

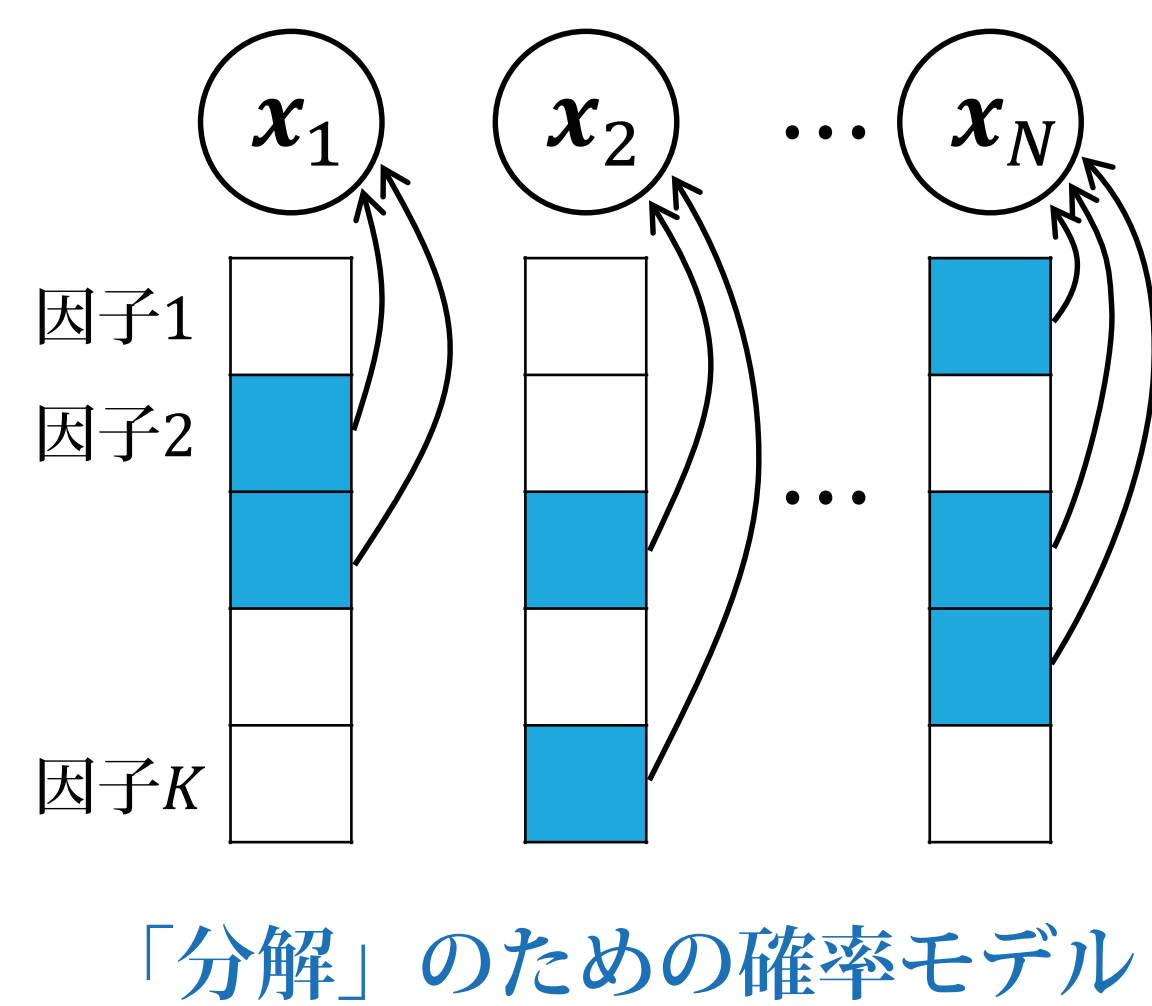
音楽音響信号解析のための ディリクレ過程に基づくベイズ潜在成分分析

吉井 和佳 中村 栄太 糸山 克寿 (京大) 後藤 真孝 (産総研)

潜在要素を発見するための二つのアプローチ：因子モデルと混合モデル

因子モデル (factor model)

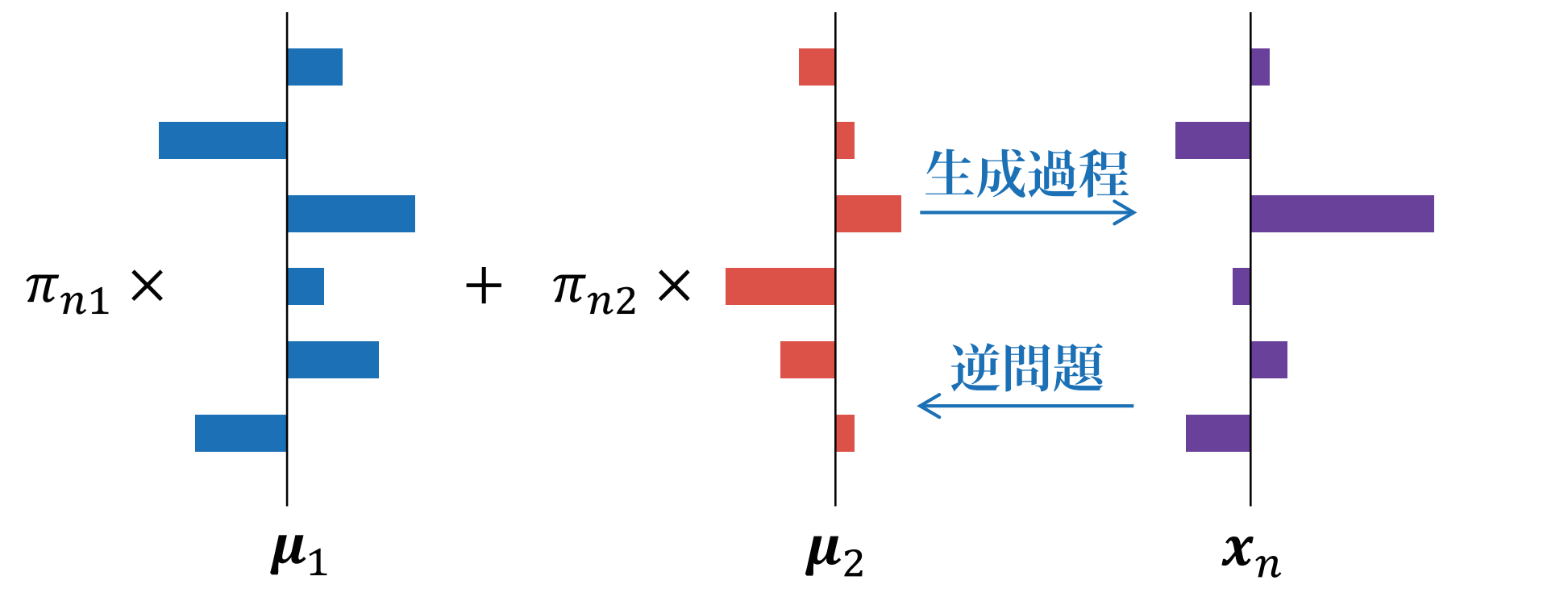
各サンプルは複数の因子の組み合わせで得られる



主成分分析 (PCA)

K個の要素ベクトルを準備
パラメータ： $\theta = \{\mu_k\}_{k=1}^K$

要素を足し合わせた分布から生成
パラメータ： $\pi = \{\pi_n\}_{n=1}^N$

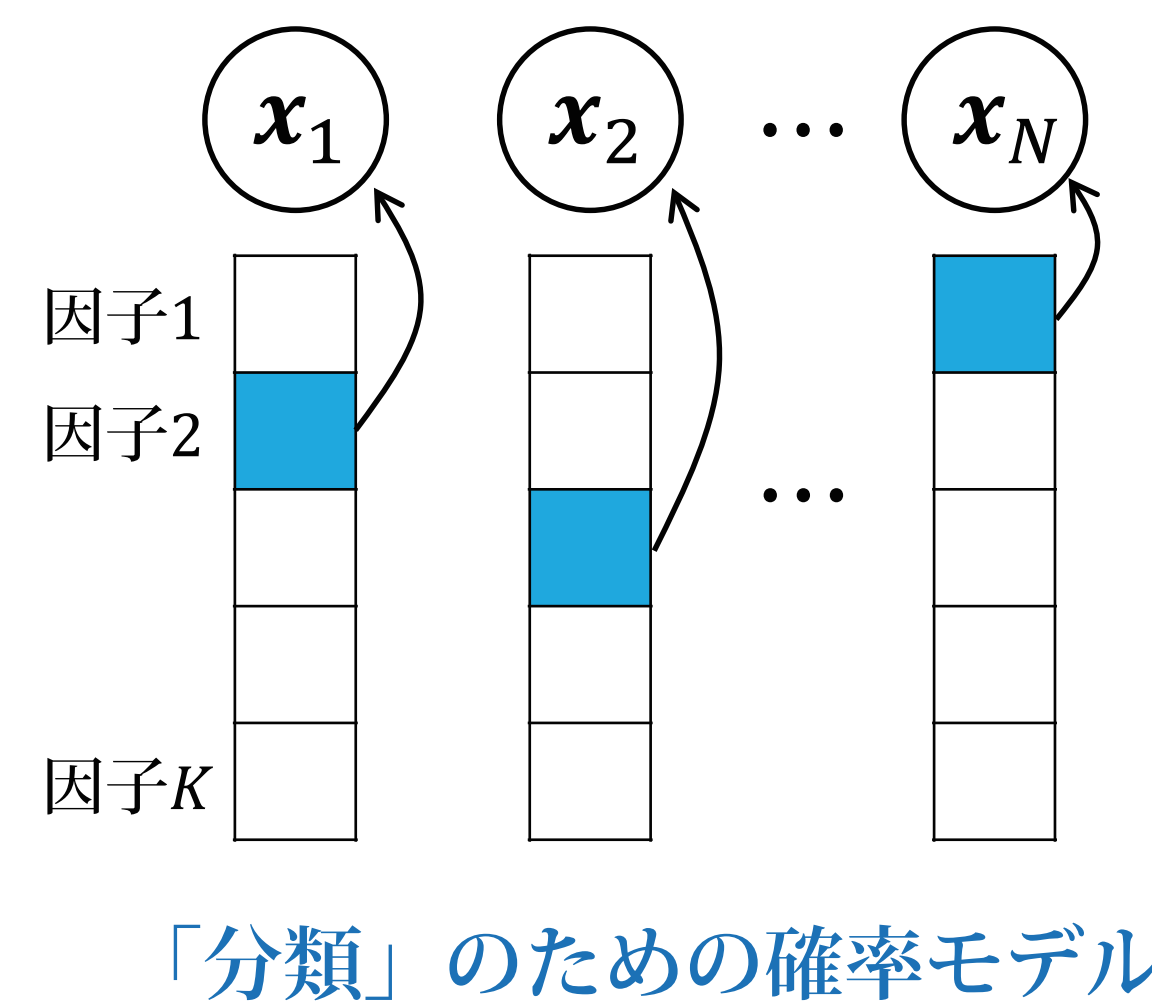


$$x_n \sim N\left(\sum_{k=1}^K \pi_{nk} \mu_k, \Sigma\right)$$

和が確率分布の内側
→ 確率変数の足し合わせ

混合モデル (mixture model)

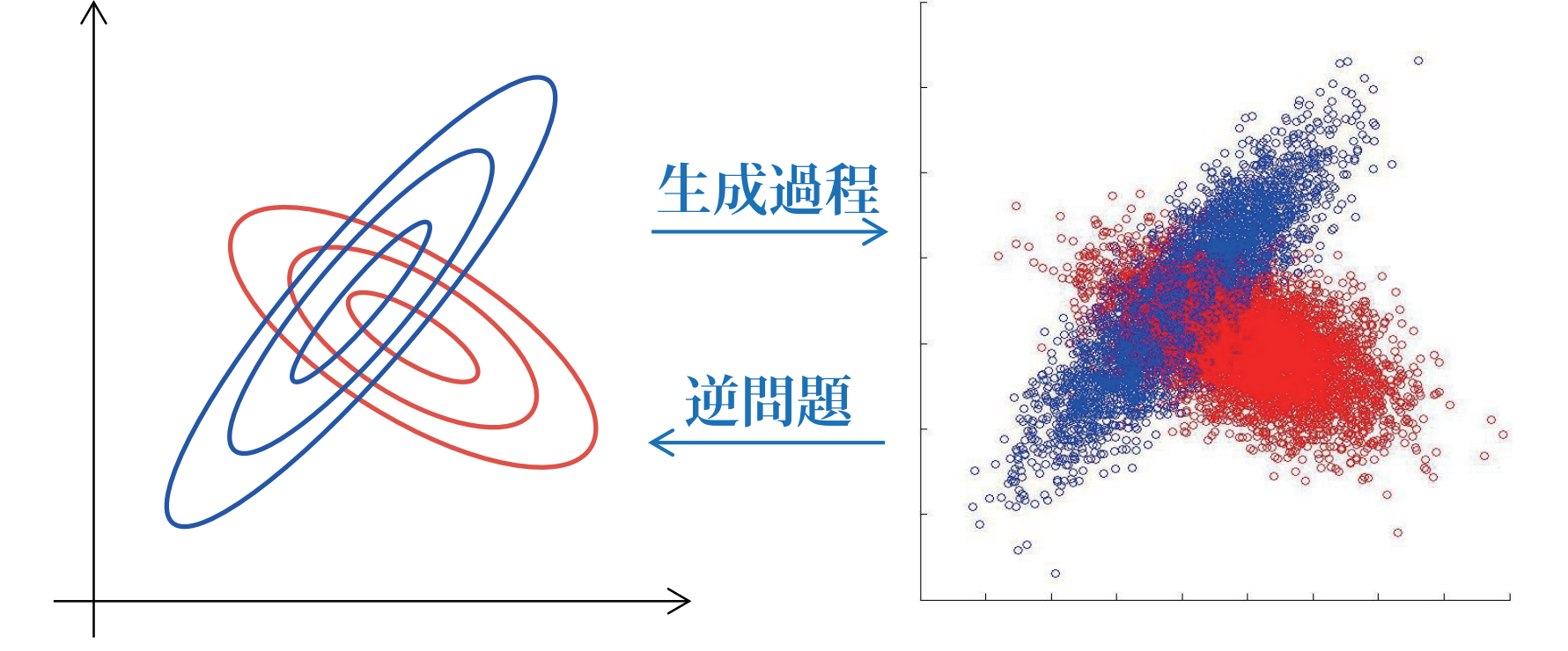
あるサンプルはいずれか一つの因子のみに依存する



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： $\theta = \{\mu_k, \Sigma_k\}_{k=1}^K$

選択した要素分布から生成
パラメータ： $\pi = \{\pi_k\}_{k=1}^K$



$$x_n \sim \sum_{k=1}^K \pi_k N(\mu_k, \Sigma_k)$$

和が確率分布の外側
→ 確率分布の足し合わせ

非負値行列因子分解 (NMF) と確率的潜在成分分析 (PLCA)

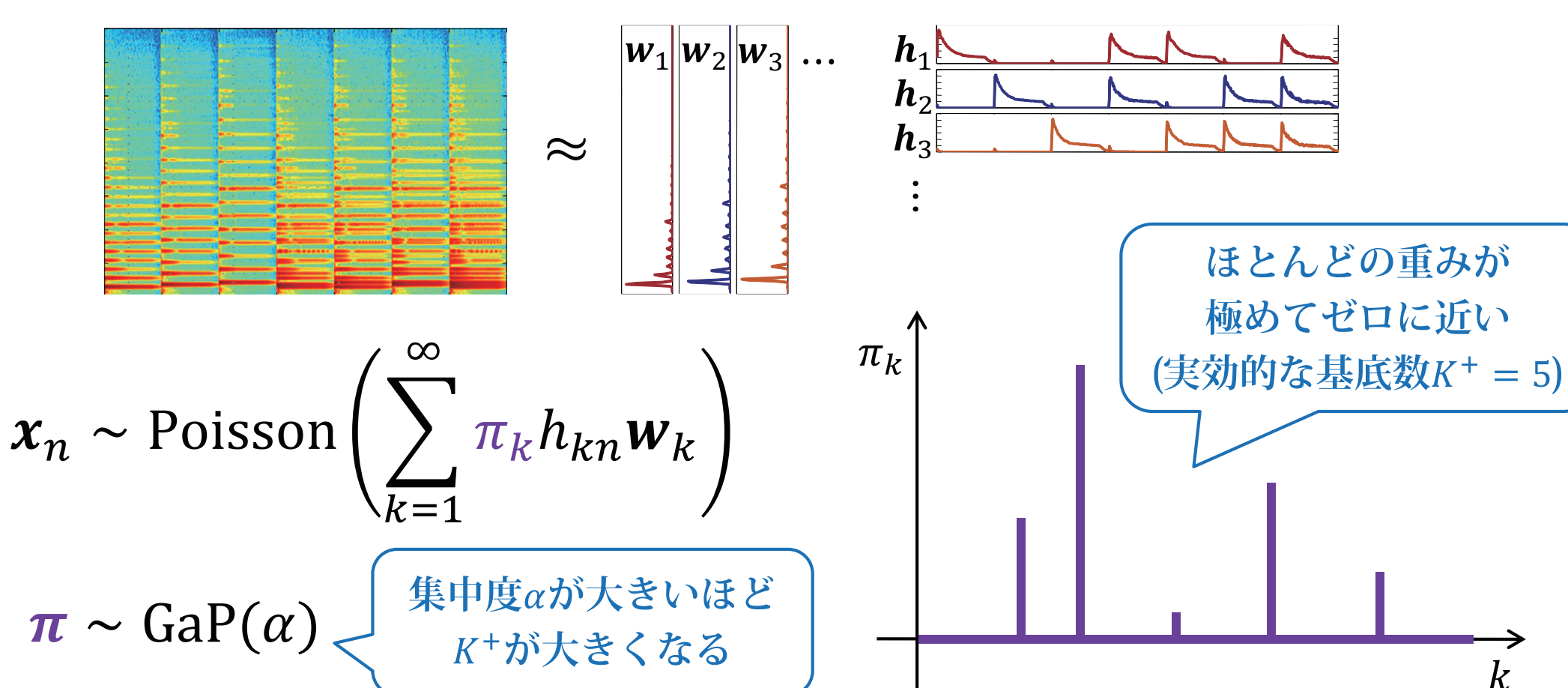
理論的には「因子モデル」が妥当 → NMF

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

ガンマ過程NMF (GaP-NMF) [Hoffman 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

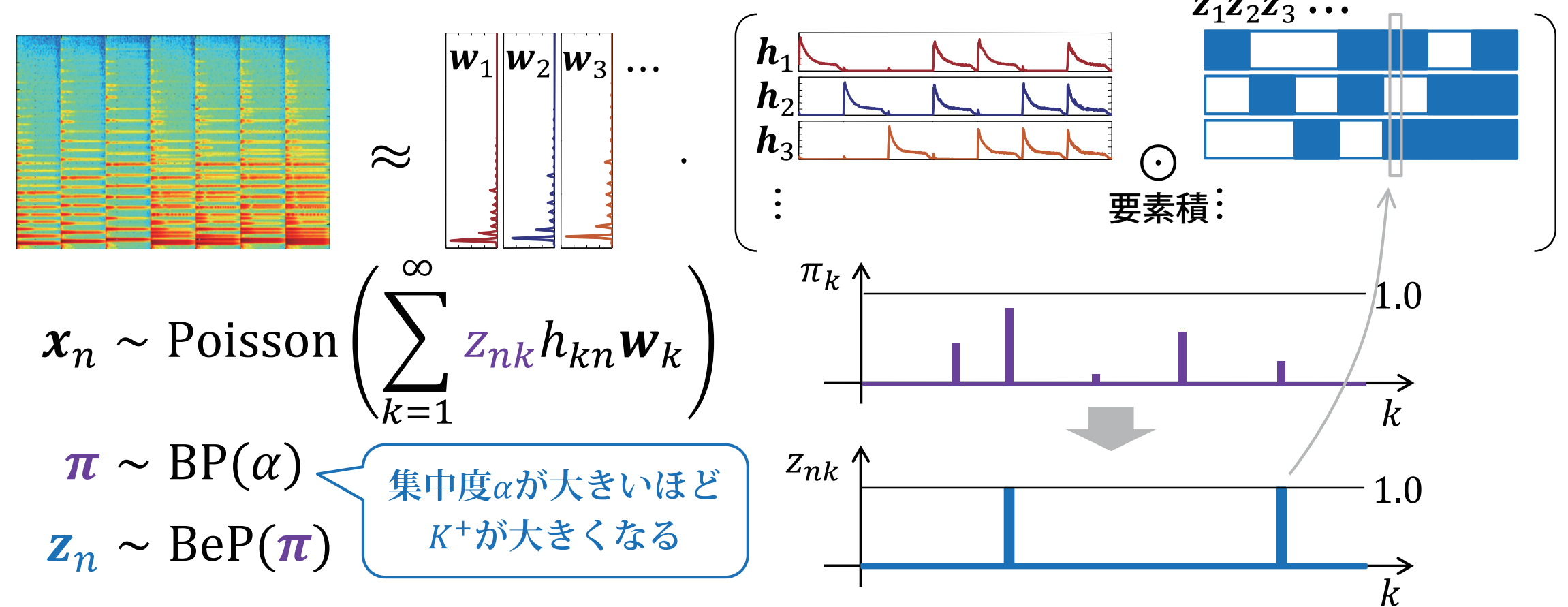


ベータ過程NMF (BP-NMF) [Liang 2014]

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0~1の非負値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

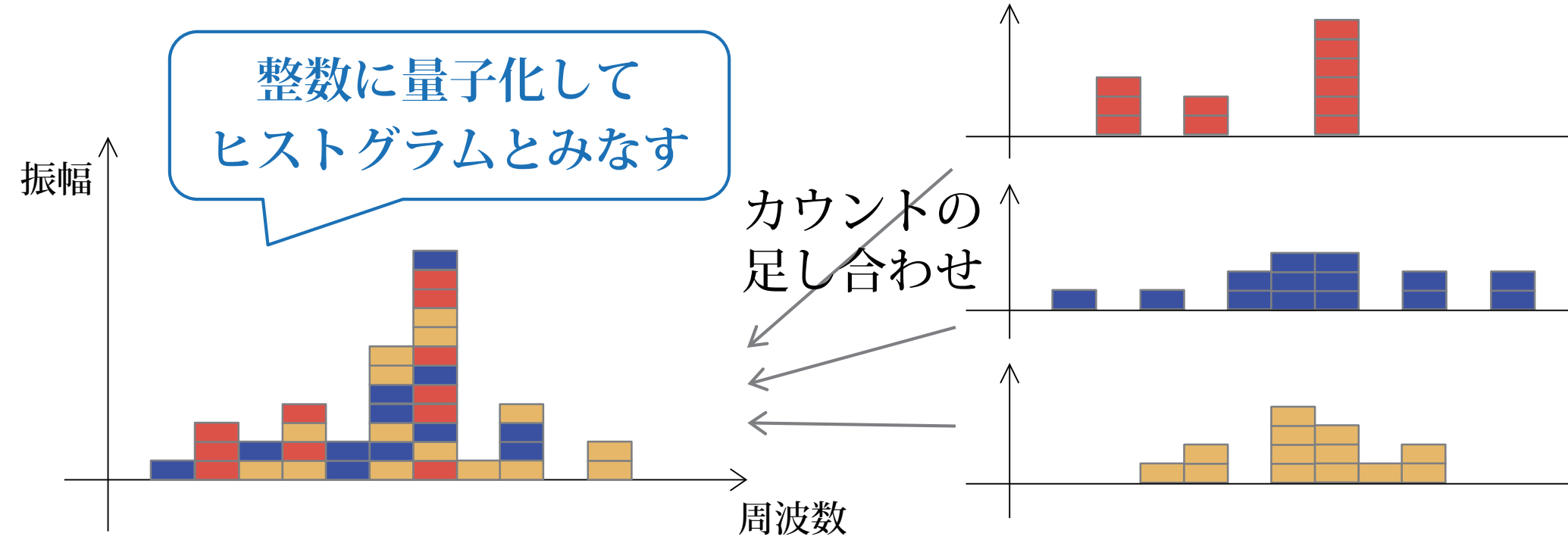


現実には「混合モデル」も提案 → PLCA

トリック！

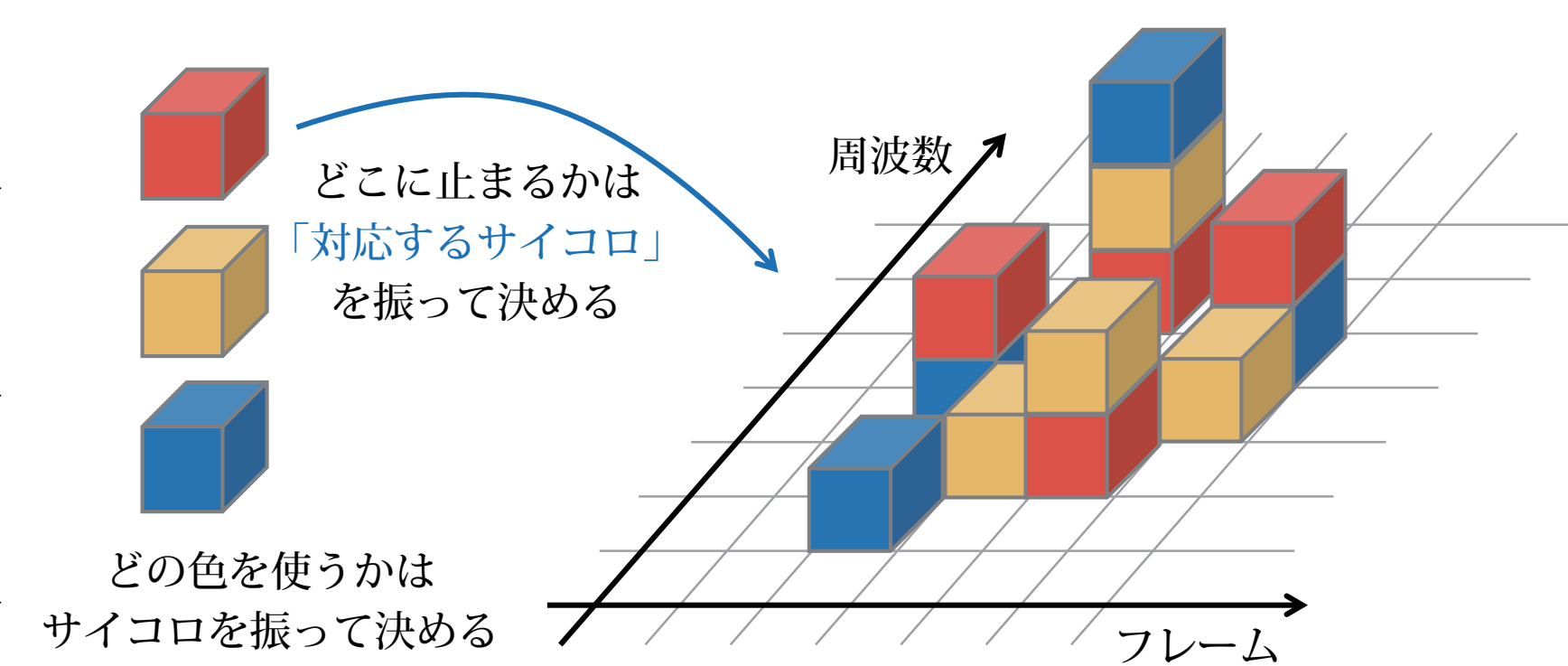
混合音のスペクトル = 「サンプル」の集合
スペクトル全体としてみたときには複数の音源が含まれる

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



Poisson分布に基づくNMF (KL-NMF) も PLCAもエネルギー量子化の粒度を適切に決める必要がある

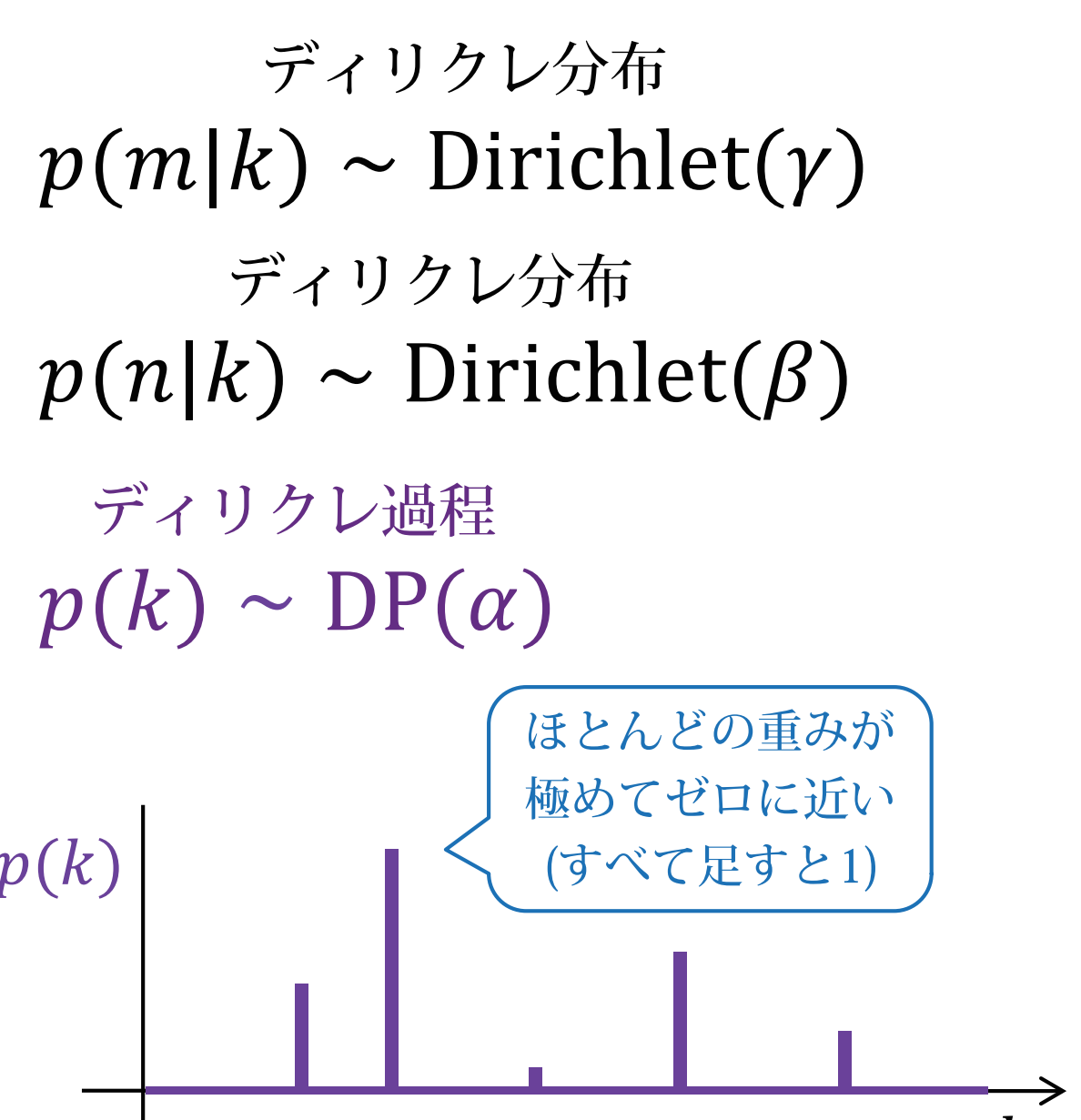
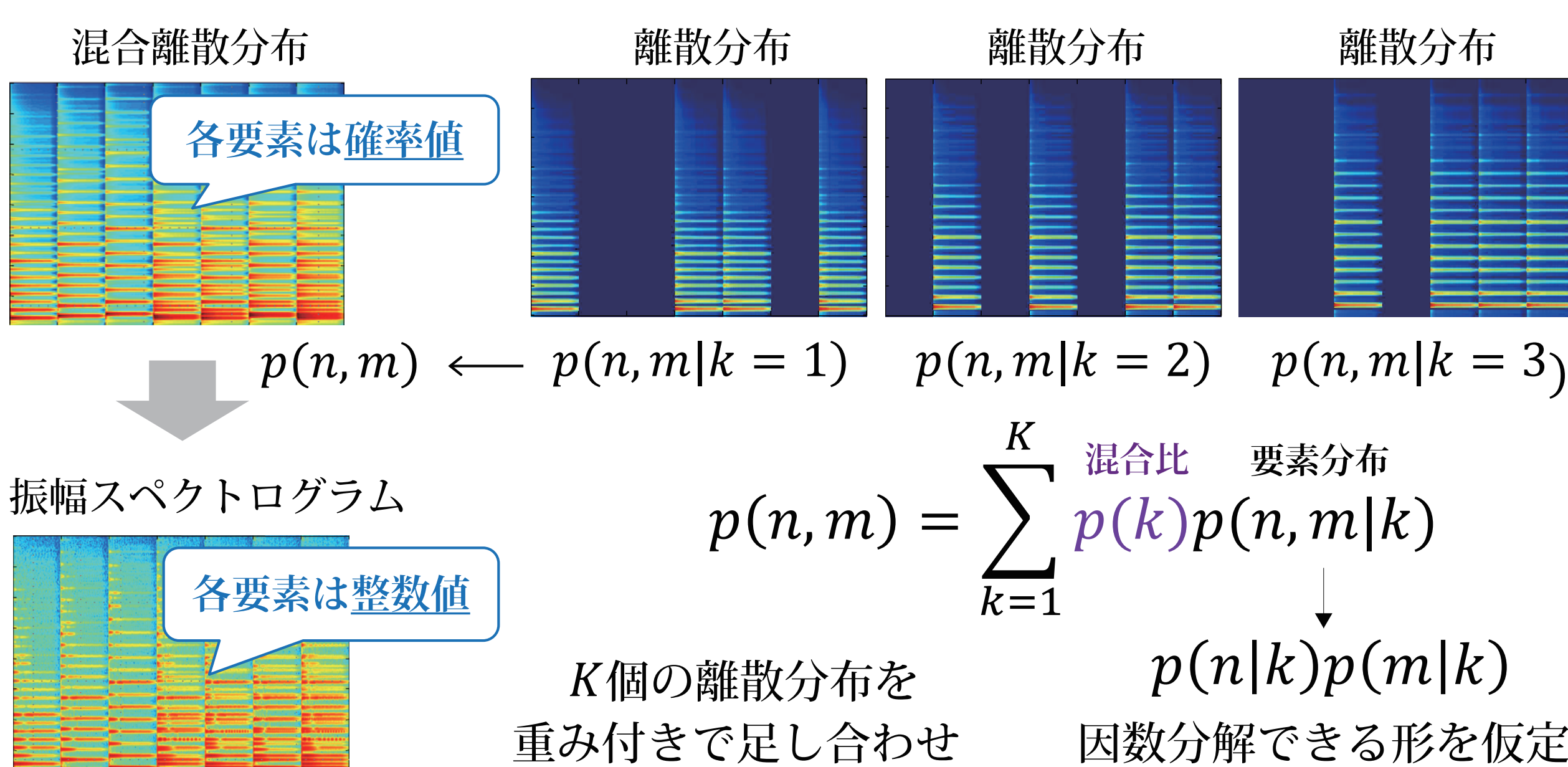
“音量子”を投げ入れると平面内のどこかで止まる
二次元平面はグリッドに分割されている



ディリクレ過程PLCA (DP-PLCA) [提案法]

無限個の基底に対する正規化された重み (足して1) を導入してスパース学習

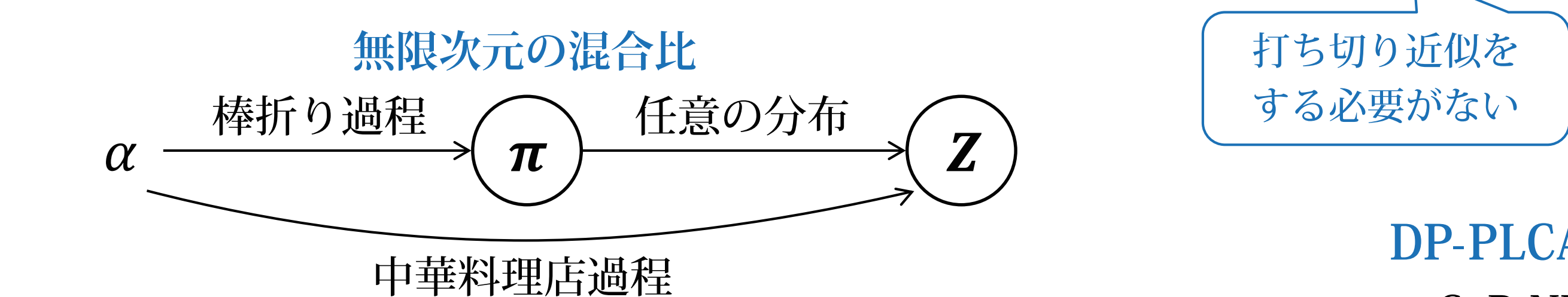
ディリクレ過程 (DP)：無限次元の正規化された非負値ベクトル (離散分布) に対する事前分布



潜在的ディリクレ配分法 (LDA) と類似しているが、階層ディリクレ過程 (HDP) を利用しなくても定式化可能

ディリクレ過程PLCAのベイズ学習：周辺化ギブスサンプリング

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	インド料理店過程 (IBP)	中華料理店過程 (CRP)



DP-PLCAにおいては、実効的な基底数 K^+ を増減させながらパラメータの事後分布計算が可能

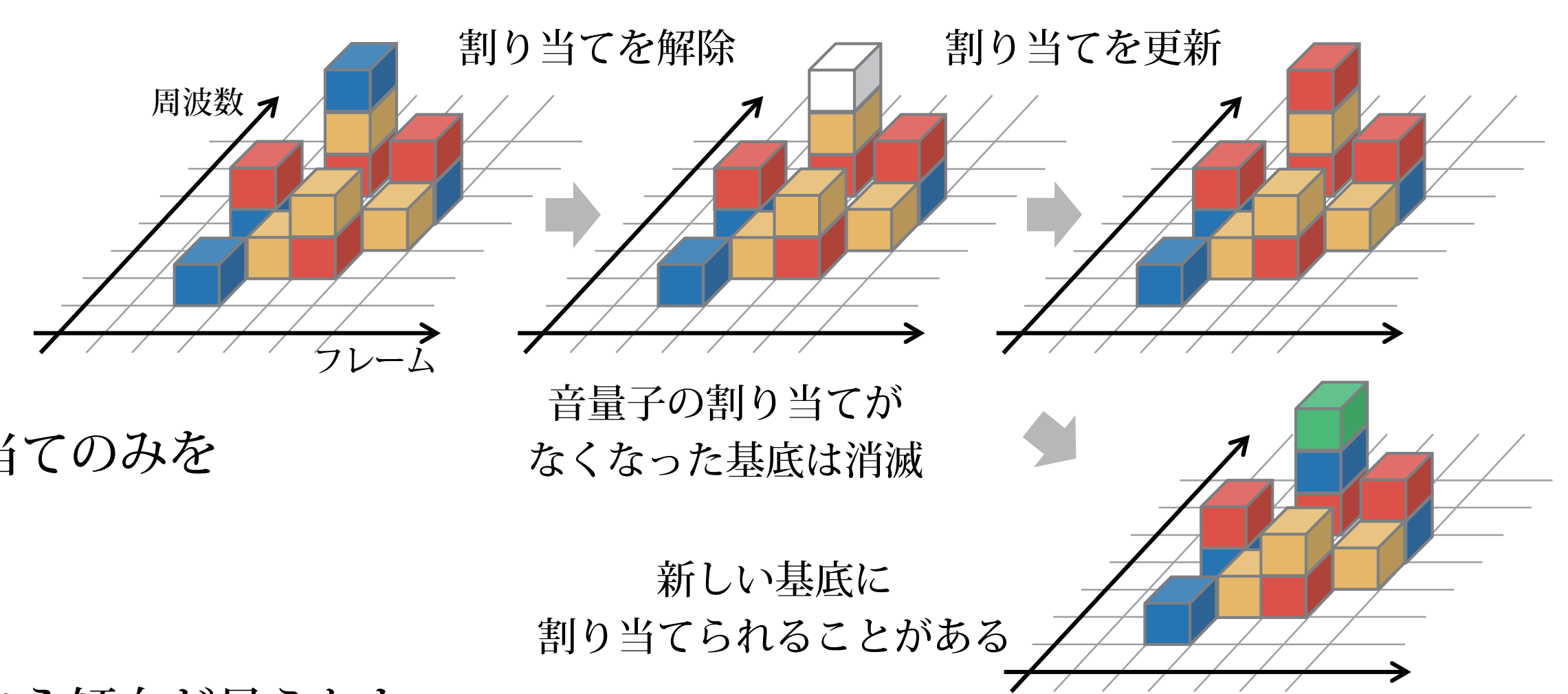
最大基底数で打ち切る近似が不要 (GaP-NMFやBP-NMFでは打ち切りが必要な変分ベイズ法が提案)

全てのパラメータが周辺化可能

中華料理店過程のおかげで、 $p(k)$ を考えなくてOK

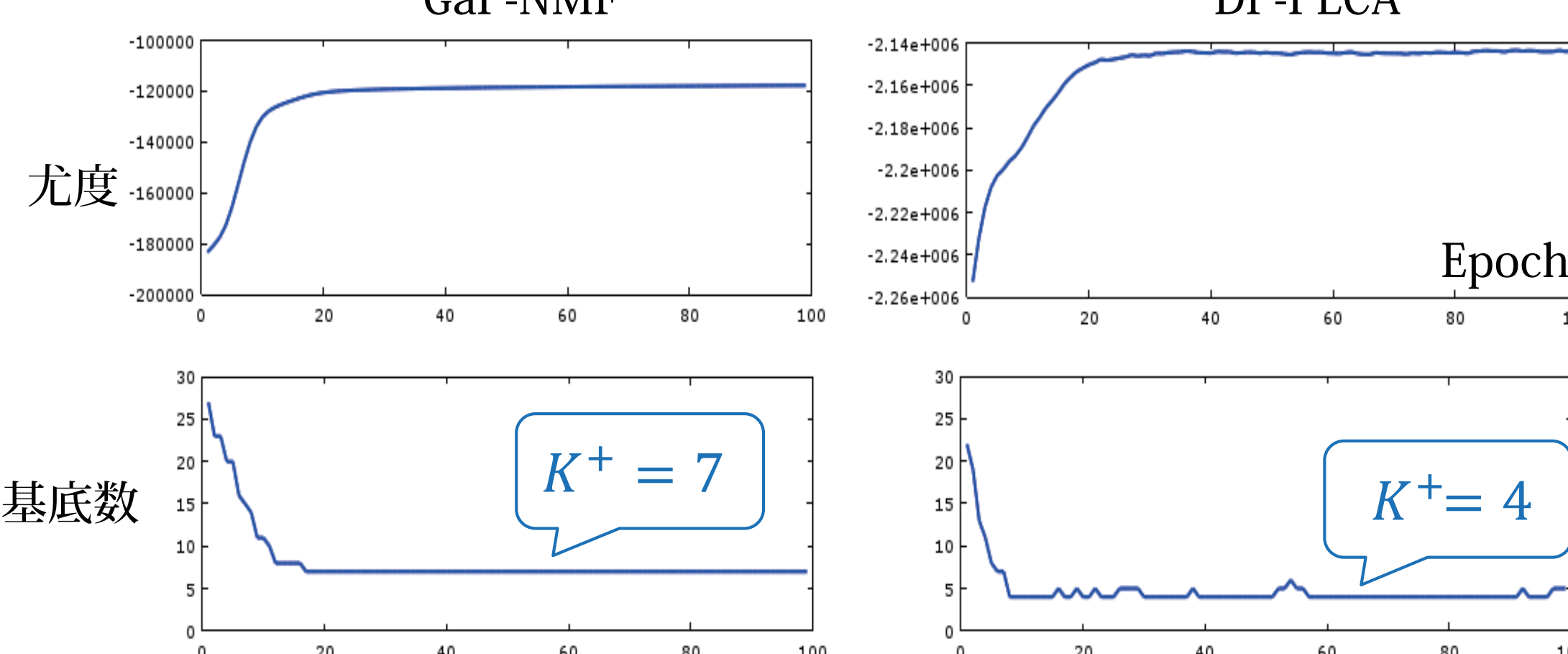
事前分布の共役性のおかげで、 $p(n|k)p(m|k)$ を考えなくてOK

各音量子に関して、音源への割り当てのみを順番にサンプリングしていけばOK



DP-PLCA (Gibbs) は基底数を適切に推定

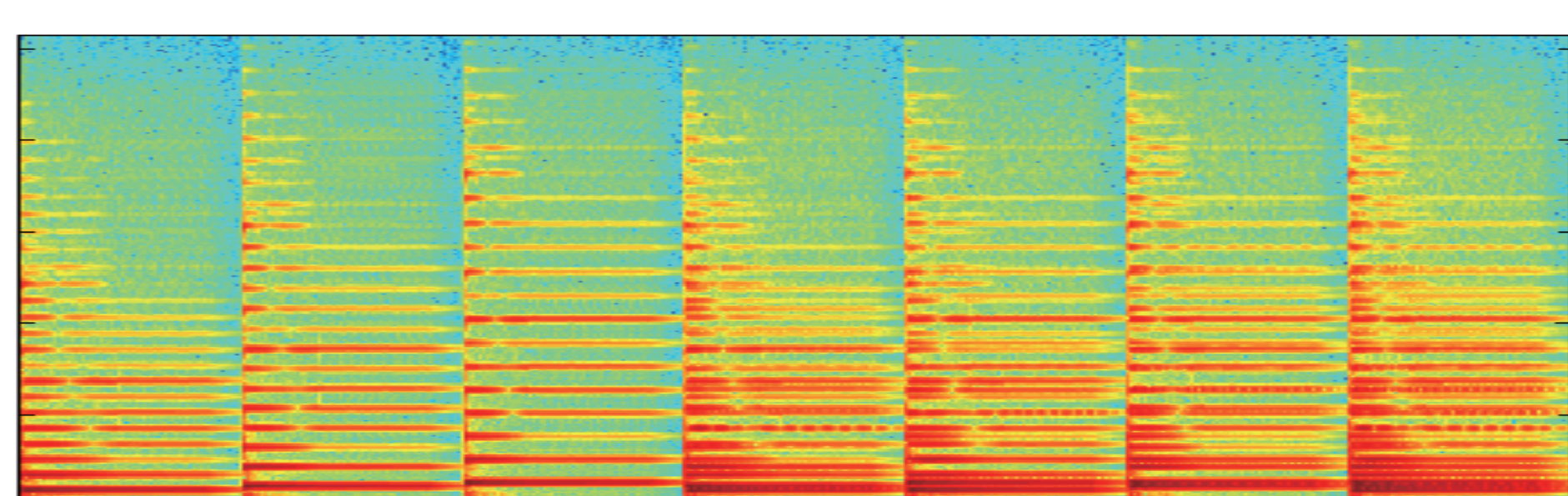
GaP-NMF (VB) は基底数を過剰に見積もってしまう傾向が見られた



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FFT



今後の課題：“音量子”の物理的な裏付け

最尤推定の場合、十分細かくしておけばよい (サンプル数を十分に大きくする)

ベイズ推定の場合、サンプル数が多いほど事後分布における不確実性が低下

ノンパラメトリックベイズモデルだと、推定される基底数が増加してしまうのが問題

フォノンのような概念？