Bayesian Nonparametric Approaches for Reconstruction of Dynamical PET Data.

Éric Barat[1]
Claude Comtat[2], Thomas Dautremer[1], Diarra Fall[3],
Thierry Montagu[1] and Simon Stute[2]

[1]Laboratory of Modeling, Simulation and Systems, CEA, LIST
[2]Frédéric Joliot Hospital Service, CEA, SHFJ
[3]MAP5, Paris-Descartes University
eric.barat@cea.fr

ISSSMA Workshop
June 3, 2013

3D usual iterative reconstruction
Parametric indirect regression with basis functions

- Assume a set of normalized basis functions $\phi_1, \ldots, \phi_J$ (e.g. voxels, blobs) and parametrize a function with fixed finite size $\mathbf{g} = \{g_1, \ldots, g_J\}$

$$G(x; \mathbf{g}) = \sum_{j=1}^{J} g_j \, \phi_j(x)$$

- Find optimal parameters (*optimize*) from data $\mathbf{n} = n_1, \ldots, n_I$ where $n_i | \mathbf{g} \overset{iid}{\sim} \text{Poisson}(\sum_{i=1}^{I} p_{ij} g_j)$

$$\hat{\mathbf{g}} = \underset{\mathbf{g}>0}{\text{argmin}}(-\log \mathcal{L}(\mathbf{g}|\mathbf{n}) + \lambda \Psi(\mathbf{g}))$$

- Expectation-Maximization "family" algorithm.
  - ML estimator ($\lambda = 0$): [Vardi et al., 1985].
  - MAP (aka Bayesian) estimator: prior on $\mathbf{g} = \exp(-\lambda\Psi(\mathbf{g}))$, e.g. Gibbs field, see [Green, 1990].

Same thing in 4D

Additional set of basis functions

- Assume another finite set of temporal basis functions $B_1, \ldots, B_K$ (e.g. spline, *etc*) and set $\mathbf{g} = \{g_{11}, \ldots, g_{JK}\}$

$$G(x, t; \mathbf{g}) = \sum_{k=1}^{K} \sum_{j=1}^{J} g_{jk}\, \phi_j(x)\, B_k(t)$$

- Find optimal parameters from data $\boldsymbol{\tau} = \tau_{11}, \ldots, \tau_{1n_1}, \ldots, \tau_{In_I}$ with $\tau_{i1}, \ldots, \tau_{in_i} | \mathbf{g} \sim$Poisson process $\left(\sum_{i=1}^{I} p_{ij} \sum_{k=1}^{K} g_{jk} B_k(t)\right)$

$$\hat{\mathbf{g}} = \underset{\mathbf{g}>0}{\operatorname{argmin}}(-\log \mathcal{L}(\mathbf{g}|\boldsymbol{\tau}) + \lambda \Psi(\mathbf{g}))$$

- Expectation-Maximization “family” algorithm.
  - ML, MAP, penalized likelihood...
  - [Nichols et al., 2002, Reader et al., 2006].

Open questions.

- Choice of ϕ_j and B_k (e.g voxel size) ?
- How many basis functions J and K ?
- Do we trust that function of interest $G^\star(x,t)$ can be expressed as $G^\star(x,t) = G(x,t;\mathbf{g})$ for some $\widehat{\mathbf{g}}$
  - Do we trust in a digitized brain structure ?
  - Do Gibbs fields correspond to biological structures prior ?
- Can we give an interpretation to models with several millions (3D) or billions (4D) of parameters ?

Model selection

- Models have deep influence on inverse problem regularization.
- Models are almost never correct for real world data...
- Model selection and averaging are suitable to prevent over and under-fitting.

Open questions.

- Choice of ϕ_j and B_k (e.g voxel size) ?
- How many basis functions J and K ?
- Do we trust that function of interest $G^\star(x,t)$ can be expressed as $G^\star(x,t) = G(x,t;\mathbf{g})$ for some $\widehat{\mathbf{g}}$
  - Do we trust in a digitized brain structure ?
  - Do Gibbs fields correspond to biological structures prior ?
- Can we give an interpretation to models with several millions (3D) or billions (4D) of parameters ?

Model selection

- Models have deep influence on inverse problem regularization.
- Models are almost never correct for real world data...
- Model selection and averaging are suitable to prevent over and under-fitting.

Open questions.

- Choice of ϕ_j and B_k (e.g voxel size) ?
- How many basis functions J and K ?
- Do we trust that function of interest $G^\star(x,t)$ can be expressed as $G^\star(x,t) = G(x,t;\mathbf{g})$ for some $\widehat{\mathbf{g}}$
  - Do we trust in a digitized brain structure ?
  - Do Gibbs fields correspond to biological structures prior ?
- Can we give an interpretation to models with several millions (3D) or billions (4D) of parameters ?

Model selection

- Models have deep influence on inverse problem regularization.
- Models are almost never correct for real world data...
- Model selection and averaging are suitable to prevent over and under-fitting.

Nonparametric *vs.* parametric Models

Parametric models with infinitely many parameters...

Parametric models

- Characterized by a *fixed-size vector* of real-valued parameters.
- Basis functions (reconstruction grid) do not depend on data.

Nonparametric models.

- $\neq$ *no* parameter !
- The number of parameters can *grow unboundedly* with the dataset length.
- *Let the data choose* the appropriate complexity of the model.
- A model over *infinite dimensional* function or measure spaces.
- Side-step model selection and averaging.
- From discrete–discrete to discrete–continuous reconstruction.

Why Bayesian nonparametrics ?
First, why to be Bayesian...

Bayes' rule

$$\mathrm{P}\left(\Theta|\mathbf{Y}\right) = \frac{\mathrm{P}\left(\mathbf{Y}|\Theta\right)\mathrm{P}\left(\Theta\right)}{\int_{\mathcal{S}_\theta}\mathrm{P}\left(\mathbf{Y}|\theta\right)\mathrm{P}\left(\theta\right)\mathrm{d}\theta}$$

Prior knowledges.

- Statistical knowledges on objects : e.g. probability measure on $\mathrm{R}^3 \times \mathrm{R}^+$.
- Field specific knowledges : e.g. biological, physical.
- Explicit degree of belief in priors.

"Honest" estimation.

- Whole set of solutions via posterior distribution ($\neq$ MAP).
- $\rightarrow$ Posterior uncertainty e.g. highest probability density (HPD) interval of activity concentration for any ROI.

Contrast with parametric priors

- Priors on infinite-dimensional objects (here probability measure) $\rightarrow$ stochastic processes.
- Prior give insight to correlation structure (smoothness, *etc*).
  - Regularization
- Solutions set *dense* in infinite-dimensional spaces.

Difficulties

- How to elicit prior $\mathcal{G}$ for nonparametric $G(x, t)$?
- How to infer on infinite dimensional objects in real life ?

Nonparametric Bayesian Poisson inverse problem framework

$$G \sim \mathcal{G}$$
$$F(\cdot, \mathbf{t}) = \int_{\mathcal{X}} \mathcal{P}(\cdot|\mathbf{x}) \; G(\mathrm{d}\mathbf{x}, \mathbf{t}) \quad (1)$$
$$Y_i, T_i | F \overset{\text{iid}}{\sim} F, \;\text{for } i = 1, \ldots, n$$

- $G(\cdot)$: $\mathcal{G}$-distributed random probability measure (RPM), defined on $(\mathcal{X} \times \mathcal{T}, \sigma(\mathcal{X}) \otimes \sigma(\mathcal{T}))$.
- Objective: estimate the posterior distribution of $G(\cdot)$ from the observed F-distributed dataset $(\mathbf{Y}, \mathbf{T})' = \{(Y_1, T_1), \ldots, (Y_n, T_n)\}$.
- $\mathcal{P}(\cdot|\mathbf{x})$: given probability distribution, indexed by $\mathbf{x}$, defined on $(\mathcal{Y}, \sigma(\mathcal{Y}))$.

Emission Tomography context $\mathcal{X} \subseteq \mathrm{R}^3$, $\mathcal{T} \subseteq \mathrm{R}^+$.

- Y_i: index of the tube of response (TOR) and T_i: arrival time of the i^{th} observed event.
- Radon: $\mathcal{P}(\mathbf{y} = l|\mathbf{x}) \propto \delta(\langle \vec{\phi}_l, \mathbf{x} \rangle - u_l)$

Probability measure of annihilations events ("origins set")

- Define space-time clusters of annihilations events.
- Think about (overlapping) blobs whose number, different shapes and locations may be driven by data.
- See $G(\cdot)$ as the (nonparametric) probability distribution of clustered origins.
- E.g. voxels are replaced by data driven components.

Questions

- How to control (*regularize*) the number of components ?
- How to introduce annihilations events ?

Dirichlet process

The cornerstone of Bayesian nonparametrics [Ferguson, 1973]

Definition

- G_0 be a probability measure over $(\mathcal{X}, \mathcal{B})$ and $\alpha \in \mathrm{R}^{+\star}$.
- A *Dirichlet process* is the distribution of a random measure G over $(\mathcal{X}, \mathcal{B})$ s.t., for any finite partition $(B_1, \ldots, B_r)$ of $\mathcal{X}$,

$$(G(B_1), \ldots, G(B_r)) \sim \mathrm{Dir}(\alpha G_0(B_1), \ldots, \alpha G_0(B_r))$$

- G_0 is the mean distribution, α the concentration parameter.
- We write $G \sim \mathrm{DP}(\alpha, G_0)$.

Representations of Dirichlet processes

- Pólya urns (DP arises here as the De Finetti measure of the *exchangeable* sequence).
- Stick-breaking representation (constructive).
- Chinese restaurant (prior over partitions).

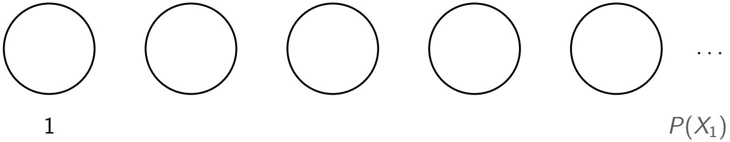


Figure: Assignment probability for customer 1.

- $\mathbf{X}_n = X_1, \dots, X_n$ take on $K < n$ distinct values $\theta_1, \dots, \theta_K$.
- This defines a partition of $\{1, \dots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n + 1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K + 1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

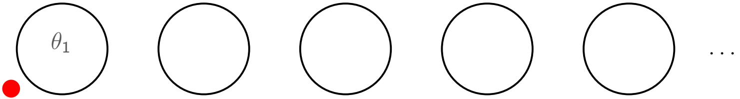


Figure: Table draw for customer 1.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n + 1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K + 1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

Chinese restaurant process

A worthy allegory for *partition prior* construction.

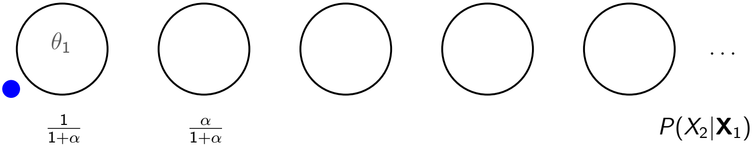


Figure: Assignment probability for customer 2.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

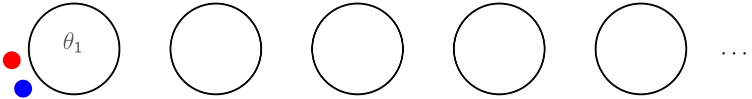


Figure: Table draw for customer 2.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

Chinese restaurant process

A worthy allegory for *partition prior* construction.

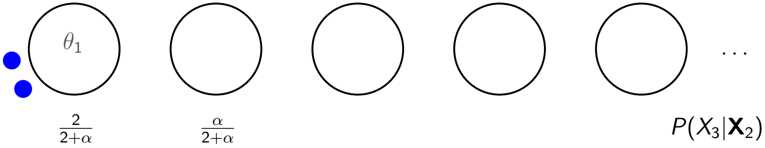


Figure: Assignment probability for customer 3.

- $\mathbf{X}_n = X_1, \dots, X_n$ take on $K < n$ distinct values $\theta_1, \dots, \theta_K$.
- This defines a partition of $\{1, \dots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

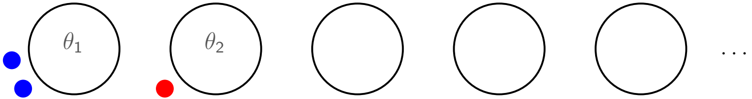


Figure: Table draw for customer 3.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n + 1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K + 1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

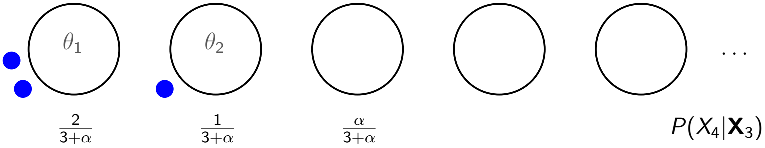


Figure: Assignment probability for customer 4.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n + 1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K + 1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

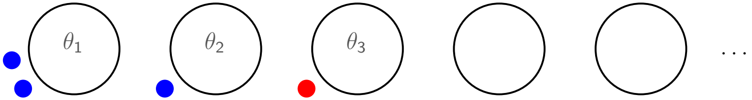


Figure: Table draw for customer 4.

- $\mathbf{X}_n = X_1, \dots, X_n$ take on $K < n$ distinct values $\theta_1, \dots, \theta_K$.
- This defines a partition of $\{1, \dots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

Chinese restaurant process

A worthy allegory for *partition prior* construction.

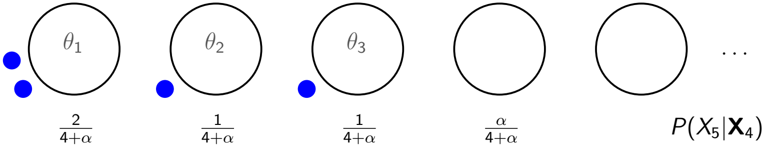


Figure: Assignment probability for customer 5.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

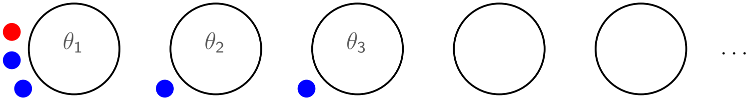


Figure: Table draw for customer 5.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

Chinese restaurant process

A worthy allegory for *partition prior* construction.

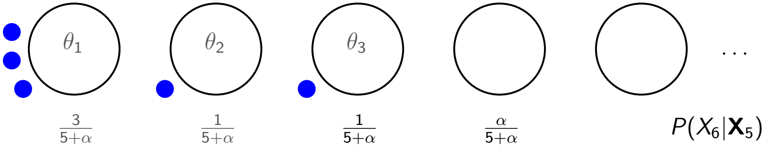
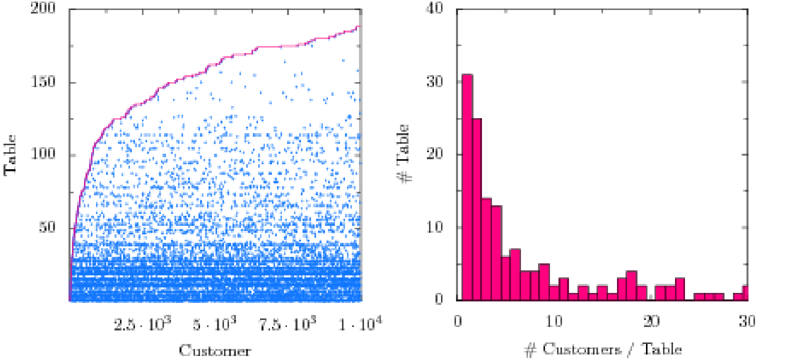


Figure: Assignment probability for customer 6.

- $\mathbf{X}_n = X_1, \ldots, X_n$ take on $K < n$ distinct values $\theta_1, \ldots, \theta_K$.
- This defines a partition of $\{1, \ldots, n\}$ into K clusters, s.t. i belongs to cluster k iff $X_i = \theta_k$.
- Sequentially generating from a CRP
  - First customer sits at table 1 and order $\theta_1 \sim G_0$.
  - Customer $n+1$ sits at:
    - Table k with probability $\frac{n_k}{n+\alpha}$ with n_k the number of customers at table k.
    - A new table $K+1$ with probability $\frac{\alpha}{n+\alpha}$ and order $\theta_{K+1} \sim G_0$

Chinese restaurant process

Clustering behaviour ($\alpha = 30$).



- The CRP exhibits the clustering property of the DP.
  - Expected number of clusters $K = O(\alpha \log n)$.
  - *Rich-gets-richer* effect $\rightarrow$ *Reinforcement* (small number of large clusters).
  - E.g.: Ewens sampling formula, species sampling.

Stick-breaking representation.

- $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \overset{\text{iid}}{\sim} G_0$
- $\mathbf{V} = (V_1, V_2, \ldots) \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha)$
- $\mathbf{p} = (p_1, p_2, \ldots)$, s.t. $p_1 = V_1$ and $p_k = V_k \prod_{i=1}^{k-1} (1 - V_i)$.
- Then,

$$G(\cdot) \triangleq \sum_{k=1}^{\infty} p_k \, \delta_{\theta_k}(\cdot)$$

  is a $\text{DP}(\alpha, G_0)$-distributed random probability distribution.
- We say that: $\mathbf{p} \sim \text{GEM}(\alpha)$.
- Almost sure truncation, [Ishwaran and James, 2001]: $\mathcal{P}_N(\cdot) = \sum_{k=1}^{N} p_k \, \delta_{\theta_k}(\cdot)$ with $V_N = 1$ converges a.s. to a $\text{DP}(\alpha G_0)$ random probability measure.

Stick-breaking representation
Example of construction

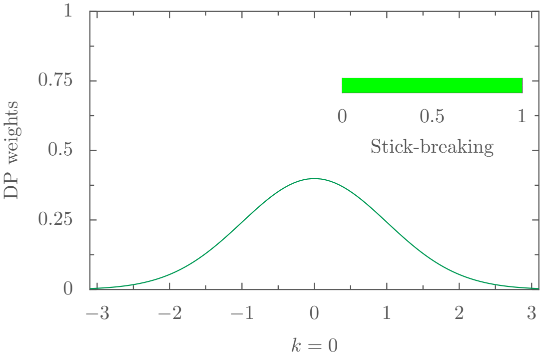


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0, 1)$).

Stick-breaking representation
Example of construction

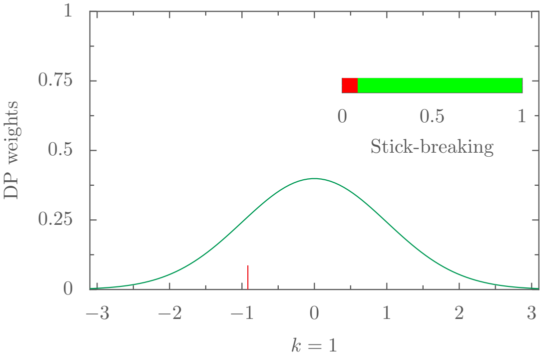


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0, 1)$).

Stick-breaking representation
Example of construction

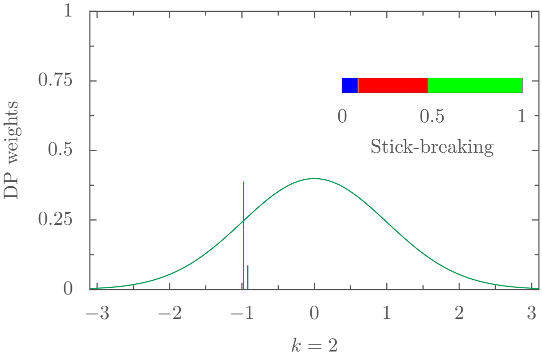


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0,1)$).

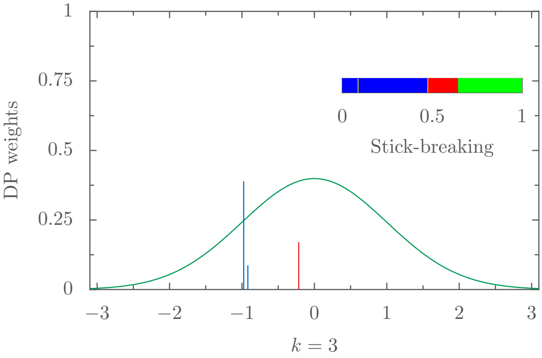


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0, 1)$).

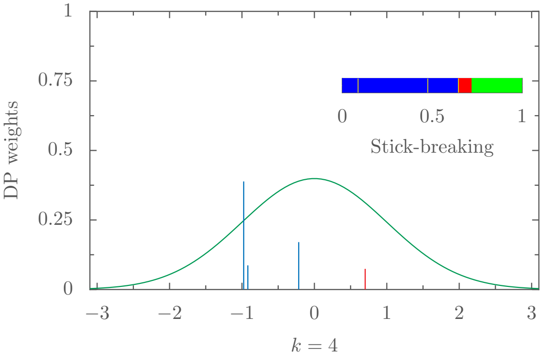


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0, 1)$).

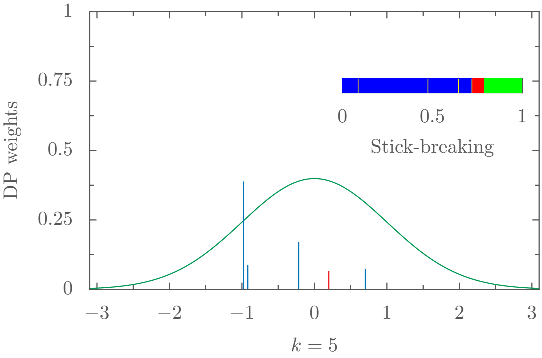


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0, 1)$).

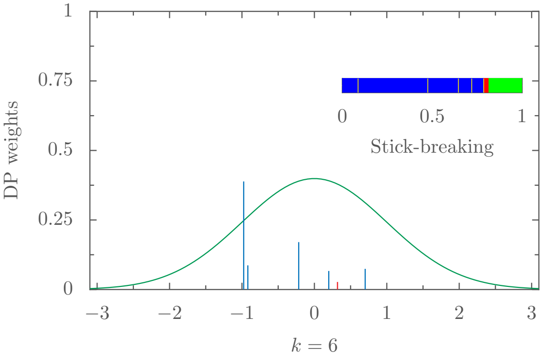


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0,1)$).

Stick-breaking representation
Example of construction

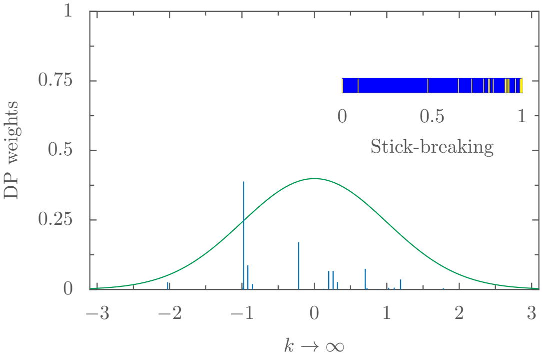


Figure: Dirichlet process GEM construction ($\alpha = 3,\ G_0 = \mathcal{N}(0,1)$).

Discreteness of DP(α, G_0) generated measures

- Cannot be used for probability density functions estimation !
- $\rightarrow$ Hierarchical mixture model with continuous distribution ϕ.

- Hierarchical data model

$$
\begin{aligned}
Y_i | X_i &\sim \phi(Y_i | X_i) \\
X_i &\sim H(\cdot) \\
H &\sim \mathrm{DP}(\alpha, G_0)
\end{aligned}
$$

- Data distribution

$$
y | H \sim \sum_{k=1}^{\infty} p_k \, \phi(y | \theta_k) = G(y)
$$

- E.g.: Dirichlet mixture of Normals with G_0 taken as Normal-Inverse Wishart, s.t. $\theta_k = (\mu_k, \Sigma_k)$.

Posterior sampling of DPM
Specific random schemes

- How to infer on infinite dimensional objects in a real world (and in a decent time) ?
- Sampling from the posterior: specific MCMC techniques.
  - Integrate out the random distribution: [Escobar, 1994], [Mac Eachern and Müller, 1998], [Neal, 2000].
    - side-step infiniteness by marginalization, only the allocation to *occupied* clusters (finite number) is sampled (Pólya Urn scheme).
    - Collapsing → good mixing properties.
    - Gives *only* access to sequences generated from the RPM.
  - Almost sure truncation: [Ishwaran and James, 2001].
    - Easy implementation.
  - Slice sampling: [Walker, 2007], [Kalli et al., 2011].
    - Conditional approach: inference retains whole distribution.
    - Use of auxiliary variables: only a finite pool of atoms are involved at each iteration, *without truncation*.
    - Gives access to posterior of any functional of the RPM (mean, variance, credible intervals, *etc.*).
- Variational techniques: [Blei and Jordan, 2006].

Definition

Let $E = \{0, 1\}$, $E^m = E \times \cdots \times E$ and $E^\star = \cup_{m=0}^{\infty} E^m$.
Let $\pi_m = \{B_\epsilon : \epsilon \in E^m\}$ be a partition of $\mathcal{T}$ and $\Pi = \cup_{m=0}^{\infty} \pi_m$.
A probability distribution Q on $\mathcal{T}$ has a Pólya tree distribution $\mathrm{PT}(\Pi, \mathcal{A})$ if there are nonnegative numbers $\mathcal{A} = \{\alpha_\epsilon : \epsilon \in E^\star\}$ and r.v. $\mathcal{W} = \{W_\epsilon : \epsilon \in E^\star\}$ s.t.

- $\mathcal{W}$ is a sequence of independent random variables,
- for all ϵ in $E^\star$, $W_\epsilon \sim \mathrm{Beta}(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$, and
- for all integer m and $\epsilon = \epsilon_1 \cdots \epsilon_m$ in E^m,

$$Q(B_{\epsilon_1 \cdots \epsilon_m}) = \prod_{\substack{j=1 \\ \epsilon_j = 0}}^{m} W_{\epsilon_1 \cdots \epsilon_{j-1}} \times \prod_{\substack{j=1 \\ \epsilon_j = 1}}^{m} (1 - W_{\epsilon_1 \cdots \epsilon_{j-1}})$$

Note that for $\epsilon \in E^\star$, $W_{\epsilon 0} = Q\left(B_{\epsilon 0} \middle| B_\epsilon\right)$

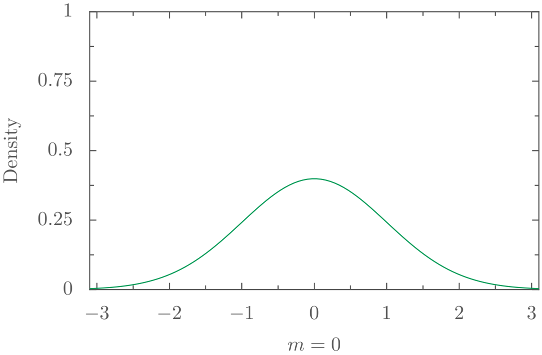


Figure: Pólya tree sequence construction (normal mean).

Pólya tree construction
Dyadic partition

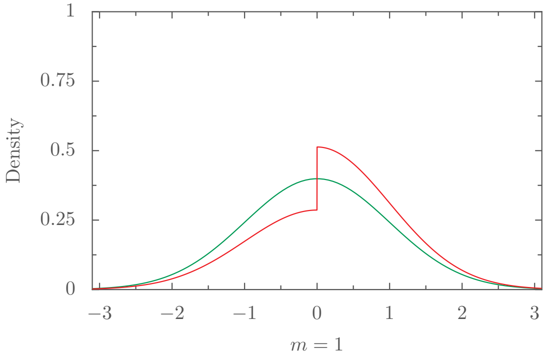


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

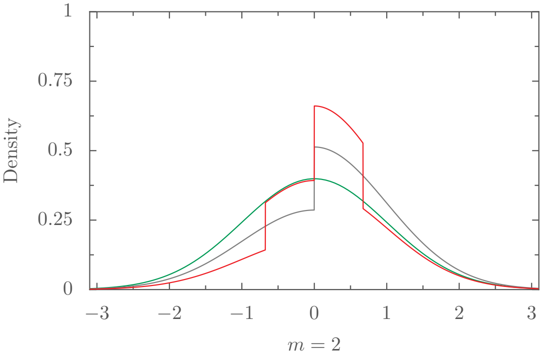


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

Pólya tree construction

Dyadic partition

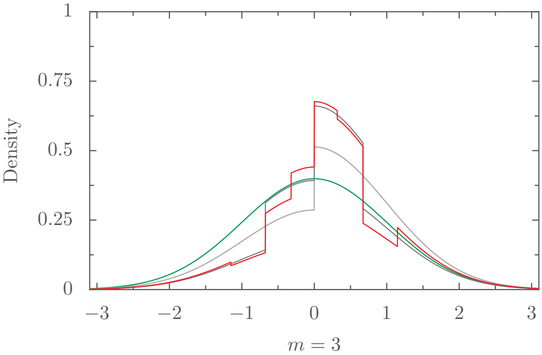


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

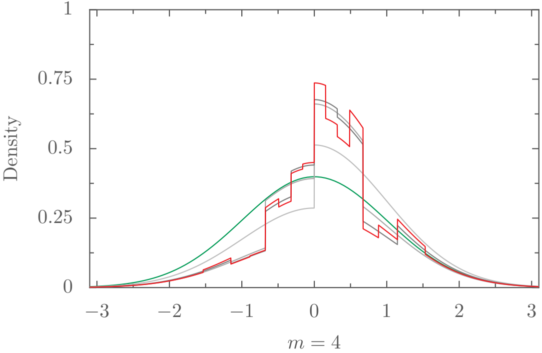


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

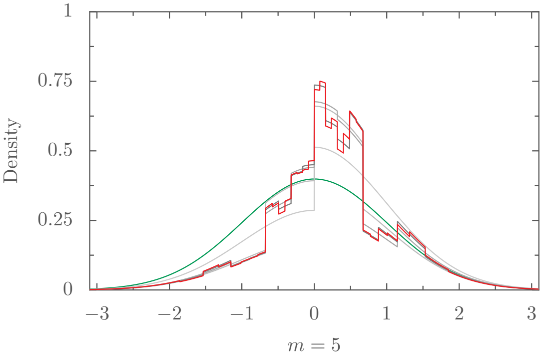


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

Pólya tree construction
Dyadic partition

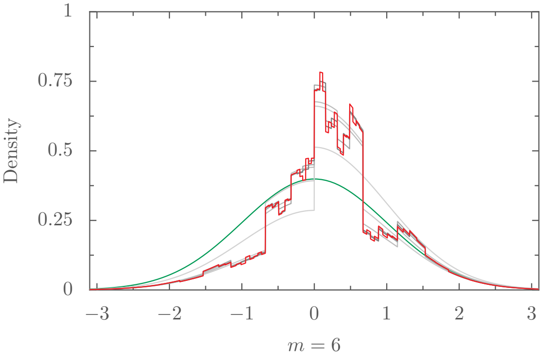


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

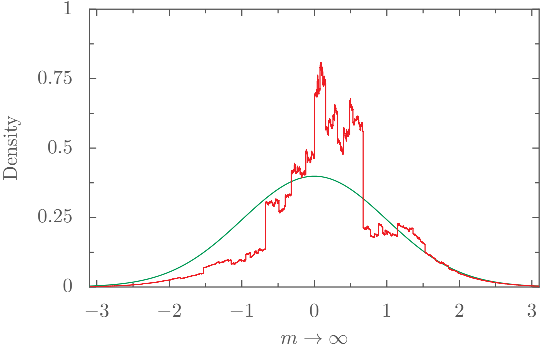


Figure: Pólya tree sequence construction ($\mathcal{A} = \{\alpha_m = 3^m\}$).

Spatial hierarchical model

$$\begin{aligned} Y_i|X_i &\overset{\text{ind}}{\sim} \mathcal{P}\left(Y_i|X_i\right) \\ X_i|Z_i &\overset{\text{ind}}{\sim} \mathcal{N}\left(X_i|Z_i\right) \\ Z_i|H &\overset{\text{iid}}{\sim} H \\ H &\sim \text{DP}\left(\alpha, \mathcal{NIW}\right) \end{aligned} \tag{2}$$

Remarks

- Tomography: Only Y_i is observed, thus X_i (the emission location) is introduced as latent variable (origin).
- In EM approach, latent variables are the number of emissions from voxel v which are recorded in line of response l.
- Compared to BNP density estimation, PET reconstruction mainly involves a sampling step from conditional $(X_i|Y_i, \mathbf{p}, \boldsymbol{\theta})$.
- Spatial distribution: $G(\cdot) = \int_\Theta \mathcal{N}(\cdot|\theta)H(d\theta) = \sum_{k=1}^{\infty} p_k \mathcal{N}(\cdot|\theta_k)$.

Inference by Gibbs sampling

Sampling from conditional distributions

Sampling from the posterior

- Let introduce $\mathbf{C} = C_1, C_2, \ldots, C_n$, the classification of emissions to DP components s.t. $Z_i = \theta_{C_i}$ for all $i < n$.
- Let $\mathbf{u} = u_1, u_2, \ldots, u_n$ uniform auxiliary variables.
- Successively draw samples from the following conditionals

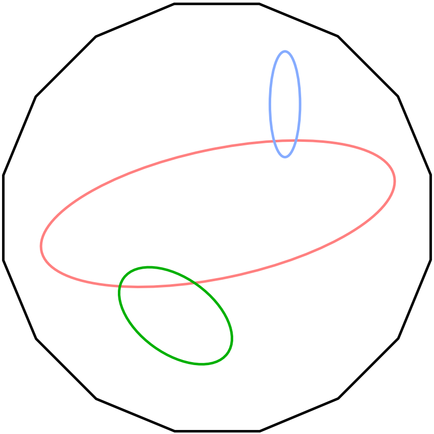
Annihilation location :	$(\mathbf{X}\mid\mathbf{Y}, \mathbf{p}, \boldsymbol{\theta}, \mathbf{u})$
DPM component parameters :	$(\boldsymbol{\theta}\mid\mathbf{C}, \mathbf{X})$
Emission allocation to DP atoms :	$(\mathbf{C}\mid\mathbf{p}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{u})$
DP weights & auxiliary variables :	$(\mathbf{p}, \mathbf{u}\mid\mathbf{C})$

Sampling $\mathbf{X}|\mathbf{Y}, \mathbf{p}, \boldsymbol{\theta}, \mathbf{u}$: Metropolis (independent MH) within Gibbs

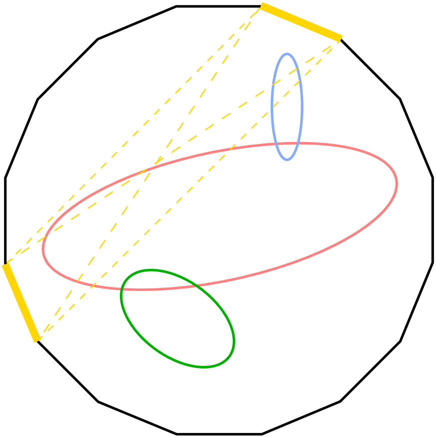
- $(X_i|Y_i, \mathbf{p}, \boldsymbol{\theta}, \mathbf{u}) \overset{\propto}{\sim} \mathcal{P}(Y_i|X_i)\, G(X_i|\mathbf{p}, \boldsymbol{\theta}, \mathbf{u})$
- $\mathcal{P}(Y_i|X_i)$ accounts for physical and geometrical properties of PET system $\rightarrow$ no hope for conjugacy...
- Candidate: $X_i^\star|Y_i, \mathbf{p}, \boldsymbol{\theta}, \mathbf{u} \overset{\propto}{\sim} \mathcal{N}(X_i^\star|\mu_{Y_i}, \Sigma_{Y_i})\, G(X_i^\star|\mathbf{p}, \boldsymbol{\theta}, \mathbf{u})$

Gibbs sampler in action

A toy system...



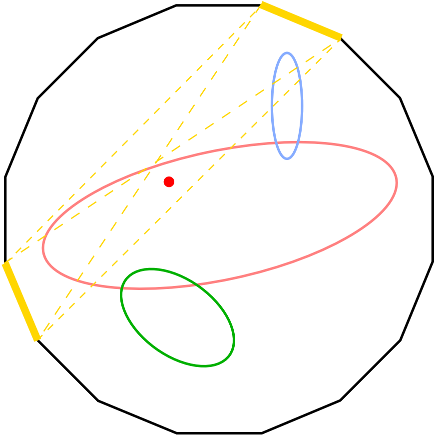
Iteration k, $(\mathbf{p}, \mathbf{u}|\mathbf{C})$, $(\theta|\mathbf{C}, \mathbf{X})$



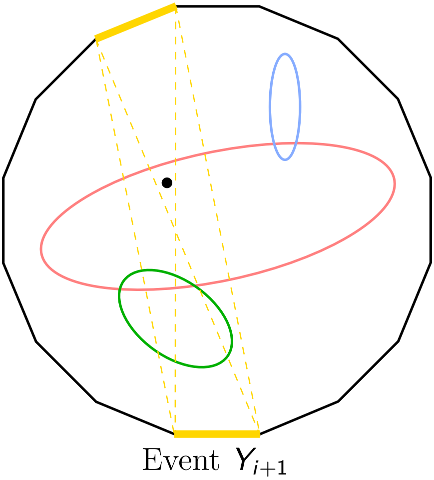
Event Y_i

Gibbs sampler in action

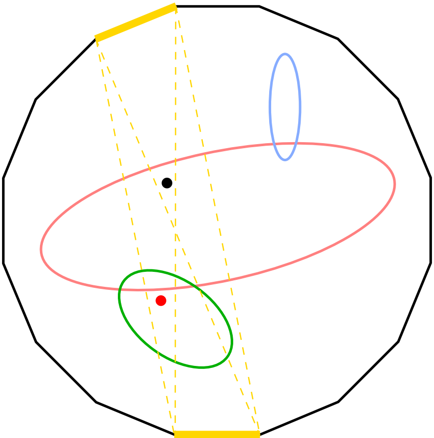
A toy system...



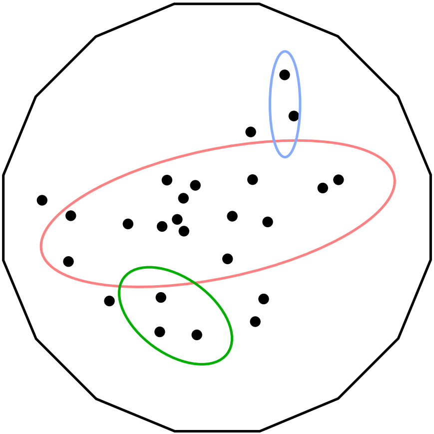
Back-projection $X_i | Y_i$, $\mathbf{p}$, $\mathbf{u}$, θ



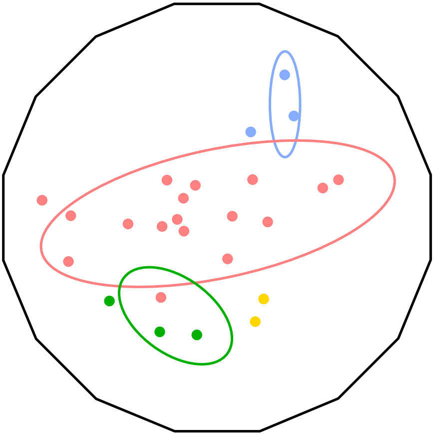
Event Y_{i+1}



Back-projection $X_{i+1} | Y_{i+1}, \mathbf{p}, \mathbf{u}, \boldsymbol{\theta}$

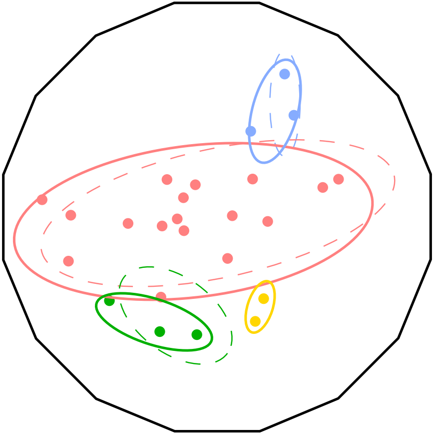


Back-projections $\mathbf{X}|\mathbf{Y}$, $\mathbf{p}$, $\mathbf{u}$, θ



Cluster allocations $\mathbf{C}|\theta, \mathbf{p}, \mathbf{u}, \mathbf{X}$

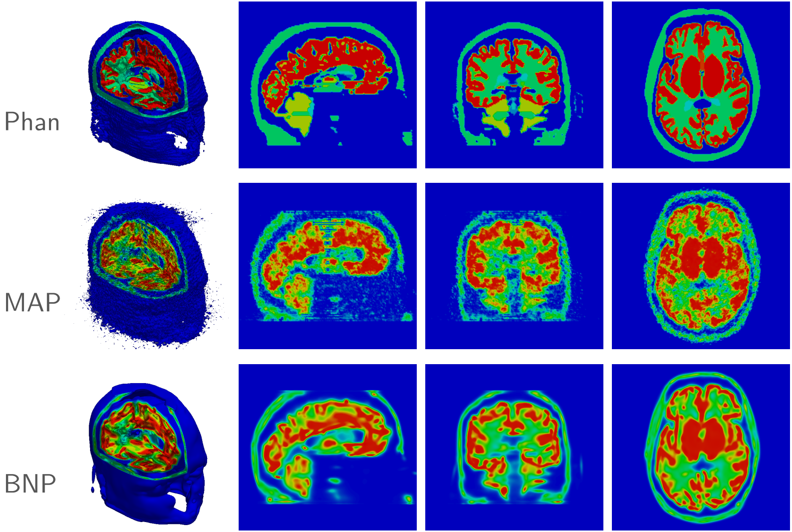
Gibbs sampler in action
A toy system...



Iteration $k+1$, $(\mathbf{p}, \mathbf{u}|\mathbf{C})$, $(\theta|\mathbf{C}, \mathbf{X})$

Synthetic data

True coincidences from a realistic digital phantom



Posterior uncertainty
Statistical coverage

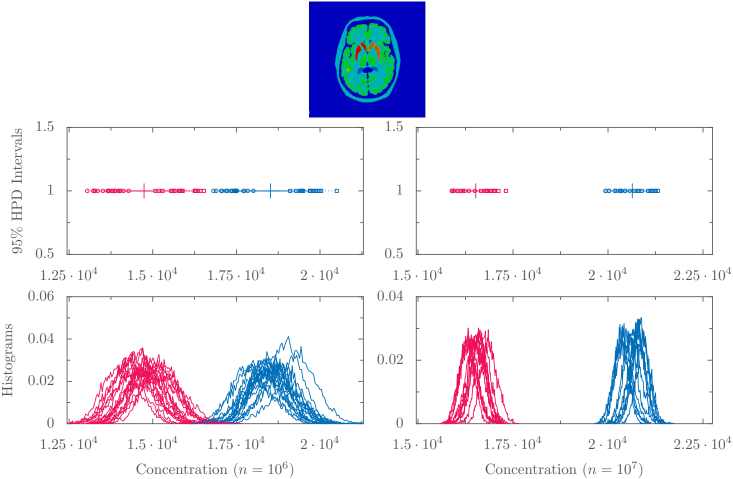


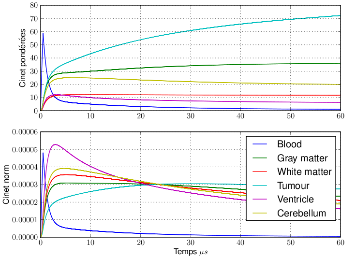
Figure: 20% decreased uptake in left putamen concentration (red) *vs.* right putamen (blue) for 20 replicas and 2 dataset lengths. Concentration = (total activity on volume V)/V

Dynamic PET Data

Tissue kinetics: time dependency

Modeling metabolic activity

- Biokinetic: tissue dependent.
- Functional volume (FV): spatial region characterized by a particular kinetic.
- Radioactive decay.



Separable space-time activity distribution

$$G(\mathbf{x}, \mathbf{t}) = \sum_{k=1}^{\infty} p_k \mathcal{N}(\mathbf{x}|\theta_k) \widetilde{Q}_k(\mathbf{t})$$

Kinetics RPM

- Each event Y_i is time stamped (T_i).
- Continuous measure with compact support (right truncation).

Space-time hierarchical model

$$
\begin{aligned}
Y_i|X_i &\overset{\text{ind}}{\sim} \mathcal{P}\left(Y_i|X_i\right) \\
X_i, T_i|Z_i, Q_i &\overset{\text{ind}}{\sim} \mathcal{N}\left(X_i|Z_i\right) \times Q_i\left(T_i\right) \\
Z_i, Q_i|H &\overset{\text{iid}}{\sim} H \\
H &\sim \text{DP}\left(\alpha, \mathcal{NIW} \times \mathcal{K}_0\right) \\
\mathcal{K}_0 &\sim \text{DP}\left(\beta, \text{PT}\left(\mathcal{A}, Q_0\right)\right)
\end{aligned}
$$

- With $H = \sum_{k=1}^{\infty} w_k \, \delta_{\theta_k, \widetilde{Q}_k}$, where $\widetilde{\mathbf{Q}}$ are i.i.d. $\mathcal{K}_0$
- $\mathcal{K}_0 = \sum_{j=1}^{\infty} \pi_j \, \delta_{Q_j^\star}$ with $\boldsymbol{\pi} \sim \text{GEM}(\beta)$, $\mathbf{Q}^\star$ are i.i.d. $\text{PT}(\mathcal{A}, Q_0)$, a Pólya tree with parameters $\mathcal{A}$ and mean Q_0.
- $\mathcal{K}_0$: DP process with PT process as base distribution $\rightarrow$ nested RPM (cf. nested DP, [Rodriguez et al., 2008])
- Distinct θ_k may share the same $Q_j^\star$ ($\mathcal{K}_0$ is discrete) $\rightarrow$ partial Hierarchical DP [Teh et al., 2006]; (diffuse $\mathcal{NIW} \times \mathcal{K}_0$).

Additional latent variables

- Allocation variable: $D_k = j$ iff $\widetilde{Q}_k = Q_j^\star$ (kinetics clustering).
- Auxiliary variables $\mathbf{v}$ for slice sampling of $\mathcal{K}_0$.

Posterior computations

- Gibbs sampling of additional conditionals is straightforward.

Functional volumes distribution

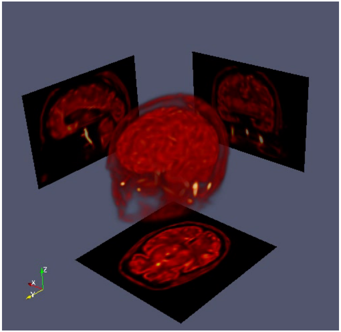
- For all j (label of $\mathcal{K}_0$ atoms),

$$\mathrm{FV}_j(\mathbf{x}) = \sum_{k:\,\widetilde{Q}_k = Q_j^\star} p_k \,\mathcal{N}(\mathbf{x}|\theta_k)$$

4D synthetic data
Application and results

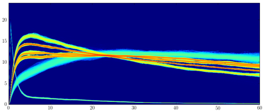
Data generation (FDG tracer)

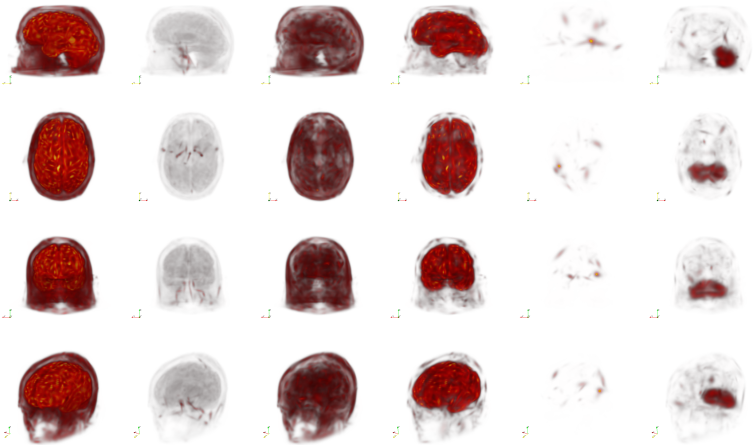
- 4 FV : gray matter, white matter, cerebellum, tumors.
- Blood pool and blood fraction in tisssues (5% to 10%).
- $n = 10^7$ events ($\approx \frac{1}{10}$ usual dose for 4D PET).
- Same 3D phantom.



4D activity

4D synthetic data
Estimated kinetics and functional volumes





4D clinical data
Comparison with ML-EM + post-smoothing

Cardiac data

- Biograph scanner, bone metabolism tracer, evaluation of potential interest in cardiology.
- BNP : dose ÷30; EM+smoothing : dose ÷1
- $\Delta T = 11$ min, sharp kinetic during first minute.

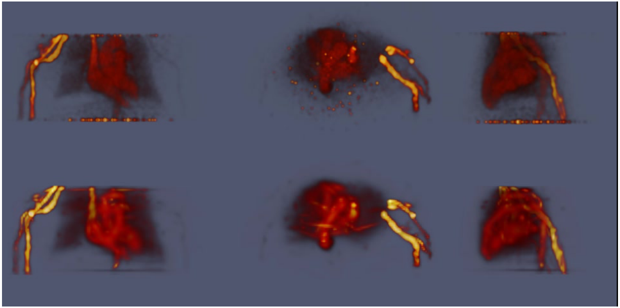


Figure: up: EM (÷1); bottom: BNP (÷30) 4D BNP vs EM

Conclusion
and perspectives.

Some observations...

- Suitable nonparametric framework for discrete-continous 3D/4D PET.
- Flexible data modeling : hierarchical, dependent, *etc.*
- From *low* level data to *high* level structures in a unified framework.
- Access to posterior intervals (uncertainty).
- Sampling schemes have to be carefully considered.

...and perspectives

- Prior refinements.
- Computing architecture consistent with MCMC approach.

For Further Reading. I

Blei, D. M. and Jordan, M. I. (2006).
Variational inference for dirichlet process mixtures.
Bayesian Analysis, 1(1):121–144.

Escobar, M. D. (1994).
Estimating normal means with a dirichlet process prior.
J. Am. Stat. Assoc., 89:268–277.

Ferguson, T. S. (1973).
A bayesian analysis of some nonparametric problems.
Ann. Statist., 1:209–230.

Ghosh, J. K. and Ramamoorthi, R. V. (2003).
Bayesian Nonparametrics.
Springer.

Green, P. J. (1990).
Bayesian reconstructions from emission tomography data using a modified EM algorithm.
IEEE Trans. Med. Imaging, 9(1):84–93.

Hjort, N. L., Holmes, C., Müller, P., Walker, S. G., Ghosal, S., Lijoi, A., Prünster, I., Teh, Y. W., Jordan, M. I., Griffin, J., Dunson, D. B., and Quintana, F. (2010).
Bayesian Nonparametrics.
Cambridge Series in Statistical and Probabilistic Mathematics.

For Further Reading. II

Ishwaran, H. and James, L. F. (2001).
Gibbs sampling methods for stick-breaking priors.
J. Am. Stat. Assoc., 96:161–173.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011).
Slice sampling mixture models.
Statistics and Computing, 21(1):93–105.

Mac Eachern, S. N. and Müller, P. (1998).
Estimating mixture of dirichlet process models.
J. Comput. Graph. Stat., 7:223–238.

Neal, R. M. (2000).
Markov chain sampling methods for dirichlet process mixture models.
J. Comput. Graph. Stat., 9(2):249–265.

Nichols, T. E., Qi, J., Asma, E., Member, S., Leahy, R. M., and Member, S. (2002).
Spatiotemporal reconstruction of list-mode pet data.
IEEE Trans. Med. Imag, 21:396–404.

Pitman, J. (2006).
Combinatorial stochastic processes.
Technical Report 621, Dept. Statistics, U.C. Berkeley.

For Further Reading. III

Reader, A., Sureau, F., Comtat, C., Trébossen, R., and Buvat, I. (2006).
Joint estimation of dynamic PET images and temporal basis functions using fully 4D ML-EM.
Physics in Medicine and Biology, 51(21):5455–5474.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008).
The nested dirichlet process.
J. Am. Stat. Assoc., 103:1131–1154.

Sethuraman, J. (1994).
A constructive definition of Dirichlet priors.
Stat. Sinica, 4:639–650.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006).
Hierarchical Dirichlet processes.
J. Am. Stat. Assoc., 101(476):1566–1581.

Vardi, Y., Shepp, L., and Kaufman, L. (1985).
A statistical model for Positron Emission Tomography.
J. Am. Stat. Assoc., 80:8–20.

Walker, S. G. (2007).
Sampling the Dirichlet mixture model with slices.
Comm. Statist., 36:45–54.