

ノンパラメトリックベイズによるメディア処理

2012. 11. 15 AIチャレンジ研究会

NTTコミュニケーション科学基礎研究所

中野允裕

nakano.masahiro@lab.ntt.co.jp

5, 6年前であれば...

教科書に載っているような各種ツールの
ノンパラベイズ化が話題の中心になっていたが...

主成分分析

独立成分分析

非負値行列分解

隠れマルコフモデル

確率文脈自由文法

n -gram

ダイナミックベイジアンネット

ディープニューラルネット

...

2006

2012

Infinite HMM
[Beal+2002]

Sticky HDP-HMM
[Fox+2007]

Hidden Markov model

HDP-HMM
[Teh+2006]

Block diagonal iHMM
[Stepleton+2009]

Mondrian iHMM
[Nakano+2012]

IBP NMF
[2007以降多数]

GaP NMF
[Hoffman+2010]

Probabilistic
context
free grammar

Infinite PCFG
[Liang+2007]

Nonnegative matrix factorization

Infinite ICA
[2007以降多数]

Independent
component analysis

Infinite-gram model
[Mochihashi & Sumita2007]

Sequence memoizer
[2009以降多数]

N-gram model

Infinite hierarchical HMM
[Heller+2008]

Infinite dynamic Bayesian net.
[Finale+2011]

Combinatorial

State space model

Infinite factorial HMM
[Van Gael+2008]

Infinite latent event model
[Wingate+2009]

本発表の流れ

1. アプリケーション

2. ベイズとは？ ノンパラベイズとは？

2002

2012

既存ツールのノンパラベイズ化

モデルの拡張法

3. 混合モデル
と因子モデル

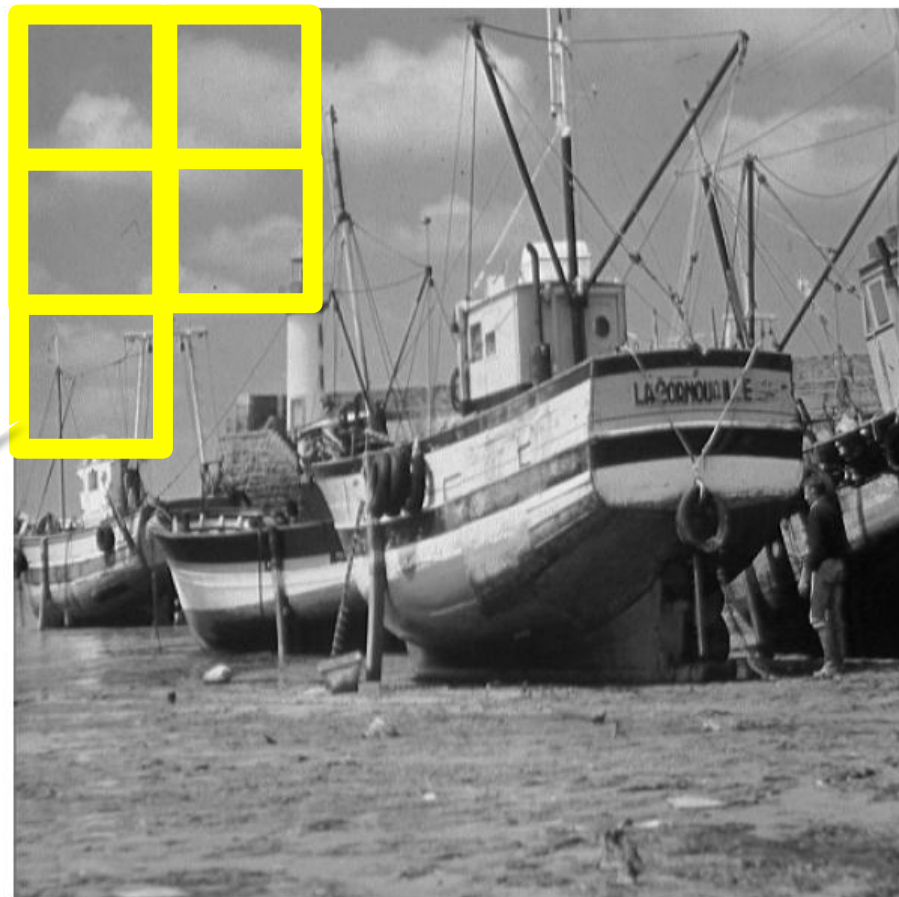
4. 階層
入れ子
相関

例: 電子透かし [Shterev&Dunson2012]

目的: 画像・動画像からのwatermarking

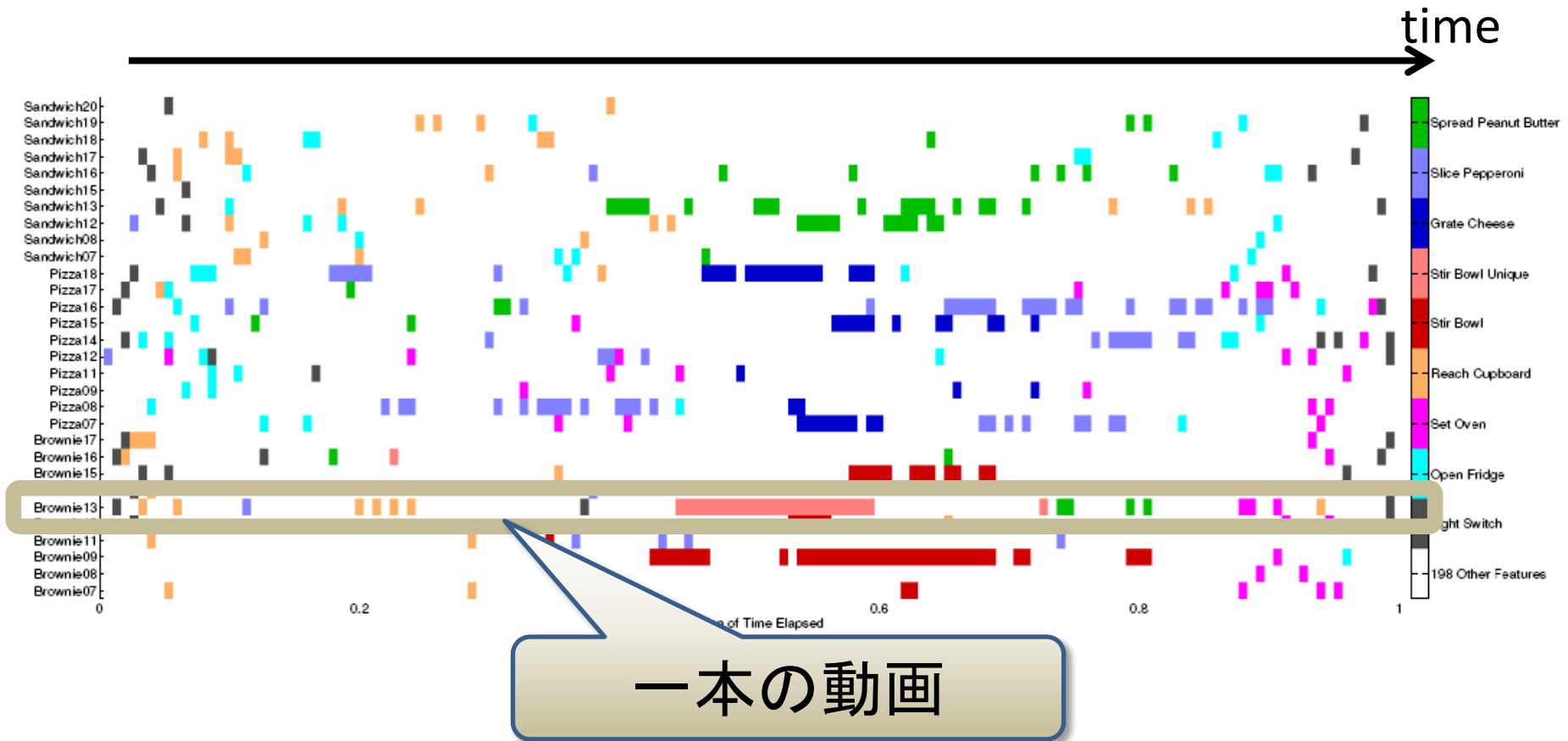
- ・埋め込み箇所未知
- ・埋め込み信号未知
- ・画像中に繰り返し登場

どこに埋め込んだのかは未知



例: 動画系列群からパターン発見 [Hughes&Sudderth2012]

目的: 動画系列群をbehaviorごとにセグメント分け

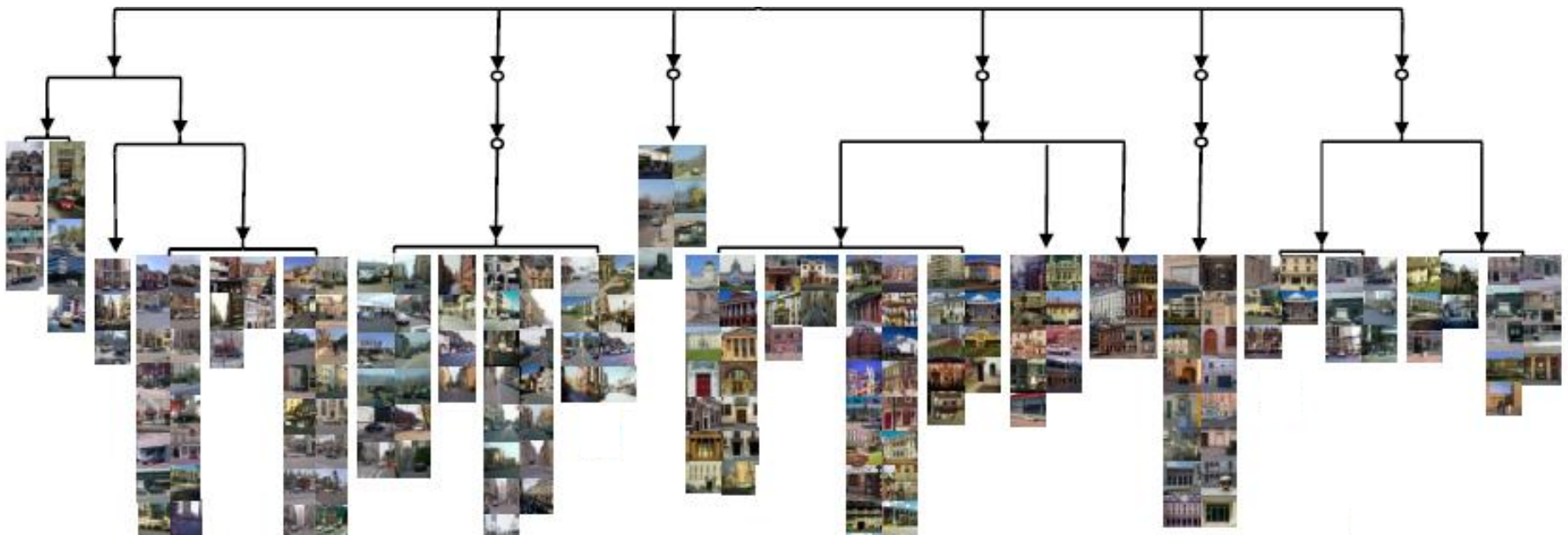


M. Hughes and E. Sudderth (2012),

Nonparametric discovery of active patterns from video collections, CVPR

例: 画像群のパーツ分解 + 階層的な分類 [Li+2012]

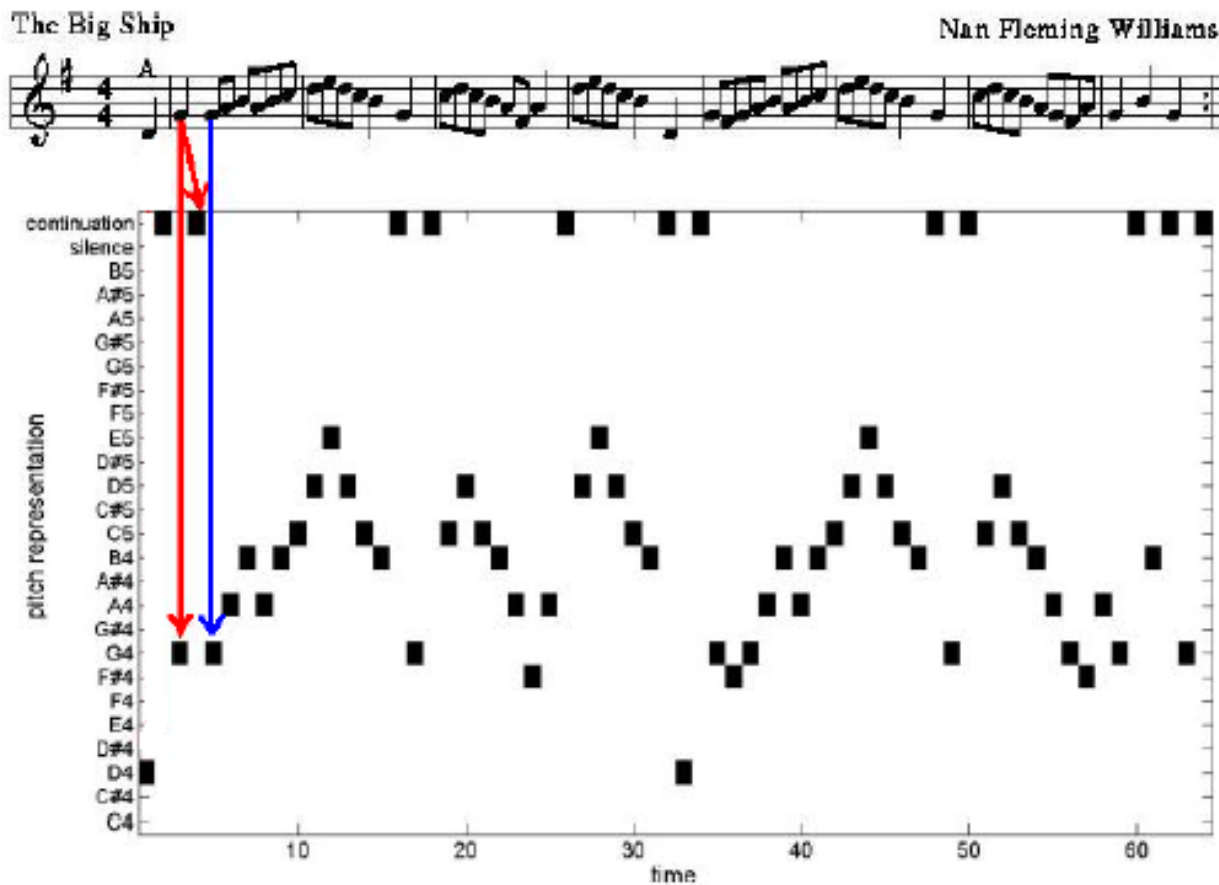
目的: 画像 + アノテーションの手がかりから、画像とそのパッチの階層クラスタリングをしつつ、画像のパーツ分解



L. Li, X. Zhang, M. Zhou and L. Carin (2012),
Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI.

例: 音楽のメロディーの予測・生成 [Spiliopoulou & Storkey2012]

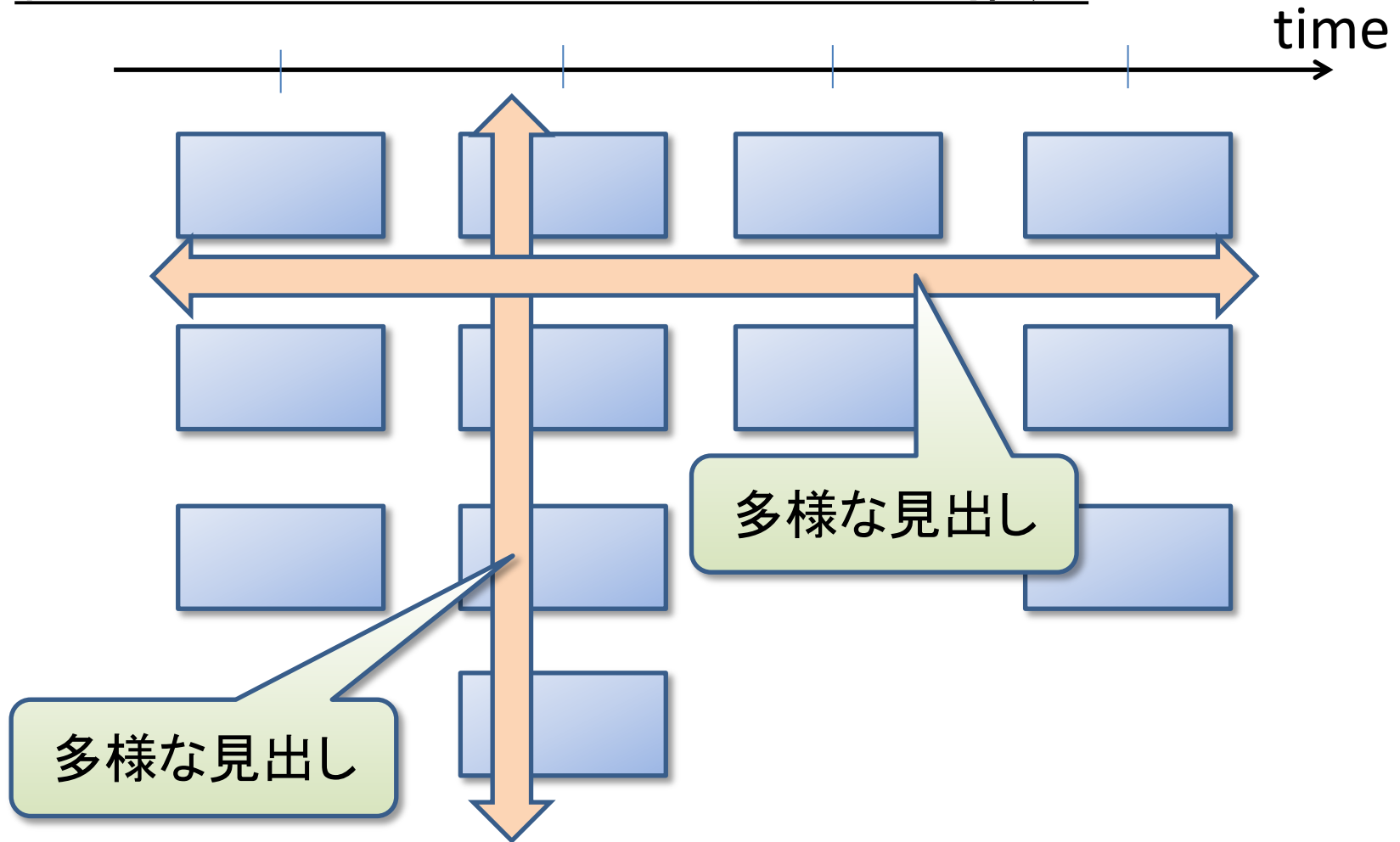
目的: メロディーのような時間方向に長い依存関係があるものの予測や自動生成



A. Spiliopoulou and A. Storkey (2012), A topic model for melody sequences, ICML.

例: ニュースヘッドラインの生成 [Affandi+2012]

目的: 日々のニュースヘッドラインの推定



ベイズ的なメディア処理の基本的な戦略

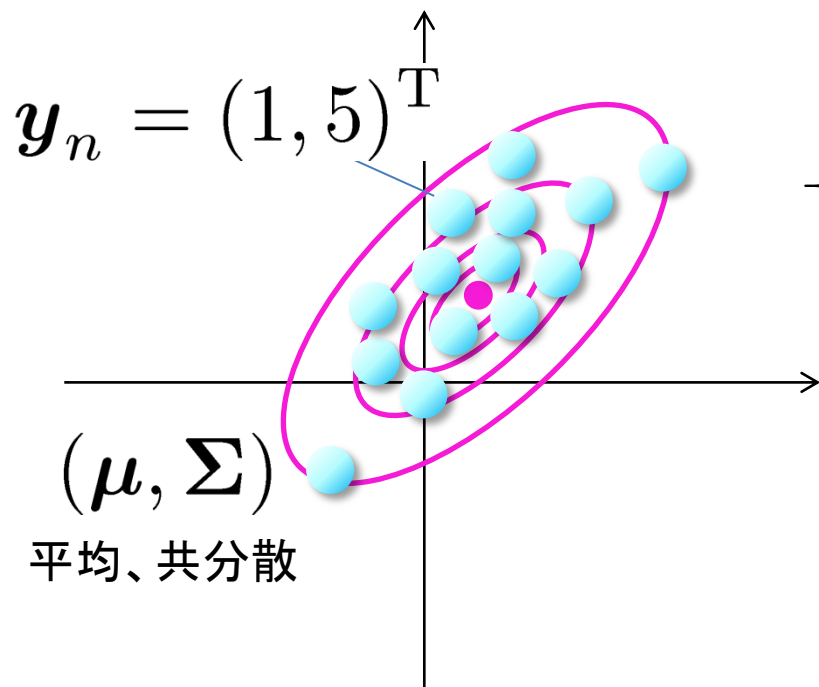
観測データの確率的生成モデル

$$\{\mathbf{y}_1, \dots, \mathbf{y}_N\} \sim f(\{\mathbf{y}_1, \dots, \mathbf{y}_N\}; \Theta)$$

N 個の観測データ

隠れ変数

→ $p(\{\Theta \mid \mathbf{y}_1, \dots, \mathbf{y}_N\})$ を推定



尤度関数: $\mathbf{y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

事前分布: $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim F(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, n_0)$$

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_k/m)$$

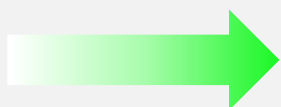
ベイズ的なメディア処理の基本的な戦略

観測データの確率的生成モデル

$$\{y_1, \dots, y_N\} \sim f(\{y_1, \dots, y_N\}; \Theta)$$

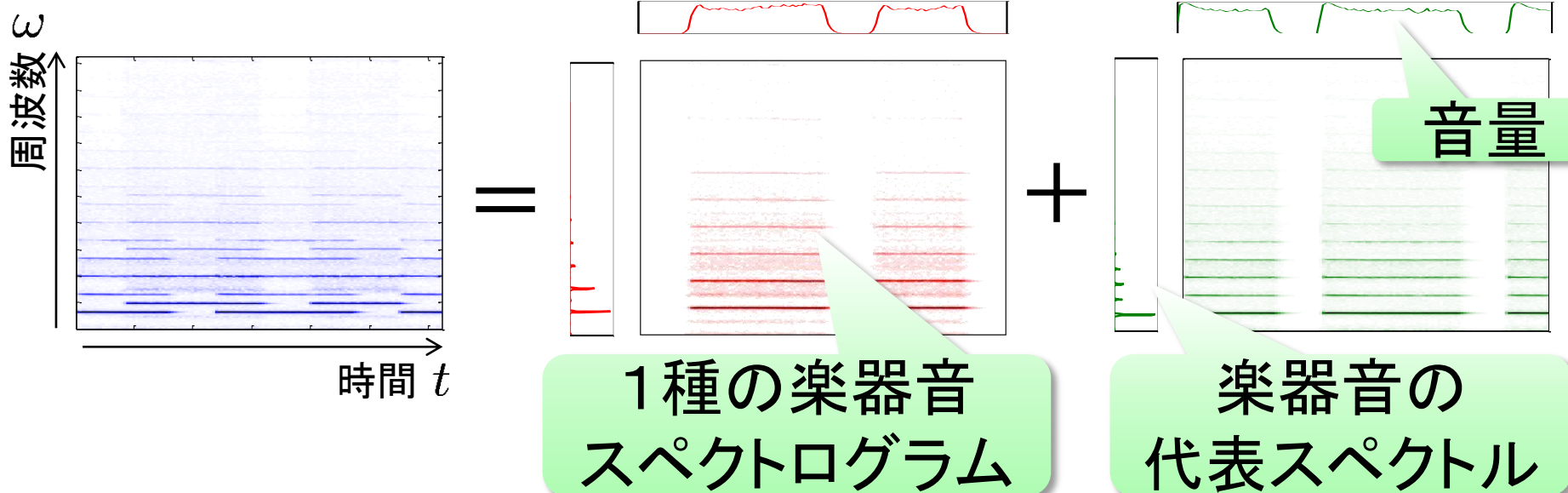
N 個の観測データ

隠れ変数



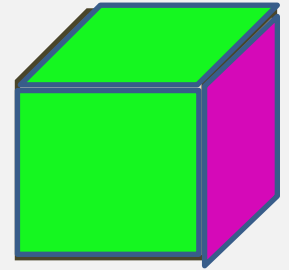
$p(\{\Theta \mid y_1, \dots, y_N\})$ を推定

例) 音楽信号分解



確率空間 (Ω, \mathcal{B}, P) : 可算集合の場合

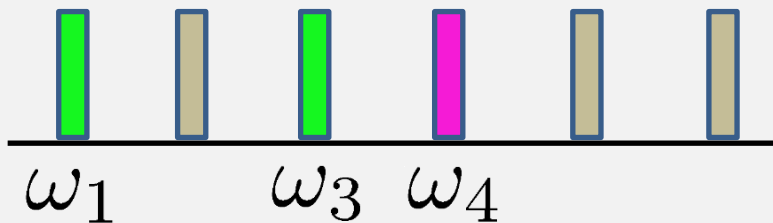
■ 標本空間 Ω 、標本点 ω



■ σ 加法族 \mathcal{B} : 標本空間の部分集合の集合族

$\phi, \{\omega_1\}, \dots, \{\omega_1, \omega_2\}, \dots, \{\omega_1, \omega_4, \omega_5\}, \dots,$
べき集合!!!

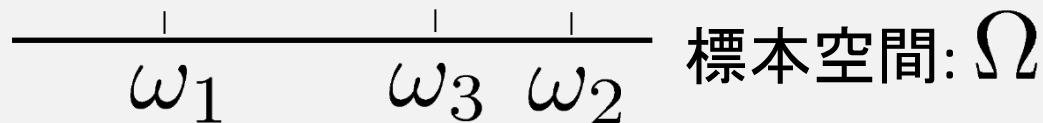
■ 確率測度 P : \mathcal{B} 上の非負値関数



$$P(\underbrace{\{\omega_1, \omega_3\}}_{\text{green}}) + P(\underbrace{\{\omega_4\}}_{\text{pink}}) \\ = P(\{\omega_1, \omega_3, \omega_4\})$$

確率空間 (Ω, \mathcal{B}, P)

- 標本空間 Ω 、標本点 ω

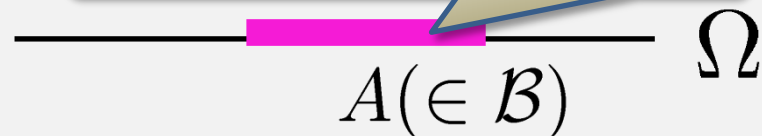


- σ 加法族 \mathcal{B} : 標本空間の部分集合の集合族

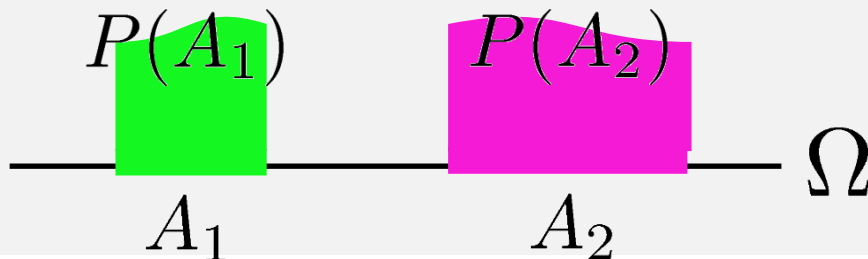
$$A_n \in \mathcal{B} \quad (n = 1, 2, \dots)$$

$$\Rightarrow \cup_n A_n \in \mathcal{B}$$

「長さ」が測れる!



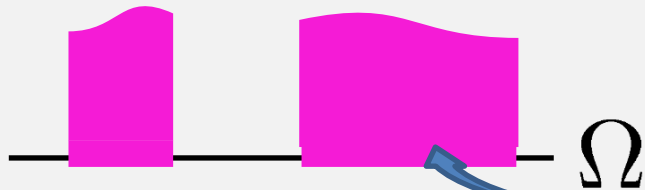
- 確率測度 P : \mathcal{B} 上の非負値関数



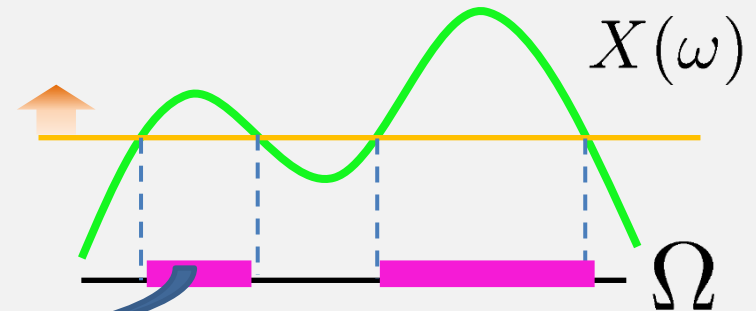
$$P(A_1) + P(A_2)$$
$$= P(A_1 \cup A_2)$$

確率変数と確率分布

確率空間: (Ω, \mathcal{B}, P)

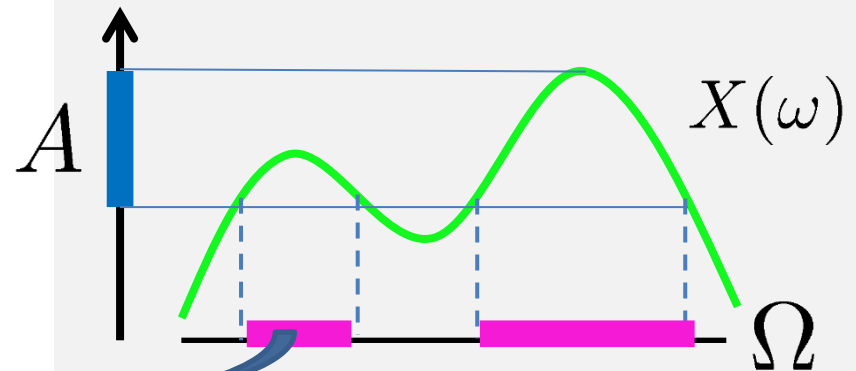
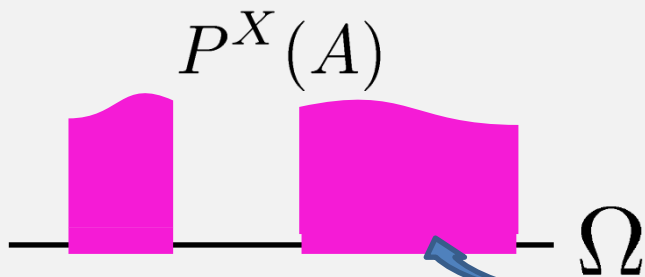


確率変数



「確率変数」とはとる値に対して確率が計算できるもの！

確率空間: (Ω, \mathcal{B}, P)



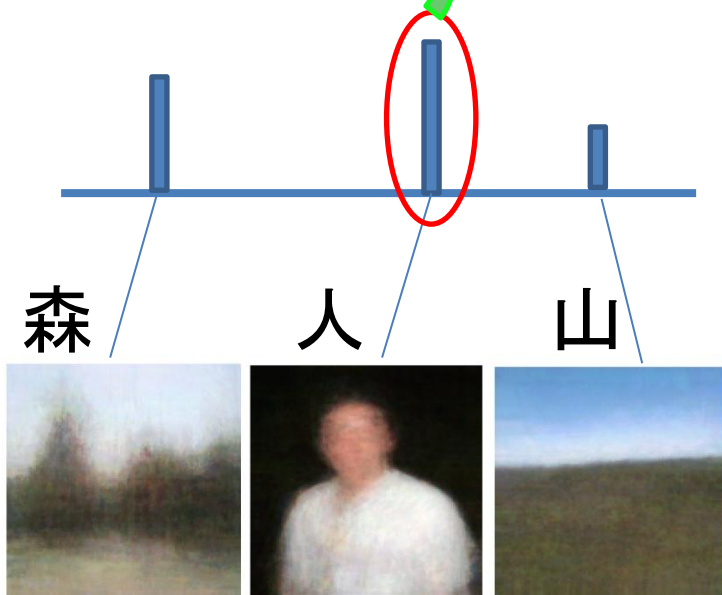
確率分布: $P^X(A) = P(\{\omega; X(\omega) \in A\})$ ¹⁴

二つの基本的な確率的生成モデル

混合モデル

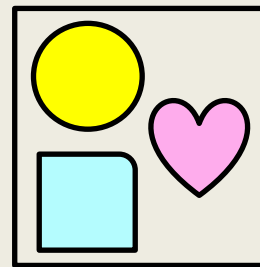


単一のクラスターがデータを生成

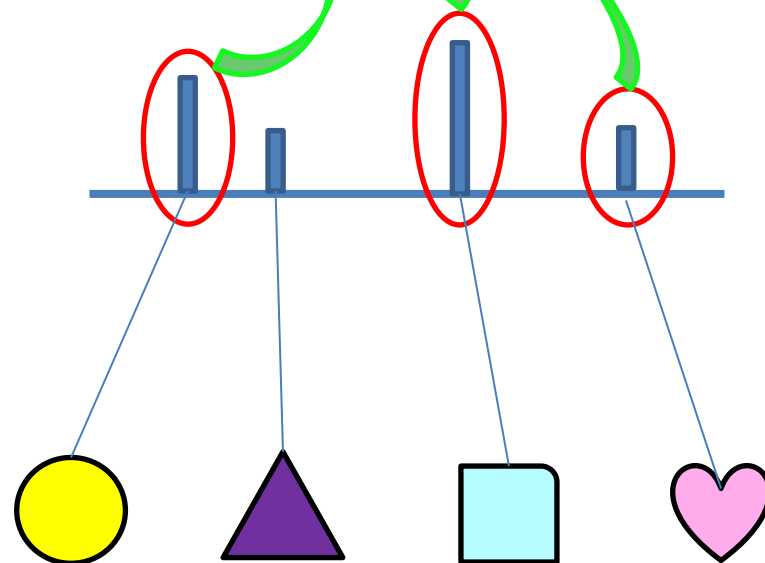


[Torralba+2008]

因子モデル



複数の因子がデータを生成



混合モデルの設計法

混合モデル



森

人

山



観測データの生成

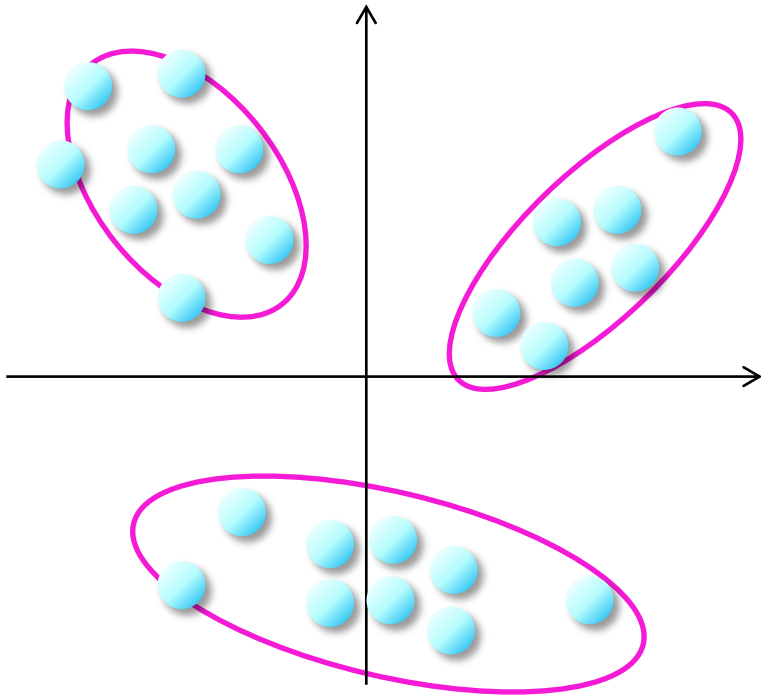
ラベルの割り当て

離散分布の生成

各ラベルごとの
パラメータの生成

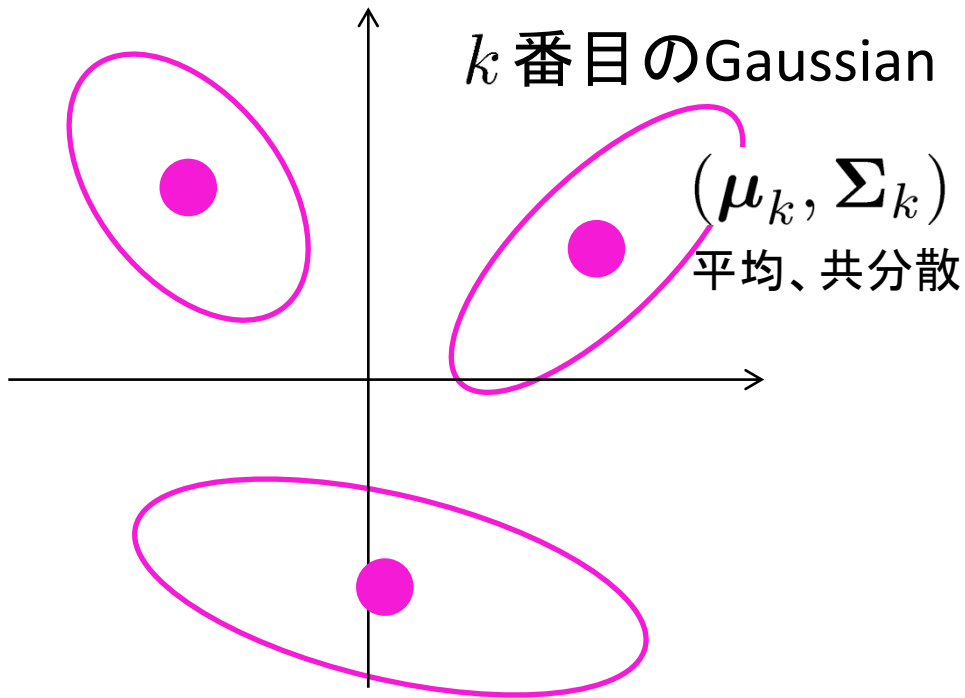
例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



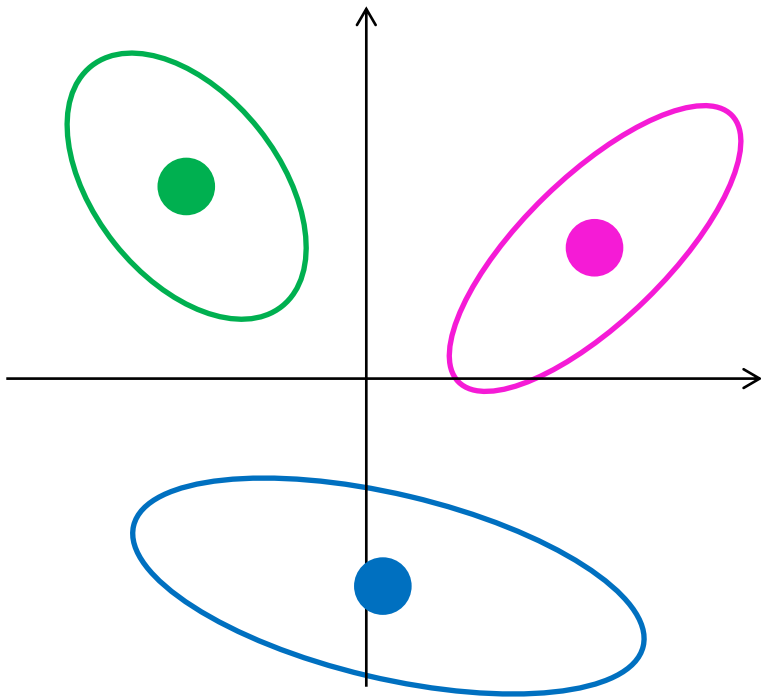
各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



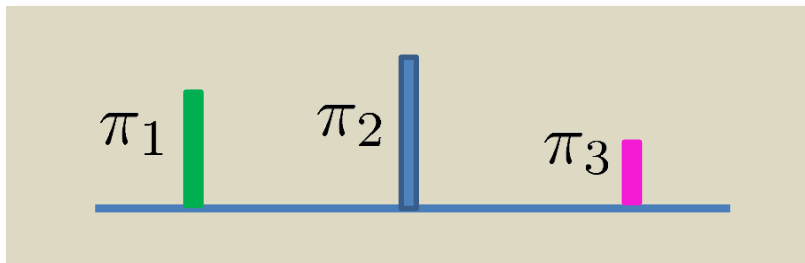
各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

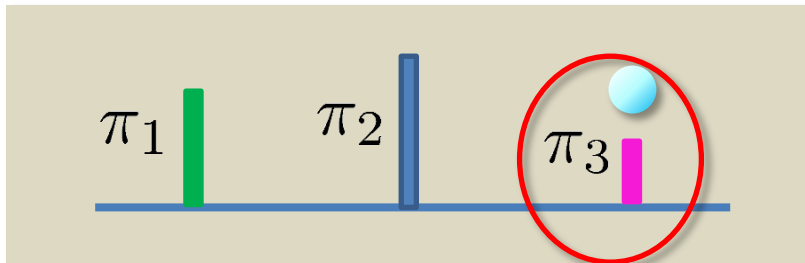
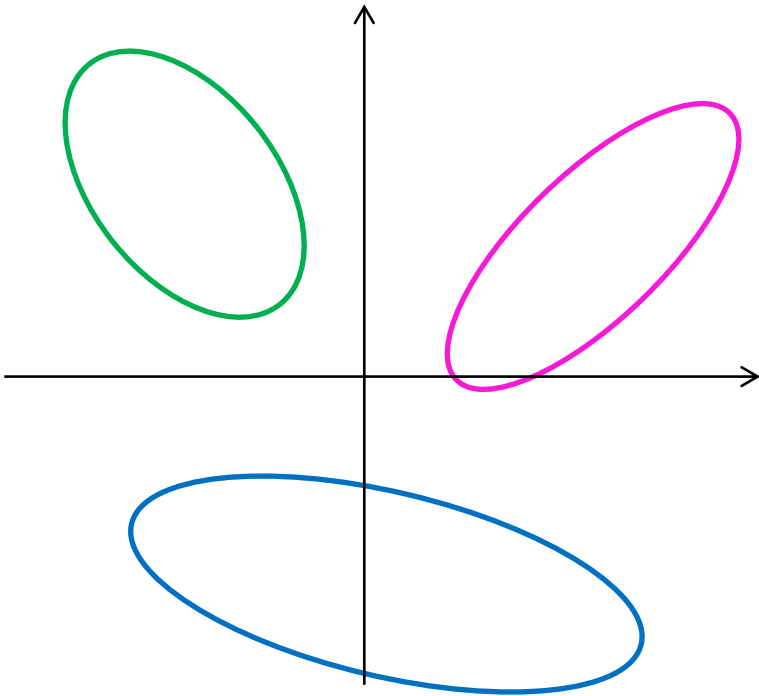
離散分布

$$\underline{\pi} \sim \text{Dirichlet}(\alpha)$$



例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

離散分布

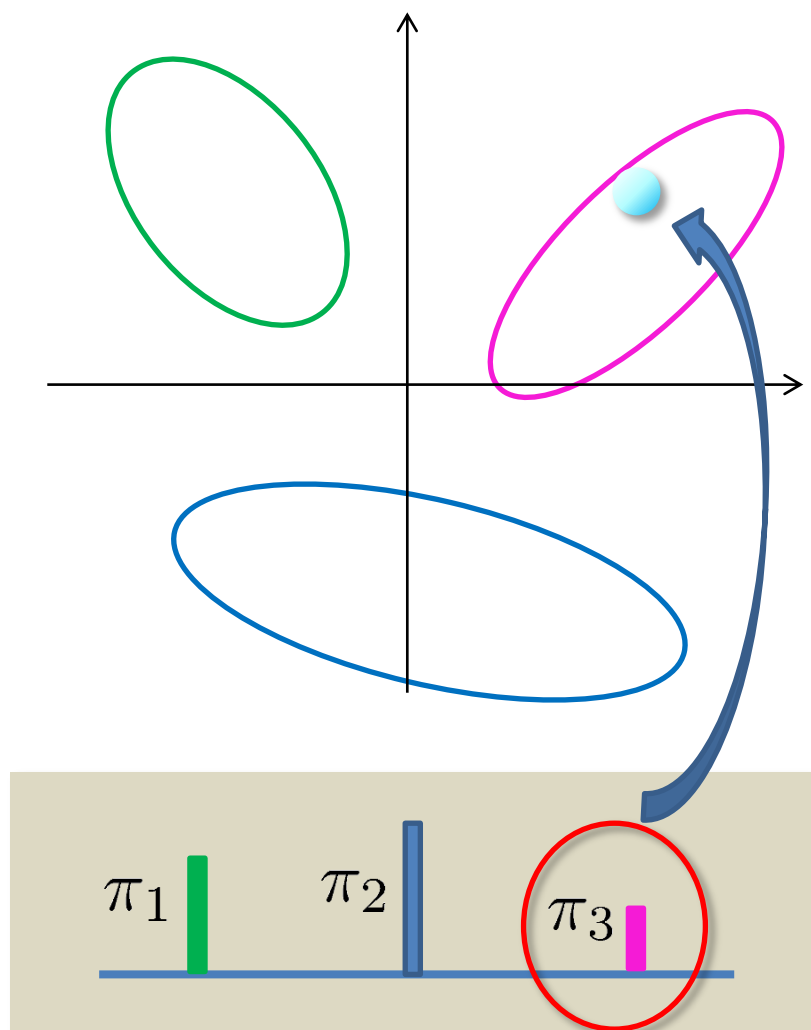
$$\underline{\pi} \sim \text{Dirichlet}(\alpha)$$

クラスタ割り当て

$$\underline{z}_n \sim \text{Discrete}(\pi)$$

例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

離散分布

$$\underline{\pi} \sim \text{Dirichlet}(\alpha)$$

クラスタ割り当て

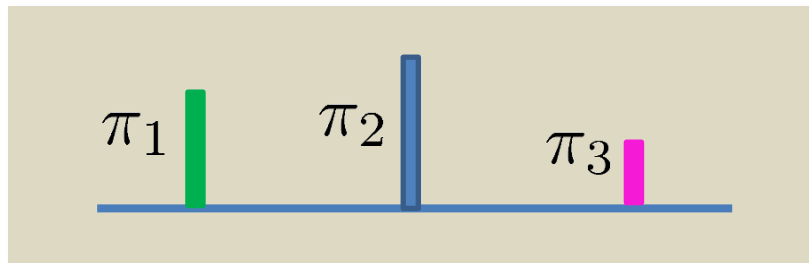
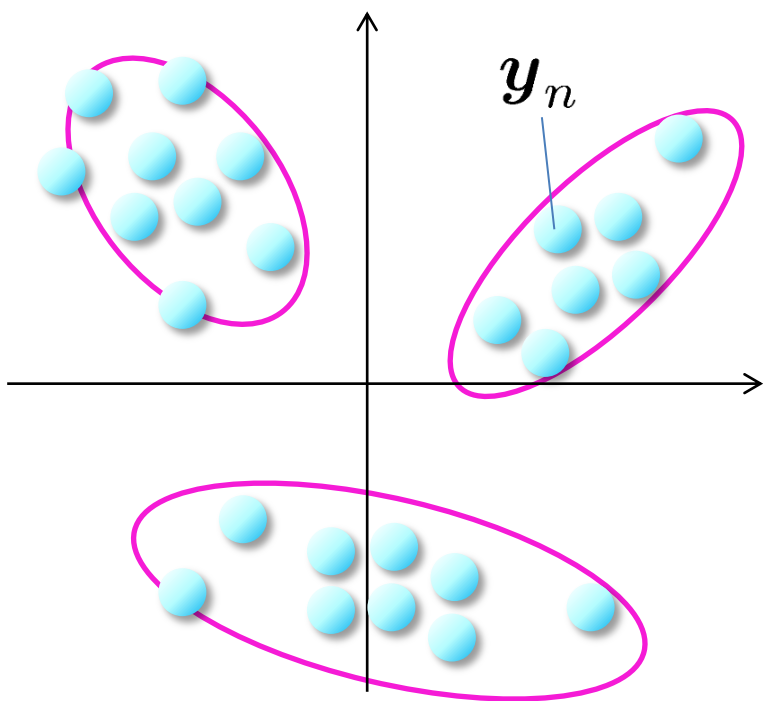
$$\underline{z}_n \sim \text{Discrete}(\pi)$$

データ生成

$$\mathbf{y}_n \sim \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})$$

例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

多項分布

$$\underline{\pi} \sim \text{Dirichlet}(\alpha)$$

クラスタ割り当て

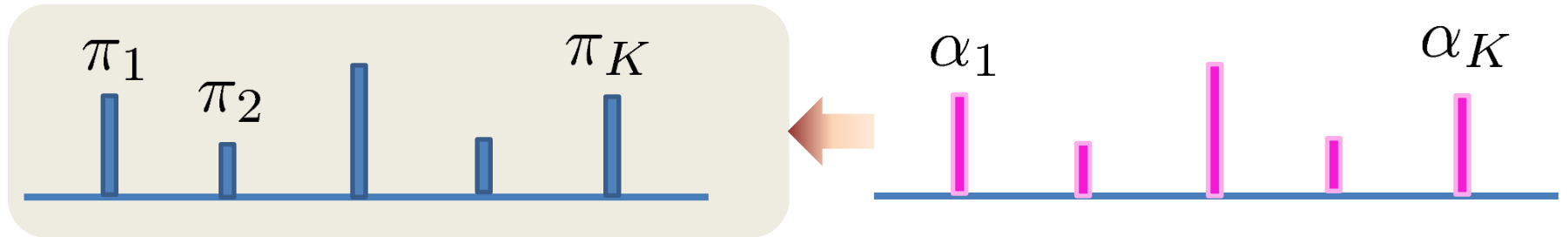
$$\underline{z}_n \sim \text{Discrete}(\pi)$$

データ生成

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

Dirichlet分布

総和が1となるベクトルを生成するための分布



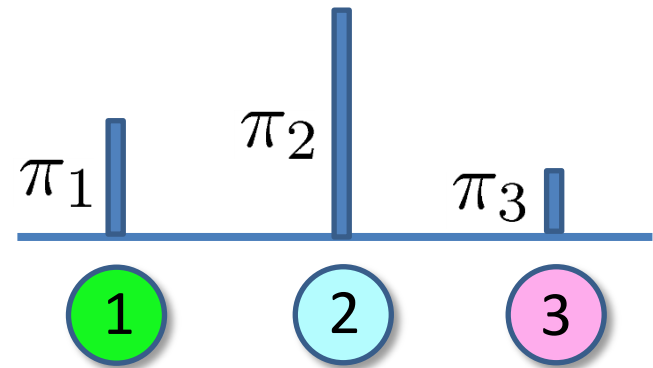
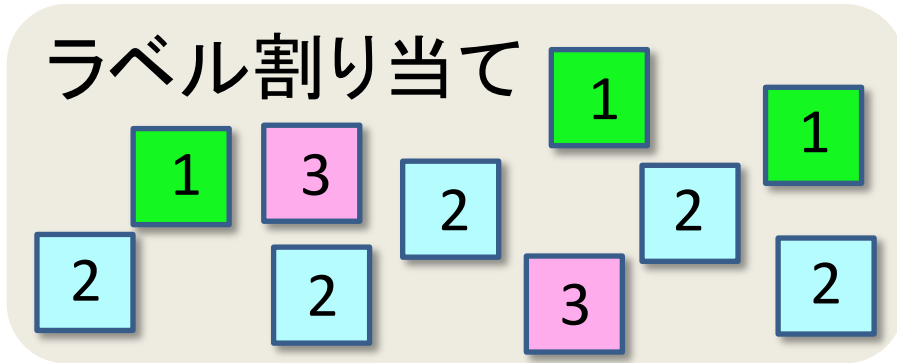
$$\text{Dirichlet}(\pi_1, \dots, \pi_K; \alpha_1, \dots, \alpha_K) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \pi_k^{\alpha_k - 1}$$

Dirichlet変数の作り方1

1. Gamma変数を生成: $\psi_k \sim \text{Gamma}(\alpha_k, 1)$

2. それらを総和が1になるよう正規化: $\pi_k = \frac{\phi_k}{\sum_i \phi_i}$

ディリクレ分布-離散分布(1/2)



ラベル割当確率: $P(z_1, \dots, z_{10} | \pi_1, \pi_2, \pi_3) = \prod_n \prod_k \pi_k^{\mathbb{I}[z_n=k]}$

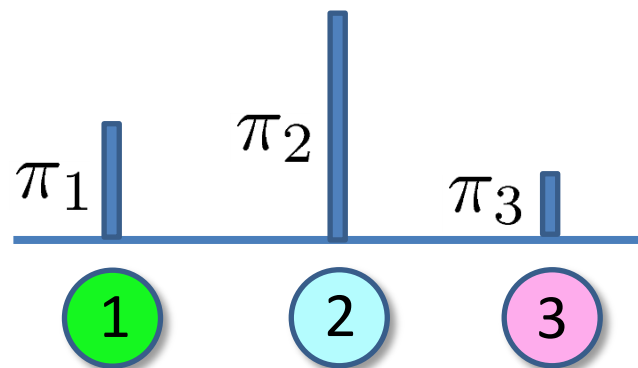
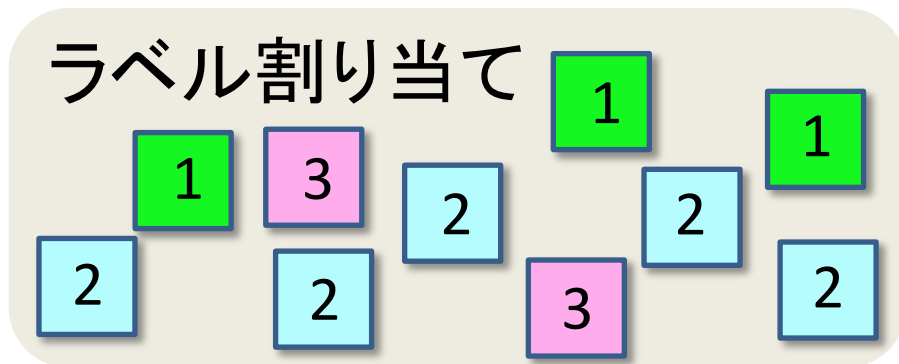
π に関して何の事前知識(事前分布)もなければ

$$p(\pi | z) \propto p(z | \pi) p(\pi)$$

$$= \frac{1}{\text{Normalize}} \prod_k \pi_k^{\mathbb{I}[z_n=k]} = \text{Dirichlet}(N_1 + 1, N_2 + 1, N_3 + 1)$$

① が何回使われたか

ディリクレ分布-離散分布(2/2)



ラベル割当確率: $P(z_1, \dots, z_{10} | \pi_1, \pi_2, \pi_3) = \prod_n \prod_k \pi_k^{\mathbb{I}[z_n=k]}$

$\pi \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)$ の事前分布を考えると

$$p(\pi | z) \propto p(z | \pi) p(\pi)$$

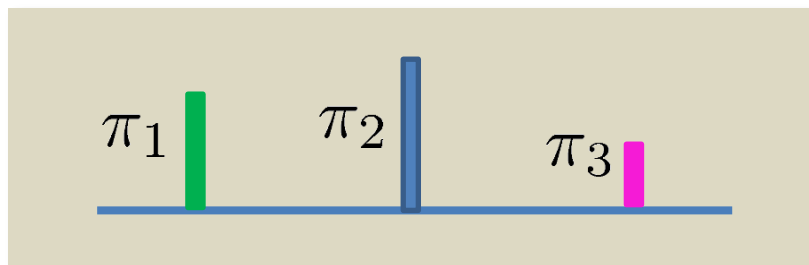
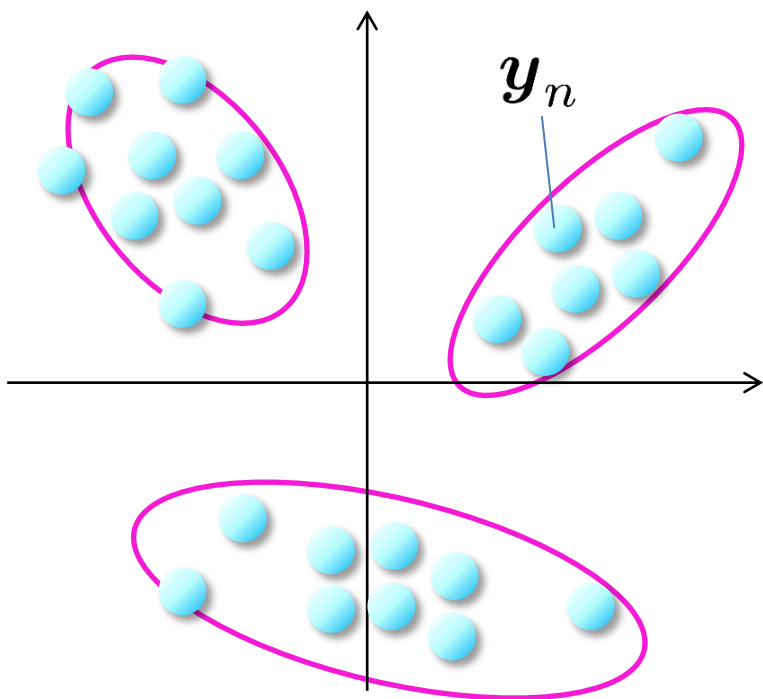
$$= \text{Dirichlet}(N_1 + \alpha_1 + 1, N_2 + \alpha_2 + 1, N_3 + \alpha_3 + 1)$$

① が何回使われたか

事前分布は各ラベルの使用回数へ下駄を履かせたもの！！

例: 3-Gaussian mixture model

課題: 平面上の点のクラスタリング



各ラベルごとのパラメータ

$$\underline{\Sigma}_k \sim \mathcal{W}^{-1}(\Psi, n_0)$$

$$\underline{\mu}_k \sim \mathcal{N}(\mu_0, \Sigma_k/m)$$

多項分布

$$\underline{\pi} \sim \text{Dirichlet}(\alpha)$$

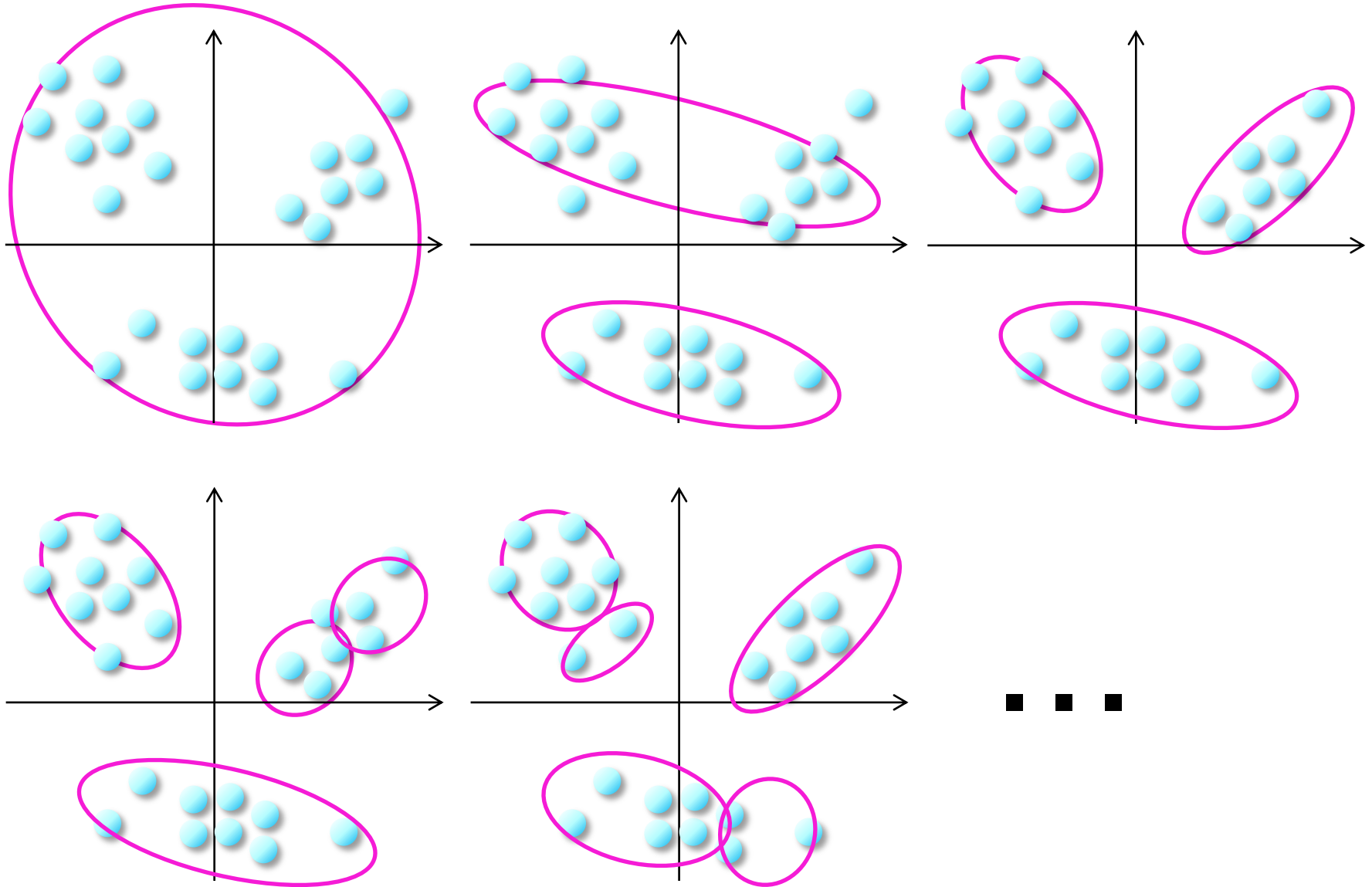
クラスタ割り当て

$$\underline{z}_n \sim \text{Discrete}(\pi)$$

データ生成

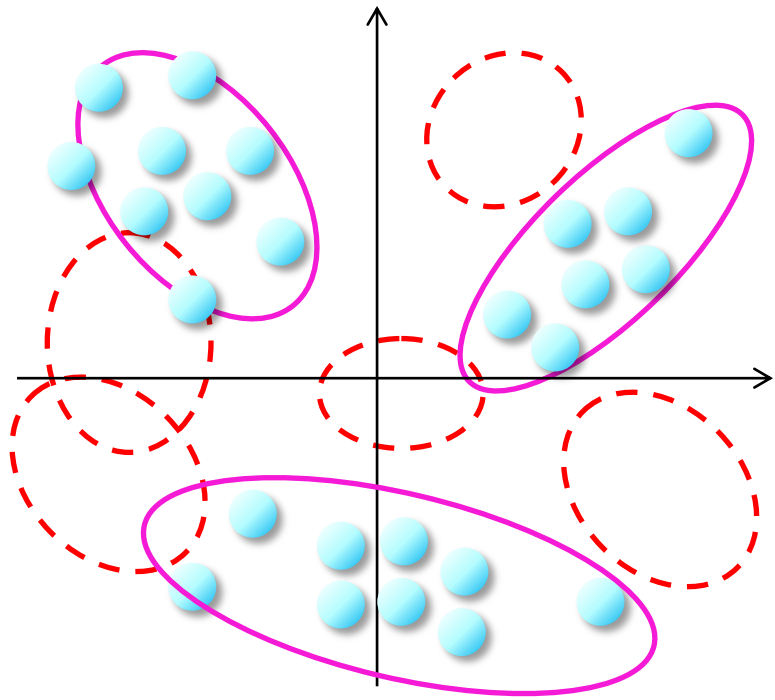
$$\mathbf{y}_n \sim \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})$$

モデルの複雑度の設定:



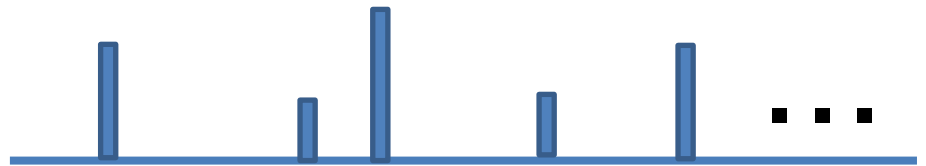
■ ■ ■

モデルの複雑度の設定：無限モデルの導入



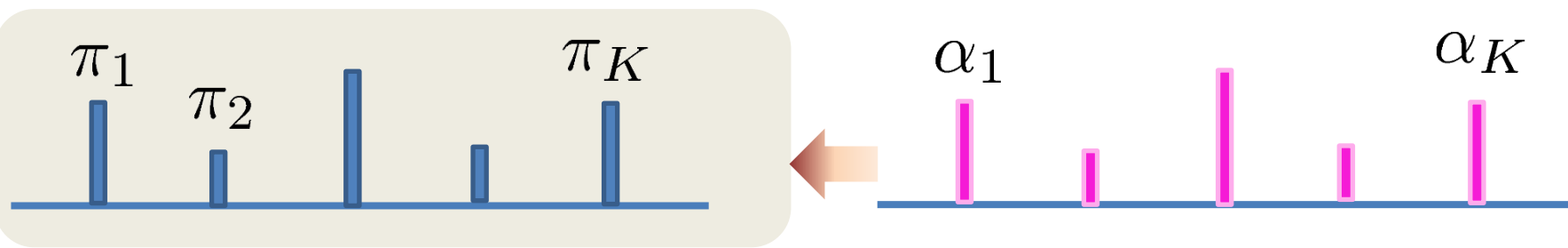
膨大なパラメータのうち
観測データを説明するの
に必要な分だけが
機能するようにしたい！！

- ・ラベルの重み (総和1)
- ・ラベル付随のパラメータ
の可算無限個のペア

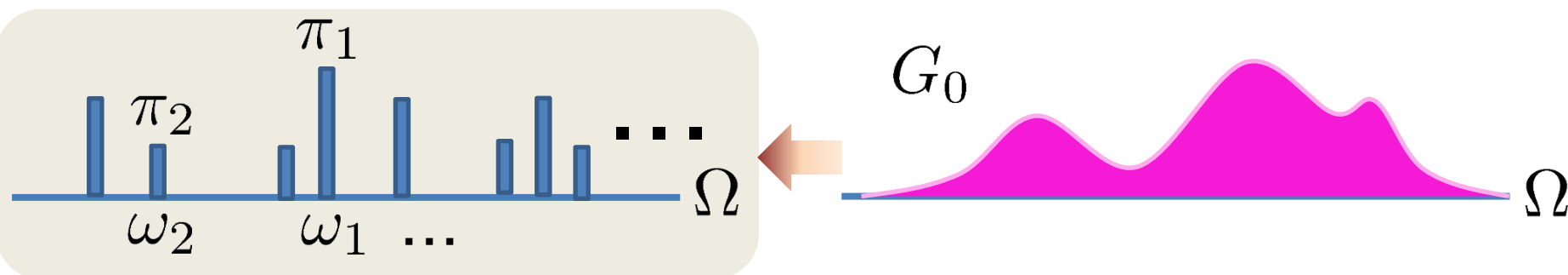


Dirichlet変数とDirichlet過程

Dirichlet変数



Dirichlet過程



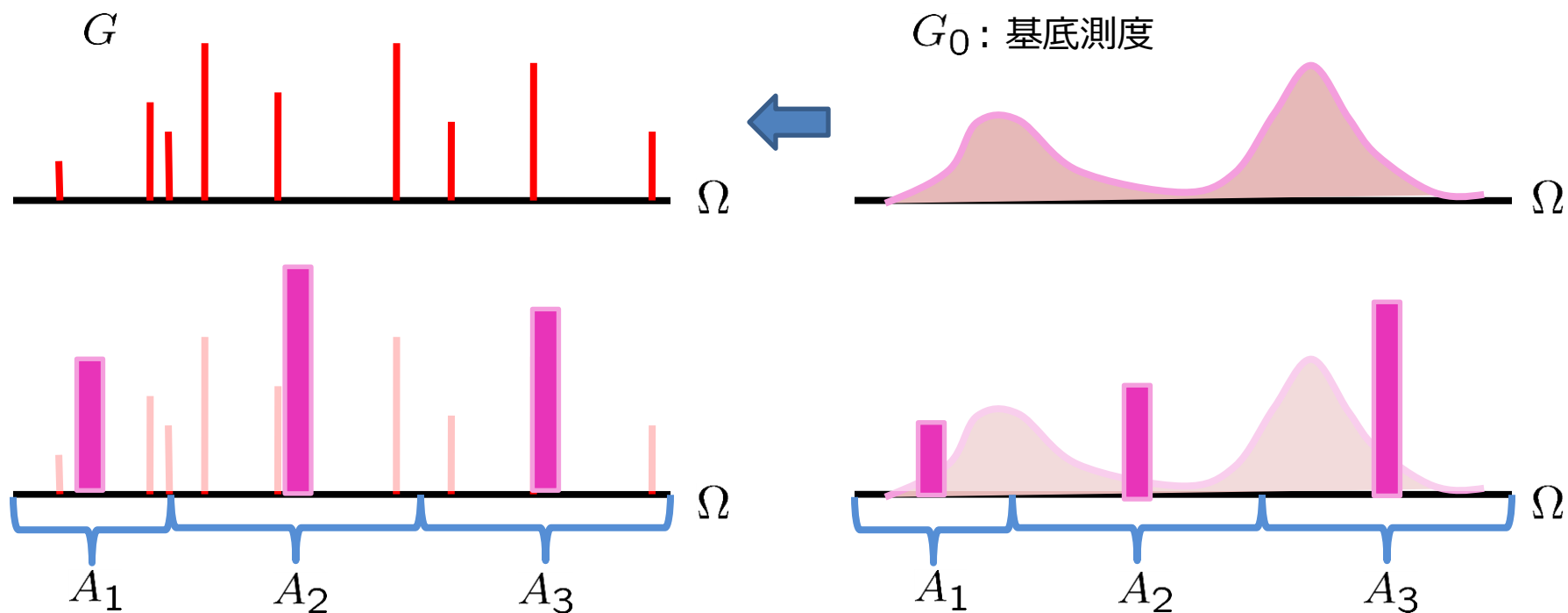
Dirichlet processesの定義

可測空間 (Ω, \mathcal{F}) の基底測度を G_0 、集中度を α とする。確率測度 G が

$$G \sim \text{DP}(\alpha, G_0)$$

に従うとき、任意の disjoint な Ω の分割 (A_1, \dots, A_K) に対して

$$(G(A_1), \dots, G(A_K)) \sim \text{Dirichlet}(\alpha G_0(A_1), \dots, \alpha G_0(A_K))$$



Dirichlet processに関するFAQ

- Dirichlet processはなぜ無限のアトムを持つような確率測度になるのか？
- Dirichlet process混合モデルは「コンポーネント数」を推定することが出来るのか？？

Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]

$$(\pi_1, \dots, \pi_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$$

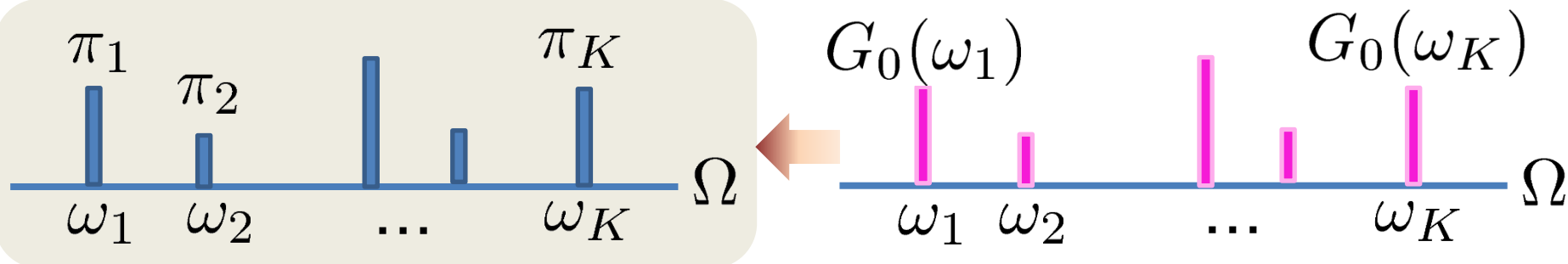
α_k は k 番目のラベルに対する事前の「価値」

パラメータへの事前の価値

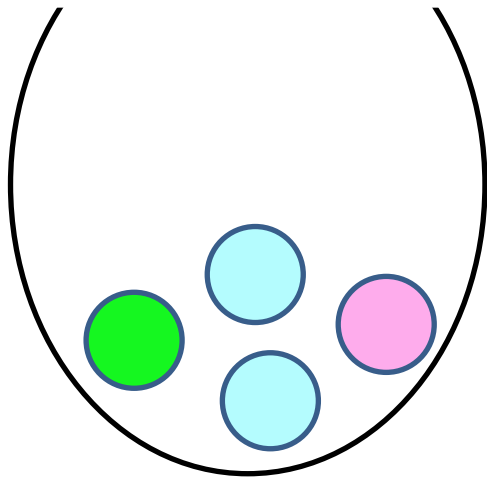
可測空間 (Ω, \mathcal{F}) 上の測度 G_0 に対して

$$(\pi_1, \dots, \pi_K) \sim \text{Dirichlet}(G_0(\omega_1), \dots, G_0(\omega_K))$$


のようなDirichlet変数を作りたい！！





Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]

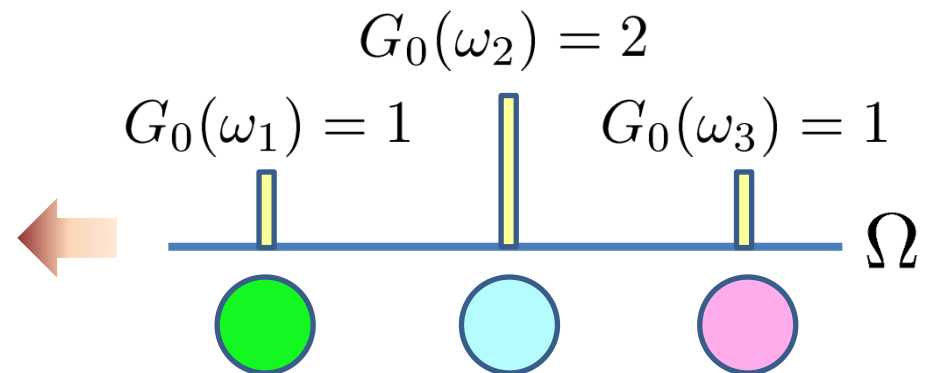
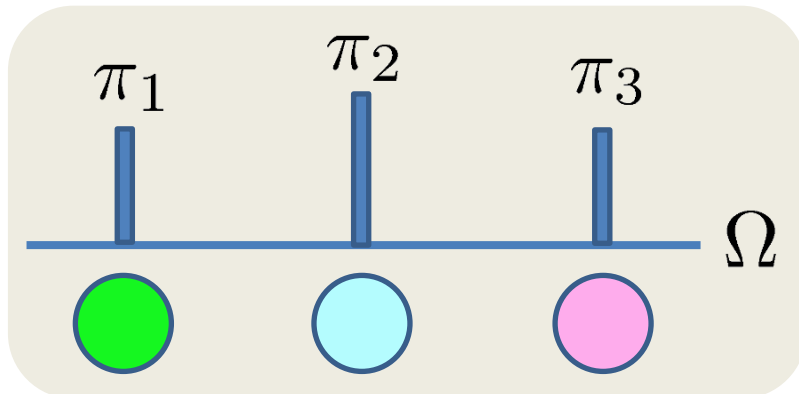


0回の試行

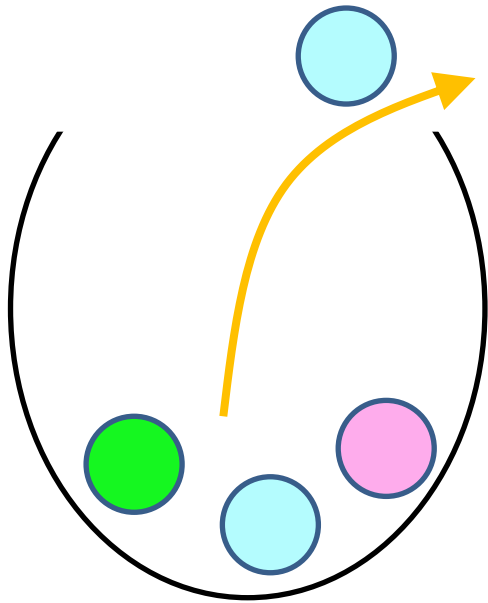
 : $G_0(\omega_1) = 1$

 : $G_0(\omega_2) = 2$

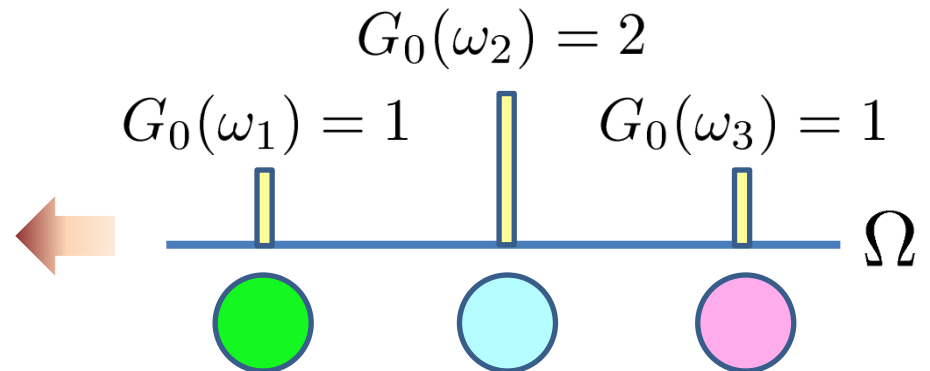
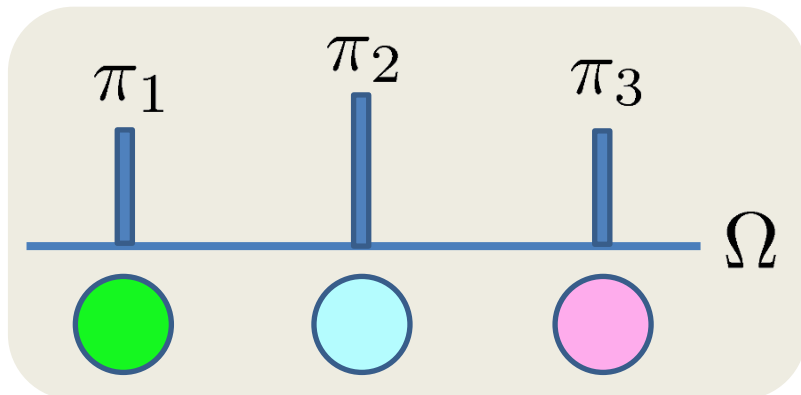
 : $G_0(\omega_3) = 1$



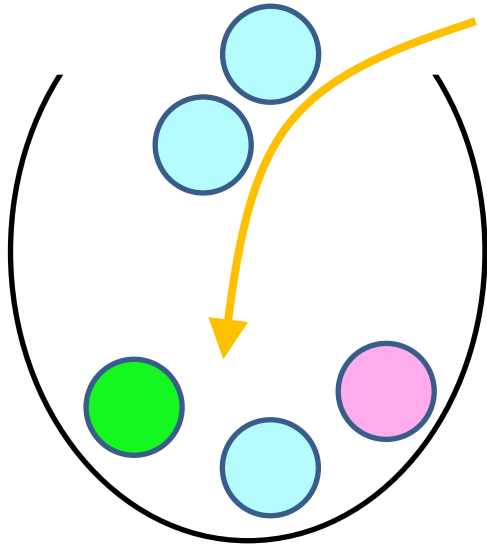
Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]



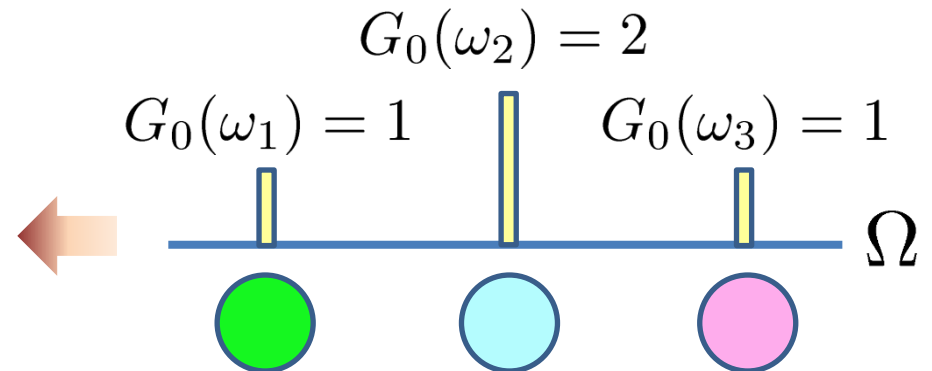
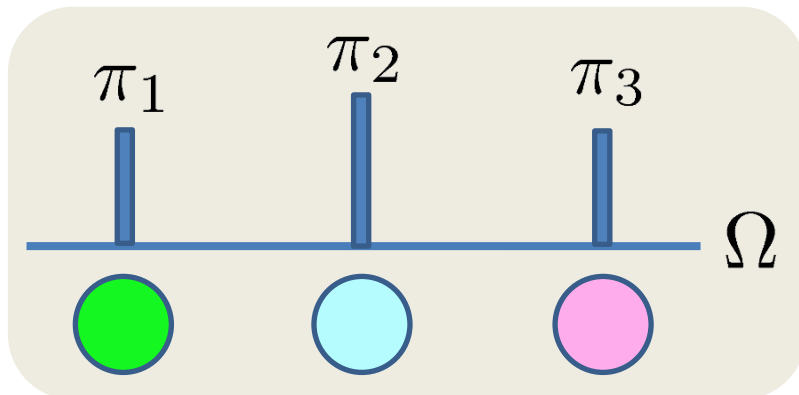
1つのボールを取り出す



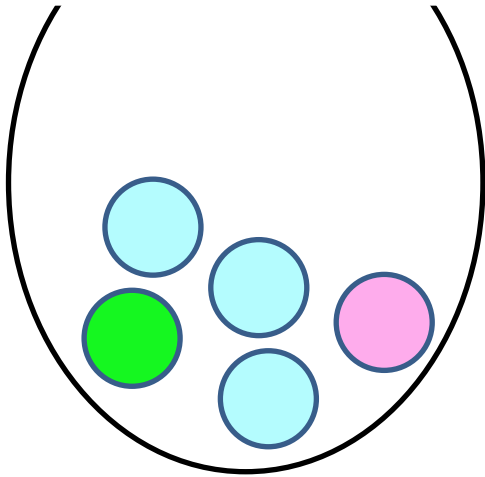
Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]



取り出したボールと同色のボールを加えて2つを戻す



Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]

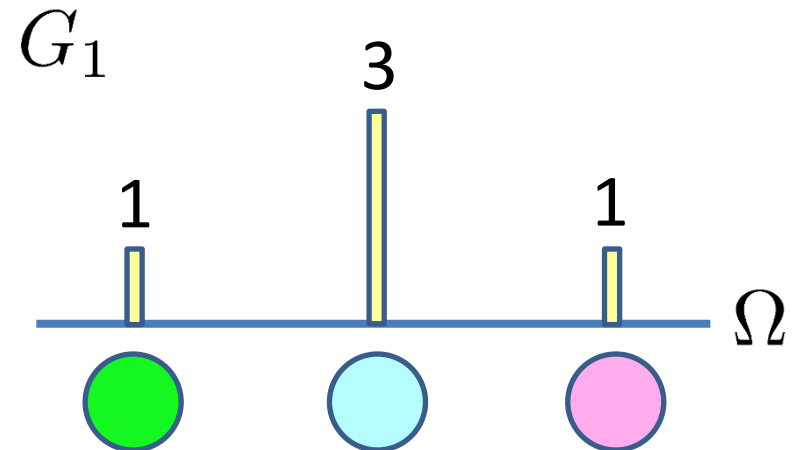
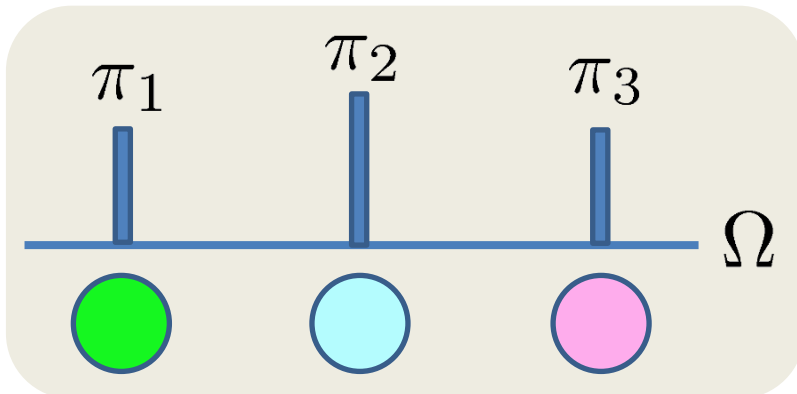


1回の試行

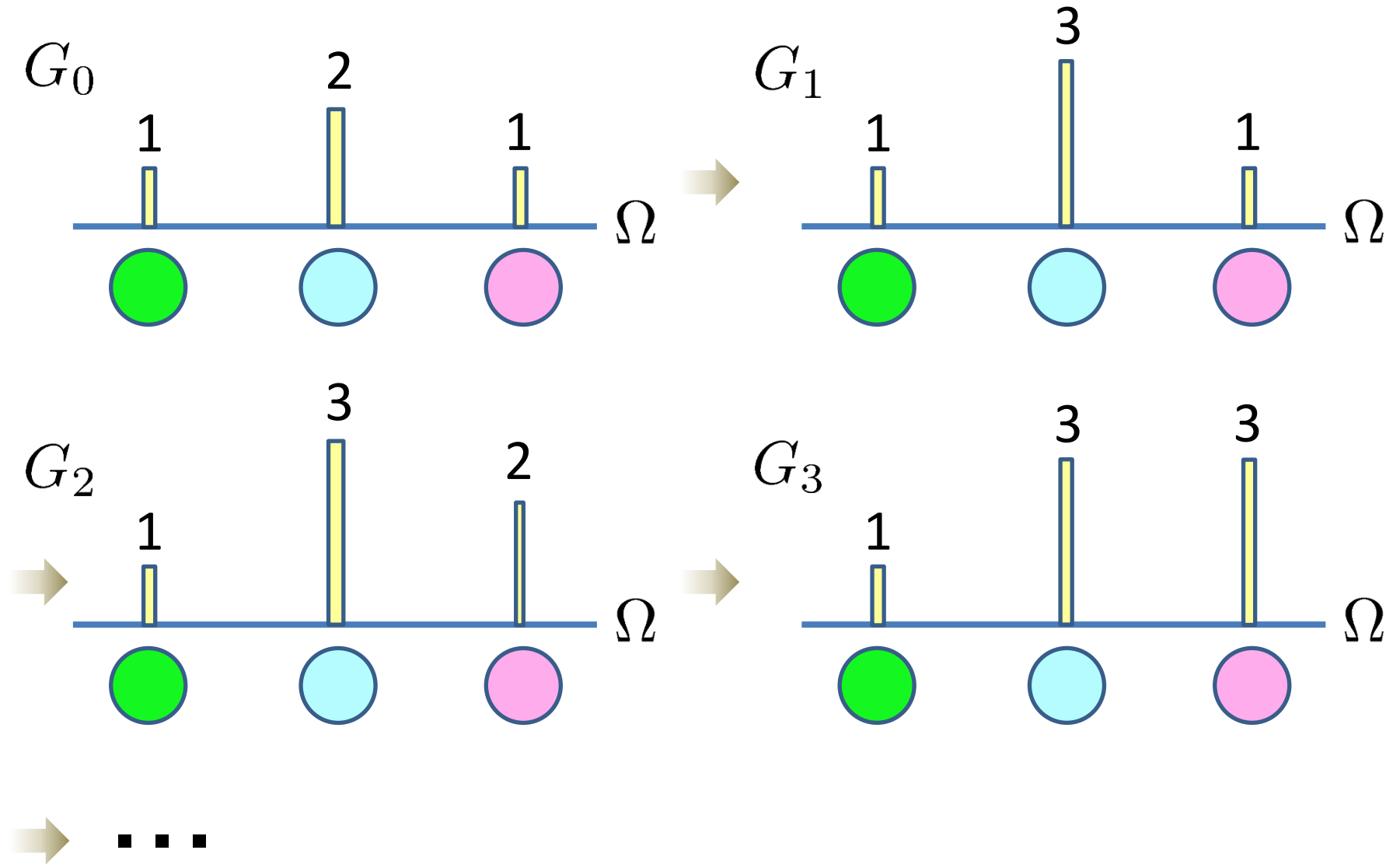
● : $G_1(\omega_1) = 1$

● : $G_1(\omega_2) = 3$

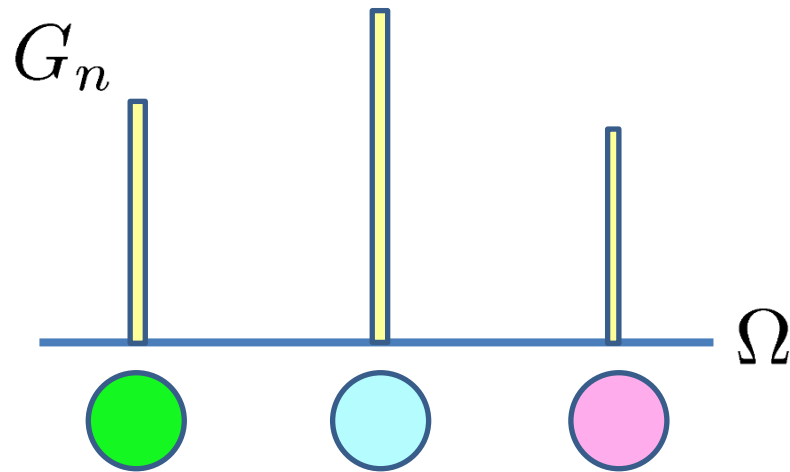
● : $G_1(\omega_3) = 1$



Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]

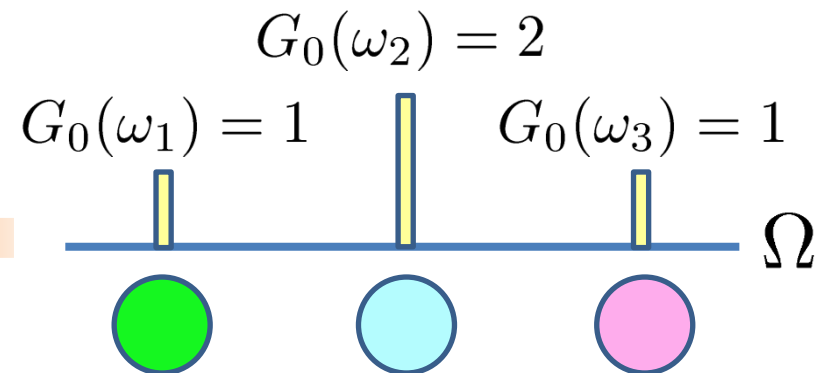
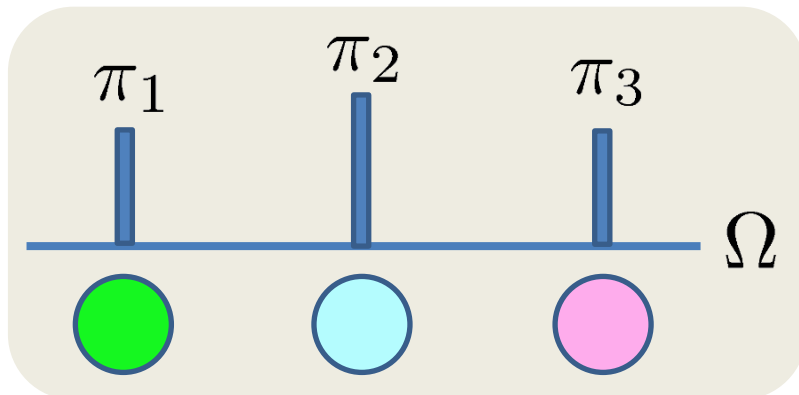


Dirichlet変数の作り方2: ポリアの壺 [Blackwell&Macqueen1973]



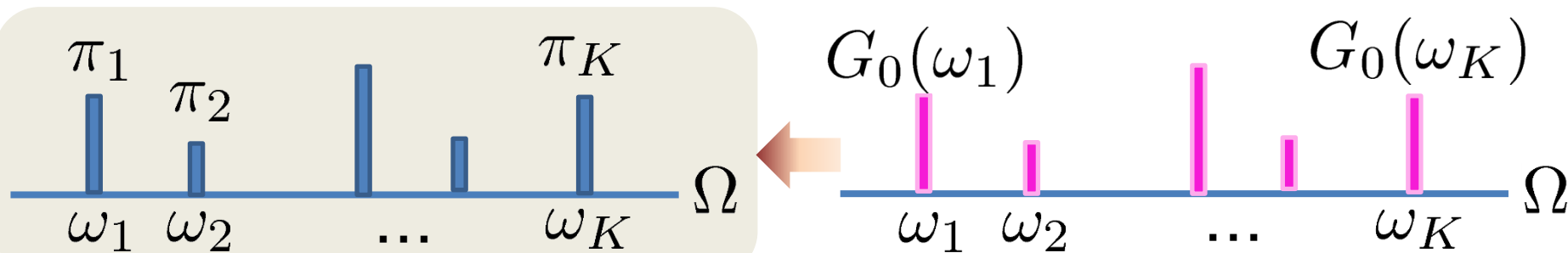
$$G_n^* = \frac{G_n}{G_n(\Omega)}$$

$$G_n^*(\omega_k) \rightarrow \pi_k \quad (n \rightarrow \infty)$$

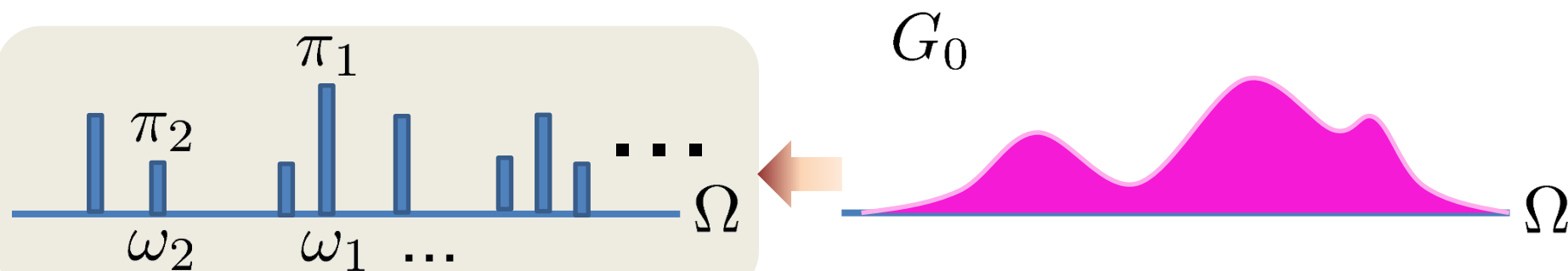


Dirichlet変数とDirichlet過程

Dirichlet変数

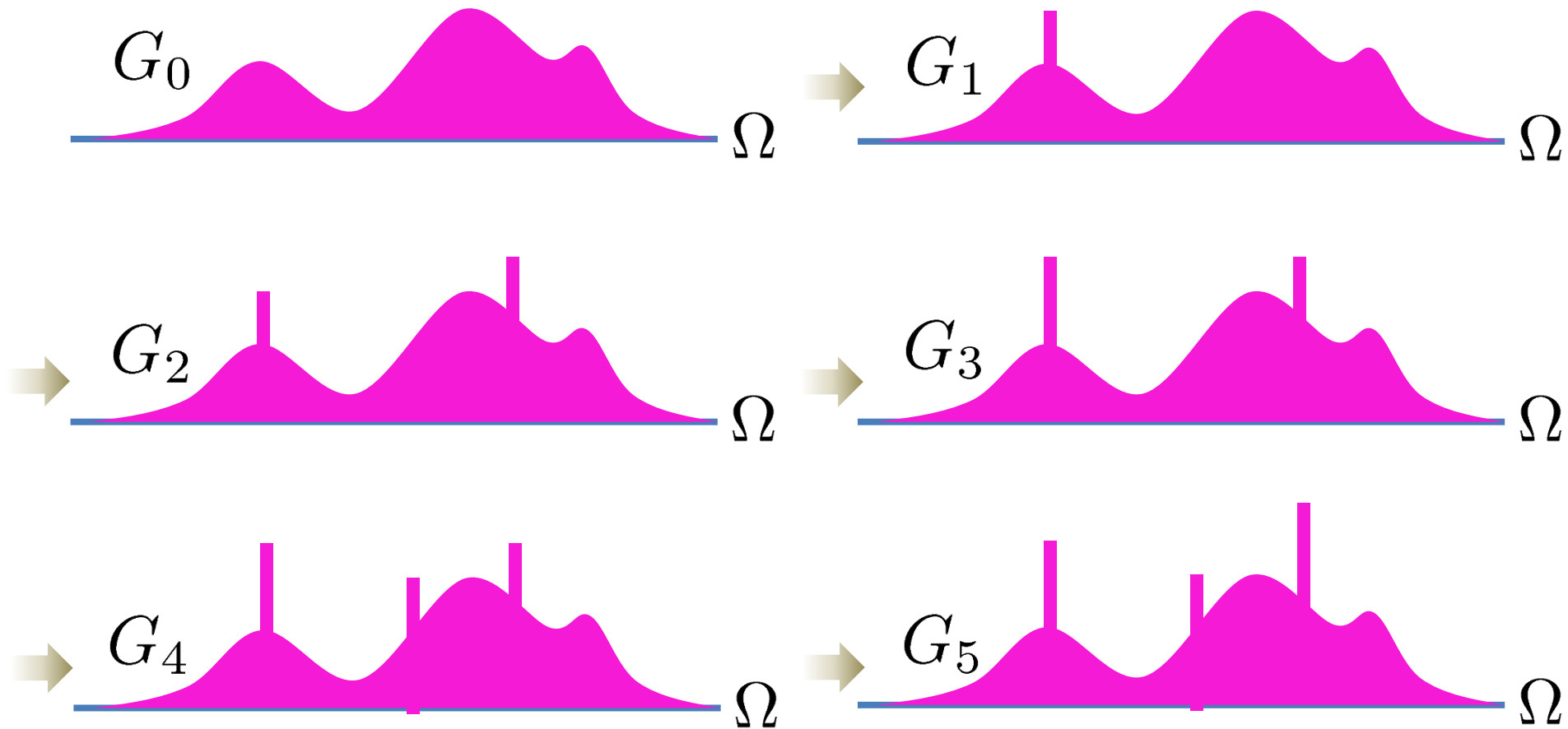


Dirichlet過程



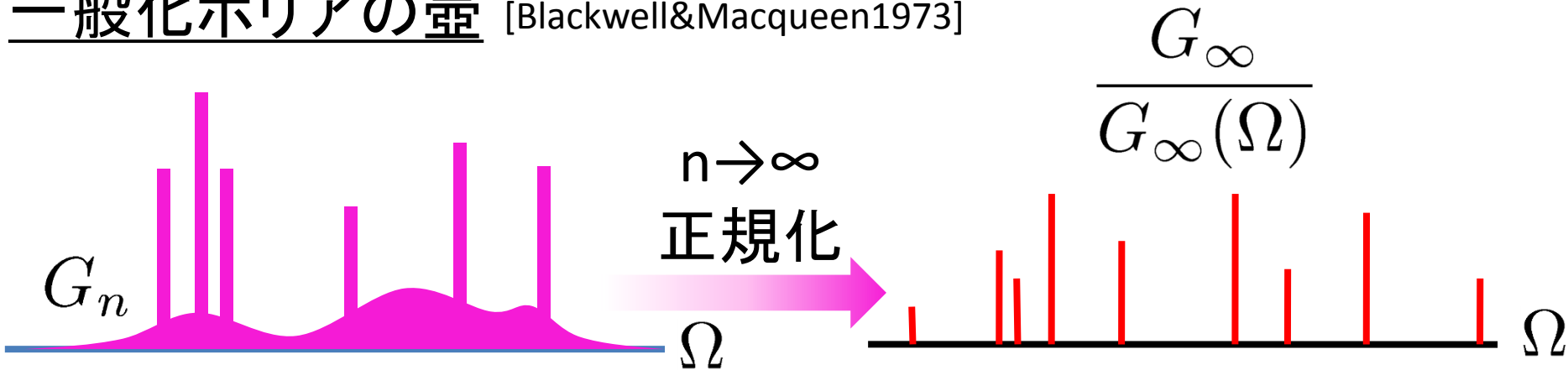
Dirichlet過程が**離散(無限のatom)**的なのは何故？

一般化ポリアの壺 [Blackwell&Macqueen1973]



Dirichlet過程が**離散(無限のatom)**的なのは何故？

一般化ポリアの壺 [Blackwell&Macqueen1973]



[証明スケッチ]

1. **標本空間が有限のポリア列**はディリクレ変数を作る
2. M回目までに登場したatom以外を引いた時にはnullを返すような新しい**有限のポリアの壺**を作る
3. $M \rightarrow \infty$ にしたとき、それらだけで**確率1を占有**する

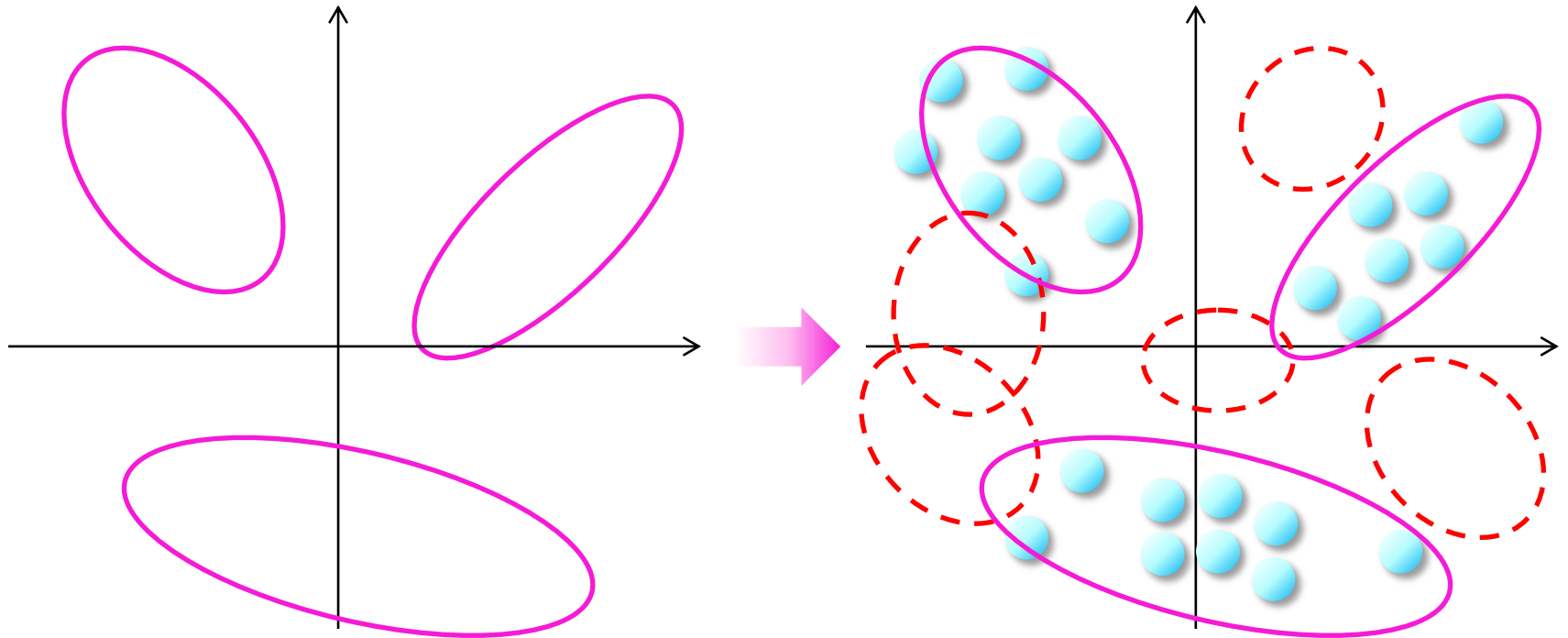
[Blackwell&Macqueen1973]

Dirichlet processに関するFAQ

- Dirichlet processはなぜ無限のアトムを持つような確率測度になるのか？
- Dirichlet process混合モデルは「コンポーネント数」を推定することができるのか？？

Dirichlet過程はコンポーネント数を推定可能か？

[Miller2012]

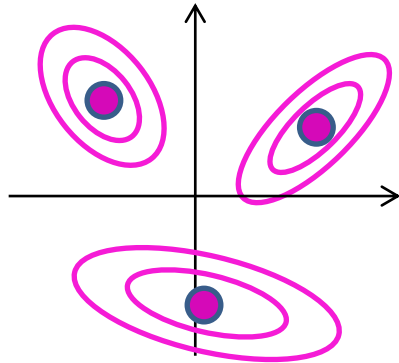


3-componentのGMMから人工的に生成したデータにDirichlet process mixture modelを適用すると、「3」componentになるか??

Dirichlet過程はコンポーネント数を推定可能か？

[Miller2012]

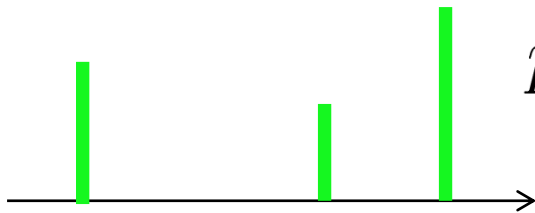
1. 密度関数



$$p_{\text{model}}(\mathcal{D}(f_{\text{true}}, f_{\text{model}}) < \epsilon) \rightarrow 1$$

(観測データ数 $\rightarrow \infty$)

2. 混合



$$p_{\text{model}}(\mathcal{D}(\beta_{\text{true}}, \beta_{\text{model}}) < \epsilon) \rightarrow 1$$

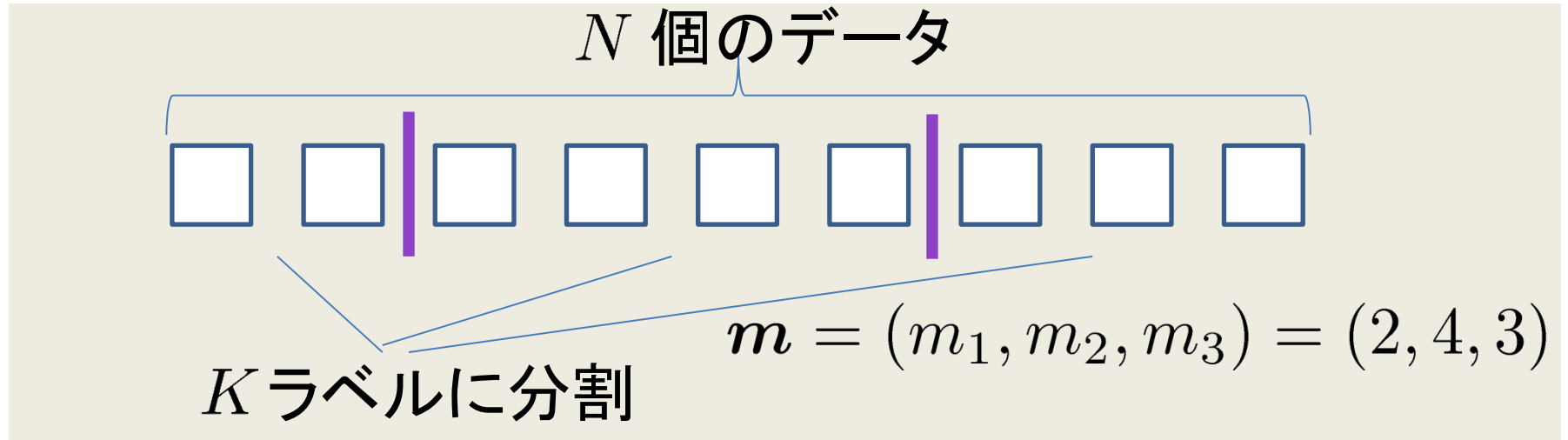
(観測データ数 $\rightarrow \infty$)

3. コンポーネント数

「3」



Dirichlet過程はコンポーネント数を推定可能か？



$$P_{\text{DPM}}(\mathbf{m} \mid K) \propto \frac{1}{m_1 m_2 \cdots m_K} \left(\sum_k m_k = N \right)$$

例) 100個のデータを3クラスタに分割

$$\frac{1}{1 * 1 * 98} > \frac{1}{25 * 25 * 50}$$

Dirichlet過程混合モデルの構成法と推論

構成法

- Chinese restaurant process [Aldous1981]
- Stick-breaking process [Sethuraman1994]
- Normalized Gamma process [Ferguson1973]
- Poisson process (beta intensity) [Roy&Teh2009]

推論法

- Gibbs sampler
- VB, collapsed VB [多数]
- Slice sampler [Walker2007]
- Retrospective sampler [Papaspiliopoulos2008]

Stick-breaking construction

Stick-breaking construction

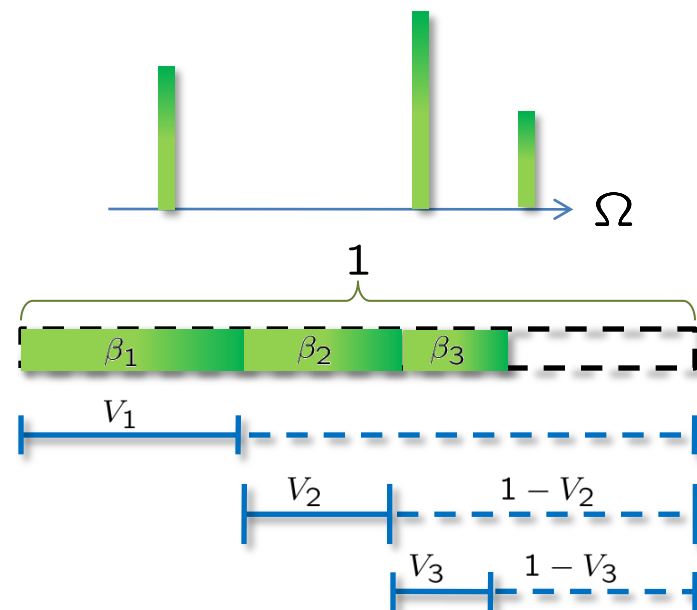
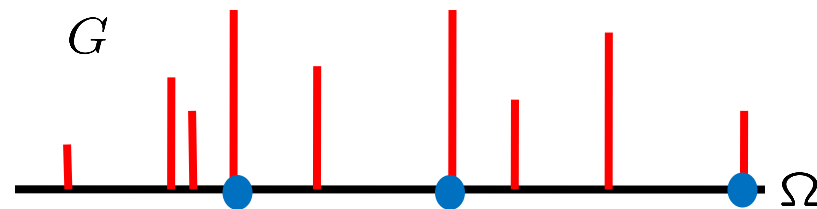
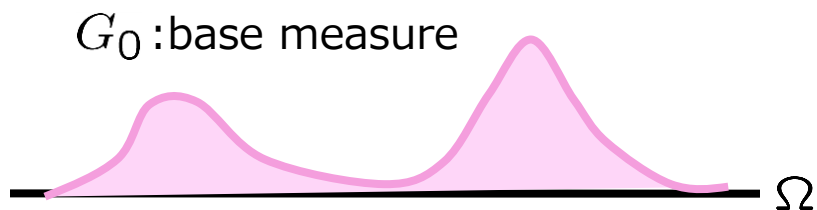
確率測度 $G \sim \text{DP}(\alpha, G_0)$ は次のように構成できる:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

$$\phi_k \sim G_0$$

$$V_k \sim \text{Beta}(1, \alpha)$$

$$\pi_k = V_k \prod_{i=1}^{k-1} (1 - V_i)$$



Stick-breaking constructionの略証 [Paisely2008]

補題1

確率変数 W_1, W_2, V が

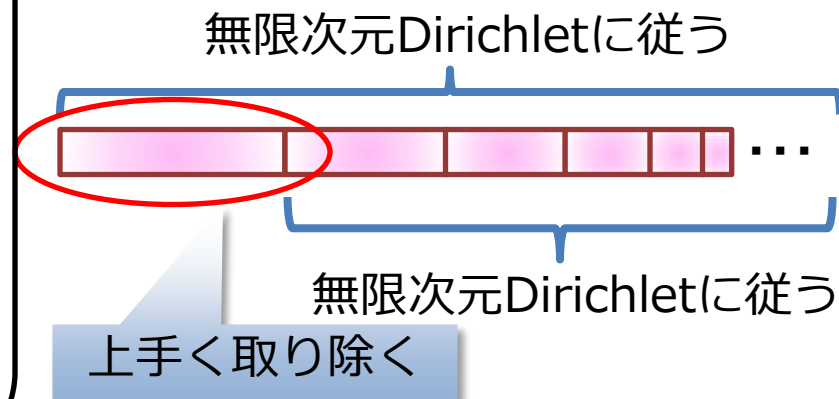
$$W_1 \sim \text{Dir}(w_1, w_2 \dots)$$

$$W_2 \sim \text{Dir}(u_1, u_2 \dots)$$

$$V \sim \text{Beta}(\sum_k w_k, \sum_k u_k)$$

に従うとき、 $X = VW_1 + (1 - V)W_2$ は

$$X \sim \text{Dir}(w_1 + u_1, w_2 + u_2 \dots)$$



正規化Gamma processを使って書き直すと簡単に確認出来る！

$$W_1 = \left(\frac{\gamma_1}{\sum_k \gamma_k}, \dots \right) \quad W_2 = \left(\frac{\gamma'_1}{\sum_k \gamma'_k}, \dots \right) \quad V = \frac{\sum_k \gamma_k}{\sum_k \gamma_k + \sum_k \gamma'_k}$$

Stick-breaking construction

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

$$V_k \sim \text{Beta}(1, \alpha)$$

$$\pi_k = V_k \sum_{i=1}^{k-1} (1 - V_i)$$

$$\phi_k \sim G_0$$

Stick-breaking constructionの略証 [Paisely2008]

補題2

確率変数 $\pi \sim \text{Dir}(\alpha g_0)$ は

$$Z \sim \text{Mult}(\{1, 2, \dots\}, g_0)$$

$$\pi \sim \text{Dir}(\alpha g_0 + (0, \dots, 0, 1, 0, \dots, 0)^T)$$

から作れる。

Z番目だけ1

→ ダミーデータを一個先に作る

→ ダミーデータ一個分を Dirichlet分布のパラメータに加えることができる！

補題1が使える！！

※ $\sum_k g_{0,k} = 1$ として一般性を失わない

Stick-breaking construction

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

$$V_k \sim \text{Beta}(1, \alpha)$$

$$\pi_k = V_k \sum_{i=1}^{k-1} (1 - V_i)$$

$$\phi_k \sim G_0$$

Stick-breaking constructionの略証 [Paisely2008]

補題1

確率変数 W_1, W_2, V が

$$W_1 \sim \text{Dir}(w_1, w_2 \dots)$$

$$W_2 \sim \text{Dir}(u_1, u_2 \dots)$$

$$V \sim \text{Beta}(\sum_k w_k, \sum_k u_k)$$

に従うとき、 $X = VW_1 + (1 - V)W_2$ は
 $X \sim \text{Dir}(w_1 + u_1, w_2 + u_2 \dots)$

$$W \sim \text{Dir}((0, \dots, 0, 1, 0, \dots, 0))$$

$$\pi' \sim \text{Dir}(\alpha g_0)$$

$$V \sim \text{Beta}(1, \alpha)$$

$$\pi = V \mathbf{W} + (1 - V)\pi'$$

Z番目のatomが自動的に採用される

補題2

確率変数 $\pi \sim \text{Dir}(\alpha g_0)$ は

$$Z \sim \text{Mult}(\{1, 2, \dots\}, g_0)$$

$$\pi \sim \text{Dir}(\alpha g_0 + \underbrace{(0, \dots, 0, 1, 0, \dots, 0)^T}_{Z\text{番目だけ1}})$$

から作れる。

Beta変数による可算無限回の棒折りでDPMが作れる

メディア処理頻出のツール

主成分分析

独立成分分析

非負値行列分解

隠れマルコフモデル

確率文脈自由文法

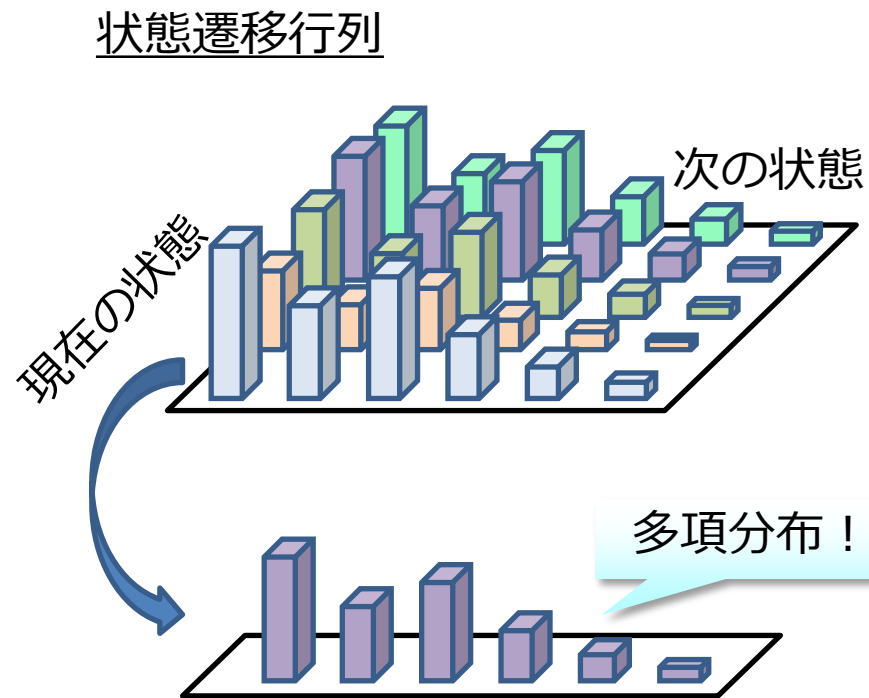
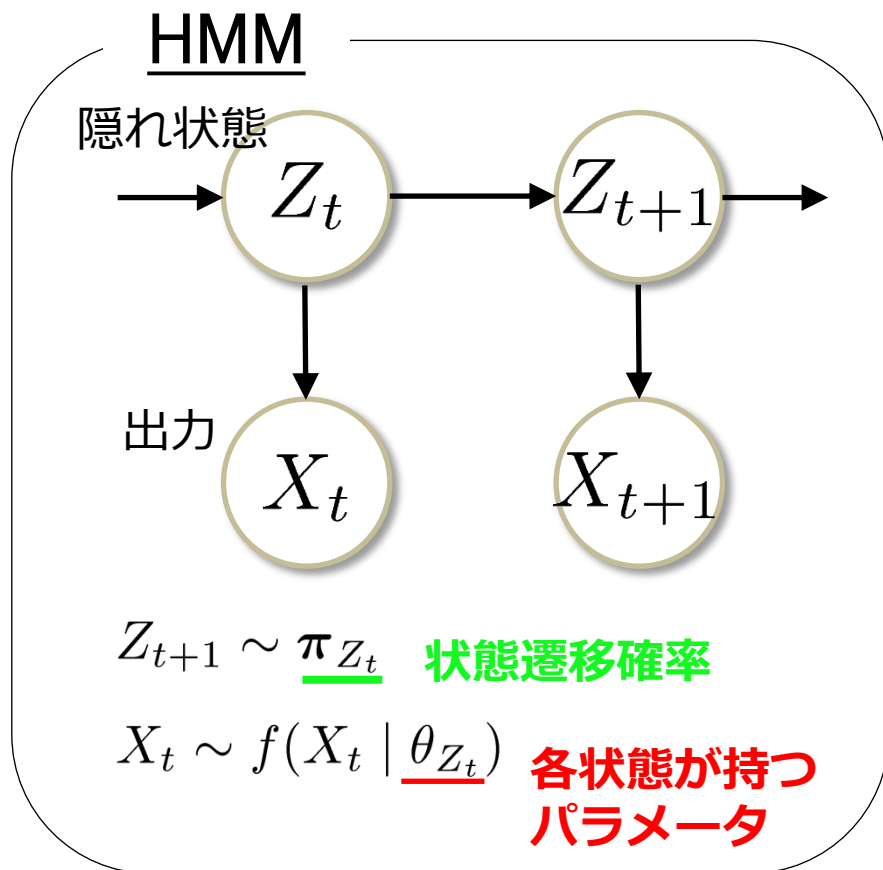
n -gram

ダイナミックベイジアンネット

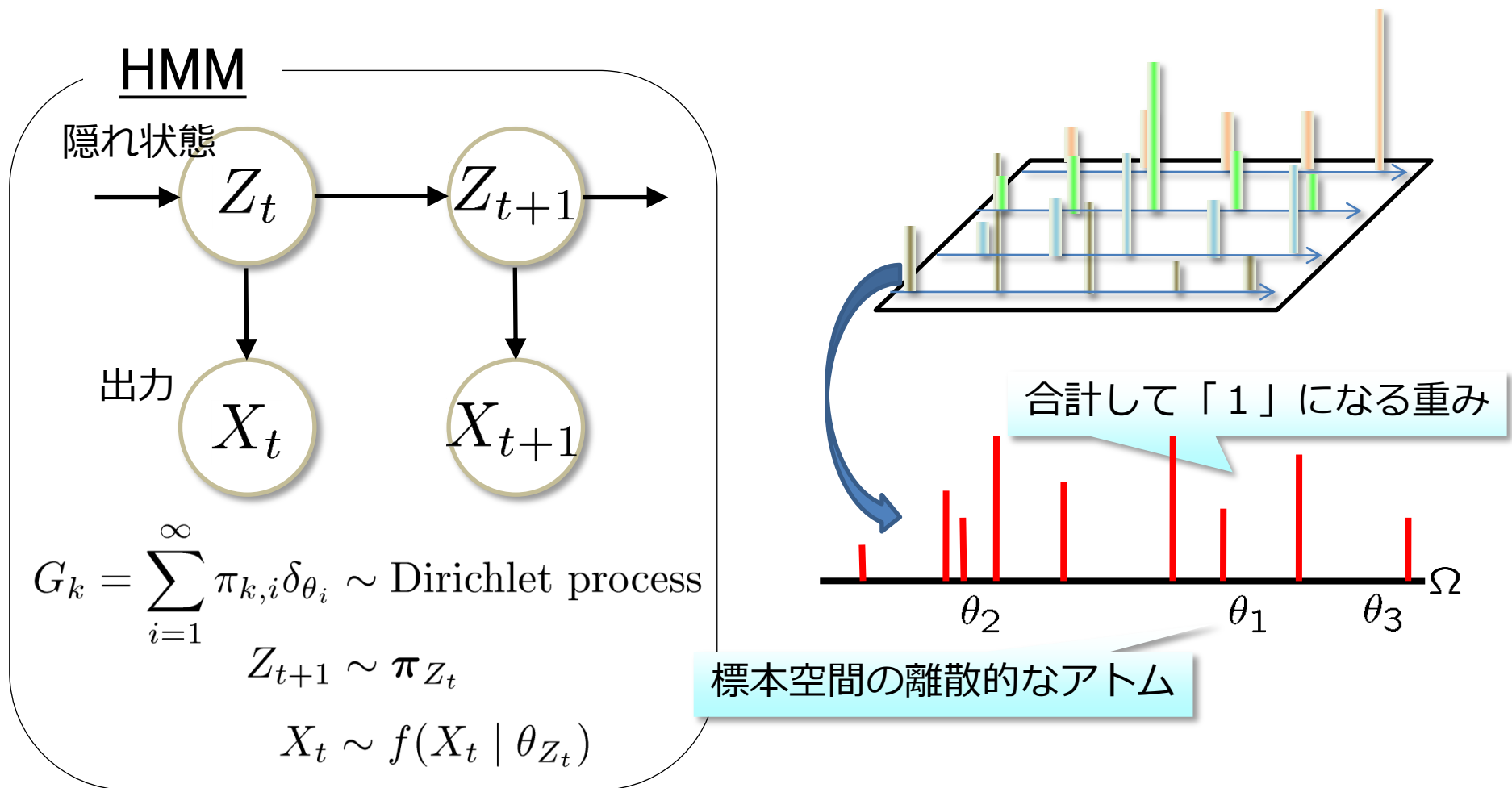
ディープニューラルネット

...

Hidden Markov models



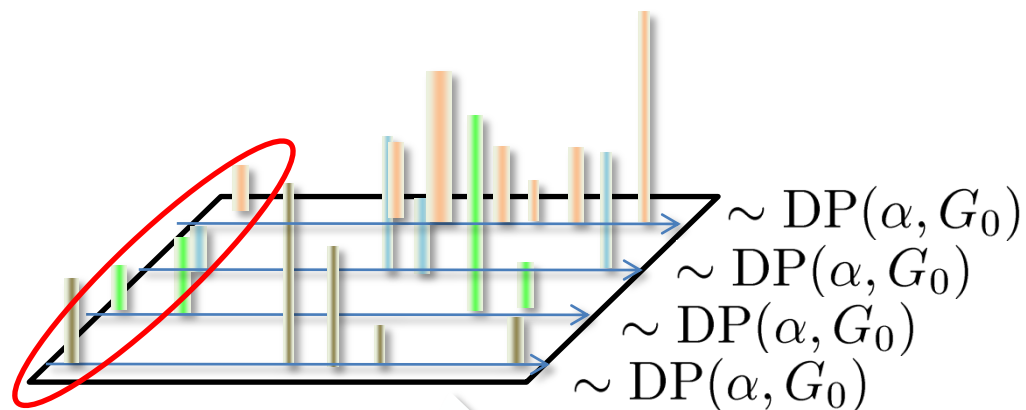
Hierarchical Dirichlet process HMM [Teh+2006]



Dirichlet processを用いた無限混合モデルを用いる！

Hierarchical Dirichlet process HMM [Teh+2006]

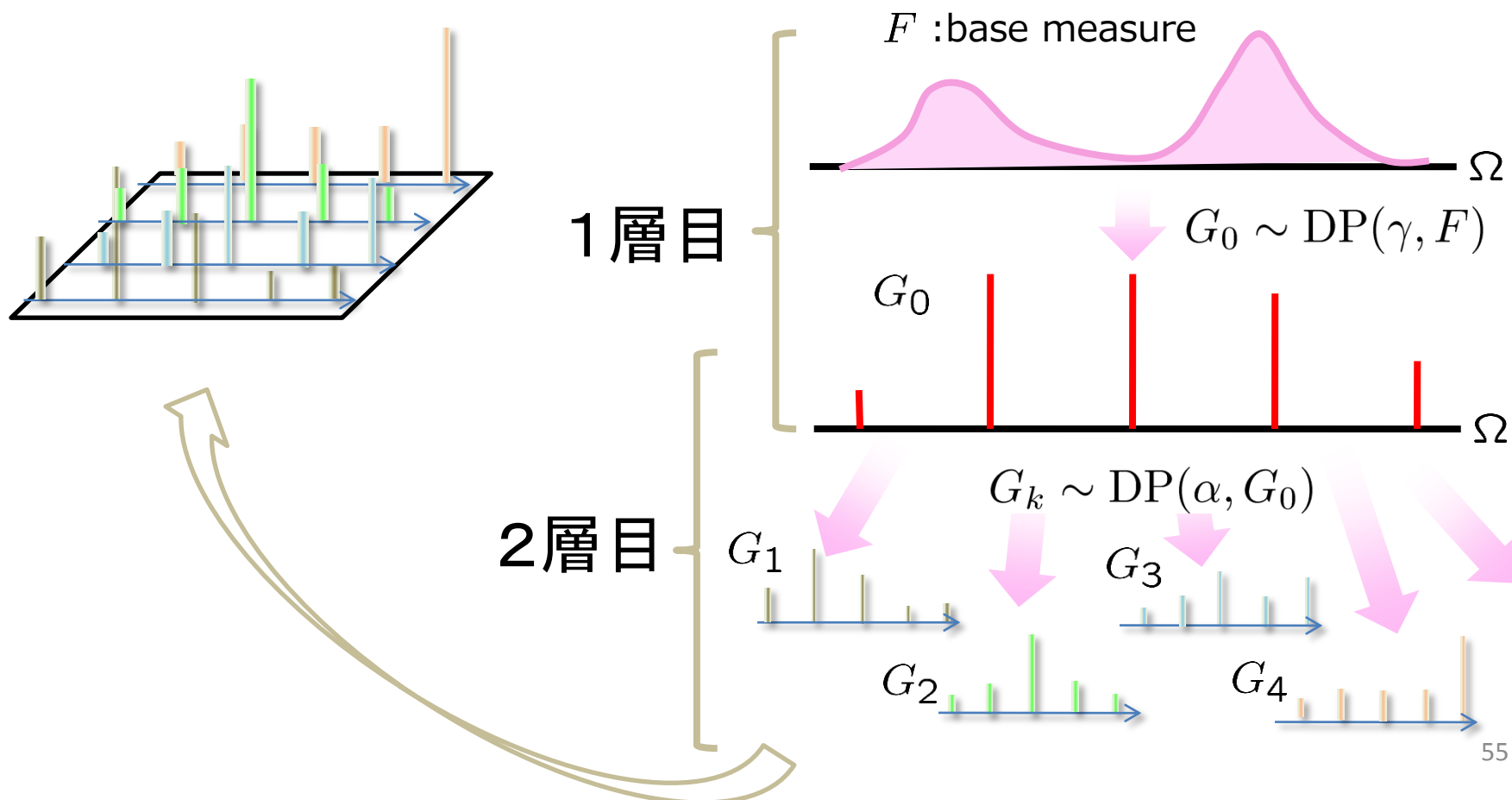
何故 Dirichlet process を階層化する必要があるのか？



アトム達を陽に共有させたい！

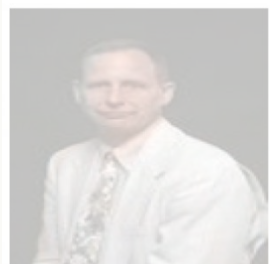
**アトムを陽に共有化するために、
Dirichlet processを階層化する！！！！**

アトムを陽に共有化するために、
Dirichlet processを階層化する！！！！

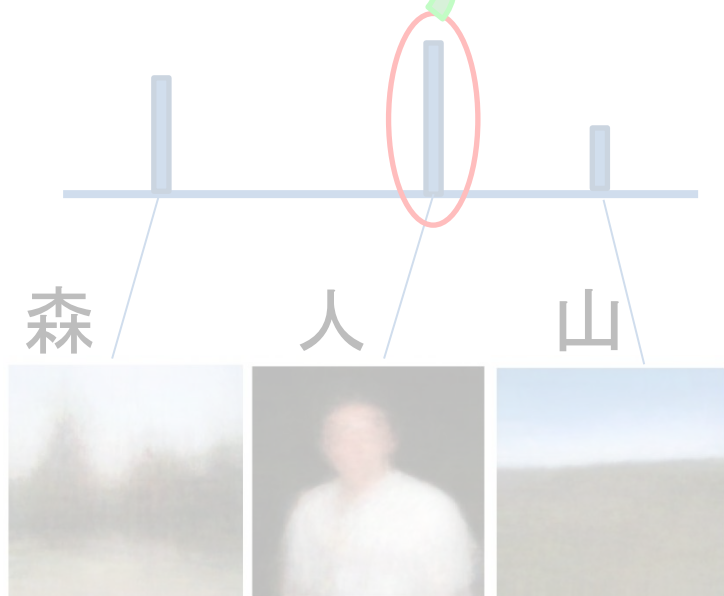


二つの基本的な確率的生成モデル

混合モデル

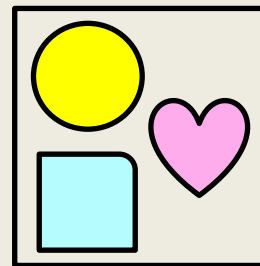


単一のクラスター
がデータを生成

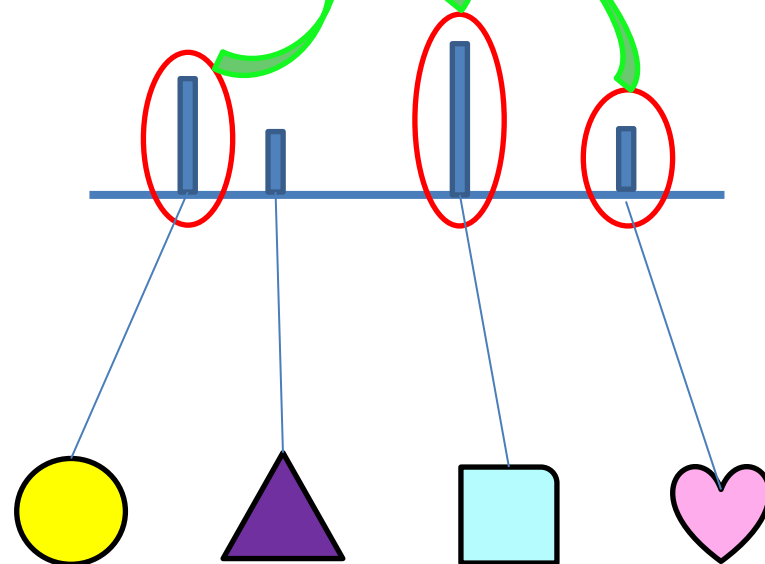


[Torralba+2008]

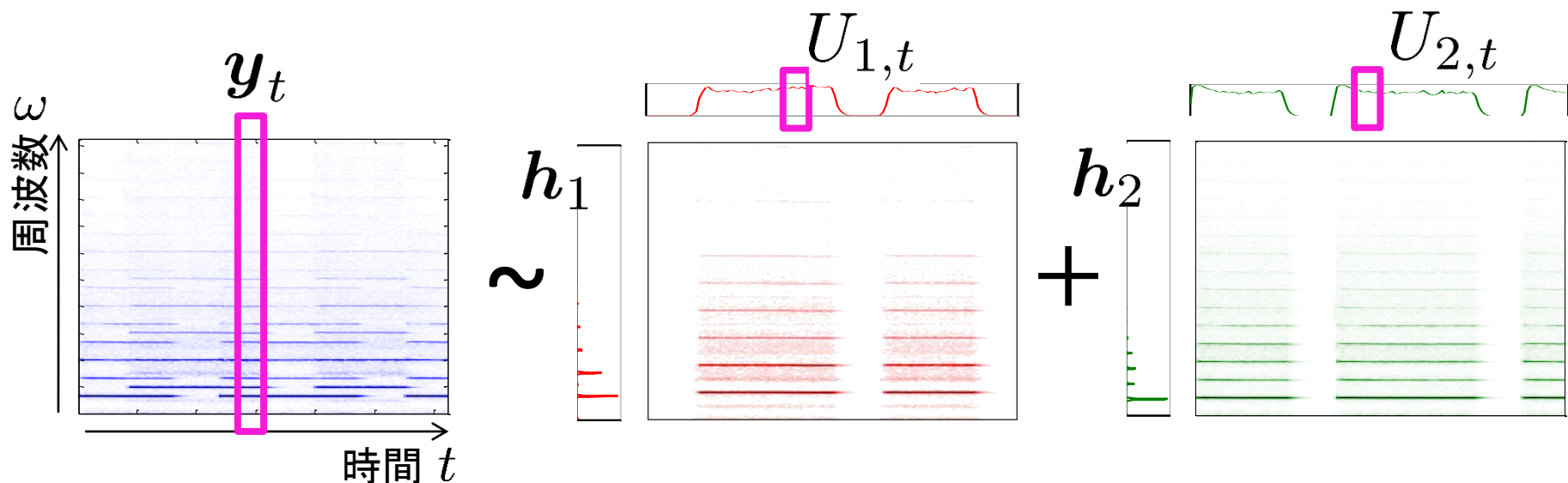
因子モデル



複数の因子が
データを生成



例: nonnegative matrix factorization



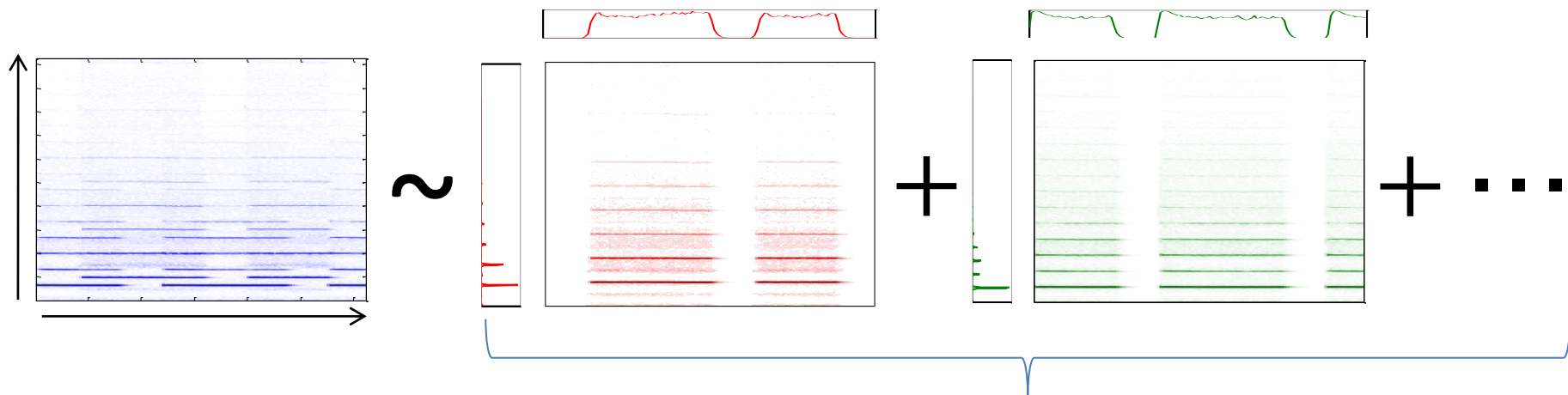
$$\text{線形モデル: } y_t \sim f(h_1 U_{1,t} + h_2 U_{2,t})$$

Bayesian NMF

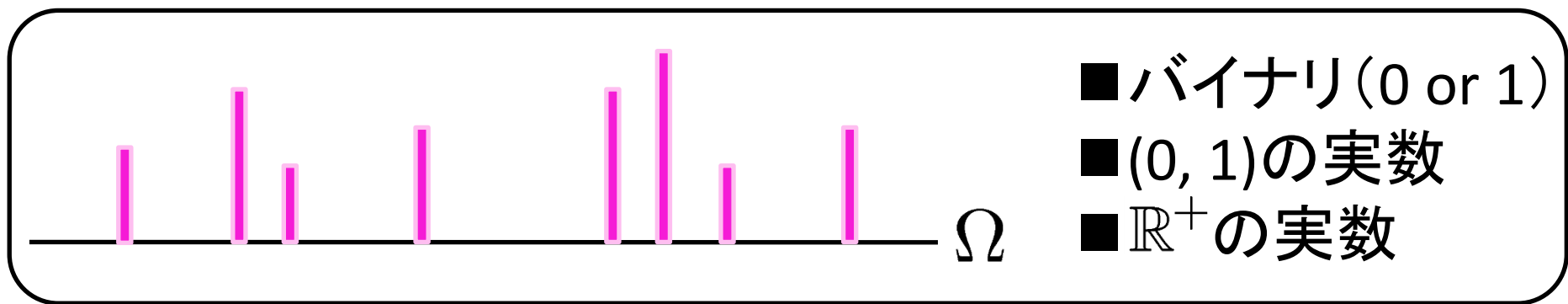
$$Y_{\omega,t} \sim \text{Exp}\left(1 / \sum_d H_{\omega,d} U_{d,t}\right)$$

$$H_{\omega,d} \sim \text{Gamma}(a_H, b_H) \quad U_{d,t} \sim \text{Gamma}(a_U, b_U)$$

モデルの複雑度の設定: 無限因子モデル



潜在的に無限の因子を考えたい！！

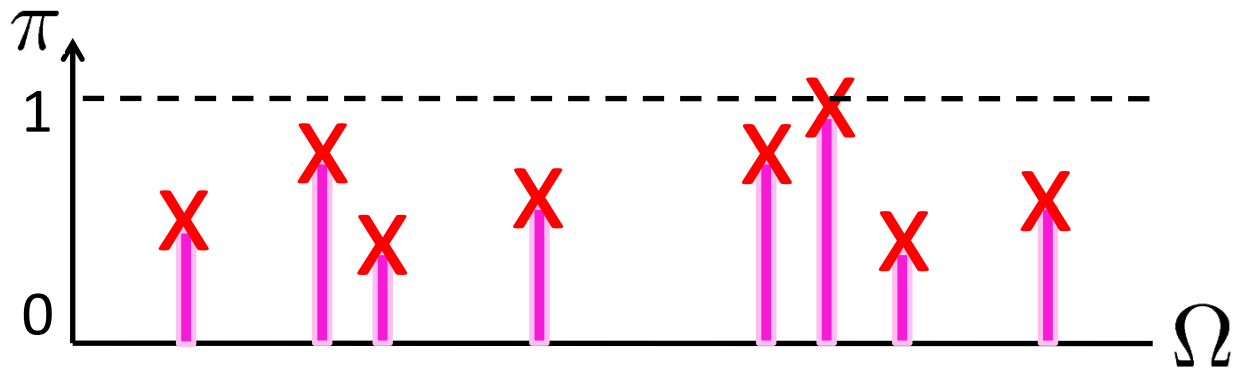


Lévy processの特別な(離散的な)場合

Machine learning の文脈でよく登場するのは

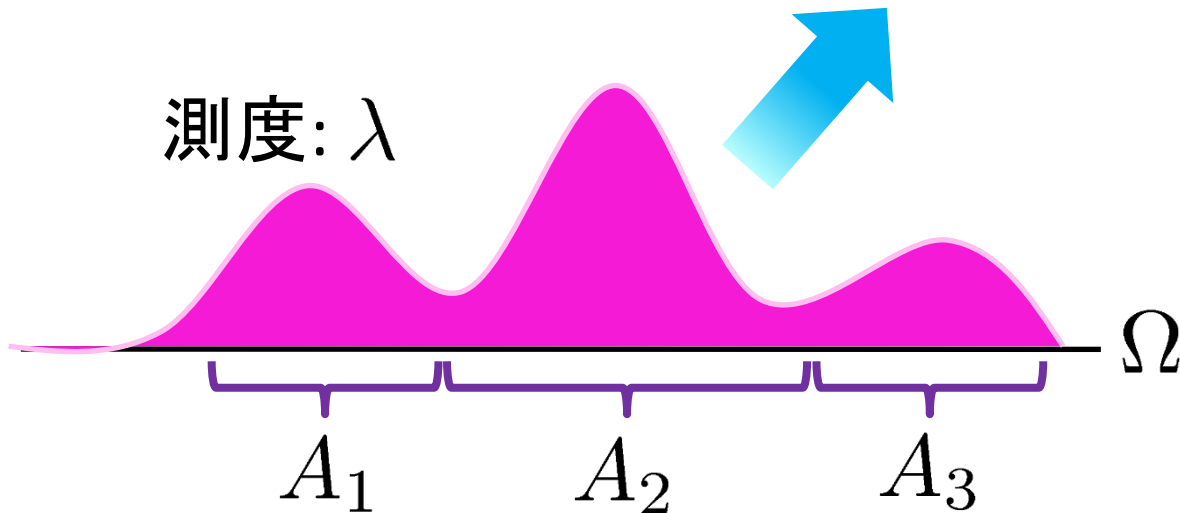
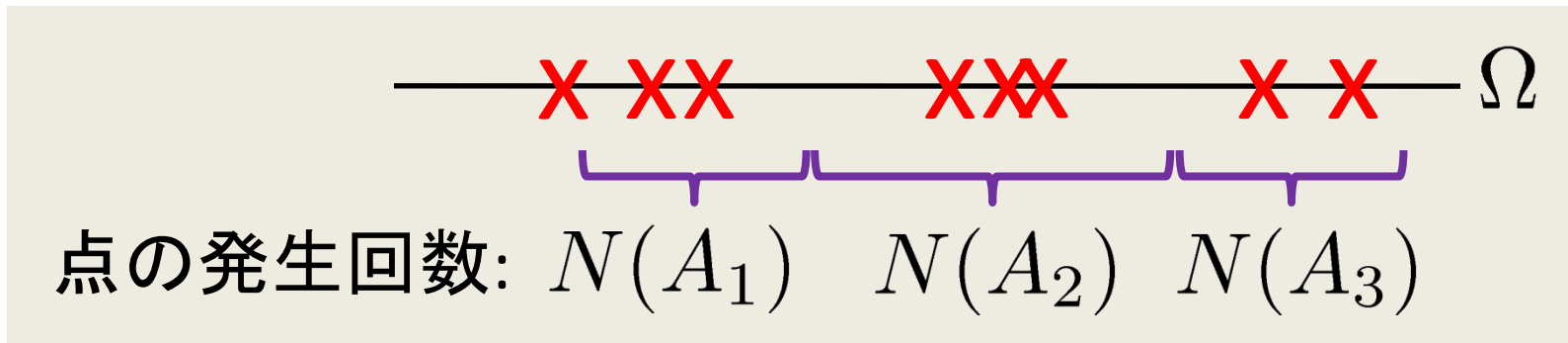
1. Beta process
2. Bernoulli process
3. Gamma process
4. Negative binomial process

例) Beta process:



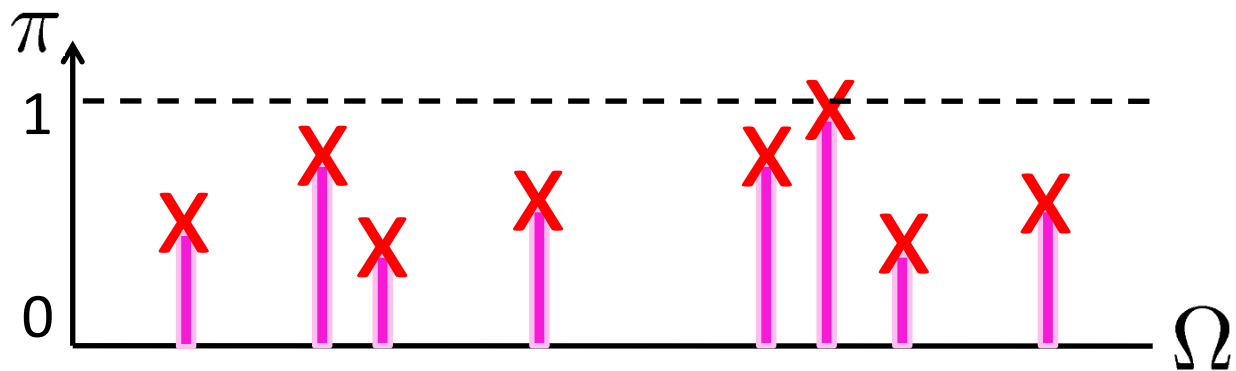
(準備) Poisson process

$$N(A_k) \sim \text{Poisson}(\lambda(A_k)) \text{ (i.i.d.)}$$



Lévy processの特別な(離散的な)場合 [Kingman1967]

Beta process: $(0, 1) \times \Omega$ のPoisson process



Levy measure

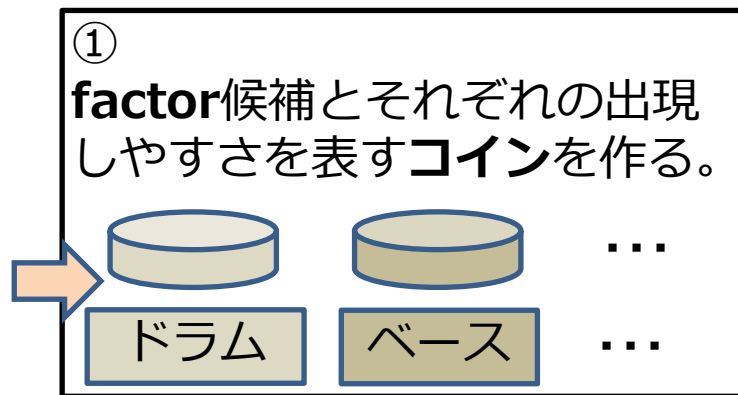
$$\nu(d\pi, d\omega) = c \pi^{-1} (1 - \pi)^{c-1} d\pi \mu(d\omega)$$

Beta(0, c)

$$\int \text{Beta}(0, c) d\pi \rightarrow \infty$$

Beta – Bernoulli process [Thibaux & Jordan2007]

事前知識

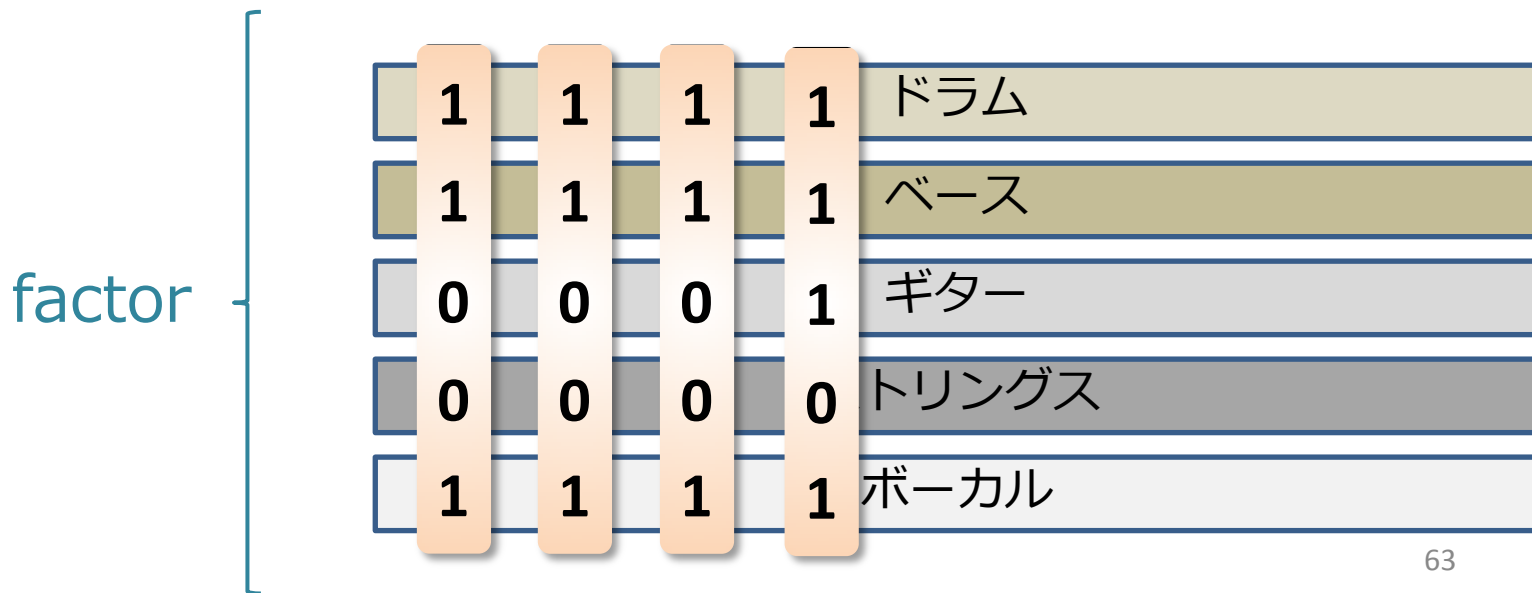
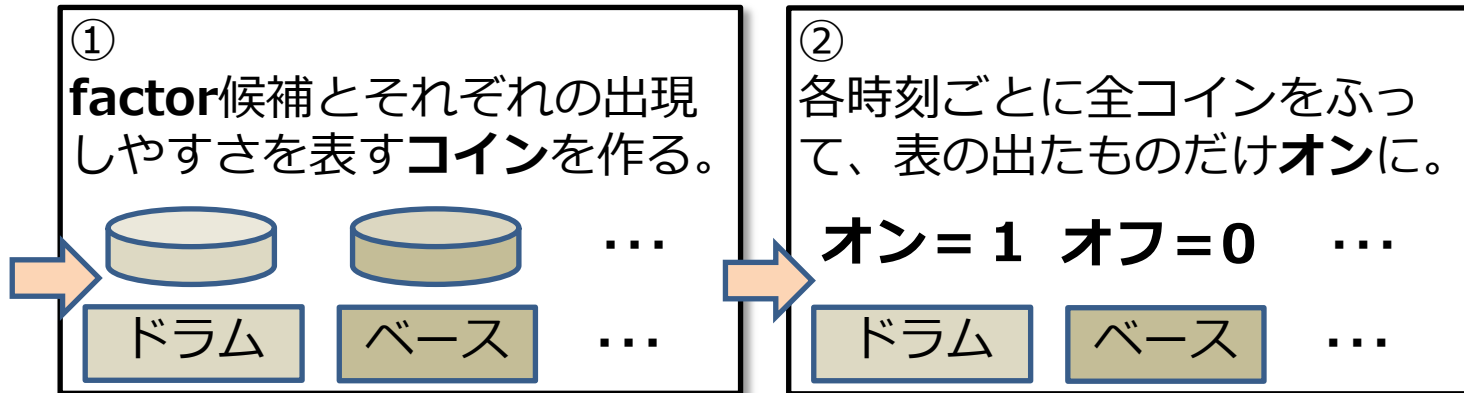


factor



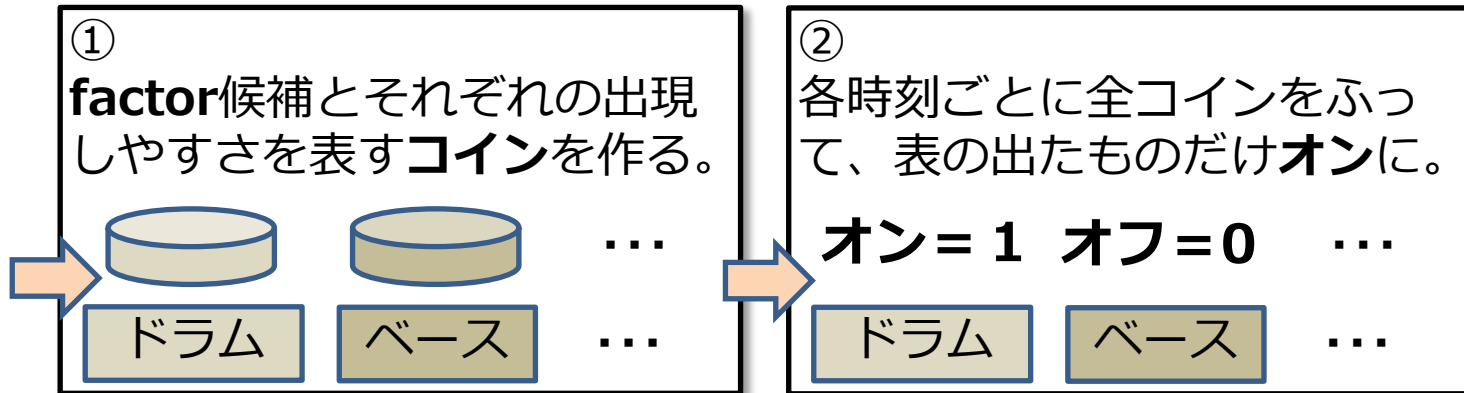
Beta – Bernoulli process [Thibaux & Jordan2007]

事前知識



Beta – Bernoulli process [Thibaux & Jordan2007]

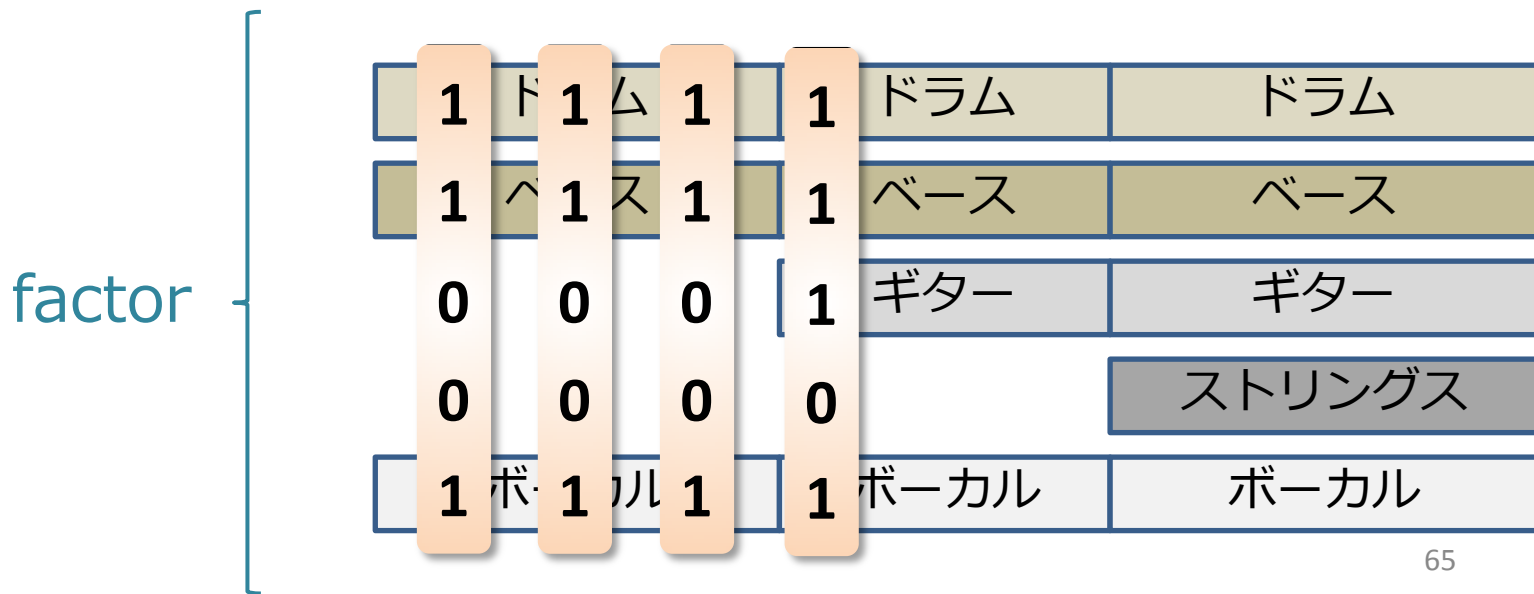
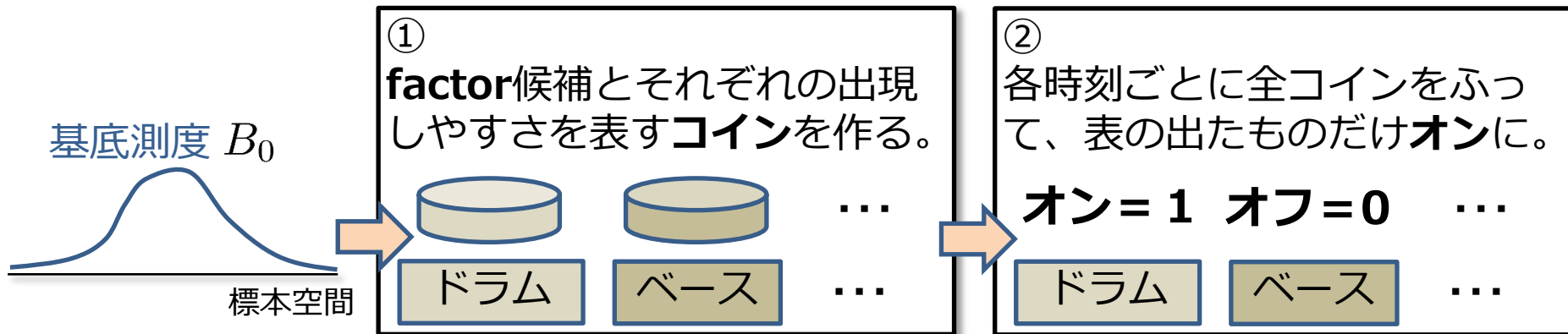
事前知識



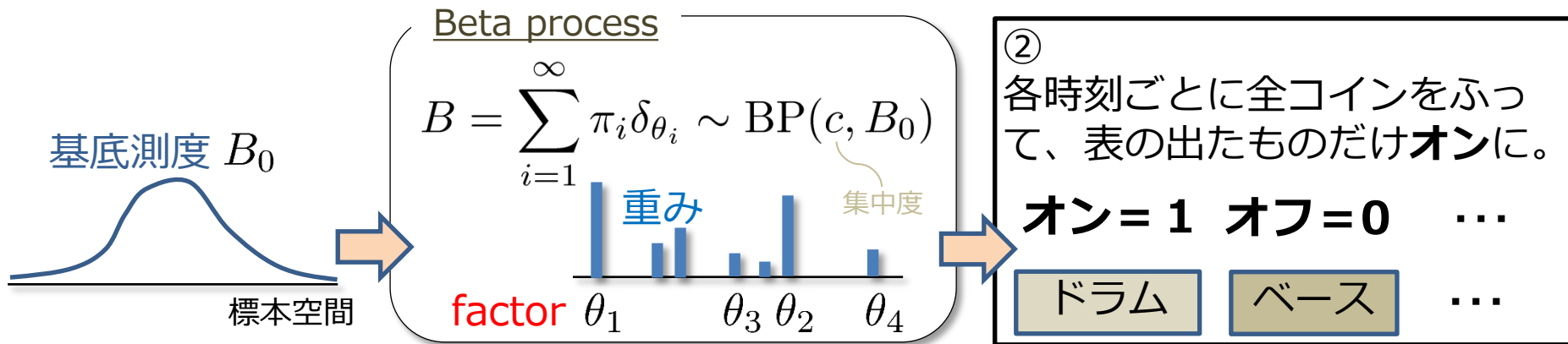
factor

1	ド	1	ム	1	1	ドラム	ドラム
1	ベ	1	ス	1	1	ベース	ベース
0		0		0	1	ギター	ギター
0		0		0			ストリングス
1	ボ	1	カ	1	1	ボーカル	ボーカル

Beta – Bernoulli process [Thibaux & Jordan2007]



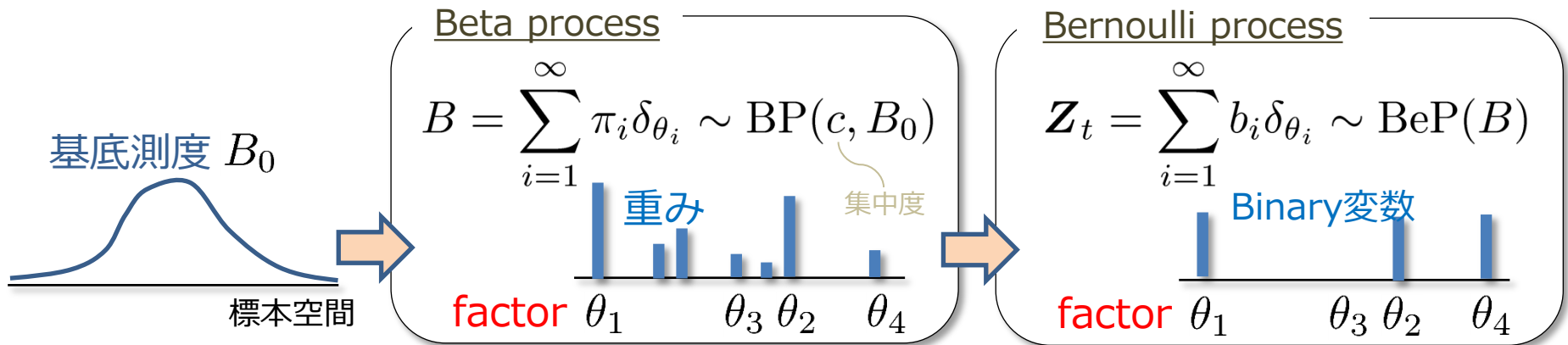
Beta – Bernoulli process [Thibaux & Jordan2007]



factor

θ_1	1	1	1	1	ドラム	ドラム
θ_2	1	1	1	1	ベース	ベース
θ_3	0	0	0	1	ギター	ギター
θ_4	0	0	0	0		ストリングス
⋮	1	1	1	1	ボーカル	ボーカル
					⋮	

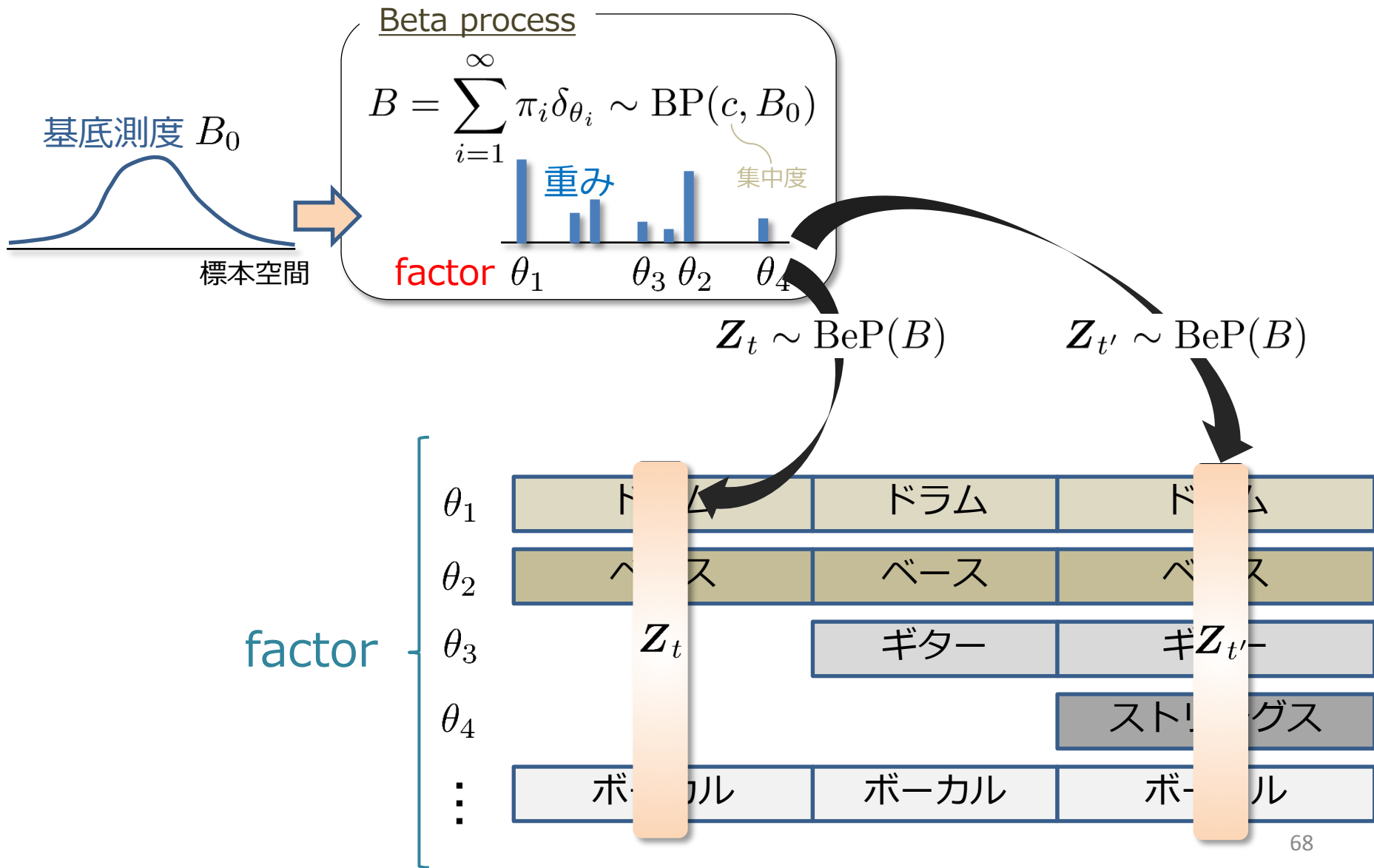
Beta – Bernoulli process [Thibaux & Jordan2007]



factor

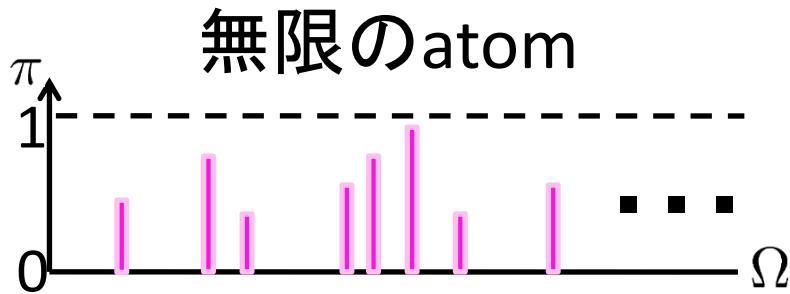
θ_1	ト	ム	ドラム	ドラム
θ_2	ヘ	ス	ベース	ベース
θ_3	Z_t		ギター	ギター
θ_4	Z_t			ストリングス
⋮	ボ	ー	ボーカル	ボーカル

Beta – Bernoulli process [Thibaux & Jordan2007]



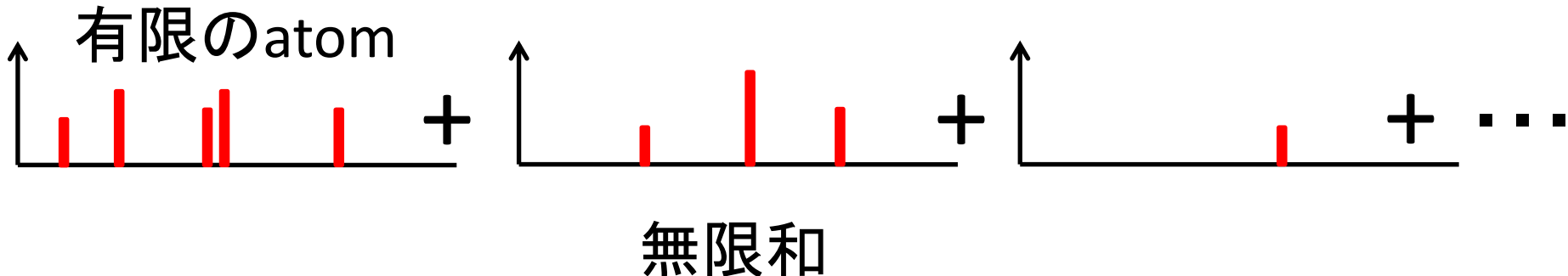
Beta processのLevy measure decomposition [Ren+2012]

Beta process

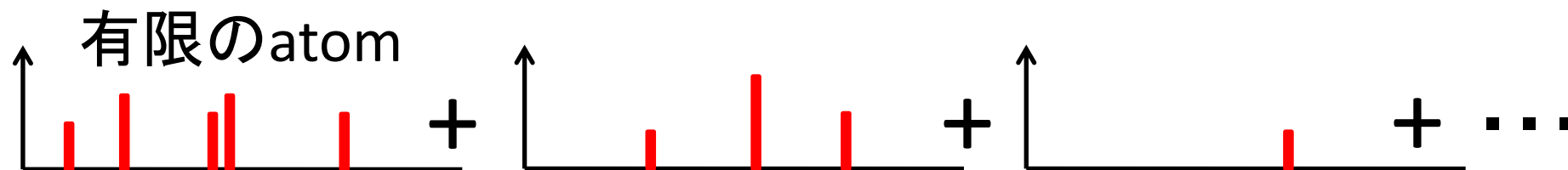


Levy measure: $\nu(d\pi, d\omega)$
 $= c(\omega)\pi^{-1}(1 - \pi)^{c(\omega)-1}d\pi\mu(d\omega)$
Beta(0, c(ω)) improper

Improper beta の解消！！



Beta processのLevy measure decomposition [Ren+2012]



$$\nu(d\pi, d\omega) = c(\omega) \pi^{-1} (1 - \pi)^{c(\omega)-1} d\pi \mu(d\omega)$$

Taylor展開しただけ

$$= c(\omega) \sum_{k=0}^{\infty} (1 - \pi)^k (1 - \pi)^{c(\omega)-1} d\pi \mu(d\omega)$$

$$= \sum_{k=0}^{\infty} \frac{\Gamma(c(\omega) + k + 1)}{\Gamma(c(\omega) + k) \Gamma(1)} \pi^{1-1} (1 - \pi)^{c(\omega)+k-1} d\pi \frac{c(\omega)}{c(\omega) + k} \mu(d\omega)$$

$$= \sum_{k=0}^{\infty} \text{Beta}(1, c(\omega) + k) d\pi \frac{c(\omega)}{c(\omega) + k} \mu(d\omega)$$

Improperではない！！

モデルの拡張法

2012

Infinite HMM
[Beal+2002]

Hierarchical DP
[Teh+2006]

Infinity-gram
[Mochihashi+2007]

階層化

Infinite PCFG
[Liang+2007]

Hierarchical BP
[Jordan2007]

Nested
hierarchical DP
[Paisley+2012]

入れ子

Nested DP
[Rodriguez+2008]

Nested BP
[Jordan2009]

相関

Kernel SBP
[Dumson2008]

Nested GaP
[Jordan2009]

Nested partition model
[Rodriguez+2012]

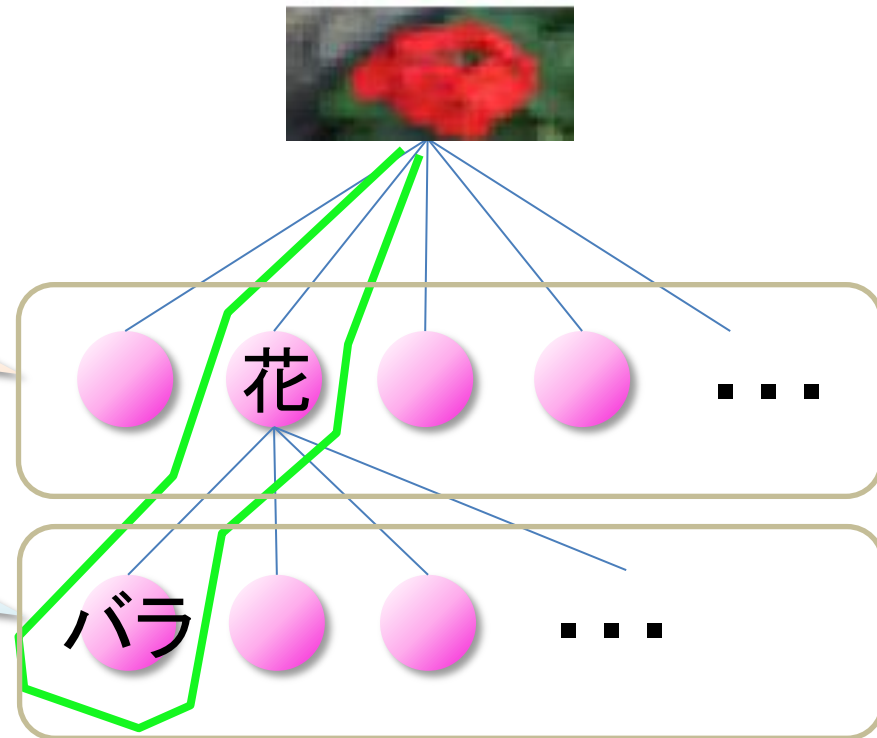
logistic SBP
[Ren+2011]

Kernel BP
[Ren+2011]

Nested Dirichlet process [Rodriguez+2008]

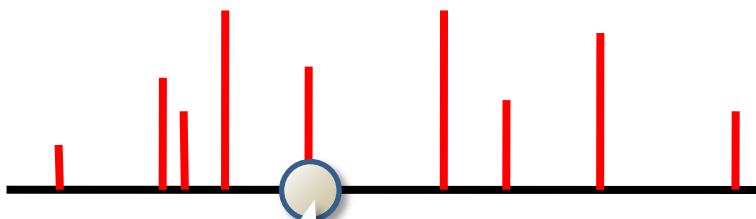
おおまかなクラスタリング

詳細なクラスタリング

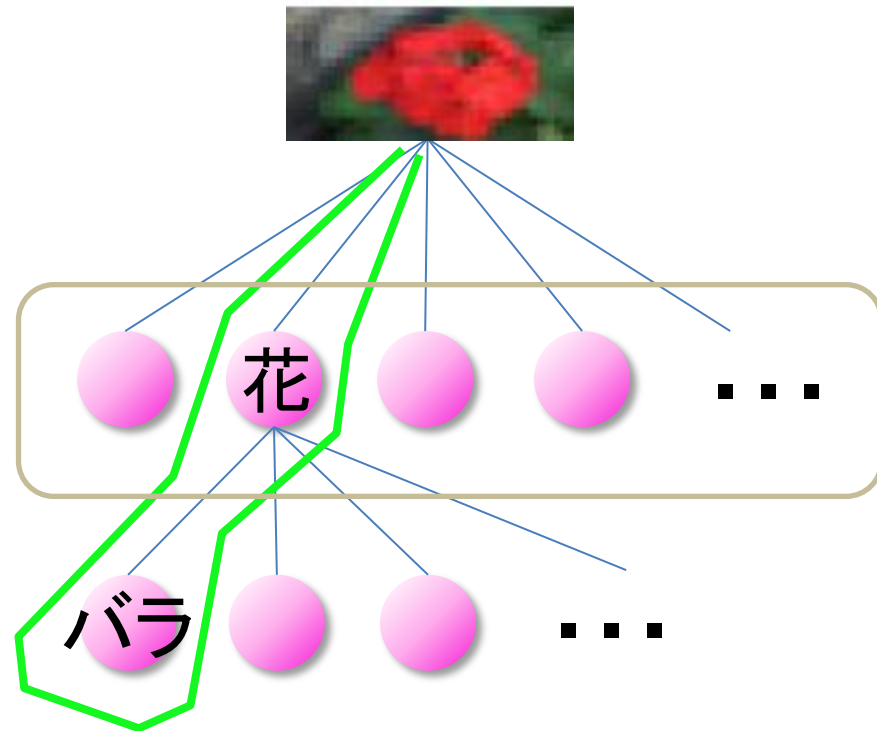


Nested Dirichlet process [Rodriguez+2008]

無限混合: $G \sim \text{DP}(\gamma, G_0)$

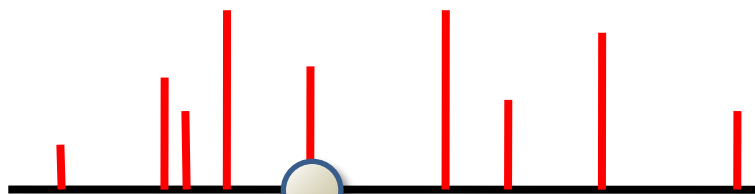


次の詳細なクラスタリング
用の手がかり

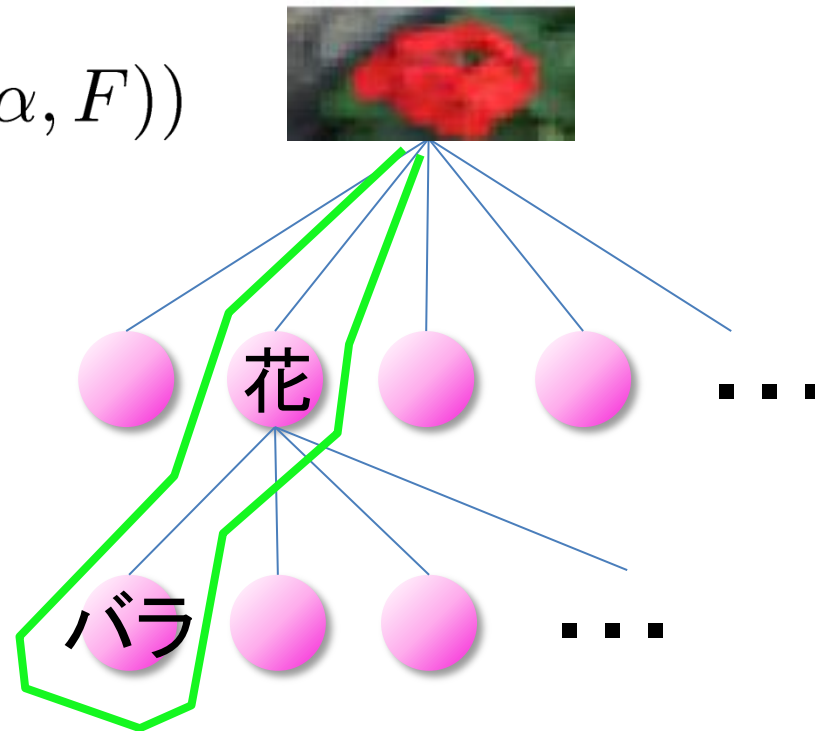


Nested Dirichlet process [Rodriguez+2008]

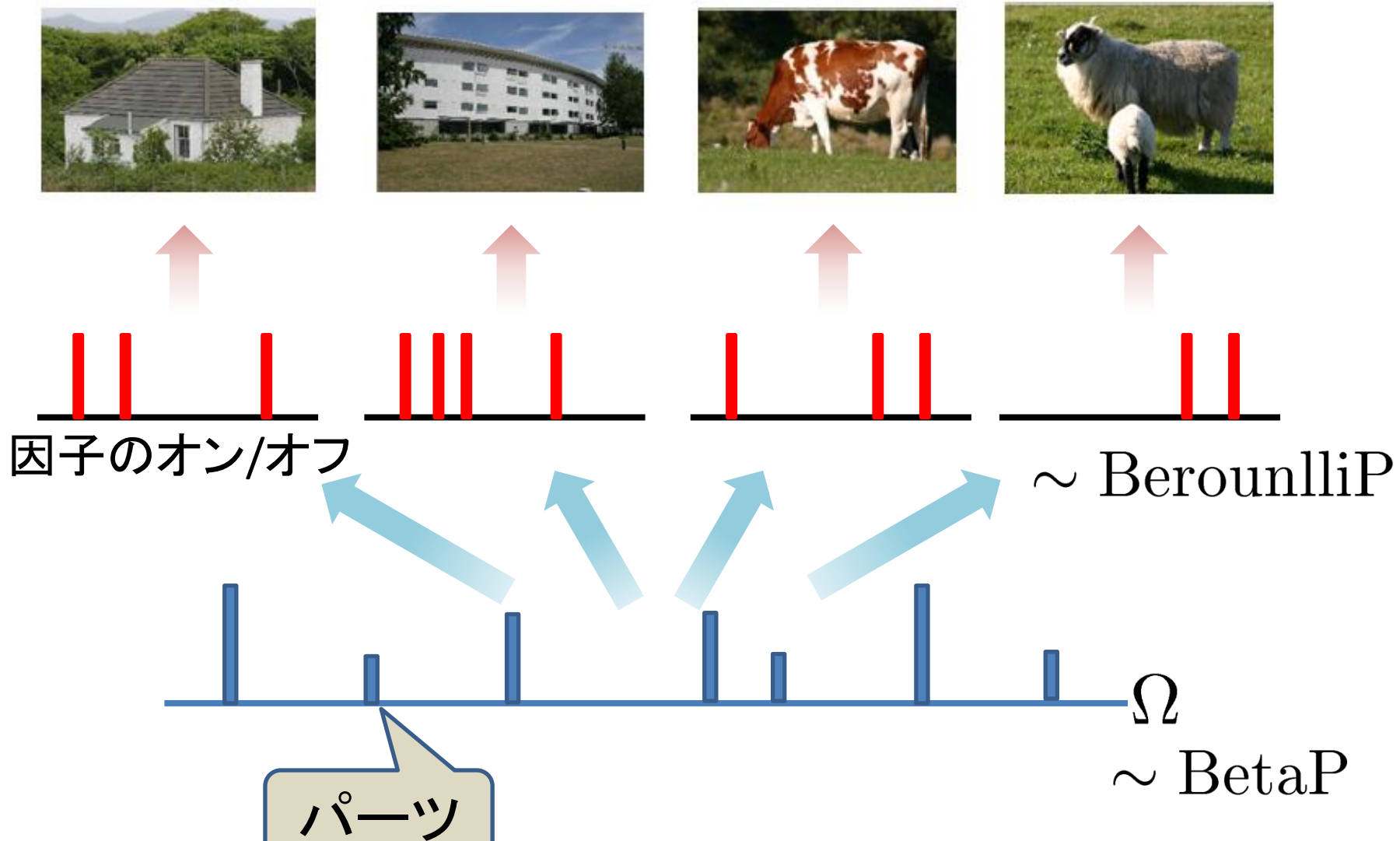
$$G = \sum_k \pi_k \delta_{G_k^*} \sim \text{DP}(\gamma, \text{DP}(\alpha, F))$$



$$G_k^* \sim \text{DP}(\alpha, F)$$



画像群に対する木構造の分類+パーツ分解 [Li+2012]



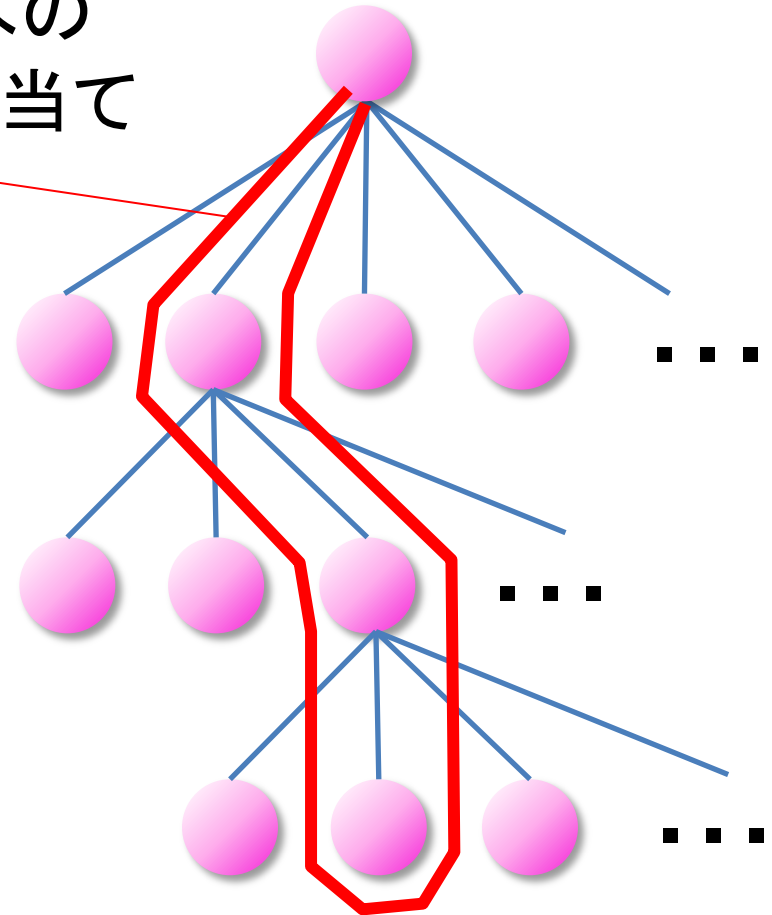
L. Li, X. Zhang, M. Zhou and L. Carin (2012),

Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI.

画像群に対する木構造の分類+パーツ分解 [Li+2012]



各画像を木の
pathに割り当て

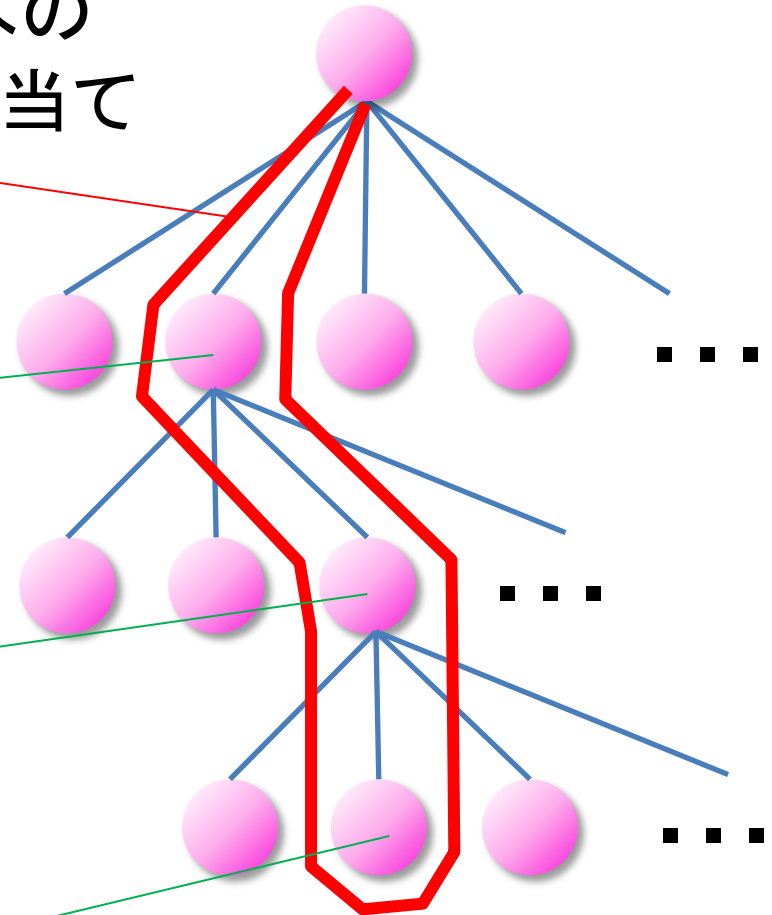
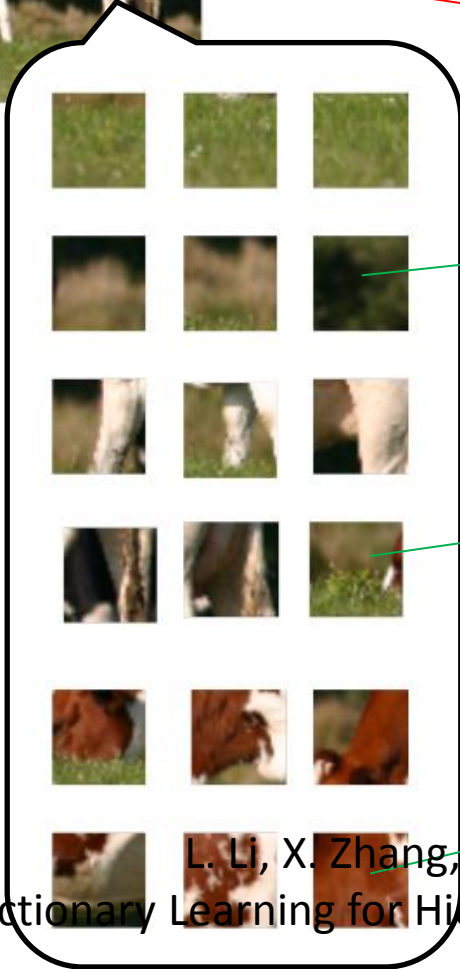


L. Li, X. Zhang, M. Zhou and L. Carin (2012),
Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI.

画像群に対する木構造の分類+パーツ分解 [Li+2012]



各画像を木の path に割り当て



L. Li, X. Zhang, M. Zhou and L. Carin (2012),
Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI.

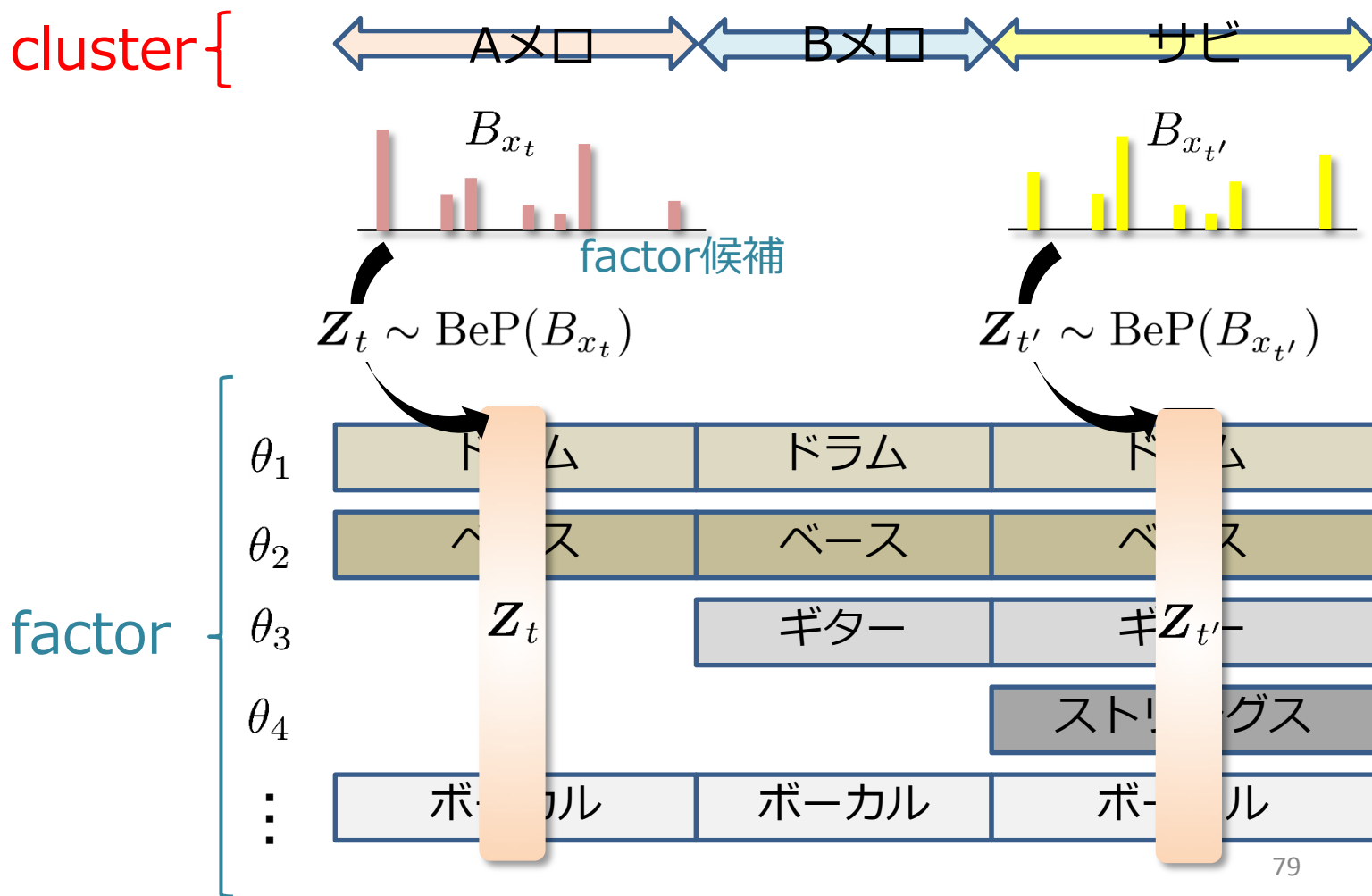
画像群に対する木構造の分類+パーツ分解

[Li+2012]



L. Li, X. Zhang, M. Zhou and L. Carin (2012),
Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI.

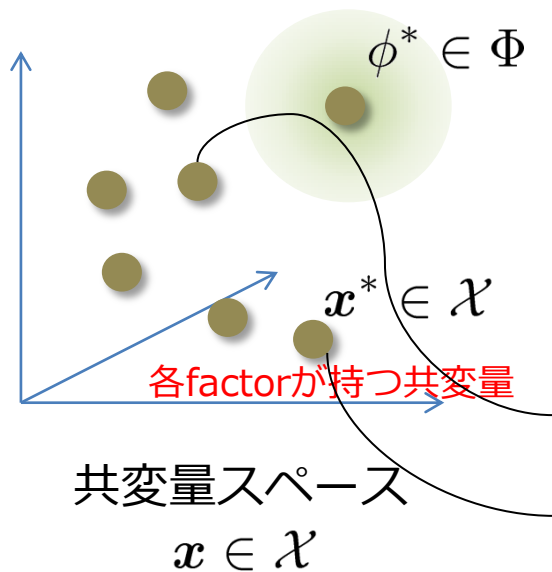
Cluster と *factor* を結びつけるために共変量を導入！



Cluster と *factor* を結びつけるために共変量を導入！



各factorが共変量スペースの局所的な計量を定めるパラメータ



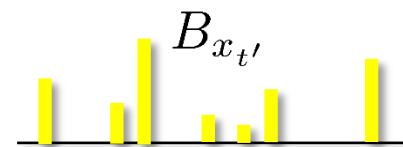
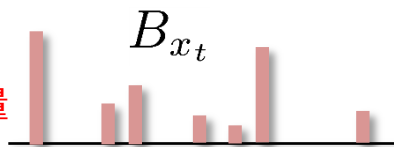
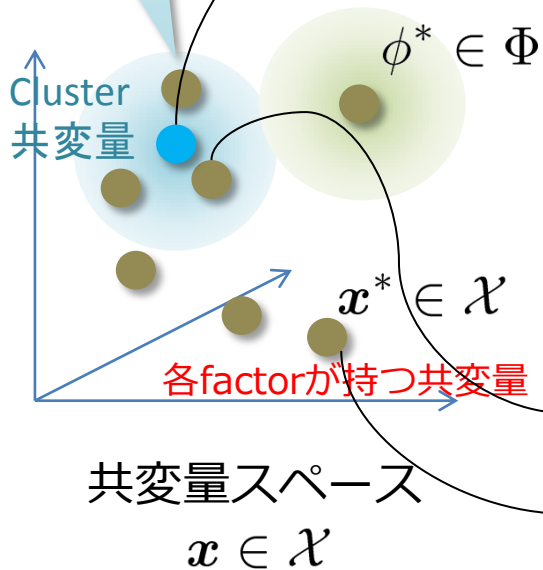
θ_1	ドラム	ドラム	ドラム
θ_2	ベース	ベース	ベース
θ_3	Z_t	ギター	ギター
θ_4			ストリングス
\vdots	ボーカル	ボーカル	ボーカル

Kernel Beta process [Ren+2011]

Cluster と factor を結びつけるために共変量を導入!

Cluster共変量に近いfactorがactiveになりやすいように!

各factorが共変量スペースの局所的な計量を決めるパラメータ

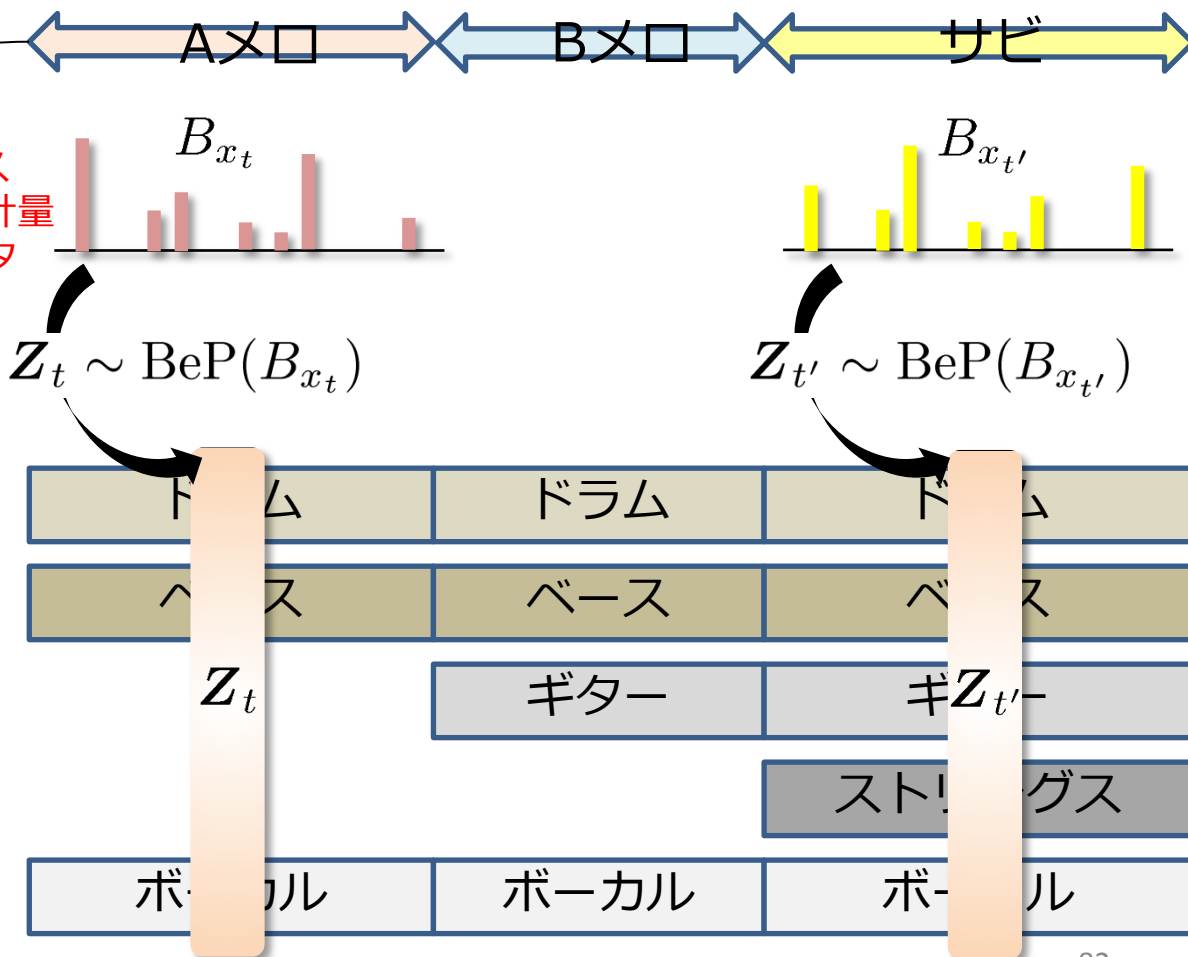
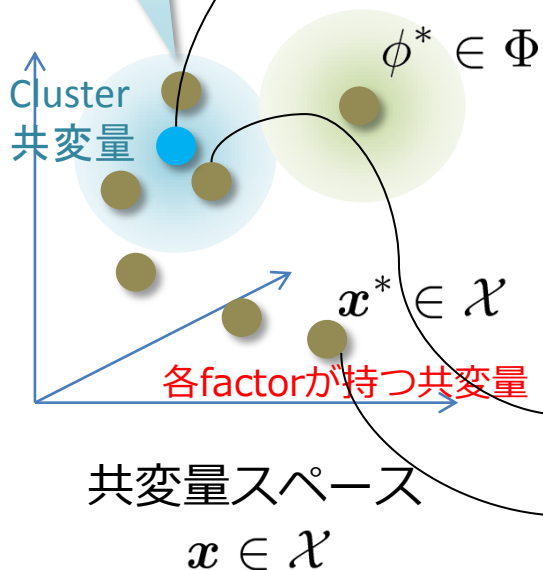


θ_1	ドラム	ドラム	ドラム
θ_2	ベース	ベース	ベース
θ_3	Z_t	ギター	ギター
θ_4			ストリングス
\vdots	ボーカル	ボーカル	ボーカル

Cluster と **factor** を結びつけるために共変量を導入!

Cluster共変量に近いfactorがactiveになりやすいように!

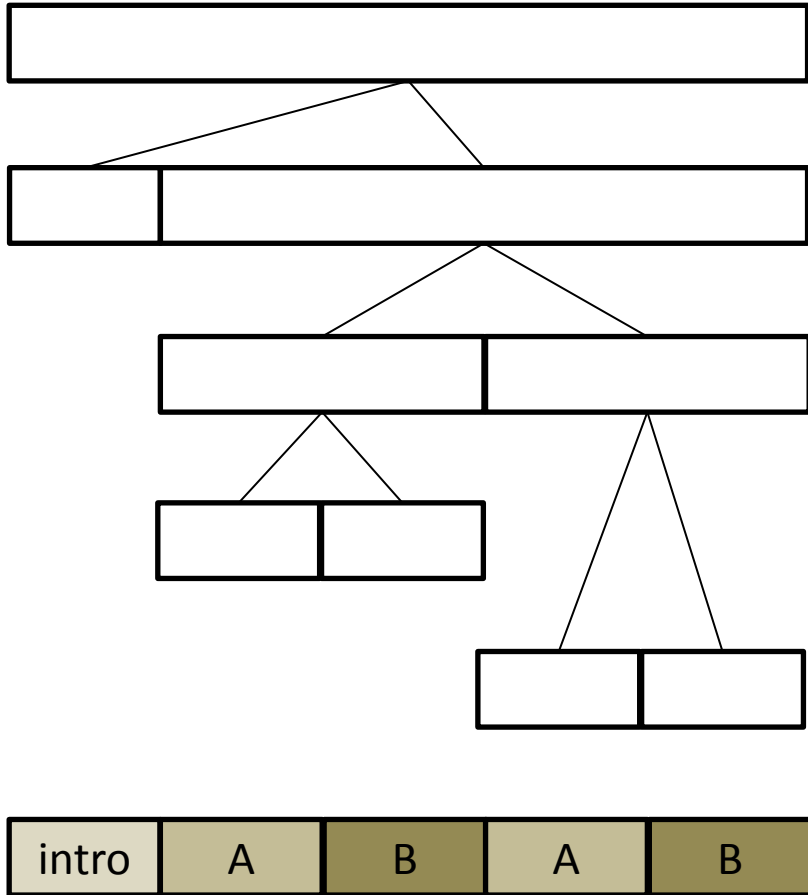
各factorが共変量スペースの局所的な計量を定めるパラメータ



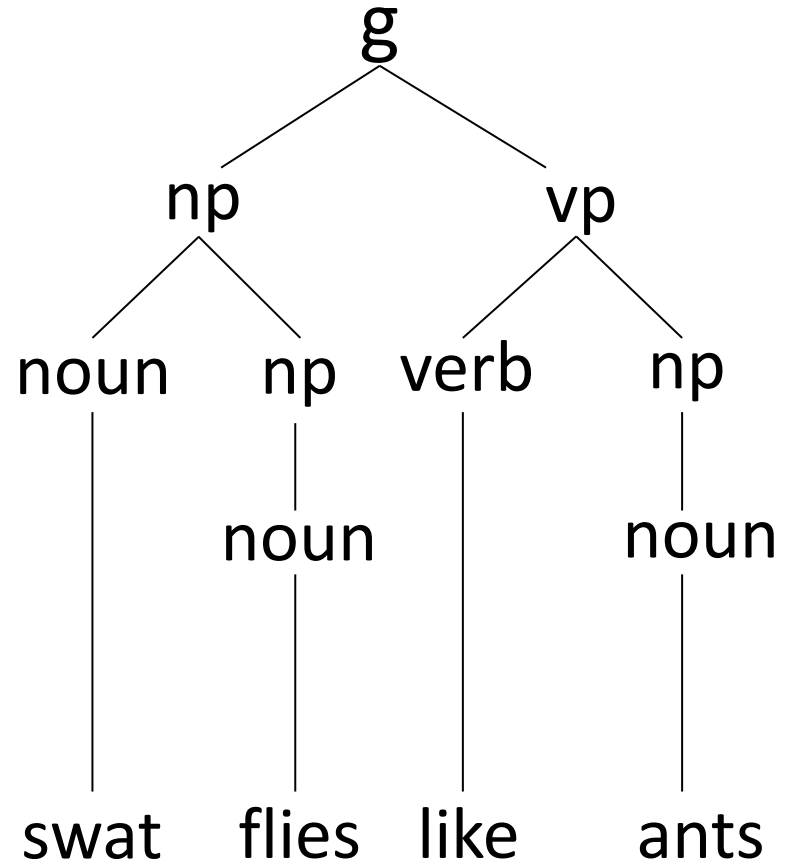
音楽信号からの構文解析

[Nakano+2011, Kameoka+2012]

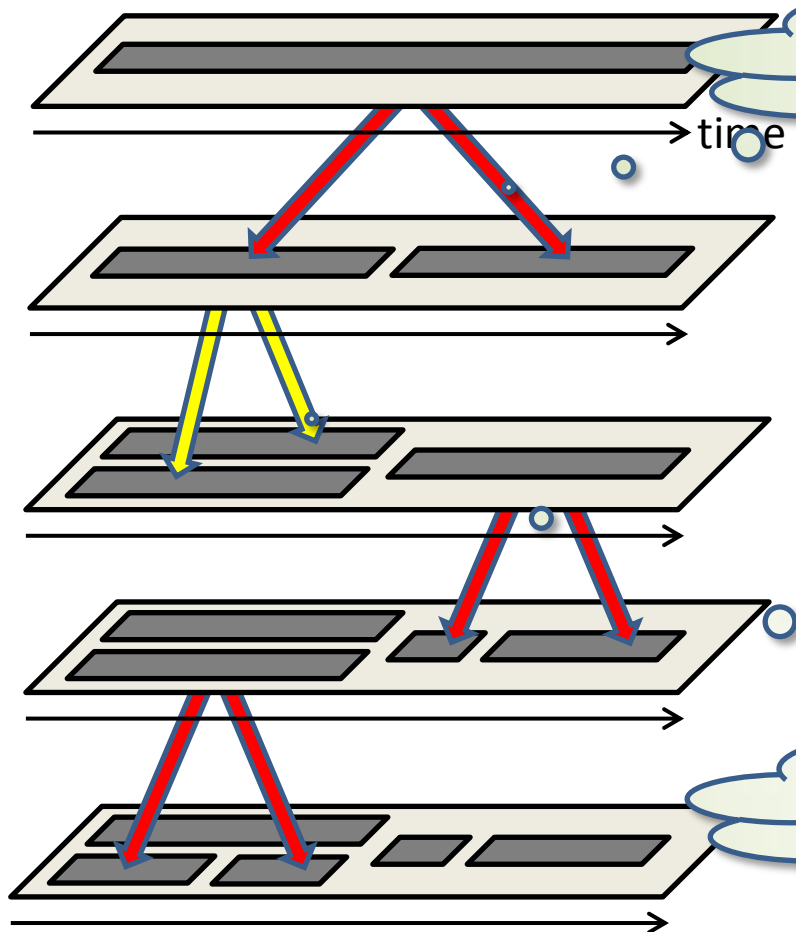
音楽



自然言語



音楽の構造には「時間」の情報が重要な役割を果たす

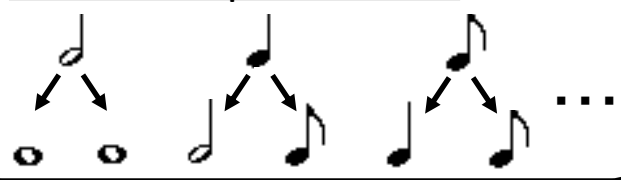


時間分割の分岐規則

Realistic productions



Unrealistic productions

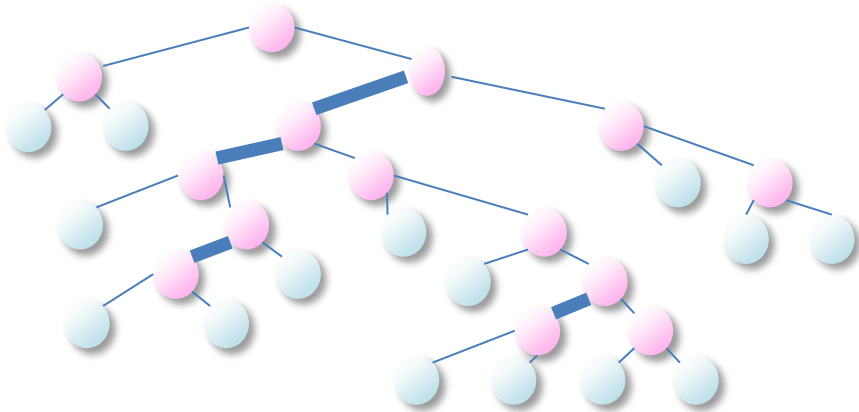


同期の分岐規則

音楽信号からの構文解析

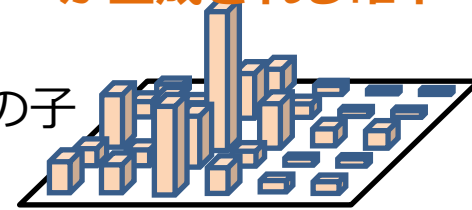
[Nakano+2011, Kameoka+2012]

シンボルの木構造に対する確率分布を作りたい！

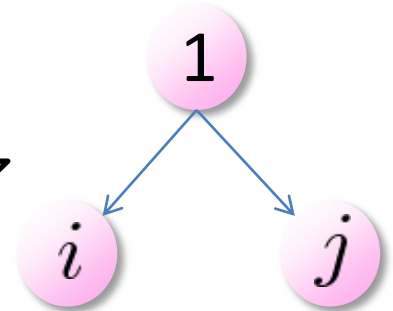


親"1"から子(i, j)
が生成される確率

左の子



右の子

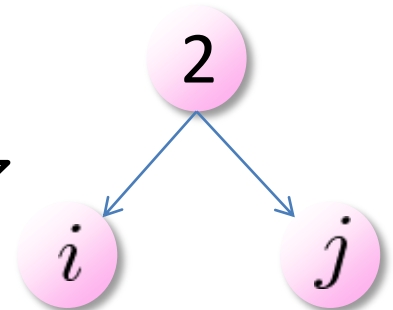


親"2"から子(i, j)
が生成される確率

左の子



右の子



■
■
■

音楽信号からの構文解析

[Nakano+2011, Kameoka+2012]

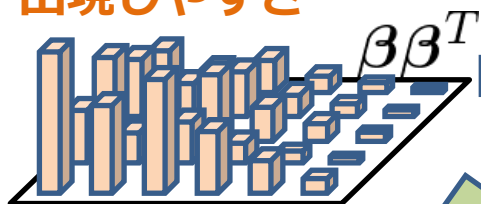
シンボル候補とその出現しやすさ



ϕ_k

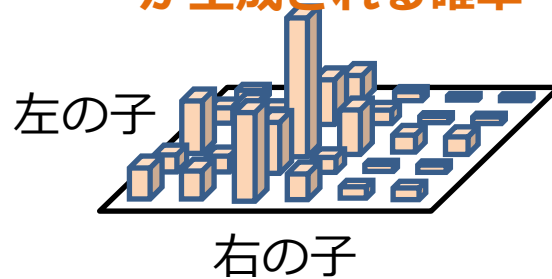
シンボル候補

シンボル2つ組の出現しやすさ

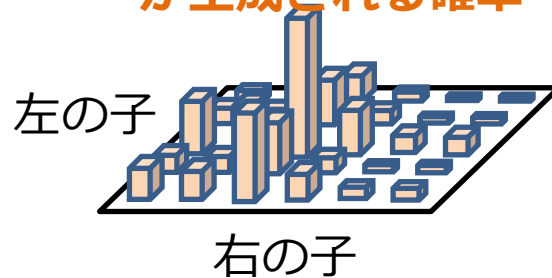


DirichletProcess

親"1"から子(i, j)が生成される確率



親"2"から子(i, j)が生成される確率



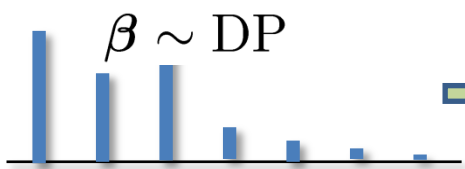
■
■
■

音楽信号からの構文解析

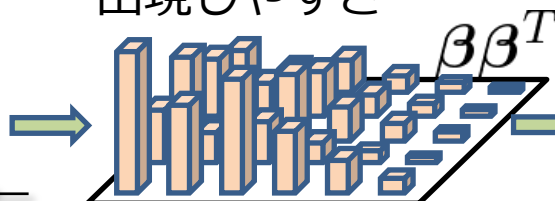
[Nakano+2011, Kameoka+2012]

従来のinfinite PCFG

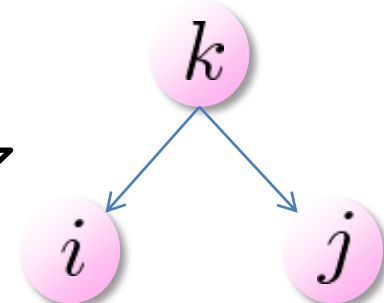
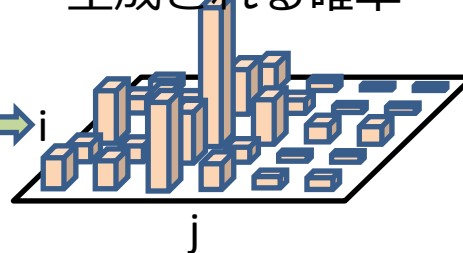
シンボル候補とその出現しやすさ



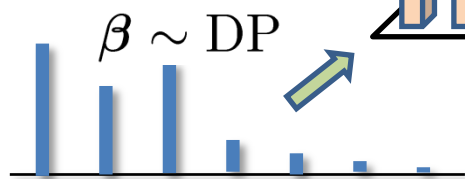
シンボル2つ組の出現しやすさ



親kから子(i, j)が生成される確率

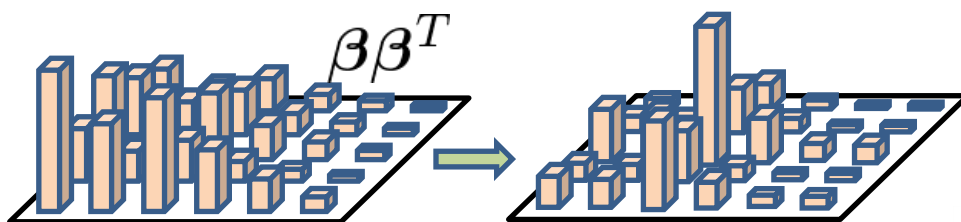


提案モデル

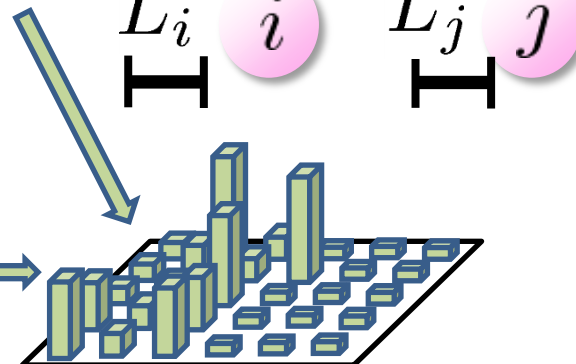


音長

$$w_{i,j}^{(k)} = \exp\left(-\frac{(L_k - L_i - L_j)^2}{\sigma^2}\right)$$

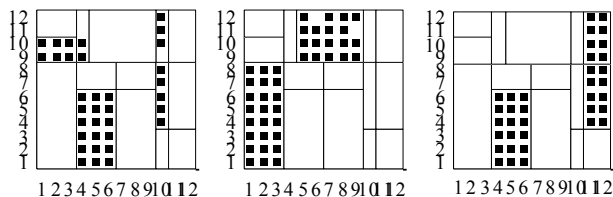
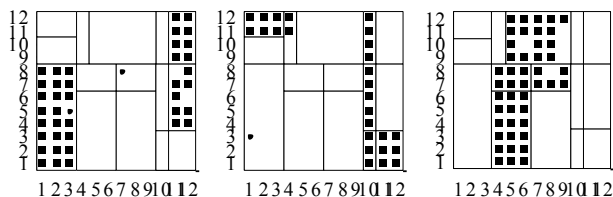
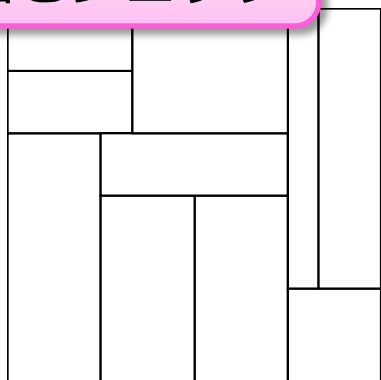


親子間で音長を保存するよう働くバイアス



Mondrian HMM (モンドリアン模様の状態遷移確率の生成モデル)

隠れ状態2-gram
表に潜むブロック

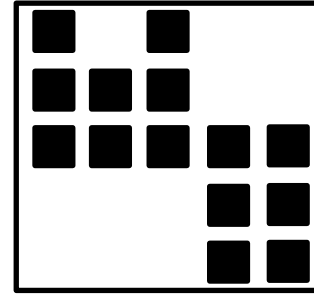
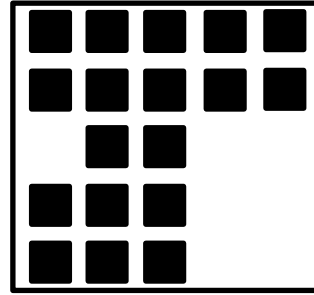
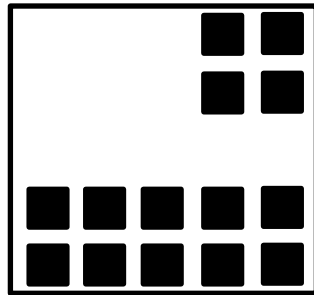
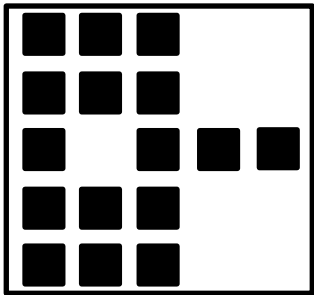
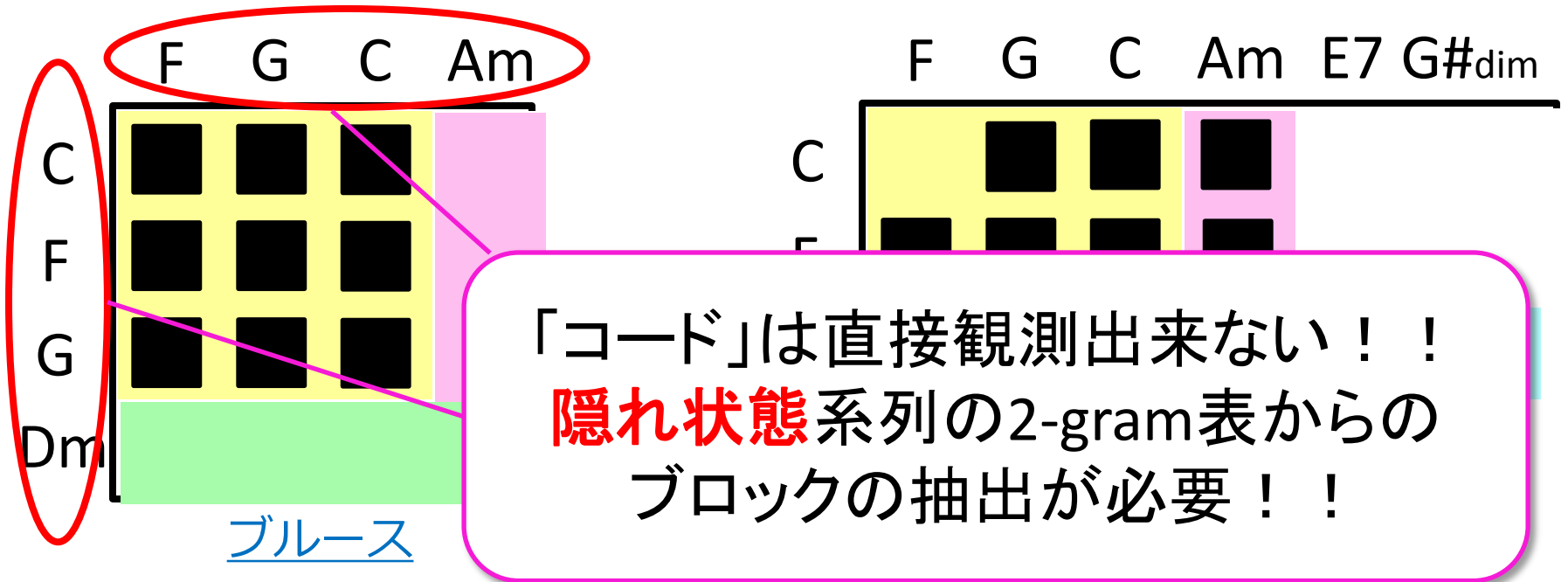


楽曲ごとの遷移のオン/オフ



楽曲ごとの状態遷移確率

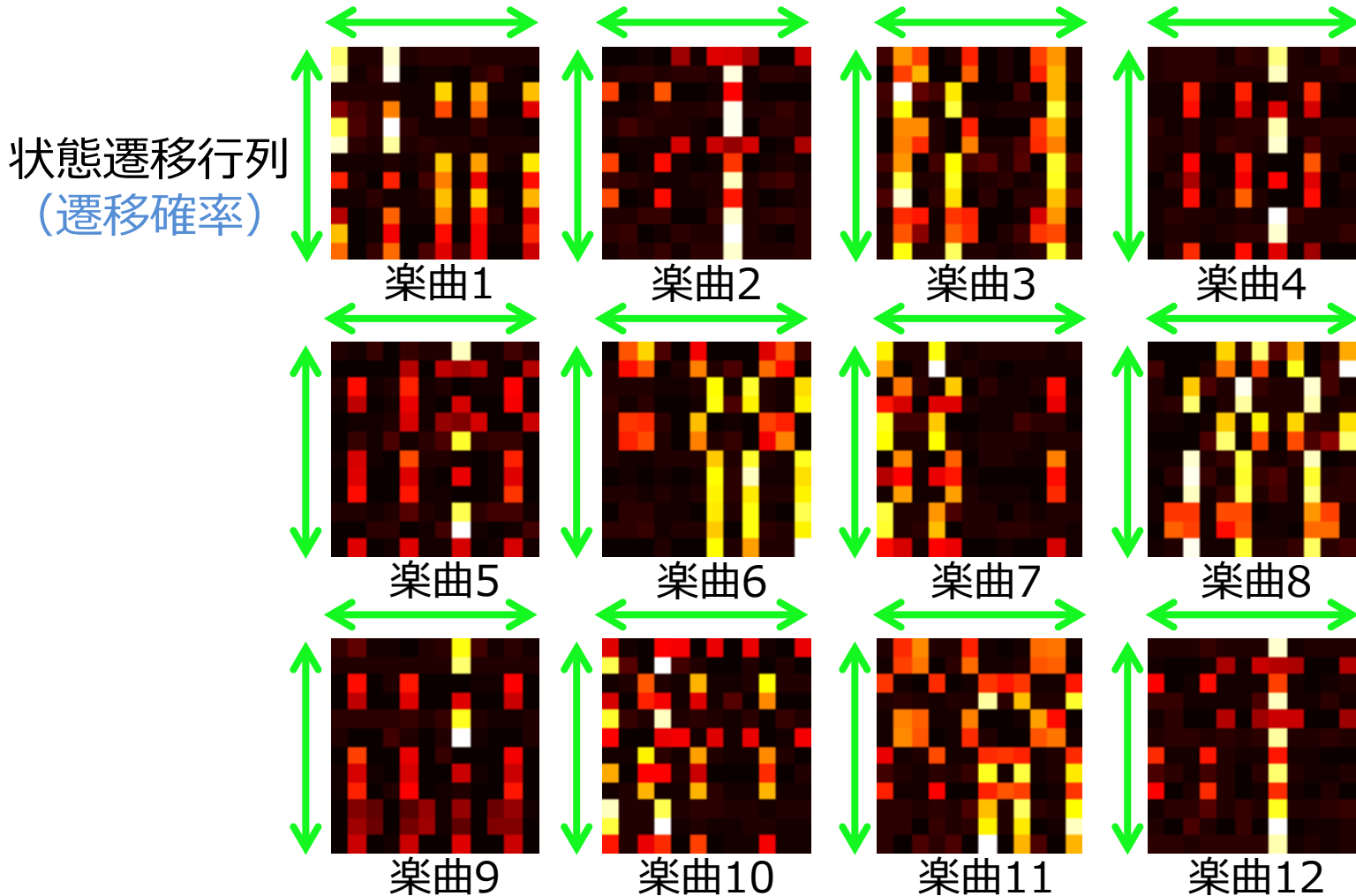
動機: コード進行2-gram表からブロックを見つけたい



...

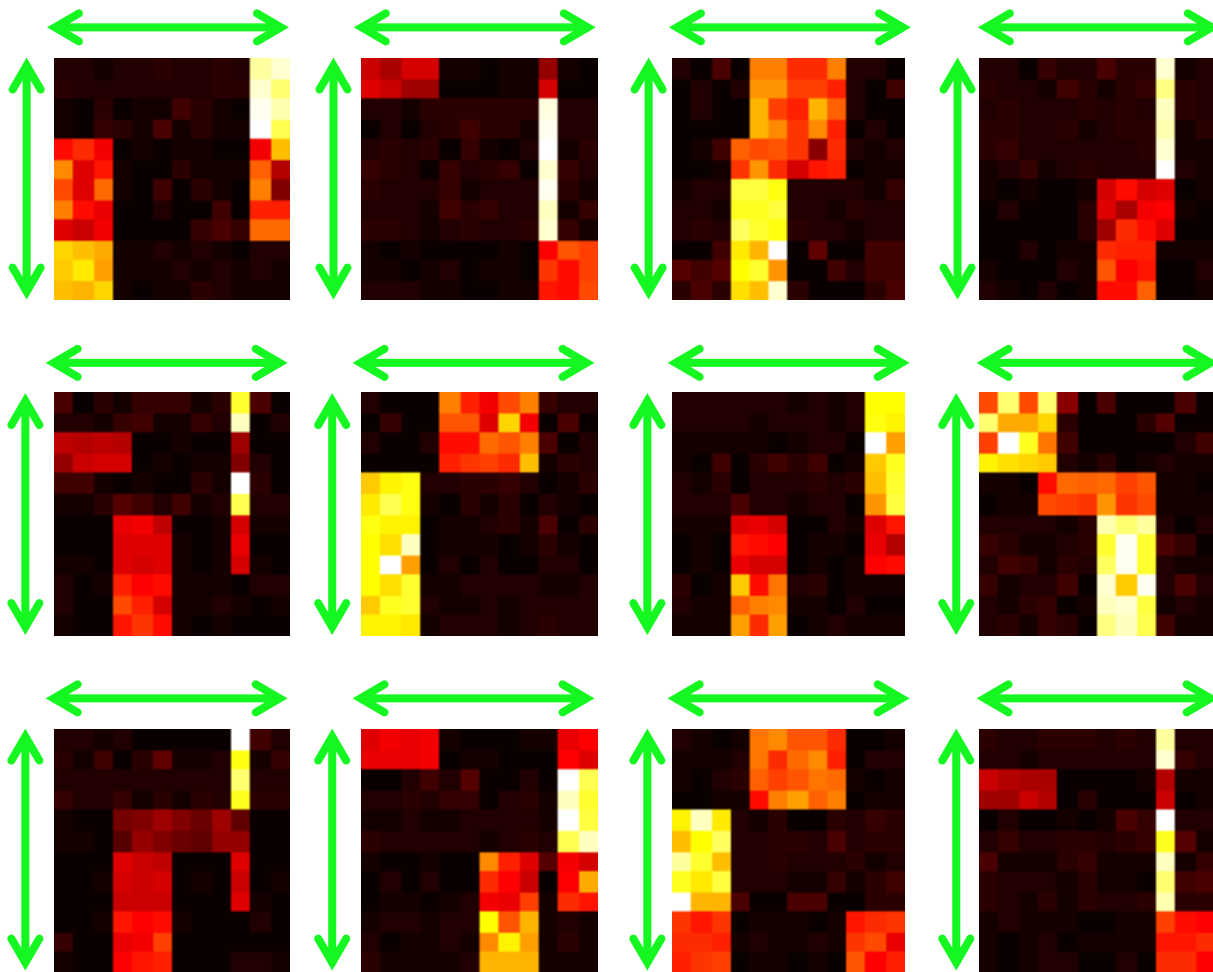
目的: 状態遷移行列内のブロックの発見

複数の楽曲に隠れマルコフモデルを適用する際に...



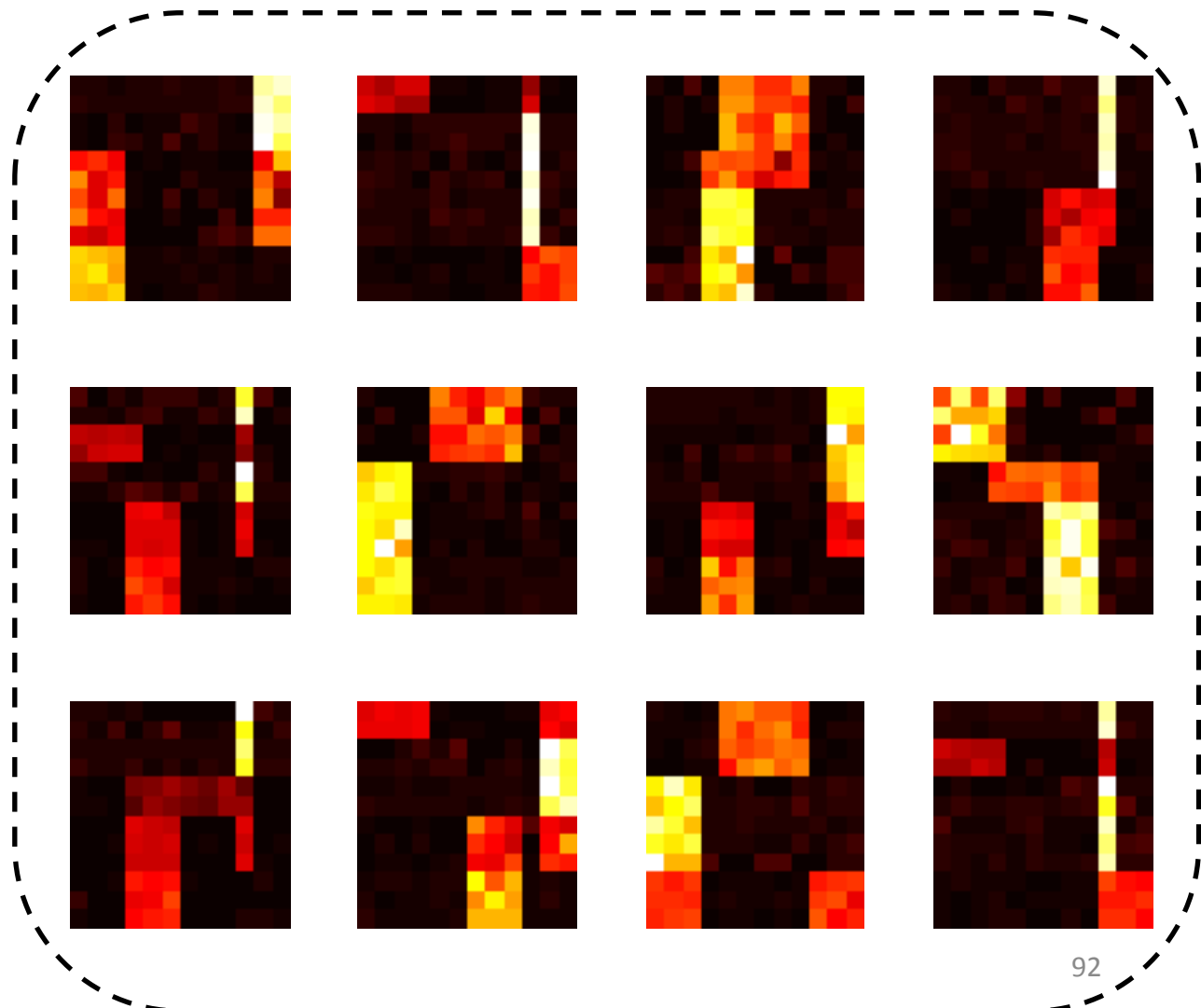
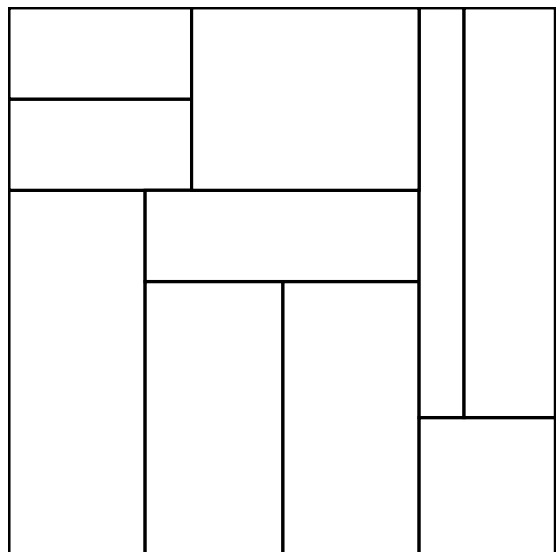
目的: 状態遷移行列内のブロックの発見

状態の並び順を上手く誘導しつつ...



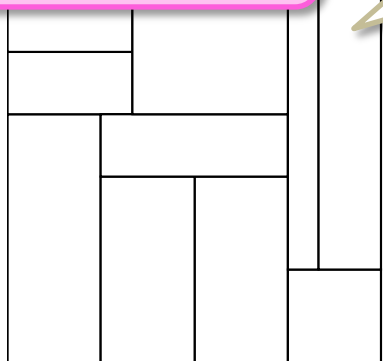
目的: 状態遷移行列内のブロックの発見

状態遷移配列の中に潜むブロックを見つけたい!



Mondrian HMM (モンドリアン模様の状態遷移確率の生成モデル)

隠れ状態2-gram
表に潜むブロック



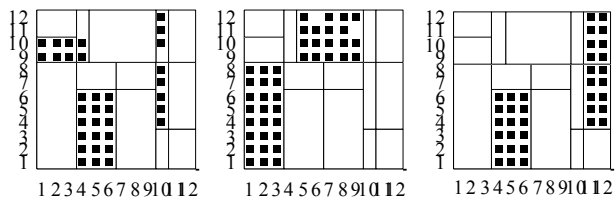
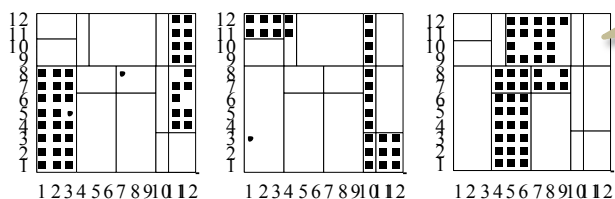
Unit squareへのパーティションの生成
 $\mathcal{M} \sim \text{MP}(\lambda, [0, 1], [0, 1])$

隠れ状態に関する縦横の並び順の生成

$S_k \sim \text{Uniform}([0, 1]), S'_k \sim \text{Uniform}([0, 1])$

オン/オフを表すバイナリ変数の生成

$R_{d,i,j} \sim \text{Bernoulli}(\phi_{B_{i,j}}), \phi_B \sim \text{Beta}(a_0, a_1)$



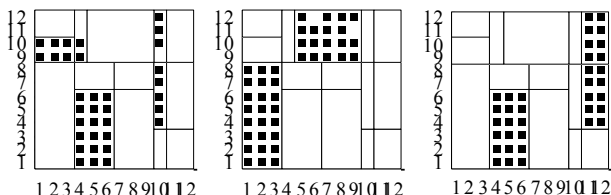
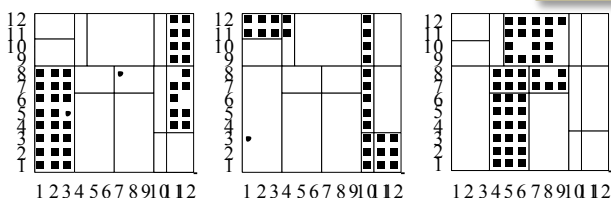
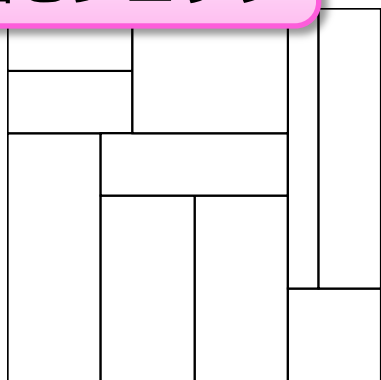
楽曲ごとの遷移のオン/オフ



楽曲ごとの状態遷移確率

Mondrian HMM (モンドリアン模様の状態遷移確率の生成モデル)

隠れ状態2-gram
表に潜むブロック



楽曲ごとの遷移のオン/オフ

無限混合のためのTop-level Dirichlet process

$$G_0 = \sum_{i=1}^{\infty} \beta_i \delta_{\theta_i} \sim \text{DP}(\gamma, F)$$

各遷移のオン/オフに基づく重みの修正

$$\beta_{d,i,j}^* = \frac{1}{1 + \xi_d} \beta_j (\xi_{d,i}^*)^{R_{d,i,j}}, \quad \xi_{d,i}^* = 1 + \frac{\xi_d}{\sum_l R_{d,i,l} \beta_l}$$

無限HMMのためのsecond-level Dirichlet process

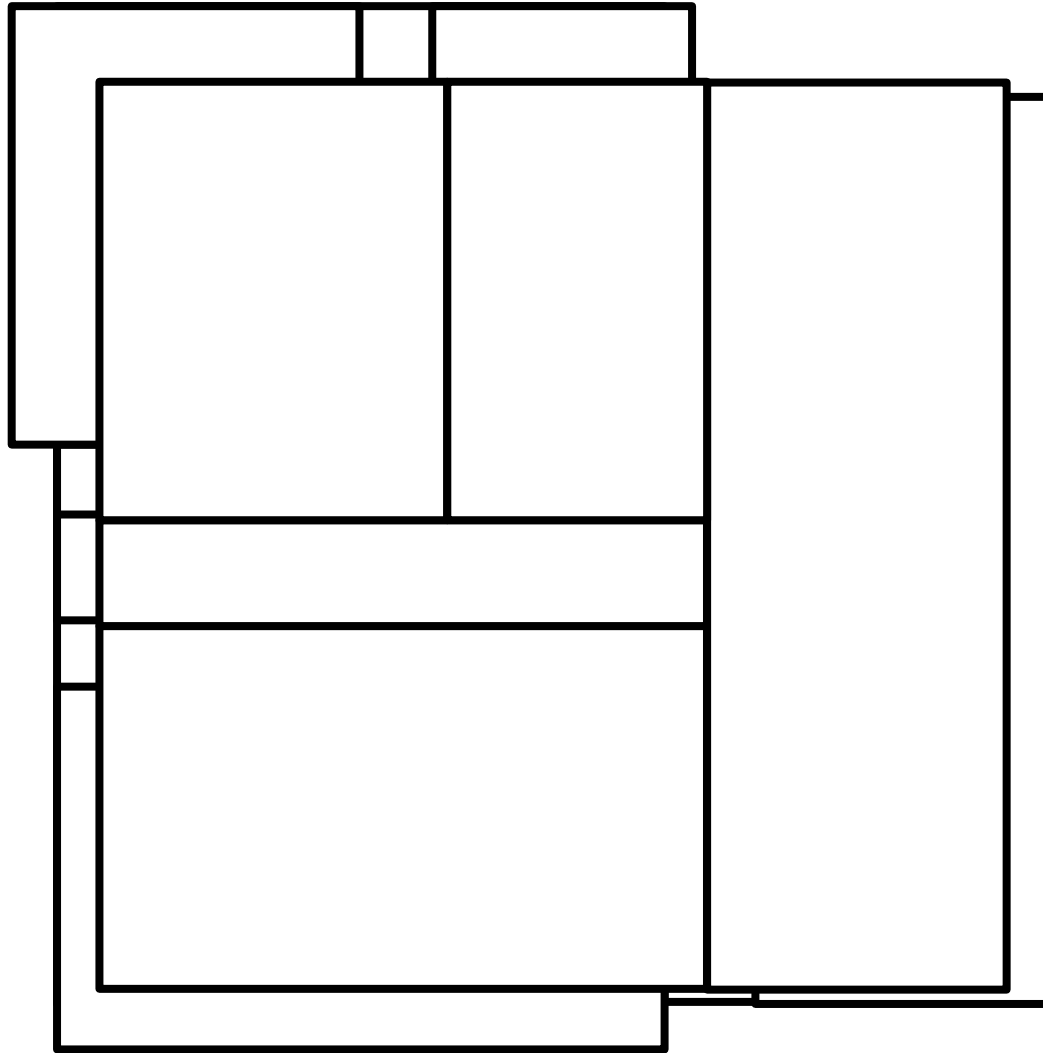
$$\pi_{d,i} \sim \text{DP}(\alpha, \beta_{d,i}^*)$$



楽曲ごとの状態遷移確率

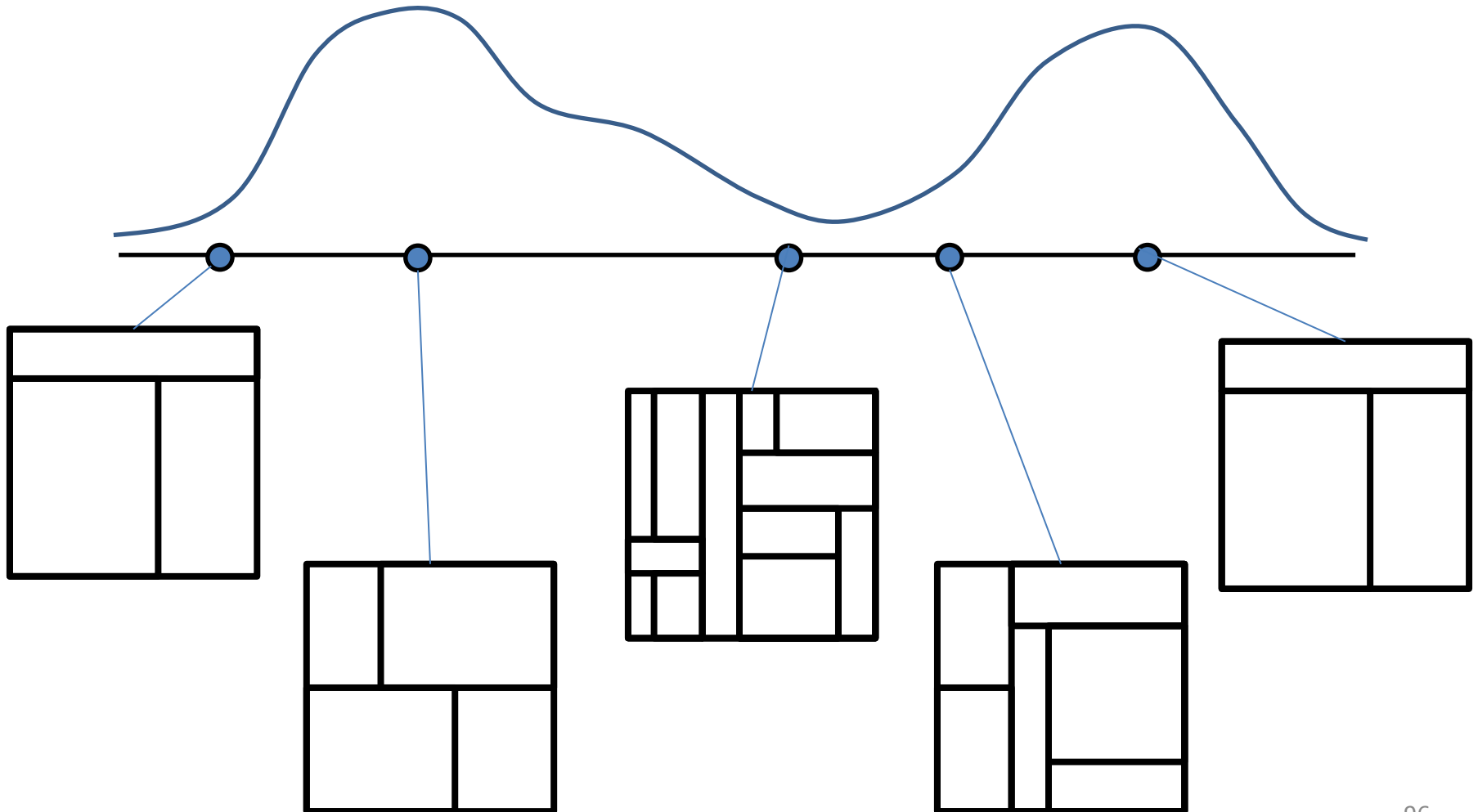
Guillotine partitions

[Gonzales & Zheng1989]

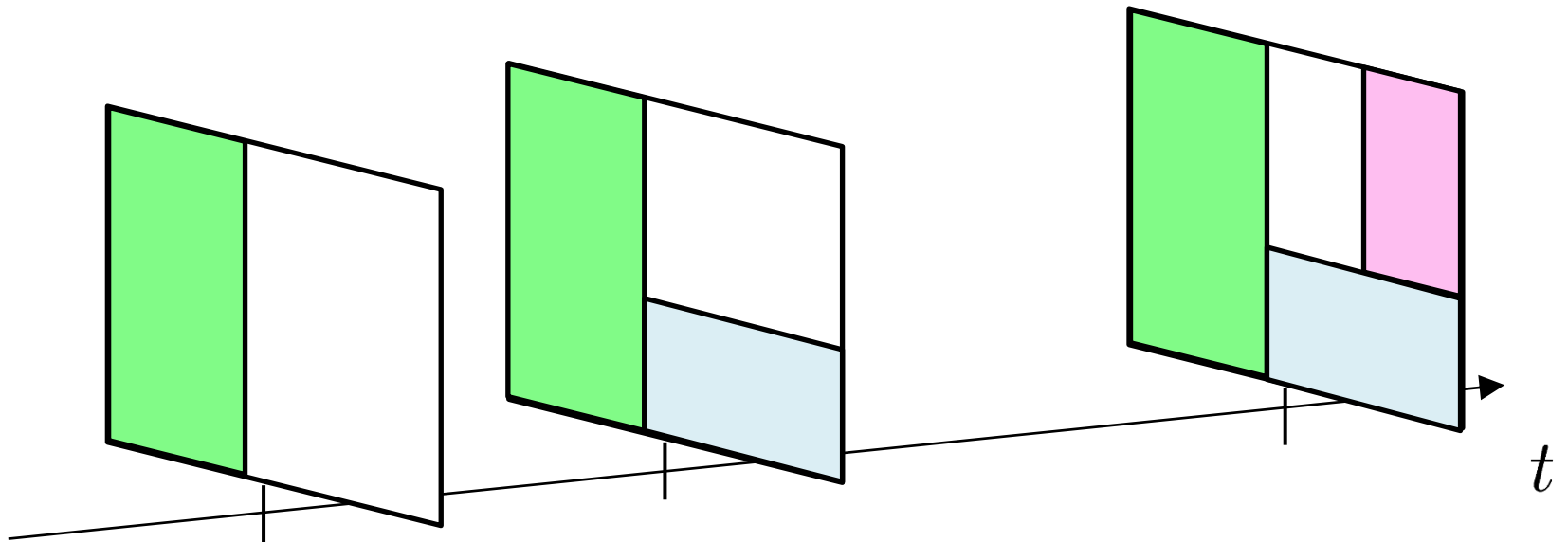


Mondrian process [Roy & Teh2009]

Guillotine partitionsを与える確率過程



Guillotine partitions の発展を表現したマルコフ過程



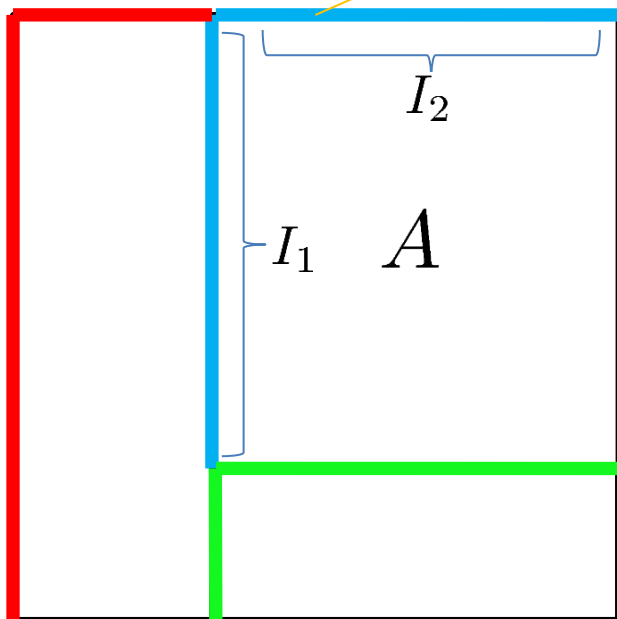
Mondrian process [Roy 2011]

Guillotine partitions の発展を表現したマルコフ過程

レートカーネル $\alpha(\pi, B) = c(\pi) \bar{\nu}_\pi(B \cap R_\pi)$

$$c(\pi) = \sum_{A \in \pi} \sum_d |I_d|$$

カットの起こりやすさ



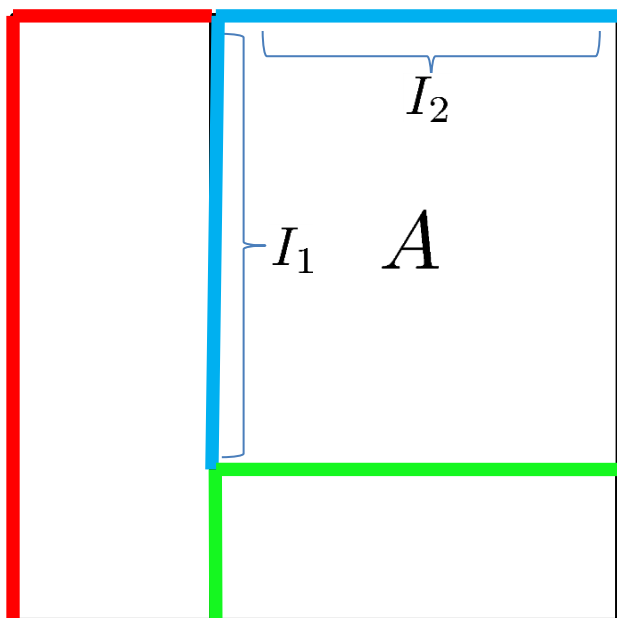
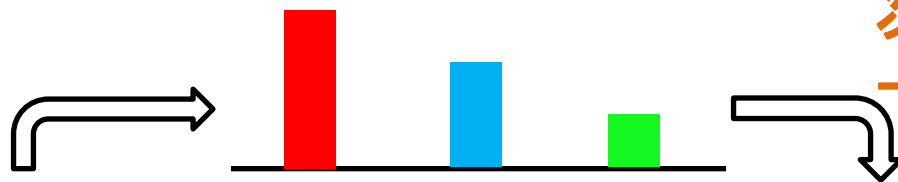
現在のパーティション: π

Mondrian process [Roy 2011]

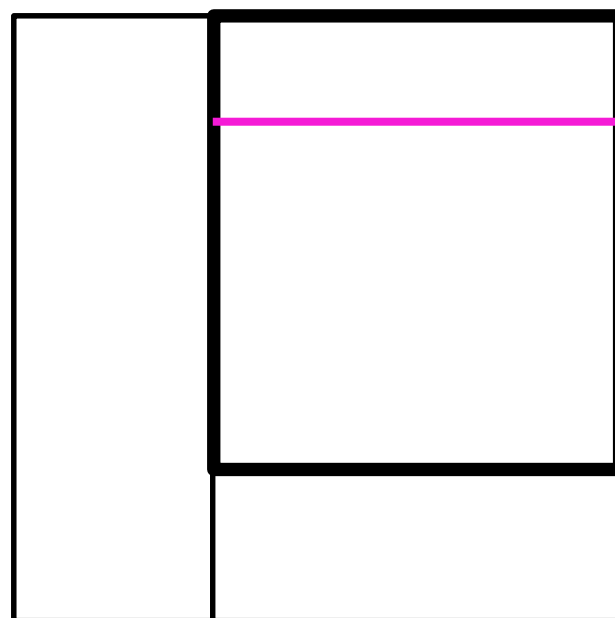
Guillotine partitions の発展を表現したマルコフ過程

レートカーネル $\alpha(\pi, B) = c(\pi) \bar{\nu}_\pi(B \cap R_\pi)$ π から作れる Guillotine partitions

次のpartitionを作る
一様分布



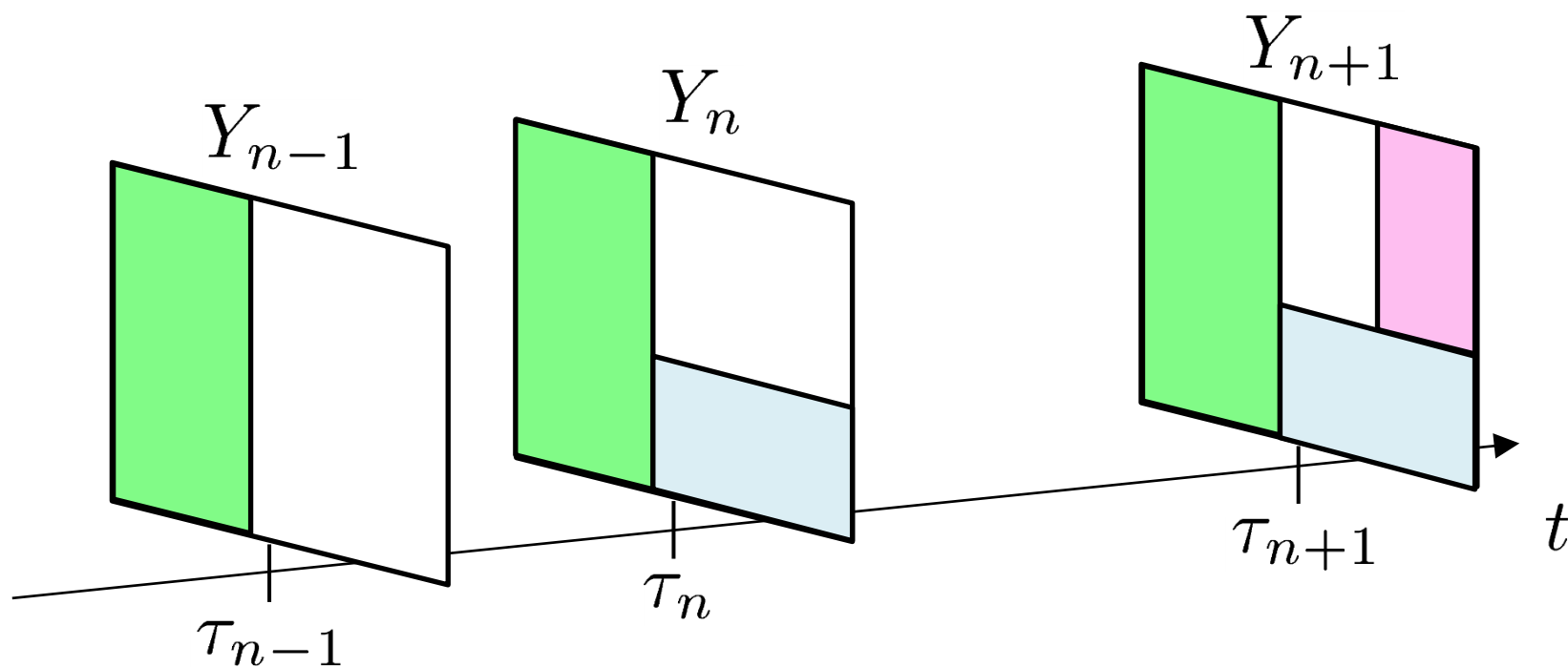
現在のパーティション: π



次のパーティション

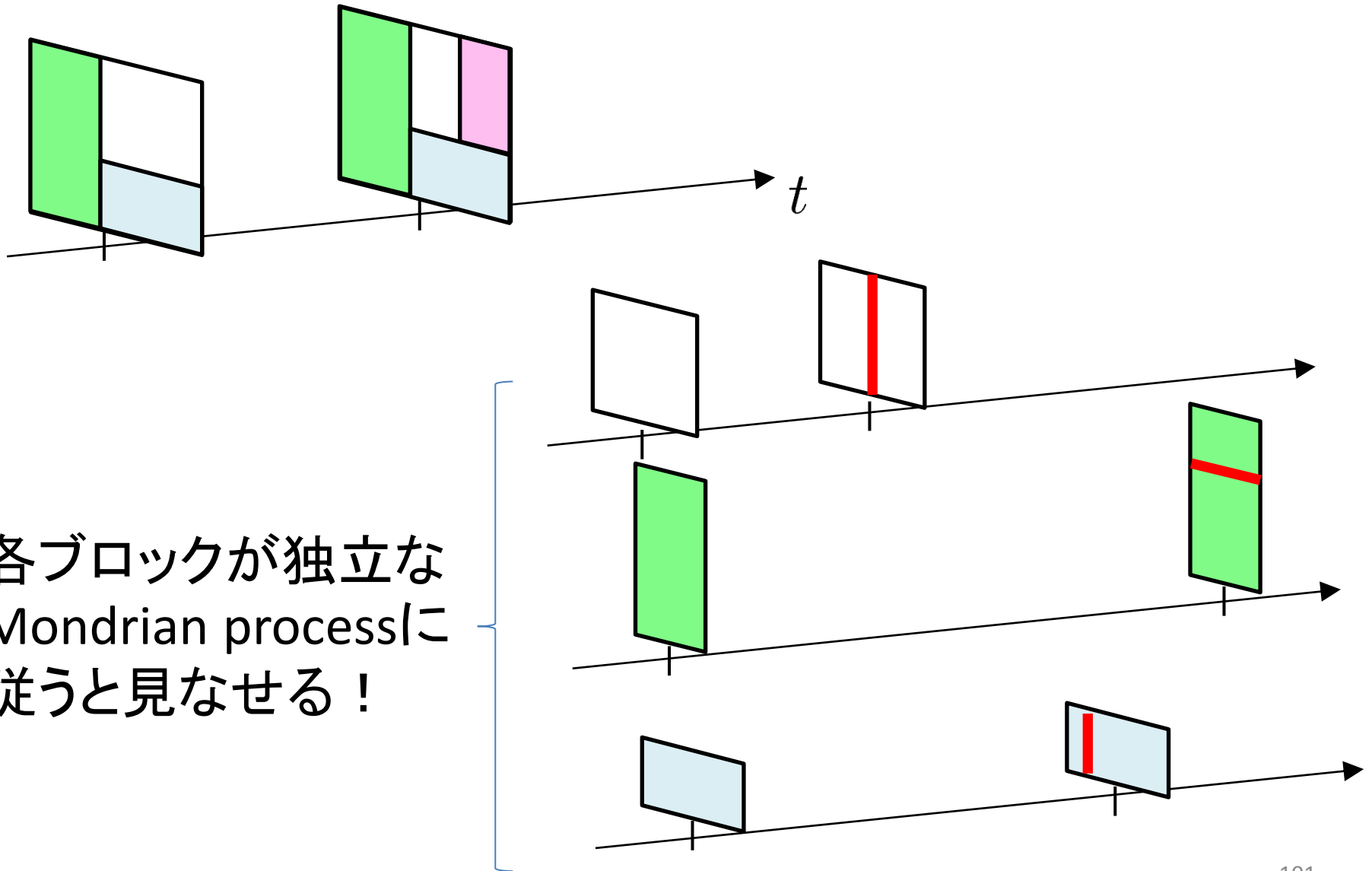
Mondrian process [Roy 2011]

離散時間マルコフ過程



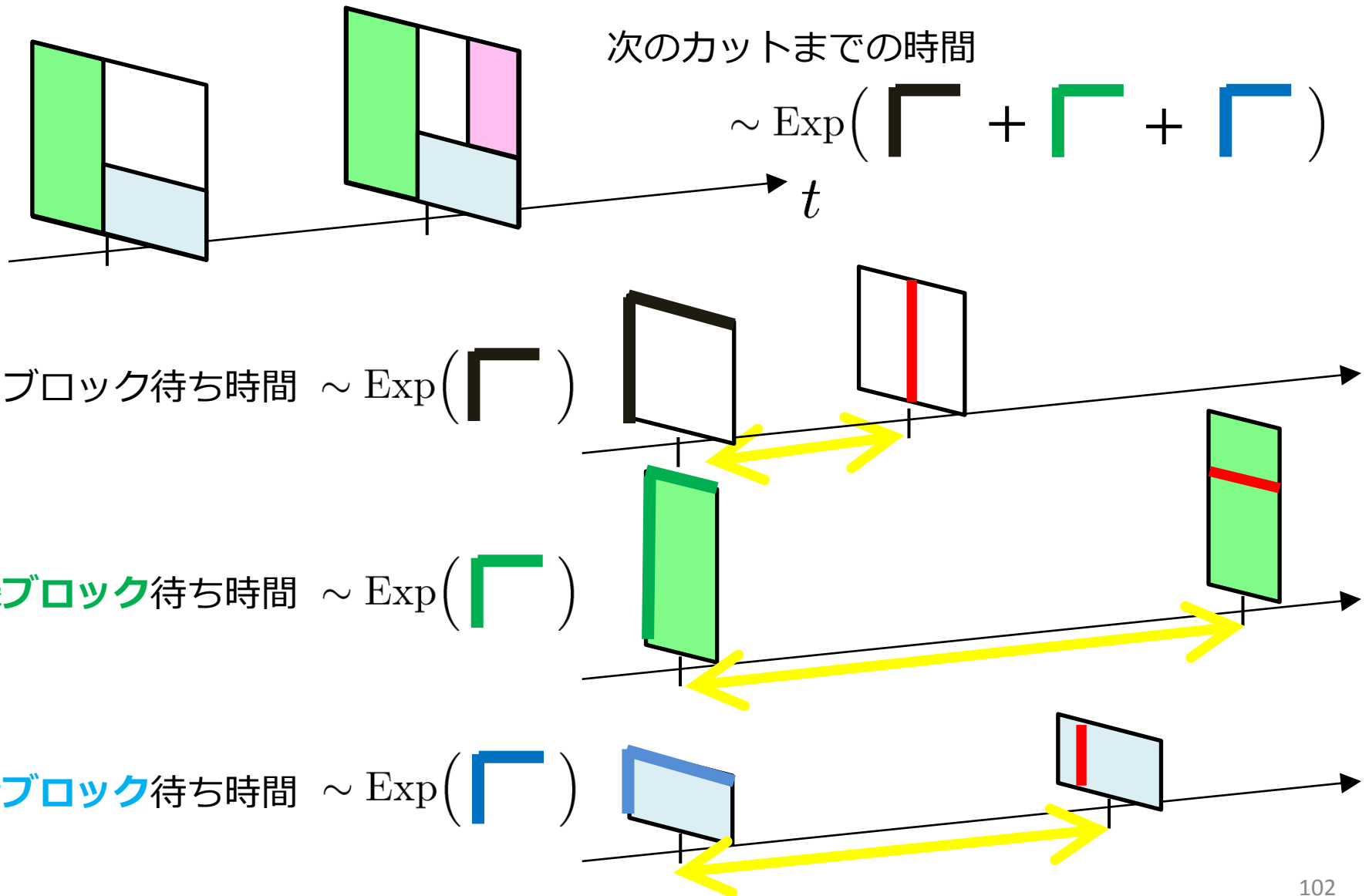
$$\mu_j \sim \text{Exp}(1) \quad \tau_n = \sum_{j=1}^n \frac{\mu_j}{c(Y_{j-1})}$$

Mondrian process の構成法

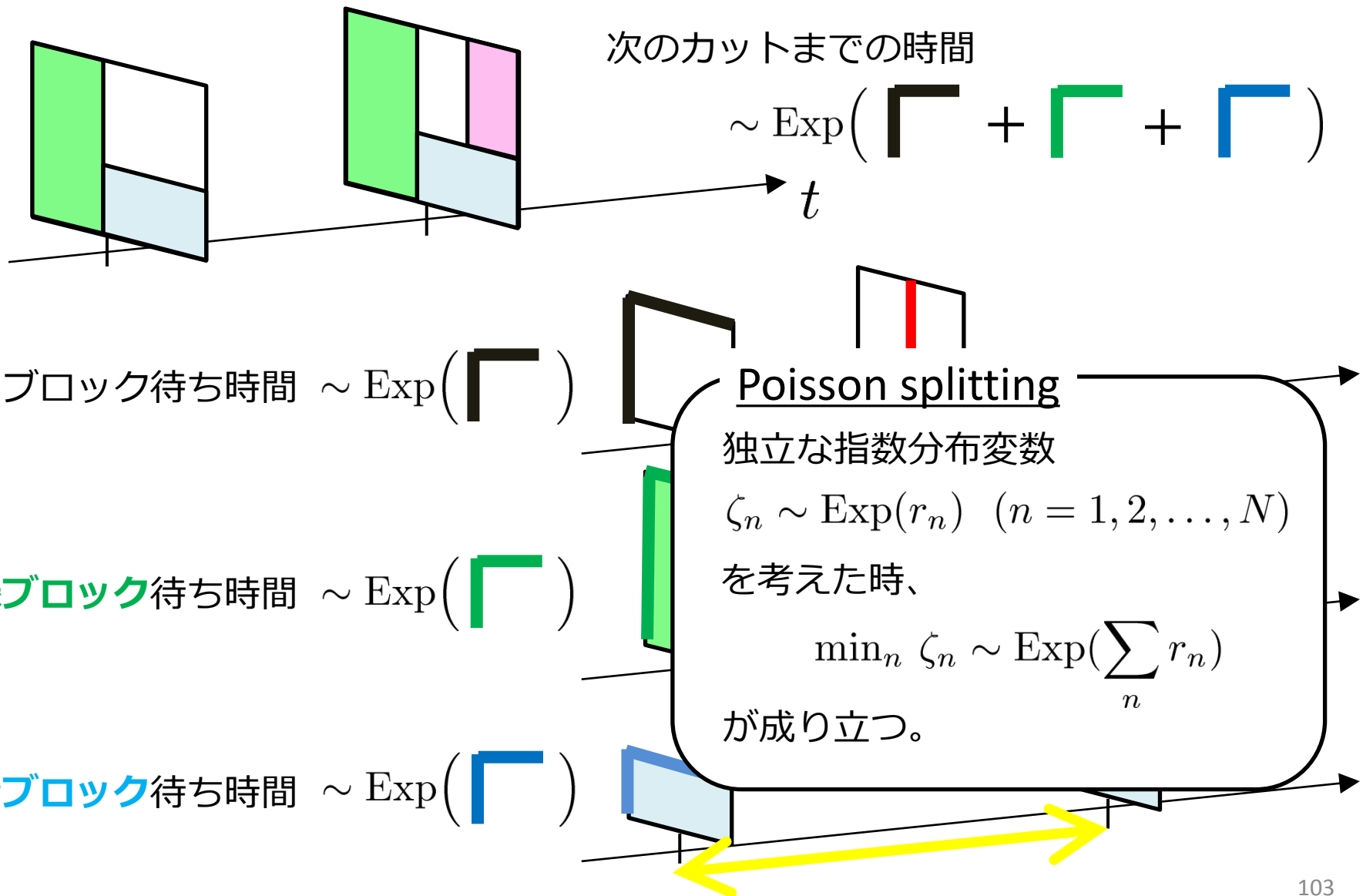


各ブロックが独立な
Mondrian processに
従うと見なせる！

Mondrian process の構成法

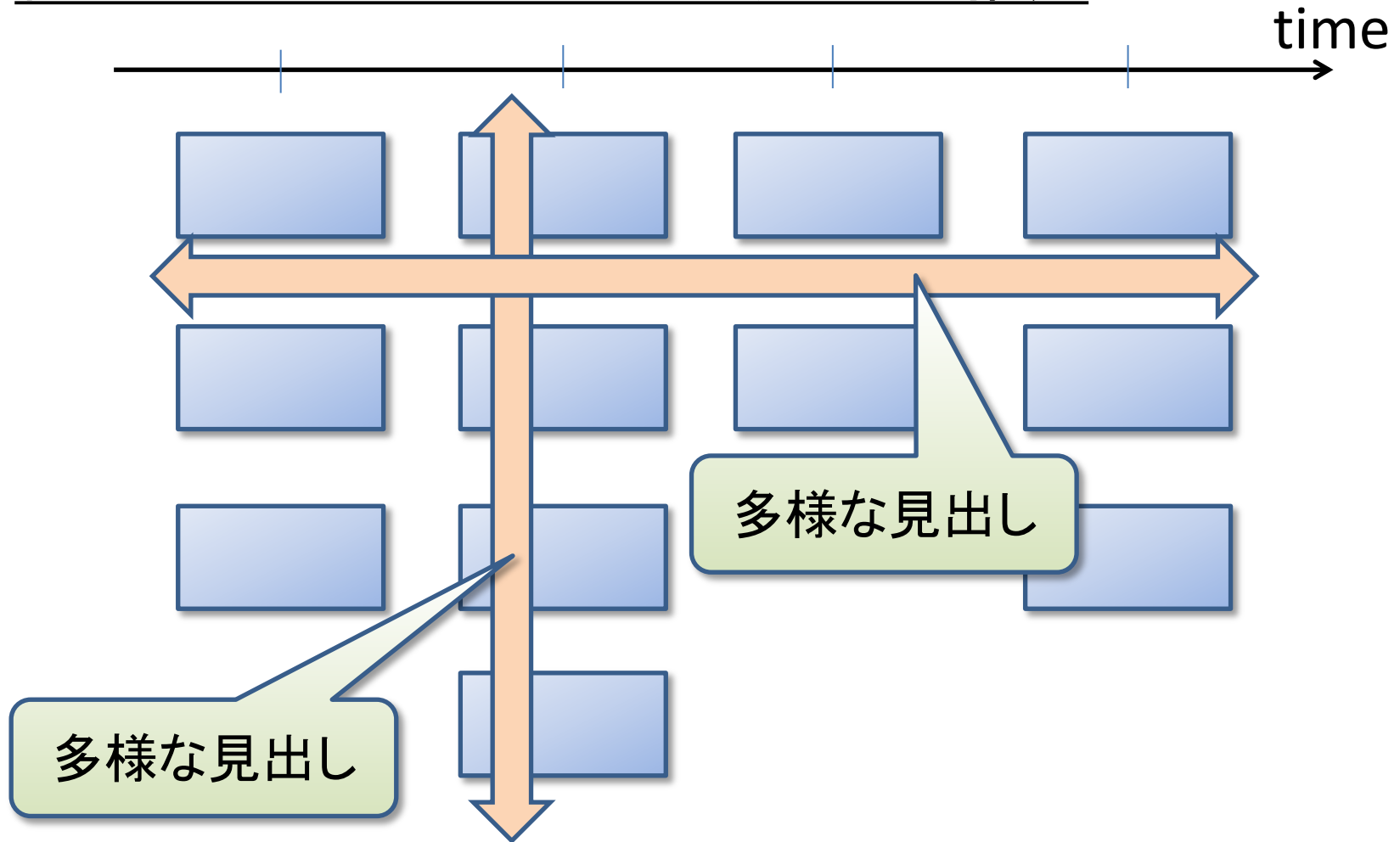


Mondrian process の構成法



例: ニュースヘッドラインの生成 [Affandi+2012]

目的: 日々のニュースヘッドラインの推定



まとめ

2002

2012

既存ツールのノンパラベイズ化

モデルの拡張法

参考文献

- I. V. Shterev and D. B. Dunson, Bayesian watermark attacks, ICML, 2012.
- M. Hughes and E. Sudderth , Nonparametric discovery of active patterns from video collections, CVPR, 2012.
- L. Li, X. Zhang, M. Zhou and L. Carin, Nested Dictionary Learning for Hierarchical Organization of Imagery and Text, UAI, 2012.
- A. Spiliopoulou and A. Storkey, A topic model for melody sequences, ICML, 2012.
- R. H. Affandi, A. Kulesza, and E. B. Fox, Markov determinantal point process, UAI, 2012
- T. S. Ferguson, A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1(2): pp. 209-230, 1973.
- J. Sethuraman, A constructive definition of Dirichlet priors, *Statistica Sinica*: 4, pp. 639-650, 1994.
- J. W. Miller and M. T. Harrison, Dirichlet process mixtures are inconsistent for the number of components in a finite mixture, in *ICERM*, 2012.
- D. J. Aldous, Representations for Partially Exchangeable Arrays of Random Variables, *Journal of Multivariate Analysis*, 11: pp. 581-598, 1981.
- S. G. Walker, Sampling the Dirichlet mixture model with slices, *Communications in Statistics - Simulation and Computation*, 36:45, 2007.
- O. Papaspiliopoulos and G. O. Roberts, Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models, *Biometrika*, 95(1): pp. 169-186, 2008.

参考文献

- Y. Wang and L. Carin, Levy Measure Decompositions for the Beta and Gamma Processes, in *Proc. of ICML*, 2012.
- J. F. C. Kingman, Completely random measure, *Pacific Journal of Mathematics*, vol. 21(1): pp. 59-78, 1967.
- M. I. Jordan, Hierarchical models, nested models and completely random measures, *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger*. New York: Springer, 2009.
- M. Hoffman, D. Blei and P. Cook, Bayesian nonparametric matrix factorization for recorded music in *Proc. ICML*, pp. 641-648, 2010.
- T. Stepleton, Z. Ghahramani, G. Gordon and T. S. Lee, The block diagonal infinite hidden Markov model, in *Proc. of the International Conference on Artificial Intelligence and Statistics*, 2009.
- R. Thibaux, and M. I. Jordan, Hierarchical beta processes and the indian buffet process," in *Proc. of International Conference on Artificial Intelligence and Statistics*, 2007.
- K. A. Heller, Y. W. Teh and D. Gorur, Infinite hierarchical hidden Markov models, in *Proc. of the International Conference on Artificial Intelligence and Statistics*, 2009.
- J. Van Gael, Y. W. Teh and Z. Ghahramani, The infinite factorial hidden Markov model, in *Advances in Neural Information Processing Systems*, 2009.

参考文献

- D. Wingate, N. D. Goodman, D. M. Roy, D and J. B. Tenenbaum, The infinite latent events model," in *Proc. of the International Conference on Uncertainty in Artificial Intelligence*, 2009.
- F. Doshi-Velez, D. Wingate, N. Roy and J. Tenenbaum, Infinite dynamic Bayesian networks," in *Proc. of International Conference in Machine Learning*, 2011.
- Y. W. Teh, M. I. Jordan, M. Beal and D. Blei, Hierarchical Dirichlet processes, *Journal of the American Statistical Association*, 101, 1566-1581, 2006.
- M. Beal, Z. Ghahramani and C. Rasmussen, The infinite hidden Markov model, in *Advances in Neural Information Processing Systems*, 2002.
- D. M. Blei, A. Y. Ng and M. I. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3:993-1022, 2003.
- P. Liang, S. Petrov, M. I. Jordan, and D. Klein, The infinite PCFG using hierarchical Dirichlet processes, " in *Proc. of EMNLP*, pp. 688-697, 2007.
- H. Kameoka, K. Ochiai, M. Nakano, M. Tsuchiya, S. Sagayama, Context-free 2D tree structure model of musical notes for Bayesian modeling of polyphonic spectrograms," in *Proc. of ISMIR*, 2012.
- M. Nakano, Y. Ohishi, H. Kameoka, R. Mukai, K. Kashino, Bayesian nonparametric music parser," in *Proc. of ICASSP*, 2011.

参考文献

- D. Mochihashi and E. Sumita, The infinite Markov model, in *Advances in Neural Information Processing Systems*, 2007.
- F. Wood, C. Archambeau, J. Gasthaus, L. F. James and Y. W. Teh, A Stochastic Memoizer for Sequence Data, in *Proc. of ICML*, 2009.
- A. Rodriguez, D. B. Dunson and A. E. Gelfand, The nested Dirichlet process," *Journal of American Statistics Association* 103, 1131-1154, 2008.
- A. Rodriguez and K. Ghosh, Nested partition models, *Jack Baskin School of Engineering*, Technical report, 2009.
- E. B. Fox, E. B. Sudderth, M.I. Jordan, A.S. Willsky, A Sticky HDP-HMM with Application to Speaker Diarization, *Annals of Applied Statistics*, 2011.
- Lu Ren, Y. Wang, D. Dunson, and L. Carin, The Kernel Beta process, in *Proc. NIPS*, 2011.