Discreteness and dependence: an effective interplay in Bayesian nonparametrics

2022 ISBA World Meeting – Montréal, Canada

Antonio Lijoi

Bocconi University and Institue of Data Science and Analytics, Milano

Outline

- Discrete priors, random partitions and prediction
- Partial exchangeability and dependence
- Multivariate species sampling processes
- Flexible dependence structures & discreteness
  - Correlation structure
  - Random partitions
  - Atoms
- Measuring dependence

Discrete priors, random partitions and prediction

Exchangeable sequences

- Sequence $\boldsymbol{X} = (X)_{n\geq 1}$ of observations or latent features
- **Standard assumption**: exchangeability of $\boldsymbol{X}$

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \pi\boldsymbol{X} = (X_{\pi(i)})_{i\geq 1} \qquad \text{for any finite permutation } \pi \text{ of } \mathbb{N}$$

Exchangeable sequences

- Sequence $\boldsymbol{X} = (X)_{n\geq 1}$ of observations or latent features
- **Standard assumption**: exchangeability of $\boldsymbol{X}$

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \pi\boldsymbol{X} = (X_{\pi(i)})_{i\geq 1} \qquad \text{for any finite permutation } \pi \text{ of } \mathbb{N}$$

Conditional i.i.d. characterization

For any $n \geq 1$ and $A_1, \ldots, A_n$

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\mathsf{P}} \prod_{j=1}^{n} p(A_j)\, Q(\mathrm{d}p)$$

where P is the space of probability measures on the sample space or, equivalently,

$$X_i | \tilde{p} \stackrel{\text{iid}}{\sim} \tilde{p} \qquad \tilde{p} \sim Q$$

$$Q = \begin{cases} \text{probability measure on } \mathsf{P} \\ \text{de Finetti measure of } \boldsymbol{X} \text{ / prior distribution} \end{cases}$$

Prediction

In B. de Finetti's own words

> *"Exchangeability" is the name and the notion proposed to replace "independent with unknown constant probability". Such terminology is clearly contradictory:* [...] *let us think of drawing successively – with replacement – balls from an urn with unknown composition. Are the successive drawings independent?*
> *They would be, for someone informed about the unknown (for other people) composition; for such a people the drawings are not informative, and the probability, for him, does not change. But, for any other observer, independence cannot, obviously, hold: the observed frequency is* [...] *informative ...* [de Finetti (1979)]

Prediction

In B. de Finetti's own words

> *"Exchangeability" is the name and the notion proposed to replace "independent with unknown constant probability". Such terminology is clearly contradictory:* [...] *let us think of drawing successively – with replacement – balls from an urn with unknown composition. Are the successive drawings independent?*
> *They would be, for someone informed about the unknown (for other people) composition; for such a people the drawings are not informative, and the probability, for him, does not change. But, for any other observer, independence cannot, obviously, hold: the observed frequency is* [...] *informative ...* [de Finetti (1979)]

- Notion of "homogeneity" or "analogy" among observations for prediction

Prediction

In B. de Finetti's own words

> *"Exchangeability" is the name and the notion proposed to replace "independent with unknown constant probability". Such terminology is clearly contradictory:* [...] *let us think of drawing successively – with replacement – balls from an urn with unknown composition. Are the successive drawings independent?*
> *They would be, for someone informed about the unknown (for other people) composition; for such a people the drawings are not informative, and the probability, for him, does not change. But, for any other observer, independence cannot, obviously, hold: the observed frequency is* [...] *informative ...* [de Finetti (1979)]

- Notion of "homogeneity" or "analogy" among observations for prediction
- In view of the exchangeability assumption, for any $m \geq 1$

$$\mathbb{P}[X_{n+1} \in A_1, \ldots, X_{n+m} \in A_m \mid X_1, \ldots, X_n] = \int_{\mathsf{P}_{\mathbb{X}}} \prod_{i=1}^{m} p(A_i)\, Q(\mathrm{d}p \mid X_1, \ldots, X_n)$$

Prediction

In B. de Finetti's own words

> *"Exchangeability" is the name and the notion proposed to replace "independent with unknown constant probability". Such terminology is clearly contradictory:* [...] *let us think of drawing successively – with replacement – balls from an urn with unknown composition. Are the successive drawings independent?*
> *They would be, for someone informed about the unknown (for other people) composition; for such a people the drawings are not informative, and the probability, for him, does not change. But, for any other observer, independence cannot, obviously, hold: the observed frequency is* [...] *informative ...* [de Finetti (1979)]

- Notion of "homogeneity" or "analogy" among observations for prediction
- In view of the exchangeability assumption, for any $m \geq 1$

$$\mathbb{P}[X_{n+1} \in A_1, \ldots, X_{n+m} \in A_m \,|\, X_1, \ldots, X_n] = \int_{\mathrm{P}_{\mathbb{X}}} \prod_{i=1}^{m} p(A_i)\, Q(\mathrm{d}p \,|\, X_1, \ldots, X_n)$$

- Most often Q selects discrete distributions

$$Q\Big(\Big\{p: \quad p = \sum_i \omega_i \delta_{\theta_i} \quad \text{for some } (\omega_i)_i \text{ and } (\theta_i)_i\Big\}\Big) = 1$$

Species sampling

Species sampling process

- $\theta_i \overset{\text{iid}}{\sim} H$ with $H(\{x\}) = 0$ for any x
- $(w_i)_{i\geq 1}$ non–negative and such that $\sum_i w_i \leq 1$, almost surely
- $(\theta_i)_{i\geq 1} \perp (w_i)_{i\geq 1}$

The random probability measure

$$\tilde{p} = \sum_{i\geq 1} w_i \, \delta_{\theta_i} + \left(1 - \sum_{i\geq 1} w_i\right) H$$

is a *species sampling process* (SSP). It is a *proper* SSP if $\sum_{i\geq 1} w_i = 1$ (a.s.)

Species sampling

Species sampling process

- $\theta_i \stackrel{\text{iid}}{\sim} H$ with $H(\{x\}) = 0$ for any x
- $(w_i)_{i \geq 1}$ non–negative and such that $\sum_i w_i \leq 1$, almost surely
- $(\theta_i)_{i \geq 1} \perp (w_i)_{i \geq 1}$

The random probability measure

$$\tilde{p} = \sum_{i \geq 1} w_i \, \delta_{\theta_i} + \Big(1 - \sum_{i \geq 1} w_i\Big) \, H$$

is a *species sampling process* (SSP). It is a *proper* SSP if $\sum_{i \geq 1} w_i = 1$ (a.s.)

Species sampling sequence

A sequence $(X_i)_{i \geq 1}$ of random variables such that

$$X_i \,|\, \tilde{p} \stackrel{\text{iid}}{\sim} \tilde{p} \qquad \& \qquad \tilde{p} = \text{SSP}$$

is termed a *species sampling sequence* (SSS).

A characterization

- $X^*_{1,n}, \ldots, X^*_{k_n,n}$ be the k_n distinct values in $\boldsymbol{X}^{(n)} = (X_1, \ldots, X_n)$ with respective frequencies $\boldsymbol{n} = (n_1, \ldots, n_{k_n})$
- $\theta_i \overset{\text{iid}}{\sim} H$, with $H(\{x\}) = 0$ for any x.

A characterization

- $X^*_{1,n}, \ldots, X^*_{k_n,n}$ be the k_n distinct values in $\boldsymbol{X}^{(n)} = (X_1, \ldots, X_n)$ with respective frequencies $\boldsymbol{n} = (n_1, \ldots, n_{k_n})$
- $\theta_i \overset{\text{iid}}{\sim} H$, with $H(\{x\}) = 0$ for any x.

Theorem (Pitman, 1996)

A sequence $(X_n)_{n\geq 1}$ is a SSS if and only if there exist weights $\{p_{j,n}(n_1, \ldots, n_k) : \ 1 \leq j \leq k \leq n\}$ such that $X_1 = \theta_1$ and for any $n \geq 1$

$$X_{n+1} \mid \boldsymbol{X}^{(n)} = \begin{cases} \theta_{n+1} & \text{with prob } \ p_{k_n+1,n}(\boldsymbol{n}^+) \\ X^*_{j,n} & \text{with prob } \ p_{k_n,n}(\boldsymbol{n}^{+j}) \end{cases}$$

where $\boldsymbol{n}^+ = (n_1, \ldots, n_{k_n}, 1)$ and $\boldsymbol{n}^{+j} = (n_1, \ldots, n_j + 1, \ldots, n_{k_n})$.

A characterization

- $X^*_{1,n}, \dots, X^*_{k_n,n}$ be the k_n distinct values in $\boldsymbol{X}^{(n)} = (X_1, \dots, X_n)$ with respective frequencies $\boldsymbol{n} = (n_1, \dots, n_{k_n})$
- $\theta_i \stackrel{\text{iid}}{\sim} H$, with $H(\{x\}) = 0$ for any x.

Theorem (Pitman, 1996)

A sequence $(X_n)_{n\geq 1}$ is a SSS if and only if there exist weights $\{p_{j,n}(n_1, \dots, n_k) : \ 1 \leq j \leq k \leq n\}$ such that $X_1 = \theta_1$ and for any $n \geq 1$

$$X_{n+1} \mid \boldsymbol{X}^{(n)} = \begin{cases} \theta_{n+1} & \text{with prob } \ p_{k_n+1,n}(\boldsymbol{n}^+) \\ X^*_{j,n} & \text{with prob } \ p_{k_n,n}(\boldsymbol{n}^{+j}) \end{cases}$$

where $\boldsymbol{n}^+ = (n_1, \dots, n_{k_n}, 1)$ and $\boldsymbol{n}^{+j} = (n_1, \dots, n_j + 1, \dots, n_{k_n})$.

- Predictive probability function $\implies \{p_{j,n}(n_1, \dots, n_k) : \ 1 \leq j \leq k+1, \ k \leq n\}$
- If one is able to define a predictive probability function, then a SSS $\boldsymbol{X}$ is uniquely identified, without the need of specifying a prior Q. Very challenging a task!
- See Fortini, Ladelli & Regazzini (2002), Lee, Quintana, Müller & Trippa (2013).

Determining the prediction rule

- If $(X_n)_{n\geq 1}$ is a SSS, then $\mathbb{P}[X_i = X_j] > 0$ for any $i \neq j$.
- Random partition Ψ_n of $[n] = \{1, \ldots, n\}$, namely $i \sim j$ if and only if $X_i = X_j$

Determining the prediction rule

- If $(X_n)_{n\geq 1}$ is a SSS, then $\mathbb{P}[X_i = X_j] > 0$ for any $i \neq j$.
- Random partition Ψ_n of $[n] = \{1, \ldots, n\}$, namely $i \sim j$ if and only if $X_i = X_j$

Exchangeable partition probability function (EPPF)

$$\mathbb{P}[\Psi_n = \{C_1, \ldots, C_k\}] = \Phi_k^{(n)}(n_1, \ldots, n_k), \qquad n_i = \mathrm{card}(C_i)$$

For example, with $n = 4$ and $k = 2$

$$\mathrm{Prob}(X_1\, X_2\, X_3\, X_4) = \Phi_2^{(4)}\left(\boldsymbol{n}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$$

Determining the prediction rule

- If $(X_n)_{n\geq 1}$ is a SSS, then $\mathbb{P}[X_i = X_j] > 0$ for any $i \neq j$.
- Random partition Ψ_n of $[n] = \{1, \dots, n\}$, namely $i \sim j$ if and only if $X_i = X_j$

Exchangeable partition probability function (EPPF)

$$\mathbb{P}[\Psi_n = \{C_1, \dots, C_k\}] = \Phi_k^{(n)}(n_1, \dots, n_k), \qquad n_i = \mathrm{card}(C_i)$$

For example, with $n = 4$ and $k = 2$

$$\mathrm{Prob}(\, X_1 \; X_2 \; X_3 \; X_4 \,) = \Phi_2^{(4)}\left(\boldsymbol{n}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$$

Consistency condition & **Symmetry**

$$\Phi_k^{(n)}(n_1, \dots, n_k) = \sum_{j=1}^{k} \Phi_k^{(n+1)}(n_1, \dots, n_j + 1, \dots, n_k) + \Phi_{k+1}^{(n+1)}(n_1, \dots, n_k, 1) \qquad (*)$$

$$\Phi_k^{(n)}(n_1, \dots, n_k) = \Phi_k^{(n)}(n_{\sigma(1)}, \dots, n_{\sigma(k)}) \quad \text{for any finite permutation } \sigma \text{ of } [k] \qquad (\Delta)$$

Random partitions and prediction

Prediction rule

The weights defined by

$$p_{j,n}(\boldsymbol{n}^{+j}) = \frac{\Phi_k^{(n+1)}(\boldsymbol{n}^{+j})}{\Phi_k^{(n)}(\boldsymbol{n})}, \qquad p_{k+1,n}(\boldsymbol{n}^{+}) = \frac{\Phi_{k+1}^{(n+1)}(\boldsymbol{n}^{+})}{\Phi_k^{(n)}(\boldsymbol{n})}$$

identify the prediction rule of a SSS $(X_n)_{n\geq 1}$.

Random partitions and prediction

Prediction rule

The weights defined by

$$p_{j,n}(\boldsymbol{n}^{+j}) = \frac{\Phi_k^{(n+1)}(\boldsymbol{n}^{+j})}{\Phi_k^{(n)}(\boldsymbol{n})}, \qquad p_{k+1,n}(\boldsymbol{n}^{+}) = \frac{\Phi_{k+1}^{(n+1)}(\boldsymbol{n}^{+})}{\Phi_k^{(n)}(\boldsymbol{n})}$$

identify the prediction rule of a SSS $(X_n)_{n\geq 1}$.

- If one can directly specify $\Phi_k^{(n)}$ that satisfies the conditions $(*)$–(Δ), then a SSS $(X_n)_{n\geq 1}$ is identified
- A prior Q is not involved in this construction, though it is uniquely identified by the EPPF $\{\Phi_k^{(n)} : \ 1 \leq k \leq n\}$
- **Pro:** assigning probability to observable events / quantities

An example: Gibbs–type priors (Gnedin & Pitman, 2006)

- Array of weights $\{V_{k,n} : \; 1 \le k \le n\}$ such that $V_{n,k} = V_{n+1,k+1} + (n - k\sigma) V_{n+1,k}$
- $\sigma < 1$
- $\Phi_k^{(n)}(n_1, \dots, n_k) = V_{n,k} \prod_{j=1}^{k} (1-\sigma)_{n_j - 1}$, with $(a)_q = \Gamma(a+q)/\Gamma(a)$

An example: Gibbs–type priors (Gnedin & Pitman, 2006)

- Array of weights $\{V_{k,n}: \ 1 \leq k \leq n\}$ such that $V_{n,k} = V_{n+1,k+1} + (n - k\sigma)V_{n+1,k}$
- $\sigma < 1$
- $\Phi_k^{(n)}(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^{k} (1-\sigma)_{n_j - 1}$, with $(a)_q = \Gamma(a+q)/\Gamma(a)$

Prediction rule

$$p_{j,n}(\boldsymbol{n}^{+j}) = \mathbb{P}[X_{n+1} = X_j^* \mid \boldsymbol{X}^{(n)}] = \frac{V_{n+1,k}}{V_{n,k}}(n_j - \sigma)$$

$$p_{k+1,n}(\boldsymbol{n}^{+}) = \mathbb{P}[X_{n+1} = \text{"new"} \mid \boldsymbol{X}^{(n)}] = \frac{V_{n+1,k+1}}{V_{n,k}}$$

An example: Gibbs–type priors (Gnedin & Pitman, 2006)

- Array of weights $\{V_{k,n}: \ 1 \leq k \leq n\}$ such that $V_{n,k} = V_{n+1,k+1} + (n - k\sigma)V_{n+1,k}$
- $\sigma < 1$
- $\Phi_k^{(n)}(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^k (1-\sigma)_{n_j - 1}$, with $(a)_q = \Gamma(a+q)/\Gamma(a)$

Prediction rule

$$p_{j,n}(\boldsymbol{n}^{+j}) = \mathbb{P}[X_{n+1} = X_j^* \mid \boldsymbol{X}^{(n)}] = \frac{V_{n+1,k}}{V_{n,k}}(n_j - \sigma)$$

$$p_{k+1,n}(\boldsymbol{n}^{+}) = \mathbb{P}[X_{n+1} = \text{"new"} \mid \boldsymbol{X}^{(n)}] = \frac{V_{n+1,k+1}}{V_{n,k}}$$

If K_n is the number of distinct values in $X_1, \ldots, X_n$ and

$$c_n(\sigma) = \begin{cases} 1 & \sigma < 0 \\ \log n & \sigma = 0 \\ n^\sigma & \sigma \in (0,1) \end{cases}$$

then

$$\frac{K_n}{c_n(\sigma)} \xrightarrow{\text{a.s.}} S_\sigma$$

for random variable $S_\sigma > 0$

Gibbs–type priors (ctd)

Dirichlet process: $\sigma = 0$ and $V_{n,k} = \theta^k/(\theta)_n$

$$\mathbb{P}[X_{n+1} = \text{''}\,new\text{''} \mid \boldsymbol{X}^{(n)}] = \frac{\theta}{\theta + n}$$

Gibbs–type priors (ctd)

Dirichlet process: $\sigma = 0$ and $V_{n,k} = \theta^k/(\theta)_n$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta}{\theta + n}$$

Pitman–Yor process: $\sigma \in (0,1)$ and $V_{n,k} = \prod_{i=1}^{k-1}(\theta + i\sigma)/(\theta + 1)_{n-1}$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta + K_n\sigma}{\theta + n}$$

Gibbs–type priors (ctd)

Dirichlet process: $\sigma = 0$ and $V_{n,k} = \theta^k/(\theta)_n$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta}{\theta + n}$$

Pitman–Yor process: $\sigma \in (0,1)$ and $V_{n,k} = \prod_{i=1}^{k-1}(\theta + i\sigma)/(\theta+1)_{n-1}$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta + K_n\sigma}{\theta + n}$$

Finite–dimensional priors: $\sigma < 0$ and

$$V_{n,k} = \sum_{j \geq k} \frac{|\sigma|^{k-1}\prod_{i=1}^{k-1}(j-i)}{(j|\sigma|+1)_{n-1}}\,\pi(j) = \text{mixture of } \mathrm{Dir}_j(|\sigma|, \ldots, |\sigma|)$$

with π a prior on the number of components. See De Blasi, L. and Prünster (2013) for some examples with different choices of π.

Gibbs–type priors (ctd)

Dirichlet process: $\sigma = 0$ and $V_{n,k} = \theta^k/(\theta)_n$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta}{\theta + n}$$

Pitman–Yor process: $\sigma \in (0,1)$ and $V_{n,k} = \prod_{i=1}^{k-1}(\theta + i\sigma)/(\theta + 1)_{n-1}$

$$\mathbb{P}[X_{n+1} = \text{"}new\text{"} \mid \boldsymbol{X}^{(n)}] = \frac{\theta + K_n\sigma}{\theta + n}$$

Finite–dimensional priors: $\sigma < 0$ and

$$V_{n,k} = \sum_{j \geq k} \frac{|\sigma|^{k-1}\prod_{i=1}^{k-1}(j-i)}{(j|\sigma|+1)_{n-1}}\,\pi(j) = \text{mixture of } \mathrm{Dir}_j(|\sigma|, \ldots, |\sigma|)$$

with π a prior on the number of components. See De Blasi, L. and Prünster (2013) for some examples with different choices of π.

- Gibbs–type priors and corresponding prediction rules introduced in BNP by L., Mena and Prünster (2007a, 2007b).
- Frequentist large sample properties in De Blasi, L. and Prünster (2013).
- Review in De Blasi et al. (2015)

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

Definition $\Longrightarrow$ $\Phi_k^{(n)}(n_1, \ldots, n_k) = \mathbb{E} \sum_{i_1 \neq \cdots \neq i_k} w_{i_1}^{n_1} \cdots w_{i_k}^{n_k}$

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

Definition $\Longrightarrow$ $\Phi_k^{(n)}(n_1, \ldots, n_k) = \mathbb{E} \sum_{i_1 \neq \cdots \neq i_k} w_{i_1}^{n_1} \cdots w_{i_k}^{n_k}$

- Not helpful for actual evaluation of $\Phi_k^{(n)}$

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

Definition $\Longrightarrow$ $\Phi_k^{(n)}(n_1, \dots, n_k) = \mathbb{E} \sum_{i_1 \neq \cdots \neq i_k} w_{i_1}^{n_1} \cdots w_{i_k}^{n_k}$

- Not helpful for actual evaluation of $\Phi_k^{(n)}$
- General representation of the EPPF

$$\Phi_k^{(n)}(n_1, \dots, n_k) = \mathbb{E} \int_{\mathbb{X}^k} \tilde{p}^{n_1}(\mathrm{d}x_1) \cdots \tilde{p}^{n_k}(\mathrm{d}x_k)$$

more convenient if $\tilde{p}$ has a nice probabilistic structure such as with completely random measure based models

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

Definition $\Longrightarrow$ $\Phi_k^{(n)}(n_1, \dots, n_k) = \mathbb{E} \sum_{i_1 \neq \cdots \neq i_k} w_{i_1}^{n_1} \cdots w_{i_k}^{n_k}$

- Not helpful for actual evaluation of $\Phi_k^{(n)}$
- General representation of the EPPF

$$\Phi_k^{(n)}(n_1, \dots, n_k) = \mathbb{E} \int_{\mathbb{X}^k} \tilde{p}^{n_1}(\mathrm{d}x_1) \cdots \tilde{p}^{n_k}(\mathrm{d}x_k)$$

  more convenient if $\tilde{p}$ has a nice probabilistic structure such as with completely random measure based models
- Symmetry and, more importantly, consistency conditions are automatically satisfied in this case: no need to check their validity!

The role of the prior Q

- In principle, a SSS can be defined without specifying a prior Q
- Even identifying an EPPF $\Phi_k^{(n)}$ that satisfies the consistency and symmetry conditions is a daunting task.
- What if a prior Q is specified?

Definition $\Longrightarrow$ $\Phi_k^{(n)}(n_1, \ldots, n_k) = \mathbb{E} \sum_{i_1 \neq \cdots \neq i_k} w_{i_1}^{n_1} \cdots w_{i_k}^{n_k}$

- Not helpful for actual evaluation of $\Phi_k^{(n)}$
- General representation of the EPPF

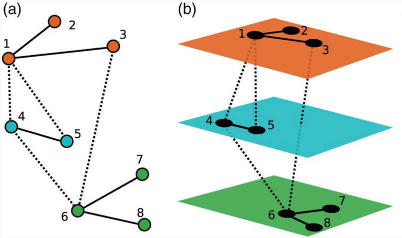
$$\Phi_k^{(n)}(n_1, \ldots, n_k) = \mathbb{E} \int_{\mathbb{X}^k} \tilde{p}^{n_1}(\mathrm{d}x_1) \cdots \tilde{p}^{n_k}(\mathrm{d}x_k)$$

  more convenient if $\tilde{p}$ has a nice probabilistic structure such as with completely random measure based models
- Symmetry and, more importantly, consistency conditions are automatically satisfied in this case: no need to check their validity!
- Consistency is crucial for prediction and validity of inferential procedures. Lack of it, as in some finitely exchangeable models, only allows exploratory data analysis.

Partial exchangeability and dependence

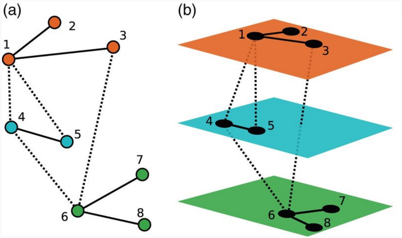
Beyond exchangeability: an example

- Exchangeability is a simplistic form of symmetry
- Not a realistic assumption in several applications



Beyond exchangeability: an example

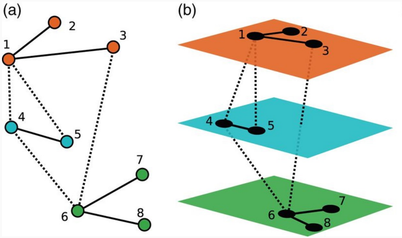
- Exchangeability is a simplistic form of symmetry
- Not a realistic assumption in several applications



- Nodes: $V = \{1, 2, \ldots, 8\}$ Layers: $L = \{\text{Green}, \text{Blue}, \text{Orange}\}$
- Edges between nodes in the same layer or across different layers
- Clustering of nodes based on connections

Beyond exchangeability: an example

- Exchangeability is a simplistic form of symmetry
- Not a realistic assumption in several applications



- Nodes: $V = \{1, 2, \ldots, 8\}$ Layers: $L = \{\text{Green}, \text{Blue}, \text{Orange}\}$
- Edges between nodes in the same layer or across different layers
- Clustering of nodes based on connections

Examples:

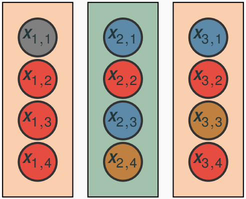
- Bill co-sponsorship ($L =$ political parties)
- Criminal networks ($L =$ main "clan" affiliation)
- ....
- **Francesco Gaffi**'s talk yesterday & **Filippo Ascolani**'s talk tomorrow

Beyond exchangeability

- Data are recorded under different experimental conditions: multicenter studies, change-point problems, topic modeling, ...

Beyond exchangeability

- Data are recorded under different experimental conditions: multicenter studies, change-point problems, topic modeling, ...



- Observations from different samples/groups
- reasonable dependence assumption
  - Homogeneity within each group
  - Heterogeneity across different groups
- Two levels of clustering: samples/groups $\tilde{p}_j$ and observations $X_{i,j}$

Multiple samples and partial exchangeability

- **Assumption:** J samples $\boldsymbol{X}_1 = (X_{1,i})_{i\geq 1}, \ldots, \boldsymbol{X}_J = (X_{J,i})_{i\geq 1}$ are *partially exchangeable*, i.e. for any finite permutations $\pi_1, \ldots, \pi_J$ of $\mathbb{N}$

$$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J) \stackrel{\mathrm{d}}{=} (\pi_1 \boldsymbol{X}_1, \ldots, \pi_J \boldsymbol{X}_J)$$

Multiple samples and partial exchangeability

- **Assumption:** J samples $\boldsymbol{X}_1 = (X_{1,i})_{i\geq 1}, \ldots, \boldsymbol{X}_J = (X_{J,i})_{i\geq 1}$ are *partially exchangeable*, i.e. for any finite permutations $\pi_1, \ldots, \pi_J$ of $\mathbb{N}$

$$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J) \stackrel{\mathrm{d}}{=} (\pi_1 \boldsymbol{X}_1, \ldots, \pi_J \boldsymbol{X}_J)$$

With $J = 2$ samples (red & green): $\Longrightarrow$

$$(X_{1,1}, X_{1,2}, X_{2,1}) \stackrel{\mathrm{d}}{=} (X_{1,2}, X_{1,1}, X_{2,1})$$

$$(X_{1,1}, X_{1,2}, X_{2,1}) \stackrel{\mathrm{d}}{\neq} (X_{1,1}, X_{2,1}, X_{1,2})$$

Multiple samples and partial exchangeability

- **Assumption:** *J samples* $\boldsymbol{X}_1 = (X_{1,i})_{i\geq 1}, \dots, \boldsymbol{X}_J = (X_{J,i})_{i\geq 1}$ are *partially exchangeable*, i.e. for any finite permutations $\pi_1, \dots, \pi_J$ of $\mathbb{N}$

$$(\boldsymbol{X}_1, \dots, \boldsymbol{X}_J) \stackrel{\mathrm{d}}{=} (\pi_1 \boldsymbol{X}_1, \dots, \pi_J \boldsymbol{X}_J)$$

With $J = 2$ samples (red & green): $\Longrightarrow$

$$(X_{1,1}, X_{1,2}, X_{2,1}) \stackrel{\mathrm{d}}{=} (X_{1,2}, X_{1,1}, X_{2,1})$$

$$(X_{1,1}, X_{1,2}, X_{2,1}) \stackrel{\mathrm{d}}{\neq} (X_{1,1}, X_{2,1}, X_{1,2})$$

Conditional independence characterization ($J = 2$)

$$\mathbb{P}\Big[X_{1,1} \in A_1, \cdots, X_{1,N_1} \in A_{N_1}, X_{2,1} \in B_1, \cdots, X_{2,N_2} \in B_{N_2}\Big]$$

$$= \int_{\mathrm{P}^2_{\mathbb{X}}} \prod_{i=1}^{N_1} p_1(A_i) \prod_{j=1}^{N_2} p_2(B_j)\, Q_2(\mathrm{d}p_1, \mathrm{d}p_2)$$

$$(X_{1,i}, X_{2,j}) \,|\, (\tilde{p}_1, \tilde{p}_2) \stackrel{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2, \qquad (\tilde{p}_1, \tilde{p}_2) \sim Q_2$$

de Finetti (1938), Cifarelli & Regazzini (1978), MacEachern (1999, 2000), ...

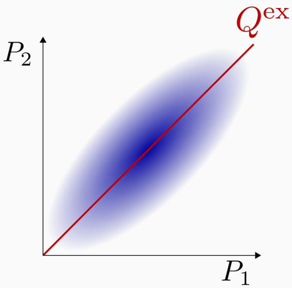
Range of dependence

$Q_2(\{p_1 = p_2\}) = 1$	$\Longrightarrow$	homogeneity across samples (maximal dependence)
$Q_2(\mathrm{d}p_1, \mathrm{d}p_2) = Q_1^*(\mathrm{d}p_1)\, Q_2^*(\mathrm{d}p_2)$	$\Longrightarrow$	unconditional independence (maximal heterogeneity)

Range of dependence

$Q_2(\{p_1 = p_2\}) = 1$	$\Longrightarrow$	homogeneity across samples (maximal dependence)
$Q_2(\mathrm{d}p_1, \mathrm{d}p_2) = Q_1^*(\mathrm{d}p_1)\, Q_2^*(\mathrm{d}p_2)$	$\Longrightarrow$	unconditional independence (maximal heterogeneity)

Exchangeability $\underbrace{\cdots \quad \cdots \quad \cdots}_{\text{Intermediate cases: partial exchangeability}}$ Independence



Discrete priors and random partitions ($J = 2$)

In this talk $\tilde{p}_i$'s are discrete

$$\tilde{p}_1 = \sum_{j \geq 1} w_{1,j}\, \delta_{Z_{1,j}} \qquad \tilde{p}_2 = \sum_{j \geq 1} w_{2,j}\, \delta_{Z_{2,j}}$$

Discrete priors and random partitions ($J = 2$)

In this talk $\tilde{p}_i$'s are discrete

$$\tilde{p}_1 = \sum_{j\geq 1} w_{1,j}\,\delta_{Z_{1,j}} \qquad \tilde{p}_2 = \sum_{j\geq 1} w_{2,j}\,\delta_{Z_{2,j}}$$

- $k \leq N_1 + N_2$ distinct values in the samples

$$\boldsymbol{X}_1^{(N_1)} = (X_{1,1}, \ldots, X_{1,N_1}) \qquad \boldsymbol{X}_2^{(N_2)} = (X_{2,1}, \ldots, X_{2,N_2})$$

Discrete priors and random partitions ($J = 2$)

In this talk $\tilde{p}_i$'s are discrete

$$\tilde{p}_1 = \sum_{j \geq 1} w_{1,j}\, \delta_{Z_{1,j}} \qquad \tilde{p}_2 = \sum_{j \geq 1} w_{2,j}\, \delta_{Z_{2,j}}$$

- $k \leq N_1 + N_2$ distinct values in the samples

$$\boldsymbol{X}_1^{(N_1)} = (X_{1,1}, \ldots, X_{1,N_1}) \qquad \boldsymbol{X}_2^{(N_2)} = (X_{2,1}, \ldots, X_{2,N_2})$$

- $\boldsymbol{n}_1 = (n_{1,1}, \ldots, n_{1,k})$ & $\boldsymbol{n}_2 = (n_{2,1}, \ldots, n_{2,k})$ frequency vectors

$$n_{i,j} = \sharp \text{ elements in sample } i \text{ that equal the } j\text{th distinct value}$$

Discrete priors and random partitions ($J = 2$)

In this talk $\tilde{p}_i$'s are discrete

$$\tilde{p}_1 = \sum_{j\geq 1} w_{1,j}\, \delta_{Z_{1,j}} \qquad \tilde{p}_2 = \sum_{j\geq 1} w_{2,j}\, \delta_{Z_{2,j}}$$

- $k \leq N_1 + N_2$ distinct values in the samples

$$\boldsymbol{X}_1^{(N_1)} = (X_{1,1}, \ldots, X_{1,N_1}) \qquad \boldsymbol{X}_2^{(N_2)} = (X_{2,1}, \ldots, X_{2,N_2})$$

- $\boldsymbol{n}_1 = (n_{1,1}, \ldots, n_{1,k})$ & $\boldsymbol{n}_2 = (n_{2,1}, \ldots, n_{2,k})$ frequency vectors

$$n_{i,j} = \sharp \text{ elements in sample } i \text{ that equal the } j\text{th distinct value}$$

- $n_{1,j} n_{2,j} \neq 0$ whenever the jth distinct value is shared across the two samples

Discrete priors and random partitions ($J = 2$)

In this talk $\tilde{p}_i$'s are discrete

$$\tilde{p}_1 = \sum_{j \geq 1} w_{1,j}\, \delta_{Z_{1,j}} \qquad \tilde{p}_2 = \sum_{j \geq 1} w_{2,j}\, \delta_{Z_{2,j}}$$

- $k \leq N_1 + N_2$ distinct values in the samples

$$\boldsymbol{X}_1^{(N_1)} = (X_{1,1}, \ldots, X_{1,N_1}) \qquad \boldsymbol{X}_2^{(N_2)} = (X_{2,1}, \ldots, X_{2,N_2})$$

- $\boldsymbol{n}_1 = (n_{1,1}, \ldots, n_{1,k})$ & $\boldsymbol{n}_2 = (n_{2,1}, \ldots, n_{2,k})$ frequency vectors

$$n_{i,j} = \sharp \text{ elements in sample } i \text{ that equal the } j\text{th distinct value}$$

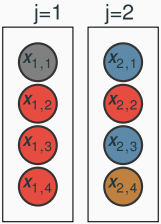
- $n_{1,j} n_{2,j} \neq 0$ whenever the jth distinct value is shared across the two samples
- Induced random partition Ψ_N of $[N] = \{1, \ldots, N_1 + N_2\}$, characterized by the *partially exchangeable partition probability function*

$$\text{pEPPF} = \mathbb{P}[\Psi_n = \{C_1, \ldots, C_k\}] = \Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

where $\text{card}(C_j) = n_{1,j} + n_{2,j} \geq 1$.

Discrete priors and random partitions

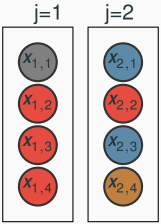
Joint distribution of the induced random partition:



$$\mathbb{P}(\quad) = \Pi_4^{(8)}\left(\boldsymbol{n}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{n}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}\right)$$

Discrete priors and random partitions

Joint distribution of the induced random partition:

$\mathbb{P}($ 

$) = \Pi_4^{(8)}\left(\boldsymbol{n}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{n}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}\right)$

Marginal distribution of the single sample partition:

$\mathbb{P}($ 

$) = \Phi_2^{(4)}\left(\boldsymbol{n}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$

Evaluating the pEPPF

- By definition

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{E} \sum_{\substack{i_1 \neq \cdots \neq i_k \\ j_1 \neq \cdots \neq j_k}} w_{1,i_1}^{n_{1,1}} \cdots w_{1,i_k}^{n_{1,k}} \, w_{2,j_1}^{n_{2,1}} \cdots w_{2,j_k}^{n_{2,k}}$$

  which is impossible to evaluate

Evaluating the pEPPF

- By definition

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{E} \sum_{\substack{i_1 \neq \cdots \neq i_k \\ j_1 \neq \cdots \neq j_k}} w_{1,i_1}^{n_{1,1}} \cdots w_{1,i_k}^{n_{1,k}} \, w_{2,j_1}^{n_{2,1}} \cdots w_{2,j_k}^{n_{2,k}}$$

  which is impossible to evaluate

The most convenient form for actual evaluation is

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{E} \int_{\mathbb{X}_d^k} \prod_{j=1}^{k} \tilde{p}_1^{n_{1,j}}(\mathrm{d}x_j) \, \tilde{p}_2^{n_{2,k}}(\mathrm{d}x_k)$$

with closed form expressions available for models based on completely random measures

- GM–dependent NRMIs (L., Nipoti and Prünster, 2014)
- Hierarchical NRMIs (Camerlenghi et al., 2019)
- Latent nested processes (Camerlenghi et al., 2019)
- Bivariate Pitman–Yor process (Leisen and L., 2011)

Approaching dependence

In this talk: general framework

$(\tilde{p}_1, \ldots, \tilde{p}_J)$ multivariate species sampling process

- Which dependence is induced among
  - the random probability measures $\tilde{p}_j$?
  - the observations, within and between samples?
- Discreteness is helpful in shaping the dependence structure, which may be characterized through
  - correlation
  - the induced random partition
  - the atoms of the $\tilde{p}_j$'s

Multivariate species sampling process

Multivariate species sampling processes

Definition

A vector of random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a *multivariate species sampling process* (mSSP) if for any $j \in \{1, \ldots, J\}$

$$\tilde{p}_j = \sum_{h \geq 1} w_{j,h} \, \delta_{\theta_h} + \Big(1 - \sum_{h \geq 1} w_{j,h}\Big) H$$

- θ_h are iid from the non–atomic probability measure H
- $(w_{1,h}, \ldots, w_{J,h})_{h \geq 1} \perp (\theta_h)_{h \geq 1}$

If $\sum_{h \geq 1} w_{j,h} = 1$, a.s., for every j, then $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is said *proper*.

Multivariate species sampling processes

Definition

A vector of random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a *multivariate species sampling process* (mSSP) if for any $j \in \{1, \ldots, J\}$

$$\tilde{p}_j = \sum_{h \geq 1} w_{j,h} \, \delta_{\theta_h} + \Big(1 - \sum_{h \geq 1} w_{j,h}\Big) H$$

- θ_h are iid from the non–atomic probability measure H
- $(w_{1,h}, \ldots, w_{J,h})_{h \geq 1} \perp (\theta_h)_{h \geq 1}$

If $\sum_{h \geq 1} w_{j,h} = 1$, a.s., for every j, then $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is said *proper*.

Multivariate species sampling sequence

The array of random elements $\{\boldsymbol{X}_j = (X_{j,i})_{i \geq 1} : \ j = 1, \ldots, J\}$ is a *multivariate species sampling sequence* (mSSS) if

- it is partially exchangeable
- its de Finetti measure is a mSSP

A predictive characterization (with $J = 2$)

- Does there exist a predictive characterization that mimics the one seen for SSS?
- If yes, does the prediction rule relate to the induced random partition?

A predictive characterization (with $J = 2$)

- Does there exist a predictive characterization that mimics the one seen for SSS?
- If yes, does the prediction rule relate to the induced random partition?

Let $\theta_h \stackrel{\text{iid}}{\sim} H$ and $X_1^*, \ldots, X_k^*$ the k distinct values observed in the samples $\boldsymbol{X}_j^{(N_j)} = (X_{1,j}, \ldots, X_{N_j,j})$, for $j = 1, 2$, with frequencies $\boldsymbol{n} = (\boldsymbol{n}_1, \boldsymbol{n}_2)$.

A predictive characterization (with $J = 2$)

- Does there exist a predictive characterization that mimics the one seen for SSS?
- If yes, does the prediction rule relate to the induced random partition?

Let $\theta_h \stackrel{\text{iid}}{\sim} H$ and $X_1^*, \ldots, X_k^*$ the k distinct values observed in the samples $\boldsymbol{X}_j^{(N_j)} = (X_{1,j}, \ldots, X_{N_j,j})$, for $j = 1, 2$, with frequencies $\boldsymbol{n} = (\boldsymbol{n}_1, \boldsymbol{n}_2)$.

Predictive characterization

The array of sequences of random variables $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a mSSS if an only if there exists weights $\{p_{i,j} : 1 \leq j \leq k+1, \; n \geq k\}$ such that $X_{1,1} = \theta_1$ and

$$X_{1,N+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) = \begin{cases} \theta_{N+1} & \text{with prob } p_{1,k+1}(\boldsymbol{n}^{+}) \\ X_\ell^* & \text{with prob } p_{1,\ell}(\boldsymbol{n}^{+\ell}) \end{cases}$$

where $N = N_1 + N_2$ and

$$\boldsymbol{n}^{+} = (n_{1,1}, \ldots, n_{k,1}, 1, \boldsymbol{n}_2), \quad \boldsymbol{n}^{+\ell} = (n_{1,1}, \ldots, n_{\ell,1} + 1, \ldots, n_{k,1}, \boldsymbol{n}_2)$$

Similar expression if the $(N+1)$–th observation is from sample 2.

Prediction and random partition

- If one is able to determine a multivariate predictive probability function $\{p_{i,j} : 1 \leq j \leq k+1 \leq n+1;\ i = 1, 2\}$ that satisfies some suitable conditions then a mSSS is identified (Franzolini et al., 2022+).
- Is there a connection between the prediction rule and the distribution of the partially exchangeable partition probability function (pEPPF) that extends what is known in the exchangeable case?
- Can one deduce a Pólya urn scheme from the pEPPF $\Pi_k^{(N)}$?

Prediction and random partition

- If one is able to determine a multivariate predictive probability function $\{p_{i,j} : 1 \leq j \leq k+1 \leq n+1;\ i = 1,2\}$ that satisfies some suitable conditions then a mSSS is identified (Franzolini et al., 2022+).
- Is there a connection between the prediction rule and the distribution of the partially exchangeable partition probability function (pEPPF) that extends what is known in the exchangeable case?
- Can one deduce a Pólya urn scheme from the pEPPF $\Pi_k^{(N)}$?

$$p_{1,k+1}(\boldsymbol{n}) = \frac{\Pi_{k+1}^{(N+1)}(\boldsymbol{n}_1^{+}, \boldsymbol{n}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)} = \mathbb{P}[X_{1,N_1+1} = \text{"new"} \mid \boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}]$$

$$p_{1,\ell}(\boldsymbol{n}) = \frac{\Pi_k^{(N_1+1)}(\boldsymbol{n}_1^{+\ell}, \boldsymbol{n}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)} = \mathbb{P}[X_{1,N_1+1} = X_\ell^* \mid \boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}]$$

Prediction and random partition

- If one is able to determine a multivariate predictive probability function $\{p_{i,j}: 1 \leq j \leq k+1 \leq n+1;\ i = 1, 2\}$ that satisfies some suitable conditions then a mSSS is identified (Franzolini et al., 2022+).
- Is there a connection between the prediction rule and the distribution of the partially exchangeable partition probability function (pEPPF) that extends what is known in the exchangeable case?
- Can one deduce a Pólya urn scheme from the pEPPF $\Pi_k^{(N)}$?

$$p_{1,k+1}(\boldsymbol{n}) = \frac{\Pi_{k+1}^{(N+1)}(\boldsymbol{n}_1^+, \boldsymbol{n}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)} = \mathbb{P}[X_{1,N_1+1} = \text{"new"} \mid \boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}]$$

$$p_{1,\ell}(\boldsymbol{n}) = \frac{\Pi_k^{(N_1+1)}(\boldsymbol{n}_1^{+\ell}, \boldsymbol{n}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)} = \mathbb{P}[X_{1,N_1+1} = X_\ell^* \mid \boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}]$$

- Marginally each $\boldsymbol{X}_j$ is a SSS, for $j = 1, 2$.

From prediction to prior & viceversa

(1) Prediction to prior

Predictive $p_{j,n}$ or pEPPF $\Pi_k^{(N)}$ $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$

prior (or de Finetti measure) Q_J

(2) Prior to prediction

prior (or de Finetti measure) Q_J $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$

Predictive $p_{j,n}$ & pEPPF $\Pi_k^{(N)}$

From prediction to prior & viceversa

(1) Prediction to prior

Predictive $p_{j,n}$ or pEPPF $\Pi_k^{(N)}$ $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$

prior (or de Finetti measure) Q_J

(2) Prior to prediction

prior (or de Finetti measure) Q_J $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$

Predictive $p_{j,n}$ & pEPPF $\Pi_k^{(N)}$

Approach (1)

- Assess probabilities related only to observable quantities/events
- Very challenging (not clear whether even possible) to implement!

From prediction to prior & viceversa

(1) Prediction to prior

Predictive $p_{j,n}$ or pEPPF $\Pi_k^{(N)}$ $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$ prior (or de Finetti measure) Q_J

(2) Prior to prediction

prior (or de Finetti measure) Q_J $\Longrightarrow$ mSSS $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J)$ Predictive $p_{j,n}$ & pEPPF $\Pi_k^{(N)}$

Approach (1)

- Assess probabilities related only to observable quantities/events
- Very challenging (not clear whether even possible) to implement!

Approach (2)

- It amounts to specifying a probability measure Q_J on an infinite-dimensional space
- It can be actually implemented

Flexible dependence structures and discreteness

Correlation structure

$\mathbf{Corr}(\tilde{p}_i(A), \tilde{p}_j(A))$ is available for several special cases of multivariate species sampling proceses: HDP, HNRMI, NDP, LNP, NCoRM, ...

Correlation structure

$\mathbf{Corr}(\tilde{\boldsymbol{p}}_{\boldsymbol{i}}(\boldsymbol{A}), \tilde{\boldsymbol{p}}_{\boldsymbol{j}}(\boldsymbol{A}))$ is available for several special cases of multivariate species sampling proceses: HDP, HNRMI, NDP, LNP, NCoRM, ...

Main findings common to all these cases:

- For any A

  $$\boxed{\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) \quad \text{does not depend on } A}$$

  and it is, then, used as an overall measure of pairwise dependence among random probabilities in $(\tilde{p}_1, \ldots, \tilde{p}_J)$.

Correlation structure

$\mathbf{Corr}(\tilde{\boldsymbol{p}}_{\boldsymbol{i}}(\boldsymbol{A}), \tilde{\boldsymbol{p}}_{\boldsymbol{j}}(\boldsymbol{A}))$ is available for several special cases of multivariate species sampling proceses: HDP, HNRMI, NDP, LNP, NCoRM, ...

Main findings common to all these cases:

- For any A

  $$\boxed{\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) \quad \text{does not depend on } A}$$

  and it is, then, used as an overall measure of pairwise dependence among random probabilities in $(\tilde{p}_1, \ldots, \tilde{p}_J)$.

  Extreme cases

  - $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ fully exchangeable $\Longrightarrow \mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 1$
  - $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ unconditionally independent $\Longrightarrow \mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 0$

Correlation structure

$\mathbf{Corr}(\tilde{\boldsymbol{p}}_i(\boldsymbol{A}), \tilde{\boldsymbol{p}}_j(\boldsymbol{A}))$ is available for several special cases of multivariate species sampling proceses: HDP, HNRMI, NDP, LNP, NCoRM, ...

Main findings common to all these cases:

- For any A

  $$\boxed{\text{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) \quad \text{does not depend on } A}$$

  and it is, then, used as an overall measure of pairwise dependence among random probabilities in $(\tilde{p}_1, \dots, \tilde{p}_J)$.

  Extreme cases

  - $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ fully exchangeable $\Longrightarrow$ $\text{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 1$
  - $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ unconditionally independent $\Longrightarrow$ $\text{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 0$

- For any A

  $$\boxed{\text{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) \geq 0}$$

Open problems of interest

- Is $\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A))$ not dependent on A a general property that holds true beyond the specific examples of priors studied in the literature?
- What is the implied correlation for the observables from any two samples, i.e. $\mathrm{Corr}(X_{i,j}, X_{\ell,j'})$?
  - Can $\mathrm{Corr}(X_{i,j}, X_{\ell,j'})$ be either positive or negative?
  - What is the role played by the atoms of the underlying discrete random probability measures?
- Are there models for which
  - $\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 1$ if and only if $\tilde{p}_i = \tilde{p}_j$
  - $\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = 0$ if and only if $\tilde{p}_i \perp \tilde{p}_j$
- How is dependence reflected by the induced random partitions?
- Is it possible to provide and overall measure of dependence between $\tilde{p}_i$ and $\tilde{p}_j$ that captures the distance between the actual prior specification and the extremes of exchangeability and unconditional independence?

  Ideally, an index $\mathcal{I} \in [0, 1]$ such that $\mathcal{I} = 0$ is equivalent to independence and $\mathcal{I} = 1$ identifies the other extreme of exchangeability

Correlation with multivariate species sampling processes

Correlation structure with mSSP's

(1) If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a mSSS identified by (P_1, P_2), then

$$\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = \frac{\mathbb{P}[X_{1,i} = X_{2,j}]}{\sqrt{\mathbb{P}[X_{1,i} = X_{1,\ell}]\,\mathbb{P}[X_{2,j} = X_{2,\kappa}]}} \geq 0$$

Hence

$$\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_J(A)) = 0 \qquad \text{if and only if} \qquad \mathbb{P}[\text{"ties across samples"}] = 0$$

(2) If $\tilde{p}_1$ and $\tilde{p}_2$ have the same marginal distribution, then

$$\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = \frac{\mathbb{P}[X_{1,i} = X_{2,j}]}{\mathbb{P}[X_{1,i} = X_{1,\ell}]} = \frac{\text{Prob["tie across samples"]}}{\text{Prob["tie within a sample"]}}$$

Correlation with multivariate species sampling processes

Correlation structure with mSSP's

(1) If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a mSSS identified by (P_1, P_2), then

$$\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = \frac{\mathbb{P}[X_{1,i} = X_{2,j}]}{\sqrt{\mathbb{P}[X_{1,i} = X_{1,\ell}]\,\mathbb{P}[X_{2,j} = X_{2,\kappa}]}} \geq 0$$

Hence

$$\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_J(A)) = 0 \qquad \text{if and only if} \qquad \mathbb{P}[\text{"ties across samples"}] = 0$$

(2) If $\tilde{p}_1$ and $\tilde{p}_2$ have the same marginal distribution, then

$$\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = \frac{\mathbb{P}[X_{1,i} = X_{2,j}]}{\mathbb{P}[X_{1,i} = X_{1,\ell}]} = \frac{\text{Prob["tie across samples"]}}{\text{Prob["tie within a sample"]}}$$

- $\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A))$ does not depend on A for the general class of mSSP's
- $\mathrm{Corr}(\tilde{p}_i(A), \tilde{p}_j(A))$ exclusively depends on the probabilities of sharing atoms, regardless of their specific value

$$\mathbb{P}[X_{1,i} = X_{2,j}] = \Pi_1^{(2)}(1,1)$$

$$\mathbb{P}[X_{j,i} = X_{j,\ell}] = \Phi_{1,j}^{(2)}(2) \qquad (j = 1, 2).$$

Regular multivariate species sampling processes

A proper mSSP can be further rewritten as

$$\tilde{p}_j = \sum_{\ell=1}^{L_0} w_\ell^{(0)} \, \delta_{\theta_\ell} + \sum_{k=1}^{K_j} w_k^{(j)} \delta_{\eta_{j,k}}$$

- $\theta_\ell \overset{\text{iid}}{\sim} H$ are shared atoms
- $\eta_{j,k} \overset{\text{iid}}{\sim} H$ are sample–specific atoms
- $L_0, K_j \in \{0, 1, \ldots\} \cup \{\infty\}$

Regular multivariate species sampling processes

A proper mSSP can be further rewritten as

$$\tilde{p}_j = \sum_{\ell=1}^{L_0} w_\ell^{(0)}\, \delta_{\theta_\ell} + \sum_{k=1}^{K_j} w_k^{(j)} \delta_{\eta_{j,k}}$$

- $\theta_\ell \overset{\text{iid}}{\sim} H$ are shared atoms
- $\eta_{j,k} \overset{\text{iid}}{\sim} H$ are sample–specific atoms
- $L_0, K_j \in \{0, 1, \dots\} \cup \{\infty\}$

Regular mSSP

A mSSP $(\tilde{p}_1, \tilde{p}_2)$ is said *regular* if either

- $K_1 = K_2 = 0$ (no sample–specific atoms)

or

- $(w_k^{(1)})_{k\geq 1} \perp (w_k^{(2)})_{k\geq 1}$ (independent frequencies of non–shared atoms)

Examples and property

- **Hierarchical NRMIs & hierarchical Pitman–Yor processes**
  regular mSSP since $K_1 = K_2 = 0$
  Teh et al. (2006), Camerlenghi et al. (2019)
- **Nested Dirichlet processes and latent nested processes**
  regular mSSP since $K_1 = K_2 = 0$
  Rodríguez et al. (2008), Camerlenghi et al. (2019)
- **GM–dependent processes:**
  regular mSSP since $K_1 = K_2 = \infty$ and $(w_k^{(1)})_{k\geq 1} \perp (w_k^{(2)})_{k\geq 1}$
  Müller, Quintana and Rosner (2004), L., Nipoti and Prünster (2014)

Examples and property

- **Hierarchical NRMIs & hierarchical Pitman–Yor processes**
  regular mSSP since $K_1 = K_2 = 0$
  Teh et al. (2006), Camerlenghi et al. (2019)
- **Nested Dirichlet processes and latent nested processes**
  regular mSSP since $K_1 = K_2 = 0$
  Rodríguez et al. (2008), Camerlenghi et al. (2019)
- **GM–dependent processes:**
  regular mSSP since $K_1 = K_2 = \infty$ and $(w_k^{(1)})_{k\geq 1} \perp (w_k^{(2)})_{k\geq 1}$
  Müller, Quintana and Rosner (2004), L., Nipoti and Prünster (2014)

Property

If (P_1, P_2) is a regular mSSP, then

- $\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = 1$ if and only if $\tilde{p}_1 = \tilde{p}_2$ almost surely
- $\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) = 0$ if and only if $\tilde{p}_1 \perp \tilde{p}_2$

Dependence and random partitions

Mixture representation of the pEPPF

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \int_\Lambda \Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda)\, \pi(\mathrm{d}\lambda)$$

Dependence and random partitions

Mixture representation of the pEPPF

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \int_\Lambda \Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda)\, \pi(\mathrm{d}\lambda)$$

For a large family of discrete priors

- there exists $\lambda^* \in \mathrm{supp}(\pi)$ such that $\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda^*) = \Phi_k^{(n)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)$
- $\pi_1 = \pi(\{\lambda^*\}) > 0$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \pi_1 \underbrace{\Phi_k^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}_{\text{joint exchangeability}} + (1 - \pi_1)\, f_{k,N_1,N_2}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

Dependence and random partitions

Mixture representation of the pEPPF

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \int_\Lambda \Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda)\, \pi(\mathrm{d}\lambda)$$

For a large family of discrete priors

- there exists $\lambda^* \in \text{supp}(\pi)$ such that $\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda^*) = \Phi_k^{(n)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)$
- $\pi_1 = \pi(\{\lambda^*\}) > 0$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \pi_1 \underbrace{\Phi_k^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}_{\text{joint exchangeability}} + (1 - \pi_1)\, f_{k,N_1,N_2}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

Dependence and random partitions

Mixture representation of the pEPPF

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \int_\Lambda \Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda)\, \pi(\mathrm{d}\lambda)$$

For a large family of discrete priors

- there exists $\lambda^* \in \mathrm{supp}(\pi)$ such that $\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda^*) = \Phi_k^{(n)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)$
- $\pi_1 = \pi(\{\lambda^*\}) > 0$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \pi_1 \underbrace{\Phi_k^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}_{\text{joint exchangeability}} + (1 - \pi_1)\, f_{k,N_1,N_2}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

- Nested Dirichlet Process (Rodriguez, Dunson & Gelfand, 2008) and Latent Nested Processes (Camerlenghi et al., 2019)
- Semi–Hierarchical Dirichlet Process (Beraha, Guglielmi & Quintana, 2020)

Dependence and random partitions

Mixture representation of the pEPPF

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \int_\Lambda \Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda)\, \pi(\mathrm{d}\lambda)$$

For a large family of discrete priors

- there exists $\lambda^* \in \mathrm{supp}(\pi)$ such that $\Pi_k^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2; \lambda^*) = \Phi_k^{(n)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)$
- $\pi_1 = \pi(\{\lambda^*\}) > 0$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \pi_1 \underbrace{\Phi_k^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}_{\text{joint exchangeability}} + (1 - \pi_1)\, f_{k,N_1,N_2}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

- Nested Dirichlet Process (Rodriguez, Dunson & Gelfand, 2008) and Latent Nested Processes (Camerlenghi et al., 2019)
- Semi–Hierarchical Dirichlet Process (Beraha, Guglielmi & Quintana, 2020)

An example: *Hidden Hierarchical Processes* (L. Prünster and Rebaudo, 2022; Fasano, L., Prünster and Rebaudo, 2022+)

Hidden hierarchical Pitman–Yor processes

$$\text{Nested component } \Rightarrow \begin{cases} \tilde{p}_j \mid G \overset{\text{iid}}{\sim} G \\ G \mid \tilde{p}_0 \sim \text{PY}(\alpha, \gamma; \text{PY}(\beta, \sigma; \tilde{p}_0)) \end{cases}$$

$$\text{Hierarchical component } \Rightarrow \quad \tilde{p}_0 \sim Q = \text{PY}(\beta_0, \sigma_0; H)$$

Hidden hierarchical Pitman–Yor processes

Nested component $\Rightarrow \begin{cases} \tilde{p}_j \mid G \overset{\text{iid}}{\sim} G \\ G \mid \tilde{p}_0 \sim \mathrm{PY}(\alpha, \gamma; \mathrm{PY}(\beta, \sigma; \tilde{p}_0)) \end{cases}$

Hierarchical component $\Rightarrow$ $\tilde{p}_0 \sim Q = \mathrm{PY}(\beta_0, \sigma_0; H)$

Definition

A vector $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a *hidden hierarchical Pitman–Yor process* (HHPY) if

$$\tilde{p}_j \mid G \overset{\text{iid}}{\sim} G = \sum_{k \geq 1} \omega_k \, \delta_{P_k^*} \qquad (\omega_k)_{k \geq 1} \sim \mathrm{GEM}(\alpha, \gamma)$$

$$P_k^* \mid \tilde{p}_0 \overset{\text{iid}}{\sim} \mathrm{PY}(\beta, \sigma; \tilde{p}_0) \qquad \tilde{p}_0 \sim \mathrm{PY}(\beta_0, \sigma_0; H)$$

Notation: $(\tilde{p}_1, \ldots, \tilde{p}_J) \sim \mathrm{HHPY}(\psi; P_0)$, where $\psi = (\alpha, \gamma, \sigma, \beta, \sigma_0, \beta_0)$

Hidden hierarchical Pitman–Yor processes

Nested component $\Rightarrow \begin{cases} \tilde{p}_j \mid G \overset{\text{iid}}{\sim} G \\ G \mid \tilde{p}_0 \sim \text{PY}(\alpha, \gamma; \text{PY}(\beta, \sigma; \tilde{p}_0)) \end{cases}$

Hierarchical component $\Rightarrow \quad \tilde{p}_0 \sim Q = \text{PY}(\beta_0, \sigma_0; H)$

Definition

A vector $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a *hidden hierarchical Pitman–Yor process* (HHPY) if

$$\tilde{p}_j | G \overset{\text{iid}}{\sim} G = \sum_{k \geq 1} \omega_k \, \delta_{P_k^*} \qquad (\omega_k)_{k \geq 1} \sim \text{GEM}(\alpha, \gamma)$$

$$P_k^* | \tilde{p}_0 \overset{\text{iid}}{\sim} \text{PY}(\beta, \sigma; \tilde{p}_0) \qquad \tilde{p}_0 \sim \text{PY}(\beta_0, \sigma_0; H)$$

Notation: $(\tilde{p}_1, \ldots, \tilde{p}_J) \sim \text{HHPY}(\psi; P_0)$, where $\psi = (\alpha, \gamma, \sigma, \beta, \sigma_0, \beta_0)$

$\boxed{\gamma = \sigma = \sigma_0 = 0} \Longrightarrow$ Hidden hierarchical Dirichlet process (HHDP)

pEPPF

When $J = 2$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \frac{1-\gamma}{\alpha+1} \Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + \frac{\alpha+\gamma}{\alpha+1} \Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

pEPPF

When $J = 2$

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \frac{1-\gamma}{\alpha+1} \Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + \frac{\alpha+\gamma}{\alpha+1} \Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)$$

With $\mathscr{C}(n, k; \sigma)$ denoting the generalized factorial coefficient

$$\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) = \frac{\prod_{r=1}^{k-1}(\beta_0 + r\sigma_0)}{(\beta+1)_{N-1}} \sum_{\boldsymbol{\ell}} \frac{\prod_{s=1}^{|\boldsymbol{\ell}|-1}(\beta + s\sigma)}{(\beta_0+1)_{|\boldsymbol{\ell}|-1}}$$

$$\times \prod_{j=1}^{k} \frac{\mathscr{C}(n_{1,j} + n_{2,j}, \ell_j; \sigma)}{\sigma^{\ell_j}} (1-\sigma_0)_{\ell_j - 1}$$

$$\Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \text{ available, though it does not correspond to independence}$$

$$\neq \Phi_{k,1}^{(N_1)}(\boldsymbol{n}_1)\Phi_{k,2}^{(N_2)}(\boldsymbol{n}_2)$$

Some distributional properties

Posterior probability of exchangeability

$$\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \mid \boldsymbol{X}_1, \boldsymbol{X}_2] = \frac{(1-\gamma)\,\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}{(1-\gamma)\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + (\alpha+\gamma)\Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)}$$

Some distributional properties

Posterior probability of exchangeability

$$\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \mid \boldsymbol{X}_1, \boldsymbol{X}_2] = \frac{(1-\gamma)\,\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}{(1-\gamma)\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + (\alpha+\gamma)\Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)}$$

Number of clusters K_N

- Distribution of the number of clusters K_N out of $N = N_1 + N_2$ observations from the two samples

Some distributional properties

Posterior probability of exchangeability

$$\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \mid \boldsymbol{X}_1, \boldsymbol{X}_2] = \frac{(1-\gamma)\,\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}{(1-\gamma)\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + (\alpha+\gamma)\Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)}$$

Number of clusters K_N

- Distribution of the number of clusters K_N out of $N = N_1 + N_2$ observations from the two samples
- Asymptotics of K_N
  - If $N_1 = N_2 = N/2$ and $\sigma, \sigma_0 \in (0, 1)$, as $N \to \infty$

$$K_N \simeq N^{\sigma\sigma_0} \qquad \text{a.s.}$$

Some distributional properties

Posterior probability of exchangeability

$$\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \mid \boldsymbol{X}_1, \boldsymbol{X}_2] = \frac{(1-\gamma)\,\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2)}{(1-\gamma)\Phi_{k,1}^{(N)}(\boldsymbol{n}_1 + \boldsymbol{n}_2) + (\alpha+\gamma)\Phi_{k,2}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2)}$$

Number of clusters K_N

- Distribution of the number of clusters K_N out of $N = N_1 + N_2$ observations from the two samples
- Asymptotics of K_N
  - If $N_1 = N_2 = N/2$ and $\sigma, \sigma_0 \in (0,1)$, as $N \to \infty$

    $$K_N \simeq N^{\sigma\sigma_0} \qquad \text{a.s.}$$

  - If $N_1 = N_2 = N/2$ and $\sigma = \sigma_0 = 0$, as $N \to \infty$

    $$K_N \simeq \log\log N \qquad \text{a.s.}$$

Dependence and random atoms: beyond mSSP's

If one sticks to a mSSP for partially exchangeable data $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, then

(1) $\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) \geq 0$

(2) $\mathrm{Corr}(X_{1,\ell}, X_{2,\kappa}) \geq 0$

Dependence and random atoms: beyond mSSP's

If one sticks to a mSSP for partially exchangeable data $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, then

(1) $\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) \geq 0$

(2) $\mathrm{Corr}(X_{1,\ell}, X_{2,\kappa}) \geq 0$

How recover a wider range of correlations?

- Bivariate SSP on $\mathbb{X}^2$

$$p_1^* = \sum_{h\geq 1} w_h^{(1)}\, \delta_{(\theta_h, \phi_h)}, \quad p_2^* = \sum_{h\geq 1} w_h^{(2)}\, \delta_{(\theta_h, \phi_h)}, \quad (\theta_h, \phi_h) \overset{\text{iid}}{\sim} G_0$$

- Projections

$$\boxed{\tilde{p}_1(\,\cdot\,) = p_1^*(\,\cdot\, \times \mathbb{X}) = \sum_{h\geq 1} w_h^{(1)}\, \delta_{\theta_h}} \qquad \boxed{\tilde{p}_2(\,\cdot\,) = p_2^*(\mathbb{X} \times \,\cdot\,) = \sum_{h\geq 1} w_h^{(2)}\, \delta_{\phi_h}}$$

Dependence and random atoms: beyond mSSP's

If one sticks to a mSSP for partially exchangeable data $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, then

(1) $\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(A)) \geq 0$

(2) $\mathrm{Corr}(X_{1,\ell}, X_{2,\kappa}) \geq 0$

How recover a wider range of correlations?

- Bivariate SSP on $\mathbb{X}^2$

$$p_1^* = \sum_{h\geq 1} w_h^{(1)}\, \delta_{(\theta_h, \phi_h)}, \quad p_2^* = \sum_{h\geq 1} w_h^{(2)}\, \delta_{(\theta_h, \phi_h)}, \quad (\theta_h, \phi_h) \overset{\text{iid}}{\sim} G_0$$

- Projections

$$\boxed{\tilde{p}_1(\,\cdot\,) = p_1^*(\,\cdot\, \times \mathbb{X}) = \sum_{h\geq 1} w_h^{(1)}\, \delta_{\theta_h}} \qquad \boxed{\tilde{p}_2(\,\cdot\,) = p_2^*(\mathbb{X} \times \,\cdot\,) = \sum_{h\geq 1} w_h^{(2)}\, \delta_{\phi_h}}$$

$(\tilde{p}_1, \tilde{p}_2)$ is a bivariate SSP iff: **(1)** $G_0 = P_0^2$ (independence)
or
(2) $G_0(\{\theta_h = \phi_h, \text{ for any } h\}) = 1$ (shared atoms)

Correlations with any sign

- **Probability of a tie within sample 1 or sample 2**

$$\beta = \mathbb{P}[X_{1,i} = X_{1,j}] = \mathbb{P}[X_{2,i'} = X_{2,j'}] = \sum_{k \geq 1} \mathbb{E}\{w_k^{(1)}\}^2$$

- **Probability of a hyper–tie between sample 1 and sample 2**

$$\gamma = \mathbb{P}[(X_{1,1}, X_{2,1}) = (\theta_k, \phi_k) \text{ for some } k] = \sum_{k \geq 1} \mathbb{E}\, w_k^{(1)} w_k^{(2)}$$

Correlations with any sign

- **Probability of a tie within sample 1 or sample 2**

$$\beta = \mathbb{P}[X_{1,i} = X_{1,j}] = \mathbb{P}[X_{2,i'} = X_{2,j'}] = \sum_{k \geq 1} \mathbb{E}\{w_k^{(1)}\}^2$$

- **Probability of a hyper–tie between sample 1 and sample 2**

$$\gamma = \mathbb{P}[(X_{1,1}, X_{2,1}) = (\theta_k, \phi_k) \text{ for some } k] = \sum_{k \geq 1} \mathbb{E}\, w_k^{(1)} w_k^{(2)}$$

Ties vs hyper–ties

- $0 \leq \gamma \leq \beta$
- $\gamma = \beta$ if and only if $w_k^{(1)} \overset{\text{a.s.}}{=} w_k^{(2)}$, for any k.

Correlations with any sign

- **Probability of a tie within sample 1 or sample 2**

$$\beta = \mathbb{P}[X_{1,i} = X_{1,j}] = \mathbb{P}[X_{2,i'} = X_{2,j'}] = \sum_{k\geq 1} \mathbb{E}\{w_k^{(1)}\}^2$$

- **Probability of a hyper–tie between sample 1 and sample 2**

$$\gamma = \mathbb{P}[(X_{1,1}, X_{2,1}) = (\theta_k, \phi_k) \text{ for some } k] = \sum_{k\geq 1} \mathbb{E}\, w_k^{(1)} w_k^{(2)}$$

Ties vs hyper–ties

- $0 \leq \gamma \leq \beta$
- $\gamma = \beta$ if and only if $w_k^{(1)} \overset{\text{a.s.}}{=} w_k^{(2)}$, for any k.

How do the probabilities of tie (β) and of hyper–tie (γ) relate to correlations within and between samples?

Ties and hyper–ties probabilities

Correlations and ties/hyper–ties

$$\mathrm{Corr}(X_{1,i}, X_{1,j}) = \mathrm{Corr}(X_{2,i'}, X_{2,j'}) = \beta$$

$$\mathrm{Corr}(X_{1,1}, X_{2,1}) = \gamma\rho_0 \qquad \text{where} \quad \rho_0 = \mathrm{Corr}(\theta_1, \phi_1)$$

Ties and hyper–ties probabilities

Correlations and ties/hyper–ties

$$\text{Corr}(X_{1,i}, X_{1,j}) = \text{Corr}(X_{2,i'}, X_{2,j'}) = \beta$$

$$\text{Corr}(X_{1,1}, X_{2,1}) = \gamma\rho_0 \qquad \text{where} \quad \rho_0 = \text{Corr}(\theta_1, \phi_1)$$

Corollary

1) $\text{Corr}(X_{1,1}, X_{2,1}) \in [-\beta, \beta]$

2) $|\text{Corr}(X_{1,1}, X_{2,1})| = \beta$ if and only if: **2.1)** $\omega_k^{(1)} \overset{\text{a.s.}}{=} \omega_k^{(2)}$ for any k
2.2) $|\rho_0| = 1$

Ties and hyper–ties probabilities

Correlations and ties/hyper–ties

$$\mathrm{Corr}(X_{1,i}, X_{1,j}) = \mathrm{Corr}(X_{2,i'}, X_{2,j'}) = \beta$$

$$\mathrm{Corr}(X_{1,1}, X_{2,1}) = \gamma\rho_0 \qquad \text{where} \qquad \rho_0 = \mathrm{Corr}(\theta_1, \phi_1)$$

Corollary

1) $\mathrm{Corr}(X_{1,1}, X_{2,1}) \in [-\beta, \beta]$

2) $|\mathrm{Corr}(X_{1,1}, X_{2,1})| = \beta$ if and only if: **2.1)** $\omega_k^{(1)} \overset{\text{a.s.}}{=} \omega_k^{(2)}$ for any k
2.2) $|\rho_0| = 1$

Full Range of Borrowing Information (FuRBI): independence
shrinkage
repulsion / negative correlation

An effective proposal: nFuRBI priors

- Completely random vector

$$(\tilde{\mu}_1, \tilde{\mu}_2) = \Big(\sum_{h \geq 1} J_h \, \delta_{(\theta_h, \phi_h)}, \sum_{h \geq 1} W_h \, \delta_{(\theta_h, \phi_h)} \Big)$$

  where $(J_h, W_h, \theta_h, \phi_h)_{h \geq 1}$ are points of a Poisson process with Lévy intensity

$$\nu(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}\theta, \mathrm{d}\phi) = \rho(s_1, s_2)\mathrm{d}s_1 \mathrm{d}s_2 \, G_0(\mathrm{d}\theta, \mathrm{d}\phi)$$

  and $G_0(\cdot \times \mathbb{X}) = G_0(\mathbb{X} \times \cdot) = P_0(\cdot)$

- Projections of $(\tilde{\mu}_1, \tilde{\mu}_2)$

$$\mu_1(\cdot) = \tilde{\mu}_1(\cdot \times \mathbb{X}), \quad \mu_2(\cdot) = \tilde{\mu}_2(\mathbb{X} \times \cdot)$$

An effective proposal: nFuRBI priors

- Completely random vector

$$(\tilde{\mu}_1, \tilde{\mu}_2) = \Big(\sum_{h \geq 1} J_h \, \delta_{(\theta_h, \phi_h)}, \sum_{h \geq 1} W_h \, \delta_{(\theta_h, \phi_h)} \Big)$$

where $(J_h, W_h, \theta_h, \phi_h)_{h \geq 1}$ are points of a Poisson process with Lévy intensity

$$\nu(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}\theta, \mathrm{d}\phi) = \rho(s_1, s_2) \mathrm{d}s_1 \mathrm{d}s_2 \, G_0(\mathrm{d}\theta, \mathrm{d}\phi)$$

and $G_0(\,\cdot\, \times \mathbb{X}) = G_0(\mathbb{X} \times \,\cdot\,) = P_0(\,\cdot\,)$

- Projections of $(\tilde{\mu}_1, \tilde{\mu}_2)$

$$\mu_1(\,\cdot\,) = \tilde{\mu}_1(\,\cdot\, \times \mathbb{X}), \quad \mu_2(\,\cdot\,) = \tilde{\mu}_2(\mathbb{X} \times \,\cdot\,)$$

nFuRBI priors

If ρ is such that $\int_{\mathbb{R}_+^2} \rho(s_1, s_2) \, \mathrm{d}s_1, \mathrm{d}s_2 = \infty$, then

$$\boxed{(\tilde{p}_1, \tilde{p}_2) = (\mu_1/\mu_1(\mathbb{X}), \, \mu_2/\mu_2(\mathbb{X}))}$$

is a *normalized FuRBI CRM* or, in short, *nFuRBI prior*.

Prediction rule

With $k_i \leq N_1$ and $k_2 \leq N_2$, let

(1) $X^*_{1,1}, \ldots, X^*_{1,k_1}$ distinct values in $\boldsymbol{X}_1^{(N_1)}$

(2) $X^*_{2,1}, \ldots, X^*_{2,k_2}$ distinct values in $\boldsymbol{X}_2^{(N_2)}$

Predictive distribution

Conditional on samples $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ as in (1)–(2), one has

$$X_{1,N_1+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \xi_0 P_0 + \sum_{i=1}^{k_1} \xi_{1,i} \delta_{X^*_{1,i}} + \sum_{i=1}^{k_2} \xi_{2,i}\, P_{X^*_{2,i}}$$

$$X_{2,N_2+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \eta_0 P_0 + \sum_{i=1}^{k_2} \eta_{2,i}\, \delta_{X^*_{2,i}} + \sum_{i=1}^{k_1} \eta_{1,i} P_{X^*_{1,i}}$$

Prediction rule

With $k_i \leq N_1$ and $k_2 \leq N_2$, let

(1) $X^*_{1,1}, \ldots, X^*_{1,k_1}$ distinct values in $\boldsymbol{X}_1^{(N_1)}$

(2) $X^*_{2,1}, \ldots, X^*_{2,k_2}$ distinct values in $\boldsymbol{X}_2^{(N_2)}$

Predictive distribution

Conditional on samples $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ as in (1)–(2), one has

$$X_{1,N_1+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \xi_0 P_0 + \sum_{i=1}^{k_1} \xi_{1,i} \delta_{X^*_{1,i}} + \sum_{i=1}^{k_2} \xi_{2,i}\, P_{X^*_{2,i}}$$

$$X_{2,N_2+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \eta_0 P_0 + \sum_{i=1}^{k_2} \eta_{2,i}\, \delta_{X^*_{2,i}} + \sum_{i=1}^{k_1} \eta_{1,i} P_{X^*_{1,i}}$$

Predictive distributions are linear combinations of: (a) baseline distribution P_0; (b) weighted empirical distribution of the sample data; (c) conditional distributions given data from the other sample

Prediction rule

With $k_i \leq N_1$ and $k_2 \leq N_2$, let

(1) $X^*_{1,1}, \ldots, X^*_{1,k_1}$ distinct values in $\boldsymbol{X}_1^{(N_1)}$

(2) $X^*_{2,1}, \ldots, X^*_{2,k_2}$ distinct values in $\boldsymbol{X}_2^{(N_2)}$

Predictive distribution

Conditional on samples $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ as in (1)–(2), one has

$$X_{1,N_1+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \xi_0 P_0 + \sum_{i=1}^{k_1} \xi_{1,i} \delta_{X^*_{1,i}} + \sum_{i=1}^{k_2} \xi_{2,i}\, P_{X^*_{2,i}}$$

$$X_{2,N_2+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \eta_0 P_0 + \sum_{i=1}^{k_2} \eta_{2,i}\, \delta_{X^*_{2,i}} + \sum_{i=1}^{k_1} \eta_{1,i} P_{X^*_{1,i}}$$

Predictive distributions are linear combinations of: (a) baseline distribution P_0; (b) weighted empirical distribution of the sample data; (c) conditional distributions given data from the other sample

Prediction rule

With $k_i \leq N_1$ and $k_2 \leq N_2$, let

(1) $X^*_{1,1}, \ldots, X^*_{1,k_1}$ distinct values in $\boldsymbol{X}_1^{(N_1)}$

(2) $X^*_{2,1}, \ldots, X^*_{2,k_2}$ distinct values in $\boldsymbol{X}_2^{(N_2)}$

Predictive distribution

Conditional on samples $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ as in (1)–(2), one has

$$X_{1,N_1+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \xi_0 P_0 + \sum_{i=1}^{k_1} \xi_{1,i} \delta_{X^*_{1,i}} + \sum_{i=1}^{k_2} \xi_{2,i} P_{X^*_{2,i}}$$

$$X_{2,N_2+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \eta_0 P_0 + \sum_{i=1}^{k_2} \eta_{2,i} \delta_{X^*_{2,i}} + \sum_{i=1}^{k_1} \eta_{1,i} P_{X^*_{1,i}}$$

Predictive distributions are linear combinations of: (a) baseline distribution P_0; (b) weighted empirical distribution of the sample data; (c) conditional distributions given data from the other sample

Prediction rule

With $k_i \leq N_1$ and $k_2 \leq N_2$, let

(1) $X^*_{1,1}, \ldots, X^*_{1,k_1}$ distinct values in $\boldsymbol{X}_1^{(N_1)}$

(2) $X^*_{2,1}, \ldots, X^*_{2,k_2}$ distinct values in $\boldsymbol{X}_2^{(N_2)}$

Predictive distribution

Conditional on samples $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ as in (1)–(2), one has

$$X_{1,N_1+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \xi_0 P_0 + \sum_{i=1}^{k_1} \xi_{1,i} \delta_{X^*_{1,i}} + \sum_{i=1}^{k_2} \xi_{2,i}\, P_{X^*_{2,i}}$$

$$X_{2,N_2+1} \,|\, (\boldsymbol{X}_1^{(N_1)}, \boldsymbol{X}_2^{(N_2)}) \;\sim\; \eta_0 P_0 + \sum_{i=1}^{k_2} \eta_{2,i}\, \delta_{X^*_{2,i}} + \sum_{i=1}^{k_1} \eta_{1,i} P_{X^*_{1,i}}$$

Predictive distributions are linear combinations of: (a) baseline distribution P_0; (b) weighted empirical distribution of the sample data; (c) conditional distributions given data from the other sample

Posterior distribution

Posterior distribution

If $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ are samples as in (1)–(2), conditional on a suitable set of latent variables, the posterior distribution of P_1 equals the distribution of the random probability measure

$$w_{1,n}P_1^* + w_{2,n}\sum_i \pi_{1,i}\,\delta_{X_{1,i}^*} + (1 - w_{2,n} - w_{3,n})\sum_j \pi_{2,j}\delta_{Z_{X_{2,j}^*}}$$

where $Z_{X_{2,j}^*} \overset{\text{ind}}{\sim} P_{X_{2,j}^*}$. Similar representation holds true for P_2.

Posterior distribution

Posterior distribution

If $\boldsymbol{X}_1^{(N_1)}$ and $\boldsymbol{X}_2^{(N_2)}$ are samples as in (1)–(2), conditional on a suitable set of latent variables, the posterior distribution of P_1 equals the distribution of the random probability measure

$$w_{1,n} P_1^* + w_{2,n} \sum_i \pi_{1,i}\, \delta_{X_{1,i}^*} + (1 - w_{2,n} - w_{3,n}) \sum_j \pi_{2,j} \delta_{Z_{X_{2,j}^*}}$$

where $Z_{X_{2,j}^*} \overset{\text{ind}}{\sim} P_{X_{2,j}^*}$. Similar representation holds true for P_2.

Mixture of

- an updated nFURBI (P_1^*)
- a weighted empirical distribution of the sample data
- a weighted empirical distribution of a transformation of the other sample's data $\Rightarrow G_0\{\theta_k = \phi_k,$ for any $k\} = 1$, then $Z_{X_{2,j}^*} = X_{2,j}^*$: standard posterior characterization of random measure–beased models.
- Talk by **Beatrice Franzolini** (Monday)

Measuring dependence

An approach for measuring dependence

- Several priors for partially exchangeable data are transformations of completely random vectors $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_J)$, i.e.
$$\tilde{p}_j = T(\tilde{\mu}_j), \qquad i = 1, \dots, J$$

An approach for measuring dependence

- Several priors for partially exchangeable data are transformations of completely random vectors $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_J)$, i.e.
$$\tilde{p}_j = T(\tilde{\mu}_j), \qquad i = 1, \ldots, J$$
- Two extreme situations
  - Complete dependence or co-monotonicity $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$
  $$\tilde{\mu}_1 = \cdots = \tilde{\mu}_J \qquad \text{(a.s.)}$$
  - Independence $\tilde{\boldsymbol{\mu}}^{\perp}$, with independent components of $\tilde{\boldsymbol{\mu}} = (\mu_1, \ldots, \mu_j)$

An approach for measuring dependence

- Several priors for partially exchangeable data are transformations of completely random vectors $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_J)$, i.e.

$$\tilde{p}_j = T(\tilde{\mu}_j), \qquad i = 1, \ldots, J$$

- Two extreme situations
  - Complete dependence or co-monotonicity $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$

$$\tilde{\mu}_1 = \cdots = \tilde{\mu}_J \qquad \text{(a.s.)}$$

  - Independence $\tilde{\boldsymbol{\mu}}^{\perp}$, with independent components of $\tilde{\boldsymbol{\mu}} = (\mu_1, \ldots, \mu_j)$

A possible index

For some distance d and completely random vector $\tilde{\boldsymbol{\mu}}$

$$\mathcal{I}_d(\tilde{\boldsymbol{\mu}}) = 1 - \frac{d^2(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})}{\sup_{\boldsymbol{\mu}'} d^2(\boldsymbol{\mu}', \tilde{\boldsymbol{\mu}}^{\mathrm{co}})}$$

An approach for measuring dependence

- Several priors for partially exchangeable data are transformations of completely random vectors $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_J)$, i.e.

$$\tilde{p}_j = T(\tilde{\mu}_j), \qquad i = 1, \dots, J$$

- Two extreme situations
  - Complete dependence or co-monotonicity $\tilde{\boldsymbol{\mu}}^{\mathrm{CO}}$

$$\tilde{\mu}_1 = \cdots = \tilde{\mu}_J \qquad \text{(a.s.)}$$

  - Independence $\tilde{\boldsymbol{\mu}}^{\perp}$, with independent components of $\tilde{\boldsymbol{\mu}} = (\mu_1, \dots, \mu_j)$

A possible index

For some distance d and completely random vector $\tilde{\boldsymbol{\mu}}$

$$\mathcal{I}_d(\tilde{\boldsymbol{\mu}}) = 1 - \frac{d^2(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{CO}})}{\sup_{\boldsymbol{\mu}'} d^2(\boldsymbol{\mu}', \tilde{\boldsymbol{\mu}}^{\mathrm{CO}})}$$

Which d? Based on an extended Wasserstein distance d_W between the corresponding Lévy intensities ν^1 and ν^2

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\mu}}_2) = \sup_A d_W(\nu_A^1, \nu_A^2)$$

The Wasserstein based index of dependence

If for some completely random vector $\tilde{\boldsymbol{\mu}}$, a prior is defined through $\tilde{p}_i = T(\tilde{\mu}_i)$, how close is such a specification from the exchangeability extreme?

For any completely random vector $\tilde{\boldsymbol{\mu}}$, $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \in [0, 1]$. Moreover

- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\text{CO}}$
- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$

The Wasserstein based index of dependence

If for some completely random vector $\tilde{\boldsymbol{\mu}}$, a prior is defined through $\tilde{p}_i = T(\tilde{\mu}_i)$, how close is such a specification from the exchangeability extreme?

For any completely random vector $\tilde{\boldsymbol{\mu}}$, $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \in [0, 1]$. Moreover

- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\text{CO}}$
- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$

- Evaluation for some classes of priors
  - $\tilde{\boldsymbol{\mu}}$ a completely random vector with an additive structure (L., Nipoti and Prünster, 2014)
  - Compound random measures (Griffin and Leisen, 2017)
- Possible extension to posterior distribution would allow testing for distributional homogeneity

The Wasserstein based index of dependence

If for some completely random vector $\tilde{\boldsymbol{\mu}}$, a prior is defined through $\tilde{p}_i = T(\tilde{\mu}_i)$, how close is such a specification from the exchangeability extreme?

For any completely random vector $\tilde{\boldsymbol{\mu}}$, $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \in [0, 1]$. Moreover

- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\mathrm{CO}}$
- $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$

- Evaluation for some classes of priors
  - $\tilde{\boldsymbol{\mu}}$ a completely random vector with an additive structure (L., Nipoti and Prünster, 2014)
  - Compound random measures (Griffin and Leisen, 2017)
- Possible extension to posterior distribution would allow testing for distributional homogeneity

Talks by **Marta Catalano** (Wednesday) and **Hugo Lavenant** (Tuesday)

References

ASCOLANI, FRANZOLINI, LIJOI and PRÜNSTER (2022). Nonparametric priors with full-range borrowing of information. *Under revision*.

CATALANO, LIJOI and PRÜNSTER (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *Ann. Statist.* **49**, 2916–2947.

CATALANO, LAVENANT, LIJOI and PRÜNSTER (2022). A Wasserstein index of dependence for random measures. *arxiv.2109.06646* [math.ST], under revision.

DURANTE, GAFFI, LIJOI and PRÜNSTER (2022). Partially–exchangeable multilayer stochastic block models. *In progress*.

FASANO, LIJOI, PRÜNSTER and REBAUDO (2022+). Prediction in species sampling problems via hidden hierarchical Pitman–Yor processes. *In progress*.

FRANZOLINI, LIJOI, PRÜNSTER and REBAUDO (2022+). Multivariate species sampling proceses. *In progress*.

LIJOI, PRÜNSTER and REBAUDO (2022). Flexible clustering via hidden hierarchical Dirichlet priors. *Scand. J. Statist.*, to appear.

Thank you!