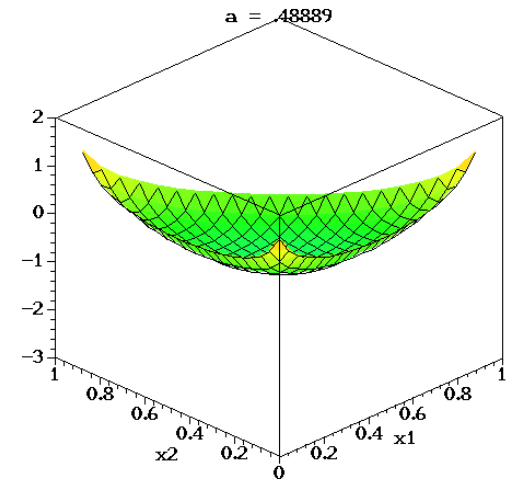
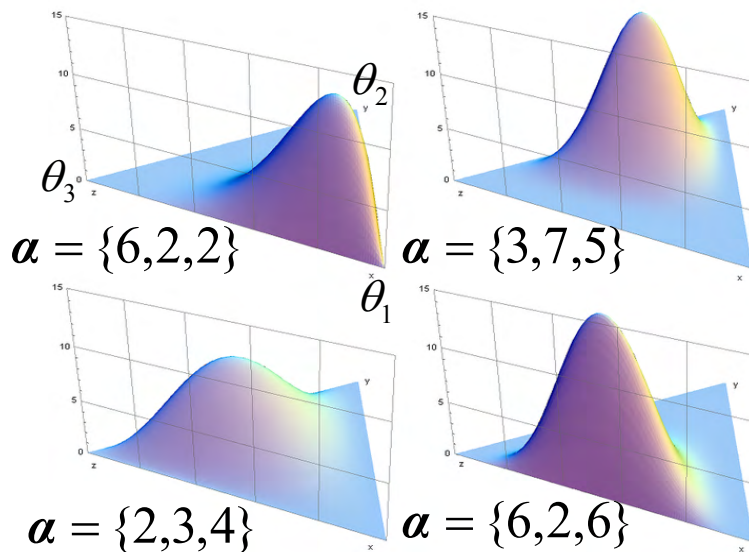
# 事前分布

- 事前分布とは確率分布上の確率分布
  - Distribution over distributions
    - 確率分布を記述するパラメータが従う確率分布とも言える
  - 共役事前分布の利用が便利
    - 事前分布  $p(\Theta)$  と事後分布  $p(\Theta | X)$  が同じ形になる
      - 離散分布上の確率分布：ディリクレ分布



$\Theta = \{\theta_1, \theta_2, \theta_3\}$  の分布  $\text{Dir}(\Theta | \alpha)$

$0 < \alpha < 2$  における変化





# ベイズ推定

- 適切な事前分布を与えることで過学習が抑制
  - ディリクレ事前分布 = 下駄をはかせる

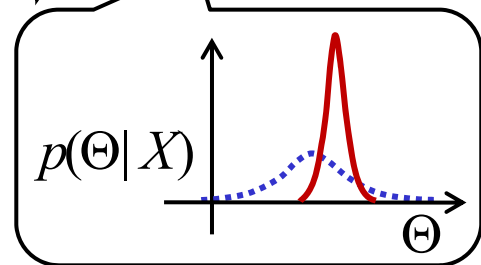
本当にこれだけ!

ディリクレ分布のパラメータ

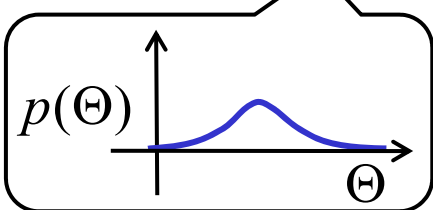
$k$	事前の観測回数 $\alpha_k$	実際の観測回数 $n_k$	事後の観測回数 $\alpha_k + n_k$
1	3	4	7
2	3	1	4
3	3	0	3
4	3	0	3
5	3	2	5
6	3	3	6

ベイズ推定による不確実性の保持

$\text{Dir}(\Theta | \alpha + n)$



$\text{Dir}(\Theta | \alpha)$



点推定

最尤推定値

MAP推定値

$$\Theta = \{0.4, 0.1, 0.0, 0.0, 0.2, 0.3\}$$

$$\hat{\Theta} = \{0.25, 0.14, 0.11, 0.11, 0.18, 0.21\}$$

# ベイズ推定 (詳細)

- 共役事前分布を用いて事後分布を計算

例：サイコロの各面が出る確率をベイズ推定する

$$X = \{x_1, x_2, \dots, x_N\} \quad \Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$$

$$p(\Theta) = \text{Dir}(\Theta | \alpha) = \frac{\Gamma(\sum_{k=1}^6 \alpha_k)}{\prod_{k=1}^6 \Gamma(\alpha_k)} \prod_{k=1}^6 \theta_k^{\alpha_k - 1} = C(\alpha) \prod_{k=1}^6 \theta_k^{\alpha_k - 1}$$

$a_k$  : 仮想的な観測回数に相当 (事前の信念の強さ)

ある程度真っ当なサイコロだと信じるなら例えば  $a_k = 3$

信念が変化

$$p(X | \Theta) = \prod_{k=1}^6 \theta_k^{n_k}$$

$$p(\Theta | X) = \frac{p(X | \Theta) p(\Theta)}{p(X)} \propto p(X | \Theta) p(\Theta) = C(\alpha) \prod_{k=1}^6 \theta_k^{\alpha_k + n_k - 1}$$

$$p(\Theta | X) = C(\alpha + n) \prod_{k=1}^6 \theta_k^{\alpha_k + n_k - 1} = \text{Dir}(\Theta | \alpha + n)$$

正規化係数は分布が正しく正規化されるようにあとで計算すればよい

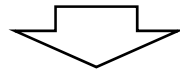
# ベイズモデル再考

- これまで確率モデルの複雑さは既知と仮定
  - 複雑さは「観測データに合わせて」手動で指定
    - 例：サイコロの面数 (複雑さ) は6と指定  
= 6次元ディリクレ分布を事前分布として利用

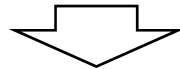
疑問：サイコロの面数が不明な場合はどうすべきか？

観測データ  $X = \{1, 2, 2, 1, 2, 1, 3, 2\}$

観測データに対しては3面サイコロを考えれば良さそうに思えるが  
このサイコロを振ると次に「4」や「5」が出るかもしれない



4次元ディリクレ分布を仮定すべき？ → 将来「5」が出たら困る  
5次元ディリクレ分布を仮定すべき？ → 将来「6」が出たら困る  
...



無限次元ディリクレ分布 = ディリクレ過程 (DP)

# ノンパラメトリックベイズモデル

- 無限の複雑さを持つベイズモデル

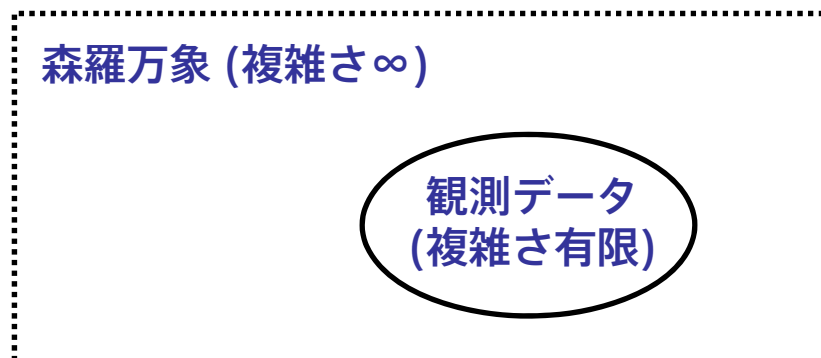
- 「ノンパラメトリック」とは

- 「無限個のパラメータをもつ」という意味
- 「パラメータがない」という意味ではない

- 観測データに限らない森羅万象を考慮

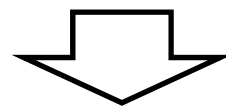
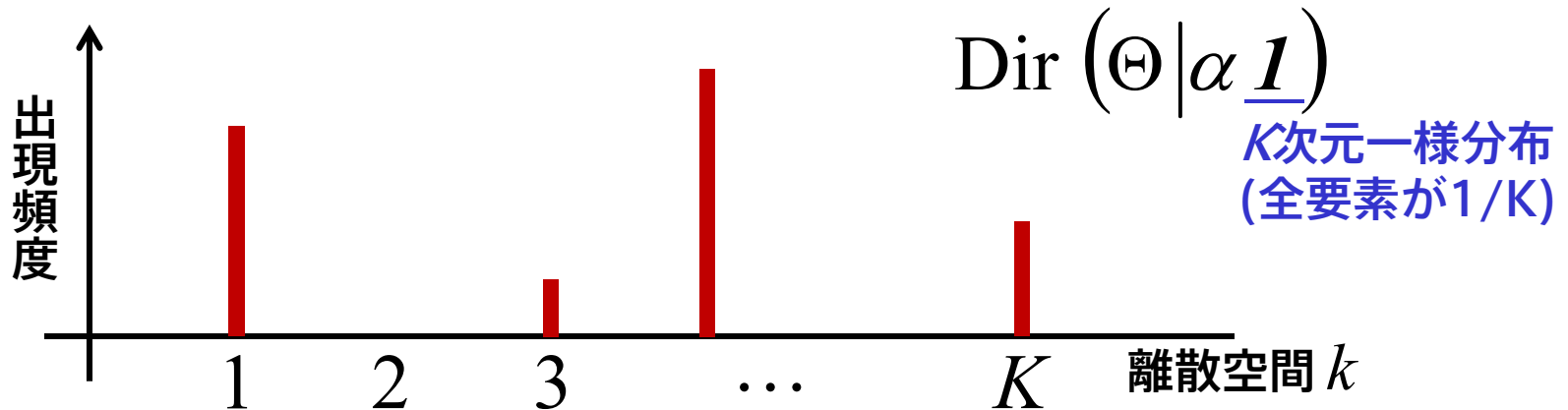
- 本質的に汎化性能に優れている
- 無限集合である「森羅万象」は無数のバリエーションに富む
  - 無限個のデータがあれば無限個のパラメータが必要
- 有限集合である「観測データ」はそのごく一部
  - 有限個のデータであれば有限個のパラメータで十分

→ 計算機で実現可能！

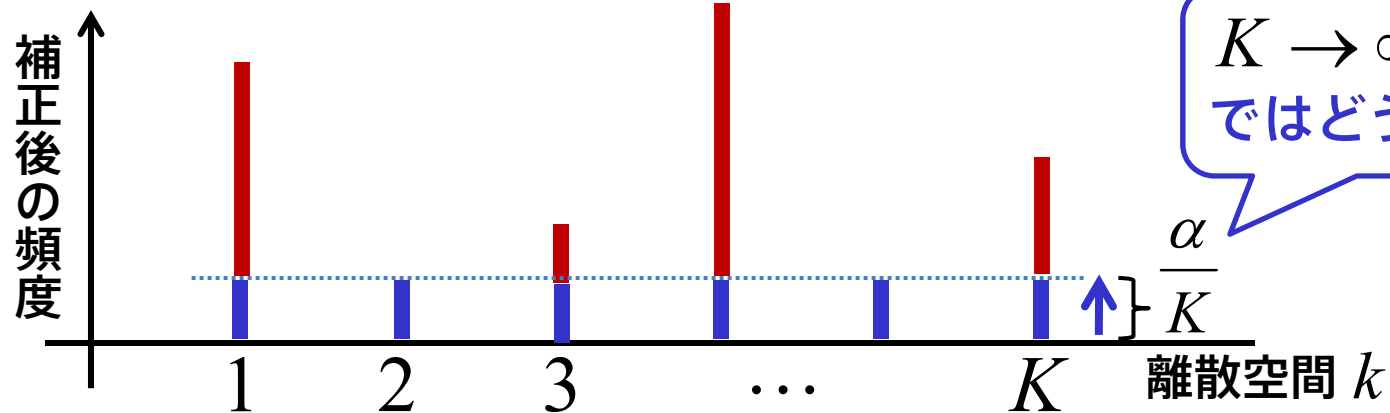


# K次元ディリクレ分布

- ディリクレ事前分布 = かさ上げスムージング
  - ディリクレ分布のパラメータ：はかせる下駄の高さ



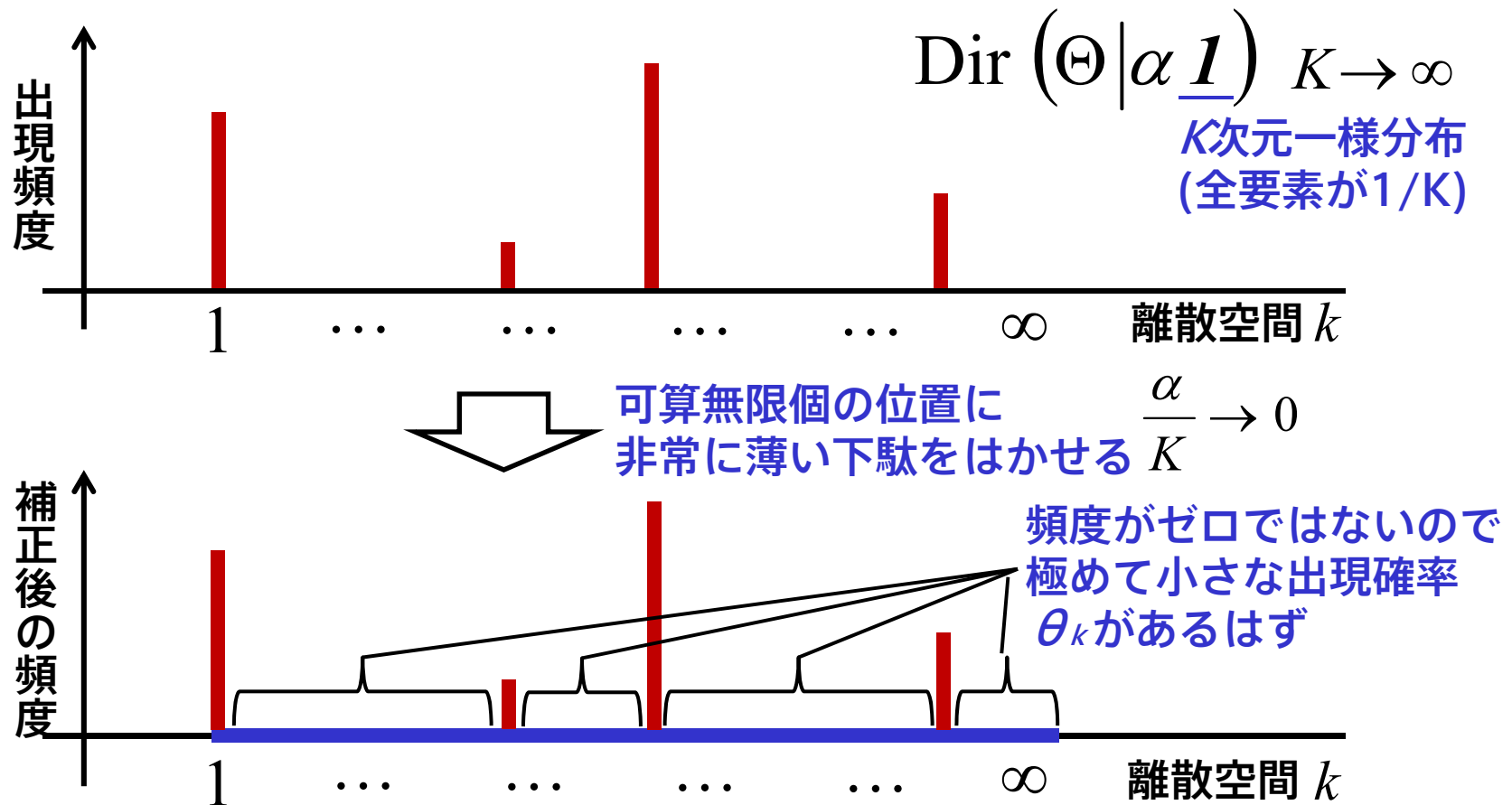
各頻度到下駄をはかせて  
ゼロ頻度問題を解消



$K \rightarrow \infty$   
ではどうなる？

# 無限次元ディリクレ分布

- ディリクレ分布の次元  $K$  を無限に発散させる
  - 下駄の高さが非常に薄くなる
  - 無限個ある面の出現確率は厳密にはゼロではない

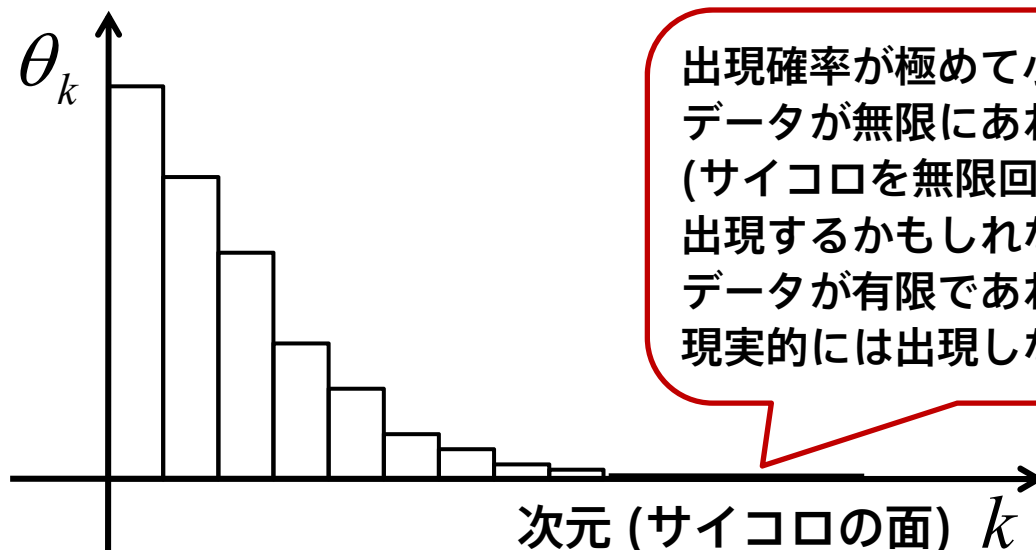


# ディリクレ過程 (DP)

- 無限次元ディリクレ分布と等価

- 無限次元離散分布に対する確率分布  $\Theta \sim \text{DP}(\alpha, I)$

- 無限個のパラメータの和は1  $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_\infty\}$
- ほとんどのパラメータはほとんどゼロ (ゼロではない)



- 無限次元離散分布を直接モデル化 (棒折り過程)
- 無限次元離散分布から得られるサンプルをモデル化 (中華料理店過程)

# 棒折り過程 (SBP)

- 無限次元の離散分布を直接表現
  - 長さ1の棒を無限回折りとっていく
  - どこで折りとるかは確率的に決まる

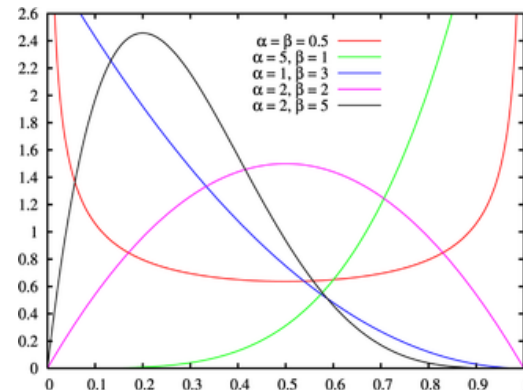
$$\tilde{\theta}_k \sim \text{Beta}(1, \alpha)$$

平均的には  
1: $\alpha$ で折る

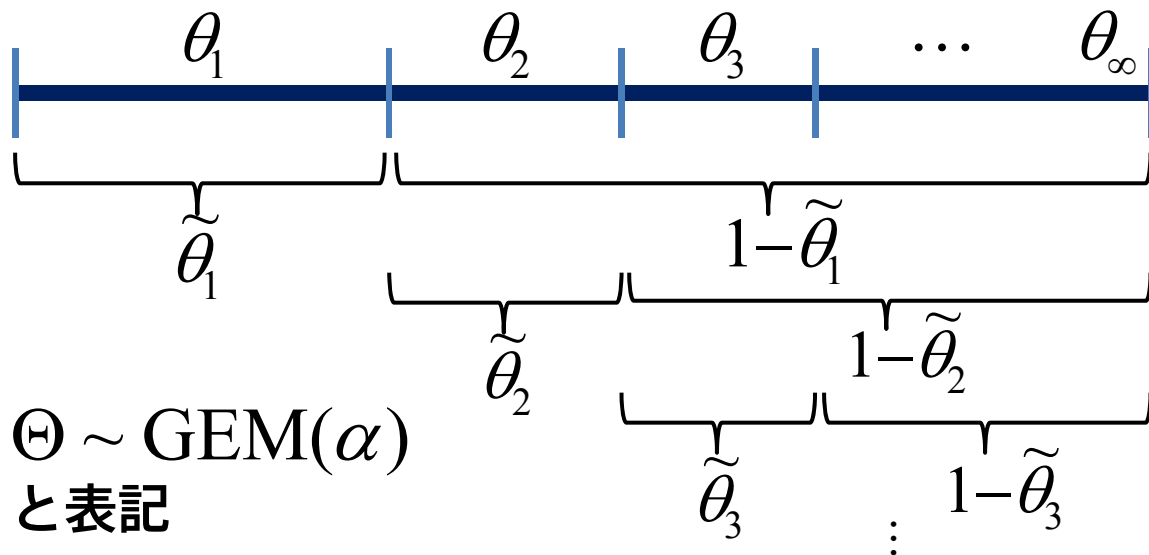
$$\theta_k = \tilde{\theta}_k \prod_{l=1}^{k-1} (1 - \tilde{\theta}_l)$$

DPのパラメータ  
(超パラメータ)

Beta( $\alpha, \beta$ )



(0, 1)上の確率分布



DPの一般化

Pitman-Yor過程

Beta( $1 - d, \alpha + dk$ )

Beta two-parameter過程

Beta( $\alpha, \beta$ )

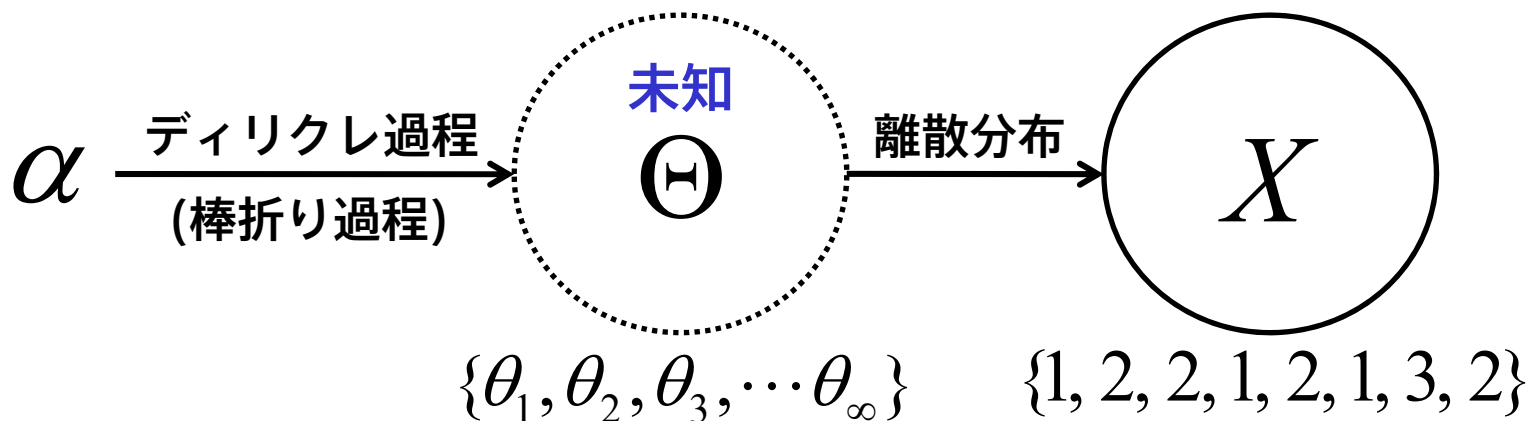
$\Theta \sim \text{GEM}(\alpha)$

と表記



# 「無限」の取り扱い

- 想定する状況：次に出る目の予測がしたい
  - 無限個のパラメータの値を求めずに済ませたい
    - Vapnikの原理
      - ある問題を解くとき、その問題よりも難しい問題を途中段階で解いてはならない



過去データから次のデータが  
いきなり予測が可能！

積分＝あらゆる可能性を考える

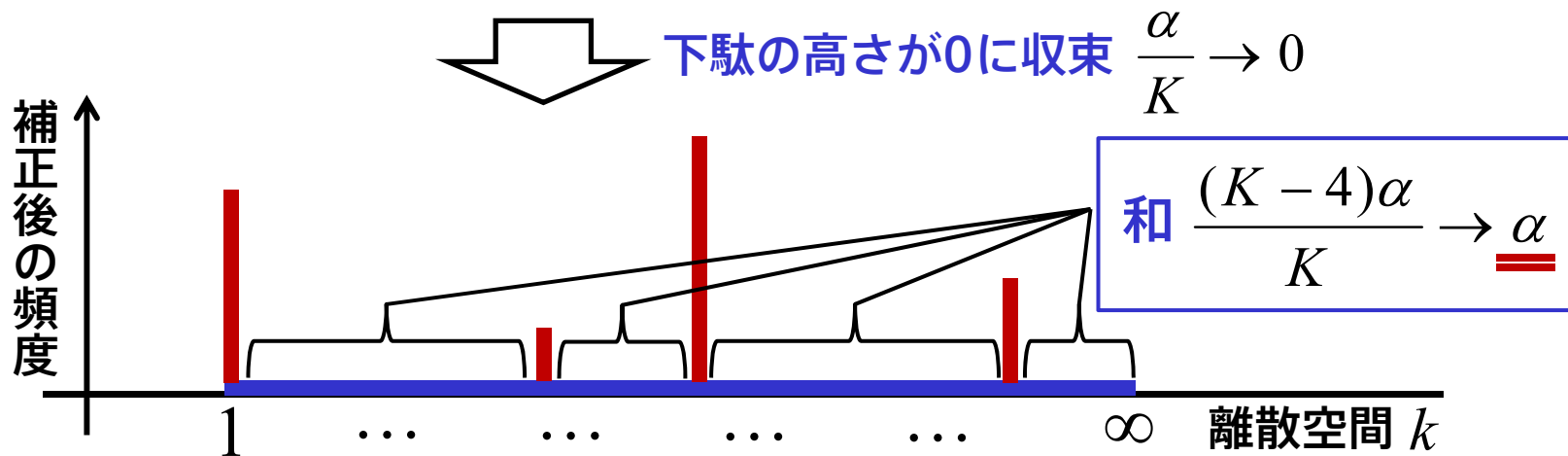
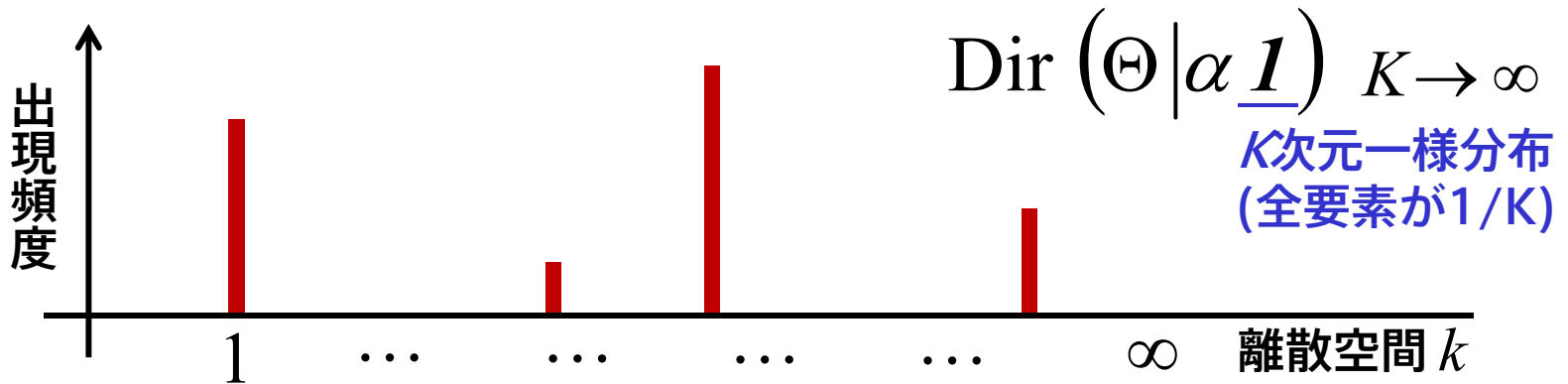
$$\underline{p(x_{\text{next}} | X)} = \int p(x_{\text{next}} | \Theta) \underline{p(\Theta | X)} d\Theta$$

解きたい問題

無限次元を扱う必要あり  
→ 難しい問題が増えた！

# 無限次元ディリクレ分布再考

- ディリクレ分布の次元  $K$  を無限に発散させる
  - 下駄の高さはゼロに収束
  - 未観測の部分の **下駄の高さ合計** はゼロではない！



# 中華料理店過程 (CRP)

- 無限次元離散分布からのサンプル (頻度)に着目

- 「The rich get richer」の法則

- 次の目の出方は過去に出た目の頻度に比例する

- N回の試行なら高々N種類しか出現しない

- 種類数の期待値は  $\log(N)$  に比例

新たなクラスのために  
小さな確率をリザーブ  
(これまでの客数が多く  
なるほど出にくくなる)

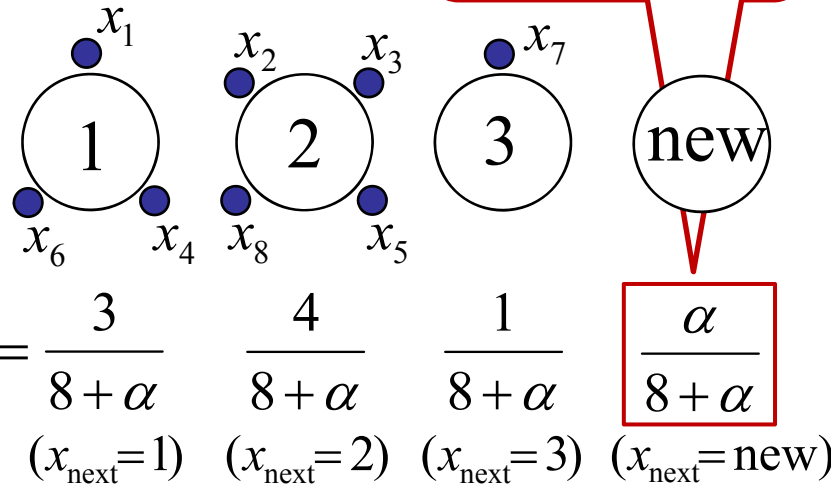
観測データ  $X = \{1, 2, 2, 1, 2, 1, 3, 2\}$

パラメータ  $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_\infty\}$

確率モデル

$$p(X | \Theta) = \prod_{k=1}^{\infty} \theta_k^{n_k}$$

$$p(\Theta) = \text{DP}(\alpha, I)$$



パラメータの事後分布  $p(\Theta | X)$  は無限次元だとしても

次サンプルの予測分布  $p(x_{\text{next}} | X) = \int p(x_{\text{next}} | \Theta) p(\Theta | X) d\Theta$  は計算可能!

# 潜在変数モデル

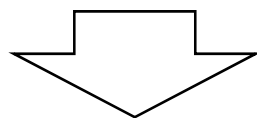
- 観測データだけでなく非観測データも考える
  - 機械学習における中心的な確率モデル (例: 混合モデル)

例：性別ラベルが分からない身長データ

観測変数 $X$ ：身長 (ガウス分布に従う)

潜在変数 $Z$ ：男 or 女 (2次元離散分布に従う)

パラメータ $\theta$ ：2つのガウス分布の平均と分散・混合比



クラス数が未知であれば？

例：クラスラベルが分からない特徴量データ

観測変数 $X$ ：特徴量 (ガウス分布に従う)

潜在変数 $Z$ ：クラス (無限次元離散分布に従う)

パラメータ $\theta$ ：無限個のガウス分布の平均と分散・混合比

潜在変数 $Z$ に対する事前分布としてDPが利用可能！  
→ 無限混合ベイズモデル

# ディリクレ過程の応用

- 無限混合ベイズモデルをどう定式化するか
  - ディリクレ過程の**基底測度**を明示的に表示

例：クラスラベルが分からない特徴量データ

観測変数 $X$ ：特徴ベクトル (ガウス分布に従う)

潜在変数 $Z$ ：クラス (無限次元離散分布に従う)

パラメータ $\Theta$ ：無限個のガウス分布の平均と分散・**混合比**

$$p(x | \Theta) = \sum_{k=1}^{\infty} \pi_k N(x | \mu, \sigma^2)$$

無限個の混合比は  
棒折り過程でOK

無限個のガウス分布は  
どうやって生成するの？

→ ガウス・ウィシャート分布  
(ガウス分布に対する共役事前分布)

$$p(\Theta) = \text{DP}(\alpha, G_0)$$

基底測度：合わせて表記しておく

# 中華料理店過程表現

- テーブルの料理を生成する機構 = 基底測度  $G_0$

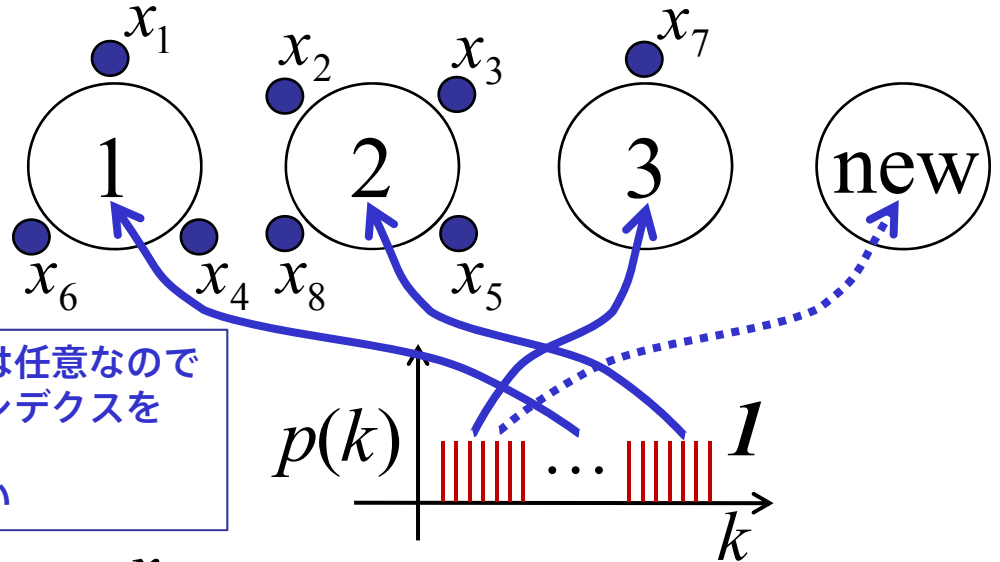
無限面サイコロの場合

$$DP(\alpha, I)$$

観測変数：サイコロの目

$$X = \{1, 2, 2, 1, 2, 1, 3, 2\}$$

正整数に対するラベル付けは任意なので  
出現した目の種類の順にインデクスを  
1, 2, 3, ... としてもよい  
= 新規の目は4としてもよい



無限混合モデルの場合

$$DP(\alpha, G_0)$$

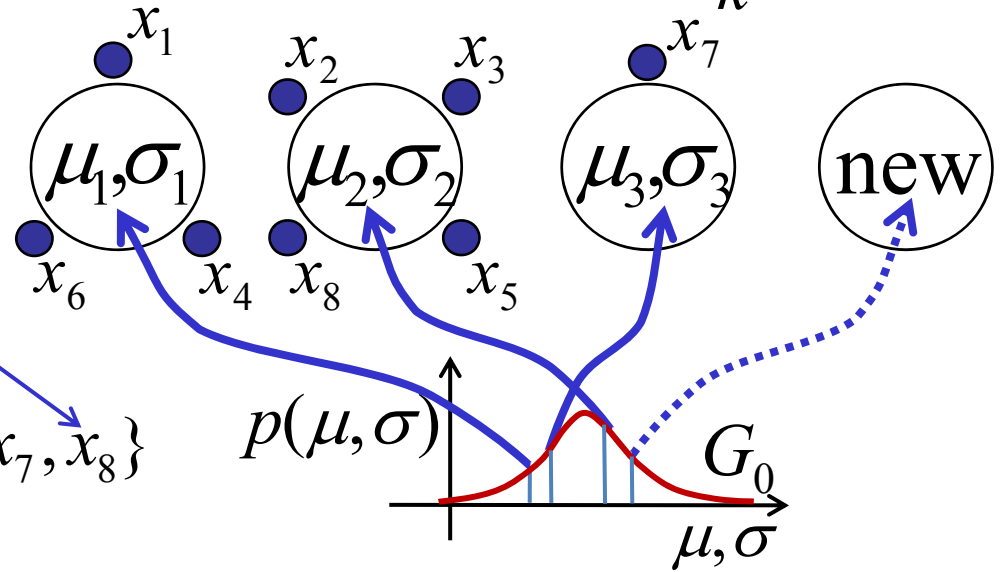
潜在変数：クラス

$$Z = \{1, 2, 2, 1, 2, 1, 3, 2\}$$

対応するガウス  
分布から生成

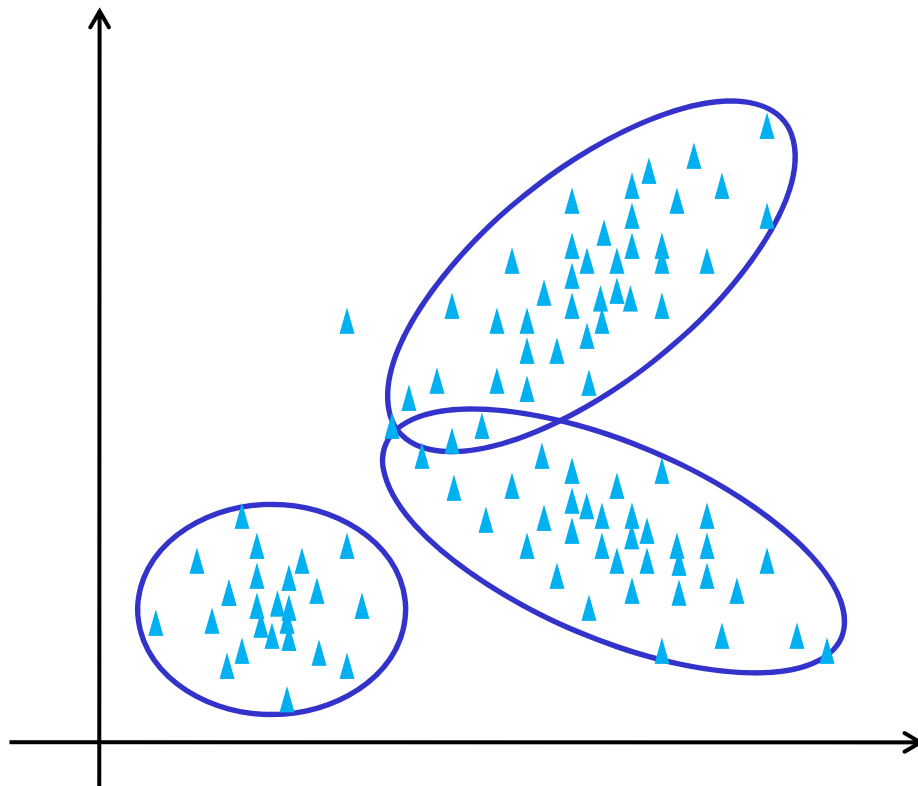
$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

観測変数：特徴量



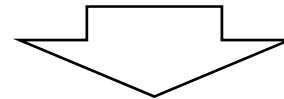
# Infinite GMM

- 可算無限個のクラスを許容する  
ベイズ混合ガウスモデル [Rasmussen2000]
  - ディリクレ過程 (DP) を利用



有限混合モデル：

1. クラス数  $K$  を変化させて  
大量のモデルを学習
2. AICやBICなどの指標を用いて  
適切なものをあとから選択



無限混合モデル：

- 一度の学習でクラス数の  
事後分布が求まる  
(必要ならば事後確率の高い  
クラス数を求めることも可能)

# Infinite HMM

- 可算無限個の状態(と出力シンボル)を許容する  
ベイズ隠れマルコフモデル
  - 階層ディリクレ過程 (HDP) を利用
  - ギブスサンプリング [Beal2001]
    - 各時刻の隠れ状態を反復的にサンプル
      - 十分時間をかければ真の事後分布に収束
      - HMMは隣同士の相関が強くて収束が非常に遅い
  - **ビームサンプリング** [VanGael2007]
    - 問題：状態数が無限なので動的計画法が実行できない
      - かといって打ち切り近似はしたくない・・・
    - 解決法：スライスサンプリングを組み合わせる
      - ある閾値以下の確率の遷移を無視する
      - 閾値も確率的にサンプルすることで真の事後分布に収束



# 確率的文脈自由文法 (PCFG)

- シンボルの導出規則に確率が付与されたもの
  - 予めチョムスキー標準系に変換しておく

$S \rightarrow NP VP$  1.0  
 $PP \rightarrow P NP$  1.0  
 $VP \rightarrow V NP$  0.7  
 $VP \rightarrow VP PP$  0.3  
 $P \rightarrow with$  1.0  
 $V \rightarrow saw$  1.0

$NP \rightarrow NP PP$  0.4  
 $NP \rightarrow astronomers$  0.1  
 $NP \rightarrow ears$  0.18  
 $NP \rightarrow saw$  0.04  
 $NP \rightarrow stars$  0.18  
 $NP \rightarrow telescopes$  0.1

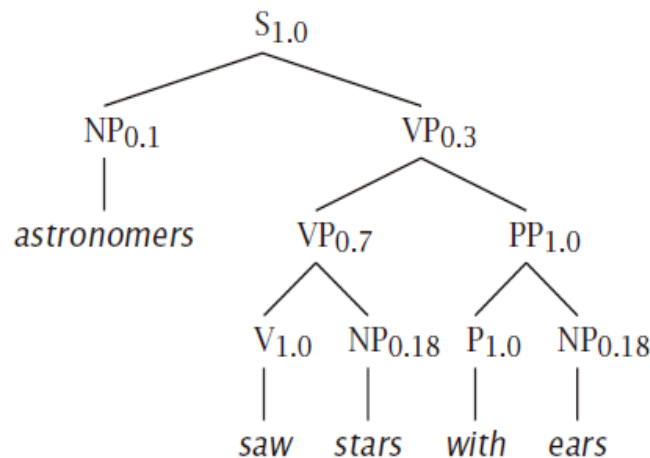
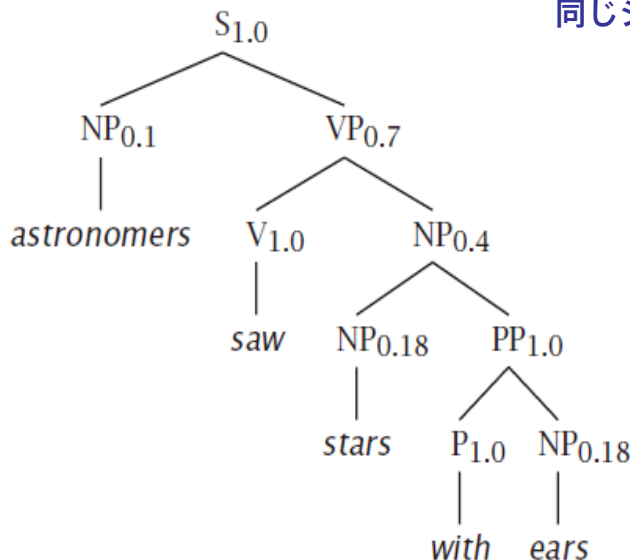
PCFGの最尤学習：  
単語列が与えられた時に  
各導出規則の確率を求める

PCFGのベイズ学習：  
各シンボルごとに導出確率の  
ディリクレ事後分布を求める

同じシンボルで開始する規則の確率の総和は1

[栗原2004]

$t_1$ :



$$\begin{aligned}
 P(t_1) &= 1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \\
 &\quad \times 0.18 \times 1.0 \times 1.0 \times 0.18 \\
 &= 0.0009072
 \end{aligned}$$

$$\begin{aligned}
 P(t_2) &= 1.0 \times 0.1 \times 0.3 \times 0.7 \times 1.0 \\
 &\quad \times 0.18 \times 1.0 \times 1.0 \times 0.18 \\
 &= 0.0006804
 \end{aligned}$$

$$P(w_{15}) = P(t_1) + P(t_2) = 0.0015876$$

# Infinite PCFG

- 可算無限個のシンボルおよび導出規則を許容する  
ベイズ文脈自由文法 [Liang2007]
  - 階層ディリクレ過程 (HDP) を利用
  - 手動で導出規則を与えなくてよい
    - 必要なシンボル・必要な導出規則が自動的に生成

## 有限モデル

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

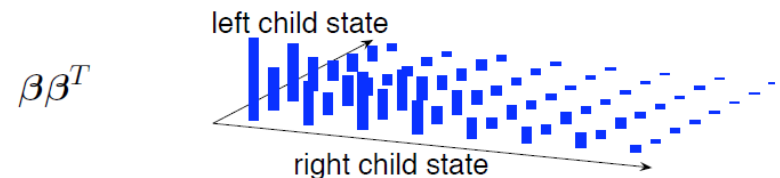
## 無限モデル

$S1 \rightarrow S1 S2$   
 $S1 \rightarrow S1 S5$   
 $S2 \rightarrow S2 S3$   
⋮

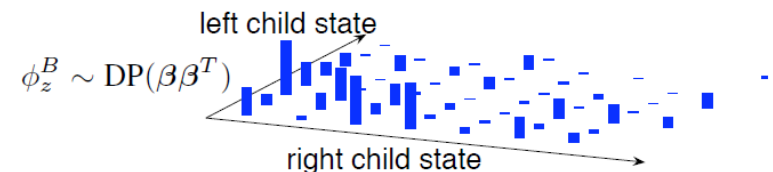
### 1. 可算無限個のシンボルを生成 (DP)



### 2. シンボルの組に対する基底測度を準備



### 3. 導出規則の右側の分布を生成 (DP)



# 最近のトレンド

- ベータ過程 (BP) を用いた研究が増加
  - 無限種類の因子 (factor/feature) を考える
    - 潜在変数：サンプルごとの各因子の有無 (無限次元)
    - 因子数がKであれば、2のK乗の状態が表現可能
  - 利用しやすいサンプリングスキームが存在
    - ディリクレ過程 (DP) → 中華料理店過程 (CRP)
    - ベータ過程 (BP) → インド料理過程 (IBP)

混合モデル (Latent Class Model)

観測データ中の各サンプルはどれか一つのクラスから生成

因子モデル (Latent Feature Model)

観測データ中の各サンプルは複数の因子から生成

例：主成分分析 (PCA) 因子分析 (FA) 独立成分分析 (ICA)

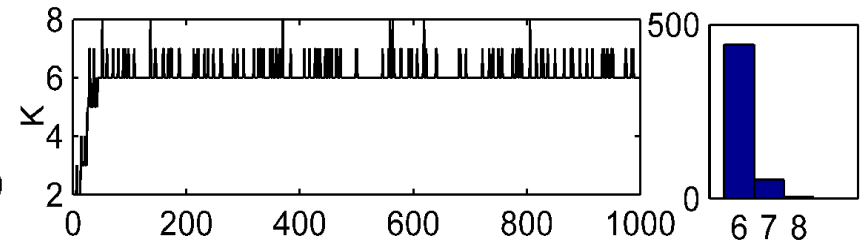
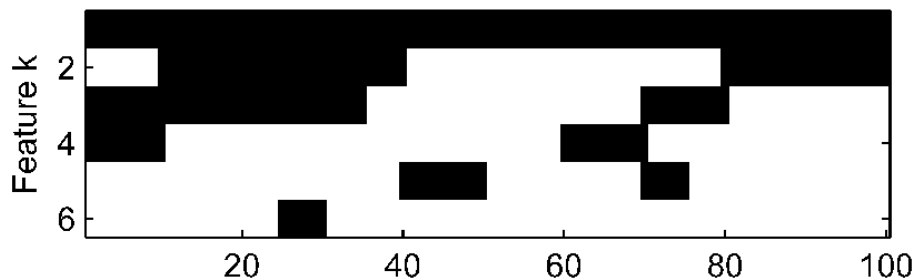
確率的行列分解 (PMF) 非負値行列分解 (NMF)

あるフレームのスペクトルは  
複数のスペクトルの足し合わせ

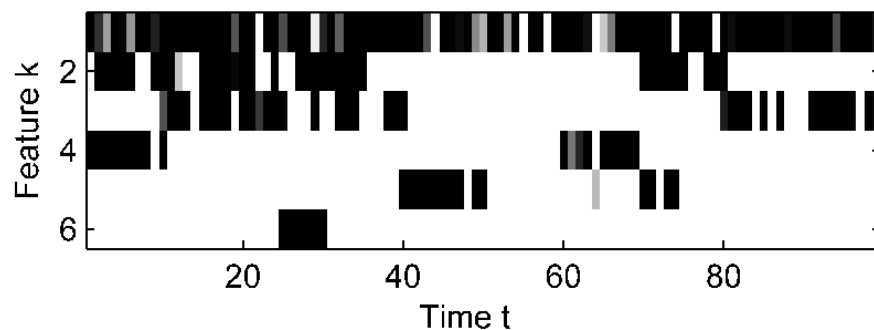
# Infinite ICA/SFA

- 可算無限個の信号源を許容する独立成分分析  
およびスパース因子分析 [Knowles2007]
  - インド料理過程 (IBP) を利用
    - ある客はそれまでの人気に比例した確率で料理を**複数**選ぶ
    - さらに新しい料理にも確率的にチャレンジする

正解



推定結果



MCMCによる推論：  
必要な信号源の数 (料理の種類数) を  
増減させながら事後分布空間を探索

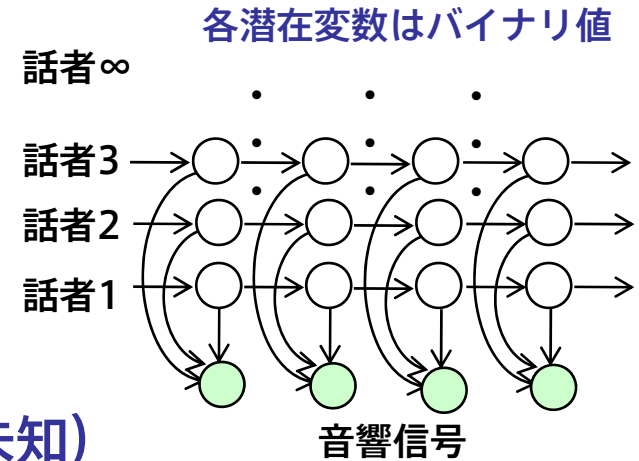
# Infinite Factorial HMM

- 可算無限個の隠れマルコフチェーンを許容するHMM [VanGael2009]

- マルコフインド料理過程 (mIBP)

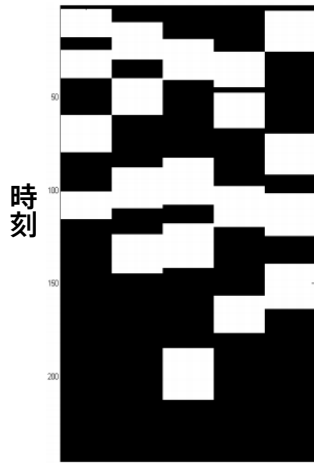
- ICAと組み合わせることが可能

- 理論的にはNMFとの組み合わせも可能



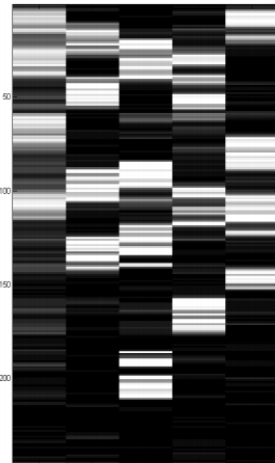
混合音からの話者区間推定実験 (話者数未知)

正解



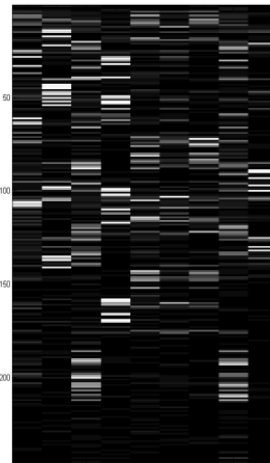
(a) Ground Truth

10マイクロホン

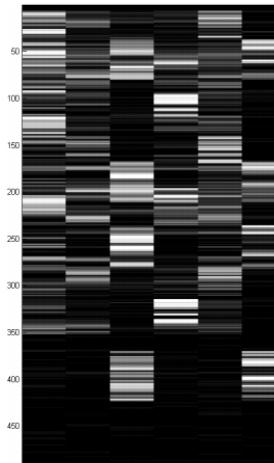


(b) ICA iFHMM

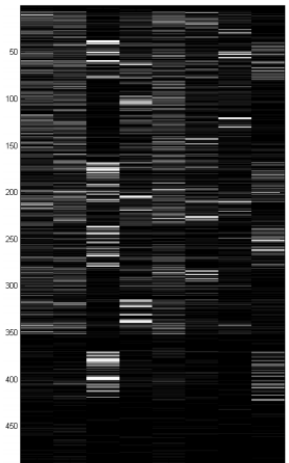
3マイクロホン



(c) iICA



(d) ICA iFHMM



(e) iICA

# おわりに

- 既存モデルの単純なNPB化はやった者勝ち
  - 有名なモデルは既にほとんどやりつくされた
- 確率モデルの定式化に活路を見いだすべき
  - NPB事前分布の設計は超高難度 → ML業界の仕事
  - 音楽情報処理独自の確率モデル+NPB化が大事！
- 音楽情報処理分野での適用例はまだ少数
  - しかしNPBに適したテーマは多岐にわたる
    - 音響信号中に音源は何個含まれているか？
      - 無限潜在的調波配分法 [吉井: SIGMUS86, ISMIR2010]
      - ガンマ過程非負値行列分解 [Hoffmann: ICML2010]
    - スペクトルの取りうる状態は何種類か？
      - 無限状態スペクトルモデル [中野: SIGMUS86, SIGMUS91, ICASSP2011]
    - 楽譜情報から演奏表現への写像関数どう設定すべきか？
      - ガウス過程に基づく名演奏の学習 [寺村: SIGMUS78]

# 参考資料

- **ディリクレ過程 (latent class modelのノンパラメトリックベイズ化)**
  - **Infinite GMM**
    - Carl Edward Rasmussen:  
The Infinite Gaussian Mixture Model (NIPS 2000)
  - **Infinite HMM**
    - Matthew J. Beal, Zoubin Ghahramani, Carl Edward Rasmussen:  
The Infinite Hidden Markov Model (NIPS 2001)
    - Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, Zoubin Ghahramani:  
Beam Sampling for the Infinite Hidden Markov Model (ICML 2008)
  - **Infinite PCFG**
    - Percy Liang, Slav Petrov, Michael I. Jordan, Dan Klein:  
The Infinite PCFG using Hierarchical Dirichlet Processes (EMNLP 2007)
  - **Infinite LDA**
    - Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, David M. Blei:  
Hierarchical Dirichlet Processes (Journal of ASA 2006)
  - **Nonparametric Bayesian N-gram Models**
    - Yee Whye Teh:  
A Bayesian Interpretation of Interpolated Kneser-Ney (Technical Report 2006)
- **ベータ過程 (latent feature modelのノンパラメトリックベイズ化)**
  - **Infinite Latent Feature Models**
    - Thomas L. Griffiths, Zoubin Ghahramani:  
Infinite Latent Feature Models and the Indian Buffet Process
  - **Infinite ICA**
    - David Knowles, Zoubin Ghahramani:  
Infinite Sparse Factor Analysis and Infinite Independent Components Analysis (ICA 2007)
  - **Infinite FHMM**
    - Jurgen Van Gael, Yee Whye Teh, Zoubin Ghahramani:  
The Infinite Factorial Hidden Markov Model (ICML 2008)

# 参考資料

- **上田 修功**
  - ノンパラメトリックベイズモデル (日本応用数理学会 Vol. 17, No. 3, pp.196-214, 2007)
    - Paper: [http://www.kecl.ntt.co.jp/as/members/yamada/dpm\\_ueda\\_yamada2007.pdf](http://www.kecl.ntt.co.jp/as/members/yamada/dpm_ueda_yamada2007.pdf)
  - ノンパラメトリックベイズモデル入門 ( CVIM3月研究会, 2009)
    - Slides: <http://www.kecl.ntt.co.jp/as/members/ueda/pdf/CVIM-ueda.pdf>
- **持橋 大地**
  - ノンパラメトリックベイズ (IBIS 2008)
    - Slides: <http://chasen.org/~daiti-m/paper/ibis2008-npbayes-tutorial.pdf>
- **Yee Whye Teh**
  - Nonparametric Bayesian Models
    - Video & Slides: [http://videlectures.net/mlss09uk\\_teh\\_nbm/](http://videlectures.net/mlss09uk_teh_nbm/)
    - Paper: <http://www.gatsby.ucl.ac.uk/~ywteh/research/npbayes/OrbTeh2010a.pdf>
  - A Tutorial on Dirichlet Processes and Hierarchical Dirichlet Processes
    - Slides: <http://mlg.eng.cam.ac.uk/tutorials/07/ywt.pdf>
  - Dirichlet Processes: Tutorial and Practical Course
    - Video & Slides & Paper: [http://videlectures.net/mlss07\\_teh\\_dp/](http://videlectures.net/mlss07_teh_dp/)
- **Zoubin Ghahramani**
  - Non-parametric Bayesian Methods (UAI 2005)
    - Slides: <http://learning.eng.cam.ac.uk/zoubin/talks/uai05tutorial-b.pdf>
  - A Brief Overview of Nonparametric Bayesian Models (NIPS 2009)
    - Slides: <http://learning.eng.cam.ac.uk/zoubin/talks/nips09npb.pdf>
- **Michael I. Jordan**
  - Dirichlet Processes, Chinese Restaurant Processes, and all that (ICML 2005)
    - Video & Slides: [http://videlectures.net/icml05\\_jordan\\_dpcrp/](http://videlectures.net/icml05_jordan_dpcrp/)
- **David M. Blei**
  - A Tutorial on Bayesian Nonparametric Models
  - Report: <http://www.princeton.edu/~sjgershm/npbayes.pdf>
- **その他**
  - 機械学習に関するトップレベルの国際会議 (NIPS, ICML, UAI, AISTATS など) は要チェック