NMF vs PLCA: 多重音生成過程に対する 無限因子モデルと無限混合モデル

吉井 和佳　糸山 克寿　中村 栄太 (京大)
後藤 真孝 (産総研)

第112回情報処理学会音楽情報科学研究会 (SIGMUS 112)

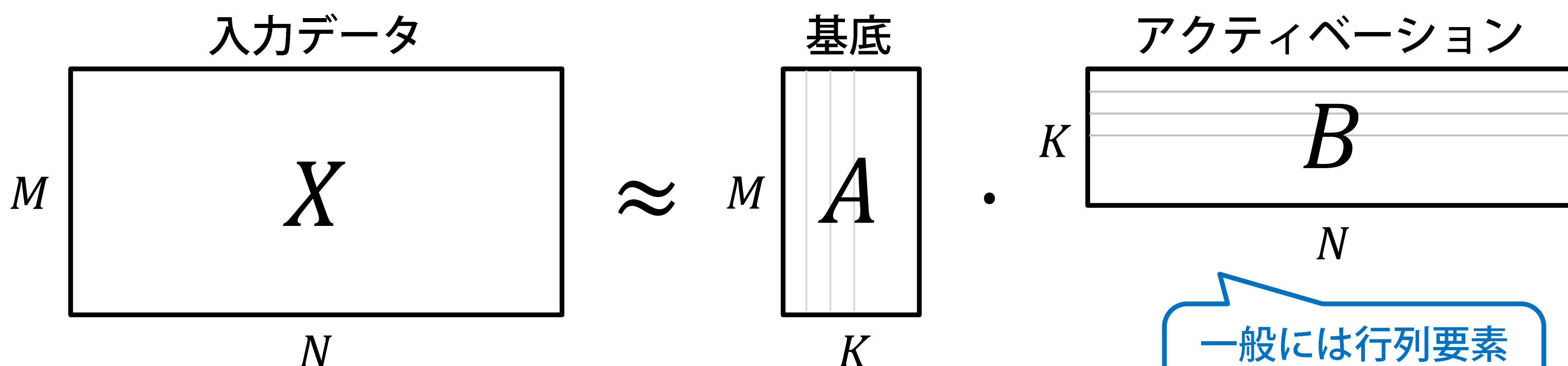
研究の背景

- 音楽音響信号の音源分離には行列因子分解が広く利用
  - 非負値行列分解 (nonnegative matrix factorization, NMF)
    - 画像処理分野で提案 [Lee@NIPS 2000]
    - 音楽音響信号の分離のために応用 [Smaragdis@WASPAA 2003]
  - 確率的潜在要素解析 (probabilistic latent component analysis, PLCA)
    - 自然言語処理分野で確率的潜在意味解析 (PLSA) が提案 [Hofmann@SIGIR 1999]
    - 音楽音響信号の分離のためにPLSAを拡張 [Smaragdis@NIPS 2006]
  - 確率モデルの定式化が可能
    - NMFもPLCAもある確率モデルの最尤推定として解釈可能
    - (ノンパラメトリック)ベイズモデルの定式化も可能

確率モデルを考えることで、手法の理論的な妥当性を議論できる！

行列因子分解 (matrix factorization)

- ある行列を二つの行列の積で近似 (低ランク近似)
  - 行列に関する制約やコスト関数の違いでさまざまな変種が存在
    - 主成分分析 (PCA)：基底が直交
    - 独立成分分析 (ICA)：基底が独立

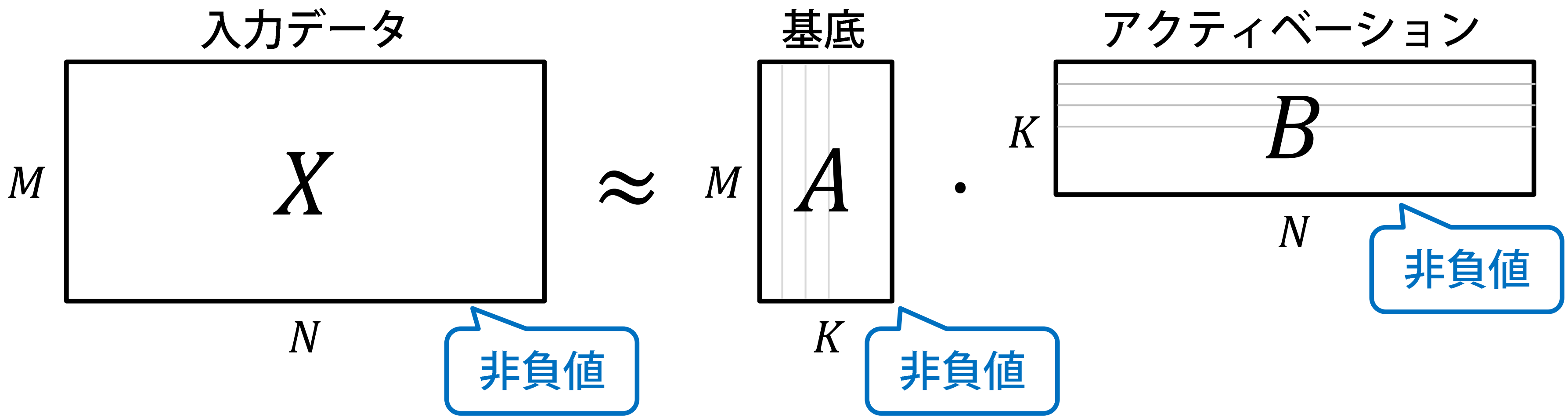


一般には行列要素
はなんでもよい

$K \ll M, N$とすることにより、冗長性が除去され、
データXを構成する「基底」が教師なし学習できる

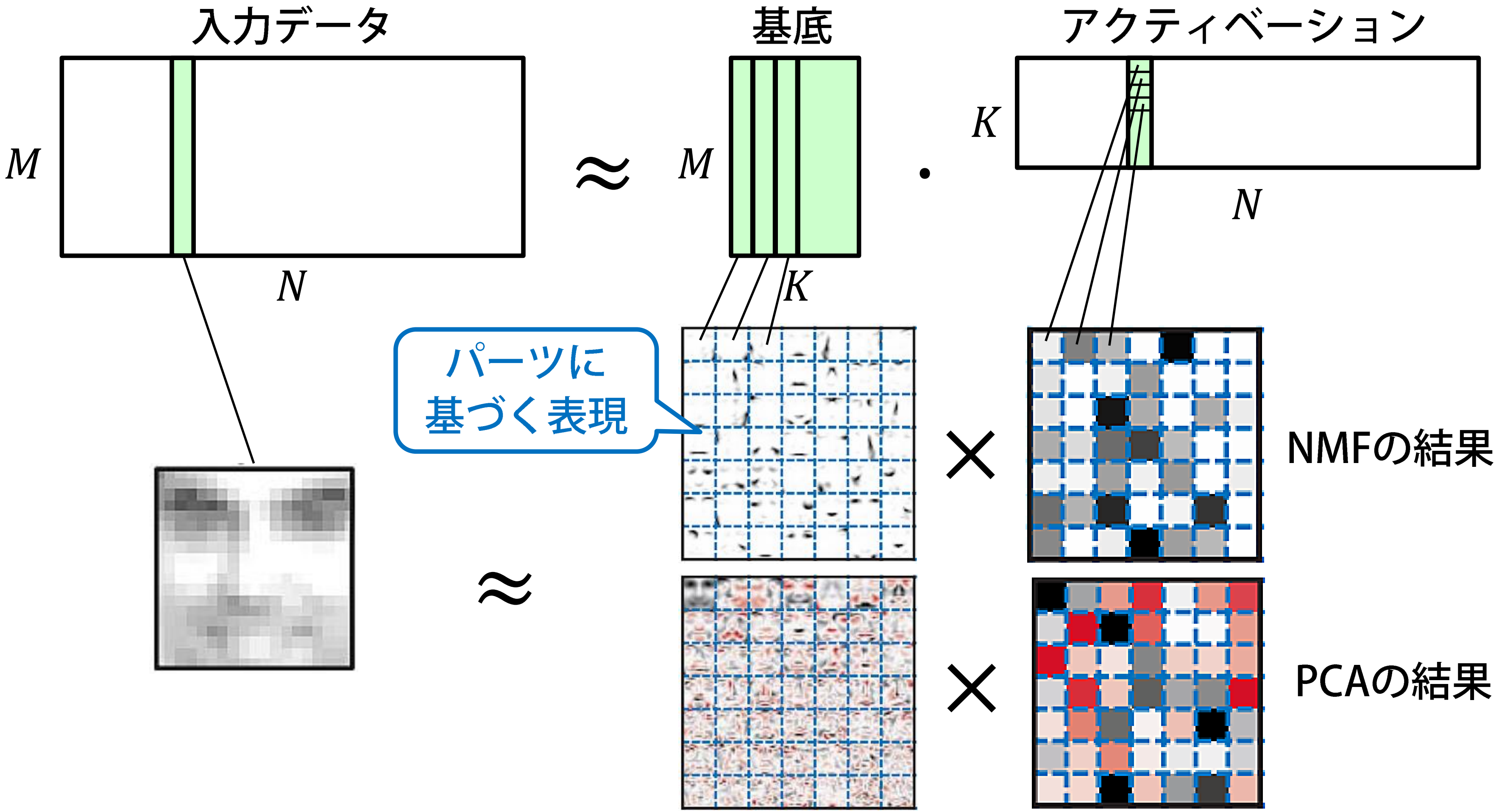
非負値制約に基づく行列分解

- ある非負値行列を二つの非負値行列の積で近似
  - 非負値行列因子分解 (NMF)
  - 確率的潜在成分分解析 (PLCA)

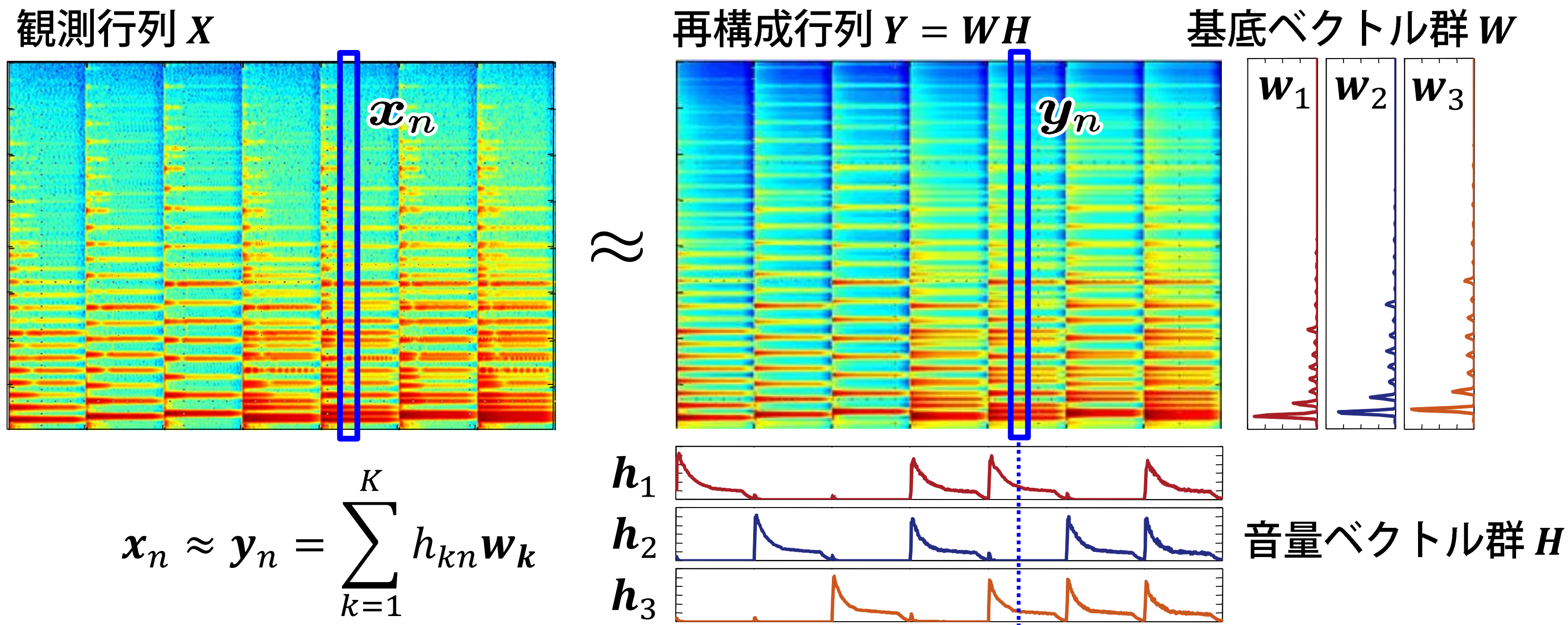


非負値制約により、基底の足し合わせのみで入力データを表現
→「パーツ」に基づく分解表現

NMF vs PCA



非負値行列因子分解 (NMF)



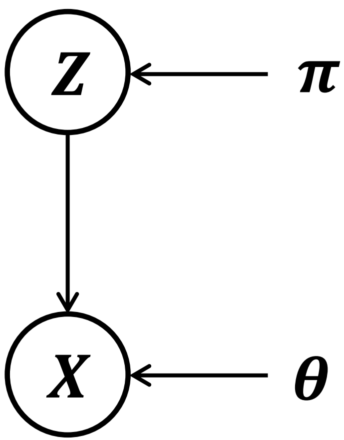
$$x_n \approx y_n = \sum_{k=1}^{K} h_{kn} w_k$$

疑問：コスト関数 $C(x_n|y_n)$ はどのように設計すればよいか？

確率モデル (観測データの生成過程) を考えることが重要！

潜在変数モデル

- 「観測変数」の背後に「潜在変数」が存在することを仮定
  - 現実の問題で利用される確率モデルはほとんどがこのタイプ
  - 段階的なデータの生成過程を表現



(1) 潜在変数を確率的に生成（$\boldsymbol{\pi}$はパラメータ）
$p(\boldsymbol{Z}|\boldsymbol{\pi})$

(2) 観測変数を確率的に生成（$\boldsymbol{\theta}$はパラメータ）
$p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\theta})$

(0) ベイズモデルではパラメータも確率的に生成
$p(\boldsymbol{\pi})p(\boldsymbol{\theta})$

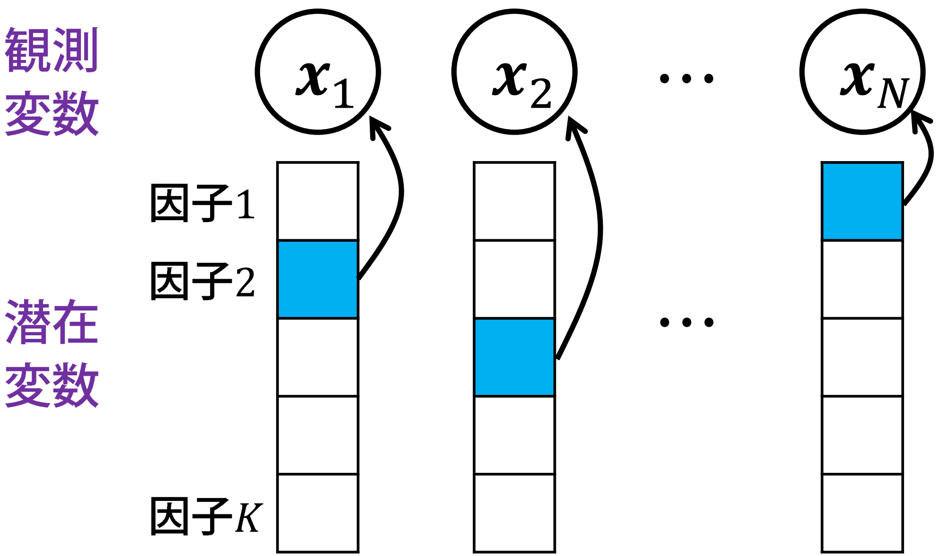
事後分布の計算が目標

$$p(\boldsymbol{Z},\boldsymbol{\pi},\boldsymbol{\theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\pi},\boldsymbol{\theta})}{p(\boldsymbol{X})} = \frac{p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\theta})p(\boldsymbol{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi})p(\boldsymbol{\theta})}{p(\boldsymbol{X})}$$

混合モデルと因子モデル

混合モデル (mixture model)

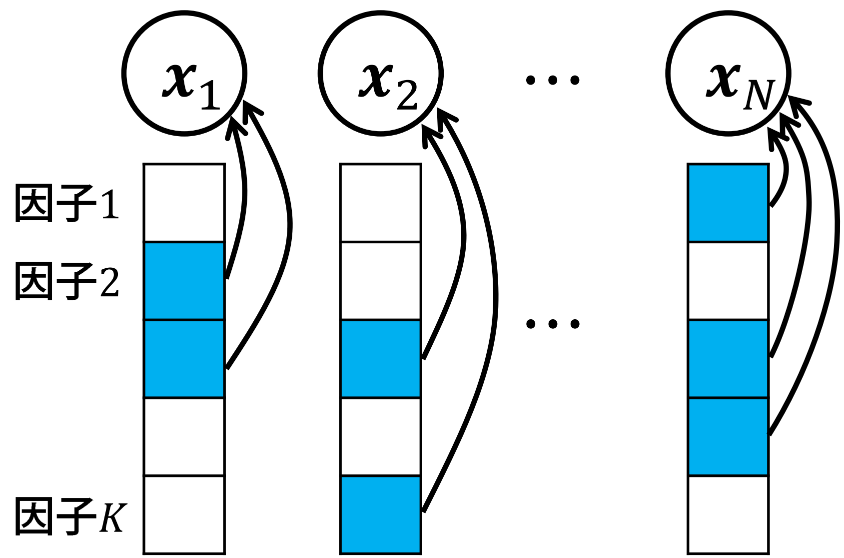
あるサンプルはいずれか一つの
因子のみに依存する



「分類」のための確率モデル

因子モデル (factor model)

各サンプルは複数の因子の
組み合わせで得られる



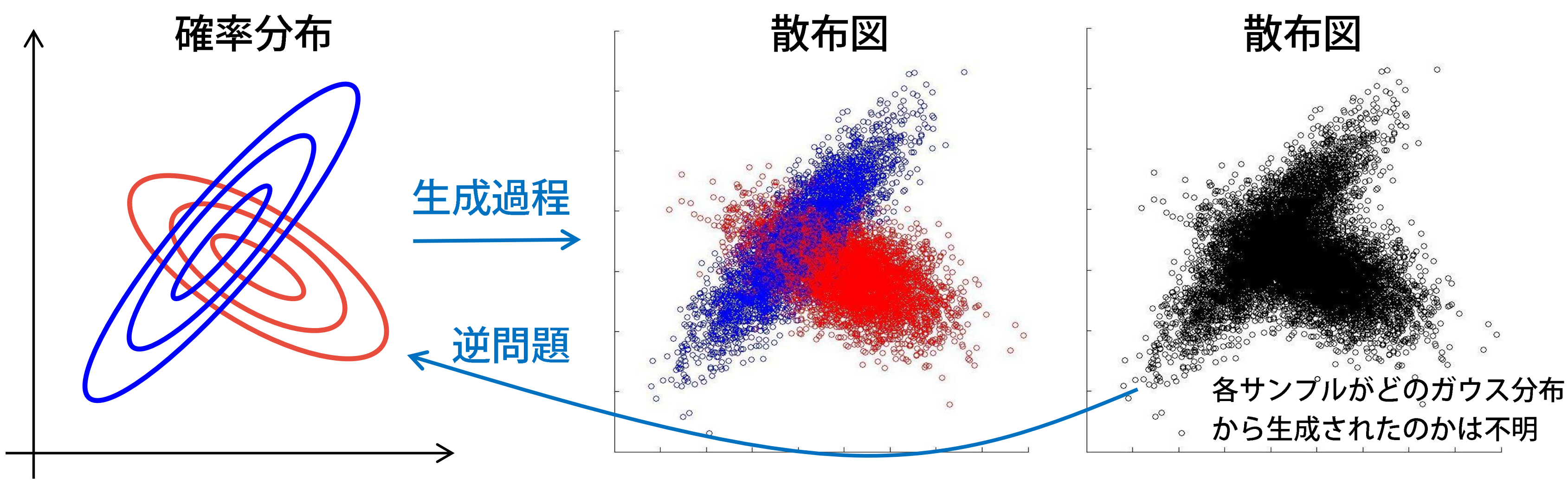
「分解」のための確率モデル

混合モデルの例

- 混合ガウスモデル (GMM)
  - K個のガウス分布を準備
    - パラメータ：$\boldsymbol{\theta} = [(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \cdots, (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)]$
  - ガウス分布を確率的に選択してサンプルを生成
    - パラメータ：$\boldsymbol{\pi} = [\pi_1, \cdots, \pi_K]$

$$\boldsymbol{x}_n \sim \sum_{k=1}^{K} \pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

和が確率分布の外側
→ 確率分布の足し合わせ

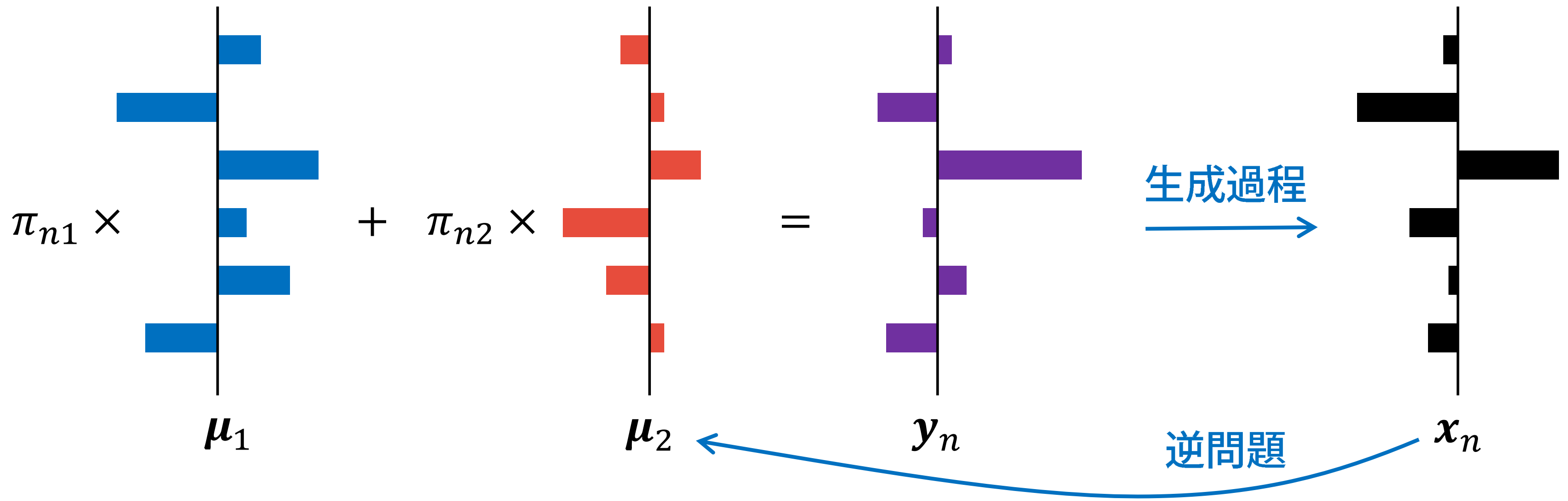


因子モデルの例

- 主成分分析 (PCA)・因子分析 (FA)
  - K個の基底ベクトルを準備
    - パラメータ：$\boldsymbol{\theta} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_K]$
  - 重み付きで足し合わせてからサンプルを生成
    - パラメータ：$\boldsymbol{\pi} = [\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_N]$

$$\boldsymbol{x}_n \sim N\left(\sum_{k=1}^{K} \pi_{nk} \boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)$$

和が確率分布の内側
→ 確率変数の足し合わせ



混合モデルと因子モデル

混合モデル (mixture model)

あるサンプルはいずれか一つの
因子のみに依存する

$$\boldsymbol{x}_n \sim \sum_{k=1}^{K} \pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

和が確率分布の外側
→ <u>確率分布の足し合わせ</u>

各サンプルを
「分類する」ための確率モデル

因子モデル (factor model)

各サンプルは複数の因子の
組み合わせで得られる

$$\boldsymbol{x}_n \sim N\left(\sum_{k=1}^{K} \pi_{nk} \boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)$$

和が確率分布の内側
→ <u>確率変数の足し合わせ</u>

各サンプルを
「分解する」ための確率モデル

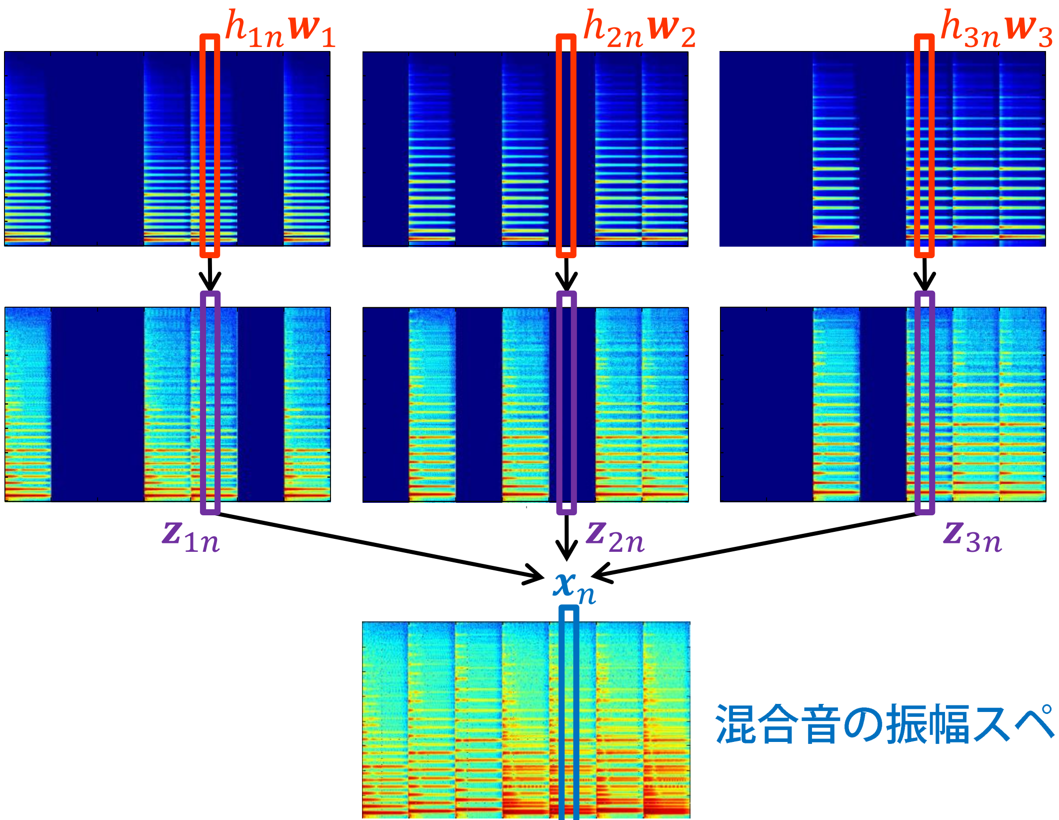
音源分離にはどちらを使う？

- 理論的には「因子モデル」が妥当 → NMF
  - 各フレームにおいて、混合音のスペクトル(「サンプル」)は、複数の音源信号のスペクトルが重畳することによって得られる
    - 各フレームが複数の音源に分解される
- 現実には「混合モデル」も提案 → PLCA
  - 何を「サンプル」ととらえるかがキモ
    - 各「サンプル」はいずれかの音源に分類される
    - 各フレームでは複数の音源が同時に存在できるようにしたい

何らかのトリックが必要

NMFの確率モデル

- Kullback-Leibler NMF (KL-NMF)：ポアソン分布に基づく因子モデル



各音源の振幅スペクトル密度

$$\boldsymbol{z}_{kn} \sim \mathrm{Poisson}(h_{kn}\boldsymbol{w}_k)$$

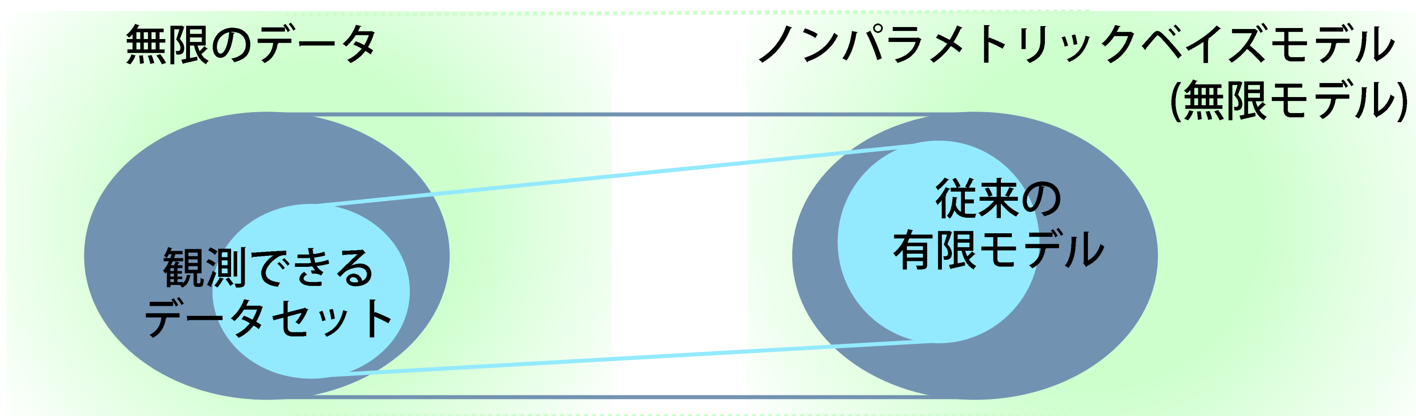
各音源の振幅スペクトル (潜在変数)

$$\boldsymbol{x}_n \sim \mathrm{Poisson}\left(\sum_{k=1}^{K} h_{kn}\boldsymbol{w}_k\right)$$

混合音の振幅スペクトル (観測変数)

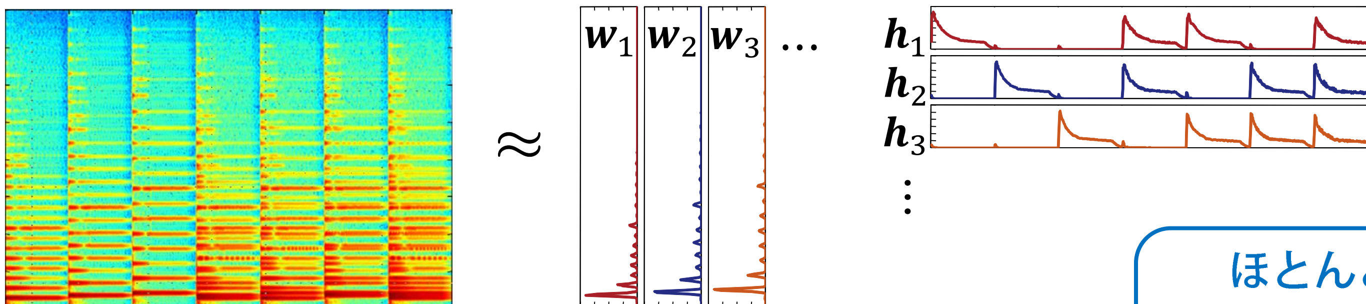
NMFのノンパラメトリックベイズ拡張

- 無限個の基底をもつNMFを定式化可能
  - 観測データに合わせて「実効的な」基底数K^+が推定される
    - 無限にデータがあれば無限個の基底が必要
    - 手持ちデータが有限であれば、高々有限個の基底ですむ
- 因子モデルを無限化する際の主な方針は二つ
  - ガンマ過程 (gamma process) とベータ過程 (beta process)



Gamma Process NMF [Hoffman@ICML 2010]

- 無限個の基底に対する重みを導入してスパース学習
  - ガンマ過程 (GaP)：無限次元の非負値ベクトルに対する事前分布

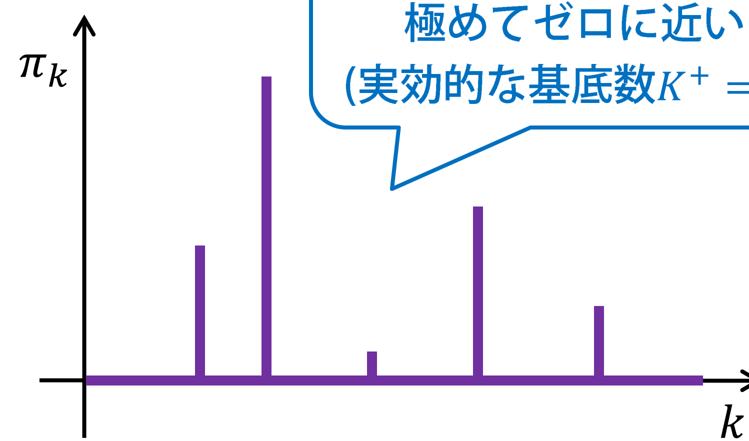


$$x_n \sim \mathrm{Poisson}\left(\sum_{k=1}^{\infty} \pi_k h_{kn} w_k\right)$$

$$\pi \sim \mathrm{GaP}(\alpha)$$

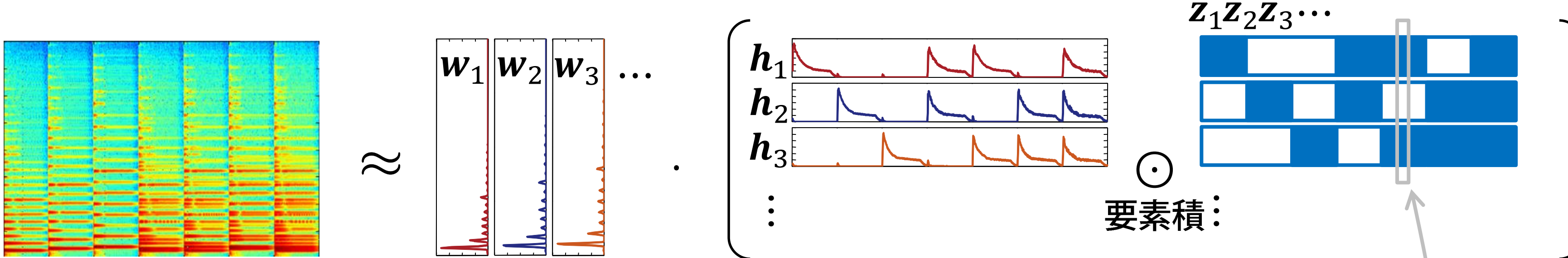
集中度αが大きいほど
K^+が大きくなる

ほとんどの重みが
極めてゼロに近い
(実効的な基底数$K^+ = 5$)



Beta Process NMF [Liang@NIPS Workshop 2014]

- 無限個のバイナリ変数を導入してスパース学習
  - ベータ過程 (BP)：無限次元のコイントス確率ベクトルに対する事前分布
  - ベルヌイ過程 (BeP)：無限次元のバイナリベクトルに対する事前分布

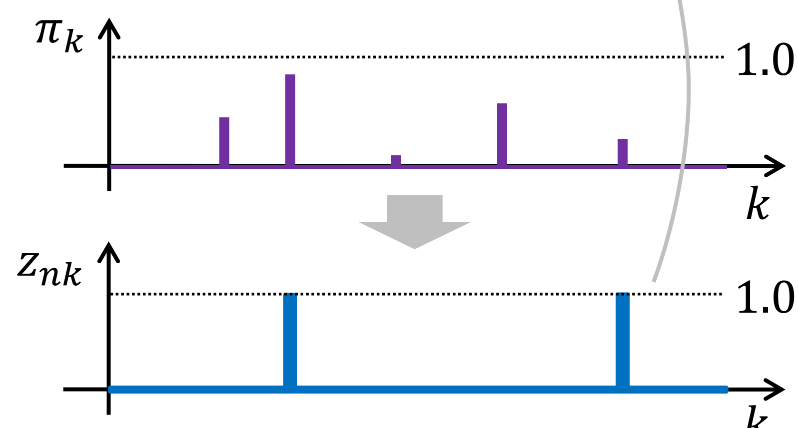


$$\boldsymbol{x}_n \sim \mathrm{Poisson}\left(\sum_{k=1}^{\infty} z_{nk} h_{kn} \boldsymbol{w}_k\right)$$

$$\boldsymbol{\pi} \sim \mathrm{BP}(\alpha)$$

$$\boldsymbol{z}_n \sim \mathrm{BeP}(\boldsymbol{\pi})$$

集中度αが大きいほど
K^+が大きくなる



ノンパラメトリックベイズNMFの問題

- 最大基底数での打ち切り近似が必要な解法しか提案されていない
  - 最初に十分に多くの基底を準備 (例えば$K = 100$)
  - 計算を反復するうちに不要になった基底を削除していく

GaP-NMF
K^+の推定には閾値処理が必要

$$\boldsymbol{x}_n \sim \mathrm{Poisson}\left(\sum_{k=1}^{\infty} \pi_k h_{kn} \boldsymbol{w}_k\right)$$

本当は「非負値」ではなくて
「正値」でないとダメ
(ゼロは許容されない)

BP-NMF
組み合わせ最適化は局所解が多い

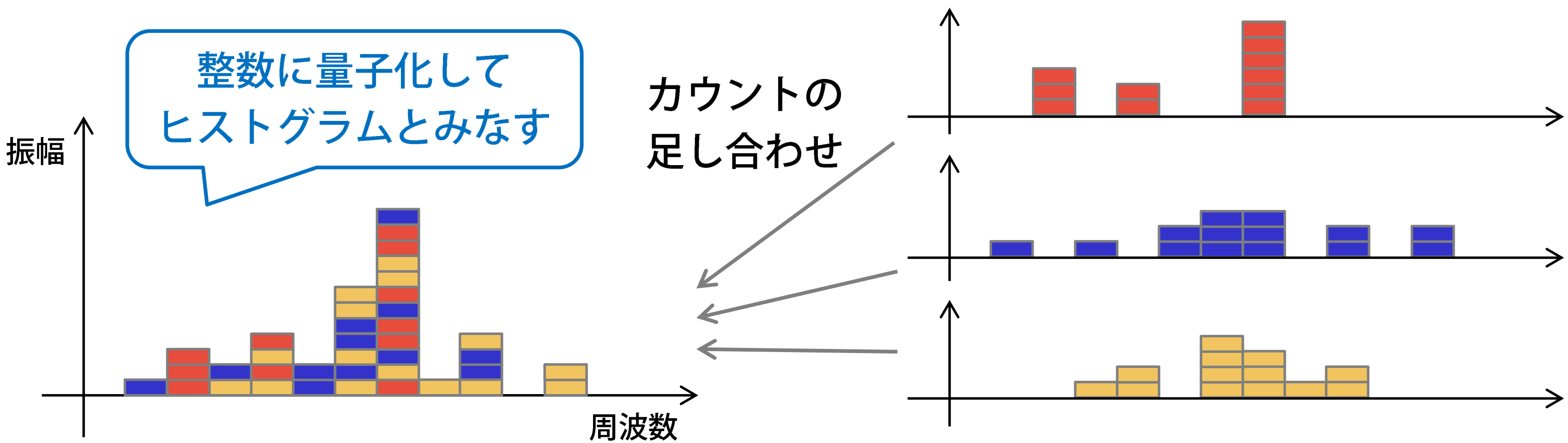
$$\boldsymbol{x}_n \sim \mathrm{Poisson}\left(\sum_{k=1}^{\infty} z_{nk} h_{kn} \boldsymbol{w}_k\right)$$

基底数をKとすると
混合モデルではK通りの可能性
因子モデルでは2^K通りの可能性

混合モデルに基づく音源分離

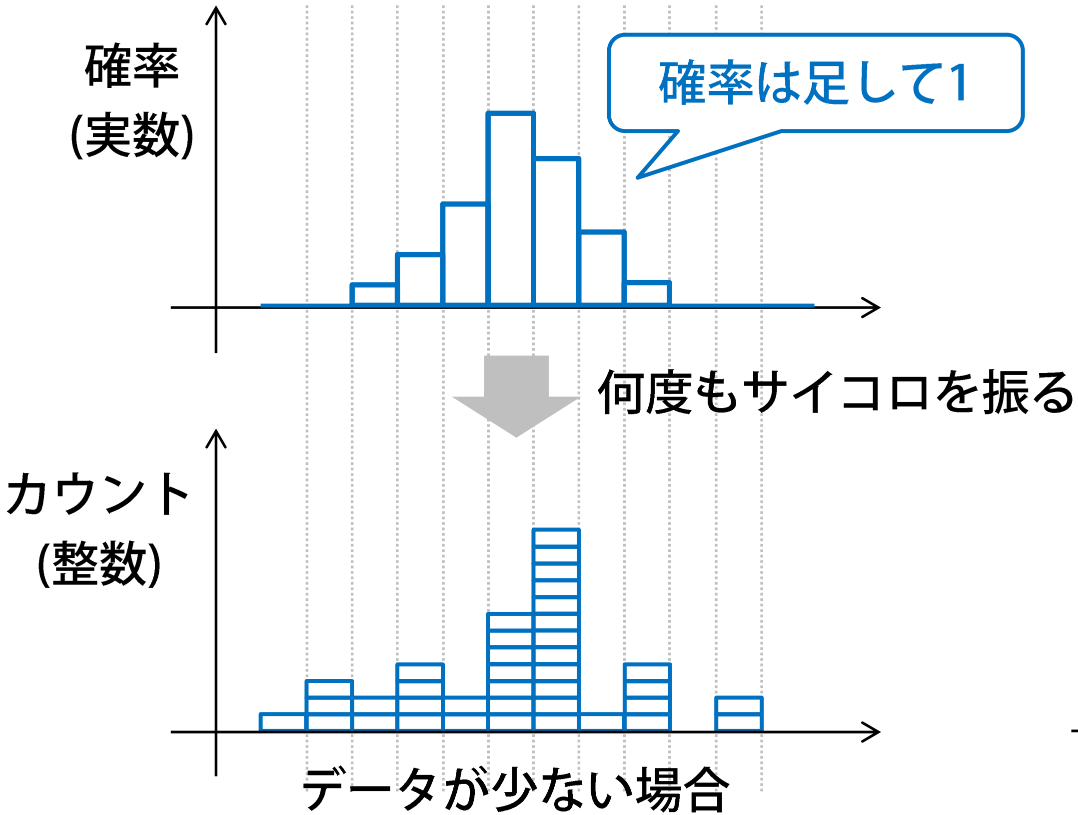
- 混合モデルを音源分離に使うにはトリックが必要
  - 各「サンプル」はいずれかの音源に分類される
  - 各フレームでは複数の音源が同時に存在できるようにしたい

混合音のスペクトル＝「サンプル」の集合
スペクトル全体としてみたときには、複数の音源が含まれる

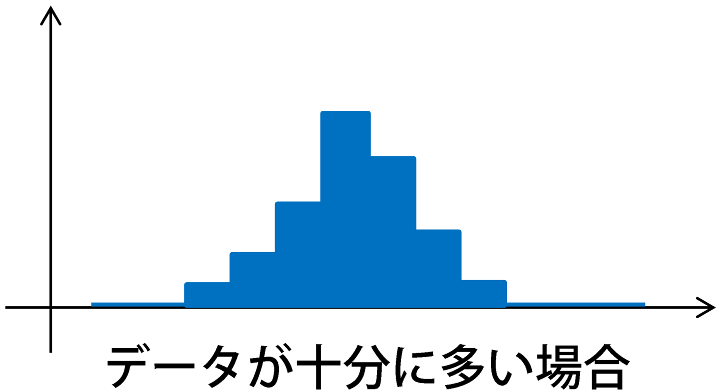


離散分布とヒストグラム

- ヒストグラムは離散分布 (サイコロ) から確率的に生成されると解釈

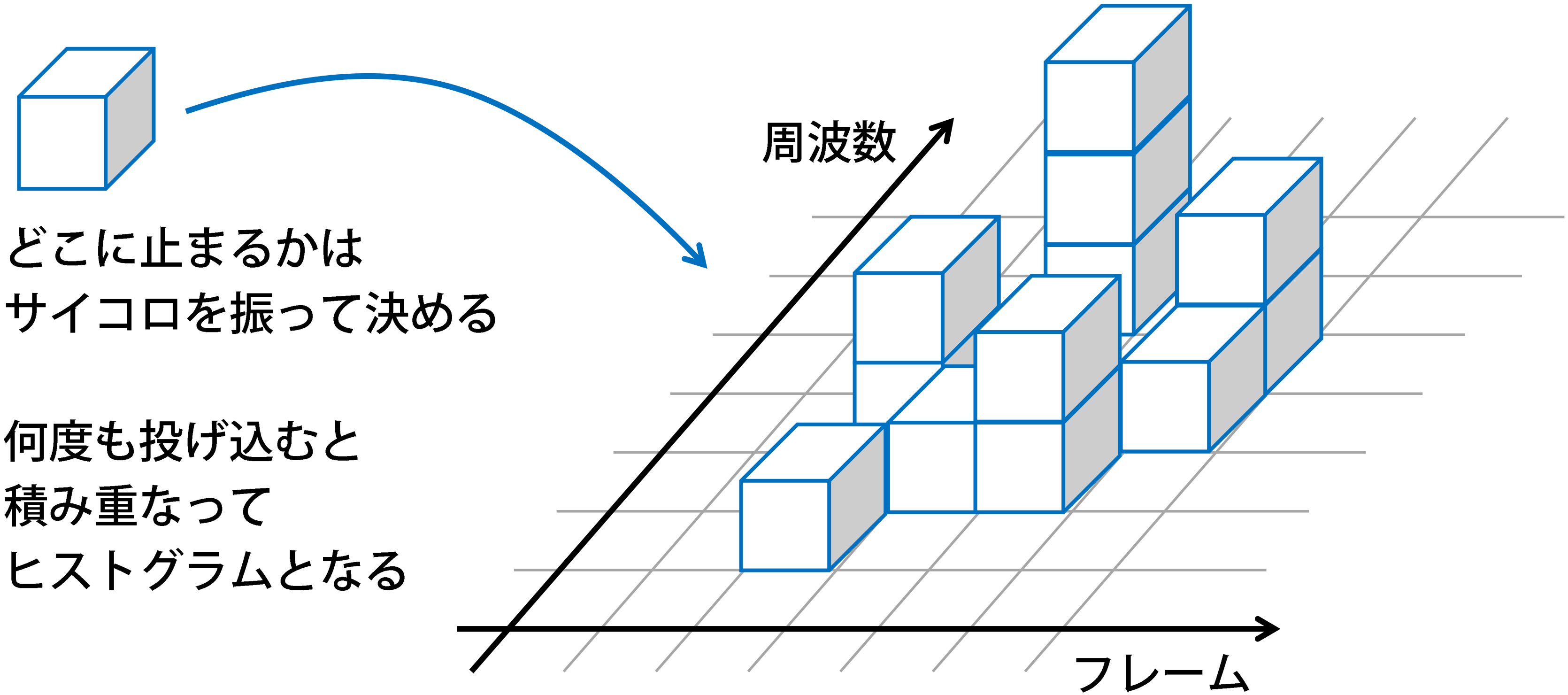


得られるヒストグラムには
「ゆらぎ」が含まれている
↓
最尤推定ではなく
ベイズ推定の必要性



二次元平面上の離散分布

- 音量子 (sound quantum) を投げ入れると、平面内のどこかで止まる
  - 二次元平面はグリッドに分割されている

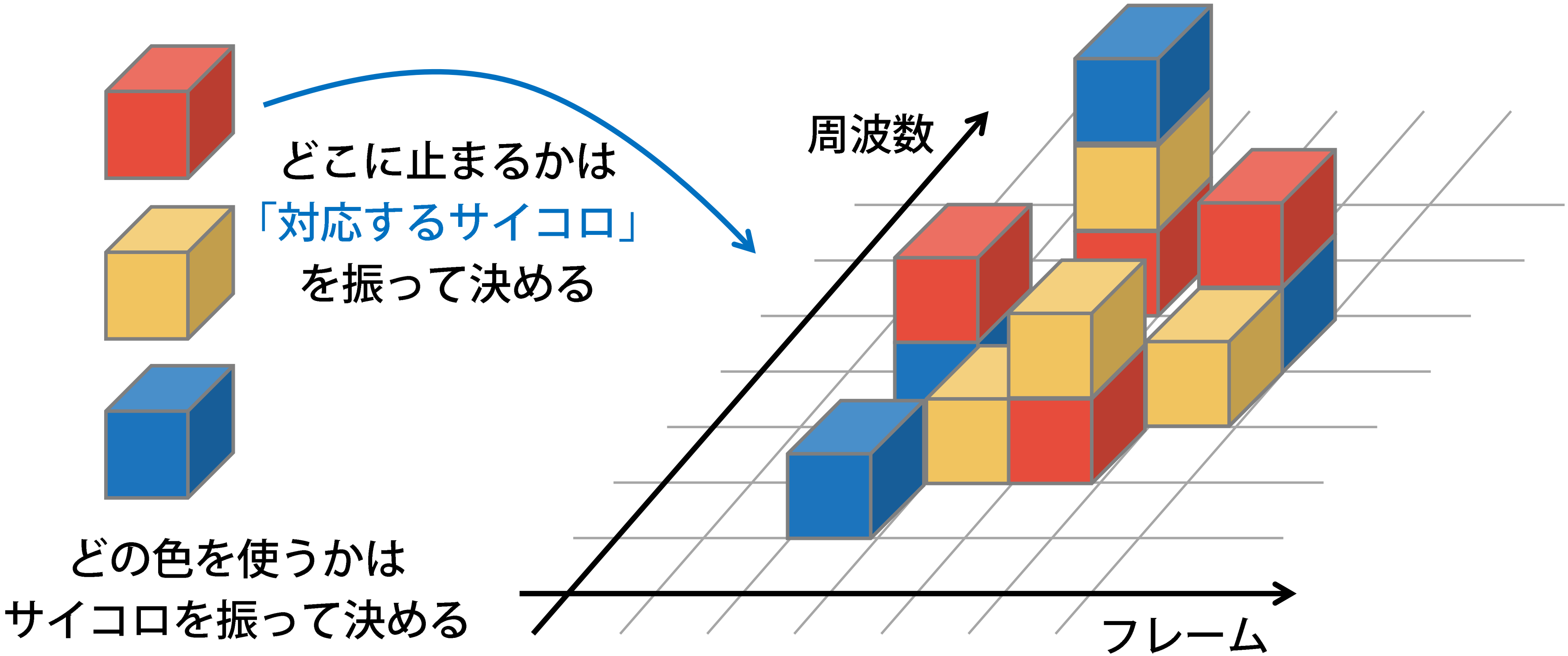


どこに止まるかは
サイコロを振って決める

何度も投げ込むと
積み重なって
ヒストグラムとなる

二次元平面上の混合離散分布

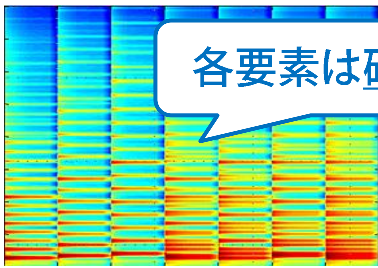
- 音量子 (sound quantum) を投げ入れると、平面内のどこかで止まる
  - 二次元平面はグリッドに分割されている



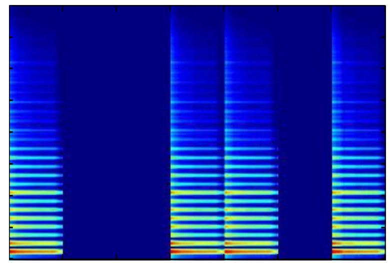
確率的潜在成分解析 (PLCA) の確率モデル

- 二次元平面上の混合離散分布を定式化

混合離散分布

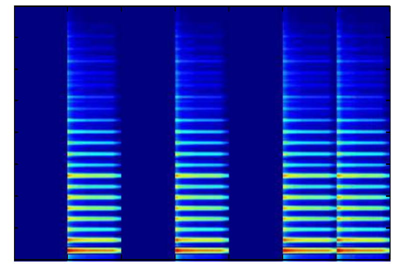


離散分布



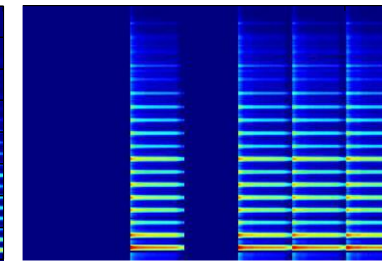
$p(n,m|k=1)$

離散分布



$p(n,m|k=2)$

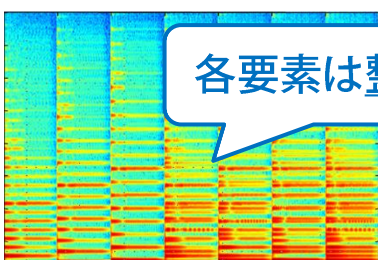
離散分布



$p(n,m|k=3)$

$p(n,m) \longleftarrow$

振幅スペクトログラム



$$p(n,m) = \sum_{k=1}^{K} \underbrace{p(k)}_{\text{混合比}} \underbrace{p(n,m|k)}_{\text{要素分布}}$$

$$p(n,m|k) \downarrow p(n|k)p(m|k)$$

K個の離散分布を重み付きで足し合わせ

因数分解できる形を仮定

ノンパラメトリックベイズ拡張

無限は扱えないので
打ち切り近似が必要

	ガンマ過程 (GaP)	ベータ過程 (BP)	ディリクレ過程 (DP)
有限次元の確率分布の極限	無限個のガンマ分布	無限個のベータ分布	無限次元のディリクレ分布
棒折り過程 (stick-breaking process)	複雑	複雑	簡単
料理店表現 (restaurant representation)	複雑	インド料理過程 (IBP)	中華料理店過程 (CRP)

打ち切り近似を
する必要がない

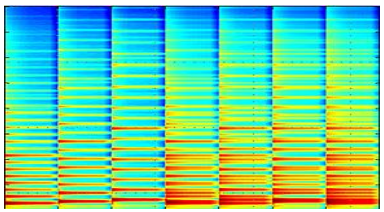
無限次元の混合比

$$\begin{tikzcd} \alpha \arrow[r, "\text{棒折り過程}"] \arrow[rr, bend right, "\text{中華料理店過程}"'] & \pi \arrow[r, "\text{任意の分布}"] & z \end{tikzcd}$$

Dirichlet Process PLCA

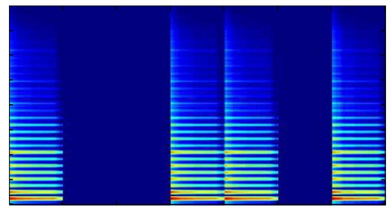
- 二次元混合離散分布の事前分布としてディリクレ過程を導入

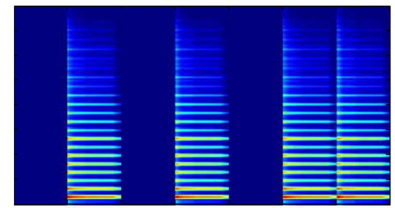
混合離散分布

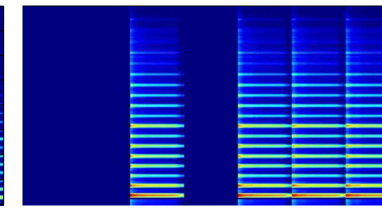


$p(n,m) \longleftarrow p(n|1)p(m|1) \quad p(n|2)p(m|2) \quad p(n|3)p(m|3)$

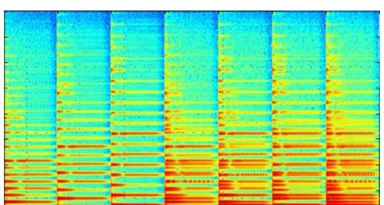
離散分布 離散分布 離散分布







振幅スペクトログラム

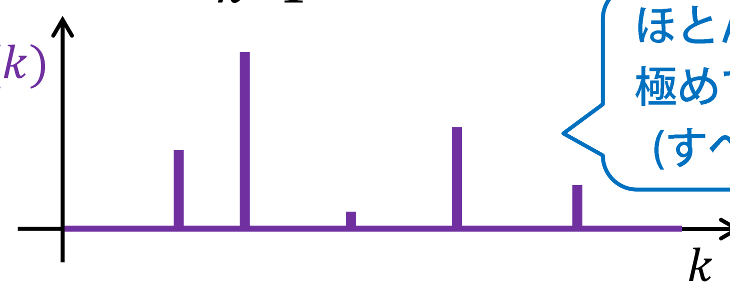


混合比 要素分布

$$p(n,m) = \sum_{k=1}^{\infty} p(k) p(n|k) p(m|k)$$

ディリクレ過程

$$p(k) \sim \mathrm{DP}(\alpha)$$



ほとんどの重みが
極めてゼロに近い
(すべて足すと1)

周辺化ギブスサンプリング

- 実効的な基底数K^+を増減させながらパラメータの推定が可能
  - 最大基底数で打ち切る近似が不要で効率的
  - 中華料理店過程のおかげで、$p(k)$を考えることが不要に
  - 事前分布の共役性のおかげで、$p(n|k)$と$p(m|k)$を考えることが不要に

$$p(n,m)=\sum_{k=1}^{\infty}p(k)p(n|k)p(m|k)$$

ディリクレ過程 $p(k)\sim \mathrm{DP}(\alpha)$

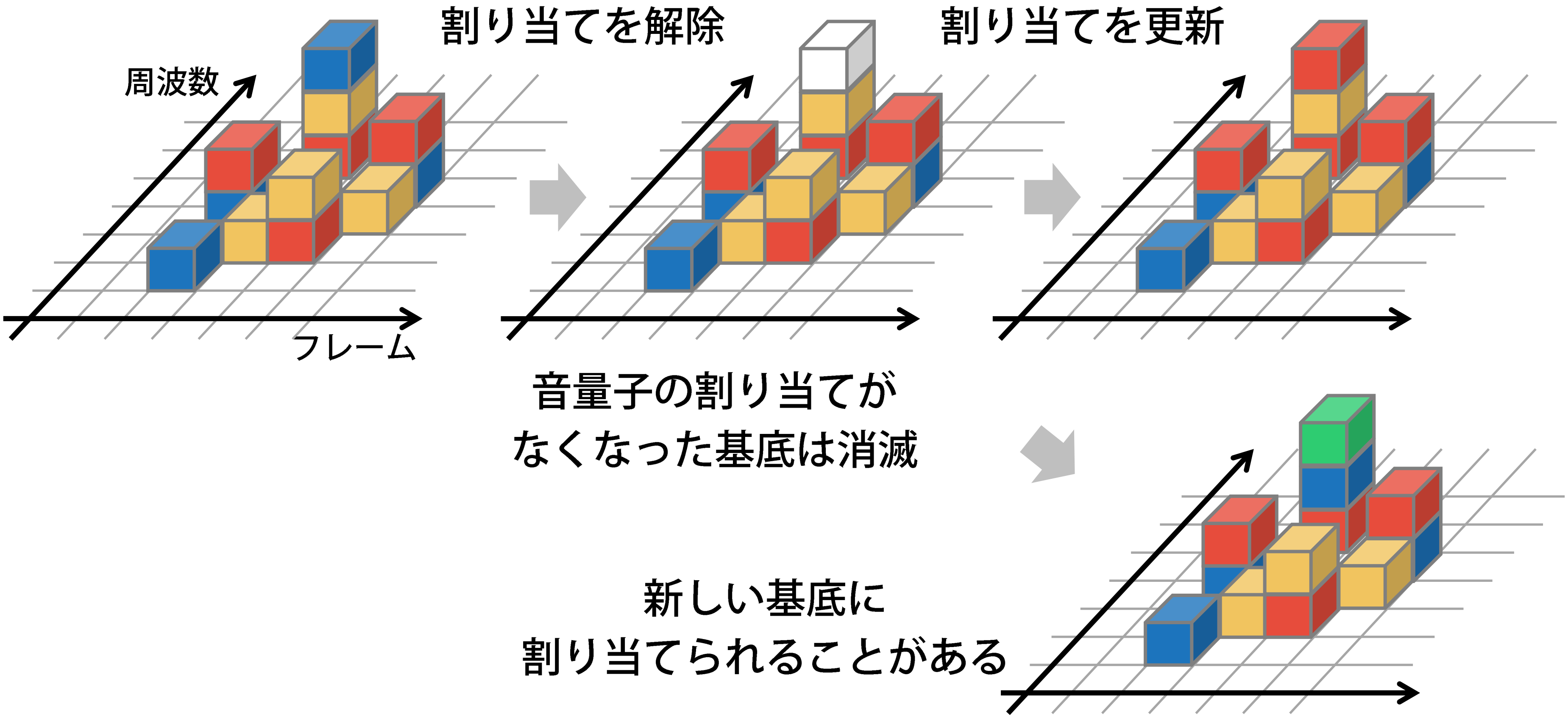
ディリクレ分布 $p(n|k)\sim \mathrm{Dirichlet}(\beta)$

ディリクレ分布 $p(m|k)\sim \mathrm{Dirichlet}(\gamma)$

本来求めたかった$p(k)p(n|k)p(m|k)$の代わりに何を求めればいいのか？

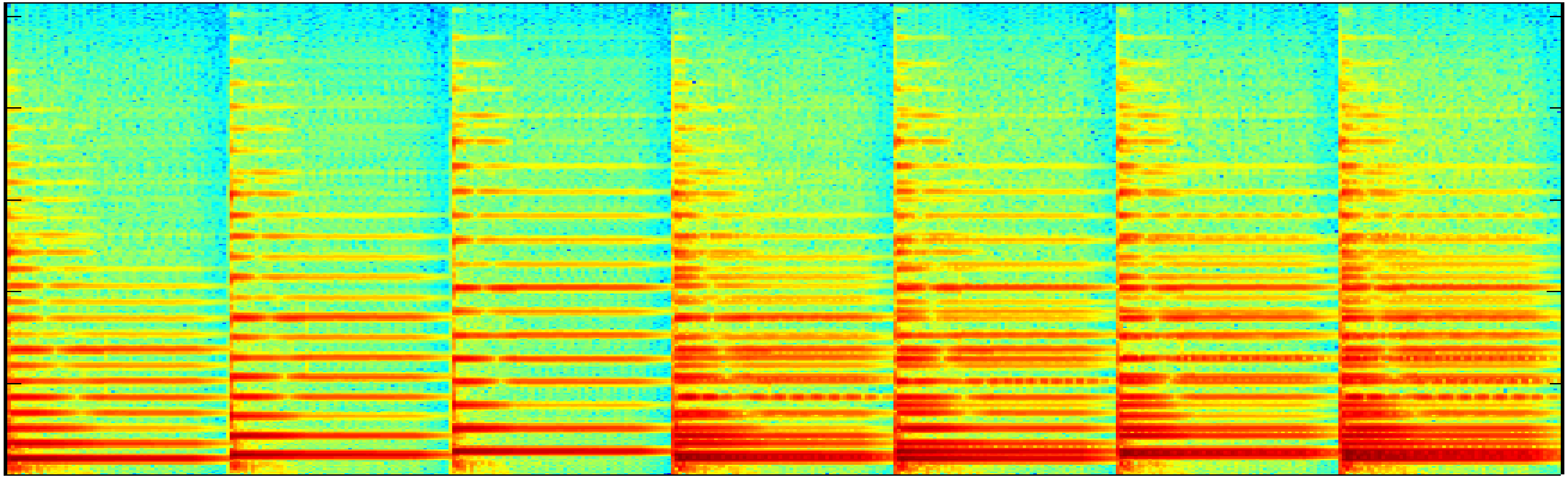
周辺化ギブスサンプリング

- 音量子の基底への割り当てを順番にサンプリング



実験

- 人工データに適用
  - C, E, Gを組み合わせを変えながら同時発音 (1.2 s x 7 = 8.4 s)
  - 16 [kHz] 16 [bits]モノラル信号を、512点窓幅・160点シフトでFFT
  - GaP-NMFとDP-PLCAを比較



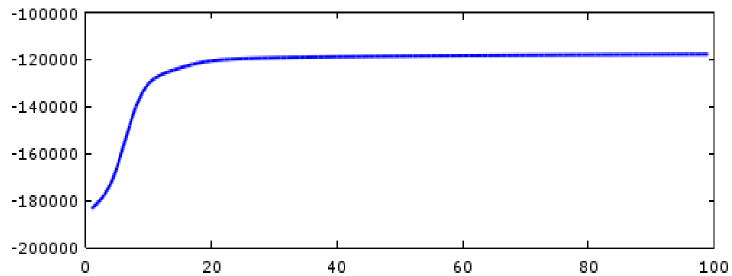
人工データ (平均振幅=1)

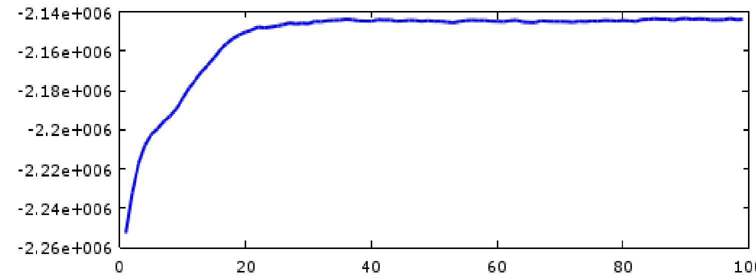
- DP-PLCA (Gibbs) は基底数を適切に推定
  - GaP-NMF (VB) は基底数を過剰に見積もってしまう傾向

GaP-NMF

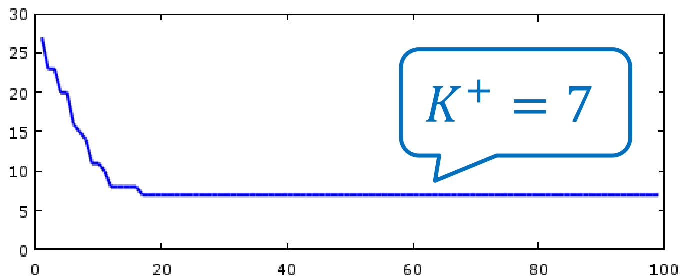
DP-PLCA

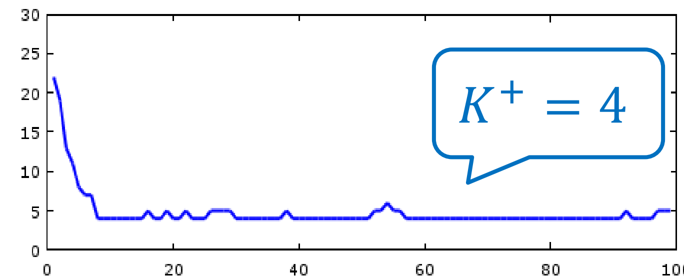
尤度





基底数



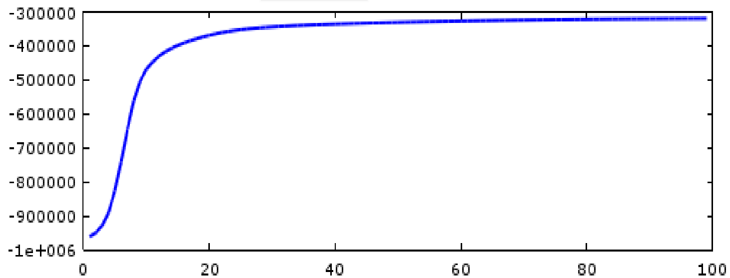


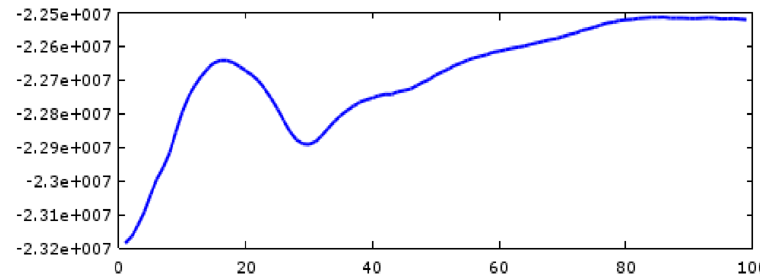
人工データ (平均振幅=10)

- DP-PLCA (Gibbs) は基底数を適切に推定
  - GaP-NMF (VB) は基底数を過剰に見積もってしまう傾向

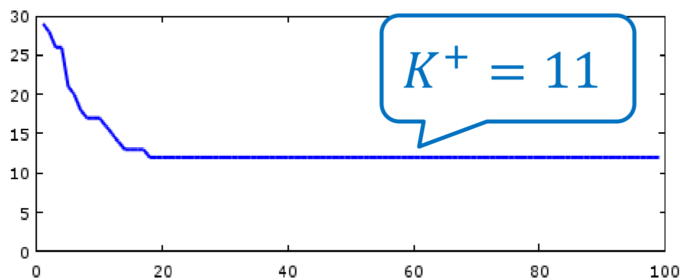
GaP-NMF　　DP-PLCA

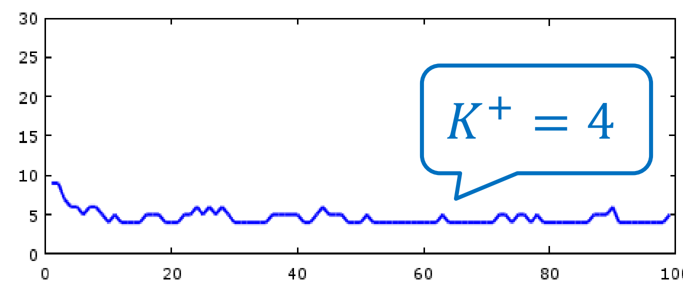
尤度





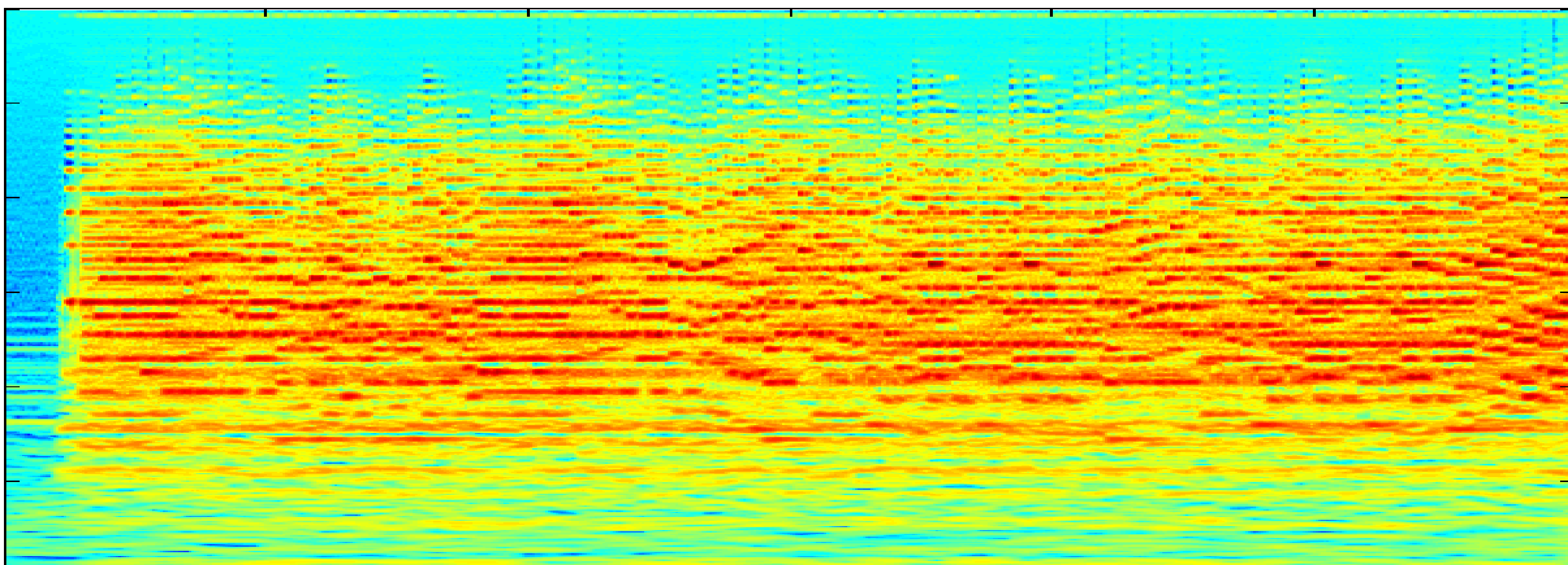
基底数





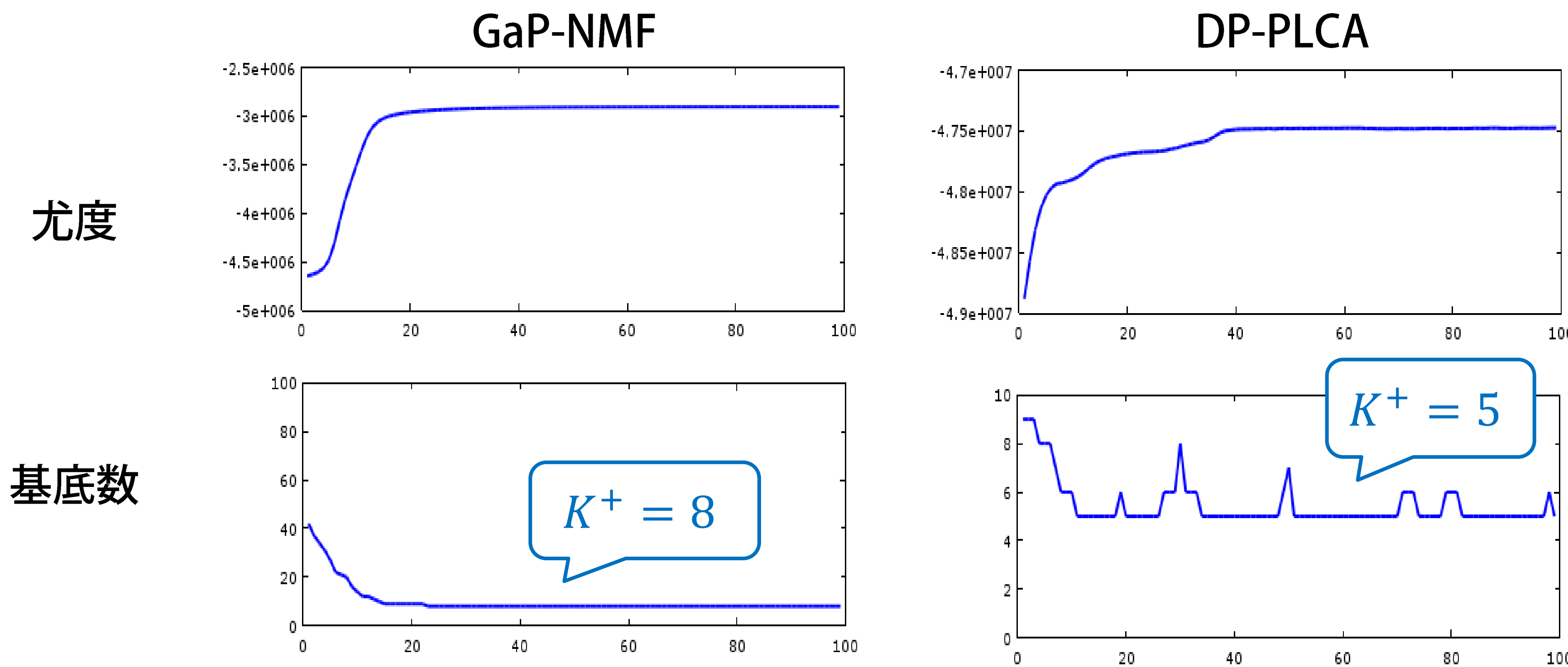
実験

- 実データに適用
  - MAPSデータベースのピアノ曲
  - 44.1 [kHz] 16 [bits]モノラル信号を、連続ウェーブレット変換
  - GaP-NMFとDP-PLCAを比較



実データ (平均振幅=1)

- データが非常に複雑でいずれも基底数を過小評価してしまう
  - DP-PLCA (Gibbs) の方がGaP-NMF (VB) より簡単なモデルを選択する傾向



まとめと今後の課題

- ノンパラメトリックベイズ拡張されたNMFとPLCAを紹介・提案
  - 非負値行列因子分解 (NMF)
    - ガンマ過程に基づく拡張 (GaP-NMF)
      - 閾値処理に依存するが、基底数を過剰に見積もる傾向
    - ベータ過程に基づく拡張 (BP-NMF)
  - 確率的潜在成分解析 (PLCA)
    - ディリクレ過程に基づく拡張 (DP-PLCA)
      - 基底数推定に関しては良好な結果
- 今後の課題：「音量子」の概念に対する物理的な裏付け
  - KL-NMFおよびDP-PLCAともに、振幅を量子化しておくことが前提
  - 量子化の最小単位をどう設定すべきか？ (e.g., フォノン)

音声/音楽/音響信号処理の解説書が出ます！

第4期刊行予定

ノンパラメトリックベイズ —点過程と統計的機械学習の数理 佐藤 一誠·著	変分ベイズ学習 中島 伸一·著
グラフィカルモデル 渡辺 有祐·著	脳画像のパターン認識 神谷 之康·著

第5期刊行予定

バンディット問題と その解法アルゴリズム 中村 篤祥 / 本多 淳也·著	データアナリティクスに おけるプライバシ保護 佐久間 淳·著
強化学習 森村 哲郎·著	オンライン予測 畑埜 晃平 / 瀧本 英二·著

第6期刊行予定

関係データ学習 石黒 勝彦 / 林 浩平·著	統計的自然言語処理 持橋 大地·著
$\boxed{\text{統計的音響信号処理 亀岡 弘和 / 吉井 和佳·著}}$	ロボットの運動学習 森本 淳·著

第7期刊行予定

道具としての情報幾何 津田 宏治·著	映像認識 篠田 浩一·著
統計的因果探索 清水 昌平·著	