多重音基本周波数解析のための無限潜在的調波配分法

吉井 和佳　後藤 真孝
産業技術総合研究所
k.yoshii@aist.go.jp

研究の意義

不確実性のベイズ的取り扱い

研究の背景

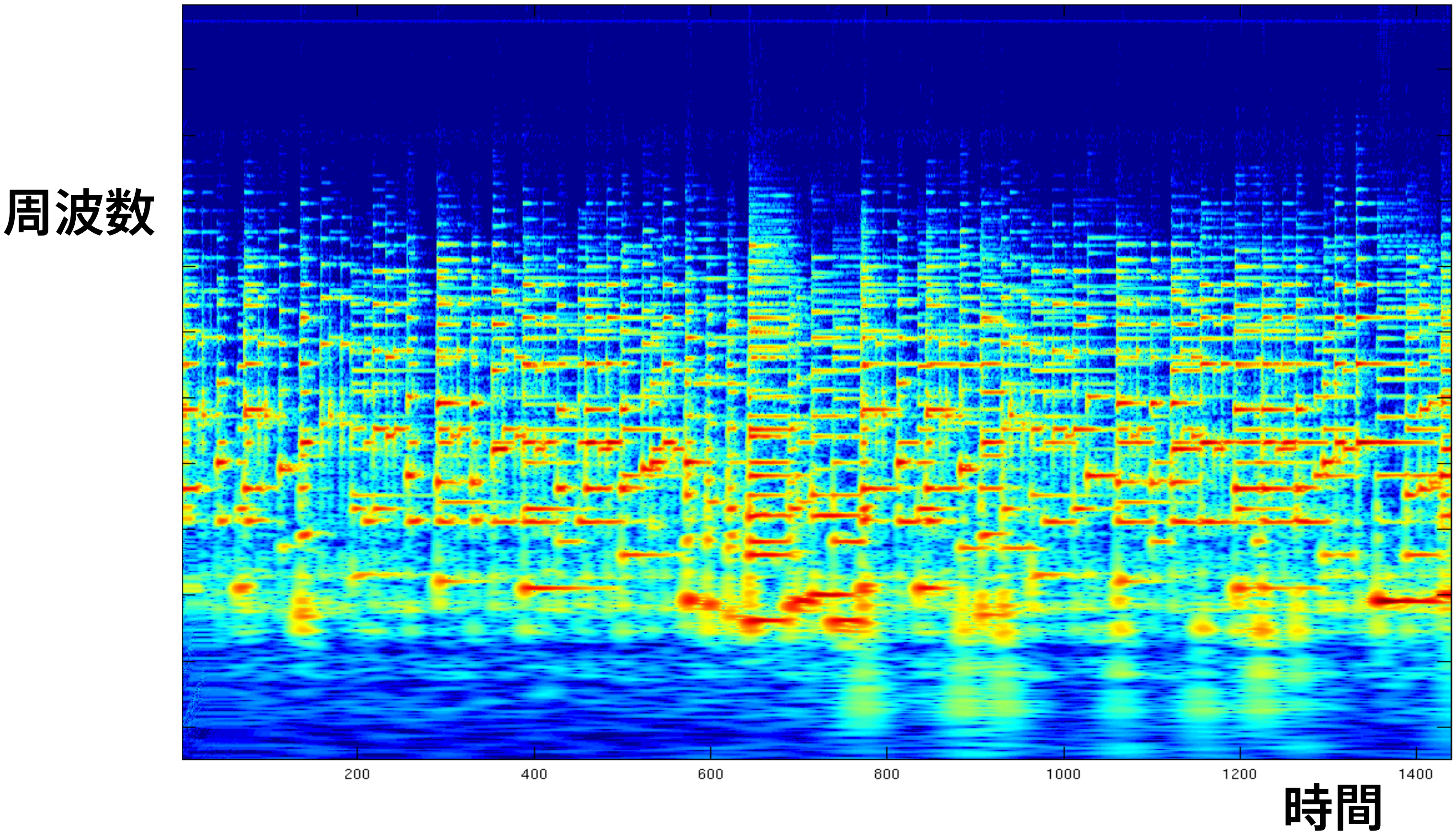
- 基本周波数解析は重要な基礎技術
  - 自動採譜・音源分離・歌手識別・楽器識別
  - 歌詞認識・歌詞アラインメント
- 行列分解手法の利用が一般的
  - 非負値行列因子分解 (NMF)
    - 因子モデル (Factorial Model)
    - 実装が簡単で高速に動作
  - 確率的潜在コンポーネント分析 (PLCA)
    - 混合モデル (Mixture Model)
    - トピックモデルの音源分離への適用
    - スペクトルを「粒子」に分解する必要あり

基底に分解後、それぞれに対してF0を後から推定
→ モデル自体にF0パラメータが入っている方がよい！

複数基本周波数解析の難しさ

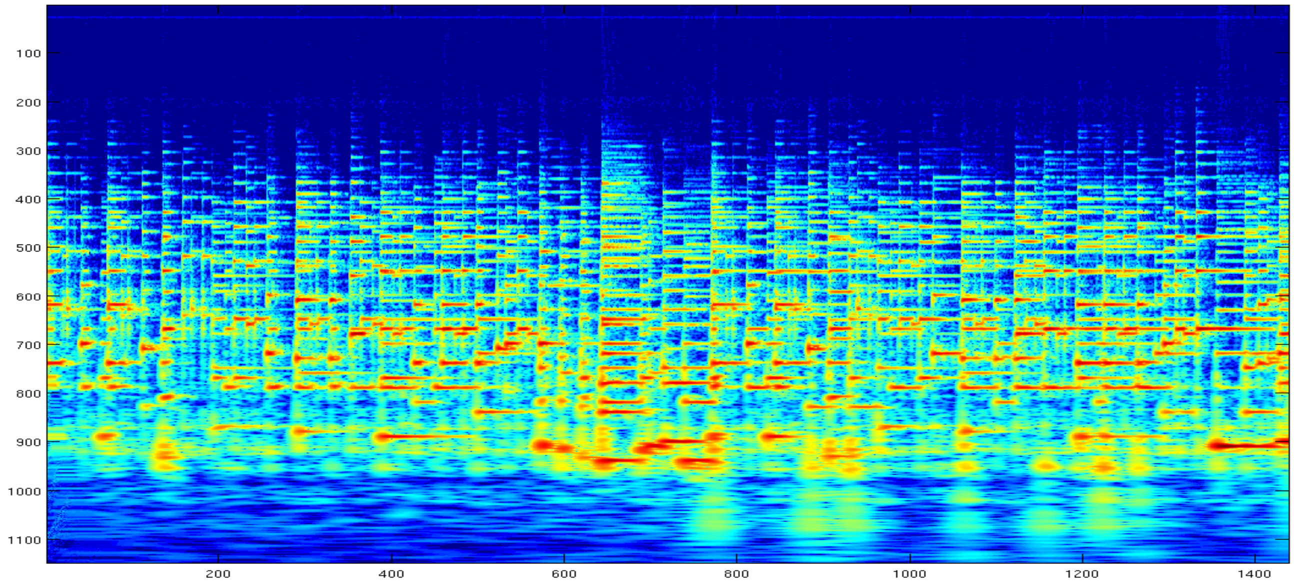
- 倍音が豊富に含まれていてそれらが重畳
  - 混合音声と異なり周波数領域でのスパース性が弱い

→ 基底の形状 (調波構造) に関するパラメトリックな表現が望まれる



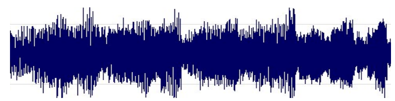
処理すべきデータ量

- 音楽音響信号処理は非常に重い！
  - 典型的な周波数解析の設定
    - 10 [ms]シフト ＝ 5分 の楽曲で 30,000フレーム
    - 2048点FFT ＝ 1フレーム当たり 1,024個のデータ
  - 1曲の持つデータ量 ＝ 3000万サンプル！
    - 並列分散処理がしやすいEMが好まれる傾向あり



研究の動機

- 不確実性 (Uncertainty) を適切に取り扱いたい
  - 音楽解析において必然的に表れる性質



音楽音響信号

例：基本周波数推定

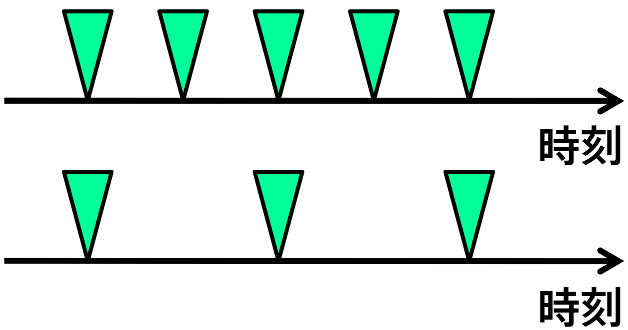
← まずまず自信あり

← あまり自信がない

← 存在を確信！

楽器音が複雑に重なり合うと
解釈の正しさを確信できない

例：ビートトラッキング



テンポ知覚に曖昧性があると
解釈を一意に決めかねる

未知事象の不確実性を表現して伝播させる方法論が必要

「完全な」ベイズ的取り扱い

- 観測データ以外はすべて未知：不確実性が存在
  - パラメータ
    - 基本周波数の値は？
  - 確率モデルの複雑さ
    - いくつの音源・倍音(混合数)が含まれているのか？
  - 超パラメータ
    - 上記に対する事前分布のパラメータの設定方法は？

未知事象の不確実性を表現して伝播させる方法論が必要

	パラメータ	複雑さ	超パラメータ	頑健性
最尤推定	一意に決定	手動で設定	不使用	×
MAP推定	一意に決定	手動で設定	手動で設定	△
古典的なベイズ推定	事後分布を推定	手動で設定	手動で設定	○
ノンパラメトリックベイズ	事後分布を推定	事後分布を推定	(手動で設定)	◎
階層ベイズモデル	事後分布を推定	(手動で設定)	事後分布を推定	◎

階層ノンパラメトリックベイズモデルは何も一意に決定しない！

研究の概要

階層ノンパラメトリックベイズモデル

本研究の概要

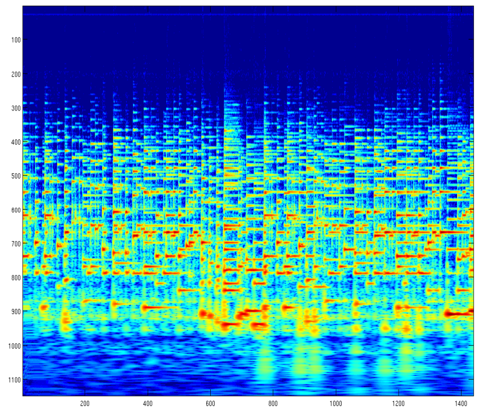
- 問題：多重音中の複数の基本周波数の推定
  - 調波構造を持たない音(ドラム音など)は対象外
  - 音源数は未知
  - 倍音数は未知

- 手法：現代的な統計的機械学習
  - 階層ノンパラメトリックベイズモデル

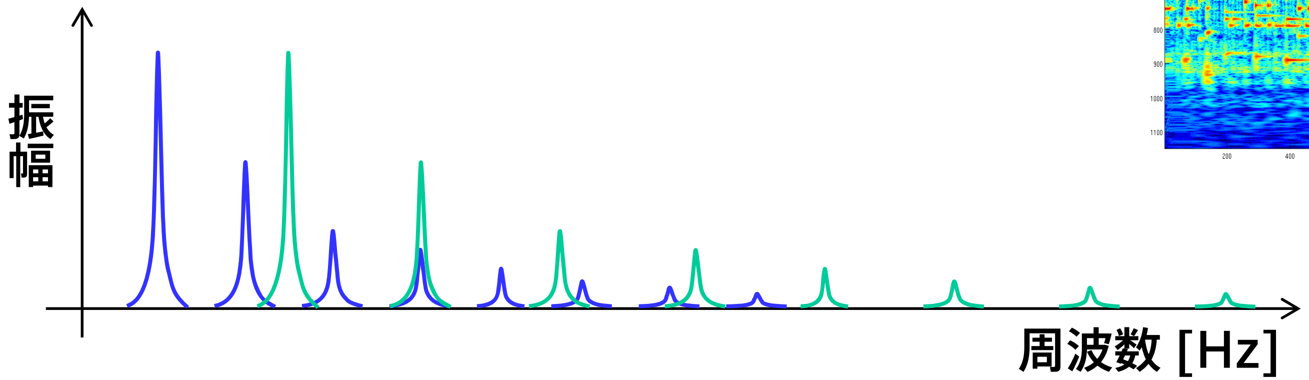
- 成果：完全自動化・高い推定精度を達成
  - 人手で最適化した従来手法と同程度

調波構造

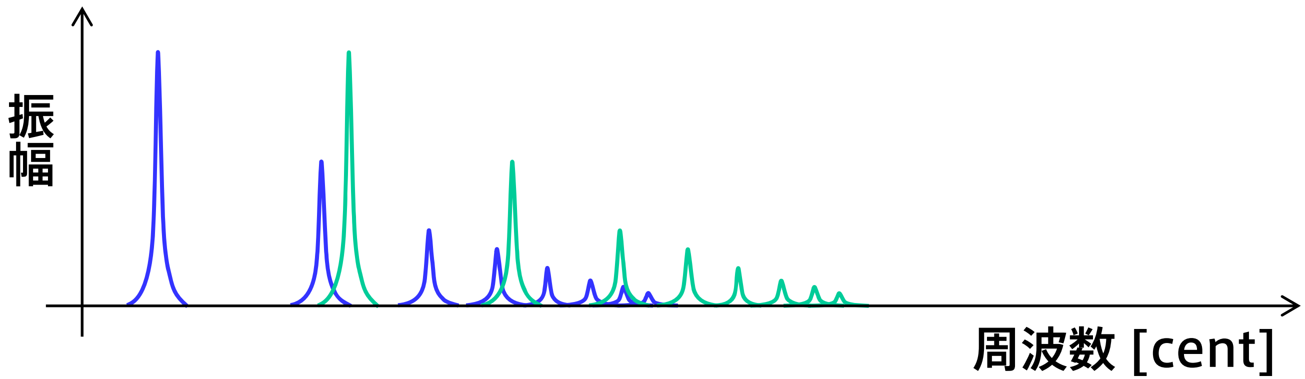
- 基本周波数 (F0) の整数倍にピークを持つ
  - 線形周波数軸だと調波構造は伸縮
  - 対数周波数軸だと調波構造は**平行移動**



フーリエ変換によるスペクトル

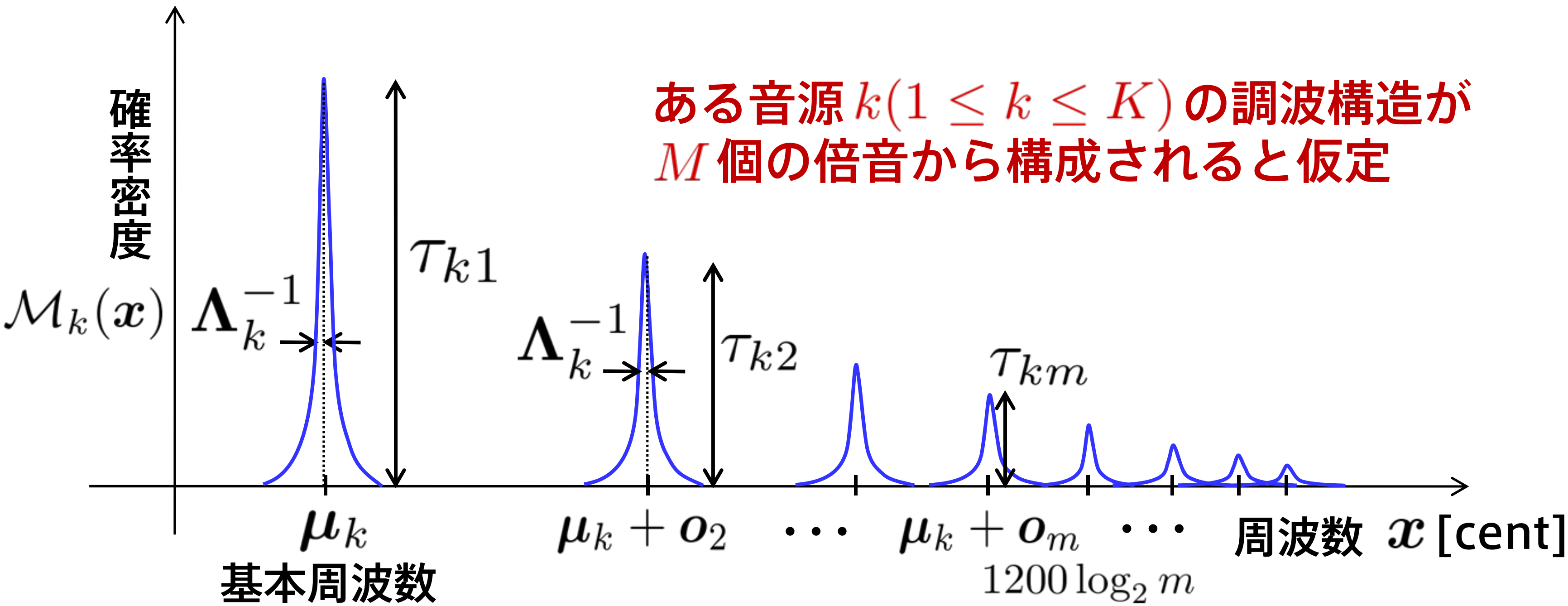


ウェーブレット変換によるスペクトル



単一音源に対する確率分布

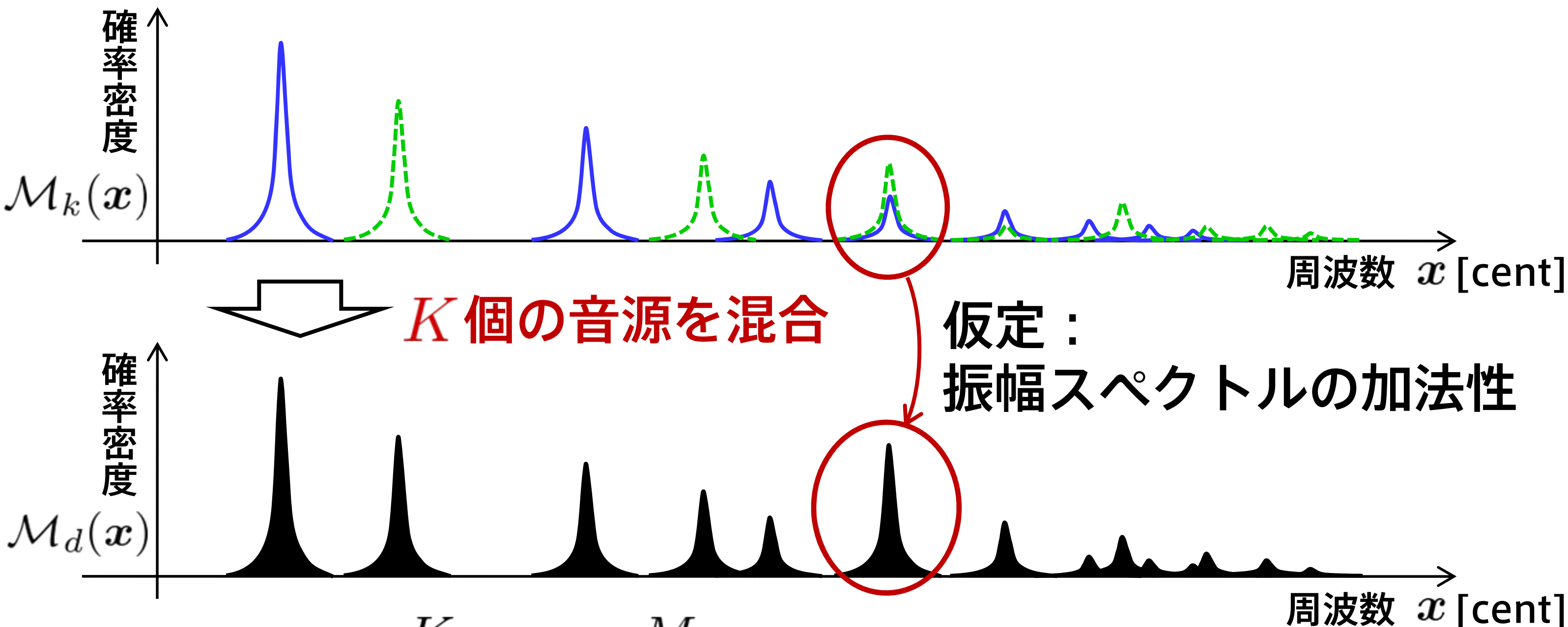
- 音源の調波構造を有限混合ガウス分布で表現
  - PreFEst [後藤1999,2004]
  - Harmonic Clustering [亀岡ら2004]



$$\mathcal{M}_k(\boldsymbol{x}) = \sum_{m=1}^{M} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

複数音源に対する確率分布

- 音源の重畳を**ネスト型**有限混合ガウス分布で表現
  - 単一音源に対する混合ガウス分布を混合する



$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_{dk} \sum_{m=1}^{M} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

フレーム d における音源 k の混合比

音源 k における倍音 m の混合比

本研究のポイント(1)

- 音源の重畳をネスト型**無限混合**ガウス分布で表現
  - 音源数K・倍音数Mを無限大に発散＝森羅万象
  - 観測データは森羅万象のほんの一部という考え方

ネスト型有限混合ガウス分布

$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_{dk} \sum_{m=1}^{M} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

フレームdにおける
音源kの混合比

音源kにおける
倍音mの混合比

ネスト型無限混合ガウス分布

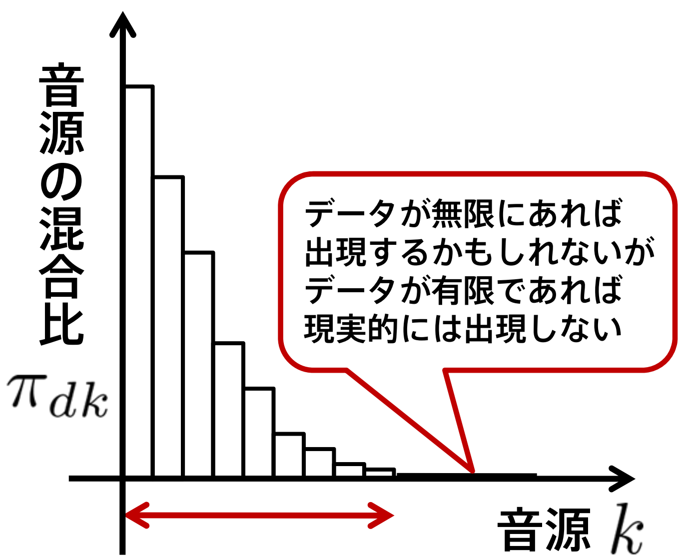
$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{\infty} \pi_{dk} \sum_{m=1}^{\infty} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

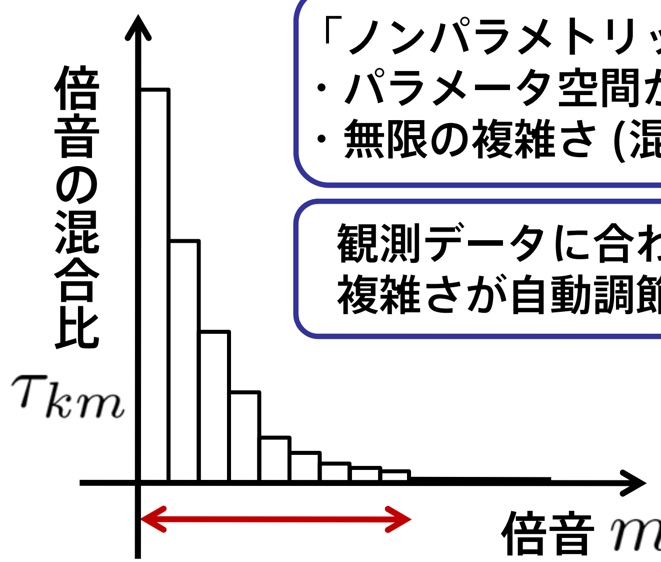
無限個の音源・倍音があるにせよ、有限の観測データ中に現れるのは
「現実的には」高々有限個の音源・倍音だけではないのか？

ノンパラメトリックベイズ

- <u>適切な事前分布</u>を導入してスパースな解に誘導
  - 無限個ある混合比が満たして欲しい性質
    - ごく一部(有限個)がある程度以上大きい
    - その他(無限個)は極めて小さい (ほぼゼロ)

$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{\infty} \pi_{dk} \sum_{m=1}^{\infty} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$



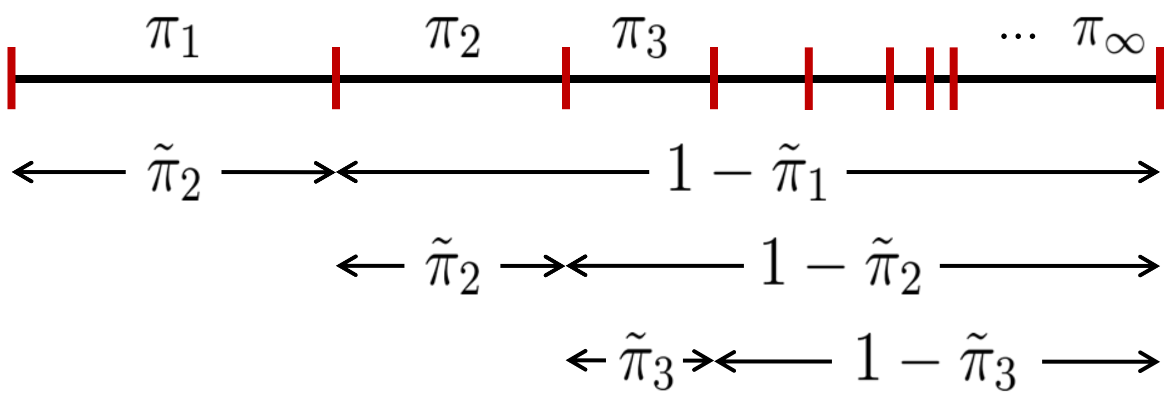


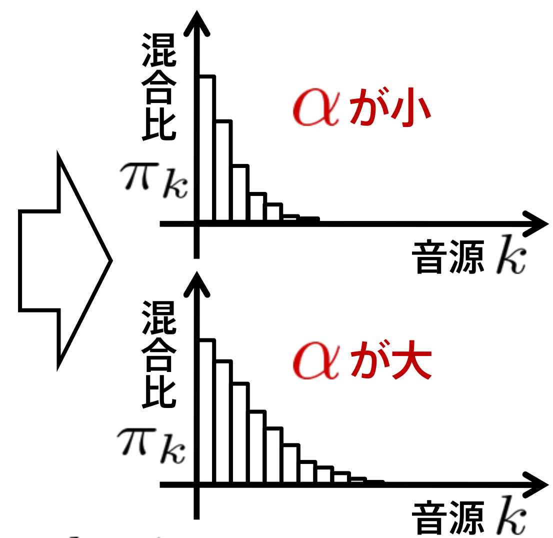
「ノンパラメトリック」の意味：
- パラメータ空間が固定されていない
- 無限の複雑さ (混合数) を考える

観測データに合わせて「実際的な」複雑さが自動調節される！

事前分布の構成法

- ディリクレ過程 (Dirichlet Process: DP)
  - 無限個の混合比に対する事前分布として利用可能
  - 棒折り過程 (Stick-Breaking Construction)
    - 長さ１の棒を無限に再帰的に折りとっていく
    - どのくらいの比で折りとるかを超パラメータ α で制御





$$\tilde{\pi}_k \sim \mathrm{Beta}(1, \alpha)$$

平均的には 1 ： α で折りとる
→ α を大きくするほど、残す棒の長さが長くなる
→ 実際的な混合数が大きくなる (heavy-tailed)

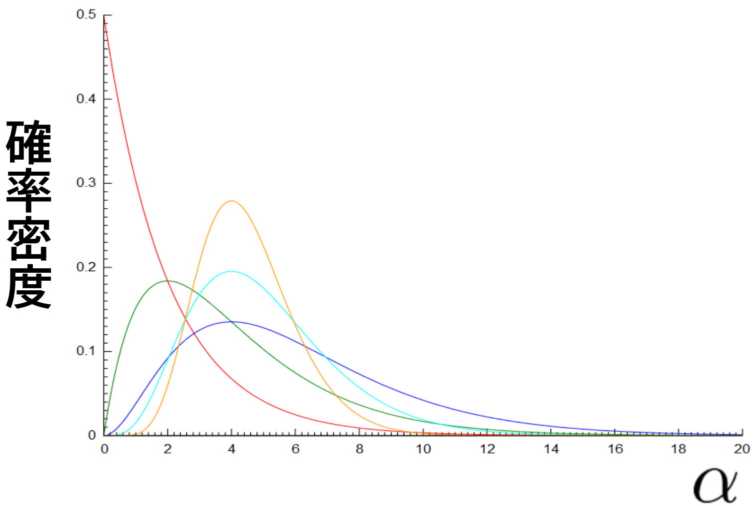
$$\pi_k = \tilde{\pi}_k \prod_{k'=1}^{k-1} (1 - \tilde{\pi}_{k'})$$

本研究のポイント(2)

- 超パラメータに対して超事前分布を設定
  - 超事前分布は一様分布(ほぼ無情報)とする
  - 超パラメータ α の手動設定の困難さを解消
    - 本来未知なのだから不確実性を適切に取り扱うべき

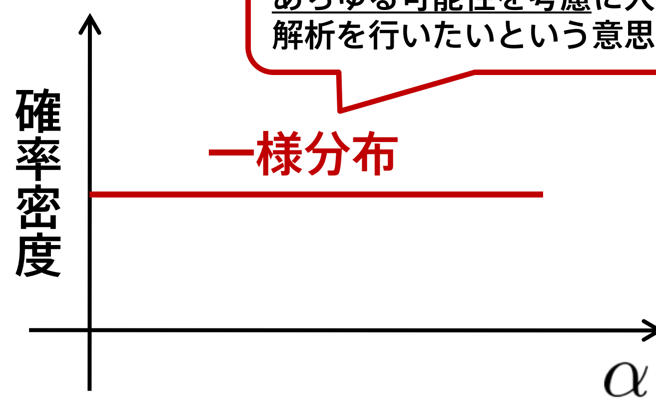
事前分布 $\tilde{\pi}_k \sim \mathrm{Beta}(1, \alpha)$

超事前分布 $\alpha \sim \mathrm{Gam}(a, b)$



ガンマ分布の一般的な形状

αを一意に決定することをせず、
あらゆる可能性を考慮に入れながら
解析を行いたいという意思表示



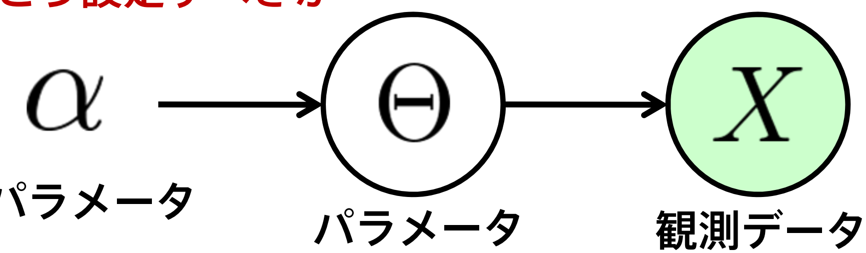
$a = 1, b = 0$ のとき

階層ベイズモデル

- 超パラメータに対して超事前分布を設定
  - パラメータだけでなく超パラメータも未知なので不確実性を適切に取り扱うのが自然

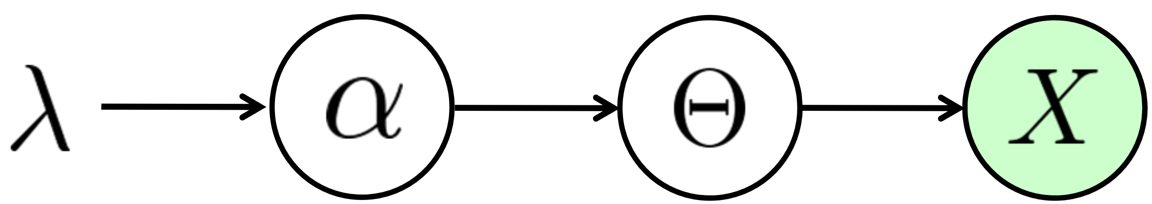
問題：事前分布をどう設定すべきか

古典的な
ベイズモデル



不確実性を考慮して
超パラメータを
確率変数として扱う

階層
ベイズモデル



新たな問題：超事前分布をどう設定すべきか

無情報事前分布：事前知識が特にない時に利用
事後分布に影響をほとんど与えない曖昧な分布

本研究の位置付け

- 混合モデルに基づく基本周波数解析法の究極形
  - 本研究の原点：後藤・亀岡らの調波構造モデル
  - 本研究の貢献：
    - ノンパラメトリックベイズ化（無限混合モデル＋ベイズ推定）
    - 超事前分布を導入した階層ベイズモデル

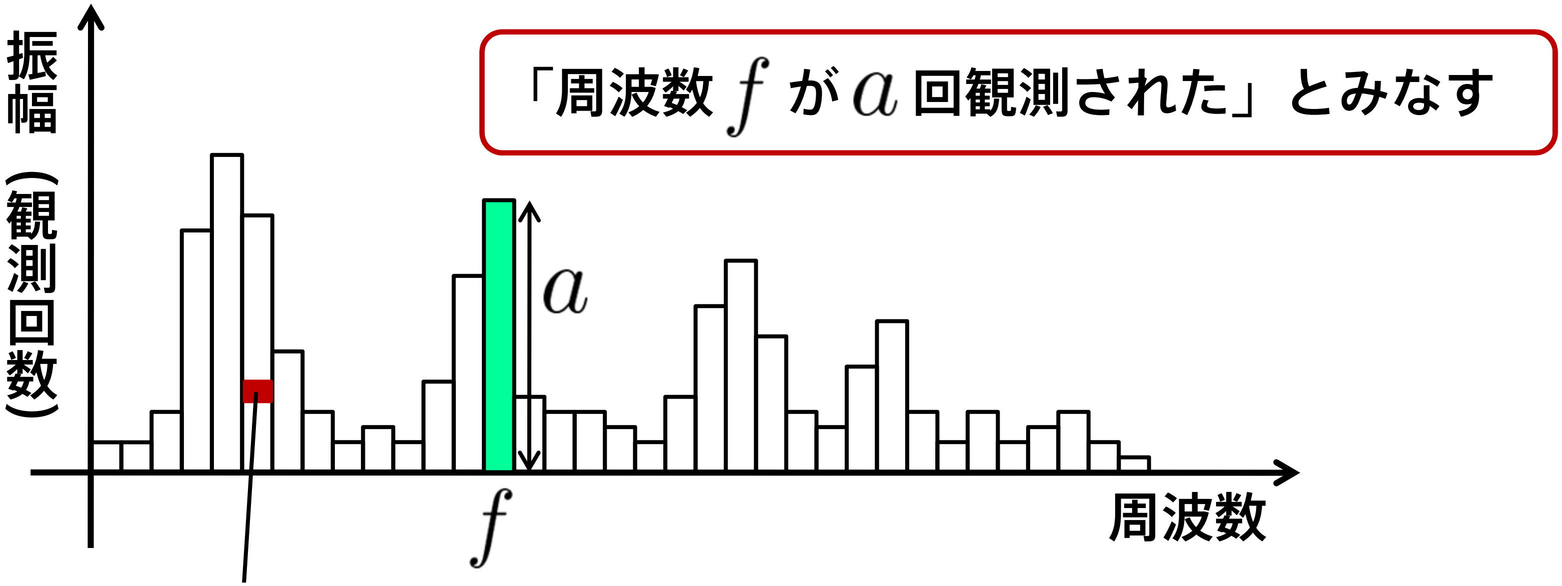
	従来研究	本研究
音源数	固定（事前に指定）	無限（指定不要）
倍音数	固定（事前に指定）	無限（指定不要）
音源の混合比に対する事前分布	―――	階層ディリクレ過程＋無情報超事前分布(チューニング不要)
倍音の混合比に対する事前分布	ディリクレ分布(チューニング必要)	ディリクレ過程＋無情報超事前分布(チューニング不要)
学習方法	MAP推定	ベイズ推定

無限潜在的調波配分法

Infinite Latent Harmonic Allocation (iLHA)

観測変数と潜在変数

- 各フレームの振幅スペクトル＝ヒストグラム
  - さまざまな周波数を大量に観測したと考える
  - 各周波数ビンごとに観測回数をカウント
    - 各周波数の生成は独立であると仮定



観測変数 $\boldsymbol{x}_{dn}$：フレーム d において n 番目に観測された周波数

潜在変数 $\boldsymbol{z}_{dn}$：$\boldsymbol{x}_{dn}$ がどの音源・倍音から生成されたか（KM＝無限個の候補から１つ選ぶ）

本研究のポイント

- 音源の重畳を**ネスト型無限混合ガウス分布**で表現
  - 音源数K・倍音数Mを無限大に発散＝森羅万象
  - 観測データは森羅万象のほんの一部という考え方

ネスト型有限混合ガウス分布

$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_{dk} \sum_{m=1}^{M} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

フレームdにおける音源kの混合比

音源kにおける倍音mの混合比

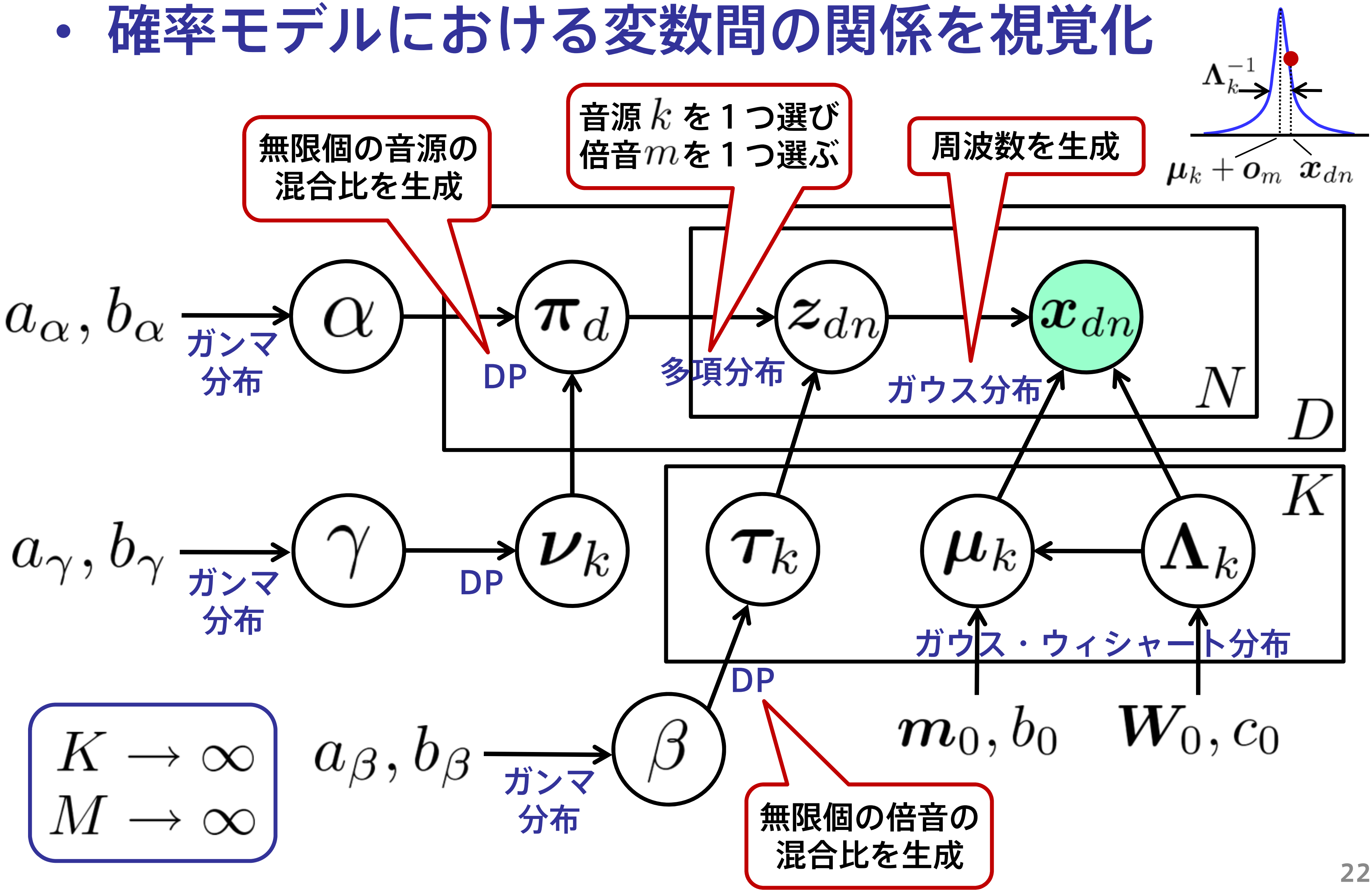
ネスト型無限混合ガウス分布

$$\mathcal{M}_d(\boldsymbol{x}) = \sum_{k=1}^{\infty} \pi_{dk} \sum_{m=1}^{\infty} \tau_{km} \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)$$

無限個の音源・倍音があるにせよ、有限の観測データ中に現れるのは「現実的には」高々有限個の音源・倍音に限られる

グラフィカル表現

- 確率モデルにおける変数間の関係を視覚化



$K \to \infty$
$M \to \infty$

確率モデルの定式化

- 同時分布

$$p(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{\mu},\boldsymbol{\Lambda}) = p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\mu},\boldsymbol{\Lambda})p(\boldsymbol{Z}|\boldsymbol{\pi},\boldsymbol{\tau})p(\boldsymbol{\pi})p(\boldsymbol{\tau})p(\boldsymbol{\mu},\boldsymbol{\Lambda})$$

  - 尤度関数

$$p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\mu},\boldsymbol{\Lambda}) = \prod_{dnkm} \mathcal{N}\left(\boldsymbol{x}_{dn}\middle|\boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{dnkm}}$$

ガウス分布

$$p(\boldsymbol{Z}|\boldsymbol{\pi},\boldsymbol{\tau}) = \prod_{dnkm} \left(\pi_{dk}\tau_{km}\right)^{z_{dnkm}}$$

多項分布

  - 事前分布

$$p(\boldsymbol{\pi}) = \prod_{d} \mathrm{Dir}(\boldsymbol{\pi}_d|\alpha\boldsymbol{\nu}) \propto \prod_{dk} \pi_{dk}^{\alpha\nu_k - 1}$$

$$p(\boldsymbol{\tau}) = \prod_{k} \mathrm{Dir}(\boldsymbol{\tau}_k|\beta\boldsymbol{\upsilon}) \propto \prod_{km} \tau_{km}^{\beta\upsilon_m - 1}$$

ディリクレ分布

$$p(\boldsymbol{\mu},\boldsymbol{\Lambda}) = \prod_{k} \mathcal{N}\left(\boldsymbol{\mu}_k\middle|\boldsymbol{m}_0, (b_0\boldsymbol{\Lambda}_k)^{-1}\right)\mathcal{W}\left(\boldsymbol{\Lambda}_k\middle|\boldsymbol{W}_0, c_0\right)$$

ガウス・ウィシャート分布

確率モデルの定式化

同時分布 $p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \alpha, \boldsymbol{\nu}, \gamma, \beta)$

$$= p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\tau})p(\boldsymbol{\pi}|\alpha, \boldsymbol{\nu})p(\boldsymbol{\nu}|\gamma)p(\boldsymbol{\tau}|\beta)p(\boldsymbol{\mu}, \boldsymbol{\Lambda})p(\alpha)p(\gamma)p(\beta)$$

尤度関数

$$p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{dnkm} \mathcal{N}\left(\boldsymbol{x}_{dn} \middle| \boldsymbol{\mu}_k + \boldsymbol{o}_m, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{dnkm}}$$ ガウス分布

$$p(\boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\tau}) = \prod_{dnkm} \left(\pi_{dk}\tau_{km}\right)^{z_{dnkm}}$$ 多項分布

事前分布

$$\pi_d \sim \mathrm{InfDir}(\alpha\nu)$$ 無限次元のディリクレ分布(DP)

$$\nu \sim \mathrm{Stick}(\gamma)$$ 棒折り過程(DP)

階層ディリクレ過程

$$\tau_k \sim \mathrm{Stick}(\beta)$$ 棒折り過程(DP)

ディリクレ過程

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_k \mathcal{N}\left(\boldsymbol{\mu}_k \middle| \boldsymbol{m}_0, (b_0\boldsymbol{\Lambda}_k)^{-1}\right) \mathcal{W}\left(\boldsymbol{\Lambda}_k \middle| \boldsymbol{W}_0, c_0\right)$$ ガウス・ウィシャート分布

超事前分布

$$\alpha \sim \mathrm{Gam}(a_\alpha, b_\alpha) \qquad \gamma \sim \mathrm{Gam}(a_\gamma, b_\gamma)$$

$$\beta \sim \mathrm{Gam}(a_\beta, b_\beta)$$

確率モデルの学習

- 周辺化変分ベイズ法 (Collapsed VB)
  - 変分ベイズ法(VB)を洗練させた手法
    - 解析的に計算可能な部分を周辺化して変数を削減
    - 変分事後分布を因子分解できる形に限定して最適化

同時分布

$$p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \alpha, \boldsymbol{\nu}, \gamma, \beta)$$

$$= \underline{p(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})}\,\underline{p(\boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\tau})}\,\underline{p(\boldsymbol{\pi}|\alpha, \boldsymbol{\nu})}\,p(\boldsymbol{\nu}|\gamma)\,\underline{p(\boldsymbol{\tau}|\beta)}\,\underline{p(\boldsymbol{\mu}, \boldsymbol{\Lambda})}\,p(\alpha)p(\gamma)p(\beta)$$

尤度関数と事前分布の間で共役性が成立 → 積分消去可能！

周辺分布

$$p(\boldsymbol{X}, \boldsymbol{Z}, \alpha, \boldsymbol{\nu}, \gamma, \beta)$$

$$= p(\boldsymbol{X}|\boldsymbol{Z})p(\boldsymbol{Z}|\alpha, \boldsymbol{\nu}, \beta)p(\boldsymbol{\nu}|\gamma)p(\alpha)p(\gamma)p(\beta)$$

求めたいもの：真の事後分布

$$p(\boldsymbol{Z}, \alpha, \boldsymbol{\nu}, \gamma, \beta|\boldsymbol{X}) \longleftarrow q(\boldsymbol{Z}, \alpha, \boldsymbol{\nu}, \gamma, \beta)$$

変分事後分布(因子分解できる形に限定)

$$q(\boldsymbol{Z}, \alpha, \boldsymbol{\nu}, \gamma, \beta) = q(\alpha)q(\boldsymbol{\nu})q(\gamma)q(\beta)\prod_{dn} q(\boldsymbol{z}_{dn})$$

できる限り真の事後分布に近づくように
変分EMアルゴリズムで反復最適化

評価実験

- 実験条件
  - 使用データ：RWC研究用音楽データベース
    - RWC-MDB-J-2001：ジャズ音楽　から6曲
    - RWC-MDB-C-2001：クラシック音楽　から2曲
    - ギター曲・ピアノ曲の冒頭23秒
  - 周波数解析
    - ガボールウェーブレット変換
    - 周波数分解能：10 [cent]
    - フレームシフト幅：16 [ms]
  - 正解データ
    - MIDIをもとに手動で同期・修正したもの
    - HTCの研究[亀岡ら2007]と大体同じ
  - 評価尺度
    - フレームレベルのF値 (再現率と適合率の調和平均)

実験結果

- iLHAが自動的に最適な性能を発揮
  - 人手チューンした従来手法と同等程度の性能を達成
  - HTCのような時間方向のモデル化で改善の余地あり

	チューンした事前分布 ＋ MAP推定	チューンした事前分布 ＋ MAP推定	無情報事前分布 ＋ ベイズモデル	無情報超事前分布 ＋ 階層ノンパラベイズモデル
	PreFEst	HTC	LHA	iLHA
Jazz No.1	75.8	79.0	70.7	82.2
Jazz No.2	78.5	78.0	69.1	77.9
Jazz No.6	70.4	78.3	49.8	71.2
Jazz No.7	83.0	86.0	70.2	85.5
Jazz No.8	85.7	84.4	55.9	84.6
Jazz No.9	85.9	89.5	68.9	84.7
Classic No.30	76.0	83.6	81.4	81.6
Classic No.35	72.8	76.0	58.9	79.6
合計	79.4	82.0	65.8	81.7

おわりに

- 従来の基本周波数解析法の究極形を提案
  - 本研究の原点
    - ネスト型有限混合ガウスモデルのMAP推定[後藤・亀岡ら]
  - 本研究の貢献
    - ノンパラメトリックベイズ
      - ネスト型無限混合ガウスモデルのベイズ推定
    - 階層ベイズモデル
      - 無情報超事前分布の導入による完全自動化
- 音楽情報処理技術の新しい方向性
  - 最新の機械学習技術の積極的導入
  - 古典的な統計的学習理論からの脱出
  - 計算音楽学 (Computational Musicology)
    - 計算言語学 (Computational Linguistics) のアナロジ