

Dependent discrete random structures in Bayesian Nonparametrics

Igor Prünster

Bocconi University

STATISTICS CONFERENCE
IN HONOR OF AAD VAN DER VAART'S 60TH BIRTHDAY

Leiden
June 17-21, 2019

Outline

EXCHANGEABILITY & PARTIAL EXCHANGEABILITY

MEASURING DEPENDENCE

DEPENDENT PRIORS VIA RANDOM MEASURES

HIERARCHICAL NRMIS

PARTITION STRUCTURE & NUMBER OF CLUSTERS

CONCLUDING REMARKS

Exchangeability

- Probabilistic statement concerning homogeneity (symmetry, analogy) among observations justifies induction

⇒ Inferences are not affected by the order of the observations

- Suitable mathematical framework for inference alternative to the classical approach with “fixed and unknown probability distribution P ”

⇒ A sequence of observations $(X_n)_{n \geq 1}$ is **exchangeable** if for any $n \geq 1$ and permutation π of $(1, \dots, n)$

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)})$$

- **de Finetti's representation theorem:** $(X_n)_{n \geq 1}$ is **exchangeable** if and only

$$\mathbb{P}[X_1 \in A_1, \dots, X_n \in A_n] = \int_{\mathcal{P}} \prod_{i=1}^n P(A_i) Q(dP)$$

where $\mathcal{P}$ is the space of probability measures on $\mathbb{X}$.

⇒ Q is the **de Finetti measure** of $(X_n)_{n \geq 1}$ and acts as a **prior distribution** for Bayesian inference being the law of a random probability measure $\tilde{P}$.

Equivalently one can state the representation theorem in hierarchical form as

$$\begin{array}{ccc} X_i | \tilde{P} & \stackrel{\text{iid}}{\sim} & \tilde{P} \\ \tilde{P} & \sim & Q \end{array} \quad i = 1, \dots, n$$

Beyond exchangeability

From de Finetti (1938):

*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**.*

In applications **dependence structures** more general than **exchangeability** are required. We focus on data collected under **different experimental conditions** s.t.

- ▶ **Homogeneity within** each experimental condition
- ▶ **Heterogeneity across** different experimental conditions

Examples: Topic modeling, Meta-Analysis, two-sample problems, nonparametric regression (covariate-indexed data), time dependent data, change-point problems ...

Partial exchangeability

The array $(\mathbf{X}_1, \mathbf{X}_2) = (X_{1,i}, X_{2,j})_{i,j \geq 1}$ is *partially exchangeable* if

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

for any $n_1, n_2 \geq 1$ and any permutations π and ϕ of $(1, \dots, n_1)$ and $(1, \dots, n_2)$.

de Finetti's representation theorem

$(\mathbf{X}_1, \mathbf{X}_2)$ is *partially exchangeable* if and only if

$$\begin{aligned} \mathbb{P}[X_{1,1} \in A_1, \dots, X_{1,n_1} \in A_{n_1}, X_{2,1} \in B_1, \dots, X_{2,n_2} \in B_{n_2}] \\ = \int_{\mathcal{P}^2} \prod_{i=1}^{n_1} P_1(A_i) \prod_{j=1}^{n_2} P_2(B_j) Q(dP_1, dP_2). \end{aligned}$$

$\Rightarrow$ This is the same as saying that

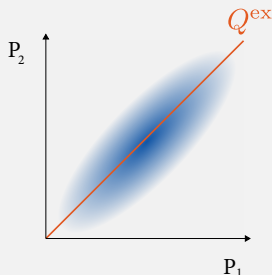
$$\begin{array}{ccc} (X_{1,i}, X_{2,j}) & \overset{\text{iid}}{\sim} & \tilde{P}_1 \times \tilde{P}_2 \\ (\tilde{P}_1, \tilde{P}_2) & \sim & Q \end{array} \quad \forall i, j \geq 1$$

with $(\tilde{P}_1, \tilde{P}_2)$ a vector of dependent random probability measures and Q the *prior*.

Measuring dependence

Extreme cases of dependence induced by the prior Q

- **Maximal dependence** $\iff$ *full exchangeability*
i.e. Q is **degenerate** on the diagonal $\{(P_1, P_2) \in \mathcal{P}^2 : P_1 = P_2\}$,
namely $\tilde{P}_1 = \tilde{P}_2$ (a.s.)



- **Independence** $\iff$ $\tilde{P}_1$ and $\tilde{P}_2$ are (unconditionally) **independent** with respect to Q .
 $\implies$ corresponds to maximal **heterogeneity**: inference on each sample is not influenced by the observations from the other sample.

Correlation as measure of dependence

Correlation

The most popular measure of dependence is **correlation**: since

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)),$$

typically does not depend on A it is taken as a **measure of overall dependence**.

Extreme cases correspond to:

- ▶ **perfect linearity**, which is implied by $\tilde{P}_1 = \tilde{P}_2$ a.s.
- ▶ **uncorrelation**, which is implied by $\tilde{P}_1 \perp\!\!\!\perp \tilde{P}_2$

$\implies$ **good proxy** for the desired measure of dependence!

Examples:

- ▶ Hierarchical Dirichlet process (HDP)

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = \frac{c + 1}{c + 1 + c_0}$$

- ▶ Hierarchical normalized stable process (HnstP)

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = \frac{1 - \sigma_0}{1 - \sigma\sigma_0}$$

Dependent priors via transformed random measures

Most popular approach to the definition of dependent nonparametric priors is **via dependent stick-breaking constructions** introduced by MacEachern (1999, 2000)

- ▶ **Pros:** “simple” implementation of “conditional” algorithms;
- ▶ **Cons:** it is **almost impossible to derive analytic expressions** for quantities of interest both marginal and conditional.

Approach based on **general random measures** with 3 possible strategies (and combinations thereof) for creating dependence:

1. Hierarchical structures

First proposed by Teh, Jordan, Beal & Blei (2006) for the Dirichlet process with stick-breaking representation: Hierarchical Dirichlet process (HDP)
 $\implies$ here defined for **general processes** and studied as **transformed random measures**

2. Additive structures

First proposed by Müller, Quintana & Rosner (2004) for the Dirichlet process with stick-breaking representation. For general normalized random measures theory developed in Lijoi, Nipoti and P. (2014).

3. Nested structures

First proposed by Rodriguez, Dunson & Gelfand (2008) for the Dirichlet process with stick-breaking representation. Development of nested structures for general random measures and generalizations to incorporate an additive layer in Camerlenghi, Dunson, Lijoi, P. and Rodriguez (2019).

Completely random measures

Completely random measures (Kingman, 1967)

A random element $\tilde{\mu}$ taking values in the space of boundedly finite measures $\mathbb{M}$ such that, for any $d \geq 1$ and collection of pairwise disjoint sets $A_1, \dots, A_d$, the random variables

$\tilde{\mu}(A_1), \tilde{\mu}(A_2), \dots, \tilde{\mu}(A_d)$ are mutually independent

is said to be a **completely random measure** (CRM).

Key properties

Assume $\tilde{\mu}$ has no fixed points of discontinuity

- ▶ The **realizations** of a CRM are a.s. **discrete** i.e. $\tilde{\mu}(\cdot) = \sum_{i=1}^{\infty} J_i \delta_{Z_i}(\cdot)$
- ▶ A CRM $\tilde{\mu}$ is uniquely characterized by its *Laplace functional*

$$\mathbb{E} \left[e^{-\int_{\mathbb{X}} g(x) \tilde{\mu}(dx)} \right] = e^{-\int_{\mathbb{R}^+ \times \mathbb{X}} [1 - e^{-v g(x)}] \nu(dv, dx)}$$

with ν indicating the **Lévy intensity**, which **characterizes the CRM** $\tilde{\mu}$.

CRM-based nonparametric priors

Normalized completely random measures (Regazzini, Lijoi and P., 2003)

Let $\tilde{\mu}$ be a CRM on $\mathbb{X}$ such that $0 < \tilde{\mu}(\mathbb{X}) < \infty$ a.s. Then

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

is a *normalized completely random measure* (NRMI=Normalized Random Measure with Independent increments).

In the following we will consider a.s. finite homogeneous CRMs i.e.

$\nu(d\nu, dx) = \rho(d\nu)cP(dx)$ and write $\tilde{\mu} \sim \text{CRM}(\rho, c, P)$ and $\tilde{P} \sim \text{NRMI}(\rho, c, P)$.

Pitman–Yor process (Pitman & Yor, 1997)

A Pitman–Yor process with parameters $\sigma \in (0, 1)$ and $\theta > 0$ can be defined via normalization as

$$\tilde{P} = \frac{\tilde{\mu}_{\sigma, \theta}}{\tilde{\mu}_{\sigma, \theta}(\mathbb{X})} \sim \text{PY}(\sigma, \theta; P)$$

where $\tilde{\mu}_{\sigma, \theta}$ is a suitable transformation of a specific CRM (but not a CRM).

An alternative measure of dependence

- For $(\tilde{P}_1, \tilde{P}_2) = (\tilde{\mu}_1/\mu_1(\mathbb{X}), \tilde{\mu}_2/\mu_2(\mathbb{X}))$, maximal dependence corresponds to **full exchangeability**, i.e. $\tilde{P}_1 = \tilde{P}_2$ a.s. $\iff \tilde{\mu}_1 = c\tilde{\mu}_2$ a.s., and denote such a vector by $(\tilde{\mu}_1^{\text{ex}}, \tilde{\mu}_2^{\text{ex}})$
- Idea: measure **dependence of $(\tilde{P}_1, \tilde{P}_2)$** as **distance of $(\tilde{\mu}_1, \tilde{\mu}_2)$ from $(\tilde{\mu}_1^{\text{ex}}, \tilde{\mu}_2^{\text{ex}})$** .

Distance and measure of dependence

Assuming $\tilde{\mu}_1 \stackrel{d}{=} \tilde{\mu}_2$ and some mild technical conditions

$$d_{\mathcal{W}}\left(\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}, \begin{pmatrix} \tilde{\mu}_1^{\text{ex}} \\ \tilde{\mu}_2^{\text{ex}} \end{pmatrix}\right) = \sup_{A \in \mathcal{X}} \mathcal{W}\left(\begin{pmatrix} \tilde{\mu}_1(A) \\ \tilde{\mu}_2(A) \end{pmatrix}, \begin{pmatrix} \tilde{\mu}_1^{\text{ex}}(A) \\ \tilde{\mu}_2^{\text{ex}}(A) \end{pmatrix}\right)$$

is a well-defined distance, with $\mathcal{W}$ the **Wasserstein distance** of order 2. Moreover, a measure of dependence is given by

$$D(\tilde{P}_1, \tilde{P}_2) = 1 - \frac{d_{\mathcal{W}}\left(\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}, \begin{pmatrix} \tilde{\mu}_1^{\text{ex}} \\ \tilde{\mu}_2^{\text{ex}} \end{pmatrix}\right)}{d_{\mathcal{W}}\left(\begin{pmatrix} \tilde{\mu}_1^{\text{ind}} \\ \tilde{\mu}_2^{\text{ind}} \end{pmatrix}, \begin{pmatrix} \tilde{\mu}_1^{\text{ex}} \\ \tilde{\mu}_2^{\text{ex}} \end{pmatrix}\right)}$$

with $(\tilde{\mu}_1^{\text{ind}}, \tilde{\mu}_2^{\text{ind}})$ corresponding to $\tilde{\mu}_1 \perp\!\!\!\perp \tilde{\mu}_2$.

Goals

1. Express d_W in terms of the underlying Lévy intensities and obtain bounds (via compound Poisson approximation)
2. Evaluate d_W for several popular BNP models $\implies$ quantification of heuristic notion of closeness to exchangeability or “approximate exchangeability” (de Finetti 1938; Bacallado, Diaconis and Holmes, 2015) and interpretation for roles of model parameters.

To date preliminary results for:

- GM-dependence (Griffiths, Milne, 1978; Lijoi, Nipoti, P, 2014);
- Compound random measures (Griffin, Leisen, 2017);
- Clayton Lévy copula (Cont, Tankov, 2004; Epifani, Lijoi, 2010).

3. If the posterior distributions

$$\left(\begin{array}{c} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{array} \right) \Big| \mathbf{x}_1, \mathbf{x}_2 \quad \left(\begin{array}{c} \tilde{\mu}_1^{\text{ex}} \\ \tilde{\mu}_2^{\text{ex}} \end{array} \right) \Big| \mathbf{x}_1, \mathbf{x}_2$$

are available (in terms of CRMs and linear combinations of random variables with known distributions) one may evaluate

$$D(\tilde{P}_{1, \mathbf{x}_1, \mathbf{x}_2}, \tilde{P}_{2, \mathbf{x}_1, \mathbf{x}_2}),$$

or a bound for it, in order to build some Bayesian test for distributional homogeneity.

Hierarchical NRM1 processes

Hierarchical NRM1s for partially exchangeable data

$$(X_{1,i}, X_{2,j}) \mid (\tilde{P}_1, \tilde{P}_2) \stackrel{\text{iid}}{\sim} \tilde{P}_1 \times \tilde{P}_2 \quad \forall i, j \geq 1$$

$$(\tilde{P}_1, \tilde{P}_2) \mid \tilde{P}_0 \stackrel{\text{iid}}{\sim} \text{NRM1}(\rho, c, \tilde{P}_0)$$

$$\tilde{P}_0 \sim \text{NRM1}(\rho_0, c_0, P_0)$$

with P_0 a non-atomic measure on $\mathbb{X}$.

Special cases

► Hierarchical Dirichlet process (HDP)

⇒ NRM1s coincide with the Dirichlet process, i.e.

$$\rho(d\nu) = \rho_0(d\nu) = \frac{e^{-\nu}}{\nu} d\nu$$

► Hierarchical normalized stable process (HnstP)

⇒ NRM1s coincide with the normalized stable process, i.e.

$$\rho(d\nu) = \frac{\sigma}{\Gamma(1-\sigma)\nu^{1+\sigma}} d\nu \quad \rho_0(d\nu) = \frac{\sigma_0}{\Gamma(1-\sigma_0)\nu^{1+\sigma_0}} d\nu$$

Correlation structure

Correlation structure for hierarchical NRMIs

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A))$$

$$= \left\{ 1 + c_0 c \frac{\int_0^\infty u e^{-c\psi(u)} \tau_2(u) du \int_0^\infty u e^{-c_0\psi_0(u)} \tau_{1,0}^2(u) du}{\int_0^\infty u e^{-c_0\psi_0(u)} \tau_{2,0}(u) du} \right\}^{-1} > 0$$

$$\text{with } \psi(u) = \int_0^\infty [1 - e^{-us}] \rho(s) ds \text{ and } \tau_q(u) = \int_0^\infty s^q e^{-us} \rho(s) ds.$$

Special cases

► Hierarchical Dirichlet process (HDP)

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = \frac{c + 1}{c + 1 + c_0}$$

$\implies$ corr increasing in c and decreasing in c_0 : if $c \uparrow \infty$ ($c_0 \uparrow \infty$), then $\text{corr} \uparrow 1$ ($\text{corr} \downarrow 0$).

► Hierarchical normalized stable process (HnstP)

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = \frac{1 - \sigma_0}{1 - \sigma\sigma_0}$$

$\implies$ corr increasing in σ and decreasing in σ_0 : if $\sigma \uparrow 1$ ($\sigma_0 \uparrow 1$), then $\text{corr} \uparrow 1$ ($\text{corr} \downarrow 0$).

Induced partition structure

- ▶ Two samples $\mathbf{X}_1 = \{X_{1,1}, \dots, X_{1,n_1}\}$ and $\mathbf{X}_2 = \{X_{2,1}, \dots, X_{2,n_2}\}$ from partially exchangeable sequences $(X_{1,j})_{j \geq 1}$ and $(X_{2,i})_{i \geq 1}$
- ▶ Hierarchical NRMI prior selects a.s. **discrete random probabilities** $(\tilde{P}_1, \tilde{P}_2)$
 $\implies$ **ties within each sample and possibly also across different samples.**
- ▶ **Partition** of $[N] = \{1, \dots, n_1 + n_2\}$ induced by $\mathbf{X}_1$ and $\mathbf{X}_2$ into
 - ▶ k_1 distinct values specific to $\mathbf{X}_1$
 - ▶ k_2 distinct values specific to $\mathbf{X}_2$
 - ▶ k_0 distinct values shared by the two samples
 - ▶ the corresponding frequencies are best recorded as $\mathbf{n}_i = (n_{1,i}, n_{2,i})$ for $i = 1, \dots, k$, with k_2 $n_{1,j}$'s and k_1 $n_{2,j}$'s being 0.

partially Exchangeable Partition Probability Function (pEPPF)

$$\Pi_k^{(N)}(\mathbf{n}_1, \dots, \mathbf{n}_k) = \mathbb{E} \int_{\mathbb{X}^k} \prod_{j=1}^k \tilde{P}_1^{n_{1,j}}(dx_j) \tilde{P}_2^{n_{2,j}}(dx_j)$$

where $N = n_1 + n_2$ and $k = k_1 + k_2 + k_0$.

Chinese restaurant franchise metaphor

Observable level: customers and dishes

- ▶ There are $d = 2$ restaurants sharing the same menu.
- ▶ $X_{i,j}$: label of the **dish** served at restaurant i to customer j
- ▶ Sample information: $N = n_1 + n_2$ customers eat $k = k_1 + k_2 + k_0$ distinct dishes with frequencies $\mathbf{n}_1, \dots, \mathbf{n}_k$.

Latent level: tables (governed by $\tilde{P}_0$)

- ▶ Customers eating dish j in restaurant i are further partitioned into tables.
- ▶ $\ell_{i,j}$ is the number of tables in restaurant i serving dish j whose range is $\{1, \dots, n_{i,j}\}$ if dish j is served at restaurant i and 0 otherwise.
- ▶ $\bar{\ell}_{\bullet,j} = \sum_{i=1}^2 \ell_{i,j}$ is then the total number of tables serving dish j for $j = 1, \dots, k$
- ▶ $|\ell|$: total number of tables in the two restaurants

Augmented partition structure

- ▶ $q_{i,j,t}$: frequency of customers at restaurant i eating dish j and sitting at table t
- ▶ $\mathbf{q}_{i,j} = (q_{i,j,1}, \dots, q_{i,j,\ell_{i,j}})$: frequency vector of customers in restaurant i eating dish j at each of the $\ell_{i,j}$ tables.
- ▶ By marginalizing over the tables one re-obtains the observed frequencies

$$n_{i,j} = |\mathbf{q}_{i,j}| = \sum_{t=1}^{\ell_{i,j}} q_{i,j,t}.$$

Partially exchangeable partition probability function

pEPPF of a hierarchical NRM

$$\Pi_k^{(N)}(n_1, \dots, n_k) = \sum_{\ell} \sum_q \Phi_{k,0}^{(|\ell|)}(\bar{\ell}_{\bullet 1}, \dots, \bar{\ell}_{\bullet k}) \prod_{i=1}^2 \Phi_{\bar{\ell}_{i\bullet}, i}^{(n_i)}(q_{i,1}, \dots, q_{i,k})$$

- ▶ $\sum_q$ is a sum over all partitions
- ▶ $\sum_{\ell}$ is a sum over all compatible table configurations, i.e. over $\ell_{i,j} \in \{1, \dots, n_{i,j}\}$ with $\ell_{i,j} = 0$ if $n_{i,j} = 0$.
- ▶ Partition probability function with the constraint $|q_{i,j}| = n_{i,j}$

$$\Phi_{\bar{\ell}_{i\bullet}, i}^{(n_i)}(q_{i,1}, \dots, q_{i,k}) = \frac{\theta_{\bar{\ell}_{i\bullet}}}{\Gamma(n_i)} \int_0^\infty u^{n_i-1} e^{-\theta_{\bar{\ell}_{i\bullet}} \psi(u)} \prod_{j=1}^k \prod_{t=1}^{\ell_{i,j}} \tau_{q_{i,j,t}}(u) du.$$

- ▶ $\Phi_{k,0}^{(|\ell|)}(\bar{\ell}_{\bullet 1}, \dots, \bar{\ell}_{\bullet k})$ is the EPPF associated to $\tilde{P}_0 \sim NRM(\rho, c_0, P_0)$

$$\Phi_{k,0}^{(|\ell|)}(\bar{\ell}_{\bullet 1}, \dots, \bar{\ell}_{\bullet k}) = \frac{c_0^k}{\Gamma(|\ell|)} \int_0^\infty u^{|\ell|-1} e^{-c_0 \psi_0(u)} \prod_{j=1}^k \tau_{\bar{\ell}_{\bullet j}, 0}(u) du,$$

with $\psi_0(u) = \int_0^\infty [1 - e^{-uv}] \rho_0(dv)$ and $\tau_{q,0}(u) = \int_0^\infty v^q e^{-uv} \rho_0(dv)$.

Special cases

pEPPF HDP

$$\Pi_k^{(n)}(\mathbf{n}_1, \dots, \mathbf{n}_k) = \frac{\theta_0^k}{\prod_{i=1}^2 (\theta)_{n_i}} \sum_{\ell} \frac{\theta^{|\ell|}}{(\theta_0)_{|\ell|}} \prod_{j=1}^k (\bar{\ell}_{\bullet j} - 1)! \prod_{i=1}^2 |\mathfrak{s}(n_{i,j}, \ell_{i,j})|$$

where $|\mathfrak{s}(r, s)|$ is the absolute value of the Stirling number of the first kind.

pEPPF hierarchical normalized stable process (HNstP)

$$\frac{\sigma_0^{k-1} \Gamma(k)}{\prod_{i=1}^2 \Gamma(n_i)} \sum_{\ell} \frac{\sigma^{|\ell|-2} \prod_{i=1}^2 \Gamma(\bar{\ell}_{i\bullet})}{\Gamma(|\ell|)} \prod_{j=1}^k (1 - \sigma_0)_{\bar{\ell}_{\bullet j} - 1} \prod_{i=1}^2 \prod_{j=1}^k \frac{\mathcal{C}(n_{i,j}, \ell_{i,j}; \sigma)}{\sigma^{\ell_{i,j}}}$$

where $\mathcal{C}(n, \ell; \sigma)$ is the *generalized factorial coefficient*.

$\Rightarrow$ From the pEPPF a **generalized Pólya urn scheme** is obtained that can be used within MCMC samplers for density estimation, prediction problems etc.

Distribution of the number of clusters K_N

In order to derive the distribution of the distinct dishes K_N eaten by $N_1 + N_2$ customers define:

- ▶ K'_{1,N_2} and K'_{2,N_2} are the number of tables the customers seated in the two restaurants;
- ▶ $K_{0,t}$ is the number of distinct dishes (generated by $\tilde{P}_0$) served at the t tables.

Distribution of K_N for a hierarchical NRM

$$\mathbb{P}[K_N = k] = \sum_{t=k}^N \mathbb{P}[K_{0,t} = k] \mathbb{P}[K'_{1,N_1} + K'_{2,N_2} = t]$$

- ▶ The distributions of $K_{0,t}$ and of K'_{1,N_1} and K'_{2,N_2} are available once the corresponding EPPFs are known.
- ▶ The law of K_N coincides with $K_{0,(K'_{1,N_1} + K'_{2,N_2})}$ i.e. the random number of dishes served at the random number of tables $K'_{1,N_1} + K'_{2,N_2}$.

Asymptotics for the number of clusters K_N

What is the growth rate of K_N as N_1 and N_2 diverge?

The notation $Y_n \simeq \lambda(n)$, for $n \rightarrow \infty$, means that $\lim_n Y_n/\lambda(n)$ almost surely exists and equals a finite random variable.

Asymptotics of K_N for a hierarchical NRMI

Suppose $K_{0,N} \simeq \lambda_0(N)$ and $K'_{i,N} \simeq \lambda(N)$ as $N \rightarrow \infty$. Then letting $N_1 = N_2 = N/2$.

$$K_N \simeq \lambda_0(\eta \lambda(N/2)) \quad \text{as } N \rightarrow \infty,$$

for some positive and finite positive random variable η .

Special cases

- Hierarchical Dirichlet process (HDP):

$$K_N \simeq \log \log N$$

- Hierarchical normalized stable process (HnstP):

$$K_N \simeq N^{\sigma\sigma_0}$$

Hierarchical Pitman–Yor process

All previous results carry over to the case of hierarchical Pitman–Yor processes with minor modifications. For instance:

Correlation structure

$$\text{corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = \left\{ 1 + \frac{1 - \sigma}{1 - \sigma_0} \frac{\theta_0 + \sigma_0}{\theta + 1} \right\}^{-1}$$

Distribution of K_N

- ▶ exact distribution

$$\begin{aligned} \mathbb{P}[K_N = k] &= \sum_{t=k}^N \mathbb{P}[K_{0,t} = k] \mathbb{P}\left[K'_{1,N_1} + K'_{2,N_2} = t\right] \\ &= \sum_{t=k}^N \frac{\prod_{r=1}^{k-1} (\theta_0 + r\sigma_0)}{(\theta_0 + 1)_{t-1}} \frac{\mathcal{C}(t, k; \sigma_0)}{\sigma_0^k} \sum_{(\zeta_1, \zeta_2) \in \Delta_t} \prod_{i=1}^2 \frac{\prod_{r=1}^{\zeta_i-1} (\theta + r\sigma)}{(\theta + 1)_{N_i-1}} \frac{\mathcal{C}(N_i, \zeta_i; \sigma)}{\sigma^{\zeta_i}} \end{aligned}$$

- ▶ asymptotic growth $K_N \simeq N^{\sigma\sigma_0}$

Concluding remarks

- ▶ The study of nonparametric models for non-exchangeable data is analytically challenging but not impossible!
- ▶ The availability of the **pEPPF** allows to study important quantities such as the **number of clusters K_N** , which have an intuitive interpretation. Combined with results on the **correlation structure** and the **quantification of the dependence structure**, **methodological guidelines** on the choice of the parameters can be deduced (as well as soon hopefully also on the type of dependence structure).
- ▶ **Posterior characterizations for hierarchical NRMI**s are available and are the **first completely explicit** posterior representations in the partially exchangeable case.
- ▶ Thanks to the distributional results derivation of **marginal and conditional sampling schemes** is quite **straightforward**.
- ▶ **Technique for deriving marginal properties and posterior distributions is general** and needs only adaptations depending on the specific transformations of the CRMs. For instance, it works for mixture hazards with hierarchical dependence and for nested processes (possibly including an additive layer).

References

- Bacallado, Diaconis & Holmes (2015). de Finetti priors using Markov chain Monte Carlo computations. *Statistics & Computing*, **25**, 797-808.
- Camerlenghi, Lijoi, Orbanz & Prünster (2018). Distribution theory for Hierarchical Processes. *Annal. Statist.*, **47**, 67-92.
- Camerlenghi, Dunson, Lijoi, Prünster & Rodriguez (2019). Latent nested nonparametric priors. (With discussion) *Bayesian Analysis*, forthcoming.
- Catalano, Lijoi & Prünster (2019). Quantifying the dependence structure of vectors of random measures. *Work in progress*.
- de Finetti (1938). Sur la condition d'équivalence partielle. *Act.sci. ind.* **739**, 5-18.
- Griffin & Leisen (2017). Compound random measures and their use in bayesian non-parametrics. *JRSS B*, **79**, 525-545.
- Kingman (1967). Completely random measures. *Pacific J. Math.* **21**, 59-78.
- Lijoi, Nipoti & Prünster (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260-1291.
- MacEachern (1999). Dependent nonparametric processes. In *ASA Proceedings*.
- MacEachern (2000). Dependent Dirichlet processes. *OSU Tech. Report*.
- Müller, Quintana & Rosner (2004). A method for combining inference across related nonparametric Bayesian models. *J. Roy. Statist. Soc. Ser. B* **66**, 735-749.
- Pitman & Yor (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855-900.
- Regazzini, Lijoi & Prünster (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560-585.
- Rodríguez, Dunson & Gelfand (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103**, 1131-1144.
- Teh, Jordan, Beal & Blei (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 1566-1581.

“ With this form of statistics [Bayesian Statistics ed.] we describe the world in probabilities, even before the data has been collected. And when there is new data, we update the probabilities. ” (Aad, 2017)



HAPPY BIRTHDAY