Nonparametric priors for covariate-dependent data

@ Bayesian Statistics in the Big Data Era – *CIRM, Luminy, 26 Nov 2018*

Antonio Lijoi

Bocconi University and Institue of Data Science and Analytics, Milano

Outline

- Heterogeneity & compositions of random measures
- Latent nested processes
- Testing distributional homogeneity
- Illustrations

Heterogeneity & compositions of random measures

Heterogeneous data

- Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

Heterogeneous data

- Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

More generally:

- Covariate $\boldsymbol{z} \in \mathcal{Z}$
- Data $(X_{j,\boldsymbol{z}})_{j \geq 1}$

Heterogeneous data

- Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

More generally:

- Covariate $\boldsymbol{z} \in \mathcal{Z}$
- Data $(X_{j,\boldsymbol{z}})_{j \geq 1}$

Heterogeneity

For any $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$

$$\mathrm{Prob}[X_{i,\boldsymbol{z}_1} \in A,\ X_{j,\boldsymbol{z}_2} \in B] \neq \mathrm{Prob}[X_{i,\boldsymbol{z}_1} \in B,\ X_{j,\boldsymbol{z}_2} \in A]$$

Heterogeneous data

- Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

More generally:

- Covariate $\boldsymbol{z} \in \mathcal{Z}$
- Data $(X_{j,\boldsymbol{z}})_{j \geq 1}$

Heterogeneity

For any $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$

$$\text{Prob}[X_{i,\boldsymbol{z}_1} \in A,\ X_{j,\boldsymbol{z}_2} \in B] \neq \text{Prob}[X_{i,\boldsymbol{z}_1} \in B,\ X_{j,\boldsymbol{z}_2} \in A]$$

- Which dependence between $X_{i,\boldsymbol{z}_1}$ and $X_{j,\boldsymbol{z}_2}$?

Multiple samples and partial exchangeability

$$\mathcal{Z} := \{1, \dots, d\} \quad \Longrightarrow \quad \boldsymbol{X}_1 = (X_{j,1})_{j \geq 1}, \ \dots \ , \ \boldsymbol{X}_d = (X_{j,d})_{j \geq 1}$$

Multiple samples and partial exchangeability

$$\mathcal{Z} := \{1, \dots, d\} \quad \Longrightarrow \quad \boldsymbol{X}_1 = (X_{j,1})_{j \geq 1}, \ \dots \ , \ \boldsymbol{X}_d = (X_{j,d})_{j \geq 1}$$

Partial exchangeability

For any collection $(\pi_1, \dots, \pi_d)$ of finite permutations of $\mathbb{N}$

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

Multiple samples and partial exchangeability

$$\mathcal{Z} := \{1, \dots, d\} \quad \Longrightarrow \quad \boldsymbol{X}_1 = (X_{j,1})_{j \ge 1}, \; \dots \; , \; \boldsymbol{X}_d = (X_{j,d})_{j \ge 1}$$

Partial exchangeability

For any collection $(\pi_1, \dots, \pi_d)$ of finite permutations of $\mathbb{N}$

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

- Homogeneity within each sample

$$\begin{aligned} &\mathrm{Prob}[X_{1,1} = 0, \; X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0] \\ &\qquad = \mathrm{Prob}[X_{1,1} = 1, \; X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 0] \end{aligned}$$

Multiple samples and partial exchangeability

$$\mathcal{Z} := \{1, \ldots, d\} \quad \Longrightarrow \quad \boldsymbol{X}_1 = (X_{j,1})_{j\geq 1}, \ \ldots \ , \ \boldsymbol{X}_d = (X_{j,d})_{j\geq 1}$$

Partial exchangeability

For any collection $(\pi_1, \ldots, \pi_d)$ of finite permutations of $\mathbb{N}$

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

- Homogeneity within each sample

$$\begin{aligned}\mathrm{Prob}[&X_{1,1} = 0,\ X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0]\\ &= \mathrm{Prob}[X_{1,1} = 1,\ X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 0]\end{aligned}$$

- Lack of homogeneity across different samples

$$\begin{aligned}\mathrm{Prob}[&X_{1,1} = 0, X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0]\\ &\neq \mathrm{Prob}[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 1]\end{aligned}$$

Hierarchical model representation

Representation theorem

$\{\boldsymbol{X}_i : \ i = 1, \ldots, d\}$ *is partially exchangeable if and only if*

$$\mathrm{Prob}[\boldsymbol{X}_1 \in A_1, \ldots, \boldsymbol{X}_d \in A_d] = \int_{\mathrm{P}_{\mathbb{X}}^d} \prod_{i=1}^{d} p_i^{(\infty)}(A_i)\, {\color{red}Q_d}(\mathrm{d}p_1, \ldots, \mathrm{d}p_d)$$

where $\mathrm{P}_{\mathbb{X}}$ is the space of probability measures on the sample space $\mathbb{X}$

Hierarchical model representation

Representation theorem

$\{\boldsymbol{X}_i : i = 1, \ldots, d\}$ *is partially exchangeable if and only if*

$$\mathrm{Prob}[\boldsymbol{X}_1 \in A_1, \ldots, \boldsymbol{X}_d \in A_d] = \int_{\mathrm{P}_{\mathbb{X}}^d} \prod_{i=1}^{d} p_i^{(\infty)}(A_i)\, Q_d(\mathrm{d}p_1, \ldots, \mathrm{d}p_d)$$

where $\mathrm{P}_{\mathbb{X}}$ is the space of probability measures on the sample space $\mathbb{X}$

Case $d = 2$ samples

$$\begin{aligned} (X_{i,1}, X_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\ (\tilde{p}_1, \tilde{p}_2) &\sim Q_2 \end{aligned}$$

and Q_2 acts as a prior distribution on $\mathrm{P}_{\mathbb{X}}^2$

Compositions of random measures $(d = 2)$

- **Choice of Q_2?** Compositions of random probability measures

$$
\begin{aligned}
(X_{i,1}, X_{j,2}) | (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\
\tilde{p}_1, \tilde{p}_2 | \tilde{p}_0 &\overset{\text{iid}}{\sim} Q(\,\cdot\,|\tilde{p}_0) \\
\tilde{p}_0 &\sim Q_0
\end{aligned}
$$

Compositions of random measures $(d = 2)$

- **Choice of Q_2?** Compositions of random probability measures

$$
\begin{aligned}
(X_{i,1}, X_{j,2}) | (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\
\tilde{p}_1, \tilde{p}_2 | \tilde{p}_0 &\overset{\text{iid}}{\sim} Q(\,\cdot\,|\tilde{p}_0) \\
\tilde{p}_0 &\sim Q_0
\end{aligned}
$$

- Extreme cases of dependence induced by the prior Q_2
  - *Full exchangeability:* $Q_2(\{(p_1, p_2) \in \mathsf{P}_{\mathbb{X}}^2 : \; p_1 = p_2\}) = 1$
  - *Maximal heterogeneity:* $Q_2(A_1 \times A_2) = Q_1^*(A_1)\, Q_1^{**}(A_2)$

Compositions of random measures $(d = 2)$

- **Choice of Q_2?** Compositions of random probability measures

$$
\begin{aligned}
(X_{i,1}, X_{j,2}) | (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\
\tilde{p}_1, \tilde{p}_2 | \tilde{p}_0 &\overset{\text{iid}}{\sim} Q(\,\cdot\,|\tilde{p}_0) \\
\tilde{p}_0 &\sim Q_0
\end{aligned}
$$

- Extreme cases of dependence induced by the prior Q_2
  - *Full exchangeability:* $Q_2(\{(p_1, p_2) \in \mathsf{P}_{\mathbb{X}}^2 : \ p_1 = p_2\}) = 1$
  - *Maximal heterogeneity:* $Q_2(A_1 \times A_2) = Q_1^*(A_1)\, Q_1^{**}(A_2)$

Possible compostitions through

(1) Hierarchical structure (Igor's talk)

(2) Nested structure

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \stackrel{\text{iid}}{\sim} \tilde{p}_0,$$
$$\tilde{p}_0 \sim Q_0$$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \overset{\text{iid}}{\sim} \tilde{p}_0,$$
$$\tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j}$$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \overset{\text{iid}}{\sim} \tilde{p}_0,$$
$$\tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h \geq 1} \gamma_{h,j} \, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \overset{\text{iid}}{\sim} P_0$$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \stackrel{\text{iid}}{\sim} \tilde{p}_0, \qquad \tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h \geq 1} \gamma_{h,j}\, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \stackrel{\text{iid}}{\sim} P_0$$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \stackrel{\text{iid}}{\sim} \tilde{p}_0, \qquad \tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h \geq 1} \gamma_{h,j}\, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \stackrel{\text{iid}}{\sim} P_0$$

- $(\gamma_{h,1})_{h\geq 1},\ (\gamma_{h,2})_{h\geq 1},\ (\gamma_{h,3})_{h\geq 1},\ \ \cdots$
  independent sequences of non–negative rv's such that $\sum_{h\geq 1} \gamma_{h,j} = 1$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \stackrel{\text{iid}}{\sim} \tilde{p}_0,$$
$$\tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h \geq 1} \gamma_{h,j}\, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \stackrel{\text{iid}}{\sim} P_0$$

- $(\gamma_{h,1})_{h\geq 1},\ (\gamma_{h,2})_{h\geq 1},\ (\gamma_{h,3})_{h\geq 1},\ \ \cdots$
  independent sequences of non–negative rv's such that $\sum_{h\geq 1} \gamma_{h,j} = 1$
- $(\omega_j)_{j\geq 1} \perp (G_j)_{j\geq 1}$ and $\sum_j \omega_j = 1$

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \stackrel{\text{iid}}{\sim} \tilde{p}_0, \qquad \tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j\geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h\geq 1} \gamma_{h,j}\, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \stackrel{\text{iid}}{\sim} P_0$$

- $(\gamma_{h,1})_{h\geq 1},\ (\gamma_{h,2})_{h\geq 1},\ (\gamma_{h,3})_{h\geq 1},\ \ldots$
  independent sequences of non–negative rv's such that $\sum_{h\geq 1} \gamma_{h,j} = 1$
- $(\omega_j)_{j\geq 1} \perp (G_j)_{j\geq 1}$ and $\sum_j \omega_j = 1$
- P_0 is a non–atomic probability measure

Nested structure

Nested random probabilities

If Q_0 is a probability measure on $\mathsf{P}_{\mathsf{P}_{\mathbb{X}}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \overset{\text{iid}}{\sim} \tilde{p}_0,$$
$$\tilde{p}_0 \sim Q_0$$

Discrete case

$$\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \qquad G_j = \sum_{h \geq 1} \gamma_{h,j}\, \delta_{\theta_{h,j}} \qquad \theta_{h,j} \overset{\text{iid}}{\sim} P_0$$

- $(\gamma_{h,1})_{h\geq 1},\ (\gamma_{h,2})_{h\geq 1},\ (\gamma_{h,3})_{h\geq 1},\ \ldots$
  independent sequences of non–negative rv's such that $\sum_{h\geq 1} \gamma_{h,j} = 1$
- $(\omega_j)_{j\geq 1} \perp (G_j)_{j\geq 1}$ and $\sum_j \omega_j = 1$
- P_0 is a non–atomic probability measure

Finally

$$\text{Prob}[\{\tilde{p}_1 = G_h\} \cap \{\tilde{p}_2 = G_\kappa\} \mid \tilde{p}_0] = \omega_h\, \omega_\kappa \qquad h, \kappa \geq 1$$

Latent nested processes

Latent nested construction

Work with the set $\mathsf{M}_{\mathbb{X}}$ of boundedly finite measures on $\mathbb{X}$

Latent nested construction

Work with the set $\mathsf{M}_{\mathbb{X}}$ of boundedly finite measures on $\mathbb{X}$

Latent nested processes

(1) Nested component: discrete random probability measure on $\mathsf{M}_{\mathbb{X}}$

$$\tilde{q} = \sum_{j \geq 1} \omega_j \, \delta_{m_j}$$

- $m_j = \sum_{h \geq 1} J_{h,j} \, \delta_{\theta_{h,j}} \in \mathsf{M}_{\mathbb{X}}$
- $\theta_{h,j} \overset{\text{iid}}{\sim} P_0$

(2) Shared component: discrete random probability measure on $\mathbb{X}$

$$\mu_S = \sum_{j \geq 1} J_j^S \, \delta_{\theta_j^S} \in \mathsf{M}_{\mathbb{X}} \qquad \text{with} \qquad \theta_j^S \overset{\text{iid}}{\sim} P_0$$

Latent nested construction (ctd)

$$(\mu_1, \mu_2) \mid \tilde{q} \overset{\text{iid}}{\sim} \tilde{q} \quad \Longleftrightarrow \quad \text{Prob}[\{\mu_1 = m_h\} \cap \{\mu_2 = m_\kappa\} \mid \tilde{q}] = \omega_h \, \omega_\kappa$$

Latent nested construction (ctd)

$$(\mu_1, \mu_2) \mid \tilde{q} \overset{\text{iid}}{\sim} \tilde{q} \quad \Longleftrightarrow \quad \text{Prob}[\{\mu_1 = m_h\} \cap \{\mu_2 = m_\kappa\} \mid \tilde{q}] = \omega_h \, \omega_\kappa$$

How to choose $(\omega_j)_{j\geq 1}$, $(J_{h,j})_{h,j\geq 1}$, $(J_j^S)_{j\geq 1}$ to achieve

- analytical tractability
- modeling flexibility

Latent nested construction (ctd)

$$(\mu_1, \mu_2) \mid \tilde{q} \overset{\text{iid}}{\sim} \tilde{q} \quad \Longleftrightarrow \quad \text{Prob}[\{\mu_1 = m_h\} \cap \{\mu_2 = m_\kappa\} \mid \tilde{q}] = \omega_h \, \omega_\kappa$$

How to choose $(\omega_j)_{j \geq 1}$, $(J_{h,j})_{h,j \geq 1}$, $(J_j^S)_{j \geq 1}$ to achieve

- analytical tractability
- modeling flexibility

Points of **Poisson processes** that induce **completely random measures** (CRMs)

The role of CRMs

- $\mathscr{P} = \{(J_j, m_j) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathrm{M}_{\mathbb{X}}$ such that

$$\mathrm{card}(\mathscr{P} \cap A) \sim \mathrm{Po}(\nu(A)) \qquad \nu(A) = \int_A \rho(s)\,\mathrm{d}s\, c\, H(\mathrm{d}m)$$

  and $\omega_j = J_j / \sum_j J_j$

The role of CRMs

- $\mathscr{P} = \{(J_j, m_j) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathrm{M}_{\mathbb{X}}$ such that

$$\text{card}(\mathscr{P} \cap A) \sim \text{Po}(\nu(A)) \qquad \nu(A) = \int_A \rho(s)\,\mathrm{d}s\, c\, H(\mathrm{d}m)$$

  and $\omega_j = J_j / \sum_j J_j$

- $\mathscr{P}_j = \{(J_{h,j}, \theta_{h,j}) : h \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\text{card}(\mathscr{P}_j \cap A) \sim \text{Po}(\nu_0(A)) \qquad \nu_0(A) = \int_A \rho_0(s)\,\mathrm{d}s\, c_0\, P_0(\mathrm{d}x)$$

The role of CRMs

- $\mathscr{P} = \{(J_j, m_j) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathsf{M}_{\mathbb{X}}$ such that

$$\mathrm{card}(\mathscr{P} \cap A) \sim \mathrm{Po}(\nu(A)) \qquad \nu(A) = \int_A \rho(s)\,\mathrm{d}s\, c\, H(\mathrm{d}m)$$

  and $\omega_j = J_j / \sum_j J_j$

- $\mathscr{P}_j = \{(J_{h,j}, \theta_{h,j}) : h \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\mathrm{card}(\mathscr{P}_j \cap A) \sim \mathrm{Po}(\nu_0(A)) \qquad \nu_0(A) = \int_A \rho_0(s)\,\mathrm{d}s\, c_0\, P_0(\mathrm{d}x)$$

- $\mathscr{P}_S = \{(J_j^S, \theta_j^S) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\mathrm{card}(\mathscr{P}_S \cap A) \sim \mathrm{Po}(\nu_S(A)) \qquad \nu_S(A) = \int_A \rho_0(s)\,\mathrm{d}s\, \gamma c_0\, P_0(\mathrm{d}x)$$

Latent nested processes

Homogeneous CRMs: the measures

$$\mu = \sum_{j \ge 1} J_j \, \delta_{m_j} \qquad m_j = \sum_{h \ge 1} J_{h,j} \, \delta_{\theta_{h,j}} \qquad \mu_S = \sum_{j \ge 1} J_j^S \, \delta_{\theta_j^S}$$

are *completely random* and we use the notation

$$\mu \sim \mathrm{CRM}(\rho, c; H) \qquad m_j \stackrel{\mathrm{iid}}{\sim} \mathrm{CRM}(\rho_0, c_0; P_0) \qquad \mu_S \sim \mathrm{CRM}(\rho_0, \gamma c_0; P_0)$$

Latent nested processes

Homogeneous CRMs: the measures

$$\mu = \sum_{j \geq 1} J_j \, \delta_{m_j} \qquad m_j = \sum_{h \geq 1} J_{h,j} \, \delta_{\theta_{h,j}} \qquad \mu_S = \sum_{j \geq 1} J_j^S \, \delta_{\theta_j^S}$$

are *completely random* and we use the notation

$$\mu \sim \mathrm{CRM}(\rho, c; H) \qquad m_j \overset{\text{iid}}{\sim} \mathrm{CRM}(\rho_0, c_0; P_0) \qquad \mu_S \sim \mathrm{CRM}(\rho_0, \gamma c_0; P_0)$$

Latent nested process (Camerlenghi et al., 2018)

If $(\mu_1, \mu_2) | \tilde{q} \sim \tilde{q}^2$, the vector $(\tilde{p}_1, \tilde{p}_2)$ with

$$\tilde{p}_1 = \frac{\mu_1 + \mu_S}{\mu_1(\mathbb{X}) + \mu_S(\mathbb{X})} \qquad \tilde{p}_2 = \frac{\mu_2 + \mu_S}{\mu_2(\mathbb{X}) + \mu_S(\mathbb{X})}$$

is a *latent nested process* and denoted as $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$

Some special cases

$\gamma = 0 \implies \mu_S$ degenerates on the null measure (no shared component)

Some special cases

$\gamma = 0 \implies \mu_S$ degenerates on the null measure (no shared component)

$\implies (\tilde{p}_1, \tilde{p}_2) =$ the standard nested process

Some special cases

$\gamma = 0 \implies \mu_S$ degenerates on the null measure (no shared component)

$\implies (\tilde{p}_1, \tilde{p}_2) =$ the standard nested process

Nested Dirichlet process (Rodríguez et al., 2008)

If $\rho_0(s) = \frac{e^{-s}}{s} = \rho(s)$ and $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(0, c_0, \rho_0, c, \rho)$ then

$$(\tilde{p}_1, \tilde{p}_2) = \text{ nested Dirichlet process}$$

Some special cases

$\gamma = 0 \implies \mu_S$ degenerates on the null measure (no shared component)

$\implies (\tilde{p}_1, \tilde{p}_2) =$ the standard nested process

Nested Dirichlet process (Rodríguez et al., 2008)

If $\rho_0(s) = \frac{\mathrm{e}^{-s}}{s} = \rho(s)$ and $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(0, c_0, \rho_0, c, \rho)$ then

$$(\tilde{p}_1, \tilde{p}_2) = \text{ nested Dirichlet process}$$

Latent nested stable process (Camerlenghi et al., 2018)

For σ and σ_0 in $(0, 1)$, let

$$\rho_0(s) = \frac{\sigma_0}{\Gamma(1-\sigma_0)}\, s^{-1-\sigma_0}, \quad \rho(s) = \frac{\sigma}{\Gamma(1-\sigma)}\, s^{-1-\sigma}$$

and $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$(\tilde{p}_1, \tilde{p}_2) = \text{ latent nested stable process}$$

Discrete random probabilities

- μ_1 and μ_2 are independently drawn from

$$\tilde{q} = \sum_{j \geq 1} \omega_j \, \delta_{m_j}$$

Discrete random probabilities

- μ_1 and μ_2 are independently drawn from

$$\tilde{q} = \sum_{j \ge 1} \omega_j \, \delta_{m_j}$$

- Conditional on $\mu_1 = m_{i_1}$ and $\mu_2 = m_{i_2}$ one has

$$\tilde{p}_1 = \sum_{h \ge 1} \omega_{h,i_1} \delta_{\theta_{h,i_1}} + \sum_{j \ge 1} \omega_j^S \, \delta_{\theta_j^S}$$

$$\tilde{p}_2 = \sum_{h \ge 1} \omega_{h,i_2} \delta_{\theta_{h,i_2}} + \sum_{j \ge 1} \omega_j^S \, \delta_{\theta_j^S}$$

Discrete random probabilities

- μ_1 and μ_2 are independently drawn from

$$\tilde{q} = \sum_{j \geq 1} \omega_j \, \delta_{m_j}$$

- Conditional on $\mu_1 = m_{i_1}$ and $\mu_2 = m_{i_2}$ one has

$$\tilde{p}_1 = \sum_{h \geq 1} \omega_{h,i_1} \delta_{\theta_{h,i_1}} + \sum_{j \geq 1} \omega_j^S \, \delta_{\theta_j^S}$$

$$\tilde{p}_2 = \sum_{h \geq 1} \omega_{h,i_2} \delta_{\theta_{h,i_2}} + \sum_{j \geq 1} \omega_j^S \, \delta_{\theta_j^S}$$

- Atoms are
  - group–specific $(\theta_{h,i_1})_h$ and $(\theta_{h,i_2})_h$
  - shared $(\theta_j^S)_j$

Discrete random probabilities

- μ_1 and μ_2 are independently drawn from

$$\tilde{q} = \sum_{j \geq 1} \omega_j \, \delta_{m_j}$$

- Conditional on $\mu_1 = m_{i_1}$ and $\mu_2 = m_{i_2}$ one has

$$\tilde{p}_1 = \sum_{h \geq 1} \omega_{h,i_1} \delta_{\theta_{h,i_1}} + \sum_{j \geq 1} \omega_j^S \, \delta_{\theta_j^S}$$

$$\tilde{p}_2 = \sum_{h \geq 1} \omega_{h,i_2} \delta_{\theta_{h,i_2}} + \sum_{j \geq 1} \omega_j^S \, \delta_{\theta_j^S}$$

- Atoms are
  - group–specific $(\theta_{h,i_1})_h$ and $(\theta_{h,i_2})_h$
  - shared $(\theta_j^S)_j$
- In a partially exchangeable setting: ties
  - within samples
  - between samples

Partially exchangeable samples

$$(X_{i,1}, X_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) \overset{\text{ind}}{\sim} \tilde{p}_1 \times \tilde{p}_2$$
$$(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$$

- Two samples $\boldsymbol{X}_1 = \{X_{1,1}, \ldots, X_{N_1,1}\}$ and $\boldsymbol{X}_2 = \{X_{1,2}, \ldots, X_{N_2,2}\}$
- Random partition Ψ_N of $[N] = \{1, \ldots, N_1 + N_2\}$ induced by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ into
  - k_1 clusters $C_{1,1}, \ldots, C_{k_1,1}$ specific to $\boldsymbol{X}_1$ with $\mathrm{card}(C_{j,1}) = n_{j,1}$
  - k_2 clusters $C_{1,2}, \ldots, C_{k_2,2}$ specific to $\boldsymbol{X}_2$ with $\mathrm{card}(C_{j,2}) = n_{j,2}$
  - k_0 shared clusters $C_1, \ldots, C_{k_0}$ with $\mathrm{card}(C_j) = q_{j,1} + q_{j,2}$

Partially exchangeable samples

$$(X_{i,1}, X_{j,2}) \,|\, (\tilde{p}_1, \tilde{p}_2) \stackrel{\text{ind}}{\sim} \tilde{p}_1 \times \tilde{p}_2$$
$$(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$$

- Two samples $\boldsymbol{X}_1 = \{X_{1,1}, \ldots, X_{N_1,1}\}$ and $\boldsymbol{X}_2 = \{X_{1,2}, \ldots, X_{N_2,2}\}$
- Random partition Ψ_N of $[N] = \{1, \ldots, N_1 + N_2\}$ induced by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ into
  - k_1 clusters $C_{1,1}, \ldots, C_{k_1,1}$ specific to $\boldsymbol{X}_1$ with $\mathrm{card}(C_{j,1}) = n_{j,1}$
  - k_2 clusters $C_{1,2}, \ldots, C_{k_2,2}$ specific to $\boldsymbol{X}_2$ with $\mathrm{card}(C_{j,2}) = n_{j,2}$
  - k_0 shared clusters $C_1, \ldots, C_{k_0}$ with $\mathrm{card}(C_j) = q_{j,1} + q_{j,2}$

Partially exchangeable partition probability function (pEPPF)

$$\Pi_k^{(N)}(\boldsymbol{n}, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \mathrm{Prob}(\Psi_N = \boldsymbol{C})$$

Partition probability function

The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \text{ function of } (c_0, \rho_0)$$

Partition probability function

The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \text{ function of } (c_0, \rho_0)$$

pEPPF (Camerlenghi et al., 2018)

If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1^* \Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) + (1 - \pi_1^*) \sum_{(*)} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \boldsymbol{\zeta}^*)$$

where the sum $(*)$ runs over all vectors $\boldsymbol{\zeta}^* \in \{0,1\}^{k_1+k_2}$.

Partition probability function

The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \text{ function of } (c_0, \rho_0)$$

pEPPF (Camerlenghi et al., 2018)

If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1^* \Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) + (1 - \pi_1^*) \sum_{(*)} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \boldsymbol{\zeta}^*)$$

where the sum $(*)$ runs over all vectors $\boldsymbol{\zeta}^* \in \{0, 1\}^{k_1 + k_2}$.

- The probability π_1^* is determined by (c, ρ)

Partition probability function

The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \text{ function of } (c_0, \rho_0)$$

pEPPF (Camerlenghi et al., 2018)

If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1^* \Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) + (1 - \pi_1^*) \sum_{(*)} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \boldsymbol{\zeta}^*)$$

where the sum $(*)$ runs over all vectors $\boldsymbol{\zeta}^* \in \{0,1\}^{k_1+k_2}$.

- The probability π_1^* is determined by (c, ρ)
- The function I_γ is determined by (c_0, ρ_0)

Partition probability function

The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \text{ function of } (c_0, \rho_0)$$

pEPPF (Camerlenghi et al., 2018)

If $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) = \pi_1^* \Phi_{k,\gamma}^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2) + (1 - \pi_1^*) \sum_{(*)} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \boldsymbol{\zeta}^*)$$

where the sum $(*)$ runs over all vectors $\boldsymbol{\zeta}^* \in \{0, 1\}^{k_1 + k_2}$.

- The probability π_1^* is determined by (c, ρ)
- The function I_γ is determined by (c_0, ρ_0)
- The sum $(*)$ can be evaluated in some cases, otherwise one may sample $\boldsymbol{\zeta}^*$

Testing distributional homogeneity

Posterior probability of exchangeability across samples

$$\mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = \frac{\pi_1^* \, \Phi_{k,\gamma}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}$$

Posterior probability of exchangeability across samples

$$\mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = \frac{\pi_1^* \, \Phi_{k,\gamma}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}$$

$\Rightarrow$ The standard nested process case is recovered with $\gamma \to 0$ and

$$\lim_{\gamma \to 0} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \zeta^*) > 0 \quad \Longleftrightarrow \quad k_0 = 0$$

Posterior probability of exchangeability across samples

$$\mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = \frac{\pi_1^* \, \Phi_{k,\gamma}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}$$

$\Rightarrow$ The standard nested process case is recovered with $\gamma \to 0$ and

$$\lim_{\gamma \to 0} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \zeta^*) > 0 \qquad \Longleftrightarrow \qquad k_0 = 0$$

$\Rightarrow$ Even with a single shared cluster across samples, i.e. $k_0 > 0$, one has

$$\lim_{\gamma \to 0} \mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = 1$$

i.e. the posterior degenerates on exchangeability across samples.

Posterior probability of exchangeability across samples

$$\mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = \frac{\pi_1^* \, \Phi_{k,\gamma}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}{\Pi_k^{(N)}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2)}$$

$\Rightarrow$ The standard nested process case is recovered with $\gamma \to 0$ and

$$\lim_{\gamma \to 0} I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \zeta^*) > 0 \qquad \Longleftrightarrow \qquad k_0 = 0$$

$\Rightarrow$ Even with a single shared cluster across samples, i.e. $k_0 > 0$, one has

$$\lim_{\gamma \to 0} \mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = 1$$

i.e. the posterior degenerates on exchangeability across samples.

$\Rightarrow$ For any $\gamma > 0$, ζ^* and (k_1, k_2, k_0) one can show that

$$I_\gamma(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{q}_1, \boldsymbol{q}_2; \zeta^*) > 0$$

and the dependence structure does not degenerate to exchangeability even if shared clusters are recorded, i.e.

$$\mathrm{Prob}[\tilde{p}_1 \neq \tilde{p}_2 \,|\, \mathrm{data}] > 0 \qquad \forall k_0 \geq 0$$

Latent nested σ–stable process

σ–stable partition probability function

If $(\tilde{p}_1, \tilde{p}_2)$ is a latent nested σ–stable process, then

$$\Pi_k^{(N)}(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{\sigma_0^{k-1}\Gamma(k)}{\Gamma(N)}\, \xi_{\sigma_0} \Big\{ 1 - \sigma + \sigma\, \gamma^{k_0} \frac{B(k_1\sigma_0, k_2\sigma_0)}{B(N_1, N_2)}$$

$$\times \int_0^1 \frac{\prod_{i=1}^{k_1}(1 + \gamma w^{n_i - \sigma_0}) \prod_{j=1}^{k_2}(1 + \gamma (1-w)^{m_j - \sigma_0})}{\Big\{\gamma + w^{\sigma_0} + (1-w)^{\sigma_0}\Big\}^k} \mathrm{Beta}(\mathrm{d}w; k_1\sigma_0, k_2\sigma_0) \Big\}$$

where $\xi_{\sigma_0} = \prod (1 - \sigma_0)_{\mathrm{card}(C_j) - 1}$

Latent nested σ–stable process

σ–stable partition probability function

If $(\tilde{p}_1, \tilde{p}_2)$ is a latent nested σ–stable process, then

$$\Pi_k^{(N)}(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{\sigma_0^{k-1}\Gamma(k)}{\Gamma(N)}\, \xi_{\sigma_0} \Big\{ 1 - \sigma + \sigma\, \gamma^{k_0} \frac{B(k_1\sigma_0, k_2\sigma_0)}{B(N_1, N_2)}$$
$$\times \int_0^1 \frac{\prod_{i=1}^{k_1}(1 + \gamma w^{n_i - \sigma_0}) \prod_{j=1}^{k_2}(1 + \gamma (1-w)^{m_j - \sigma_0})}{\Big\{ \gamma + w^{\sigma_0} + (1-w)^{\sigma_0} \Big\}^k} \mathrm{Beta}(\mathrm{d}w; k_1\sigma_0, k_2\sigma_0) \Big\}$$

where $\xi_{\sigma_0} = \prod (1-\sigma_0)_{\mathrm{card}(C_j)-1}$

$\Rightarrow$ Prior probability of full exchangeability: $\pi_1^* = \mathbb{P}[\tilde{p}_1 = \tilde{p}_2] = 1 - \sigma$

Latent nested σ–stable process

σ–stable partition probability function

If $(\tilde{p}_1, \tilde{p}_2)$ is a latent nested σ–stable process, then

$$\Pi_k^{(N)}(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{\sigma_0^{k-1}\Gamma(k)}{\Gamma(N)}\, \xi_{\sigma_0} \Big\{ 1 - \sigma + \sigma\, \gamma^{k_0} \frac{B(k_1\sigma_0, k_2\sigma_0)}{B(N_1, N_2)}$$

$$\times \int_0^1 \frac{\prod_{i=1}^{k_1}(1 + \gamma w^{n_i - \sigma_0}) \prod_{j=1}^{k_2}(1 + \gamma (1-w)^{m_j - \sigma_0})}{\Big\{\gamma + w^{\sigma_0} + (1-w)^{\sigma_0}\Big\}^k} \mathrm{Beta}(\mathrm{d}w; k_1\sigma_0, k_2\sigma_0) \Big\}$$

where $\xi_{\sigma_0} = \prod (1 - \sigma_0)_{\mathrm{card}(C_j) - 1}$

$\Rightarrow$ Prior probability of full exchangeability: $\pi_1^* = \mathbb{P}[\tilde{p}_1 = \tilde{p}_2] = 1 - \sigma$

$\Rightarrow$

$$\frac{\sigma_0^{k-1}\Gamma(k)}{\Gamma(N)}\, \xi_{\sigma_0} = \text{PPF in the fully exchangeable case}$$

- Posterior probability of full exchangeability:

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) := \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \text{data}] = \frac{1-\sigma}{1-\sigma+\sigma f(\gamma, k_0)}$$

f is an increasing function of γ, for any $k_0 \geq 0$ and

$$f(0, k_0) = 0 \qquad \text{if} \qquad k_0 > 0$$

- Posterior probability of full exchangeability:

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) := \mathrm{Prob}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}] = \frac{1-\sigma}{1-\sigma+\sigma f(\gamma, k_0)}$$

  f is an increasing function of γ, for any $k_0 \geq 0$ and

$$f(0, k_0) = 0 \qquad \text{if} \qquad k_0 > 0$$

- $\gamma > 0$, for any $k_0 \geq 0$

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) < 1$$

- Posterior probability of full exchangeability:

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) := \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{1-\sigma}{1-\sigma+\sigma f(\gamma, k_0)}$$

  f is an increasing function of γ, for any $k_0 \geq 0$ and

$$f(0, k_0) = 0 \qquad \text{if} \qquad k_0 > 0$$

- $\gamma > 0$, for any $k_0 \geq 0$

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) < 1$$

- $\gamma = 0$

  $\Rightarrow$ if $k_0 = 0$ (no shared clusters), then

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) < 1$$

  $\Rightarrow$ if $k_0 > 0$ (at least one shared cluster), then

$$\pi_1^*(\boldsymbol{x}_1, \boldsymbol{x}_2) = 1$$

Illustrations

Mixture model for density estimation

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ latent variables

Mixture model for density estimation

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ latent variables
- $x \mapsto h(x; \boldsymbol{\theta})$ density on $\mathbb{X}$, i.e. for any $\boldsymbol{\theta}$

$$h(\,\cdot\,; \boldsymbol{\theta}) > 0 \qquad \int_{\mathbb{X}} h(x; \boldsymbol{\theta})\,\mathrm{d}x = 1$$

Mixture model for density estimation

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ latent variables
- $x \mapsto h(x; \boldsymbol{\theta})$ density on $\mathbb{X}$, i.e. for any $\boldsymbol{\theta}$

$$h(\,\cdot\,; \boldsymbol{\theta}) > 0 \qquad \int_{\mathbb{X}} h(x; \boldsymbol{\theta})\,\mathrm{d}x = 1$$

Mixture model

$$\begin{aligned}
(X_{i,1}, X_{j,2}) \mid (\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{j,2}) &\overset{\text{ind}}{\sim} h(\,\cdot\,; \boldsymbol{\theta}_{i,1}) \times h(\,\cdot\,; \boldsymbol{\theta}_{j,2}) \\
(\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\
(\tilde{p}_1, \tilde{p}_2) &\sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)
\end{aligned}$$

Mixture model for density estimation

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ latent variables
- $x \mapsto h(x; \boldsymbol{\theta})$ density on $\mathbb{X}$, i.e. for any $\boldsymbol{\theta}$

$$h(\,\cdot\,; \boldsymbol{\theta}) > 0 \qquad \int_{\mathbb{X}} h(x; \boldsymbol{\theta})\, \mathrm{d}x = 1$$

Mixture model

$$\begin{aligned}
(X_{i,1}, X_{j,2}) \,|\, (\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{j,2}) &\stackrel{\text{ind}}{\sim} h(\,\cdot\,; \boldsymbol{\theta}_{i,1}) \times h(\,\cdot\,; \boldsymbol{\theta}_{j,2}) \\
(\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{j,2}) \,|\, (\tilde{p}_1, \tilde{p}_2) &\stackrel{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \\
(\tilde{p}_1, \tilde{p}_2) &\sim \mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)
\end{aligned}$$

Goals:

- Estimate the individual population densities $\tilde{f}_\ell(x) = \int_\Theta h(x; \boldsymbol{\theta})\, \tilde{p}_\ell(\mathrm{d}\boldsymbol{\theta})$
- Infer the clustering of the data, i.e. $\sharp\{\text{clusters}\}$ or $\sharp\{\text{shared clusters}\}$ or ...
- Discuss homogeneity across samples, namely full vs partial exchangeability

Prior specification, with latent stable nested processes

Prior specification, with latent stable nested processes

- $h(\,\cdot\,; (M, V))$ is Gaussian with mean M and variance V
- Base measure is the usual normal/inverse–gamma

$$P_0(\mathrm{d}M, \mathrm{d}V) = P_{0,1}(\mathrm{d}V)\, P_{0,2}(\mathrm{d}M \mid V)$$

$$P_{0,1} = \text{Inv–Ga}(s_0, S_0) \qquad P_{0,2} = \mathrm{N}(m, \tau V)$$

- $\sigma, \sigma_0 \overset{\text{iid}}{\sim} \mathrm{U}(0, 1)$.

Prior specification, with latent stable nested processes

- $h(\,\cdot\,;(M,V))$ is Gaussian with mean M and variance V
- Base measure is the usual normal/inverse–gamma

$$P_0(\mathrm{d}M,\mathrm{d}V) = P_{0,1}(\mathrm{d}V)\,P_{0,2}(\mathrm{d}M\mid V)$$

$$P_{0,1} = \text{Inv–Ga}(s_0, S_0) \qquad P_{0,2} = \mathrm{N}(m, \tau V)$$

- $\sigma, \sigma_0 \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$.

Illustrations with synthetic data

- Three simulation scenarios for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$
  - (I) Sample from two mixtures with a shared component having the same weight
  - (II) Sample from the same distribution: homogeneity across samples
  - (III) Sample from two mixtures with a shared component having different weights

Simulated data

(I) First simulation scenario

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

Simulated data

(I) First simulation scenario

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

Comparison between

- $\mathrm{LNP}(0, c_0, \rho_0, c, \rho)$

  $\Rightarrow$ $\mu_S = 0$ almost surely

  $\Rightarrow$ $\tilde{p}_\ell = \mu_\ell / \mu_\ell(\mathbb{X})$ for each $\ell = 1, 2$, where

  $$(\mu_1, \mu_2) \,|\, \tilde{q} \stackrel{\text{iid}}{\sim} \tilde{q} \quad \& \quad \tilde{q} = \tilde{\mu}/\tilde{\mu}(\mathrm{M}_{\mathbb{X}})$$

  $\Rightarrow$ If there are shared values among

  $$\boldsymbol{\theta}_1 = \{(M_{i,1}, V_{i,1}) : i = 1, \ldots, N_1\} \quad \text{and} \quad \boldsymbol{\theta}_2 = \{(M_{j,2}, V_{j,2}) : j = 1, \ldots, N_2\}$$

  then the posterior degenerates on $\{(p_1, p_2) \in \mathrm{P}_{\mathbb{X}} : \; p_1 = p_2\}$

Simulated data

(I) First simulation scenario

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

Comparison between

- $\mathrm{LNP}(0, c_0, \rho_0, c, \rho)$

  $\Rightarrow$ $\mu_S = 0$ almost surely

  $\Rightarrow$ $\tilde{p}_\ell = \mu_\ell / \mu_\ell(\mathbb{X})$ for each $\ell = 1, 2$, where

  $$(\mu_1, \mu_2) \,|\, \tilde{q} \overset{\text{iid}}{\sim} \tilde{q} \quad \& \quad \tilde{q} = \tilde{\mu}/\tilde{\mu}(\mathsf{M}_{\mathbb{X}})$$

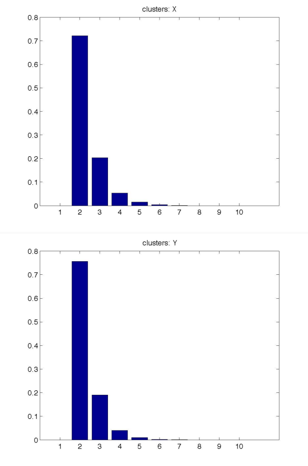
  $\Rightarrow$ If there are shared values among

  $$\boldsymbol{\theta}_1 = \{(M_{i,1}, V_{i,1}) : i = 1, \ldots, N_1\} \quad \text{and} \quad \boldsymbol{\theta}_2 = \{(M_{j,2}, V_{j,2}) : j = 1, \ldots, N_2\}$$

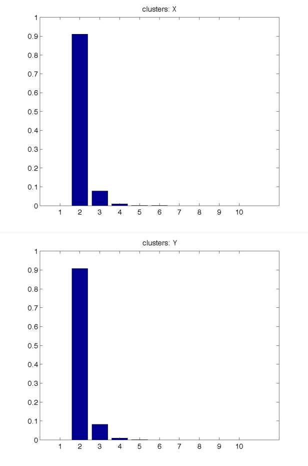
  then the posterior degenerates on $\{(p_1, p_2) \in \mathsf{P}_{\mathbb{X}} : \; p_1 = p_2\}$

- $\mathrm{LNP}(\gamma, c_0, \rho_0, c, \rho)$

Posterior of the number of clusters with $N_1 = N_2 = 50$

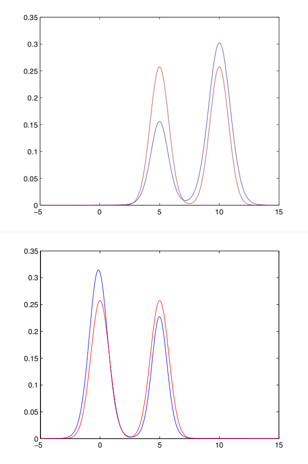


$LNP({\color{red}0}, c_0, \rho_0, c, \rho)$

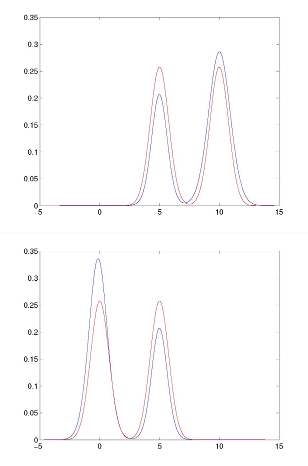


$LNP({\color{red}\gamma}, c_0, \rho_0, c, \rho)$

Density estimates with $N_1 = N_2 = 50$

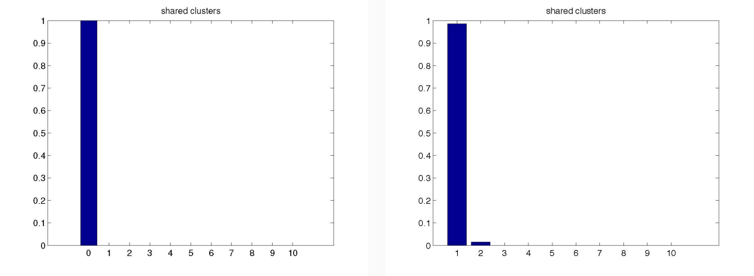


$LNP(0, c_0, \rho_0, c, \rho)$



$LNP(\gamma, c_0, \rho_0, c, \rho)$

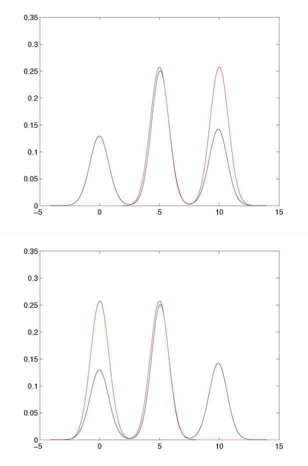
True density and estimated density



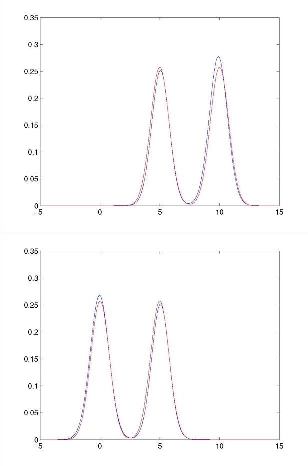
Posterior distribution of the number of shared clusters: LNP with $\gamma = 0$ *(left) and LNP with* $\gamma > 0$ *(right)*

- The latent nested process with $\gamma = 0$ identifies $\tilde{p}_1$ and $\tilde{p}_2$ as being distinct: the posterior distribution and the number of shared components which is degenerate at 0
- When the sample size is increased to $N_1 = N_2 = 100$, the latent nested model with $\gamma = 0$ identifies a common component and the density estimates get worse.

Density estimates with $N_1 = N_2 = 100$



$LNP(0, c_0, \rho_0, c, \rho)$



$LNP(\gamma, c_0, \rho_0, c, \rho)$

True density and estimated density

Simulated data

(II) Second simulation scenario

$$X_1 \overset{\mathrm{d}}{=} X_2 \sim \frac{1}{2}\,\mathrm{N}(0,1) + \frac{1}{2}\,\mathrm{N}(5,1).$$

Simulated data

(II) Second simulation scenario

$$X_1 \overset{\mathrm{d}}{=} X_2 \sim \frac{1}{2}\,\mathrm{N}(0,1) + \frac{1}{2}\,\mathrm{N}(5,1).$$

Clustering: $K_1 = \sharp\{\text{clusters specific to } \boldsymbol{X}_1\}$ $K_2 = \sharp\{\text{clusters specific to } \boldsymbol{X}_2\}$
$K_{1,2} = \sharp\{\text{clusters shared by } \boldsymbol{X}_1 \text{ and } \boldsymbol{X}_2\}$

# comp.	0	1	2	3	4	5	6	≥ 7
K_1	0	0	**0.638**	0.232	0.079	0.029	0.012	0.008
K_2	0	0	**0.635**	0.235	0.083	0.029	0.011	0.007
$K_{1,2}$	0	0	**0.754**	0.187	0.045	0.012	0.002	0.001

Simulated data

(II) Second simulation scenario

$$X_1 \stackrel{\mathrm{d}}{=} X_2 \sim \frac{1}{2}\,\mathrm{N}(0,1) + \frac{1}{2}\,\mathrm{N}(5,1).$$

Clustering: $K_1 = \sharp\{\text{clusters specific to } \boldsymbol{X}_1\}$ $K_2 = \sharp\{\text{clusters specific to } \boldsymbol{X}_2\}$

$K_{1,2} = \sharp\{\text{clusters shared by } \boldsymbol{X}_1 \text{ and } \boldsymbol{X}_2\}$

# comp.	0	1	2	3	4	5	6	≥ 7
K_1	0	0	**0.638**	0.232	0.079	0.029	0.012	0.008
K_2	0	0	**0.635**	0.235	0.083	0.029	0.011	0.007
$K_{1,2}$	0	0	**0.754**	0.187	0.045	0.012	0.002	0.001

Testing: Bayes factor for testing $H_0: p_1 = p_2$ vs $H_0: p_1 \neq p_2$

$$\mathrm{BF} = \frac{\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \,|\, \text{data}]}{\mathbb{P}[\tilde{p}_1 \neq \tilde{p}_2 \,|\, \text{data}]} \, \frac{1-\pi_1^*}{\pi_1^*} \approx 5.85$$

Simulated data

(III) Third simulation scenario

$$X_1 \sim 0.9\,\mathrm{N}(5, 0.6) + 0.1\,\mathrm{N}(10, 0.6) \quad X_2 \sim 0.1\,\mathrm{N}(5, 0.6) + 0.9\,\mathrm{N}(0, 0.6).$$

Simulated data

(III) Third simulation scenario

$$X_1 \sim 0.9\,\mathrm{N}(5, 0.6) + 0.1\,\mathrm{N}(10, 0.6) \quad X_2 \sim 0.1\,\mathrm{N}(5, 0.6) + 0.9\,\mathrm{N}(0, 0.6).$$

Clustering:

# comp.	0	1	2	3	4	5	6	≥ 7
K_1	0	0	**0.679**	0.232	0.065	0.018	0.004	0.002
K_2	0	0	**0.778**	0.185	0.032	0.004	0.001	0
$K_{1,2}$	0	**0.965**	0.034	0.001	0	0	0	0

Simulated data

(III) Third simulation scenario

$$X_1 \sim 0.9\,\mathrm{N}(5, 0.6) + 0.1\,\mathrm{N}(10, 0.6) \quad X_2 \sim 0.1\,\mathrm{N}(5, 0.6) + 0.9\,\mathrm{N}(0, 0.6).$$

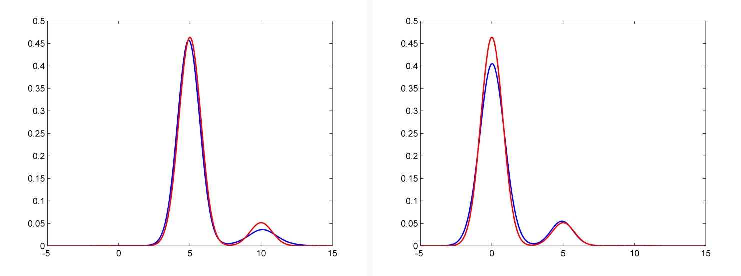
Clustering:

# comp.	0	1	2	3	4	5	6	≥ 7
K_1	0	0	**0.679**	0.232	0.065	0.018	0.004	0.002
K_2	0	0	**0.778**	0.185	0.032	0.004	0.001	0
$K_{1,2}$	0	**0.965**	0.034	0.001	0	0	0	0

Testing: Bayes factor for testing $H_0: \ p_1 = p_2$

$$\mathrm{BF} = \frac{\mathbb{P}[\tilde{p}_1 = \tilde{p}_2 \,|\, \mathrm{data}]}{\mathbb{P}[\tilde{p}_1 \neq \tilde{p}_2 \,|\, \mathrm{data}]} \, \frac{1 - \pi_1^*}{\pi_1^*} \approx 0.00022$$

Density estimates



Estimated (blue) and true densities (red) for $\boldsymbol{X}_1$ in Panel (a) and $\boldsymbol{X}_2$ in Panel (b).

Concluding remarks

- Dealing with $d > 2$ samples is more challenging
- Currently working on a hybrid hierarchical/nested model that
  - allows to deal with more than 2 samples
  - does not degenerate to exchangeability if at least one observation is shared across samples

  The main idea is to take an atomic base measure at the root of the nested model.
  Ongoing joint work with I. Prünster and G. Rebaudo (PhD student @Bocconi)

Concluding remarks

- Dealing with $d > 2$ samples is more challenging
- Currently working on a hybrid hierarchical/nested model that
  - allows to deal with more than 2 samples
  - does not degenerate to exchangeability if at least one observation is shared across samples

  The main idea is to take an atomic base measure at the root of the nested model. Ongoing joint work with I. Prünster and G. Rebaudo (PhD student @Bocconi)

Reference

- Camerlenghi, Dunson, L., Rodríguez & Prünster (2018). Latent nested nonparametric priors. *arXiv:1801.05048*. Submitted.

THANK YOU !