Nonparametric Bayes for Non-Bayesians

IBIS 2008 企画セッション
「ノンパラメトリックベイズ」

持橋大地
NTT コミュニケーション科学基礎研究所
daichi@cslab.kecl.ntt.co.jp

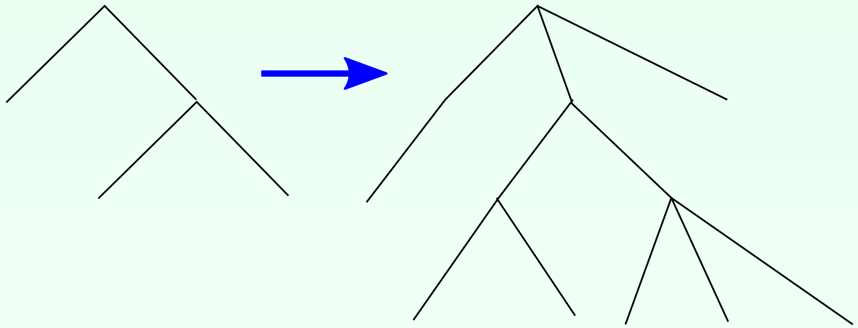
2008-10-29 (木)
仙台

チュートリアル

- ノンパラメトリックベイズって何?
- 混合モデルとディリクレ過程混合モデル
- 階層ディリクレ過程
- ノンパラメトリックベイズのポイント:
  $\cdots$ Stick-breaking process (SBP)
- Infinite Tree, Indian Buffet, The Mondrian Process, $\cdots$
- ノンパラメトリックベイズと測度論
- まとめ

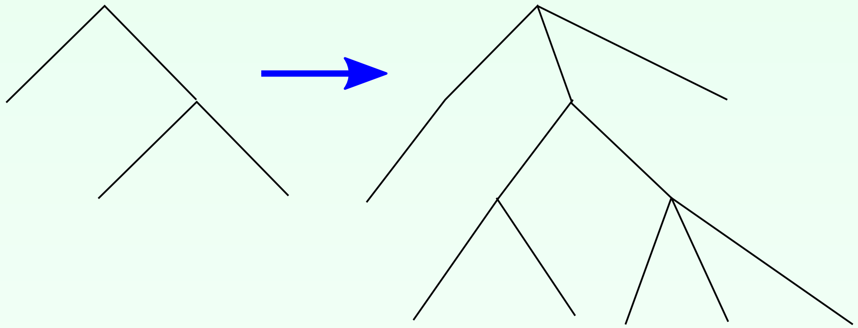
ノンパラメトリックベイズとは

- モデルをデータの複雑さに応じて増やす生成モデル
  - クラスタリング (クラスタ)
  - グラフマイニング (木, グラフ)
  - 自然言語処理 (カテゴリ, 単語)
- 理論的に可算無限個まで増やせる
  - "Infinite Models" とも呼ばれている
  - 実際には, n 個のデータに対して $O(\log n)$ 前後
    - データの複雑さによって変わる



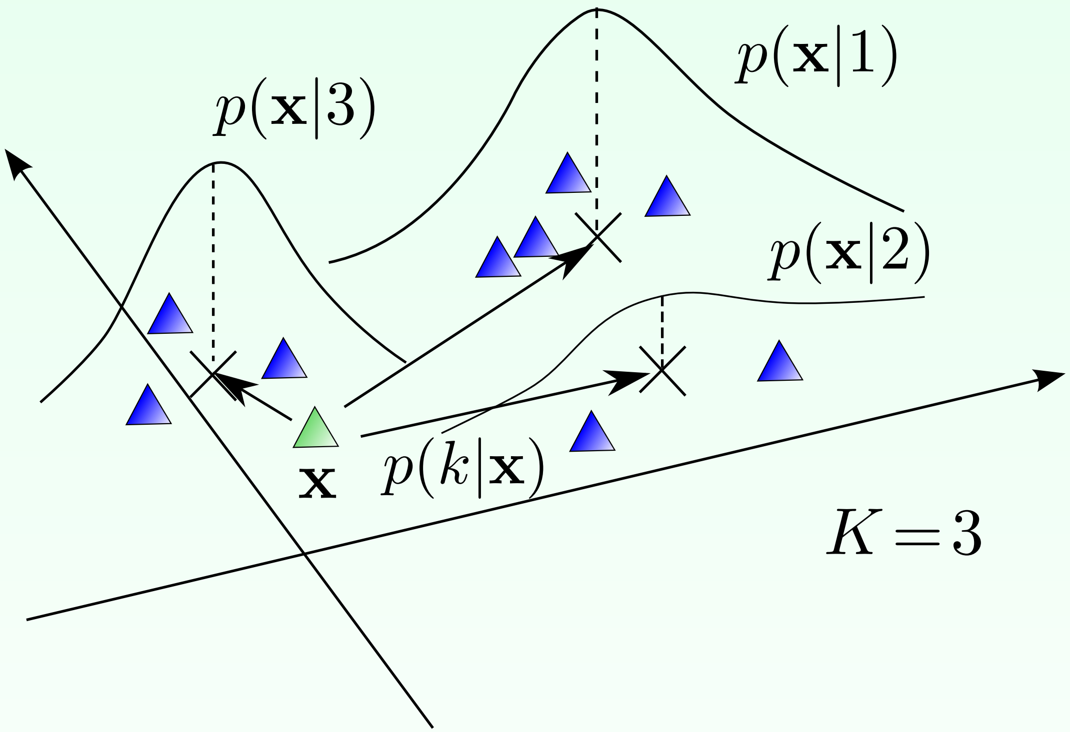
ノンパラメトリックベイズとは

- モデルをデータの複雑さに応じて増やす生成モデル
  - クラスタリング (クラスタ)
  - グラフマイニング (木, グラフ)
  - 自然言語処理 (カテゴリ, 単語)



- 理論的に可算無限個まで増やせる
  - "Infinite Models" とも呼ばれている
  - 実際には, n 個のデータに対して $O(\log n)$ 前後
    - データの複雑さによって変わる
- 注:「ノンパラメトリック」
  - パラメータがないという意味ではない
  - 少数パラメータの連続分布の推測ではない, という意味

簡単な例: 混合モデル

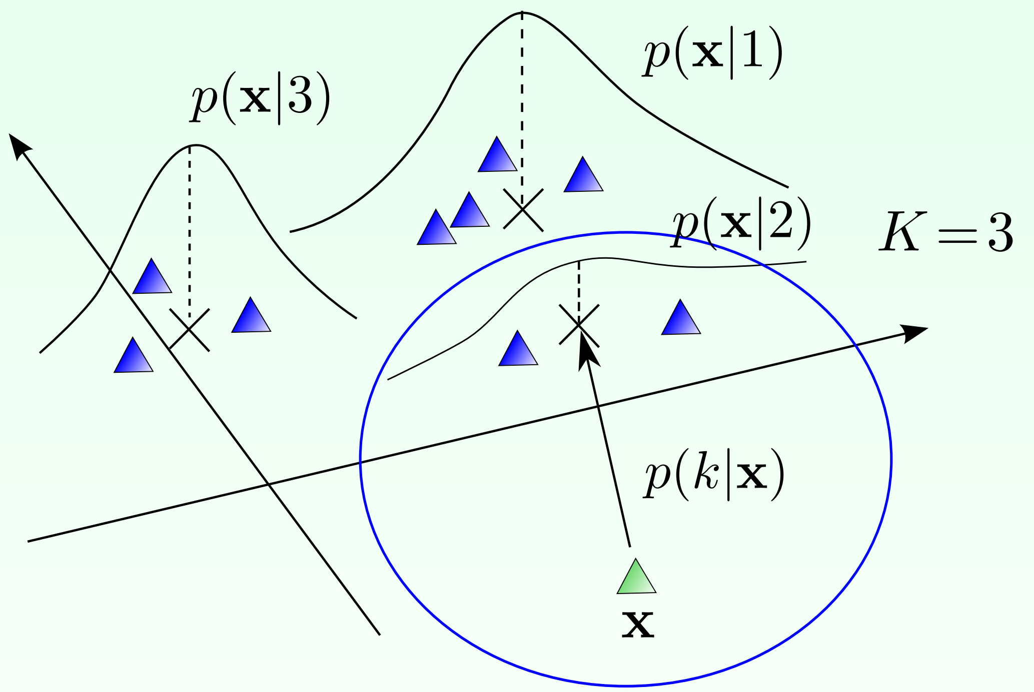


- データ $\mathbf{x}$ は確率分布 $1 \cdots K$ のどれかから生成
- EM アルゴリズム: $\mathbf{x}$ が $1 \cdots K$ のどの分布から生成されたか?

$$p(k|\mathbf{x}) \propto p(\mathbf{x}|k)\, p(k) \quad (k = 1 \cdots K) \tag{1}$$

を計算する

混合モデルの推定 (2): 問題



- 問題: 外れ値の場合はどうするか?
  - K が有限の場合, $p(k|\mathbf{x})$ が比較的大きいものに無理矢理割り当てる

    $\Downarrow$
- $p(K+1|\mathbf{x})$ (新しいクラスタ) を考えた方がいいのでは?
  - $p(K+1|\mathbf{x}) \propto p(\mathbf{x}|K+1)\,p(K+1)$ に従って, 新しいクラスタに帰属させる

混合モデルの推定 (3)

- ちょっと待った: 新しいクラスタは, $K+1, K+2, K+3, \cdots$ と無限にあるのでは?
  - EM で $p(K+1|\mathbf{x}), p(K+2|\mathbf{x})$, .. を全部持っておくのは無理

    $\Downarrow$

- MCMC 法で, 確率的にランダムにクラスタ番号を決めよう
  - “現在のクラスタ数の最大値 $+1$” まで考えればよい

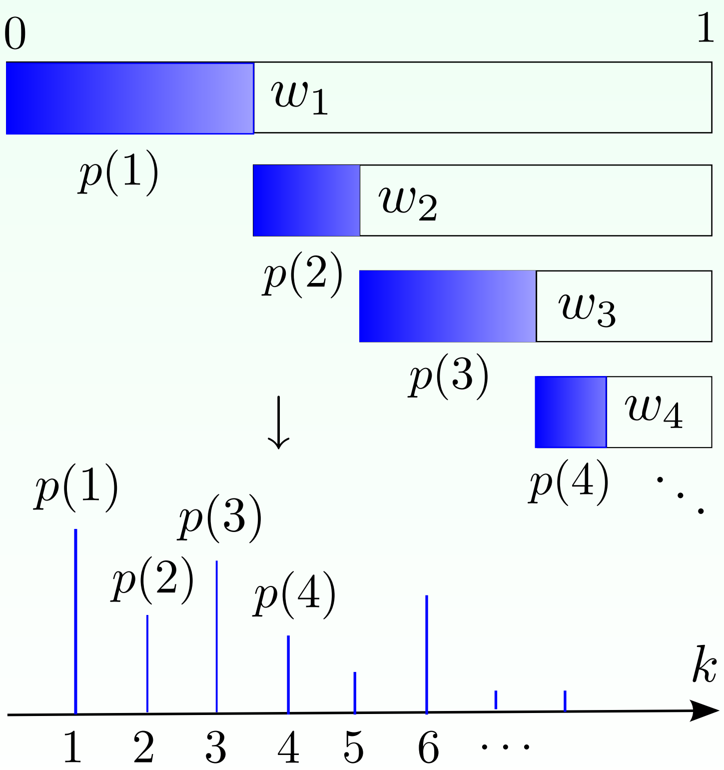
for $n = 1 \cdots N$ **do**
    for $k = 1 \cdots K+1$ **do**
        $p(k|\mathbf{x}_n) \propto p(\mathbf{x}_n|k)\,p(k)$
    end for
    $z_n \sim p(\cdot|\mathbf{x}_n)$
    if $z_n = K+1$ **then**
        $K = K+1$
    end if
end for

ディリクレ過程混合モデル (DPM)

- 今のアルゴリズムを行うためには, $p(k)\ \ (k = 1, 2, \cdots, \infty)$ が必要

  $\Downarrow$

- GEM 分布, Stick-breaking process (SBP)



$$w_k \sim \mathrm{Be}\,(1, \alpha) \quad (k = 1 \cdots \infty) \tag{2}$$

$$p(k) = w_k \prod_{i=1}^{k-1} (1 - w_i) \tag{3}$$

- ○ 幾何分布

$$\mathrm{Geo}(k) = \lambda\,(1-\lambda)^{k-1} \tag{4}$$

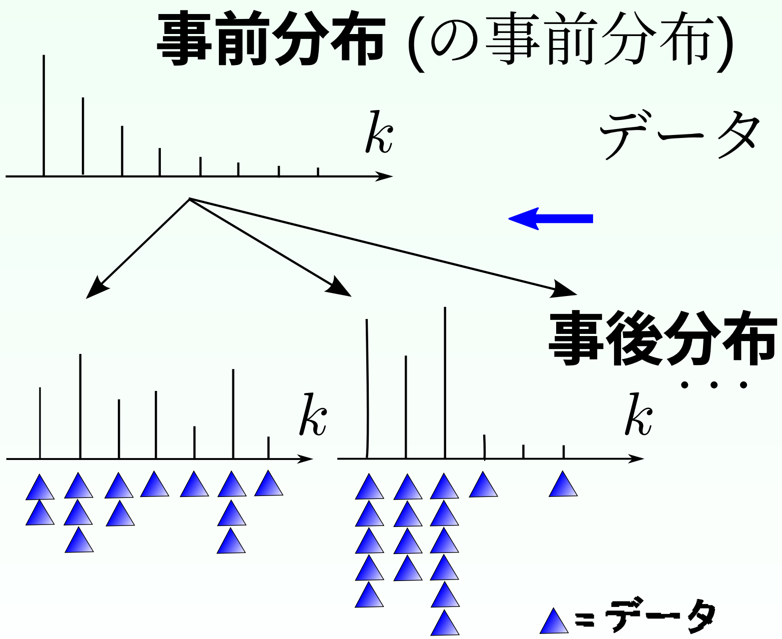
のソフト化.

Stick-breaking process: 事前分布→事後分布

- Stick-breaking prior:

$$w_k \sim \mathrm{Be}(1, \alpha), \quad p(k) = w_k \prod_{i=1}^{k-1} (1 - w_i) \qquad (5)$$

  - ○ 事前分布は最初は指数減衰→データによって変化
    - – 「事前分布の事前分布」

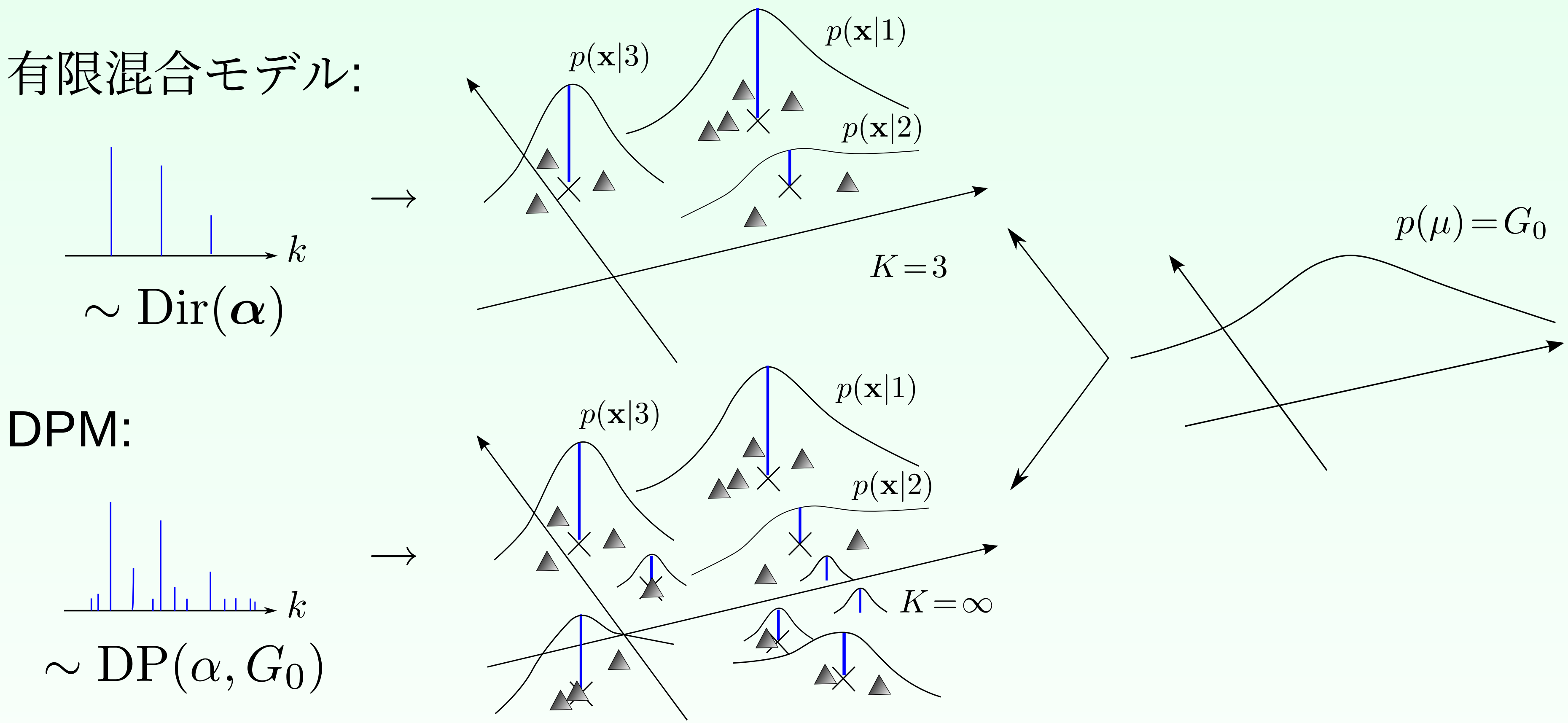


$$w_k \sim \mathrm{Be}\,(1, \alpha) \qquad (6)$$

$$\Downarrow$$

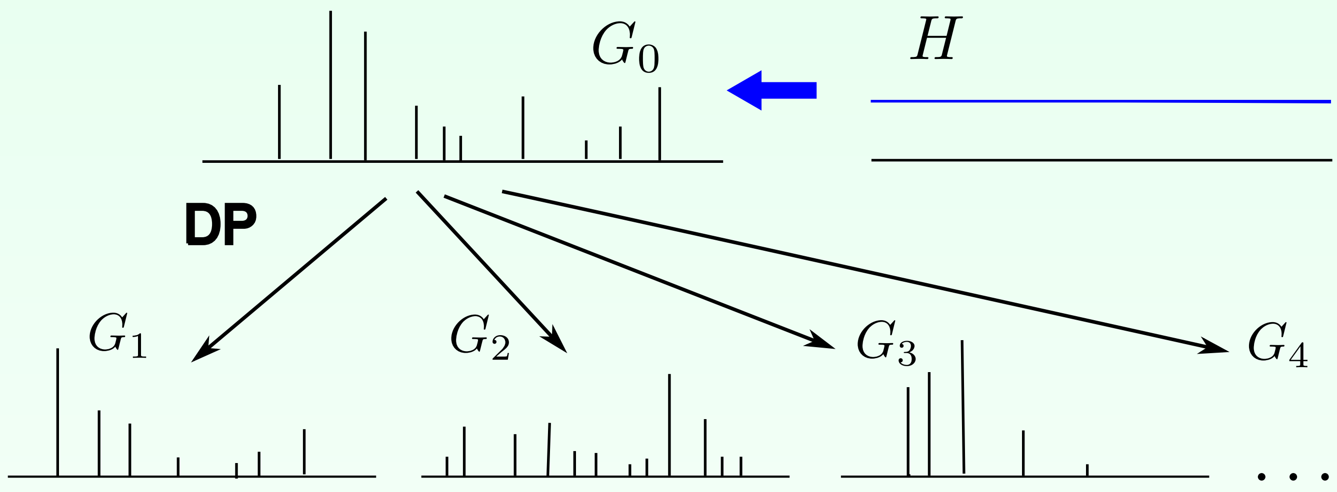
$$w_k \sim \mathrm{Be}\,(1 + \#(k\ \text{で止まったデータ数}),$$
$$\alpha + \#(k\ \text{より右のデータ数})) \qquad (7)$$

振り返ってみると



- ディリクレ過程とは…「元」となる分布 (基底測度) に似た, 無限離散分布を生成するモデル
  $\rightarrow$ データをフィッティングする
  - データが多く複雑なほど, より多くの「クラスタ」が現れる

HDP (Hierarchical Dirichlet Process) (Teh+ 2004)



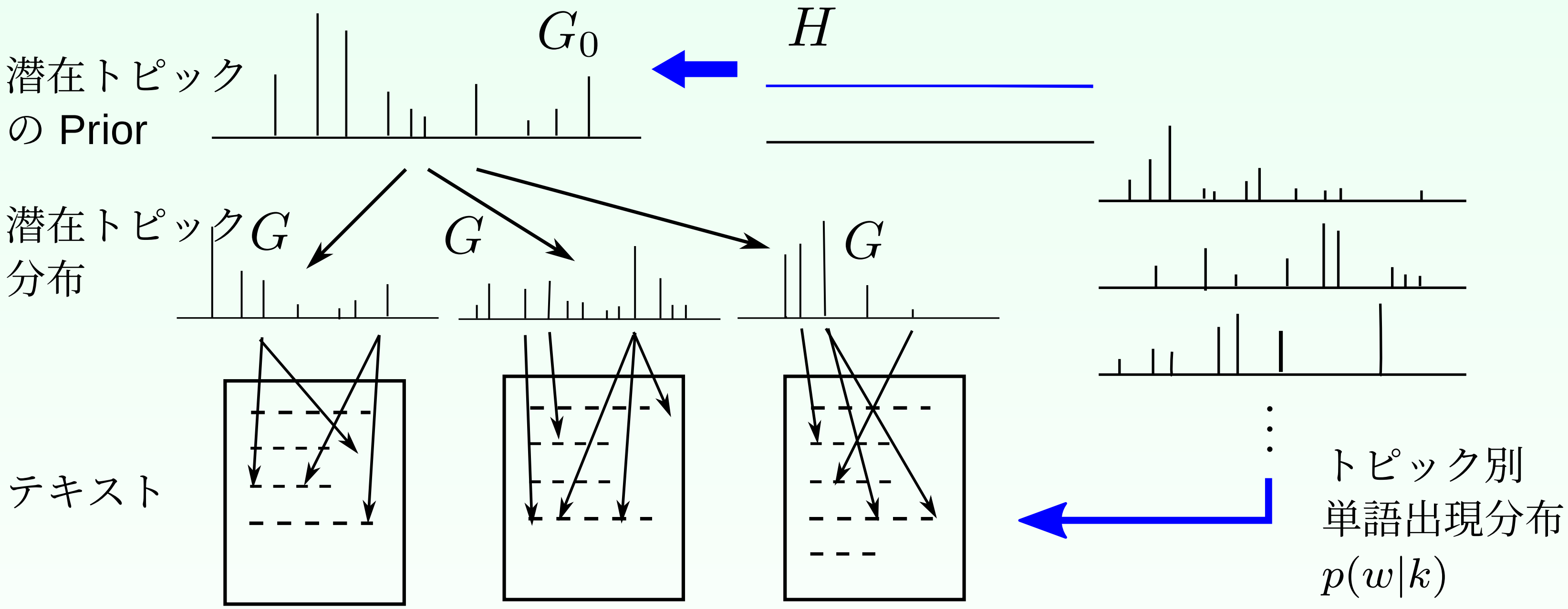
- DP の基底測度を, 親 DP から引いた無限離散分布にする

  $\Downarrow$

- ある離散分布に似た, 多数のランダムな離散分布
  - 「ある混合比に似た, 多数のランダムな混合比」
  - 技術的には: $\{G_n\}$ の中で, アトムが共有される
    - H から直接 $\{G_n\}$ を引くと, $\perp$ の立っている位置が共有されない

HDP の応用

- HDP-LDA : 無限トピックの潜在意味モデル
  - 混合モデルの混合モデル



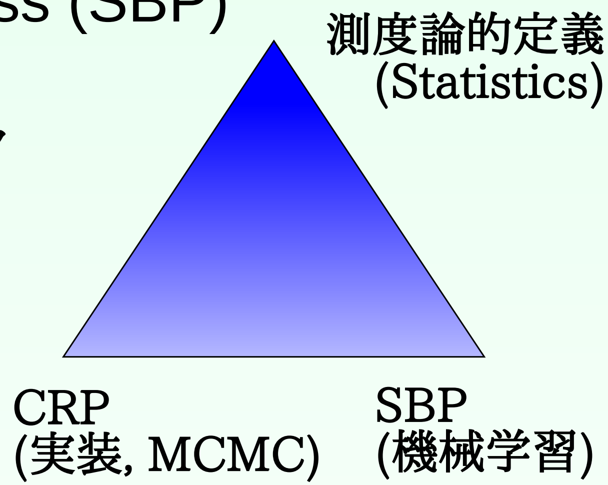
$$p(\mathbf{w}|\eta, H) = \int \prod_{n=1}^{N} \sum_{k=1}^{\infty} p(w_n|k) \underbrace{G[k]}_{\text{無限離散分布の第 } k \text{ 要素}} dG \qquad (8)$$

- 実装には Gibbs sampling を用いる
  - http://www.gatsby.ucl.ac.uk/˜ywteh/research/npbayes/npbayes-r21.tgz

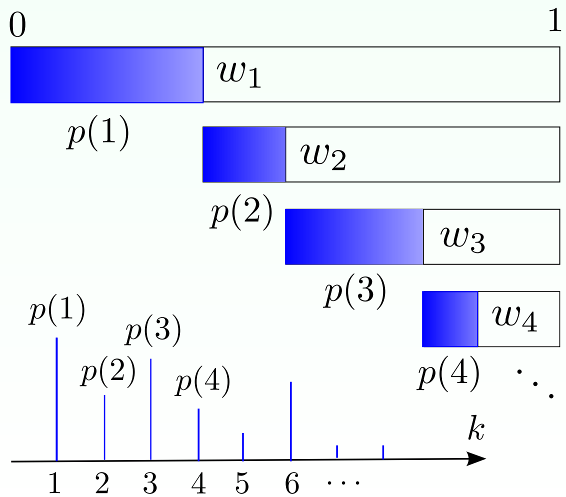
The Trinity of Nonparametric Bayes

- ノンパラメトリックベイズの本質 (機械学習として)
  $\cdots$ Stick-Breaking Process (SBP)

ノンパラメトリック
ベイズ法



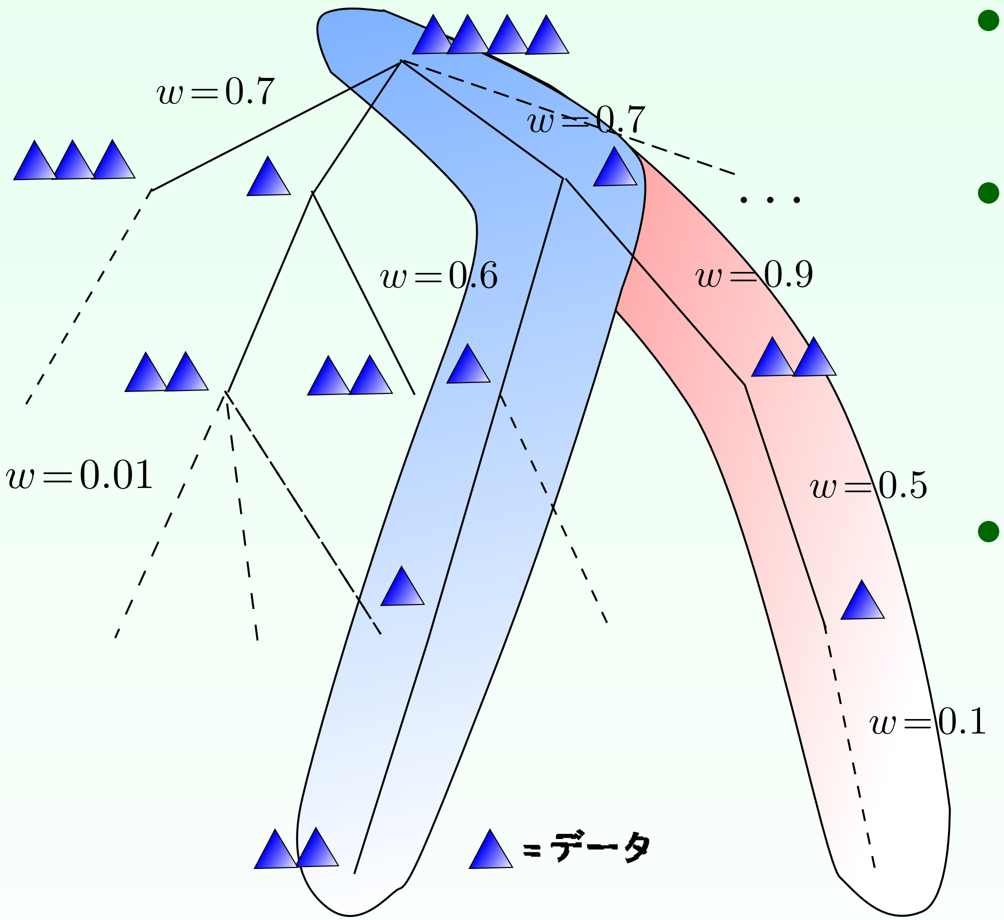
- Stick-Breaking Process の拡張



$$w_k \sim \mathrm{Be}(1, \alpha) \tag{9}$$

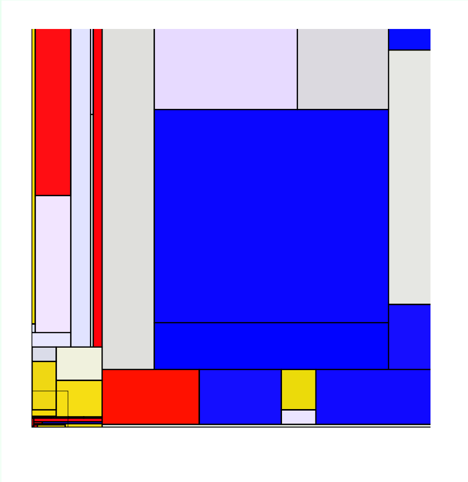
- ○ $\mathrm{Be}(1, \alpha)$ でなくともよい
  - $\mathrm{Be}(1-\alpha, \beta+k\alpha)$ $\cdots$ Pitman-Yor process (Pitman and Yor 1997)
  - $\mathrm{Be}(\alpha, \beta)$ $\cdots$ Beta two-parameter process (Ishwaran and Zarepour 2000)

Infinite Stochastic Tree

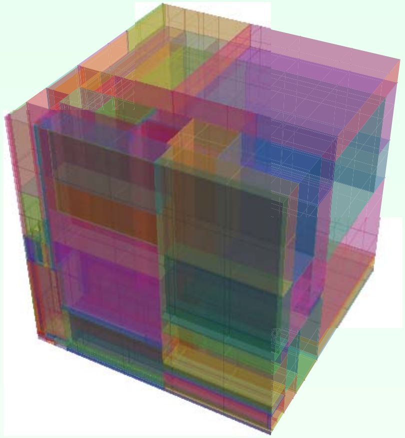


- 木の各ノードに, $w \sim \mathrm{Be}(\alpha, \beta)$ がある
- 木の各 Branch が, 一つの SBP
  - SBP の構造化
  - 深いノードほど, 到達しにくい (木が「薄い」)
- 有限のデータでの近似
  - The Infinite Markov Model : Gibbs
  - NTR (Neutral-To-Right) process と関係が深いらしい
  - 一般のグラフにも同様の無限化を考えることができそう

The Mondrian Process (Roy and Teh, NIPS 2008 to appear)

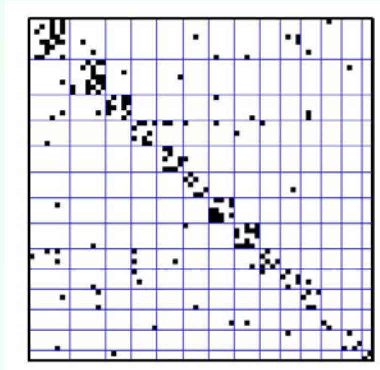


Mondrian Process



Mondrian (3D)

$\Longleftrightarrow$



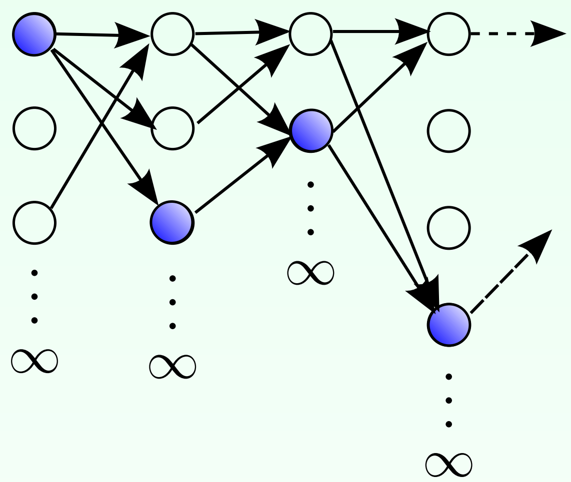
IRM

- Infinite Relational Model の最先端···確率的階層分割
  - Poisson process の階層化
  - Blockmodel $\rightarrow$ IRM (Kemp+ 2006) $\rightarrow$ Annotated Hierarchy (Roy+ 2006) $\rightarrow$ Mondrian
- 再帰的な定義

$$
\begin{aligned}
&\mathrm{MP}(\lambda, \boldsymbol{\mu}, \boldsymbol{\Theta}) \equiv \\
&\qquad \lambda' = \lambda - E, E \sim \mathrm{Exp}(\mu(\boldsymbol{\Theta})), \cdots \\
&\qquad \mathrm{return}\ \langle \lambda', d, x, \mathrm{MP}(\lambda', \boldsymbol{\mu}, \boldsymbol{\Theta}^{<_d x}), \mathrm{MP}(\lambda', \boldsymbol{\mu}, \boldsymbol{\Theta}^{>_d x}) \rangle
\end{aligned}
$$

その他の新しい話題

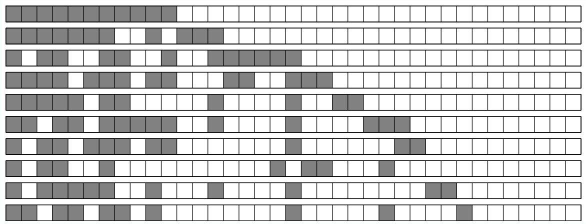
- The Infinite Hidden Markov Model (Beal+ 2002)



  - HMM の隠れ状態数が推測できる (!)
  - HDP を使って, 可算無限個の隠れ状態を生成
    - 動的計画法も提案された (ICML 2008)
    - HDP-PCFG もこの一種

- Indian Buffet Process (Griffiths and Ghahramani 2005) = Beta Process (Hjort 1990)

  - 複数の隠れ素性を持つ無限マトリクスの生成モデル

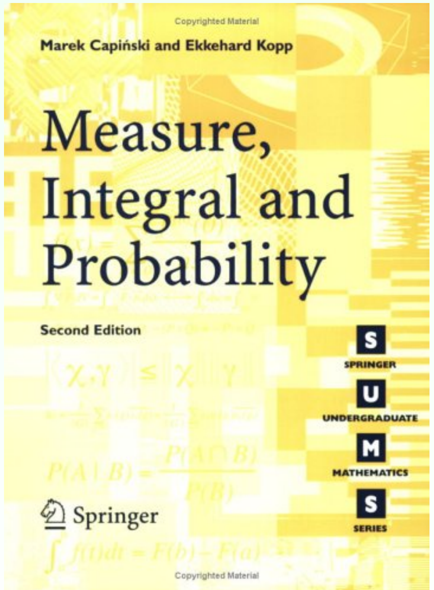


  - 階層化と HBP (Thibaux+ 2007···難解)
  - 時系列化: The Infinite Factorial HMM (NIPS 2008 to appear)

ノンパラメトリックベイズと測度論

- ノンパラメトリックベイズの学習に測度論は必要か?
  → 必要.
  - 全く基礎知識がないと, 研究のレベルが大きく制限される
  - ただし, 全部知る必要は (とりあえず) ない
  - 知らなくても, DPM+α 程度の実装は可能
- お薦め書籍:

  "Measure, Integral and Probability"
  Capinski and Kopp, Springer 2004.
  (Springer Undergraduate Mathematics series)
  ISBN 1852337818
  - ∼p70 くらいまで読めばよい
  - a.e. や $\mathcal{M}$ を見ても固まらなくて済む!



まとめ

- ノンパラメトリックベイズ…モデルを動的に増やす生成モデル
  - 複雑なデータほど, 複雑なモデル
  - クラスタリングに限らず, 木やグラフへも拡張可能
- DP, HDP, HPY, IBP, HBP, Mondrian, ⋯
  - 確率論として見ても深いモデル
  - 測度論の初歩は, 適切な本を使えば難しくない
  - NPBayes stat だけで, ご飯 20 杯くらい食べられる (ex. Newton Institute 2007, “Construction and Properties of Bayesian Nonparametric Regression Models”
    http://www.newton.ac.uk/programmes/BNR/bnrw01p.html
- 機械学習における多くの応用
  - 自然言語処理, バイオインフォマティクス
  - クラスタリング, データマイニング, ネットワーク分析
  - 画像処理, 音声処理, 音楽情報処理, ⋯

注意

- ノンパラベイズに, 汎用の「パッケージ」はない
  - 一応存在するが, 理論がわかっていないと使えない
  - ベイズ哲学…データを注意深く観察してモデル化する
  - 相手に応じて自分の振舞いを変える (動的: N. Wiener)

$$\Updownarrow$$

- どんな入力に対しても, 一定の性能を保障するミニマックス最適化 (静的: v. Neumann)
  - 入力分布を仮定しないので, パッケージ化できる
    (例: SVM)
  - ただし, ノンパラベイズのような柔軟性は難しい

参考文献 (チュートリアル類)

- NIPS 2005, Michael Jordan
  http://www.cs.berkeley.edu/˜jordan/nips-tutorial05.ps
- UAI 2005, Zoubin Ghahramani
  http://learning.eng.cam.ac.uk/zoubin/talks/uai05tutorial-b.pdf
- MLSS 2007, Yee Whye Teh
  http://www.gatsby.ucl.ac.uk/˜ywteh/research/npbayes/mlss2007.pdf
- 日本語:
  - 上田&山田, 「応用数理」2007
  - 佐藤, 中川研機械学習勉強会 2007
  - 栗原, MIRU 2008