Vol.2016-MUS-112 No.21
2016/8/1

NMF vs PLCA: 多重音生成過程のための無限因子モデルと無限混合モデル

吉井 和佳[1,a)]　中村 栄太[1,b)]　糸山 克寿[1,c)]　後藤 真孝[2,d)]

概要：本稿では，音楽音響信号に対する音源分離のための主要な二つの行列分解技法である非負値行列因子分解 (nonnegative matrix factorization: NMF) と確率的潜在要素解析 (probabilistic latent component analysis: PLCA) について，対応する確率モデルの性質を明らかにし，比較検討を行う．NMF では，各フレームにおける混合音のスペクトルを，少数の基底スペクトルの重み付き和で近似する．すなわち，各フレームにおいて複数の音源が同時に生起することが許容されており，NMF は因子モデルの一種である．一方，PLCA では，時間・周波数平面上のスペクトログラムをヒストグラムであるとみなし，その背後にある確率分布を推定する．このとき，各時間・周波数ビンの振幅値を量子化し，仮想的な音粒子の観測個数であるとみなしたうえで，各粒子をいずれかの音源に排他的に割り当てるため，PLCA は混合モデルの一種である．これまで，NMF の方が物理的に自然な解釈ができるにもかかわらず，実際には PLCA も盛んに利用されており，理論・性能の両面で十分に調査がなされているとは言えなかった．本稿では，因子モデルである NMF に対してはガンマ過程あるいはベータ過程を，混合モデルである PLCA に対してはディリクレ過程を用いることにより，基底数を自動調節するためのノンパラメトリックベイズモデルを構成できることを示し，変分ベイズ法あるいはギブスサンプリングを用いた推論方法を導出する．また，音楽音響信号に対して音源分離を行い，期待通り動作することを確かめた．

1. はじめに

行列分解 (matrix factorization) は，現在の機械学習やデータマイニング分野を支える基礎技術である．推薦システムや音源分離などさまざまな応用で，行列 $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ が入力として与えられたときに，$\boldsymbol{X} \approx \boldsymbol{AB}$ となるような二つの行列 $\boldsymbol{A} \in \mathbb{R}^{M \times K}$ および $\boldsymbol{B} \in \mathbb{R}^{K \times N}$ を求める問題が現れる．ここで重要なのは，K をもとの行列 $\boldsymbol{X}$ のサイズである M や N よりもずっと小さくしておけば，コンパクトな表現が得られることである．実際，$\boldsymbol{X}$ の自由度（独立に調節可能な変数の個数）は MN であるが，$\boldsymbol{AB}$ の自由度は $K(M+N)$ となっており，自由度が大幅に制限されている．この結果，$\boldsymbol{X}$ に含まれる冗長性が取り除かれ，$\boldsymbol{X}$ を構成する N 個の列ベクトルに含まれる典型的な K 個のパターン（基底ベクトル）が $\boldsymbol{A}$ に，それらがどの程度含まれているかの重みが $\boldsymbol{B}$ として得られる．

行列分解においては，入力行列 $\boldsymbol{X}$ や推定すべき行列 $\boldsymbol{A}$ および $\boldsymbol{B}$ が満たすべき性質を定め，$\boldsymbol{X}$ と $\boldsymbol{AB}$ との近似誤差を定式化することにより，様々な変種を得ることができる．例えば，主成分分析 (principal component analysis: PCA) では，$\boldsymbol{A}$ が固有ベクトルの集合，$\boldsymbol{B}$ がそれらの重みに対応しており，$\boldsymbol{X}$ と $\boldsymbol{AB}$ の二乗誤差が最小化されるように学習を行う．一方，独立成分分析 (independent component analysis) では，$\boldsymbol{A}$ が互いに独立なベクトルの集合，$\boldsymbol{B}$ がそれらの重みに対応しており，やはり $\boldsymbol{X}$ と $\boldsymbol{AB}$ の二乗誤差が最小化されるように学習を行うのが一般的である．

音楽音響信号の音源分離においては，$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B}$ の要素を非負値に制約した非負値行列因子分解 (nonnegative matrix factorization: NMF) や確率的潜在要素解析 (probabilistic latent component analysis: PLCA) と呼ばれる行列分解技法が広く用いられている [1–5]．過去の文献では，両者は（数式の上では）等価であると記述されたり，NMF を確率モデルとして解釈したものが PLCA であると紹介される場合があるが，このような理解は正しくない．確率モデルに基づく統計的推論の見地に立てば，NMF は因子モデル (factor model)，PLCA は混合モデル (mixture model) に対応しており，観測スペクトログラムの生成過程の捉え方に決定的な違いがある．

1　京都大学 大学院情報学研究科 知能情報学専攻
Yoshida-honmachi, Sakyo, Kyoto, Kyoto 606-8501, Japan
2　産業技術総合研究所 情報技術研究部門
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
a)　yoshii(at)i.kyoto-u.ac.jp
b)　enakamura(at)sap.ist.i.kyoto-u.ac.jp
c)　itoyama(at)kuis.kyoto-u.ac.jp
d)　m.goto(at)aist.go.jp

© 2016 Information Processing Society of Japan

一般に，観測データの背後にある基本パターン（クラス）を発見しようと思えば，因子モデルと混合モデルのうちで適切な方を選択する必要がある．混合モデルでは，観測データ中の各サンプルはひとつのクラスに排他的に属することを仮定している．したがって，混合モデルを用いると，標準的なクラスタリングの問題を解くことができる．一方，因子モデルは，各サンプルは複数の（場合によっては全ての）クラスに同時に属することを許容している．したがって，因子モデルを用いると，多重クラスタリングの問題を解くことができる．音源分離問題においては，観測される混合音スペクトログラムに含まれる基本パターンは音源に対応していることを期待したうえで，何を「サンプル」と解釈するかが重要なポイントである．

因子モデルである NMF では，観測スペクトログラムの各フレームにおけるスペクトルをサンプルとみなし，それを複数の基底スペクトルの重み付き和で近似する（図 **1**）．基底スペクトルは，ある音源の平均的なスペクトルに対応しており，各フレームは複数の音源に割り当てられることになる．このとき，観測スペクトルが従う，基底スペクトルの重み付き和をパラメータに持つ確率分布を定めることで，Euclidean NMF（EU-NMF, ガウス分布に対応）[6]，Kullback-Leiber NMF（KL-NMF, ポアソン分布に対応）[7]，Itakura-Saito NMF（IS-NMF, 複素ガウス分布に対応）[8] など様々な変種を定式化することができる．本稿では特に，最も広く利用されている点と，PLCA との関連から，KL-NMF について取り上げる．KL-NMF は，ポアソン分布に基づいているため，観測スペクトルの振幅は整数値に制限されている（実用上は実数で問題ない）．

一方，混合モデルである PLCA では，時間・周波数平面上の観測スペクトログラムをヒストグラム（カウントデータ）とみなし，そのヒストグラムを構成するひとつひとつのサンプルをある音源に排他的に割り当てていく（図 1）．混合モデルでは，複数のスペクトルの重畳を本質的に表現できない．苦肉の策として，観測スペクトルを各々が周波数をもつ音粒子の集合，すなわちヒストグラムであるとみなすことで，音粒子は一つの音源にしか属せないが，スペクトル全体としてみれば複数の音源に属していると解釈できる．これは，自然言語処理分野におけるトピックモデルの考え方に着想を得ており，ある文書に着目すると，そこに含まれる単語は一つのトピックに排他的に割り当てられるが，文書としてみれば複数のトピックに属すると解釈するのと同じである．PLCA という名前も，代表的なトピックモデルある確率的潜在意味解析 (probabilistic latent semantic analysis: PLSA) に倣って名づけられている．

このような違いを念頭におくと，観測スペクトルに含まれる音源数を自動的に推定するため，NMF や PLCA に対するノンパラメトリックベイズモデルを自然に定式化することができる．具体的には，NMF に対しては，ガンマ過程

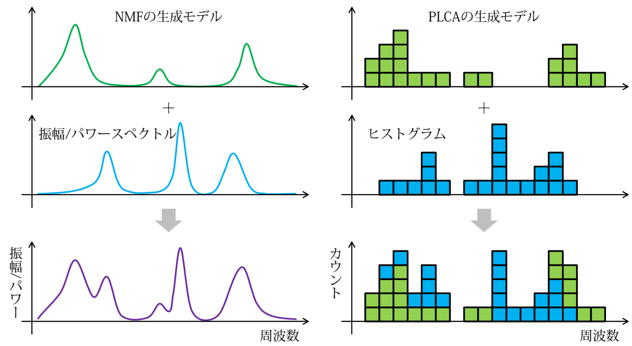

図 **1**　NMF と PLCA における観測スペクトルの生成モデル．NMF は確率変数（振幅/パワースペクトル）の和に基づいて定式化されており，PLCA は確率分布（ヒストグラム）の和に基づいて定式化されている．

(gamma process: GaP) あるいはベータ過程 (beta process: BP) を用いることにより，GaP-NMF および BP-NMF が定式化できる．特に，BP-NMF においては，各フレームにおいて各音源が存在するかどうかの二値変数が導入されており，音源区間の推定が同時に行える利点がある．一方，PLCA に対しては，ディリクレ過程 (Dirichlet process, DP) を用いることにより，DP-PLCA が定式化できる．

2. 関連研究

NMF や PLCA は，音楽音響信号の音源分離や自動採譜に最も利用されている技術であり，性能を向上させるためのさまざまな改良が試みられている．本章では，それぞれの確率モデルに関する関連技術を紹介する．

2.1 ディリクレ過程に基づく無限混合モデル

無限混合モデルは，無限個の混合比と要素分布をパラメータに持ち，それら生成するための事前分布としてディリクレ過程 (DP) を用いる．Sethuraman [9] は，無限個の混合比を明示的に生成するため，DP の構成法のひとつである棒折り過程 (stick-breaking process: SBP) を導出している．Blei ら [10] は，モデルパラメータの事後分布を近似的に計算するため，変分ベイズ法 (variational Bayes: VB) に基づく反復推定を提案している．しかし，実際の計算機では無限個のパラメータを取り扱うことはできないため，あらかじめ定めておいた要素数で計算を打ち切る必要があった．すなわち，最初に十分に大きな要素数で VB を初期化し，反復計算が進むにつれて，不要とされる要素を削除していく方法が取られていた．

DP の別の構成法として，中華料理店過程 (Chinese restaurant process: CRP) が知られている．CRP では，無限個の混合比を事前分布である DP で積分消去することで，無限個のパラメータを陽に取り扱う必要がなくなる．Neal [11] は，モデルパラメータの事後分布を近似的に計算するため，ギブスサンプリングを導出している．この方法では，

観測データを表現するのに必要な要素数が適宜増減可能なため，VB のように必要以上の個数の要素を考慮することによる計算量の無駄がない．

PLCA [12] は，基本的なトピックモデルである PLSA [13] の拡張であるため，PLSA のベイズ拡張である潜在的ディリクレ配分法 (latent Dirichlet allocation: LDA) [14] と同様のベイズ拡張が可能である．Teh ら [15] は，トピック数（PLCA における基底数に相当）を自動的に調節するため，階層ディリクレ過程 (hierarchical Dirichlet process: HDP) に基づく LDA を提案している．LDA では，異なる文書間でトピックを共有するため，一階層の DP ではなく HDP を用いる必要がある．もし，各文書のトピックの事前分布として文書ごとに独立な DP を仮定すると，推定されるトピック（単語分布）は各文書に独自のものとなってしまう．これを防ぐため，各文書の DP を束ねる上位階層の DP が導入されている．

PLCA は，調波構造を表現する基底をあらかじめ準備したり，学習しておくことによって，優れた自動採譜精度を達成している [16–18]．しかし，基底数を自動推定するノンパラメトリックベイズ拡張についてはこれまで提案されていなかった．後の章で述べるように，PLCA の無限モデルを構成する上では，HDP ではなく DP を用いれば十分であり，VB あるいはギブスサンプリングのいずれを用いても事後分布の近似計算が可能である．

2.2 ガンマ過程に基づく無限因子モデル

無限因子モデルを定式化する方法の一つは，無限個の基底に対する非負の重みを導入し，その無限次元の重みベクトルをスパースに誘導する（限られた少数の重みのみがゼロではない実効的な値を持つ）ように，事前分布としてガンマ過程 (GaP) を導入することである．ここで，無限混合モデルにおける混合比とは異なり，基底の重みの総和は 1 となるように正規化されている必要はない．Roychowdhury ら [19] は，無限個の重みを明示的に生成するため，GaP に対する SBP を提案し，行列分解モデルに対する VB を導出している．しかし，この手法の実装は複雑であり，今だほとんど利用されていない．一方，DP に対する CRP に相当する表現はまだ確立されておらず，効率的なサンプリング方法は研究の途上にある．

GaP の簡便な構成法には，弱極限近似 (weak-limit approximation) が知られており，各重みに対してガンマ事前分布を導入し，その形状パラメータを非常に小さく設定すればよい．Hoffman ら [5] は，この近似に基づく GaP-NMF を提案しており，VB を用いることで，観測データに合わせて適切に基底数が調節できることを報告している．ただし，最初に十分多くの基底を準備しておき，VB の反復に合わせて不要な基底を削除していく方式のため，計算量は比較的多くなることに注意が必要である．

2.3 ベータ過程に基づく無限因子モデル

無限因子モデルを定式化する別の方法は，各基底の生起確率（コイントスで表が出る確率）を考え，無限個ある生起確率の事前分布としてベータ過程 (BP) を導入することである．各フレームでは，その生起確率に従って，無限個の基底の中から複数の基底を同時に選ぶことができるが，生起確率は非常にスパース（ほとんどの基底の生起確率はほぼゼロ）であるため，実際には限られた少数の基底しか生起しない．Teh ら [20] は，無限個の生起確率を明示的に生成するため，BP に対する SBP を導出している．さらに，スライスサンプリングとギブスサンプリングを併用することで，打ち切り近似なしで，モデルパラメータの事後分布からサンプリングを可能にしている．Gupta ら [21] は，SBP 表現を用いた BP-NMF を提案している．Paisley ら [22] は，新たな形式を持つ SBP を導出し，VB を用いた事後分布推論方法を提案している．

BP の別の構成方法として，インド料理過程 (Indian buffet process: IBP) が知られている [23]．DP に対する CRP と同様に，無限個の生起確率を積分消去することで，BP に対する IBP が得られる．また，BP の簡便な構成法として，弱極限近似も可能であり，各基底の生起確率に対してベータ事前分布を導入し，その第一パラメータを非常に小さく設定すればよい．Liang ら [24, 25] は，この近似に基づく BP-NMF を提案しており，VB を用いた事後分布推論方法が示されている．

3. ノンパラメトリックベイズ NMF

本章では，NMF の音源分離への適用方法について述べ，ガンマ過程およびベータ過程に基づくノンパラメトリックベイズ拡張について紹介する．

3.1 NMF に基づく音源分離

混合音に対して短時間フーリエ変換を行って得られる複素スペクトログラムを $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N] \in \mathbb{C}^{M\times N}$, k 番目の音源信号の複素スペクトログラムを $\tilde{\boldsymbol{X}}_k = [\tilde{\boldsymbol{x}}_{k1}, \cdots, \tilde{\boldsymbol{x}}_{kN}] \in \mathbb{C}^{M\times N}$ とする．M は周波数ビン数，N はフレーム数である．観測した混合音が K 個の音源信号の周波数領域瞬時混合であると仮定すると，以下が成立する．

$$\tilde{\boldsymbol{X}} = \sum_{k=1}^{K} \tilde{\boldsymbol{X}}_k \quad \Leftrightarrow \quad \tilde{\boldsymbol{x}}_n = \sum_{k=1}^{K} \tilde{\boldsymbol{x}}_{kn} \tag{1}$$

いま，観測変数 $\tilde{\boldsymbol{X}}$ を潜在変数 $\tilde{\boldsymbol{X}}_k$ に分解したい．しかし，これは不良設定問題であるので，$\tilde{\boldsymbol{X}}_k$ に対応する振幅スペクトログラム $\boldsymbol{X}_k = [\boldsymbol{x}_{k1}, \cdots, \boldsymbol{x}_{kN}] \in \mathbb{R}_+^{M\times N}$ ($x_{knm} = |\tilde{x}_{knm}|$) は，ランク 1 の行列 $\boldsymbol{Y}_k = [\boldsymbol{y}_{k1}, \cdots, \boldsymbol{y}_{kN}] \in \mathbb{R}_+^{M\times N}$ で近似できるものと仮定する（図 **2**）．

$$\boldsymbol{X}_k \approx \boldsymbol{w}_k \boldsymbol{h}_k^T \stackrel{\text{def}}{=} \boldsymbol{Y}_k \tag{2}$$

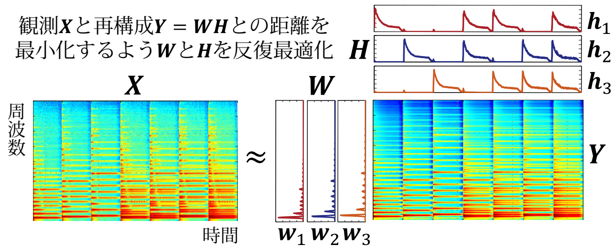

図 2　混合音スペクトログラムに対する NMF.

ここで，振幅スペクトルの加法性，すなわち $\boldsymbol{X} = \sum_{k=1}^{K} \boldsymbol{X}_k$ および $\boldsymbol{Y} = \sum_{k=1}^{K} \boldsymbol{Y}_k$ が成立することを仮定すると，

$$\boldsymbol{X} \approx \sum_{k=1}^{K} \boldsymbol{w}_k \boldsymbol{h}_k^T \stackrel{\text{def}}{=} \boldsymbol{Y} \tag{3}$$

が得られる．ここで，$\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}_+^{M \times N}$ $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N] \in \mathbb{R}_+^{M \times N}$ とすると，

$$\boldsymbol{x}_n \approx \sum_{k=1}^{K} h_{kn} \boldsymbol{w}_k \stackrel{\text{def}}{=} \boldsymbol{y}_n \tag{4}$$

が成立する．すなわち，混合音のスペクトル $\boldsymbol{x}_n$ は，K 個の基底スペクトル $\boldsymbol{w}_k$ の線形和で近似できることを意味する．

音源分離によく利用される KL-NMF に対応する確率モデルでは，潜在変数 x_{knm} が y_{knm} をパラメータに持つポアソン分布（$\mathbb{E}[x_{knm}] = y_{knm}$）に従うことを仮定する．

$$x_{knm} \sim \text{Poisson}(y_{knm}) \tag{5}$$

ここで，複数の音源信号の重畳における振幅スペクトルの加法性を仮定すると（実際には成立しないことに注意），ポアソン分布の再生性から

$$x_{nm} \sim \text{Poisson}(y_{nm}) \tag{6}$$

を得る．ここで，尤度関数である式 (6) の対数をとって符号を反転させると，

$$\begin{aligned} -\log(x_{nm}|y_{nm}) &\stackrel{c}{=} x_{nm} \log \frac{x_{nm}}{y_{nm}} - x_{nm} + y_{nm} \\ &= \mathcal{D}_{\text{KL}}(x_{nm}||y_{nm}) \end{aligned} \tag{7}$$

となり，負の対数尤度は x_{nm} と y_{nm} の KL ダイバージェンスと定数項を除いて等しいことが分かる．したがって，式 (6) の最大化（最尤推定）は式 (7) の最小化と等価である．KL-NMF の目標は，全時間・周波数平面にわたる対数尤度関数の和 $\sum_{nm} \log p(x_{nm}|y_{nm})$ を最大化するような $\boldsymbol{W}$ および $\boldsymbol{H}$ を求めることである．

NMF の実行結果を用いて，混合音の音源分離を行うには，ウィナーフィルタを用いる．

$$\mathbb{E}[\tilde{x}_{knm}|\tilde{x}_{nm}] = \frac{y_{knm}}{y_{nm}} \tilde{x}_{nm} \tag{8}$$

この処理では，$\tilde{\boldsymbol{X}}_k$ の位相は $\tilde{\boldsymbol{X}}$ の位相と同一であるという仮定がおかれている．最後に，逆フーリエ変換を用いて，$\mathbb{E}[\tilde{\boldsymbol{X}}_k|\tilde{\boldsymbol{X}}]$ から k 番目の音源信号を復元すればよい．

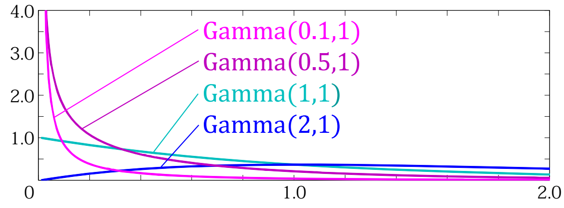

図 3　形状母数が異なる幾つかのガンマ分布．

3.2　ガンマ過程に基づく NMF

本節では，GaP-KL-NMF について説明する．まず，GaP の弱極限近似に基づくベイズモデルの定式化を説明し，VB およびギブスサンプリングの適用について述べる．

3.2.1　ベイズモデルの定式化

まず，式 (4) に対し，K 次元の非負値ベクトル $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_K]$ を導入する．

$$\boldsymbol{x}_n \approx \sum_{k=1}^{K} \theta_k h_{kn} \boldsymbol{w}_k \stackrel{\text{def}}{=} \boldsymbol{y}_n \tag{9}$$

ここで，$\theta_k \geq 0$ は基底 k の大域的な重みである．この $\boldsymbol{\theta}$ に対し，観測データ $\boldsymbol{X}$ を表現するのに必要な基底 k 以外の要素 θ_k がゼロとなるようなスパースな学習を行いたい．

次に，$\boldsymbol{\theta}$，$\boldsymbol{W}$，$\boldsymbol{H}$ に対して事前分布を導入する．まず，$\boldsymbol{W}$ 及び $\boldsymbol{H}$ の各要素は非負値であるので，ガンマ事前分布を用いると都合が良い．

$$w_{km} \sim \text{Gamma}(a_0^w, b_0^w) \tag{10}$$

$$h_{kn} \sim \text{Gamma}(a_0^h, b_0^h) \tag{11}$$

ここで，$a_0^* > 0$ 及び $b_0^* > 0$ はそれぞれ，ガンマ分布の形状パラメータと逆尺度パラメータである．更に，$\boldsymbol{\theta}$ に対しても同様にガンマ事前分布を仮定する．

$$\theta_k \sim \text{Gamma}\left(\frac{\alpha c}{K}, \alpha\right) \tag{12}$$

ここで，$\alpha > 0$ 及び $c > 0$ は超パラメータである．形状パラメータが小さくなるほど 0 が出る確率が大きくなる（図 **3**）．ただし，$\mathbb{E}_{\text{prior}}[\theta_k] = \frac{c}{K}$，$\mathbb{E}_{\text{prior}}[\sum_k \theta_k] = c$ である．

ここで，式 (10)，式 (11) 及び式 (12) で構成される有限モデルに対して，$K \to \infty$ となる極限を考えると，以下のガンマ過程が得られる．

$$G \sim \text{GaP}(\alpha, G_0) \tag{13}$$

ここで，G_0 は空間 U（$\boldsymbol{w} \in \mathbb{R}_+^M$ と $\boldsymbol{h} \in \mathbb{R}_+^N$ の直積空間）上に定義された基底測度であり，$G_0(U) = c$ を満たす（図 **4**）．このとき，G は U 上の離散測度となり，空間 U の任意の分割 $\{U_i\}_{i=1}^I$ に対して

$$G(U_i) \sim \text{Gamma}(\alpha G_0(U_i), \alpha) \tag{14}$$

が成立している．ただし，$\mathbb{E}[G] = G_0$ である．無限小区間への分割を $\{U_k\}_{k=1}^{\infty}$ とすると，$G(U_k) = \theta_k$ である．α は

© 2016 Information Processing Society of Japan

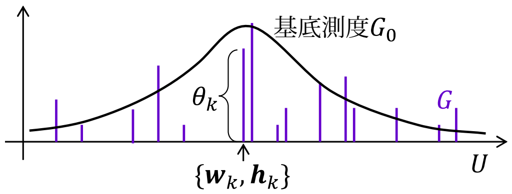

図 4　NMF のためのガンマ過程事前分布．

集中度と呼ばれ，α が小さくなるほど $\boldsymbol{\theta}$ はよりスパースになる．計算機上では $K \to \infty$ は扱えないが，K を α に比べて十分大きな値に設定すれば，式 (12) はガンマ過程の良い近似となる（弱極限近似）．

3.2.2　変分ベイズ法

式 (6), (10), (11), (12) で定義されるノンパラメトリックベイズ KL-NMF (GaP-KL-NMF) に対する VB について述べる [5]．今，観測データ $\boldsymbol{x}$ が与えられたときに，ベイズの定理を用いて未知パラメータ $\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}$ の事後分布

$$p(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})}{p(\boldsymbol{X})} \tag{15}$$

を計算したい．しかし，分母の周辺尤度 $p(\boldsymbol{X})$ は解析的に計算できないため，対数周辺尤度 $\log p(\boldsymbol{X})$ の変分下限 $\mathcal{L}(q)$ を構成し，逐次最大化を行うことで $\log p(\boldsymbol{X})$ を近似したい．具体的には，任意の分布 $q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})$ を導入し，凹関数 $\log(x)$ に対して Jensen の不等式を用いると

$$\begin{aligned}\log p(\boldsymbol{X}) &= \log \int q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) \frac{p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})}{q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})} d\boldsymbol{\theta} d\boldsymbol{W} d\boldsymbol{H} \\ &\geq \int q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) \log \frac{p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})}{q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})} d\boldsymbol{\theta} d\boldsymbol{W} d\boldsymbol{H} \\ &= \mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})] \\ &\quad - \mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})}[\log q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})] \stackrel{\text{def}}{=} \mathcal{L}(q) \end{aligned} \tag{16}$$

を得る．等号成立条件は $q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) = p(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}|\boldsymbol{X})$ であり，このとき $\mathcal{L}(q)$ が最大値をとる．しかし，真の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}|\boldsymbol{X})$ は計算困難であるため，変分事後分布を因子分解可能な形 $q(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) = q(\boldsymbol{\theta})q(\boldsymbol{W})q(\boldsymbol{H})$ に限定し，その中でも以下の通り計算できる変分下限

$$\begin{aligned}\mathcal{L}(q) &= \mathbb{E}_q[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})] + \mathbb{E}_q[\log p(\boldsymbol{\theta})] \\ &\quad + \mathbb{E}_q[\log p(\boldsymbol{W})] + \mathbb{E}_q[\log p(\boldsymbol{H})] \\ &\quad + H(q(\boldsymbol{\theta})) + H(q(\boldsymbol{W})) + H(q(\boldsymbol{H})) \end{aligned} \tag{17}$$

を最大化するものを求めたい．ここで，$H(\cdot)$ はエントロピーを表す．これは，変分事後分布 $q(\boldsymbol{\theta})q(\boldsymbol{W})q(\boldsymbol{H})$ の真の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}|\boldsymbol{X})$ に対する KL ダイバージェンスを最小化することと等価である．式 (17) を逐次最大化するには，以下の更新式を収束するまで繰り返せば良い．

$$q(\boldsymbol{\theta}) \propto \exp(\mathbb{E}_{q(\boldsymbol{H}, \boldsymbol{W})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})]) \tag{18}$$

$$q(\boldsymbol{H}) \propto \exp(\mathbb{E}_{q(\boldsymbol{H}, \boldsymbol{W})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})]) \tag{19}$$

$$q(\boldsymbol{W}) \propto \exp(\mathbb{E}_{q(\boldsymbol{H}, \boldsymbol{W})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})]) \tag{20}$$

Algorithm 1 GaP-KL-NMF に対する変分ベイズ法

Require: 非負値行列 $\boldsymbol{X} \in \mathbb{R}_+^{M \times N}$, 最大基底数 K，ガンマ過程の集中度 α，ガンマ事前分布のパラメータ $a_0^w, b_0^w, a_0^h, b_0^h$

変分事後分布 $q(\boldsymbol{\theta})$, $q(\boldsymbol{W})$, $q(\boldsymbol{H})$ をランダムに初期化
while not converged **do**
　$\lambda_{knm} \propto \exp(\log \mathbb{E}_q[y_{knm}])$
　$q(\theta_k) = \mathcal{G}(\frac{\alpha c}{K} + \sum_{nm} \lambda_{knm} x_{nm}, \alpha + \sum_{nm} \mathbb{E}_q[w_{km} h_{kn}])$
　$\lambda_{knm} \propto \exp(\log \mathbb{E}_q[y_{knm}])$
　$q(w_{km}) = \mathcal{G}(a_0^w + \sum_n \lambda_{knm} x_{nm}, b_0^w + \sum_n \mathbb{E}_q[\theta_k h_{kn}])$
　$\lambda_{knm} \propto \exp(\log \mathbb{E}_q[y_{knm}])$
　$q(h_{kn}) = \mathcal{G}(a_0^h + \sum_m \lambda_{knm} x_{nm}, b_0^h + \sum_m \mathbb{E}_q[\theta_k w_{km}])$
end while
return 変分事後分布 $q(\boldsymbol{\theta})$, $q(\boldsymbol{W})$, $q(\boldsymbol{H})$

式 (17) で与えられる変分下限 $\mathcal{L}(q)$ の第一項は式 (6) で計算できる対数ポアソン尤度の期待値であるが，依然として解析的に計算できない．そのため，凹関数 $\log(x)$ に対して Jensen の不等式を用いると更なる変分下限

$$\begin{aligned}&\mathbb{E}_q[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})] \\ &\stackrel{c}{=} \mathbb{E}_q\left[\sum_{nm}\left(x_{nm} \log \sum_k y_{knm} - \sum_k y_{knm}\right)\right] \\ &= \sum_{nm} x_{nm} \mathbb{E}_q\left[\log \sum_k \lambda_{knm} \frac{y_{knm}}{\lambda_{knm}}\right] - \sum_{knm} \mathbb{E}_q[y_{knm}] \\ &\geq \sum_{nm} x_{nm} \sum_k \lambda_{knm} \mathbb{E}_q\left[\log \frac{y_{knm}}{\lambda_{knm}}\right] - \sum_{knm} \mathbb{E}_q[y_{knm}] \\ &\stackrel{\text{def}}{=} \mathbb{E}_q[\log q(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})] \end{aligned} \tag{21}$$

を得る．ここで，λ_{knm} は $\sum_k \lambda_{knm} = 1$ を満たす補助変数である．変分下限を最大化する λ_{knm} は，ラグランジュの未定乗数法を用いて，$\lambda_{knm} \propto \exp(\log \mathbb{E}_q[y_{knm}])$ と求まる．

最後に，各パラメータの変分事後分布を導出する．実際には式 (17) で与えられる元の変分下限 $\mathcal{L}(q)$ ではなく，式 (21) を用いて得られた更なる変分下限を最大化することになる．すなわち，式 (18), (19), (20) において，$\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H})$ の代わりに次式を用いればよい．

$$\begin{aligned}\log q(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) &= \log q(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{H}) + \log p(\boldsymbol{\theta}) \\ &\quad + \log p(\boldsymbol{W}) + \log p(\boldsymbol{H}) \end{aligned} \tag{22}$$

具体的には，最適な変分事後分布 $q^*(\boldsymbol{\theta})$ は，$\boldsymbol{\theta}$ に関連する項のみを取り出すと以下の通り計算できる．

$$\begin{aligned}\log q^*(\boldsymbol{\theta}) &\stackrel{c}{=} \sum_{knm} x_{nm} \lambda_{knm} \log \theta_k - \sum_{knm} \theta_k \mathbb{E}_q[w_{km}] \mathbb{E}_q[h_{kn}] \\ &\quad + \sum_k \left(\left(\frac{\alpha c}{K} - 1\right) \log \theta_k - \alpha \theta_k\right) \end{aligned} \tag{23}$$

従って，θ_k の事後分布はガンマ分布となることが分かる．同様に，最適な $q^*(\boldsymbol{W})$ や $q^*(\boldsymbol{H})$ もガンマ分布として求まる．**Algorithm 1** に更新則を示す．反復ごとに，$\mathbb{E}[\theta_k]$ が十分に小さい基底 k を削除していけば，実効的な基底数 K_+ を自動的に推定できる．最終的に，音源分離を行う際には，各パラメータの期待値を用いるのが一般的である．

Algorithm 2 GaP-KL-NMF に対するギブスサンプリング

Require: 非負値行列 $\boldsymbol{X} \in \mathbb{R}_+^{M \times N}$, 最大基底数 K，ガンマ過程の集中度 α，ガンマ事前分布のパラメータ $a_0^w, b_0^w, a_0^h, b_0^h$

```
パラメータ θ, W, H をランダムに初期化
while not converged do
    λ_knm ∝ y_knm
    θ_k ~ G(αc/K + Σ_nm λ_knm x_nm, α + Σ_nm w_km h_kn)
    λ_knm ∝ y_knm
    w_km ~ G(a_0^w + Σ_n λ_knm x_nm, b_0^w + Σ_n θ_k h_kn)
    λ_knm ∝ y_knm
    h_kn ~ G(a_0^h + Σ_m λ_knm x_nm, b_0^h + Σ_m θ_k w_km)
end while
return パラメータ θ, W, H
```

3.2.3 ギブスサンプリング

決定的な近似推論方法である VB とは対称的に，確率的な近似推論法であるギブスサンプリングを用いれば，パラメータの事後分布からのサンプルを得ることができる．式 (15) で与えられる真の事後分布は解析的に計算できないとしても，あるパラメータ（例えば $\boldsymbol{\theta}$）の事後分布が，他のパラメータ（例えば $\boldsymbol{H}$ および $\boldsymbol{W}$）が既知であれば，解析的に計算できる場合がある．このとき，そのような「条件付き」事後分布からパラメータをサンプリングすることを各パラメータについて順番に繰り返せば，真の事後分布からのサンプルが得られる．GaP-KL-NMF においては，以下の手順を反復すればよい．

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\boldsymbol{H}, \boldsymbol{W}, \boldsymbol{X}) \tag{24}$$

$$\boldsymbol{H} \sim p(\boldsymbol{H}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X}) \tag{25}$$

$$\boldsymbol{W} \sim p(\boldsymbol{W}|\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{X}) \tag{26}$$

具体的には，$\boldsymbol{\theta}$ に着目すると，前節の VB と同様の変分下限を考えることにより

$$\begin{aligned}
&\log p(\boldsymbol{\theta}|\boldsymbol{H}, \boldsymbol{W}, \boldsymbol{X}) \propto \log p(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{W}, \boldsymbol{X}) \\
&\geq \log q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{W}, \boldsymbol{X}) \\
&\stackrel{c}{=} \sum_{knm} x_{nm} \lambda_{knm} \log \theta_k - \sum_{knm} \theta_k w_{km} h_{kn} \\
&+ \sum_k \left(\left(\frac{\alpha c}{K} - 1 \right) \log \theta_k - \alpha \theta_k \right)
\end{aligned} \tag{27}$$

を得る．ここで，λ_{knm} は $\sum_k \lambda_{knm} = 1$ を満たす補助変数である．変分下限を最大化する λ_{knm} は，ラグランジュの未定乗数法を用いて，$\lambda_{knm} \propto y_{knm}$ と求まる．VB とは異なり，λ_{knm} を最適化すれば常に等号を成立させることが可能であり，変分下限のタイトな評価が可能になっている．

最終的に，θ_k の条件付き事後分布はガンマ分布となることが分かる．**Algorithm 2** に更新則を示す．VB とは異なり，θ_k が非常に小さな値となった場合であっても，確率的に復活する可能性があるので，安易に削除してはいけない．音源分離を行う際には，各パラメータの条件付き事後分布の期待値を用いるのが一般的である．

3.3 ベータ過程に基づく NMF

本節では，BP-KL-NMF について説明する．まず，BP の弱極限近似に基づくベイズモデルの定式化を説明し，VB およびギブスサンプリングの適用について述べる．

3.3.1 ベイズモデルの定式化

まず，式 (4) に対し，K 次元の二値ベクトル $\boldsymbol{z}_n = [z_{1n}, z_{2n}, \cdots, z_{Kn}]$ を導入する．

$$\boldsymbol{x}_n \approx \sum_{k=1}^{K} z_{nk} h_{kn} \boldsymbol{w}_k \stackrel{\text{def}}{=} \boldsymbol{y}_n \tag{28}$$

ここで，$z_{nk} \in \{0, 1\}$ はフレーム n における基底 k のオン・オフを制御する二値変数である．すなわち，z_{nk} はアクティベーション h_{kn} に対するマスクとして作用する．この $\boldsymbol{Z} \in \{0, 1\}^{N \times K}$ に対し，観測データ $\boldsymbol{X}$ を表現するのに必要な基底 k 以外の変数 z_{nk} がゼロとなるようなスパースな学習を行うため，z_{nk} は非常に裏が出やすいコイン投げで確率的に定まるものとする．

$$z_{nk} \sim \text{Bernoulli}(\pi_k) \tag{29}$$

次に，$\boldsymbol{\pi}$，$\boldsymbol{W}$，$\boldsymbol{H}$ に対して事前分布を導入する．まず，$\boldsymbol{W}$ 及び $\boldsymbol{H}$ に対しては，GaP-KL-NMF と同様に，式 (10) および式 (11) を用いる．コイン投げの確率 π_k に対しては，ベータ事前分布を仮定する．

$$\pi_k \sim \text{Beta}\left(\frac{\alpha c}{K}, \frac{\alpha(K-c)}{K} \right) \tag{30}$$

ここで，$\alpha > 0$ 及び $c > 0$ は超パラメータであり，$\frac{\alpha c}{K}$ が小さくなるほど 0 が出る確率が大きくなる．ただし，$\mathbb{E}_{\text{prior}}[z_{nk}] = \frac{c}{K}$，$\mathbb{E}_{\text{prior}}[\sum_k z_{nk}] = c$ である．

ここで，式 (10)，式 (11) 及び式 (29) で構成される有限モデルに対して，$K \to \infty$ となる極限を考えると，以下のベータ過程が得られる．

$$G \sim \text{BP}(\alpha, G_0) \tag{31}$$

ここで，G_0 は空間 U（$\boldsymbol{w} \in \mathbb{R}_+^M$ と $\boldsymbol{h} \in \mathbb{R}_+^N$ の直積空間）上に定義された基底測度であり，$G_0(U) = c$ を満たす．このとき，G は U 上の離散測度となり，空間 U の無限小区間への分割 $\{U_k\}_{k=1}^{\infty}$ に対して

$$G(U_k) \sim \text{Beta}\big(\alpha G_0(U_k), \alpha(1 - G_0(U_k))\big) \tag{32}$$

が成立している．ただし，$\mathbb{E}[G] = G_0$ であり，$G(U_k) = \pi_k$ である．ベータ過程はコルモゴロフの拡張定理を満たさないため，ディリクレ過程やガンマ過程と異なり，任意の分割に対する周辺分布が陽に求まらないことに注意する．α は集中度と呼ばれ，α が小さくなるほど $\boldsymbol{\theta}$ はよりスパースになる．計算機上では $K \to \infty$ は扱えないが，K を α に比べて十分大きな値に設定すれば，式 (30) はベータ過程の良い近似となる（弱極限近似）．

3.3.2 変分ベイズ法

式 (6), (10), (11), (29), (30) で定義される BP-KL-NMF では，通常の VB の適用は困難である．一見，GaP-KL-NMF と同様に，Jensen の不等式を用いて対数尤度の変分下限が得られるように思われる．実際，計算を行うと

$$\log p(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{H}, \boldsymbol{W}) \qquad (33)$$
$$\stackrel{c}{\geq} \sum_{knm} x_{nm}\lambda_{knm} \log(z_{nk} w_{km} h_{kn}) - \sum_{knm} z_{nk} w_{km} h_{kn}$$

が得られるが，z_{nk} は二値変数であるため，第一項の対数の中の値がゼロとなってしまう場合が存在する．したがって，k, n, m に関する和に分解することができない．

3.3.3 ギブスサンプリング

本節では，ギブスサンプリングに基づく近似推論法について説明する [25]．本手法の肝は，二値行列 $\boldsymbol{Z}$ のサンプリングにある．具体的には，$\boldsymbol{\pi}$，$\boldsymbol{Z}_{\neg nk}$，$\boldsymbol{W}$, $\boldsymbol{H}$ が与えられたもとで（$\neg$ はそのインデクス以外の変数の集合を表す），あるフレーム n における基底 k の状態を表す二値変数 z_{nk} をサンプリングする際に，基底 k がオンの場合とオフの場合とに分けて，尤度を計算する．

$$\begin{aligned}
& p(z_{nk}=1|\boldsymbol{\pi}, \boldsymbol{Z}_{\neg nk}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{X}) \\
& = \gamma\pi_k \prod_m \mathrm{Poisson}(x_{nm}|y_{nm}^{\neg k} + w_{km}h_{kn}) \\
& = \gamma'\pi_k \prod_m (y_{nm}^{\neg k} + w_{km}h_{kn})^{x_{nm}} e^{-w_{km}h_{kn}} \qquad (34)
\end{aligned}$$

$$\begin{aligned}
& p(z_{nk}=0|\boldsymbol{\pi}, \boldsymbol{Z}_{\neg nk}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{X}) \\
& = \gamma(1-\pi_k) \prod_m \mathrm{Poisson}(x_{nm}|y_{nm}^{\neg k})) \\
& = \gamma'(1-\pi_k) \prod_m (y_{nm}^{\neg k})^{x_{nm}} \qquad (35)
\end{aligned}$$

ここで，$y_{nm}^{\neg k} = \sum_{i \neq k} w_{im}h_{in}$ であり，γ および γ' は正規化定数である．また，式 (29) 及び (30) には共役性が成立しているため，パラメータ $\boldsymbol{\pi}$ を積分消去することで，効率的な周辺化ギブスサンプリングを行うこともできる．

$$\begin{aligned}
& p(z_{nk}=1|\boldsymbol{Z}_{\neg nk}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{X}) \qquad (36) \\
& \propto \left(\sum_{n' \neq n} z_{n'k} + \frac{\alpha c}{K}\right) \prod_m (y_{nm}^{\neg k} + w_{km}h_{kn})^{x_{nm}} e^{-w_{km}h_{kn}}
\end{aligned}$$

$$\begin{aligned}
& p(z_{nk}=0|\boldsymbol{Z}_{\neg nk}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{X}) \\
& \propto \left(\sum_{n' \neq n} (1 - z_{n'k}) + \frac{\alpha(K-c)}{K}\right) \prod_m (y_{nm}^{\neg k})^{x_{nm}} \qquad (37)
\end{aligned}$$

その他のパラメータの推論方法は GaP-KL-NMF と同様であり，**Algorithm 3** に更新則を示す．GaP-KL-NMF と同じく，π_k が非常に小さな値となった場合であっても，確率的に復活する可能性があるので，安易に削除してはいけない．音源分離を行う際には，各パラメータの条件付き事後分布の期待値を用いるのが一般的である．

Algorithm 3 BP-KL-NMF に対するギブスサンプリング

Require: 非負値行列 $\boldsymbol{X} \in \mathbb{R}_+^{M \times N}$, 最大基底数 K，ベータ過程のパラメータ α，c，ガンマ事前分布のパラメータ $a_0^w, b_0^w, a_0^h, b_0^h$

```
パラメータ Z, W, H をランダムに初期化
while not converged do
    λ_knm ∝ y_knm
    z_nk ~ 式 (36) および (37)
    λ_knm ∝ y_knm
    w_km ~ G(a_0^w + Σ_n λ_knm x_nm, b_0^w + Σ_n z_nk h_kn)
    λ_knm ∝ y_knm
    h_kn ~ G(a_0^h + Σ_m λ_knm x_nm, b_0^h + Σ_m z_nk w_km)
end while
return パラメータ Z, W, H
```

4. ノンパラメトリックベイズ PLCA

本章では，PLCA の音源分離への適用方法について述べ，ディリクレ過程に基づくノンパラメトリックベイズ拡張について提案する．

4.1 PLCA に基づく分離

PLCA においては，混合音の振幅スペクトログラム $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots \boldsymbol{x}_n] \in \mathbb{R}^{M \times N}$ をヒストグラムデータであるとみなす．したがって，あらかじめ整数値になるように量子化しておく必要がある．経験的には，時間・周波数平面上における振幅の平均が 1 となるように混合音のスペクトログラム正規化し，四捨五入をしておくことで，PLCA が期待通り動作することが多かった．ここで，量子化されたひとつひとつの「サンプル」を x_i $(1 \leq i \leq I)$ で表すものとする．I はサンプル数であり，ヒストグラムの総和であるので $I = \sum_{nm} x_{nm}$ となる．

x_i は確率変数であり，その取りうる値は，フレームインデクスおよび周波数インデクスのペアである．具体的には，時間・周波数平面に大量の粒子がばらまかれ，積み重なってたくさんの山ができている場面を想像するとわかりやすい．ある粒子 x_i が時間・周波数平面に投げ込まれたとき，その粒子がたまたまフレーム n・周波数 m で止まったとすると，$x_i = (n, m)$ となる．このとき，x_i が従う時間・周波数平面上の確率分布を $p(n, m)$ とする．時間・周波数平面上のヒストグラム（混合音のスペクトログラム）は，$p(n, m)$ から得られたサンプル群である．

PLCA では，$p(n, m)$ が因子分解できることを仮定する．

$$p(n, m) = p(n) \sum_{k=1}^{K} p(k|n) p(m|k) \qquad (38)$$

すなわち，x_i の生成過程では，まず，$p(n)$ に従ってフレーム n が選択され，次に，$p(k|n)$ に従って基底 k が選択され，最後に，$p(m|k)$ に従って周波数 m が選択される．通常，$p(n)$ は観測データから経験的に求めることができるので，$p(k|n)$ および $p(m|k)$ を EM アルゴリズムを用いて推定することが行われている [12]．

© 2016 Information Processing Society of Japan

確率分布が推定されれば，NMF と同じくウィナーフィルタを用いて音源分離ができる．

$$\mathbb{E}[\tilde{x}_{knm}|\tilde{x}_{nm}] = \frac{p(n,m,k)}{p(n,m)}\tilde{x}_{nm} \tag{39}$$

ここで，$p(n,m,k) = p(n)p(k|n)p(m|k)$ である．

4.2 ディリクレ過程に基づく PLCA

本節では，DP-PLCA について説明する．まず，ベイズモデルを定式化し，VB および周辺化ギブスサンプリングの適用について述べる．

4.2.1 ベイズモデルの定式化

まず，観測変数 $\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_I\}$ および潜在変数 $\boldsymbol{Z} = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_I\}$ を定義しておく．ヒストグラム中の各サンプルは，離散的なフレームインデクスと離散的な周波数インデクスを取りうるため，サンプル i の値を $\boldsymbol{x}_i \in \{0,1\}^{M\times N}$ として 1-of-MN 表現で表す．すなわち，サンプル i がフレーム n・周波数 m を選択する場合，$x_{inm} = 1$ であり，$\boldsymbol{x}_i$ 中のそれ以外の要素はすべて 0 である．また，各サンプルはある基底 k をひとつ選択するため，サンプル i に対応する潜在変数 $\boldsymbol{z}_i \in \{0,1\}^K$ を 1-of-K 表現で表す．すなわち，サンプル i が基底 k を選択する場合，$z_{ik} = 1$ であり，$\boldsymbol{z}_i$ 中のそれ以外の要素はすべて 0 である．

さらに，式 (38) における確率変数 n, m, k の間の依存性を考慮して，以下の等価な表現に変換しておく．

$$p(n,m) = \sum_{k=1}^{K} p(k)p(n|k)p(m|k) \tag{40}$$

これは，k について対称な形となっており，$p(k)$ を混合比，$p(n|k)p(m|k)$ を要素分布とみなせば，$p(n,m)$ が混合分布として定式化されていることが分かる．したがって，$K \to \infty$ として無限混合分布を考える際に，ディリクレ過程を利用することができる．一方，式 (38) においては，フレーム n ごとに混合分布 $p(m|n) = \sum_{k=1}^{K} p(k|n)p(m|k)$ を考えているので，無限化する際には，LDA と同じ理由で HDP を用いる必要が出てきてしまい，複雑になる．

LDA には HDP が必要であるのに，PLCA では DP で十分である理由は，PLCA では PLSA と同様にフレームの生起確率 $p(n)$ を考慮しているからである．この結果，式 (38) から式 (40) への変換が可能になっている．しかし，$p(n)$ はすでに観測済みのフレーム N 個に対する確率分布であり，未知の観測データ（新たなフレーム）の生成機構を有しておらず，「生成モデル」としては完全とは言えない．一方，LDA では，$p(n)$ を考慮しないことにより，新たに観測される文書に対する汎化能力を獲得している点に注意が必要である．PLCA のベイズ化においても同様のアプローチが可能であるが，予備実験においては正しく音源分離を行うことができなかった．

いま，$p(k), p(n|k), p(m|k)$ はすべて標準的な離散分布であるとし，それらのパラメータを $p(k) = \pi_k, p(m|k) = \theta_{km}$, $p(n|k) = \phi_{kn}$ としておく．ここで，$\sum_k \pi_k = 1$, $\sum_m \theta_{km} = 1$，$\sum_n \phi_{kn} = 1$ を満たすものとする．これらを用いると，PLCA のベイズモデルを定式化できる．まず，$\boldsymbol{X}$ と $\boldsymbol{Z}$ に対する尤度関数は以下の通り与えられる．

$$\boldsymbol{z}_i \sim \mathrm{Categorical}(\boldsymbol{\pi}) \tag{41}$$

$$\boldsymbol{x}_i \sim \prod_{n=1}^{N}\prod_{m=1}^{M} (\theta_{km}\phi_{kn})^{x_{inm}z_{ik}} \tag{42}$$

次に，パラメータ $\boldsymbol{\pi}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ に対する事前分布を導入する．まず，$\boldsymbol{\theta}$, $\boldsymbol{\phi}$ に対しては，共役事前分布であるディリクレ分布を用いると都合がよい．

$$\boldsymbol{\theta}_k \sim \mathrm{Dir}(\beta) \tag{43}$$

$$\boldsymbol{\phi}_k \sim \mathrm{Dir}(\gamma) \tag{44}$$

ここで，β および γ は超パラメータであり，それぞれのディリクレ分布について，全ての次元で同じ値を取るものとする．さらに，$\boldsymbol{\pi}$ に対しては，DP の構成法の一つである棒折り過程 (SBP) に基づく事前分布を仮定する．

$$\pi_k = v_k \prod_{k'=1}^{k-1} (1 - v_{k'}) \tag{45}$$

$$v_k \sim \mathrm{Beta}(1, \alpha) \tag{46}$$

ここで，α は超パラメータであり，$\boldsymbol{\pi}$ が $\boldsymbol{v}$ に変数変換されていることに注意する．あるいは，より簡便にディリクレ分布を導入する方法もある．

$$\boldsymbol{\pi} \sim \mathrm{Dir}\left(\frac{\alpha}{K}\right) \tag{47}$$

ここで，式 (43)，式 (44) 及び式 (47) で構成される有限モデルに対して，$K \to \infty$ となる極限を考えると，以下のディリクレ過程が得られる．

$$G \sim \mathrm{DP}(\alpha, G_0) \tag{48}$$

ここで，G_0 は空間 U（$\boldsymbol{\theta} \in \mathbb{R}_+^M$ と $\boldsymbol{\phi} \in \mathbb{R}_+^N$ の直積空間）上に定義された確率測度であり，$G_0(U) = 1$ を満たす．このとき，G は U 上の離散測度となり，空間 U の任意の分割 $\{U_i\}_{i=1}^{I}$ に対して

$$\begin{aligned}&[G(U_1), G(U_2), \cdots, G(U_I)]\\ &\quad \sim \mathrm{Dir}(\alpha G_0(U_1), \alpha G_0(U_2), \cdots, \alpha G_0(U_I))\end{aligned} \tag{49}$$

が成立している．ただし，$\mathbb{E}[G] = G_0$ である．無限小区間への分割を $\{U_k\}_{k=1}^{\infty}$ とすると，$G(U_k) = \pi_k$ である．α は集中度と呼ばれ，α が小さくなるほど $\boldsymbol{\pi}$ はよりスパースになる．計算機上では $K \to \infty$ は扱えないが，K を α に比べて十分大きな値に設定すれば，式 (47) は DP の良い近似となる（弱極限近似）．一般には，SBP に基づく事前分布である式 (45) を用いた方が近似誤差が小さく，良い結果を与えることが多い．あとで触れるように，無限次元の $\boldsymbol{\pi}$ を積分消去した CRP 表現に基づく推論も可能である．

© 2016 Information Processing Society of Japan

4.2.2 変分ベイズ法

DP-PLCA においても，真の事後分布 $p(\boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \phi|\boldsymbol{X})$ を求めることは困難であるが，VB を用いて近似的に事後分布を推定することができる．具体的には，因子分解できる形に限定した変分事後分布 $q(\boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \phi) = q(\boldsymbol{Z})q(\boldsymbol{v})q(\boldsymbol{\theta})q(\phi)$ を考え，真の事後分布との KL ダイバージェンスを最小化するように反復最適化を行う．

$$q(\boldsymbol{Z}) \propto \exp(\mathbb{E}_{q(\boldsymbol{v},\boldsymbol{\phi},\boldsymbol{\theta})}[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\phi})]) \tag{50}$$

$$q(\boldsymbol{v}) \propto \exp(\mathbb{E}_{q(\boldsymbol{Z},\boldsymbol{\phi},\boldsymbol{\theta})}[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\phi})]) \tag{51}$$

$$q(\boldsymbol{\theta}) \propto \exp(\mathbb{E}_{q(\boldsymbol{Z},\boldsymbol{v},\boldsymbol{\phi})}[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\phi})]) \tag{52}$$

$$q(\boldsymbol{\phi}) \propto \exp(\mathbb{E}_{q(\boldsymbol{Z},\boldsymbol{v},\boldsymbol{\theta})}[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\phi})]) \tag{53}$$

これに従って計算を進めると，$\boldsymbol{Z}$ の変分事後分布は以下の通り求まる．

$$q(\boldsymbol{z}_i) = \mathrm{Categorical}(\boldsymbol{\eta}_i) \tag{54}$$

ここで，$\eta_{ik} = \frac{\rho_{ik}}{\sum_{k'=1}^{K} \rho_{ik'}}$ であり，ρ_{ik} は次式で求まる．

$$\begin{aligned} \log \rho_{ik} &= \mathbb{E}[\log v_k] + \sum_{k'=1}^{k-1} \mathbb{E}[\log(1 - v_{k'})] \\ &+ \sum_{nm} x_{inm} \mathbb{E}[\log \theta_{km}] + \sum_{nm} x_{inm} \mathbb{E}[\log \phi_{kn}] \end{aligned} \tag{55}$$

次に，$\boldsymbol{v}$ の変分事後分布は以下の通り求まる．

$$\begin{aligned} &q(v_k) \\ &= \mathrm{Beta}\left(1 + \sum_i \mathbb{E}[z_{ik}], \alpha + \sum_i \sum_{k'=k+1}^{K} \mathbb{E}[z_{ik'}]\right) \end{aligned} \tag{56}$$

最後に，$\boldsymbol{\theta}$ および $\boldsymbol{\phi}$ の変分事後分布は以下の通り求まる．

$$q(\boldsymbol{\theta}_k) = \mathrm{Dir}(\boldsymbol{\omega}_k) \tag{57}$$

$$q(\boldsymbol{\phi}_k) = \mathrm{Dir}(\boldsymbol{\lambda}_k) \tag{58}$$

ここで，$\boldsymbol{\omega}_k$ および $\boldsymbol{\lambda}_k$ は以下で求められる．

$$\omega_{km} = \beta + \sum_n x_{inm} \mathbb{E}[z_{ik}] \tag{59}$$

$$\lambda_{kn} = \beta + \sum_m x_{inm} \mathbb{E}[z_{ik}] \tag{60}$$

Algorithm 4 に更新則を示す．GaP-KL-NMF と同じく，反復ごとに，$\mathbb{E}[\pi_k]$ が十分に小さい基底 k を削除していけば，実効的な基底数 K_+ を自動的に推定できる．最終的に，音源分離を行う際には，各パラメータの期待値を用いるのが一般的である．

4.2.3 ギブスサンプリング

DP-PLCA に対しては，中華料理店過程 (CRP) に基づく効率的な周辺化ギブスサンプリングを行うことができる．この手法では，無限個の混合比である $\boldsymbol{\pi}$ が積分消去されており，明示的に実体化する必要がなくなるため，計算機上での取り扱いに都合がよい．また，ディリクレ事前分布の共役性から，他のパラメータである $\boldsymbol{\theta}$ および $\boldsymbol{\phi}$ も同時に積分消去することができ，推論すべき確率変数は $\boldsymbol{Z}$ のみになる．この結果，収束が早く，局所解にも陥りにくい性質のよいサンプラーが構成できる．推論時には，観測データを表現するのに必要な基底数 K を自動的に増減しながら適切な個数を探索することができる．

Algorithm 4 DP-PLCA に対する変分ベイズ法

Require: 観測データ $\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_I\}$, 最大基底数 K，ディリクレ過程のパラメータ α，ディリクレ事前分布のパラメータ β, γ
変分事後分布 $q(\boldsymbol{Z})$, $q(\boldsymbol{v})$, $q(\boldsymbol{\theta})$, $q(\boldsymbol{\phi})$ をランダムに初期化
while not converged **do**
  $q(\boldsymbol{z}_i) = \mathrm{Categorical}(\boldsymbol{\eta}_i)$
  $q(v_k) = \mathrm{Beta}\left(1 + \sum_i \mathbb{E}[z_{ik}], \alpha + \sum_i \sum_{k'=k+1}^{K} \mathbb{E}[z_{ik'}]\right)$
  $q(\boldsymbol{\theta}_k) = \mathrm{Dir}(\boldsymbol{\omega}_k)$
  $q(\boldsymbol{\phi}_k) = \mathrm{Dir}(\boldsymbol{\lambda}_k)$
end while
return 変分事後分布 $q(\boldsymbol{Z})$, $q(\boldsymbol{v})$, $q(\boldsymbol{\theta})$, $q(\boldsymbol{\phi})$

5. おわりに

本稿では，音楽音響信号に対する音源分離のための主要な二つの行列分解技法である NMF と PLCA について，対応する確率モデルの性質を明らかにし，比較検討を行った．具体的には，ガンマ過程およびベータ過程に基づく NMF のノンパラメトリックベイズモデルを紹介し，ディリクレ過程に基づく PLCA を提案した．今後は大規模なデータで比較実験を行っていきたい．

謝辞: 本研究の一部は，JSPS 科研費 24220006, 26700020, 26280089, 16H01744, JST CREST OngaCREST，および栢森情報科学振興財団の支援を受けた．

参考文献

[1] 亀岡弘和. 非負値行列因子分解の音響信号処理への応用. 日本音響学会誌, 68(11):559–565, 2012.
[2] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman. Dynamic source separation using non-negative factorizations: A unified view. *IEEE Signal Processing Magazine*, 31(3):66–75, 2014.
[3] H. Kameoka and K. Kashino. Composite autoregressive system for sparse source-filter representation of speech. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2477–2480, 2009.
[4] K. Yoshii and M. Goto. Infinite composite autoregressive models for music signal analysis. In *International Conference on Music Information Retrieval (ISMIR)*, pages 79–84, 2012.
[5] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning (ICML)*, pages 439–446, 2010.
[6] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Neural Information Processing Systems (NIPS)*, pages 556–562, 2000.
[7] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE*

Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 177–180, 2003.

[8] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.

[9] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[10] D. M. Blei and M. I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–144, 2006.

[11] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[12] M. Shashanka, B. Raj, and P. Smaragdis. Probabilistic latent variable models as nonnegative factorizations. *Computational Intelligence and Neuroscience*, 2008:1–8, 2008.

[13] T. Hofmann. Learning the similarity of documents: An information-geometric approach to document retrieval and categorization. In *Advances in Neural Information Processing Systems (NIPS)*, pages 914–920, 2000.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[15] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

[16] G. Grindlay and D. P. W. Ellis. Transcribing multi-instrument polyphonic music with hierarchical eigeninstruments. *Journal Selected Topics Signal Processing*, 5(6):1159–1169, 2011.

[17] B. Fuentes, R. Badeau, and G. Richard. Harmonic adaptive latent component analysis of audio and application to music transcription. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(9):1854–1866, 2013.

[18] E. Benetos and S. Dixon. Multiple-instrument polyphonic music transcription using a temporally constrained shift-invariant model. *Journal of Acoustical Society of America*, 133(3):1727–1741, 2013.

[19] A. Roychowdhury and B. Kulis. Gamma processes, stick-breaking, and variational inference. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 800–808, 2015.

[20] Y. W. Teh, D. Görür, and Z. Ghahramani. Stick-breaking construction for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 556–563, 2007.

[21] S. K. Gupta, D. Phung, and S. Venkatesh. A non-parametric Bayesian Poisson gamma model for count data. In *International Conference on Pattern Recognition (ICPR)*, pages 1815–1818, 2012.

[22] J. Paisley, L. Carin, and D. Blei. Variational inference for stick-breaking beta process priors. In *International Conference on Machine Learning (ICML)*, pages 889–896, 2011.

[23] T. L. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems (NIPS)*, pages 475–482, 2006.

[24] D. Liang, M. Hoffman, and D. Ellis. Beta process sparse nonnegative matrix factorization for music. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 375–380, 2013.

[25] D. Liang and M. D. Hoffman. Beta process non-negative matrix factorization with stochastic structured mean-field variational inference. In *NIPS Workshop on Advances in Variational Inference*, pages 1–6, 2014.

© 2016 Information Processing Society of Japan