Nested nonparametric processes

Federico Camerlenghi

September 6–10, 2022

University of Milano – Bicocca & Collegio Carlo Alberto

PAPERS AND CO-AUTHORS

We focus on the papers:

- CAMERLENGHI F., DUNSON D.B., LIJOI A., PRÜNSTER I., RODRIGUEZ A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Analysis*, **14**, 1303–1356.
- DENTI F., CAMERLENGHI F., GUINDANI M., MIRA A. (2022). A Common Atoms Model for the Bayesian Nonparametric Analysis of Nested Data. *Journal of the American Statistical Association*, to appear.

OUTLINE

INTRODUCTION

EXCHANGEABILITY

- Analogy or symmetry between observations justifies induction, i.e. the prediction of future outcomes of an experiment.
- Exchangeability is the simplest form of analogy across data: a sequence of observations $\{X_n\}_{n\geq 1}$ is exchangeable iff

$$(X_1, \cdots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \cdots, X_{\sigma(n)})$$

for every $n \geq 1$ and every permutation σ of $\{1, \ldots, n\}$.

In many applications exchangeability is too restrictive since data are affected by some sort of heterogeneity (e.g. time-dependent data, related experiments, covariate-indexed observations). Indeed (de Finetti, 1938) writes:

> *But the case of exchangeability can only be considered as a limiting case: the case in which this "analogy" is, in a certain sense, absolute for all events under consideration. [..] To get from the case of exchangeability to other cases which are more general but still tractable, we must take up the case where we still encounter "analogies" among the events under consideration, but without attaining the limiting case of exchangeability.*

PARTIAL EXCHANGEABILITY

Partial exchangeability is a more appropriate assumption in presence of heterogeneous data: data are considered exchangeable within the same group and conditional independent across different groups.

The sequences $\{(X_{i,j})_{j\geq 1}:\ i=1,2\}$ are partially exchangeable ($d=2$) iff

$$(X_{1,1},\cdots,X_{1,n_1},X_{2,1},\cdots,X_{2,n_2}) \stackrel{d}{=} (X_{1,\sigma(1)},\cdots,X_{1,\sigma(n_1)},X_{2,\pi(1)},\cdots,X_{2,\pi(n_2)})$$

for every $n_1,n_2\geq 1$ and every permutation σ and π of $\{1,\ldots,n_1\}$ and $\{1,\ldots,n_2\}$.

DE FINETTI'S REPRESENTATION THEOREM

The sequences $\{(X_{i,j})_{j\geq 1}:\ i=1,2\}$ are partially exchangeable iff there exists a vector of dependent random probability measures $(\tilde{p}_1,\tilde{p}_2)$ such that:

$$(X_{1,j_1},X_{2,j_2})|\tilde{p}_1,\tilde{p}_2 \stackrel{\text{iid}}{\sim} \tilde{p}_1\times\tilde{p}_2$$
$$(\tilde{p}_1,\tilde{p}_2)\sim Q.$$

The distribution Q is known as the de Finetti measure of the sequence.

DEPENDENT NONPARAMETRIC PRIORS

Several Bayesian nonparametric models have been proposed to accommodate for heterogeneity:

- additive structures: (Müller, Quintana & Rosner; 2004), (Lijoi, Nipoti & Prünster; 2014), (C, Lijoi, Nipoti & Prünster; 2022+);
- hierarchical structures: (Teh, Jordan, Beal & Blei; 2006) , (C, Lijoi, Orbanz & Prünster; 2019), (Colombi, Argiento, C & Paci; 2022+);
- nested structures: (Rodriguez, Dunson & Gelfand; 2008), (C, Dunson, Lijoi, Prünster & Rodriguez; 2019), (Denti, C, Guindani & Mira; 2022);
- other contributions, see (Quintana et al.; 2022) for a complete review.

Problems arising in presence of partially exchangeable observations:

- theoretical properties and clustering structures are usually complex to derive and to deal with;
- develop efficient and fast marginal or conditional algorithms for complex problems.

The nested Dirichlet process

The nested structure of (Rodriguez, Dunson & Gelfand; 2008) is as follows:

$$(X_{1,j_1}, X_{2,j_2}) | \tilde{p}_1, \tilde{p}_2 \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2$$
$$(\tilde{p}_1, \tilde{p}_2) | \tilde{q} \sim \tilde{q}^2$$

where $\tilde{q}$ is a random probability measure on the space $\mathrm{P}_{\mathbb{X}}$ (space of all random probability measures on $\mathbb{X}$), i.e.

$$\tilde{q} = \sum_{i \geq 1} \omega_i \delta_{G_i}, \quad G_i = \sum_{\ell \geq 1} w_{\ell,i} \delta_{\theta_{\ell,i}}$$

and $\theta_{\ell,i} \overset{\text{iid}}{\sim} Q_0$, for a non–atomic probability measure Q_0 on $\mathbb{X}$.

Issues with Nested Structures

If the two samples $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ share at least one value, then $\tilde{p}_1 = \tilde{p}_2$ almost surely.

1. This degeneracy property holds true for general nested processes based on Completely Random Measures;
2. There are alternative models to overcome the drawback: Latent Nested Processes and Common Atoms Model (CAM).

NESTED PROCESSES

Completely Random Measures (CRMs)

Let $\tilde{\mu}$ be a random measure on the space $\mathbb{X}$, then its law is characterized by the Laplace functional

$$L_{\tilde{\mu}}(f) := \mathbb{E}[e^{-\int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)}],$$

defined for each measurable function $f : \mathbb{X} \to \mathbb{R}^+$.

Completely Random Measures (CRMs)

$\tilde{\mu}$ is termed a completely random measure iff the random variables $\tilde{\mu}(A_1), \dots, \tilde{\mu}(A_k)$ are independent for any choice of disjoint Borel sets $A_1, \dots, A_k \in \mathscr{X}$ and for any $k \geq 1$.

We concentrate on CRMs with both random jumps and random atoms:

$$\tilde{\mu}(\,\cdot\,) = \sum_{i=1}^{\infty} J_i \delta_{Z_i}(\,\cdot\,), \text{ with Laplace functional } L_{\tilde{\mu}}(f) = e^{-\int_{\mathbb{X}\times\mathbb{R}^+}(1-e^{-sf(x)})\nu(\mathrm{d}x,\mathrm{d}s)}$$

for any measurable function $f : \mathbb{X} \to \mathbb{R}^+$. ν is termed the intensity measure and it uniquely characterizes $\tilde{\mu}$.

Priors based on completely random measures

Normalized completely random measures

Let $\tilde{\mu}$ be a CRM on $(\mathbb{X}, \mathscr{X})$ such that $\mathbb{P}(0 < \tilde{\mu}(\mathbb{X}) < \infty) = 1$, then the random probability measure

$$\tilde{p}(\,\cdot\,) = \frac{\tilde{\mu}(\,\cdot\,)}{\tilde{\mu}(\mathbb{X})}$$

is termed a Normalized Random Measure with Independent increments (NRMI). See (Regazzini, Lijoi & Prünster; 2003).

The NRMI $\tilde{p}$ is characterized by the intensity measure ν of the associated CRM $\tilde{\mu}$, noteworthy examples are:

- if $\tilde{\mu}$ is a gamma CRM, i.e. $\nu(\mathrm{d}x, \mathrm{d}s) = e^{-s}/s c \mathrm{d}s P_0(\mathrm{d}x)$, the associated NRMI $\tilde{p}$ is a Dirichlet process, denoted as $\mathscr{D}(cP_0)$;
- if $\tilde{\mu}$ is a σ–stable CRM, i.e. $\nu(\mathrm{d}x, \mathrm{d}s) = \sigma s^{-1-\sigma}/\Gamma(1-\sigma)\mathrm{d}s P_0(\mathrm{d}x)$, the associated NRMI $\tilde{p}$ is a σ–stable process, denoted as σ–stb(P_0);

We will focus on homogeneous NRMIs, i.e. whose associated CRM $\tilde{\mu}$ is homogeneous having intensity $\nu(\mathrm{d}x, \mathrm{d}s) = \rho(s)\mathrm{d}s\, cP_0(\mathrm{d}x)$, writing $\tilde{p} \sim \mathrm{NRMI}(\rho, c; P_0)$.

NESTED MODELS BASED ON NRMIs

$$(X_{1,j_1}, X_{2,j_2})|\tilde{p}_1, \tilde{p}_2 \overset{\text{ind}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \quad (j_1, j_2) \in \mathbb{N} \times \mathbb{N}$$

$$\tilde{p}_1, \tilde{p}_2 | \tilde{q} \overset{\text{iid}}{\sim} \tilde{q}, \qquad \tilde{q} \overset{d}{=} \frac{\tilde{\mu}}{\tilde{\mu}(\mathrm{P}_{\mathbb{X}})} \left(= \sum_{i=1}^{\infty} \frac{J_i}{\sum_{h \geq 1} J_h} \delta_{\tilde{q}_{0,i}}(\,\cdot\,) \right)$$

where:

- $\tilde{\mu}$ is a CRM on $(\mathrm{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ with Lévy intensity $\nu(\mathrm{d}p, \mathrm{d}s) = c\,\rho(s)\mathrm{d}sQ(\mathrm{d}p)$;
- Q is a probability measure on $(\mathrm{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ which equals the distribution of a NRMI:

$$Q(\,\cdot\,) = \mathbb{P}(\tilde{q}_0 \in \,\cdot\,), \quad \text{and } \tilde{q}_0 \sim \mathrm{NRMI}(\rho_0, c_0; Q_0).$$

Remarks:

- the model extends the Nested Dirichlet Process (Rodriguez, Dunson & Gelfand; 2008) to nested NRMIs;
- $\tilde{p}_1$ and $\tilde{p}_2$ are exchangeable and, since $\tilde{q}$ is almost surely discrete, one has

$$\pi_1 := \mathbb{P}(\tilde{p}_1 = \tilde{p}_2) > 0$$

Partition structure

- Consider $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ two samples from a partially exchangeable array of observations having size size n_1 and n_2, respectively;
- $(\tilde{p}_1, \tilde{p}_2)$ are two nested random probability measures as defined before.

Mixed moments

$$\mathbb{E} \int_{\mathsf{P}_{\mathbb{X}}^2} f_1(p_1) f_2(p_2) \tilde{q}(\mathrm{d}p_1) \tilde{q}(\mathrm{d}p_2)$$
$$= \pi_1 \int_{\mathsf{P}_{\mathbb{X}}} f_1(p) f_2(p) Q(\mathrm{d}p) + (1 - \pi_1) \int_{\mathsf{P}_{\mathbb{X}}} f_1(p) Q(\mathrm{d}p) \int_{\mathsf{P}_{\mathbb{X}}} f_2(p) Q(\mathrm{d}p),$$

for every measurable functions $f_1, f_2 : \mathsf{P}_{\mathbb{X}} \to \mathbb{R}^+$.

The moments are a convex combination of the full exchangeable situation and independence across samples.

Ties across samples

Let X_{1,j_1} (resp. X_{2,j_2}) be an observation from the first (resp. second) sample, then

$$\mathbb{P}(X_{1,j_1} = X_{2,j_2}) > 0.$$

The observations $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ may be partitioned into $k = k_0 + k_1 + k_2$ clusters according to the following scheme:

- k_1 distinct values are specific to $\boldsymbol{X}_1$, having frequencies $\boldsymbol{n}_1 = (n_{1,1}, \cdots, n_{1,k_1})$;
- k_2 distinct values are specific to $\boldsymbol{X}_2$, having frequencies $\boldsymbol{n}_2 = (n_{2,1}, \cdots, n_{2,k_2})$;
- k_0 distinct values are shared by the two samples, having frequencies $\boldsymbol{q}_1 = (q_{1,1}, \cdots, q_{1,k_0})$ and $\boldsymbol{q}_2 = (q_{2,1}, \cdots, q_{2,k_0})$.

The probability of having a specific partition of the two samples in k clusters is termed partially Exchangeable Partition Probability Function (pEPPF).

Partially exchangeable partition probability function

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1 \Phi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1 + \boldsymbol{q}_2) + (1 - \pi_1)\Phi_{k_1}^{(n_1)}(\boldsymbol{n}_1)\Phi_{k_2}^{(n_2)}(\boldsymbol{n}_2)\mathbb{1}_{\{0\}}(k_0).$$

- $\Phi_k^{(n)}$: situation of full exchangeability
- $\Phi_{k_1}^{(n_1)}\Phi_{k_2}^{(n_2)}$: product of two EPPFs in a situation of unconditional independence across samples.
- Whenever $k_0 \neq 0$ the model reduces to a situation of full exchangeability, being $\mathbb{P}(\tilde{p}_1 = \tilde{p}_2 | \boldsymbol{X}_1, \boldsymbol{X}_2) = 1$: this is too restrictive!

Application: density estimation

Model for density estimation

Data have been generated by two random dependent densities $\tilde{f}_i = \int_\Theta h(x;\theta)\tilde{p}_i(\mathrm{d}\theta)$, for $i = 1, 2$:

$$(X_{1,j_1}, X_{2,j_2})|(\theta_{1,j_1}, \theta_{2,j_2}) \sim h(\,\cdot\,;\theta_{1,j_1}) \times h(\,\cdot\,;\theta_{2,j_2})$$

$$(\theta_{1,j_1}, \theta_{2,j_2})|\tilde{p}_1, \tilde{p}_2 \overset{\mathrm{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2$$

being:

- $h(\cdot;\theta)$, where $\theta = (M, V) \in \mathbb{R} \times \mathbb{R}^+$, is a Gaussian kernel on $\mathbb{R}$ with mean M and variance V;
- $(\tilde{p}_1, \tilde{p}_2)$ is a nested process.

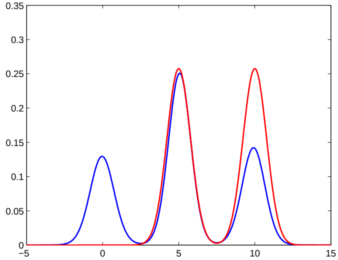
- Estimation of random dependent densities is carried out through an MCMC procedure based on the pEPPF;

TRUE AND ESTIMATED DENSITIES

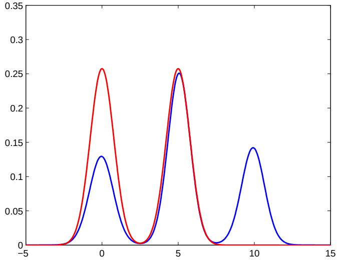
The $n_1 = n_2 = 100$ data $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have been generated from:

$$X_1 \sim \frac{1}{2}\mathrm{N}(5, 0.6) + \frac{1}{2}\mathrm{N}(10, 0.6), \qquad X_2 \sim \frac{1}{2}\mathrm{N}(5, 0.6) + \frac{1}{2}\mathrm{N}(0, 0.6).$$

(a) Density: first group



(b) Density: second group



The presence of a common component $\mathrm{N}(5, 0.6)$ forces the equality of the two random probability measures

$\Downarrow$

The two estimated densities are the same.

THE COMMON ATOMS MODEL

The CAM

The Common Atoms Model (CAM) introduced by (Denti, C, Guindani & Mira; 2022) is:

$$(X_{1,j_1}, X_{2,j_2}) | \tilde{p}_1, \tilde{p}_2 \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2$$
$$(\tilde{p}_1, \tilde{p}_2) | \tilde{q} \sim \tilde{q}^2$$

where $\tilde{q}$ is a random probability measure on the space $\mathrm{P}_{\mathbb{X}}$ defined as

$$\tilde{q} = \sum_{i \geq 1} \omega_i \delta_{G_i}, \quad G_i = \sum_{\ell \geq 1} w_{\ell,i} \delta_{\theta_\ell}.$$

- the atoms $\theta_1, \theta_2, \ldots$ are shared across the random probability measures G_i's, and $\theta_\ell \overset{\text{iid}}{\sim} Q_0$, for a non–atomic probability measure Q_0;
- the sequence of weights $(\omega_i)_{i \geq 1}$ has a GEM distribution, i.e. we consider $V_i \overset{\text{iid}}{\sim} \mathrm{Beta}(1, \alpha)$ and

$$\omega_1 = V_1, \quad \omega_i = V_i \prod_{r=1}^{i-1} (1 - V_r), i > 1,$$

  we will write $(\omega_i)_{i \geq 1} \sim \mathrm{GEM}(\alpha)$, being $\alpha > 0$;
- the sequences $(w_{\ell,i})_{\ell \geq 1}$ are i.i.d. with distribution $\mathrm{GEM}(\beta)$, being $\beta > 0$.

CAM: pEPPF

Consider two samples $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of size n_1 and n_2, respectively, and suppose they induce a partition into $k = k_0 + k_1 + k_2$ groups:

- k_1 distinct values are specific to $\boldsymbol{X}_1$, having frequencies $\boldsymbol{n}_1 = (n_{1,1}, \cdots, n_{1,k_1})$;
- k_2 distinct values are specific to $\boldsymbol{X}_2$, having frequencies $\boldsymbol{n}_2 = (n_{2,1}, \cdots, n_{2,k_2})$;
- k_0 distinct values are shared by the two samples, having frequencies $\boldsymbol{q}_1 = (q_{1,1}, \cdots, q_{1,k_0})$ and $\boldsymbol{q}_2 = (q_{2,1}, \cdots, q_{2,k_0})$.

pEPPF: CAM

Under the CAM, the pEPPF equals:

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1 \Phi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1 + \boldsymbol{q}_2) + (1 - \pi_1) I(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)$$

where $\pi_1 = \mathbb{P}(\tilde{p}_1 = \tilde{p}_2)$ and

$$I(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \int_{\mathbb{X}^{k_0+k_1+k_2}} \mathbb{E} \prod_{i=1}^{2} \prod_{j=1}^{k_i} G_i^{n_{i,j}}(\mathrm{d}x_{i,j}^*) \prod_{j=1}^{k_0} G_i^{q_{i,j}}(\mathrm{d}z_j^*)$$

We concentrate on the term

$$I(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \int_{\mathbb{X}^{k_0+k_1+k_2}} \mathbb{E} \prod_{i=1}^{2} \prod_{j=1}^{k_i} G_i^{n_{i,j}}(\mathrm{d}x_{i,j}^*) \prod_{j=1}^{k_0} G_i^{q_{i,j}}(\mathrm{d}z_j^*)$$

where the expected value is made w.r.t.

$$G_i = \sum_{\ell \geq 1} w_{\ell,i} \delta_{\theta_\ell}$$

and

- $\theta_1, \theta_2, \ldots \overset{\text{iid}}{\sim} Q_0$: common atoms;
- $(w_{\ell,i})_{\ell \geq 1} \sim \mathrm{GEM}(\beta)$: weights.

THEOREM

If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ share $k_0 > 0$ distinct values, one has

$$I(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) > 0$$

Then, the CAM does not reduce to the full exchangeable model in presence of common observations across samples.

Clustering structure

The CAM induces ties at the distributional level and observational level:

- ties among distributions are possible in view of the discreteness of $\tilde{q}$, indeed:

$$\mathbb{P}(\tilde{p}_1 = \tilde{p}_2) = \frac{1}{1+\alpha};$$

- ties across samples $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are possible with probability

$$\mathbb{P}(X_{1,j_1} = X_{2,j_2}) = \frac{1}{\alpha+1}\left[\frac{1}{1+\beta} + \alpha\frac{1}{2\beta+1}\right]$$

Thus, the CAM allows for a two-fold clustering structure:

- distributional clustering;
- observational clustering, allowing for borrowing of information across layers.

Dependence across $\tilde{p}_1$ and $\tilde{p}_2$

Covariance and correlation

- For any measurable sets A, B, the covariance equals

$$\mathrm{Cov}(\tilde{p}_1(A), \tilde{p}_2(B)) = \left(\frac{\pi_1}{1+\beta} + \frac{1-\pi_1}{1+2\beta}\right)(Q_0(A \cap B) - Q_0(A)Q_0(B))$$

  where $\pi_1 = 1/(\alpha+1)$.

- The correlation on the same set A equals

$$\rho_{1,2} := \mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = 1 - \frac{\beta}{2\beta+1} \cdot \frac{\alpha}{\alpha+1}$$

The correlation $\rho_{1,2}$:

- does not depend on the set A, it can be considered a measure of dependence across $\tilde{p}_1$ and $\tilde{p}_2$;
- lies in the interval $(1/2, 1)$, this is useful in genomics, where the experimental units are quite similar.

CAM: EXTENSIONS AND APPLICATIONS

- CAM may be easily extended to the case of $d > 2$ groups of observations:

$$\tilde{p}_1, \ldots, \tilde{p}_d | \tilde{q} \sim \tilde{q}$$

  and all the previous theoretical results can be extended to this setting;

- CAM can be used to model continuous distributions by considering a nonparametric mixture

$$(X_{1,j_1}, \ldots, X_{d,j_d}) | (\tilde{f}_1, \ldots, \tilde{f}_d) \sim \tilde{f}_1 \times \cdots \times \tilde{f}_d$$
$$\tilde{f}_i(\,\cdot\,) = \int_\Theta h(\,\cdot\,; \theta) \tilde{p}_i(\mathrm{d}\theta) \quad i = 1, \ldots, d$$

- CAM can be adapted to count data, where in group $i \in \{1, \ldots, d\}$ one observes the vector of counts

$$\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,n_i}) \in \mathbb{N}^{n_i}$$

  According to (Canale & Dunson; 2011), we embed the CAM in a rounded mixture of Gaussian framework.

Common Atoms Mixture Models

Model for density estimation

Data have been generated by random dependent densities $\tilde{f}_i = \int_\Theta h(x;\theta)\tilde{p}_i(\mathrm{d}\theta)$, for $i = 1, \ldots, d$:

$$(X_{1,j_1}, \ldots, X_{d,j_d})|(\theta_{1,j_1}, \ldots, \theta_{d,j_d}) \sim h(\,\cdot\,;\theta_{1,j_1}) \times \cdots \times h(\,\cdot\,;\theta_{d,j_d})$$

$$(\theta_{1,j_1}, \ldots, \theta_{d,j_d})|\tilde{p}_1, \ldots, \tilde{p}_d \overset{\text{iid}}{\sim} \tilde{p}_1 \times \cdots \times \tilde{p}_d$$

being:

- $h(\cdot;\theta)$, where $\theta = (M, V) \in \mathbb{R} \times \mathbb{R}^+$, is a Gaussian kernel with mean M and variance V;
- $(\tilde{p}_1, \cdots, \tilde{p}_d)$ is a CAM.

Posterior inference is carried out by implementing

- a truncated version of the Blocked-Gibbs sampler (Ishwaran & James; 2001);
- a slice-efficient sampler, extending the work of (Kalli et al.; 2011).

A SIMULATION STUDY

We consider the following scenario:

- $d = 12$ groups (or units) of observations;
- we sample two units from the following six different distributions

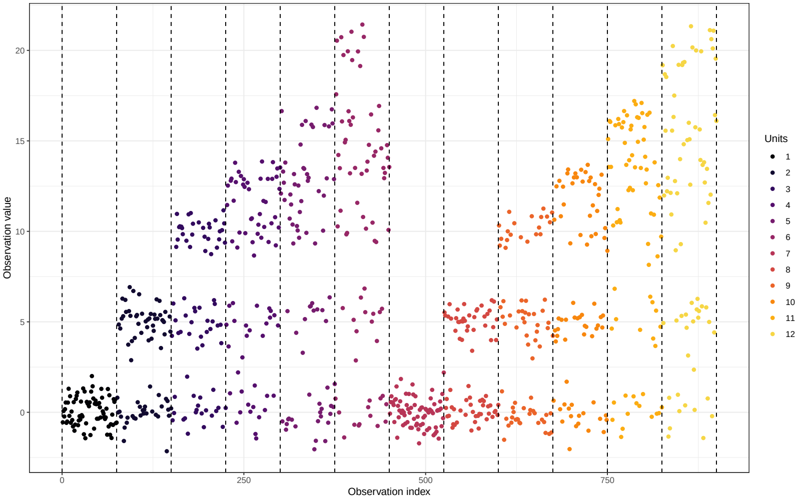
$$X_h \sim \sum_{\ell=1}^{h} \frac{1}{h} \mathrm{N}(m_h, 0.6), \quad h = 1, \ldots, 6$$

  and $(m_1, \ldots, m_6) = (0, 5, 10, 13, 16, 20)$ is the vector of means;
- all the units have the same cardinality $n_i = 75$.

Note that:

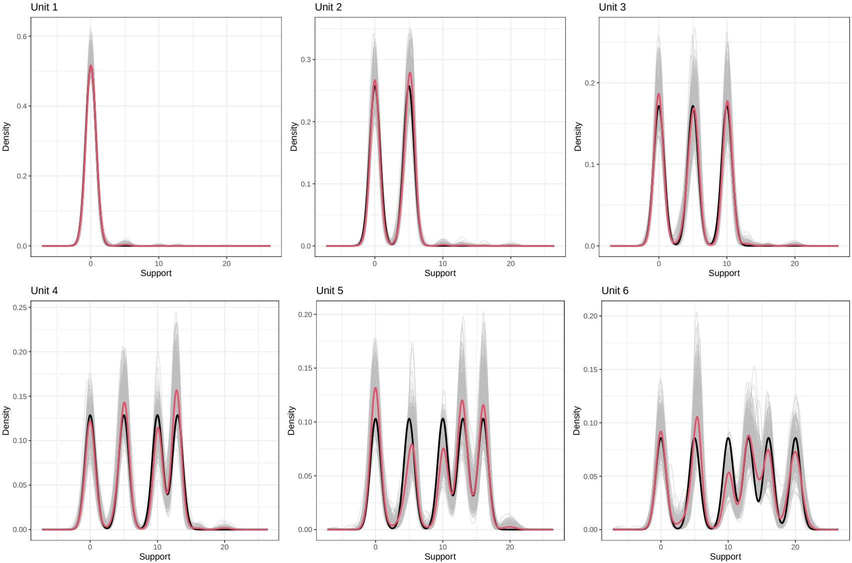
- data are generated from 6 different distributions;
- the mixture components are shared across groups, and their true number is 6.

Dataset

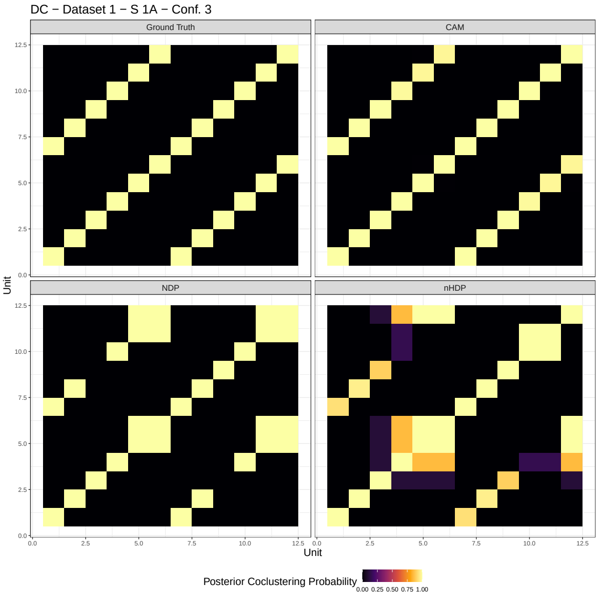


Observation value
20
15
10
5
0
0
250
500
750
Observation index
Units
1
2
3
4
5
6
7
8
9
10
11
12

True vs Estimated densities



DISTRIBUTIONAL CLUSTERING



CAM FOR MICROBIOME STUDIES

In microbiome studies, one typically deals with count data:

- d is the number of subjects in the study;
- for subject $i \in \{1, \dots, d\}$, one observe the counts of a microbial sequence

$$\boldsymbol{Z}_i = (Z_{i,1}, \dots, Z_{i,n_i}) \in \mathbb{N}^{n_i};$$

- $Z_{i,j}$ is referred to as the frequency of the jth OTU (operational taxonomic unit) in subject i.

CAM FOR COUNT DATA

For each data point $Z_{i,j}$, let us introduce a latent variable $X_{i,j}$ and assume that:

$$\mathbb{P}(Z_{i,j} = q | X_{i,j}) = \mathbb{1}_{[a_q, a_{q+1})}(X_{i,j}), \quad q \in \mathbb{N}$$

where

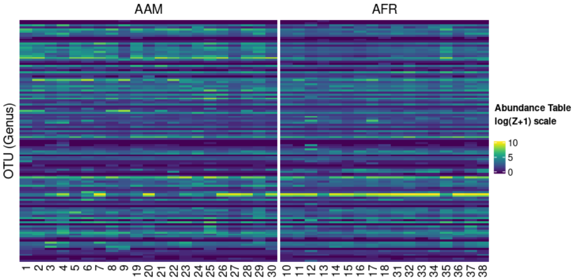
- $a_0 < a_1 < \cdots < a_\infty$ is a sequence of threshold values on the real line;
- the $X_{i,j}$'s are modelled as a CAM mixture.

See also (Canale & Dunson; 2011).

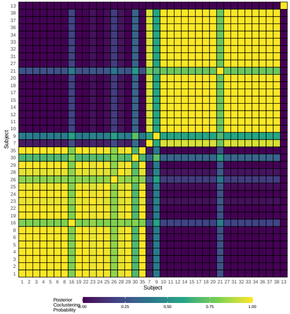
DATASET

We consider the dataset of (O'Keefe et al.; 2015):

- fecal samples of $d = 38$ subjects;
- $n_i = 119$ taxa measured for each subject;
- OTUs refer to middle-aged African Americans (AA) and rural Africans (AF).



Distributional clustering



Cluster	DC-1	DC-2	DC-3
Cardinality	18	19	1
Africans	2	14	1
Americans	16	5	0
Female	11	6	0
Male	7	13	1

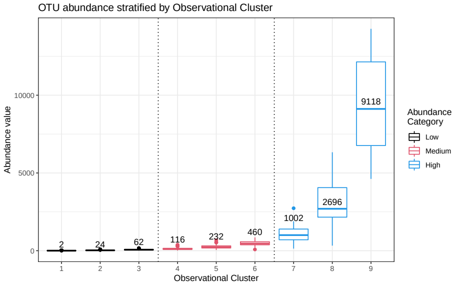
Remarks:

- the optimal partition is estimated by the approach of (Wade & Ghahramani; 2018), based on the minimization of the Variation of Information;
- the different subgroups of AA and AF are captured by the CAM, DC-3 contains only one subject with a unique microbiome distribution.

Observational clustering

As for observational clustering, we recognize 9 clusters:

- they represent intensities of the latent process underlying the counts;
- they are grouped in three macro clusters representing the abundance classes (low, medium and high);



CONCLUSIONS & FUTURE WORKS

CONCLUSIONS AND FUTURE WORKS

Conclusions:

- the Nested Dirichlet Process suffers from a degeneracy in the general setting of normalized CRMs;
- the CAM solves the degeneracy problem, allowing for both distributional and observational clustering;
- efficient algorithms have been developed for CAM.

Open problems:

- the pEPPF for the CAM is not available in closed form;
- extend the CAM to other classes of processes, e.g., Compound Random Measures (Griffin & Leisen; 2017);
- nested processes for feature sampling models, which are designed for count measurements.

COMPOUND RANDOM MEASURES (CORMS)

Compound Random Measures have been introduced by (Griffin & Leisen; 2017).
A vector $(\mu_1, \ldots, \mu_d)$ is a vector of CoRMs iff

$$\mu_j|\eta = \sum_{i\geq 1} m_{j,i} J_i \delta_{\theta_i}$$

where:

- $(m_{1,i}, \ldots, m_{d,i}) \overset{\text{iid}}{\sim} h$, and h is a score distribution;
- $\eta = \sum_{i\geq 1} J_i \delta_{\theta_i}$ is a a completely random measure with Lévy measure ν^*.

Some remarks:

- all the random measures $\mu_1, \ldots, \mu_d$ share the same atoms;
- CoRMs are analytically tractable;
- the use of CoRMs in nested structures improve flexibility.

Nested processes based on CoRMs

Nested processes based on CoRMs can be formally defined as follows:

$$(X_{1,j_1}, X_{2,j_2})|\tilde{p}_1, \tilde{p}_2 \stackrel{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2, \quad (\tilde{p}_1, \tilde{p}_2) = \left(\frac{\tilde{\mu}_1}{\tilde{\mu}_1(\mathbb{X})}, \frac{\tilde{\mu}_2}{\tilde{\mu}_2(\mathbb{X})}\right)$$

$$\tilde{\mu}_1, \tilde{\mu}_2|\tilde{q} \stackrel{\text{iid}}{\sim} \tilde{q}.$$

Here $\tilde{q}$ is a random probability measure on $\mathrm{M}_{\mathbb{X}}$ (space of all measures on $\mathbb{X}$), namely

$$\tilde{q} = \sum_{i \geq 1} \omega_i \delta_{\mu_i}$$

and

- $(\omega_i)_{i \geq 1}$ is a vector of weights summing up to 1;
- the μ_i's are CoRMs according to (Griffin & Leisen; 2017).

Ongoing project with R. Corradin and A. Ongaro.

References (I)

- Camerlenghi F., Dunson D.B., Lijoi A., Prünster I., Rodriguez A. (2019). Latent nested nonparametric priors. *Bayesian Anal.*, **14**, 1303–1356 (with discussion).
- Camerlenghi F., Lijoi A., Orbanz P., and Prünster I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.* **47**, 67–92.
- Camerlenghi F., Lijoi A., Nipoti B., Prünster I. (2022+). Posterior analysis for dependent normalized random measures. *In preparation.*
- Canale A., Dunson D.B. (2011). Bayesian kernel mixtures for counts. *J. Amer. Statist. Assoc.*, **106**, 1529–1539.
- Denti F., Camerlenghi F., Guindani M., Mira A. (2022). A Common Atom Model for the Bayesian Nonparametric Analysis of Nested Data. *J. Amer. Statist. Assoc.*, to appear.
- Escobar M.D. and West M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Stat. Assoc.* **90**, 577–588.
- Griffin J.E. and Leisen F. (2017). Compound random measures and their use in Bayesian nonparametrics. *Journal of the Royal Statistical Society-Series B*, **79**, 525–545.
- Ishwaran H. and James L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**, 161–173.
- Kalli M., Griffin J. E. and Walker S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21**, 93–105.
- Kingman J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.
- Lijoi A., Nipoti B. and Prünster I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–1291.

REFERENCES (II)

- MÜLLER P., QUINTANA F., and ROSNER G. (2004). A method for combining inference across related nonparametric Bayesian models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 735–749.
- O'KEEFE ET AL. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. *Nature Communications*, **6**.
- QUINTANA F.A., MÜLLER P., JARA A., MACEACHERN S.N. (2022). The Dependent Dirichlet Process and Related Models. *Statistical Science*, **37**, 24–41.
- REGAZZINI E., LIJOI A. and PRÜNSTER I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585.
- RODRIGUEZ A., DUNSON D.B. and GELFAND A.E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.*, **103,** 1131–1154. *Bayesian Anal.* **14**, 161–180.
- TEH Y. W., JORDAN M. I., BEAL M. J. and BLEI D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 1566–1581.
- WADE S. and GHAHRAMANI Z. (2018). Bayesian cluster analysis: point estimation and credible balls (with discussion). *Bayesian Analysis*, **13**, 559–626.