Beijing University of Posts and Telecommunications

Extended Variational Inference for Non-Gaussian Statistical Models

Zhanyu Ma
mazhanyu@bupt.edu.cn

Pattern Recognition and Intelligent System Lab.,
Beijing University of Posts and Telecommunications, Beijing, China.

VALSE Webinar

May 20, 2015

Collaborators

BUPT Beijing University of Posts and Telecommunications

KTH VETENSKAP OCH KONST

UCL

SIBS

berlin

STANFORD SCHOOL OF MEDICINE

References

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

[2] Z. Ma and A. Leijon, "Bayesian Estimation of Beta Mixture Models with Variational Inference", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 33, pp. 2160 – 2173, Nov. 2011.

[3] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition (PR)*, Volume 47, Issue 9, pp. 3143-3157, September 2014.

[4] J. Taghia, Z. Ma, A. Leijon, "Bayesian Estimation of the von-Mises Fisher Mixture Model with Variational Inference*", IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 36, Issue 9, pp. 1701-1715, September, 2014.

[5] P. K. Rana, J. Taghia, Z. Ma**,** and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP),* Volume 9, Issue 3, pp. 435-448, Apr. 2015

Outline

Non-Gaussian Statistical Models

- Non-Gaussian vs. Gaussian
- Advantages and Challenges

Variational Inference (VI) and Extended VI

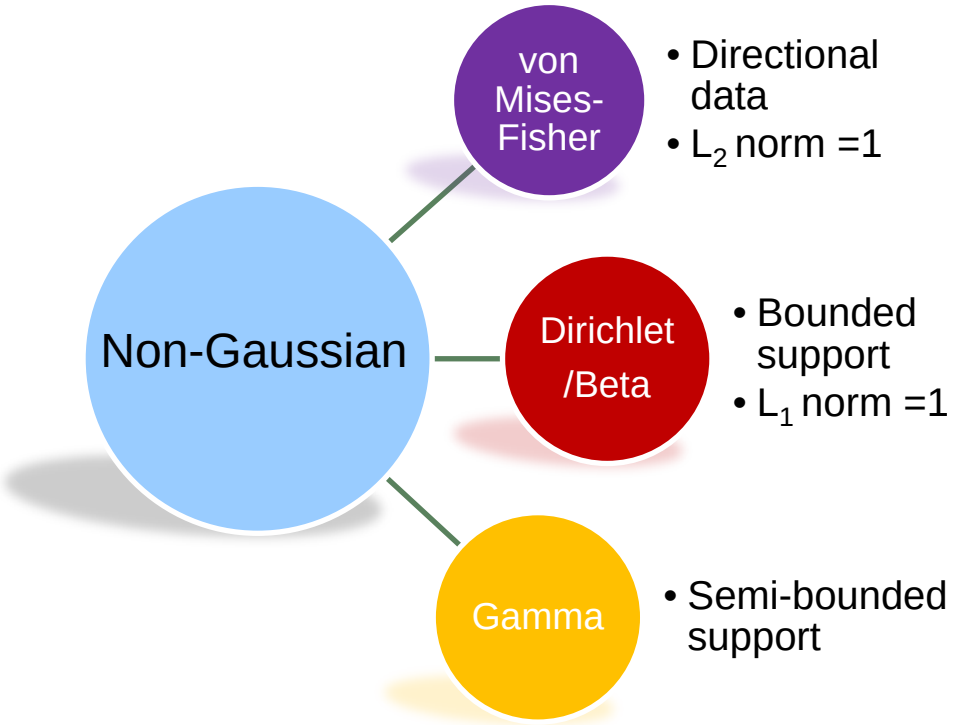
- Formulations and Conditions
- Convergence and Bias

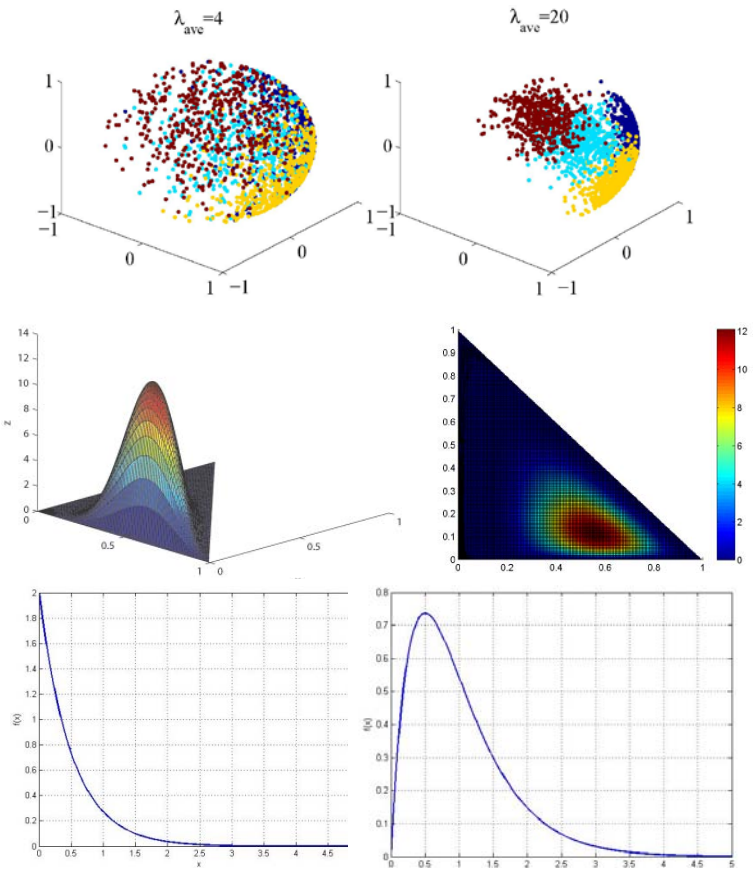
Related Applications

- Beta/Dirichlet Mixture Model
- BG-NMF

Non-Gaussian Statistical Models

- Definition
  - Statistical model for non-Gaussian data
  - Belong to exponential family





Non-Gaussian Statistical Models

Beijing University of Posts and Telecommunications

- Why non-Gaussian? OR Why *not* Gaussian?
  - ✓ Real-life data are not Gaussian
    - Speech Spectra
    - Image pixel value
    - Edge strength in complex network
    - DNA methylation level
    - …….

Non-Gaussian Statistical Models

- Gaussian distribution
  - ✓ Advantages
    - the widely used probability distribution
    - analytically tractable solution
    - Gaussian mixture model can model arbitrary distribution
    - vast applications
  - ✓ Disadvantages
    - not all the data are Gaussian distributed
    - unbounded support and symmetric shape for bounded/semi-bounded/well-structured data
    - flexible model with the cost of high model complexity

Non-Gaussian Statistical Models

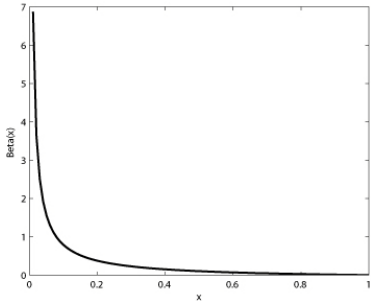
- Non-Gaussian distribution
  - ✓ Advantages
    - well defined for bounded/semi-bounded/well-structured data
    - belong to exponential family → mathematical convenience and conjugate match
    - non-Gaussian mixture model can model data more efficiently
  - ✓ Disadvantages
    - numerically challenging in parameter estimation, both ML and Bayesian estimations!
    - lack of closed-form solution for real applications

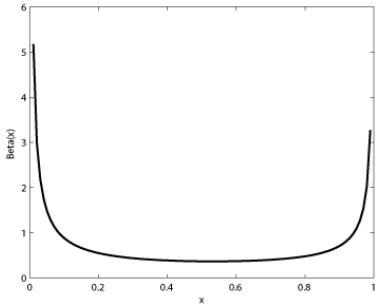
Beijing University of Posts and Telecommunications

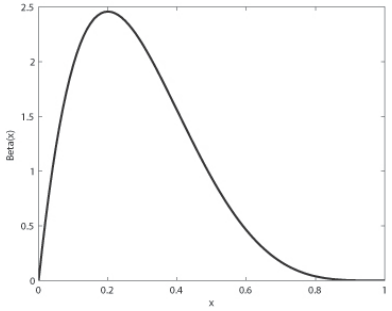
Non-Gaussian Statistical Models

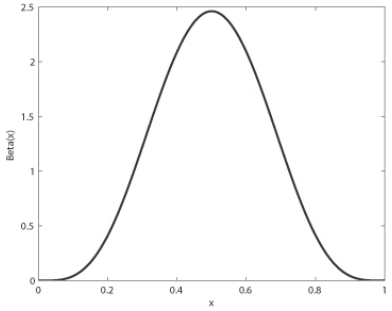
- Example 1: beta distribution

$$\text{beta}(x;u,v)=\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}x^{u-1}(1-x)^{v-1},\quad \Gamma(z)=\int_0^\infty t^{z-1}e^{-t}dt$$







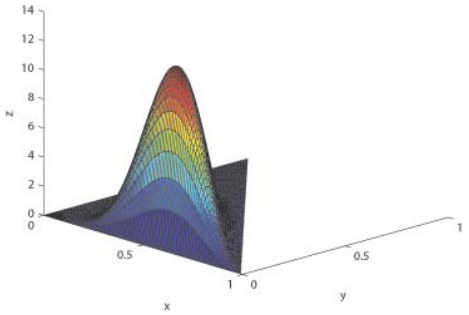


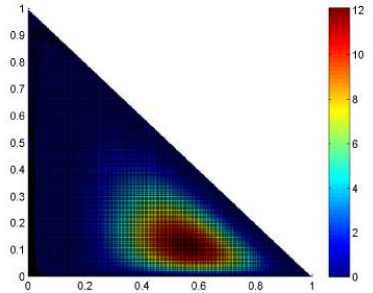
  - Bounded support and flexible shape
  - Image processing, speech coding, DNA methylation analysis

Non-Gaussian Statistical Models

- Example 2: Dirichlet distribution (neutral vector)

$$\mathrm{Dir}(\mathbf{x};\mathbf{a}) = \frac{\Gamma\left(\sum_{k=1}^{K} a_k\right)}{\prod_{k=1}^{K}\Gamma(a_k)} \prod_{k=1}^{K} x_k^{a_k-1}, \quad \sum_{k=1}^{K} x_k = 1, \quad x_k > 0, \quad a_k > 0.$$





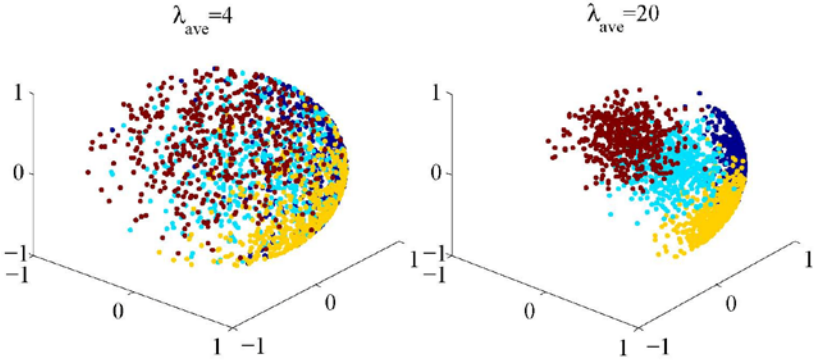
- Conventionally used as conjugate prior of multi categorical distribution or multinomial distribution, describing mixture weights in mixture modeling
- Recently, it was applied to model proportional data (i.e., data with L1 norm)
- Speech coding, skin color detection, multiview 3D enhancement, etc.

Non-Gaussian Statistical Models

- Example 3: von Mises-Fisher distribution

$$f(\mathbf{x};\lambda,\boldsymbol{\mu}) = \frac{\lambda^{\frac{K}{2}-1}}{(2\pi)^{\frac{K}{2}} I_{\frac{K}{2}-1}(\lambda)} e^{\lambda \cdot \boldsymbol{\mu}^{\mathsf{T}}\mathbf{x}}, \quad \mathbf{x}^{\mathsf{T}}\mathbf{x} = 1$$

$I_p(v)$ denotes the modified Bessel function of the first kind



  - Distributed on K-dimensional sphere
  - Two-dimensional vMF = circle
  - Directional statistics, gene expressions, speech coding

Non-Gaussian Statistical Models

- Summary
  - Non-Gaussian distribution represents a family of distributions which are not Gaussian distributed
  - Not conflicting with central limit theorem
  - Well-defined for bounded/semi-bounded/structured data
  - More efficient than Gaussian distribution
  - Hard to estimate, computationally costly, and difficult to use in practice

Outline

Non-Gaussian Statistical Models

- Non-Gaussian vs. Gaussian
- Advantages and Challenges

Variational Inference (VI) and Extended VI

- Formulations and Conditions
- Convergence and Bias

Related Applications

- Beta/Dirichlet Mixture Model
- BG-NMF

Formulation and Conditions

- Maximum likelihood (ML) estimation
  - Widely used for point estimation of the parameters
  - Expectation-maximization (EM) algorithm
  - Converge to local maxima and may yield overfitting
  - No analytically tractable solution for most non-Gaussian distributions

Formulation and Conditions

- Bayesian estimation
  - Estimating the distributions of the parameters, rather than point estimate
  - Conjugate match in exponential family
  - No overfitting, feasible for online learning
  - Without approximation, there is no analytically tractable solution for non-Gaussian distributions

Formulation and Conditions

- Example: ML estimation for beta mixture model[1]
  - M step

$$\begin{bmatrix} \psi(u+v)-\psi(u)+\frac{1}{N}\sum_{n=1}^{N}\ln x_n \\ \psi(u+v)-\psi(v)+\frac{1}{N}\sum_{n=1}^{N}\ln(1-x_n) \end{bmatrix}=0$$

$$\psi(z)=\frac{d\ln\Gamma(z)}{dz}=\int_0^\infty\left(\frac{e^{-t}}{t}-\frac{e^{-zt}}{1-e^{-t}}\right)dt$$

  - Numerical solution, Gibbs sampling, Newton-Raphson method, MCMC, etc.

[1] Z. Ma and A. Leijon, 'Beta Mixture Model and the Application to Image Classification', *IEEE International Conference on Image Processing*, pp. 2045-2048, 2009.

Formulation and Conditions

- Example: Bayesian estimation of beta distribution[1]
  - Prior

$$p(u,v;\alpha_0,\beta_0,\nu_0) \propto \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{\nu_0} e^{-\alpha_0(u-1)} e^{-\beta_0(v-1)}$$

  - Likelihood

$$\text{beta}(x;u,v) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} x^{u-1}(1-x)^{v-1}$$

  - Posterior

$$p(u,v \mid \mathbf{X};\alpha_0,\beta_0,\nu_0) \propto \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{\nu_0+N} e^{-\left(\alpha_0-\sum_{n=1}^{N}\ln x_n\right)(u-1)} e^{-\left(\beta_0-\sum_{n=1}^{N}\ln(1-x_n)\right)(v-1)}$$

  - No closed-form expression for mean, variance, etc.
  - No analytically tractable solution for mixture model
  - Not applicable in practice

[1] Z. Ma and A. Leijon, 'Bayesian Estimation of Beta Mixture Models with Variational Inference', *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 33, pp. 2160 – 2173, Nov. 2011.

Formulation and Conditions

- Variational inference[1]
  - Mean field theory in physics, dates back to 18th century, by Euler, Lagrange, etc.
  - Function over function
  - Closed form solution with certain constraints

$$f(x) = \int f(x \mid \theta) f(\theta) d\theta$$

$$\ln f(x) = \int g(\theta) \ln \frac{f(x,\theta)}{g(\theta)} d\theta + \int g(\theta) \ln \frac{g(\theta)}{f(\theta \mid x)} d\theta$$
$$= L(g) + \mathrm{KL}(g \parallel f)$$

  - Goal: approximate $f(\theta \mid x)$ by $g(\theta)$ via either maximizing $L(g)$ or minimizing $\mathrm{KL}(g \parallel f)$

[1] C. M. Bishop, 'Pattern Recognition and Machine Learning', *Springer*, 2006

Formulation and Conditions

- Factorized approximation[1]

$$g(\boldsymbol{\theta}) \approx \prod_i g_i(\boldsymbol{\theta}_i)$$

$$\ln g_i^*(\boldsymbol{\theta}_i) = \mathrm{E}_{j \neq i}\left[\ln f(x, \boldsymbol{\theta})\right] + C$$

  - No constraints on the form of $g_i(\boldsymbol{\theta}_i)$
  - Directly maximizing $L(g)$
  - Always converges but may fall in local maxima
  - Analytically tractable form solution for Gaussian

[1] C. M. Bishop, 'Pattern Recognition and Machine Learning', *Springer*, 2006

Formulation and Conditions

- Extended factorized approximation (EFA)[1,2]

$$\mathcal{L}(g) = \mathrm{E}\left[f(x,\boldsymbol{\theta})\right] - \mathrm{E}\left[g(\boldsymbol{\theta})\right]$$
$$\geq \mathrm{E}\left[\tilde{f}(x,\boldsymbol{\theta})\right] - \mathrm{E}\left[g(\boldsymbol{\theta})\right]$$

Auxiliary function

  - Optimal solution: $\ln g_i^*(\boldsymbol{\theta}_i) = \mathrm{E}_{j\neq i}\left[\ln \tilde{f}(x,\boldsymbol{\theta})\right] + C$
    - Strong requirement with larger gap: $f(x,\boldsymbol{\theta}) \geq \tilde{f}(x,\boldsymbol{\theta})$
    - Weak requirement with smaller gap: $\mathrm{E}[f(x,\boldsymbol{\theta})] \geq \mathrm{E}\left[\tilde{f}(x,\boldsymbol{\theta})\right]$
  - An efficient way to derive analytically tractable solution for non-Gaussian distribution
  - SLB vs MLB [2]

[1] Z. Ma and A. Leijon, 'Bayesian Estimation of Beta Mixture Models with Variational Inference', *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 33, pp. 2160 – 2173, Nov. 2011.
[2] Z. Ma, J. Taghia, P. K. Rana, M. Flierl, and A. Leijon, 'Bayesian Estimation of Beta Mixture Models with Variational Inference', *Pattern Recognition*, Vol 47, No. 9, pp. 3143-3157, Sep. 2014.

Convergence and Bias

- Multiple lower-bound (MLB) approximation[1][2]
  - Different auxiliary functions for different variable (group)

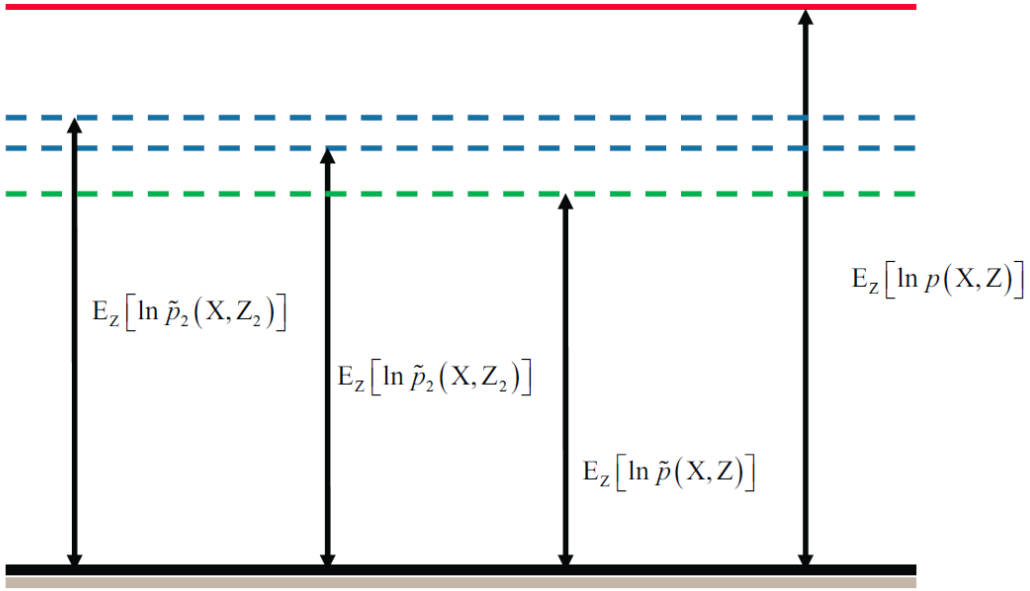
$$p(\mathbf{X},\mathbf{Z}) \geq \widetilde{p}_1(\mathbf{X},\mathbf{Z}_1)$$
$$\vdots$$
$$p(\mathbf{X},\mathbf{Z}) \geq \widetilde{p}_i(\mathbf{X},\mathbf{Z}_i)$$
$$\vdots$$
$$p(\mathbf{X},\mathbf{Z}) \geq \widetilde{p}_M(\mathbf{X},\mathbf{Z}_M)$$

  - Optimal solution for wach variable (group)

$$\ln \widetilde{q}_i^*(\mathbf{Z}_i) = \mathbf{E}_{\mathbf{Z}\setminus\mathbf{Z}_i}\left[\ln \widetilde{p}_i(\mathbf{X},\mathbf{Z}_i)\right] + \text{const.}$$

[1] Z. Ma and A. Leijon, 'Bayesian Estimation of Beta Mixture Models with Variational Inference', *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 33, pp. 2160 – 2173, Nov. 2011.
[2] W. Fan, N. Bouguila, and D. Ziou, "Variational learning for finite Dirichlet mixture models and applications," *IEEE Transactions on Neural Network and Learning Systems*, vol. 23, no. 5, pp. 762–774, May 2012

Convergence and Bias



Update Z_1 and Z_2 iteratively:

$$\ln \widetilde{q}_1^*(\mathbf{Z}_1) = \mathbf{E}_{\mathbf{Z}\setminus\mathbf{Z}_1}[\ln \widetilde{p}_1(\mathbf{X}, \mathbf{Z})] + \text{const}$$
$$\ln \widetilde{q}_2^*(\mathbf{Z}_2) = \mathbf{E}_{\mathbf{Z}\setminus\mathbf{Z}_2}[\ln \widetilde{p}_2(\mathbf{X}, \mathbf{Z})] + \text{const}$$

Convergence not guaranteed!

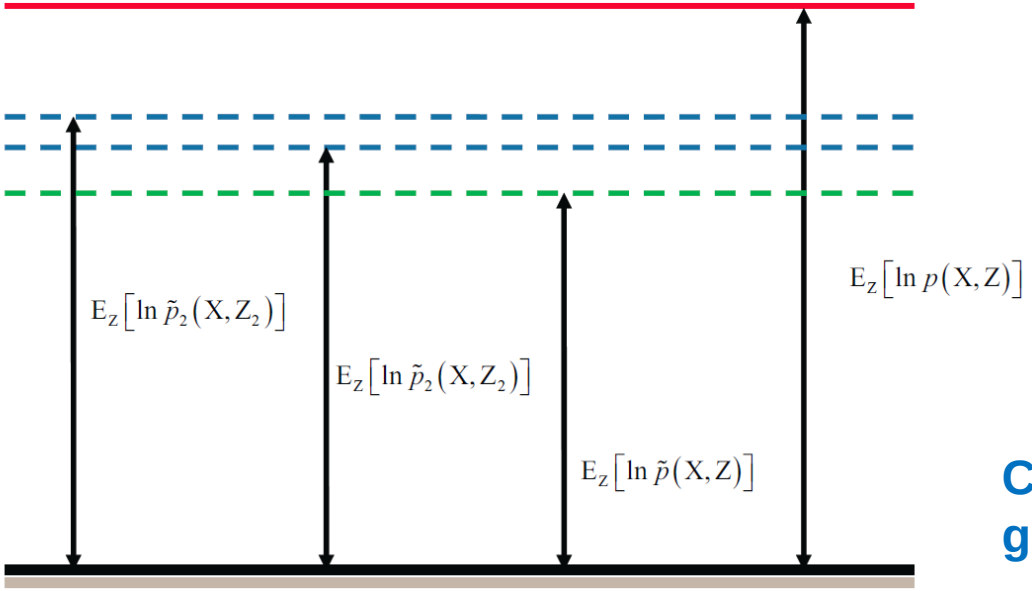
Update Z_1 and Z_2 iteratively:

$$\begin{aligned} 2 \times \mathcal{L} \geq & \widetilde{\mathcal{L}}_1 + \widetilde{\mathcal{L}}_2 \\ = & \mathbf{E}_{\mathbf{Z}}[\ln \widetilde{p}_1(\mathbf{X}, \mathbf{Z}_1)] - \mathbf{E}_{\mathbf{Z}}[\ln q(\mathbf{Z})] \\ + & \mathbf{E}_{\mathbf{Z}}[\ln \widetilde{p}_2(\mathbf{X}, \mathbf{Z}_2)] - \mathbf{E}_{\mathbf{Z}}[\ln q(\mathbf{Z})] \end{aligned}$$

[1] Z. Ma, J. Taghia, and J. Guo, "On the Convergence of Extended Variational Inference for Non-Gaussian Statistical Models", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, under review.

Convergence and Bias

- Single lower-bound (SLB) approximation[1][2]
  - One auxiliary functions for all the different variable (group)



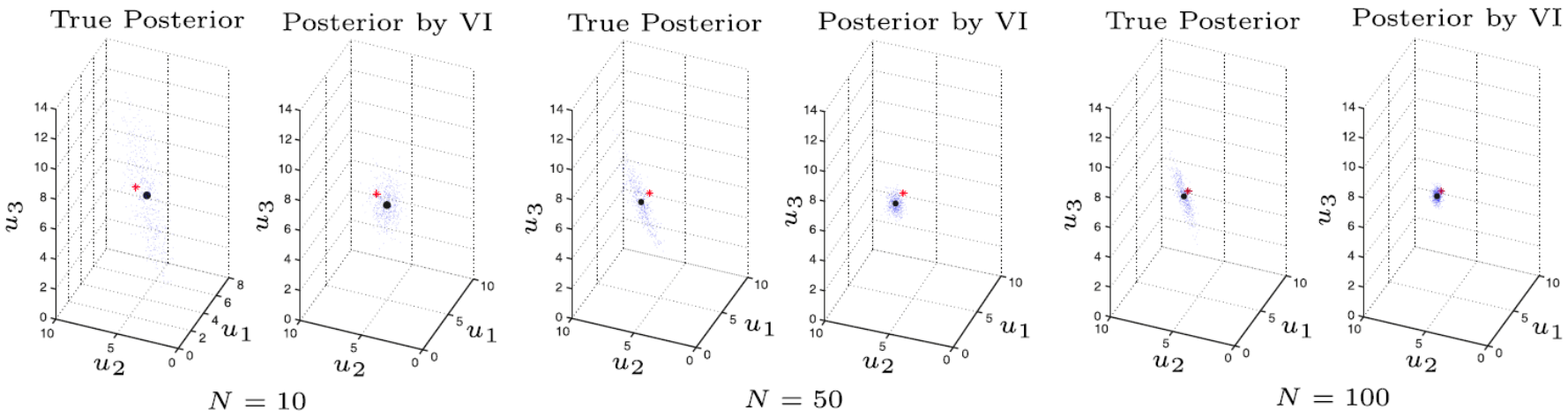
Convergence guaranteed!

  - Optimal solution

$$\ln \widetilde{q}_i^*(\mathbf{Z}_i) = \mathbf{E}_{\mathbf{Z}\setminus\mathbf{Z}_i}\left[\ln \widetilde{p}(\mathbf{X}, \mathbf{Z})\right] + \text{const}$$

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 4, pp. 876 – 889, Apr. 2015
[2] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition*, Vol. 47, No. 9, pp. 3143-3157, Sep. 2014.

Convergence and Bias



True posterior distribution vs. approximating distribution[1].Dirichlet distribution with u=[3 5 8].

- Bias always exists, due to factorized approximations and lower-bound approximation.
- Bias will vanish when increasing the amount of training data.

[1] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition*, Vol. 47, No. 9, pp. 3143-3157, Sep. 2014.

Beijing University of Posts and Telecommunications

Convergence and Bias

- Summary
  - EVI provides a flexible way to carry out Bayesian estimation of NG statistical model
  - Certain requirements should be fulfilled when implementing EVI
  - MLB vs. SLB
  - Systematic gap

Outline

Non-Gaussian Statistical Models

- Non-Gaussian vs. Gaussian
- Advantages and Challenges

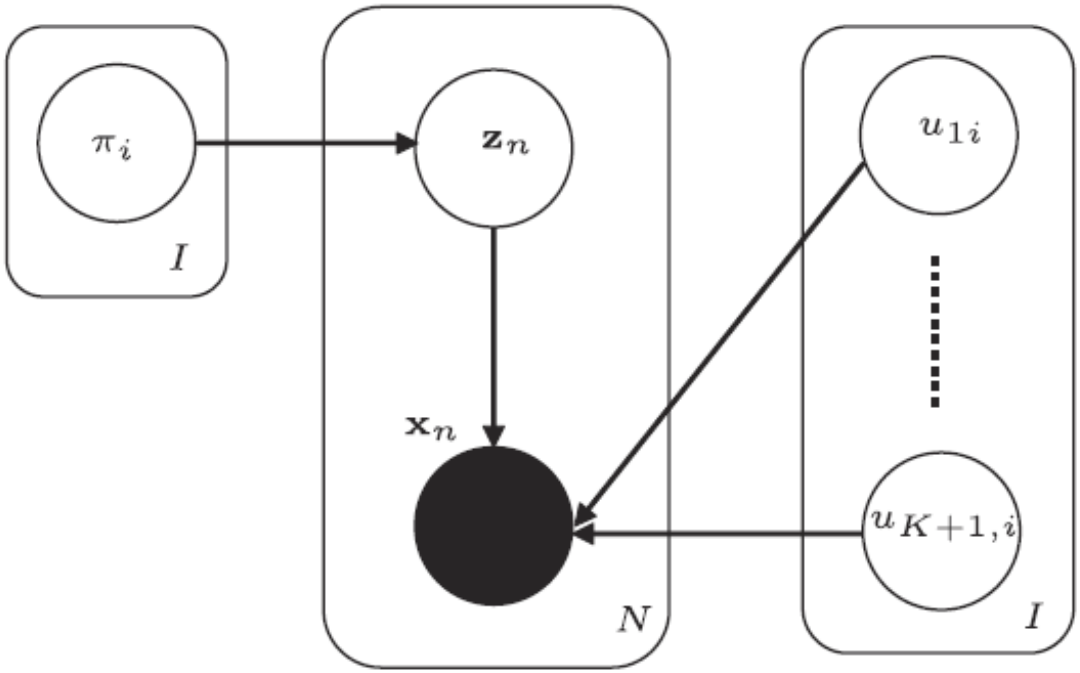
Variational Inference (VI) and Extended VI

- Formulations and Conditions
- Convergence and Bias

Related Applications

- Beta/Dirichlet Mixture Model
- BG-NMF

Dirichlet Mixture Model



Graphical Model of DMM[1]

[1] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition*, Vol. 47, No. 9, pp. 3143-3157, Sep. 2014.

Dirichlet Mixture Model

– Auxiliary function

- Step I

$$\mathbb{E}_{\mathcal{Z}}[\ln f(\mathbf{X},\mathcal{Z})] \geq \sum_{n=1}^{N}\sum_{i=1}^{I}\mathbb{E}[z_{ni}]\left(\ln\frac{\Gamma(\sum_{k=1}^{K+1}\overline{u}_{ki})}{\prod_{k=1}^{K+1}\Gamma(\overline{u}_{ki})}\right.$$
$$+\sum_{k=1}^{K+1}\mathbb{E}_{u_{p\geq k,i}}\left\{\left[\psi\left(\sum_{m=1}^{k}\overline{u}_{mi}+\sum_{l=k+1}^{K+1}u_{li}\right)-\psi(\overline{u}_{ki})\right]\right.$$
$$\left.\left.\times\overline{u}_{ki}(\ln u_{ki}-\ln\overline{u}_{ki})\right\}\right)+\mathcal{R}^{\dagger}$$

- Step II

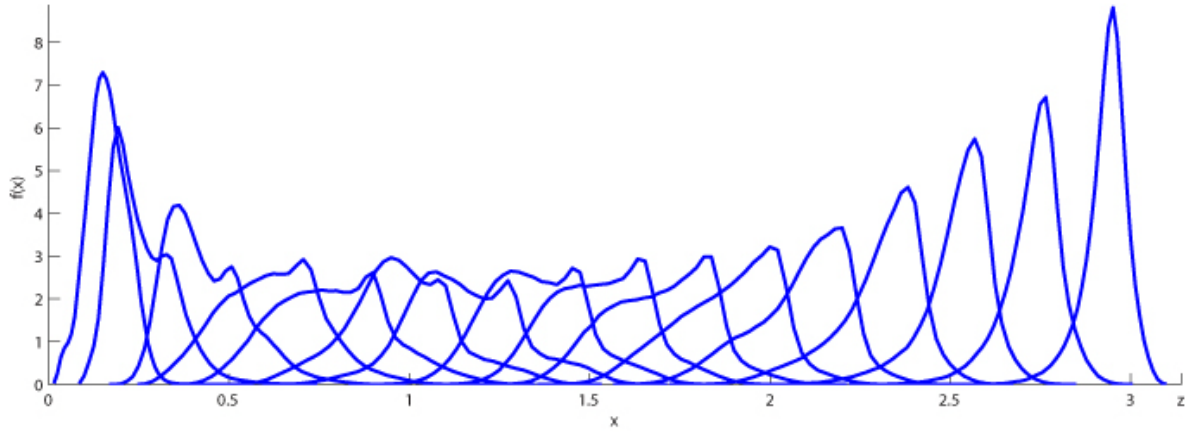
$$\mathbb{E}_{u_{p\geq k,i}}\left\{\left[\psi\left(\sum_{m=1}^{k}\overline{u}_{mi}+\sum_{l=k+1}^{K+1}u_{li}\right)-\psi(\overline{u}_{ki})\right]\overline{u}_{ki}(\ln u_{ki}-\ln\overline{u}_{ki})\right\}$$
$$=\left\{\mathbb{E}_{u_{p>k,i}}\left[\psi\left(\sum_{m=1}^{k}\overline{u}_{mi}+\sum_{l=k+1}^{K+1}u_{li}\right)\right]-\psi(\overline{u}_{ki})\right\}\overline{u}_{ki}\left(\underbrace{\mathbb{E}_{u_{ki}}[\ln u_{ki}]-\ln\overline{u}_{ki}}_{\leq 0}\right)$$
$$\geq\left[\psi\left(\sum_{m=1}^{k}\overline{u}_{mi}+\mathbb{E}_{u_{p>k,i}}\left[\sum_{l=k+1}^{K+1}u_{li}\right]\right)-\psi(\overline{u}_{ki})\right]\overline{u}_{ki}(\mathbb{E}_{u_{ki}}[\ln u_{ki}]-\ln\overline{u}_{ki})$$
$$=\left[\psi\left(\sum_{k=1}^{K+1}\overline{u}_{ki}\right)-\psi(\overline{u}_{ki})\right]\overline{u}_{ki}(\mathbb{E}_{u_{ki}}[\ln u_{ki}]-\ln\overline{u}_{ki})$$

[1] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition*, Vol. 47, No. 9, pp. 3143-3157, Sep. 2014.

Dirichlet Mixture Model

Beijing University of Posts and Telecommunications

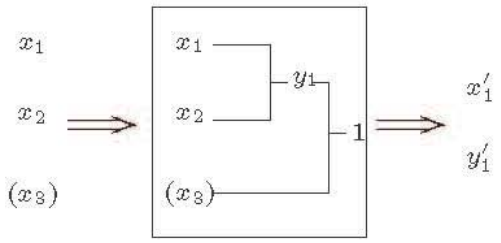
- Speech coding[1]
  - Quantization of line spectral frequency (LSF)
  - Well-structured vector
    - all the elements are in (0,π)
    - strictly ordered

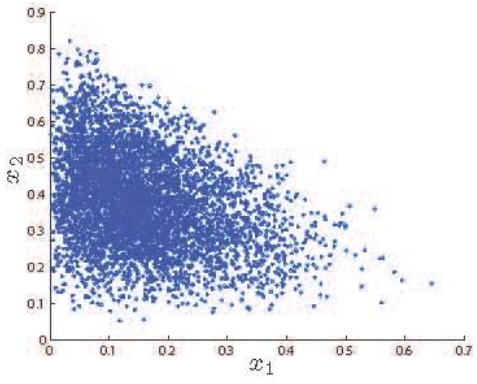


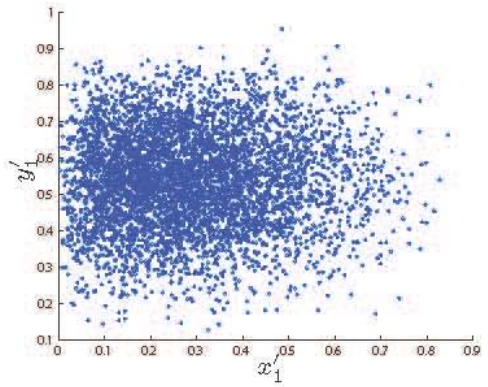
[1] Z. Ma, A. Leijon, and W. B. Kleijn, 'Vector Quantization of LSF Parameters with a Mixture of Dirichlet Distributions', *IEEE Transaction on Audio, Speech, and Language Processing*, 2013.

Dirichlet Mixture Model

- Speech coding
  - Solution: Dirichlet mixture model[1,2]
    - transfer LSF vector to ΔLSF vector
    - well-structured: nonnegative elements, L1 norm equals one
    - a neutral vector that can be nonlinearly decorrelated (comparable to KLT/PCA for Gaussian source!)







[1] Z. Ma and A. Leijon, 'Modeling Speech Line Spectral Frequencies with Dirichlet Mixture Models', *INTERSPEECH,* 2010.
[2] Z. Ma, A. Leijon, and W. B. Kleijn, 'Vector Quantization of LSF Parameters with a Mixture of Dirichlet Distributions', *IEEE Transaction on Audio, Speech, and Language Processing,* 2013.

Dirichlet Mixture Model

- Speech coding
  - Solution: Dirichlet mixture model[1,2]
    - transfer LSF vector to ΔLSF vector
    - well-structured: nonnegative elements, L1 norm equals one
    - a neutral vector that can be nonlinearly decorrelated (comparable to KLT/PCA for Gaussian source!)

TABLE V

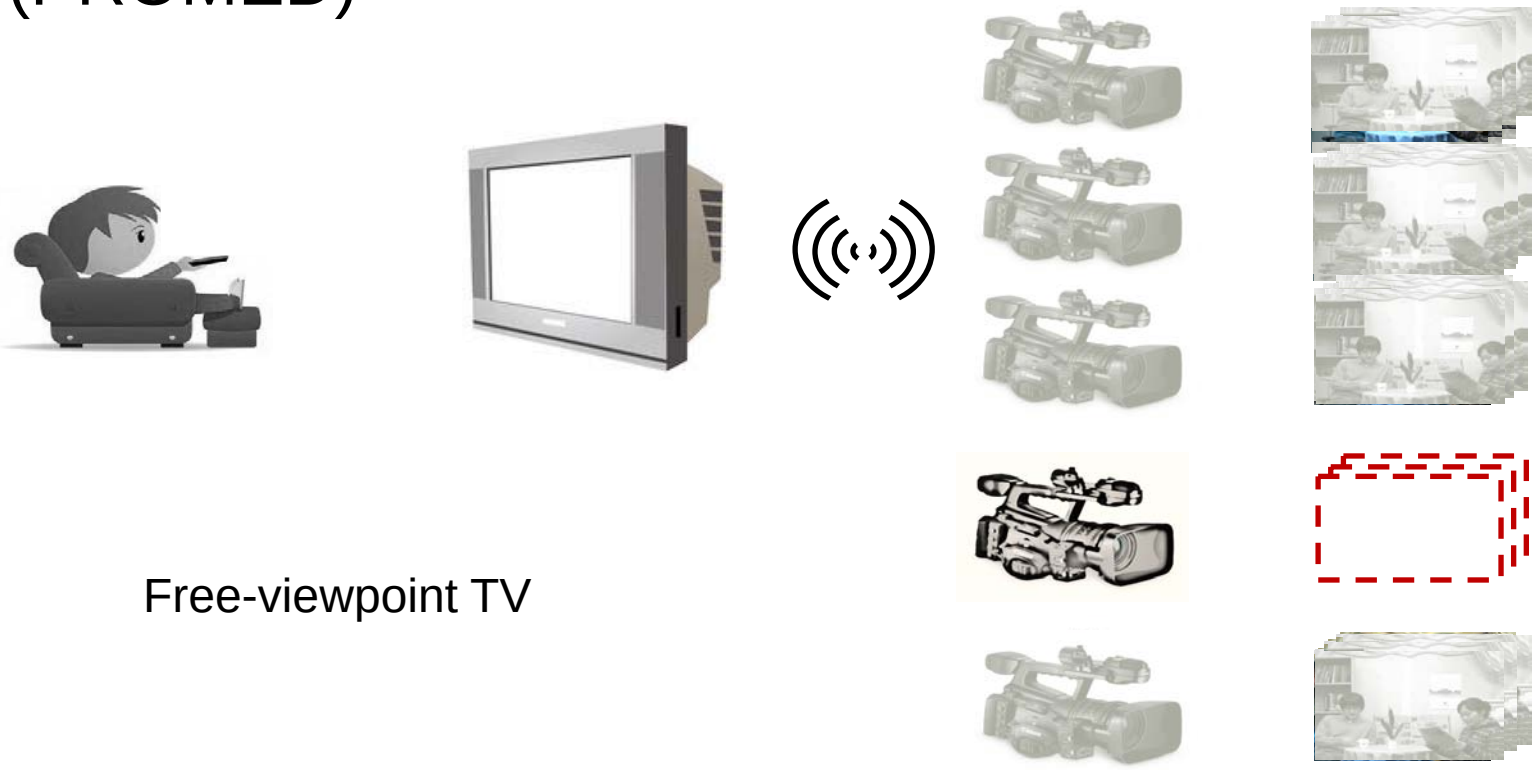
CODING PERFORMANCE OF DIFFERENT VQS, WITH DIFFERENT NUMBER OF MIXTURE COMPONENTS.

Model complexity	Coding performance of DVQ				Coding performance of BVQ				Coding performance of UGVQ			
	bits/vec.	LSD (dB)	LSD outliers (in %) 2 – 4 dB	>4 dB	bits/vec.	LSD (dB)	LSD outliers (in %) 2 – 4 dB	>4 dB	bits/vec.	LSD (dB)	LSD outliers (in %) 2 – 4 dB	>4 dB
32 mixture components	47	1.017	1.035	0.000	48	1.037	0.780	0.000	48	1.009	0.935	0.000
	48	0.982	0.825	0.000	49	0.996	0.460	0.000	49	0.971	0.700	0.000
64 mixture components	46	1.032	1.160	0.000	47	1.027	0.607	0.003	48	1.008	1.090	0.008
	47	0.987	0.840	0.000	48	0.994	0.387	0.001	49	0.963	0.765	0.005
128 mixture components	45	1.029	1.190	0.000	46	1.007	1.257	0.006	47	1.005	1.430	0.007
	46	0.989	0.880	0.000	47	0.965	1.030	0.005	48	0.956	0.700	0.003
256 mixture components	44	1.039	1.200	0.000	46	1.012	1.617	0.002	47	1.029	0.776	0.005
	45	0.997	0.830	0.000	47	0.946	0.973	0.002	48	0.971	0.920	0.003

[1] Z. Ma and A. Leijon, 'Modeling Speech Line Spectral Frequencies with Dirichlet Mixture Models', *INTERSPEECH,* 2010.
[2] Z. Ma, A. Leijon, and W. B. Kleijn, 'Vector Quantization of LSF Parameters with a Mixture of Dirichlet Distributions', *IEEE Transaction on Audio, Speech, and Language Processing,* vol.21, no.9, pp.1777-1790, Sep. 2013.

Dirichlet Mixture Model

- PRObabilistic Multiview Depth Enhancement (PROMED) [1]

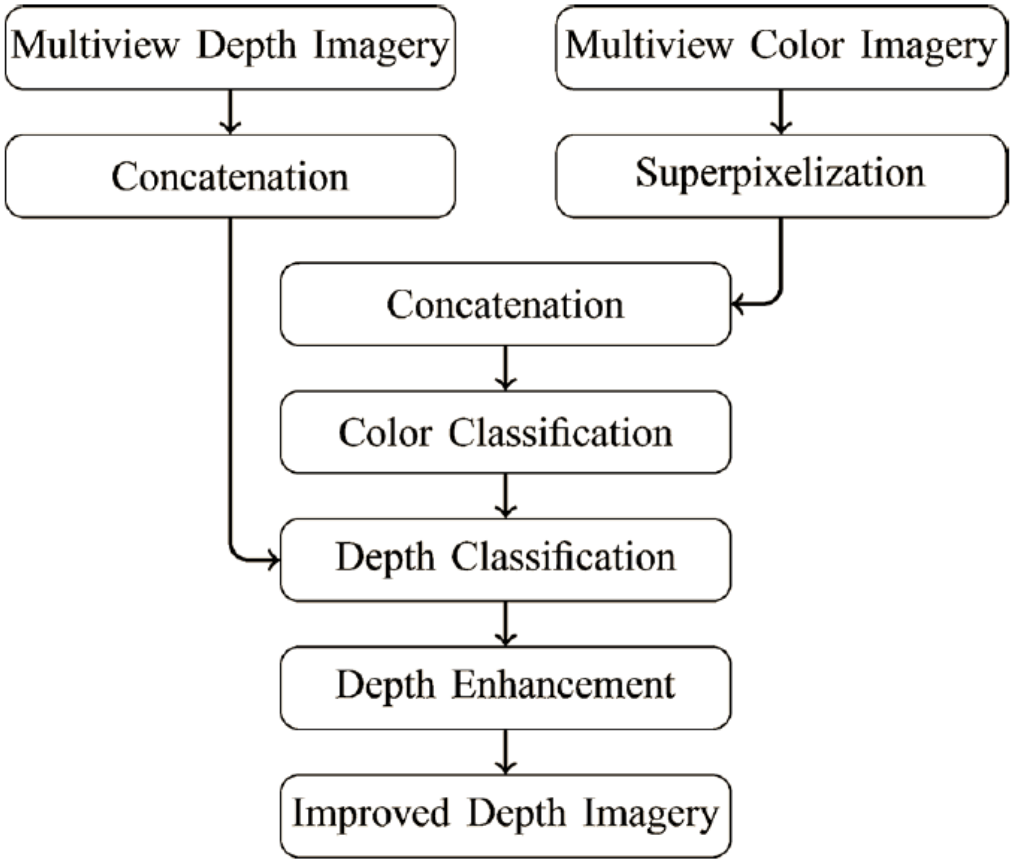


Free-viewpoint TV

Multiview video imagery

[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

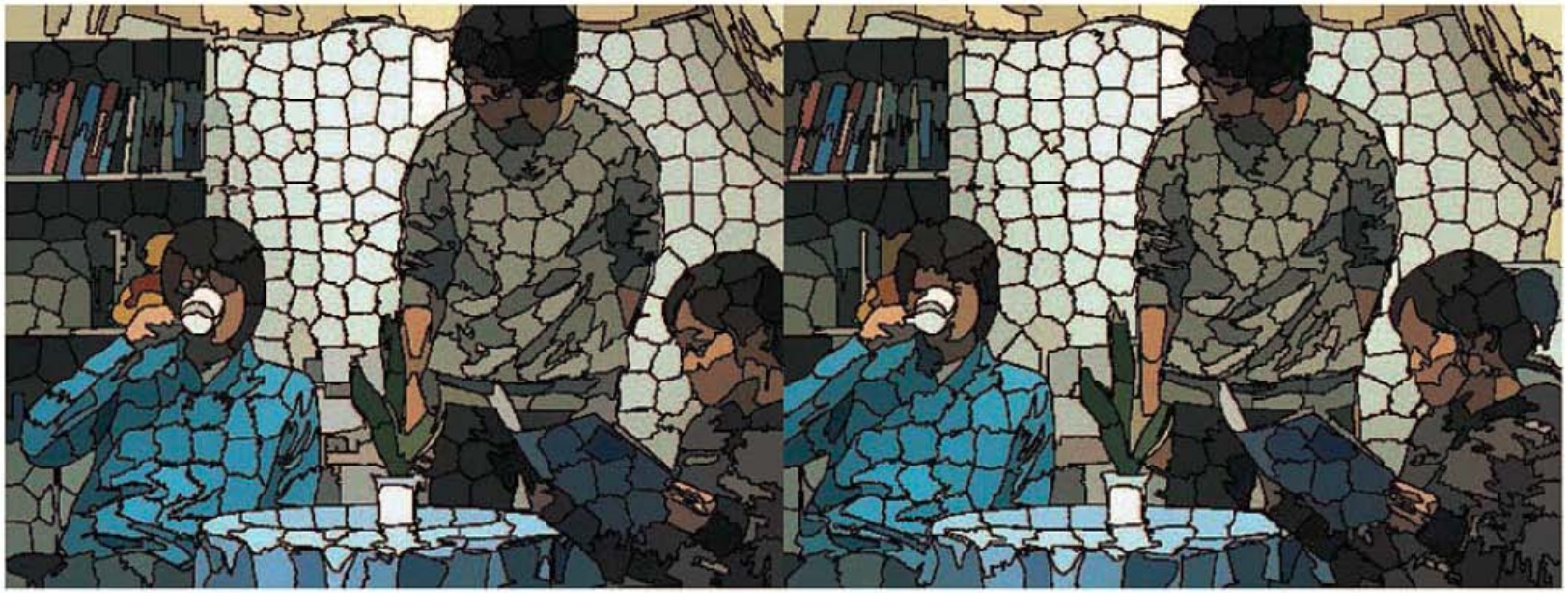
Dirichlet Mixture Model



PROMDE Flow Chart

[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

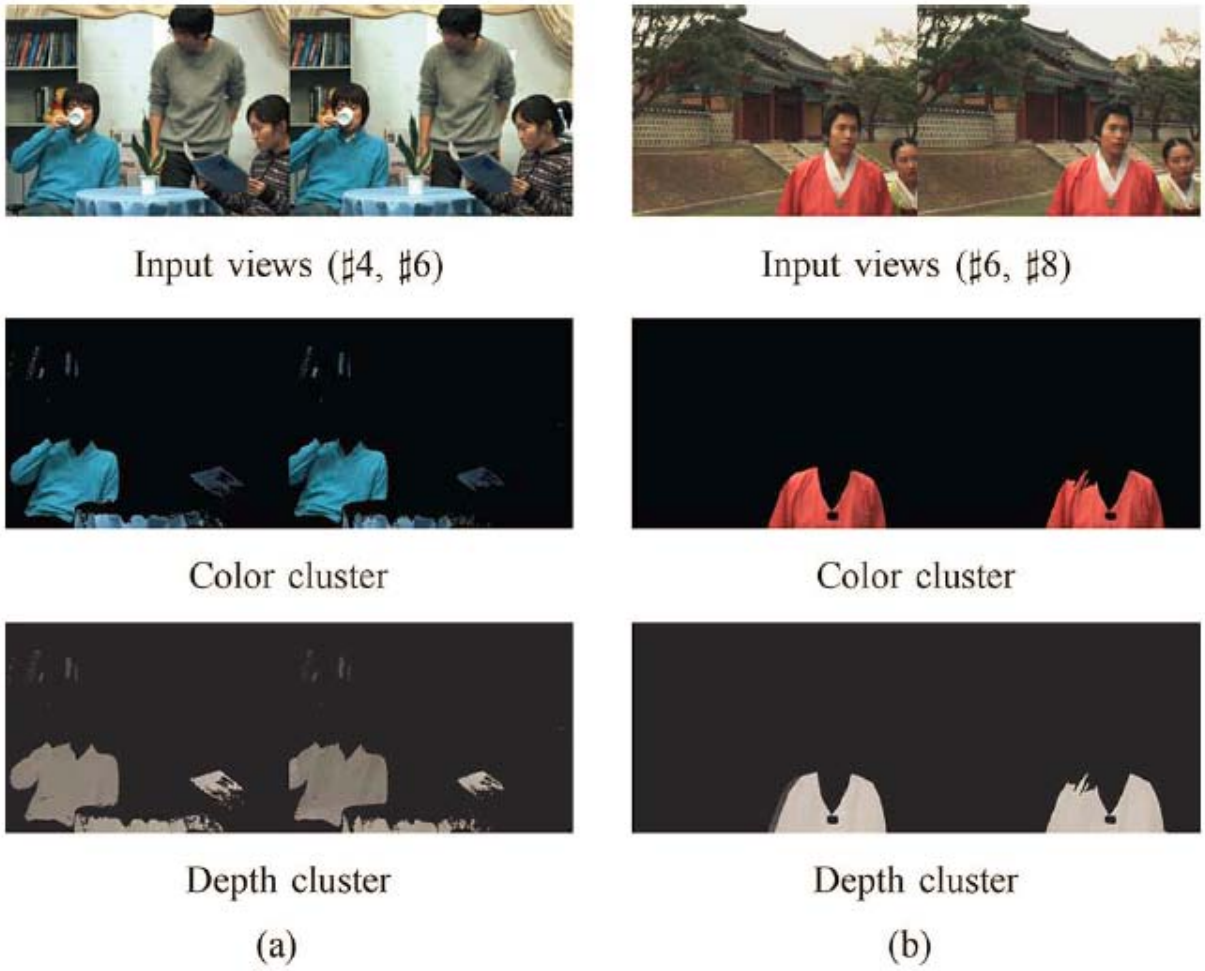
Dirichlet Mixture Model



Two concatenated Newspaper views with approximately superpixels as obtain by using SLIC[1].

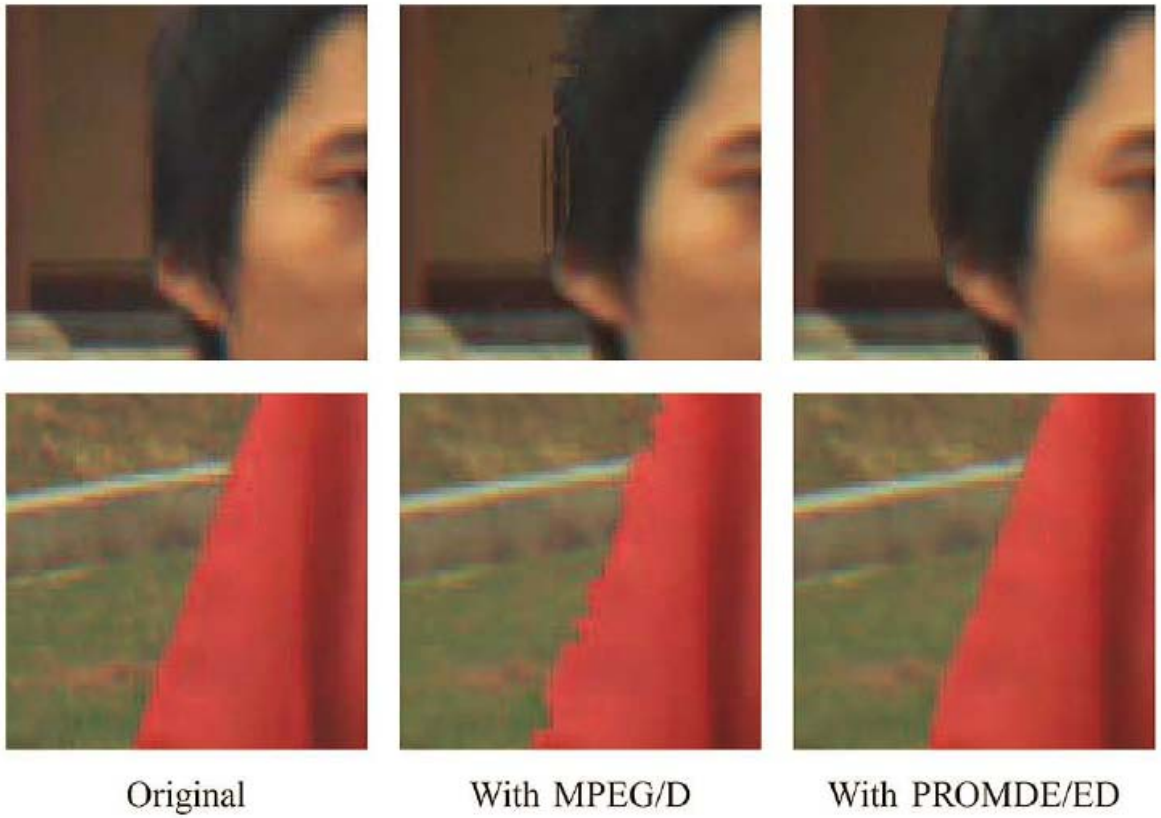
[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

Dirichlet Mixture Model



[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

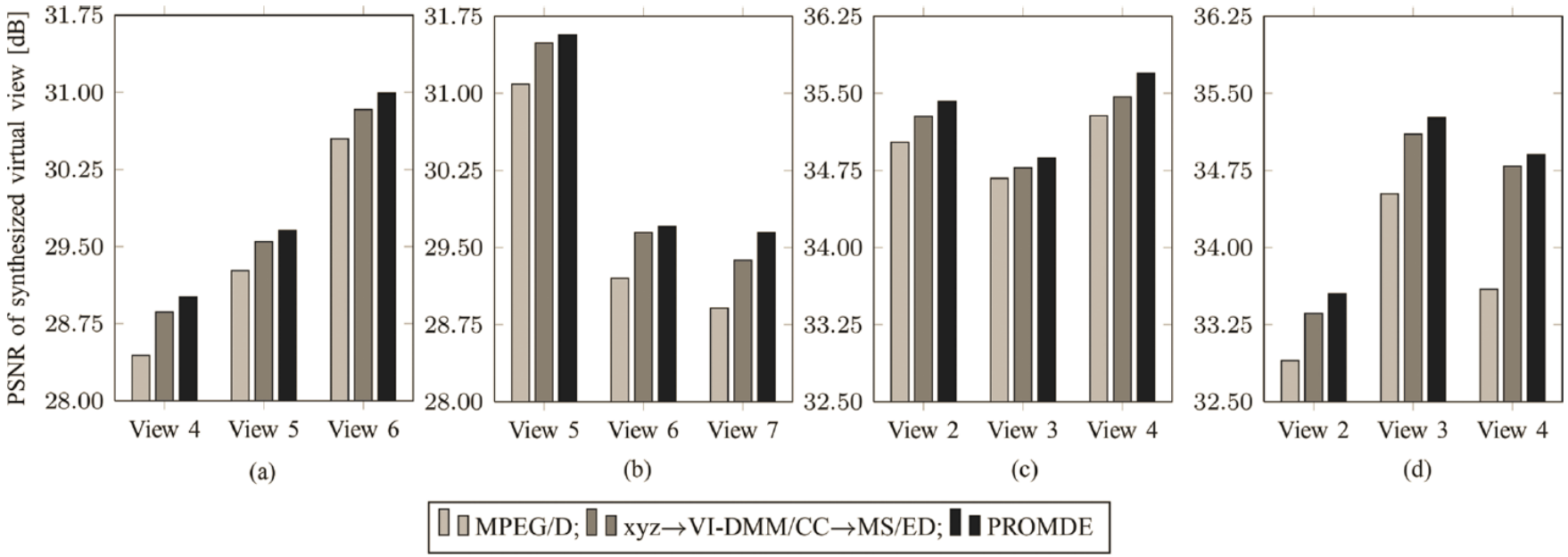
Dirichlet Mixture Model



Selected regions of synthesized virtual views of test sequences as generated by VSRS 3.5 using MPEG depth maps and enhanced depth maps from our depth enhancement algorithm.

[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

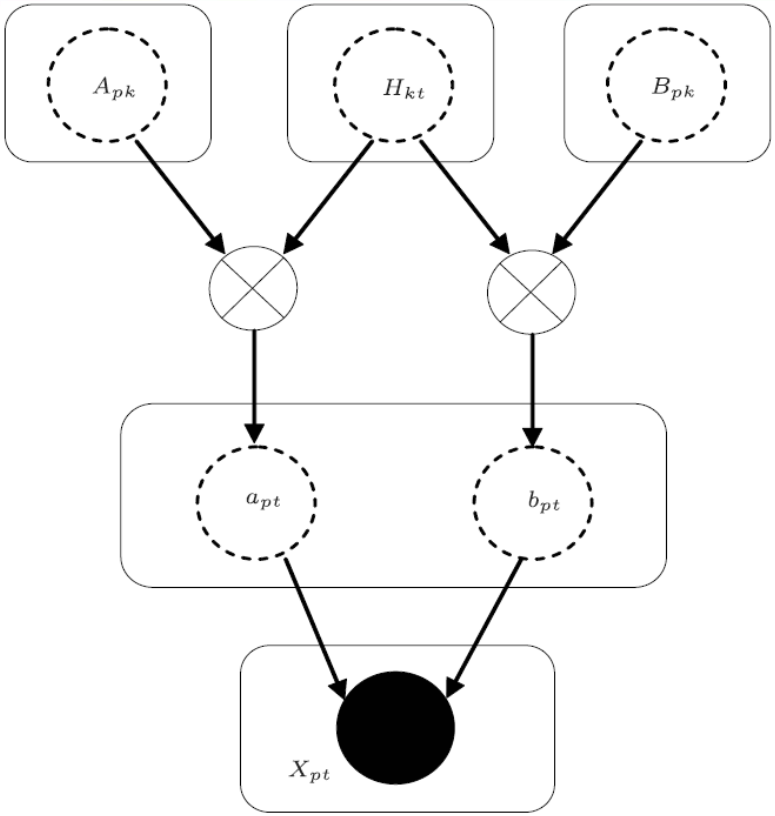
Dirichlet Mixture Model



The objective quality of three intermediate virtual views as generated by VSRS 3.5 using the large baseline setting.

[1] P. K. Rana, J. Taghia, Z. Ma, and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, Volume 9, Issue 3, pp. 435-448, Apr. 2015

Beta Gamma-NMF (BG-NMF)



Graphical Model of BG-NMF[1]

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

- Bayesian matrix factorization for bounded support data→BGNMF
- Handle highly sparse matrix → low rank matrix approximation

$$X_{p,t} \sim Beta\left(X_{p,t}; a_{p,t}, b_{p,t}\right) = \frac{\Gamma\left(a_{p,t}+b_{p,t}\right)}{\Gamma\left(a_{p,t}\right)\Gamma\left(b_{p,t}\right)} X_{p,t}^{a_{p,t}-1}\left(1-X_{p,t}\right)^{b_{p,t}-1}$$

$$\underbrace{X}_{P\times T} \begin{cases} \underbrace{a}_{P\times T} = \underbrace{A}_{P\times K} \times \underbrace{z}_{K\times T} \\ \underbrace{b}_{P\times T} = \underbrace{B}_{P\times K} \times \underbrace{z}_{K\times T} \end{cases} \Longrightarrow \begin{cases} A_{p,k} \sim Gam\left(A_{p,k}; \alpha_{p,k}, \mu_{p,k}\right) \propto A_{p,k}^{\mu_{p,k}-1} e^{-\alpha_{p,k} A_{p.k}} \\ B_{p,k} \sim Gam\left(B_{p,k}; \beta_{p,k}, \nu_{p,k}\right) \propto B_{p,k}^{\nu_{p,k}-1} e^{-\beta_{p,k} B_{p.k}} \\ z_{k,t} \sim Gam\left(z_{k,t}; \zeta_{k,t}, \rho_{k,t}\right) \propto z_{k,t}^{\rho_{k,t}-1} e^{-\zeta_{k,t} z_{k,t}} \end{cases}$$

$$\hat{X} = \frac{Az}{Az + Bz}$$

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

$$\ln q^*(A_{pk}) = \mathbf{E}_{\setminus q(A_{pk})}[\ln p(\mathbf{X}, \mathbf{Z})] + \text{const}$$

$$= \sum_t \mathbf{E}_{\setminus q(A_{pk})} \underbrace{\left[-\ln \mathcal{B}\left(\sum_k A_{pk} H_{kt}, \sum_k B_{pk} H_{kt} \right) \right]}_{\mathrm{F}(A_{p,:}, B_{p,:}, H_{:,t})^2}$$

$$+ \left(\sum_t \overline{H}_{kt} \ln X_{pt} \right) A_{pk}$$

$$+ (\mu_0 - 1) \ln A_{pk} - \alpha_0 A_{pk} + \text{const},$$

Objective function. Need to find auxiliary function for the LIB function $\mathrm{F}(A_{p,:}, B_{p,:}, H_{:,t})$

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

$$\mathbf{E}_{q(\mathbf{Z})}\left[\mathrm{F}_{pt}\right] \geq -\ln \mathcal{B}\left(\sum_k \overline{A}_{pk}\overline{H}_{kt}, \sum_k \overline{B}_{pk}\overline{H}_{kt}\right)$$
$$+\left[\psi\left(\sum_k(\overline{A}_{pk}\overline{H}_{kt}+\overline{B}_{pk}\overline{H}_{kt})\right)-\psi\left(\sum_k \overline{A}_{pk}\overline{H}_{kt}\right)\right]$$
$$\times\sum_k \overline{A}_{pk}\overline{H}_{kt}\left\{\mathbf{E}_{q(A_{p,k})q(H_{k,t})}\left[\ln(A_{pk}H_{kt})\right]-\ln(\overline{A}_{pk}\overline{H}_{kt})\right\}$$
$$+\left[\psi\left(\sum_k(\overline{A}_{pk}\overline{H}_{kt}+\overline{B}_{pk}\overline{H}_{kt})\right)-\psi\left(\sum_k \overline{B}_{pk}\overline{H}_{kt}\right)\right]$$
$$\times\sum_k \overline{B}_{pk}\overline{H}_{kt}\{\mathbf{E}_{q(B_{p,k})q(H_{k,t})}[\ln(B_{pk}H_{kt})]-\ln(\overline{B}_{pk}\overline{H}_{kt})\}$$

Auxiliary function with relative convexity,
Jensen inequality.

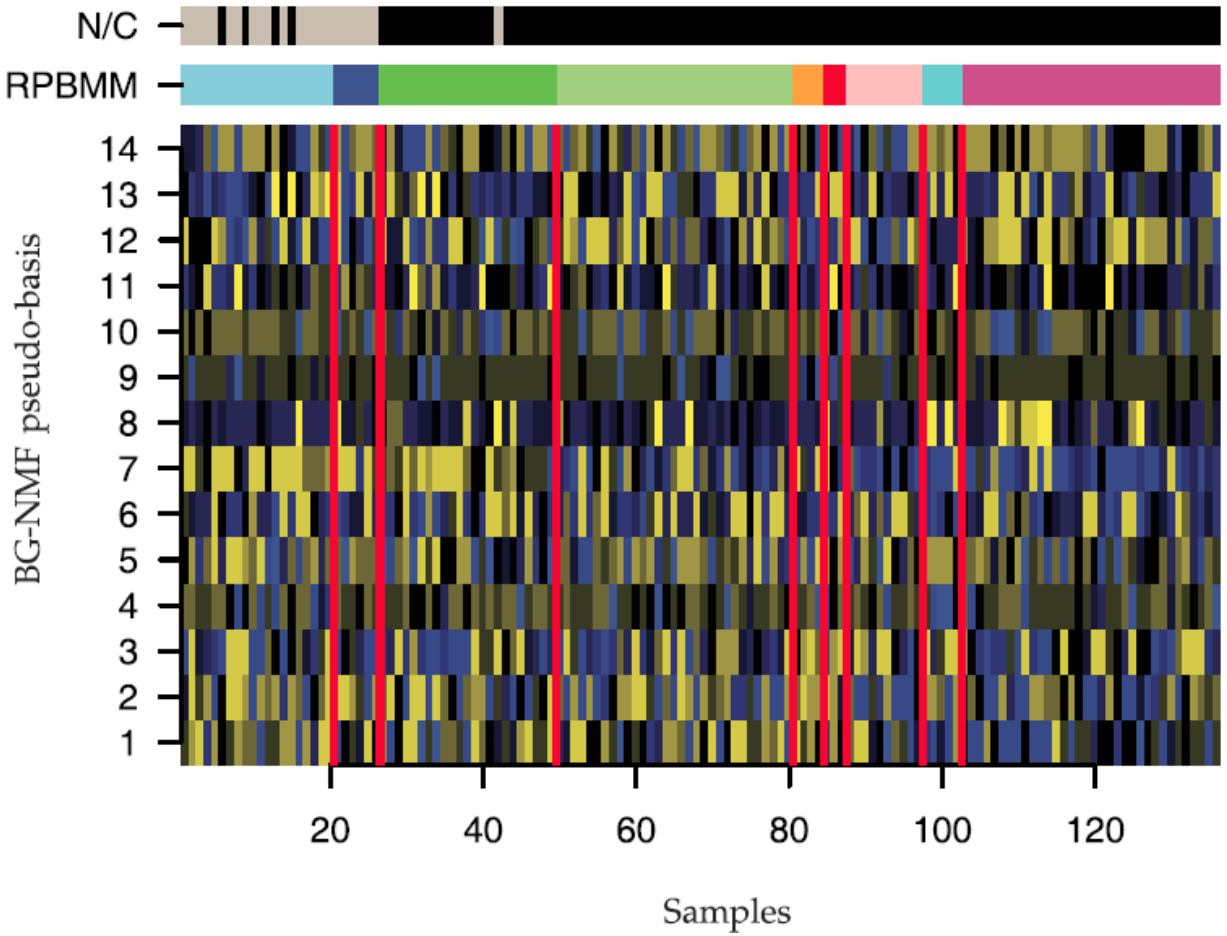
[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, “Variational Bayesian Matrix Factorization for Bounded Support Data”, *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

- DNA methylation analysis [1]
  - Motivation: using statistical model as a robust analysis tool in bioinformatics area
  - Improve analyzing performance comparing with benchmark methods
  - DNA methylation matrix of 27k ×136
  - Methylation level in [0,1]
  - Preprocessing: feature selection via variance.
  - 27k→5000

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

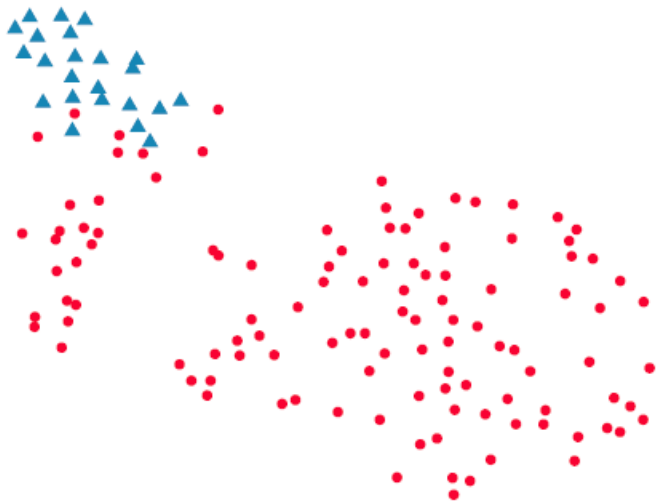


BG-NMF，5000→14.

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

Beta Gamma-NMF (BG-NMF)

– PCA + VB-GMM (5000→14)



(a) Data visualization.

(b) Clustering result via PCA+VBGMM.

9 cancers to normal
0 normal to cancer
9 out of 136

Beta Gamma-NMF (BG-NMF)

– BGNMF+ VB-BMM (5000→14)

(a) Data visualization.

(b) Clustering result via BG-NMF+VBBMM.

4 cancers to normal
1 normal to cancer
5 out of 136
124 sec. < 139 sec. (RPBMM)

Related Applications

- Summary
  - EVI-based NG statistical model shows advantages in several applications.
  - Fitting data better → improved performance
  - Needs a lot of effort to design and derive.

References

[1] Z. Ma, A.E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian Matrix Factorization for Bounded Support Data", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 37, Issue 4, pp. 876 – 889, Apr. 2015.

[2] Z. Ma and A. Leijon, "Bayesian Estimation of Beta Mixture Models with Variational Inference", *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 33, pp. 2160 – 2173, Nov. 2011.

[3] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian Estimation of Dirichlet Mixture Model with Variational Inference", *Pattern Recognition (PR)*, Volume 47, Issue 9, pp. 3143-3157, September 2014.

[4] J. Taghia, Z. Ma, A. Leijon, "Bayesian Estimation of the von-Mises Fisher Mixture Model with Variational Inference*", IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 36, Issue 9, pp. 1701-1715, September, 2014.

[5] P. K. Rana, J. Taghia, Z. Ma**,** and M. Flierl, "Probabilistic Multiview Depth Image Enhancement Using Variational Inference", *IEEE Journal of Selected Topics in Signal Processing (J-STSP),* Volume 9, Issue 3, pp. 435-448, Apr. 2015

Beijing University of Posts and Telecommunications

Thanks!