



**HAL**  
open science

# Non parametric Bayesian priors for hidden Markov random fields

Florence Forbes, Hongliang Lu, Julyan Arbel

► **To cite this version:**

Florence Forbes, Hongliang Lu, Julyan Arbel. Non parametric Bayesian priors for hidden Markov random fields. JSM 2018 - Joint Statistical Meeting, Jul 2018, Vancouver, Canada. pp.1-38. hal-01941679

**HAL Id: hal-01941679**

**<https://hal.science/hal-01941679v1>**

Submitted on 12 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation

Hongliang Lü, Julyan Arbel, [Florence Forbes](#)

*Inria Grenoble Rhône-Alpes and University Grenoble Alpes*  
*Laboratoire Jean Kuntzmann*  
*Mistis team*  
[florence.forbes@inria.fr](mailto:florence.forbes@inria.fr)

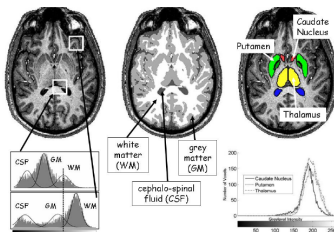
July 2018



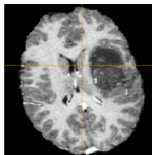
# Unsupervised image segmentation

## Challenges for mixture models (clustering)

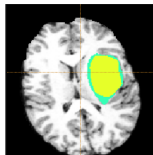
inhomogeneities, noise



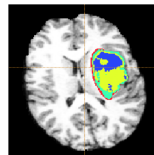
How many segments?



T1 gado



2 classes

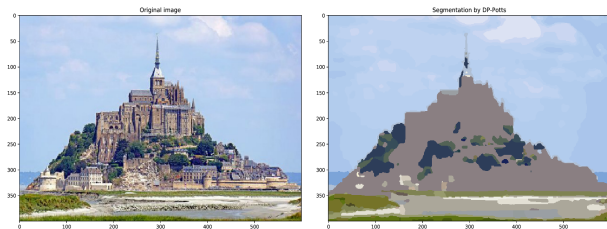


4 classes

Extensions of Dirichlet Process mixture model with spatial regularization

# Outline of the talk

- 1 Dirichlet process (DP)
- 2 Spatially-constrained mixture model: DP-Potts mixture model
  - Finite mixture model
  - Bayesian finite mixture model
  - DP mixture model
  - DP-Potts mixture model
- 3 Inference using variational approximation
- 4 Some image segmentation results
- 5 Conclusion and future work



# Dirichlet process (DP)

The DP is a central Bayesian nonparametric (BNP) prior<sup>1</sup>.

## Definition (Dirichlet process)

A **Dirichlet process** on the space  $\mathcal{Y}$  is a **random process**  $G$  such that there exist  $\alpha$  (concentration parameter) and  $G_0$  (base distribution) such that for any finite partition  $\{A_1, \dots, A_p\}$  of  $\mathcal{Y}$ , the random vector  $(P(A_1), \dots, P(A_p))$  will be Dirichlet distributed:

$$(P(A_1), \dots, P(A_p)) \sim \text{Dir}(\alpha G_0(A_1), \dots, \alpha G_0(A_p))$$

Notation:  $G \sim \text{DP}(\alpha, G_0)$

The DP is the infinite-dimensional generalization of the Dirichlet distribution.

<sup>1</sup>Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

# Dirichlet process (DP) construction

A DP prior  $G$  can be constructed using three methods:

- The Blackwell-MacQueen urn scheme
- The Chinese Restaurant Process
- The Stick-Breaking construction

---

<sup>2</sup>Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639-650.

# Dirichlet process (DP) construction

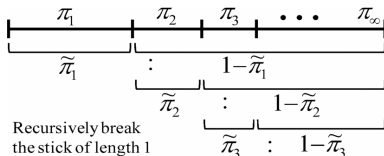
A DP prior  $G$  can be constructed using three methods:

- The Blackwell-MacQueen urn scheme
- The Chinese Restaurant Process
- **The Stick-Breaking construction**

The DP has almost surely **discrete** realizations<sup>2</sup>:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}$$

where  $\theta_k^* \stackrel{\text{iid}}{\sim} G_0$  and  $\pi_k = \tilde{\pi}_k \prod_{l < k} (1 - \tilde{\pi}_l)$  with  $\tilde{\pi}_k \stackrel{\text{iid}}{\sim} \text{Beta}(1, \alpha)$ .



<sup>2</sup>Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639-650.

# Spatially-constrained mixture model: DP-Potts mixture

Clustering/segmentation: Finite mixture models assume data are generated by a finite sum of probability distributions:

$$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N) \text{ with } \mathbf{y}_i = (y_{i1}, \dots, y_{iD}) \in \mathbb{R}^D \text{ i.i.d}$$

$$p(\mathbf{y}_i | \theta^*, \pi) = \sum_{k=1}^K \pi_k F(\mathbf{y}_i | \theta_k^*)$$

where

- $\theta^* = (\theta_1^*, \dots, \theta_K^*)$  and  $\pi = (\pi_1, \dots, \pi_K)$  with  $\theta^*$  class parameters and  $\pi$  mixture weights with  $\sum_{i=1}^K \pi_i = 1$ .
- $\theta^*$  and  $\pi$  can be estimated using EM algorithm.

Equivalently

- $G = \sum_{k=1}^K \pi_k \delta_{\theta_k^*}$  non random
- $\theta_i \sim G$  and  $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$ .



# Bayesian finite mixture model

In a Bayesian setting, a prior distribution is placed over  $\theta^*$  and  $\pi$ .

Thus, the posterior distribution of parameters given the observations is

$$p(\theta^*, \pi | \mathbf{y}) \propto p(\mathbf{y} | \theta^*, \pi) p(\theta^*, \pi)$$

To generate a data point within a **Bayesian finite mixture model**:

- $\theta_k^* \sim G_0$
- $\pi_1, \dots, \pi_K \sim \text{Dir}(\alpha/K, \dots, \alpha/K)$
- $G = \sum_{k=1}^K \pi_k \delta_{\theta_k^*}$  is a random measure
- $\theta_i | G \sim G$ , which means  $\theta_i = \theta_k^*$  with probability  $\pi_k$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

# Bayesian finite mixture model

In a Bayesian setting, a prior distribution is placed over  $\theta^*$  and  $\pi$ .

Thus, the posterior distribution of parameters given the observations is

$$p(\theta^*, \pi | \mathbf{y}) \propto p(\mathbf{y} | \theta^*, \pi) p(\theta^*, \pi)$$

To generate a data point within a **Bayesian finite mixture model**:

- $\theta_k^* \sim G_0$
- $\pi_1, \dots, \pi_K \sim \text{Dir}(\alpha/K, \dots, \alpha/K)$
- $G = \sum_{k=1}^K \pi_k \delta_{\theta_k^*}$  is a random measure
- $\theta_i | G \sim G$ , which means  $\theta_i = \theta_k^*$  with probability  $\pi_k$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

## Limitation:

Require specifying the number of components  $K$  beforehand.

## Solution:

Assume an infinite number of components using BNP priors.

# DP mixture model

From a Bayesian finite mixture model to a DP mixture model

To establish a DP mixture model, let  $G$  be a DP prior ( $K \rightarrow \infty$ ), namely

$$G \sim \text{DP}(\alpha, G_0)$$

and complement it with a likelihood associated to each  $\theta_i$

To generate a data point within a **DP mixture model**:

- $G \sim \text{DP}(\alpha, G_0)$
- $\theta_i | G \sim G$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

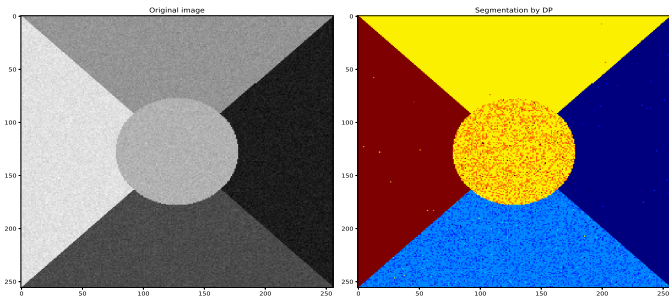
# DP mixture model

2D point clustering (unsupervised learning) based on the DP mixture model:

Let the data speak for themselves!

# DP mixture model

Application to image segmentation:



## Drawback:

Spatial constraints and dependencies are not considered.

## Solution:

Combine the DP prior with a hidden Markov random field (HMRF).

# DP-Potts mixture model

To solve the issue, we introduce a spatial Potts model component:

$$M(\theta) \propto \exp \left( \beta \sum_{i \sim j} \delta_{z(\theta_i)=z(\theta_j)} \right)$$

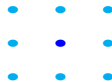
with  $\theta = (\theta_1, \dots, \theta_N)$  and  $\beta$  the interaction parameter.

The DP mixture model is thus extended:

- $G \sim \text{DP}(\alpha, G_0)$
- $\theta | M, G \sim M(\theta) \times \prod_i G(\theta_i)$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$



4-neighbours



8-neighbours

# DP-Potts mixture model

Other spatially-constrained BNP mixture models + inference algorithms:

- DP or PYP-Potts partition model + MCMC<sup>3</sup>
- Hemodynamic brain parcellation (DP-Potts) + PARTIAL VB<sup>4</sup>
- DP or PYP-Potts + Iterated Conditional Mode (ICM)<sup>5</sup>

## Markov chain Monte Carlo (MCMC):

- Advantage: asymptotically exact
- Drawback: computationally expensive

## Variational Bayes (VB):

- Advantage: much faster
- Drawback: less accurate, no theoretical guarantee

We propose a **DP-Potts mixture model** based on a **general stick-breaking construction** that allows a **natural Full VB algorithm** enabling scalable inference for large datasets and straightforward generalization to other priors.

<sup>3</sup>Orbanz & Buhmann (2008); Xu, Caron & Doucet (2016); Sodjo, Giremus, Dobigeon & Giovannelli (2017)

<sup>4</sup>Albughdadi, Chaari, Tournet, Forbes, Ciuciu (2017)

<sup>5</sup>Chatzis & Tsechpenakis (2010); Chatzis (2013)

# DP-Potts: Stick breaking construction

Stick breaking construction of DP:  $G \sim DP(\alpha, G_0)$

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, 2, \dots$
- $G = \sum_{k=1}^{\infty} \pi_k(\tau) \delta_{\theta_k^*}$

+

- $\theta_i | G \sim G$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

= Dirichlet Process Mixture Model (DPMM)



# DP-Potts: Stick breaking construction

## Stick breaking construction of DPMM

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \dots$
- $G = \sum_{k=1}^{\infty} \pi_k(\tau) \delta_{\theta_k^*} \implies$
- $\theta_i | G \sim G$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

## Stick breaking construction of DPMM

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \dots$
- $\theta_i = \theta_k^*$  with probability  $\pi_k(\tau)$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$

# DP-Potts: Stick breaking construction

Using assignment variables  $z_i$

## DPMM view

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \dots$
- $\theta_i = \theta_k^*$  with probability  $\pi_k(\tau)$
- $\mathbf{y}_i | \theta_i \sim F(\mathbf{y}_i | \theta_i)$   $\implies$

## Mixture/Clustering view

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \dots$
- $p(z_i = k | \tau) = \pi_k(\tau)$
- with  $z_i = z(\theta_i) = k$  when  $\theta_i = \theta_k^*$
- $\mathbf{y}_i | z_i, \theta^* \sim F(\mathbf{y}_i | \theta_{z_i}^*)$

# DP-Potts: Stick breaking construction

Using assignment variables  $z_i$

## Stick breaking of DPMM

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l)$
- $p(z_i = k | \tau) = \pi_k(\tau)$
- $\mathbf{y}_i | z_i, \theta^* \sim F(\mathbf{y}_i | \theta_{z_i}^*)$

## Stick breaking of DP-Potts

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \dots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l)$
- $p(\mathbf{z} | \tau, \beta) \propto \prod_i \pi_{z_i}(\tau) \exp(\beta \sum_{i \sim j} \delta_{z_i = z_j})$
- $\mathbf{z} = \{z_1, \dots, z_N\}$
- $\mathbf{y}_i | z_i, \theta^* \sim F(\mathbf{y}_i | \theta_{z_i}^*)$

**NB:** Well defined for every stick breaking construction ( $\sum_{k=1}^{\infty} \pi_k = 1$ ) :

e.g. Pitman-Yor ( $\tau_k | \alpha, \sigma) \sim \mathcal{B}(1 - \sigma, \alpha + k\sigma)$

# Inference using variational approximation

Clustering/ segmentation task:

- Estimating  $\mathbf{Z}$
- while parameters  $\Theta$  unknown , eg.  $\Theta = \{\tau, \alpha, \theta^*\}$

## Bayesian setting

Access the intractable  $p(\mathbf{Z}, \Theta | \mathbf{y}, \Phi)$  approximate as  $q(\mathbf{z}, \Theta) = q_z(\mathbf{z})q_\theta(\Theta)$

## Variational Expectation-Maximization

Alternate maximization in  $q_z$  and  $q_\theta$  ( $\phi$  are hyperparameters) of the Free Energy:

$$\begin{aligned} \mathcal{F}(q_z, q_\theta, \phi) &= E_{q_z q_\theta} \left[ \log \frac{p(\mathbf{y}, \mathbf{Z}, \Theta | \phi)}{q_z(\mathbf{z})q_\theta(\Theta)} \right] \\ &= \log p(\mathbf{y} | \phi) - KL(q_z q_\theta, p(\mathbf{Z}, \Theta | \mathbf{y}, \phi)) \end{aligned}$$

# DP-Potts Variational EM procedure

Joint DP-Potts (Gaussian) Mixture distribution

$$p(\mathbf{y}, \mathbf{z}, \boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^* | \phi) = \prod_{j=1}^N p(y_j | z_j, \boldsymbol{\theta}^*) p(\mathbf{z} | \boldsymbol{\tau}, \beta) \prod_{k=1}^{\infty} p(\tau_k | \alpha) \prod_{k=1}^{\infty} p(\boldsymbol{\theta}_k^* | \rho_k) p(\alpha | s_1, s_2)$$

- $p(y_j | z_j, \boldsymbol{\theta}^*) = \mathcal{N}(y_j | \mu_{z_j}, \Sigma_{z_j})$  is Gaussian
- $p(\mathbf{z} | \boldsymbol{\tau}, \beta)$  is a **DP-Potts model**
- $p(\tau_k | \alpha)$  is Beta  $\mathcal{B}(1, \alpha)$
- $p(\boldsymbol{\theta}_k^* | \rho_k) = \mathcal{NIW}(\mu_k, \Sigma_k | m_k, \lambda_k, \Psi_k, \nu_k)$  is Normal-inverse-Wishart
- $p(\alpha | s_1, s_2) = \mathcal{G}(\alpha | s_1, s_2)$  is Gamma

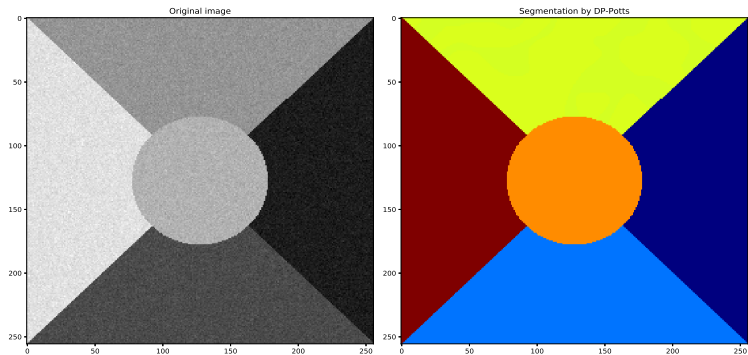
Usual **truncated** variational posterior,  $q_{\tau_k} = \delta_1$  for  $k \geq K$  (eg.  $K = 40$ )

$$q(\mathbf{z}, \boldsymbol{\Theta}) = \prod_{j=1}^N q_{z_j}(z_j) q_{\alpha}(\alpha) \prod_{k=1}^{K-1} q_{\tau_k}(\tau_k) \prod_{k=1}^K q_{\boldsymbol{\theta}_k^*}(\mu_k, \Sigma_k)$$

- E-steps: VE-Z, VE- $\alpha$ , VE- $\boldsymbol{\tau}$  and VE- $\boldsymbol{\theta}^*$
- M-step:  $\phi$  updating straightforward except for  $\beta$

# Some image segmentation results

Model validation and verification:



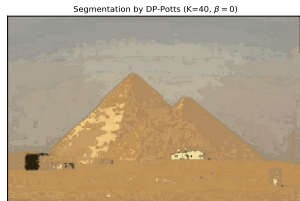
Segmented image using DP-Potts model with  $\beta = 2.5$ .

# Some image segmentation results

Convergence of the VB algorithm initialized by the k-means++ clustering:

# Some image segmentation results

Segmentation results for Berkeley Segmentation Dataset:



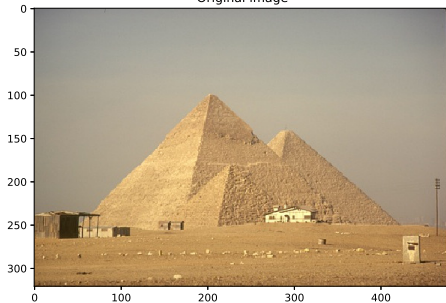
The segmentation results obtained by DP-Potts model with  $\beta = 0, 1, 5$ .



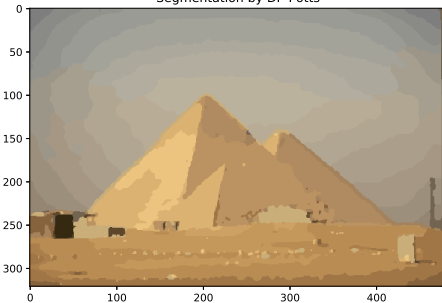
# Some image segmentation results

Segmentation with estimated  $\beta = 1.66$

Original image



Segmentation by DP-Potts



# Quantitative evaluation of the segmentations

**Probabilistic Rand Index** on 154 color (RGB) images with ground truth (several) from Berkeley dataset (1000 superpixels). But Manual ground truth segmentations are subjective !

PRI results with DP-Potts model

	Mean	Median	St.D.
K=10	71.48	72.54	0.1040
K=20	73.64	73.42	0.0935
K=40	75.33	76.47	0.0853
K=50	75.81	76.31	0.0873
K=60	76.55	77.12	0.0848
K=80	<b>77.06</b>	<b>78.30</b>	<b>0.0835</b>

PRI results from Chatzis 2013

PRI (%)	DPM	iHMRF	MRF-PYP	GC
Mean	74.15	75.50	76.49	76.10
Median	75.49	76.89	78.08	77.59
St.D.	0.084	0.082	0.079	0.083

**Computation time** : Berkeley 321x481 image reduced to 1000 superpixels takes **10-30 s** on a PC with CPU Intel(R) Core(TM) i7-5500U CPU 2.40GHz and 8GB RAM

# Conclusion and future work

- A general DP-Potts model and the associated VB algorithm were built.
- The DP-Potts model was applied to image segmentation and tested on different types of datasets.
- Impact of the interaction parameter  $\beta$  on the final results is significant.
- An estimation procedure was proposed for  $\beta$

# Conclusion and future work

- A general DP-Potts model and the associated VB algorithm were built.
  - The DP-Potts model was applied to image segmentation and tested on different types of datasets.
  - Impact of the interaction parameter  $\beta$  on the final results is significant.
  - An estimation procedure was proposed for  $\beta$
- 
- How does  $\beta$  influence the number of components?
  - Extend the model with other priors (Pitman-Yor process, Gibbs-type priors, etc.).
  - Try other variational approximations (truncation-free)
  - Investigate theoretical properties of BNP priors under structural constraints (time, spatial) ....
  - ... for other applications, such as discovery probabilities, etc.

# References

- 1 M. Albughdadi, L. Chaari, J.-Y. Tourneret, F. Forbes, and P. Ciuciu. A Bayesian non-parametric hidden Markov random model for hemodynamic brain parcellation. *Signal Processing*, 135:132-146, 2017.
- 2 S. P. Chatzis. A Markov random field-regulated Pitman-Yor process prior for spatially constrained data clustering. *Pattern Recognition*, 46(6): 1595-1603, 2013.
- 3 S. P. Chatzis and G. Tsechpenakis. The infinite hidden Markov random field model. *IEEE Trans. Neural Networks*, 21(6):1004-1014, 2010.
- 4 F. Forbes and N. Peyrard. Hidden Markov Random Field Selection Criteria based on Mean Field-like approximations. *IEEE PAMI*, 25(9):1089-1101, 2003
- 5 P. Orbanz and J. M. Buhmann. Nonparametric Bayesian image segmentation. *International Journal of Computer Vision*, 77(1-3):25-45, 2008.
- 6 J. Sodjo, A. Giremus, N. Dobigeon, J.F. Giovannelli, A generalized Swendsen-Wang algorithm for Bayesian nonparametric joint segmentation of multiple images, *ICASSP*, 2017.
- 7 R. Xu, F. Caron, and A. Doucet. Bayesian nonparametric image segmentation using a generalized Swendsen-Wang algorithm. *ArXiv e-prints*, February 2016.

*Thank you for your attention!*

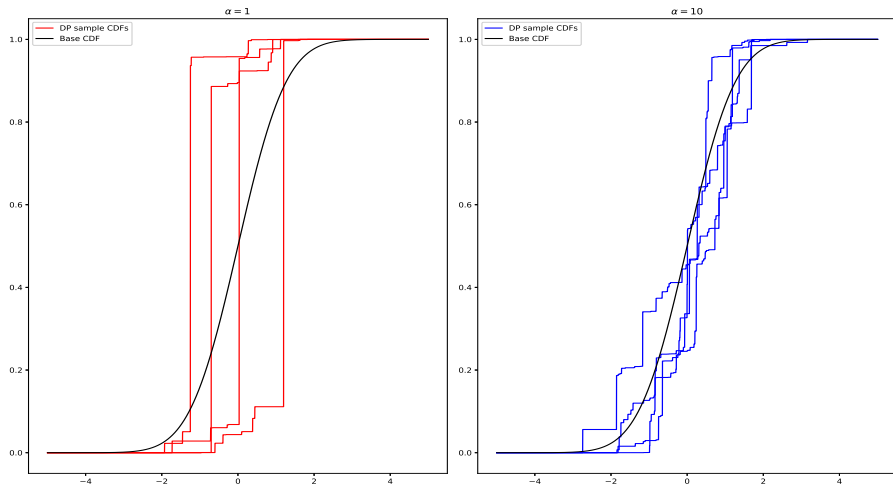
contact: [florence.forbes@inria.fr](mailto:florence.forbes@inria.fr)



Université Grenoble Alpes invites applications for a **2-year junior research chair** (post-doc) in **Data Science for Life Sciences and Health**

- Starting in October 2018
- Data science methodology and machine learning to Life Sciences and Health
- Application deadline: **August, 31 2018**
- Website: <https://data-institute.univ-grenoble-alpes.fr/>
- Contact: [florence.forbes@inria.fr](mailto:florence.forbes@inria.fr)

# Stick breaking construction



DP simulations with  $G_0$  being a standard normal distribution  $\mathcal{N}(0, 1)$  and  $\alpha = 1, 10$  using the Stick-Breaking representation.

# Variational EM

## General formulation, at iteration ( $r$ )

$$\text{E-Z } q_z^{(r)}(\mathbf{z}) \propto \exp \left( E_{q_\theta^{(r-1)}} [\log p(\mathbf{y}, \mathbf{z}, \Theta | \phi^{(r-1)})] \right)$$

$$\text{E-}\Theta \ q_\theta^{(r)}(\Theta) \propto \exp \left( E_{q_z^{(r)}} [\log p(\mathbf{y}, \mathbf{Z}, \Theta | \phi^{(r-1)})] \right)$$

$$\text{M-}\phi \ \phi^{(r)} = \arg \max_{\phi} E_{q_z^{(r)} q_\theta^{(r)}} [\log p(\mathbf{y}, \mathbf{Z}, \Theta | \phi)]$$

## VE-Z, VE- $\alpha$ , VE- $\tau$ , and VE- $\theta^*$

e.g. VE-Z step divides into  $N$  VE- $Z_j$  steps ( $q_{z_j}(z_j) = 0$  for  $z_j > K$ )

$$q_{z_j}(z_j) \propto \exp \left( E_{q_{\theta_{z_j}^*}} [\log p(y_j | \theta_{z_j}^*)] + E_{q_\tau} [\log \pi_{z_j}(\tau)] + \beta \sum_{i \sim j} q_{z_i}(z_j) \right)$$



# Estimation of $\beta$

M- $\beta$  step: involves  $p(\mathbf{z}|\boldsymbol{\tau}, \beta) = \mathcal{K}(\beta, \boldsymbol{\tau})^{-1} \exp(V(\mathbf{z}; \boldsymbol{\tau}, \beta))$

with  $V(\mathbf{z}; \boldsymbol{\tau}, \beta) = \sum_i \log \pi_{z_i}(\boldsymbol{\tau}) + \beta \sum_{i \sim j} \delta_{(z_i = z_j)}$

$$\begin{aligned}\hat{\beta} &= \arg \max_{\beta} E_{q_{\mathbf{z}} q_{\boldsymbol{\tau}}} [\log p(\mathbf{z}|\boldsymbol{\tau}; \beta)] \\ &= \arg \max_{\beta} E_{q_{\mathbf{z}} q_{\boldsymbol{\tau}}} [V(\mathbf{z}; \boldsymbol{\tau}, \beta)] - E_{q_{\boldsymbol{\tau}}} [\log \mathcal{K}(\beta, \boldsymbol{\tau})]\end{aligned}$$

## Two difficulties

- (1)  $p(\mathbf{z}|\boldsymbol{\tau}, \beta)$  is intractable (normalizing constant  $\mathcal{K}(\beta, \boldsymbol{\tau})$ , typical of MRF)
- (2) it depends on  $\boldsymbol{\tau}$  (typical of DP)

## Two approximations

- (1) "standard" Mean Field like approximation<sup>a</sup>
- (2) Replace the random  $\boldsymbol{\tau}$  by a fixed  $\tilde{\boldsymbol{\tau}} = E_{q_{\boldsymbol{\tau}}}[\boldsymbol{\tau}]$

---

<sup>a</sup>Forbes & Peyrard 2003

# Approximation of $p(\mathbf{z}|\boldsymbol{\tau}; \beta)$

$$p(\mathbf{z}|\boldsymbol{\tau}, \beta) \approx \tilde{q}_z(\mathbf{z}|\beta) = \prod_{j=1}^N \tilde{q}_{z_j}(z_j|\beta)$$

$$\tilde{q}_{z_j}(z_j = k|\beta) = \frac{\exp(\log \pi_k(\tilde{\boldsymbol{\tau}}) + \beta \sum_{i \in N(j)} q_{z_i}(k))}{\sum_{l=1}^{\infty} \exp(\log \pi_l(\tilde{\boldsymbol{\tau}}) + \beta \sum_{i \in N(j)} q_{z_i}(l))} \quad \text{and} \quad \tilde{\boldsymbol{\tau}} = E_{q_{\boldsymbol{\tau}}}[\boldsymbol{\tau}]$$

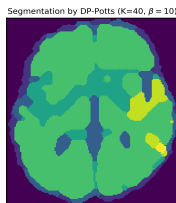
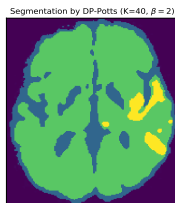
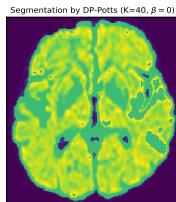
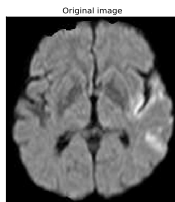
$\beta$  is estimated at each iteration by setting the approximate gradient to 0

$$E_{q_z q_{\boldsymbol{\tau}}} [\nabla_{\beta} V(\mathbf{z}; \boldsymbol{\tau}, \beta)] = \sum_{k=1}^K \sum_{i \sim j} q_{z_j}(k) q_{z_i}(k)$$

$$\nabla_{\beta} E_{q_{\boldsymbol{\tau}}} [\log \mathcal{K}(\beta, \boldsymbol{\tau})] = E_{p(\mathbf{z}|\boldsymbol{\tau}, \beta) q_{\boldsymbol{\tau}}} [\nabla_{\beta} V(\mathbf{z}; \boldsymbol{\tau}, \beta)] \approx \sum_{k=1}^K \sum_{i \sim j} \tilde{q}_{z_j}(k|\beta) \tilde{q}_{z_i}(k|\beta)$$

# Some image segmentation results

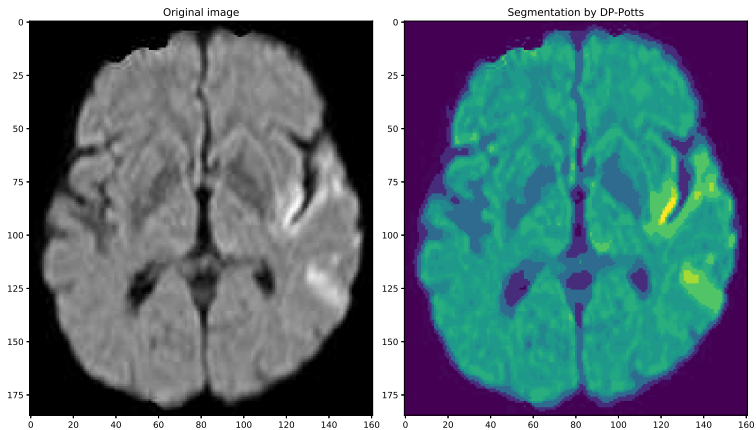
Segmentation results for medical images: all hyperparameters fixed



The segmentation results obtained by DP-Potts model with  $\beta = 0, 1, 5$ .

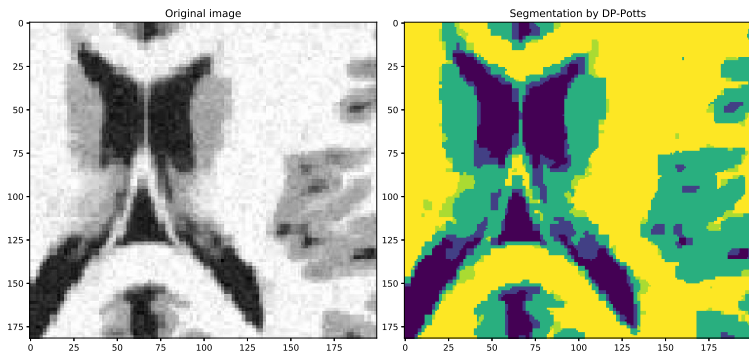
# Some image segmentation results

Segmentation with estimated hyperparameters ( $\beta = 0.75$ )



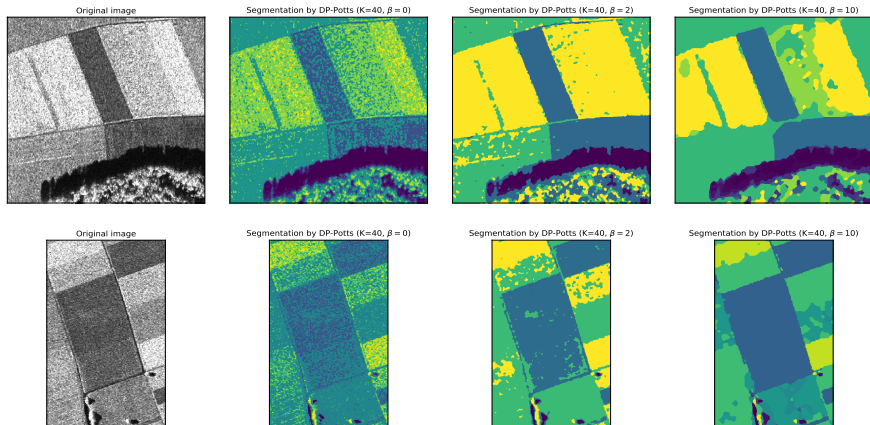
# Some image segmentation results

Segmentation with estimated  $\beta = 0.96$  (pixels with partial volume)



# Some image segmentation results

Segmentation results for SAR images:

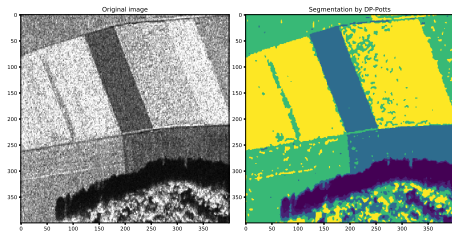


The segmentation results obtained by DP-Potts model with  $\beta = 0, 1, 5$ .

# Some image segmentation results

Segmentation results with estimated  $\beta$

$\beta = 1.11$



$\beta = 1.02$

