Nonparametric priors for covariate-dependent data

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

Antonio Lijoi

Bocconi University and Institute of Data Science and Analytics, Milano
- 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 $z \in \mathcal{Z}$
- Data $(X_{j,z})_{j \geq 1}$
− Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

More generally:

− Covariate $z \in \mathbb{Z}$
− Data $(X_j, z)_{j \geq 1}$

<table>
<thead>
<tr>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>For any $z_1 \neq z_2$</td>
</tr>
<tr>
<td>$\text{Prob}[X_i, z_1 \in A, X_j, z_2 \in B] \neq \text{Prob}[X_i, z_1 \in B, X_j, z_2 \in A]$</td>
</tr>
</tbody>
</table>
Several applied settings characterized by data heterogeneity: multi-centre studies, topic modelling, Genomics, ...

More generally:

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

Heterogeneity

For any $z_1 \neq z_2$

$$\text{Prob}[X_{i, z_1} \in A, X_{j, z_2} \in B] \neq \text{Prob}[X_{i, z_1} \in B, X_{j, z_2} \in A]$$

- Which dependence between $X_{i, z_1}$ and $X_{j, z_2}$?
Multiple samples and partial exchangeability

\[ Z := \{1, \ldots, d\} \implies X_1 = (X_{j,1})_{j \geq 1}, \ldots, X_d = (X_{j,d})_{j \geq 1} \]
Multiple samples and partial exchangeability

\[ \mathcal{Z} := \{1, \ldots, d\} \implies X_1 = (X_{j,1})_{j \geq 1}, \ldots, X_d = (X_{j,d})_{j \geq 1} \]

Partial exchangeability

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

\[
(X_1, \ldots, X_d) \overset{d}{=} (\pi_1 X_1, \ldots, \pi_d X_d)
\]
Multiple samples and partial exchangeability

\[ \mathcal{Z} := \{1, \ldots, d\} \implies \mathbf{X}_1 = (X_{j,1})_{j \geq 1}, \ldots, \mathbf{X}_d = (X_{j,d})_{j \geq 1} \]

Partial exchangeability

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

\[ (\mathbf{X}_1, \ldots, \mathbf{X}_d) \overset{\text{d}}{=} (\pi_1 \mathbf{X}_1, \ldots, \pi_d \mathbf{X}_d) \]

- Homogeneity within each sample

\[
\begin{align*}
\text{Prob}[X_{1,1} = 0, X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0] \\
= \text{Prob}[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 0]
\end{align*}
\]
Multiple samples and partial exchangeability

\[ Z := \{1, \ldots, d\} \quad \Rightarrow \quad X_1 = (X_{j,1})_{j \geq 1}, \ldots, X_d = (X_{j,d})_{j \geq 1} \]

Partial exchangeability

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

\[ (X_1, \ldots, X_d) \overset{d}{=} (\pi_1 X_1, \ldots, \pi_d X_d) \]

- Homogeneity within each sample

\[
\prob[X_{1,1} = 0, X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0] = \prob[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 0]
\]

- Lack of homogeneity across different samples

\[
\prob[X_{1,1} = 0, X_{1,2} = 1, X_{2,1} = 0, X_{2,2} = 1, X_{2,3} = 0] \neq \prob[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1, X_{2,2} = 0, X_{2,3} = 1]
\]
Hierarchical model representation

**Representation theorem**

\( \{x_i : i = 1, \ldots, d\} \) is partially exchangeable if and only if

\[
\text{Prob}[x_1 \in A_1, \ldots, x_d \in A_d] = \int_{P_X^d} \prod_{i=1}^d p_i^{(\infty)}(A_i) \ Q_d(dp_1, \ldots, dp_d)
\]

where \( P_X \) is the space of probability measures on the sample space \( X \).
Hierarchical model representation

**Representation theorem**

\[ \{X_i : i = 1, \ldots, d\} \text{ is partially exchangeable if and only if} \]

\[ \text{Prob}[X_1 \in A_1, \ldots, X_d \in A_d] = \int_{P_x} \prod_{i=1}^{d} p_i^{(\infty)}(A_i) \, Q_d(dp_1, \ldots, dp_d) \]

where \( P_x \) is the space of probability measures on the sample space \( X \)

Case \( d = 2 \) samples

\( (X_{i,1}, X_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2 \)

\( (\tilde{p}_1, \tilde{p}_2) \sim Q_2 \)

and \( Q_2 \) acts as a prior distribution on \( P_x^2 \)
Compositions of random measures \((d = 2)\)

- **Choice of \(Q_2\)?** Compositions of random probability measures

\[
(X_{i,1}, X_{j,2}) | (\tilde{p}_1, \tilde{p}_2) \overset{iid}{\sim} \tilde{p}_1 \times \tilde{p}_2
\]

\[
\tilde{p}_1, \tilde{p}_2 | \tilde{p}_0 \overset{iid}{\sim} Q(\cdot | \tilde{p}_0)
\]

\[
\tilde{p}_0 \sim Q_0
\]
Compositions of random measures \((d = 2)\)

- **Choice of \(Q_2\)?** Compositions of random probability measures
  
  \[
  (X_{i,1}, X_{j,2}) | (\bar{p}_1, \bar{p}_2) \overset{iid}{\sim} \bar{p}_1 \times \bar{p}_2 \\
  \bar{p}_1, \bar{p}_2 | \bar{p}_0 \overset{iid}{\sim} Q(\cdot | \bar{p}_0) \\
  \bar{p}_0 \sim Q_0
  \]

- **Extreme cases of dependence induced by the prior \(Q_2\)**
  
  - **Full exchangeability:** \(Q_2(\{(p_1, p_2) \in P_X^2 : p_1 = p_2\}) = 1\)
  
  - **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

\[
(X_{i,1}, X_{j,2}) | (\tilde{p}_1, \tilde{p}_2) \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2
\]

\[
\tilde{p}_1, \tilde{p}_2 | \tilde{p}_0 \overset{\text{iid}}{\sim} Q(\cdot | \tilde{p}_0)
\]

\[
\tilde{p}_0 \sim Q_0
\]

- **Extreme cases of dependence induced by the prior \(Q_2\)**
  - **Full exchangeability:** \(Q_2(\{(p_1, p_2) \in P_2^X : p_1 = p_2\}) = 1\)
  - **Maximal heterogeneity:** \(Q_2(A_1 \times A_2) = Q_1^*(A_1) \cdot Q_1^{**}(A_2)\)

Possible compositions through

1. Hierarchical structure (Igor’s talk)
2. Nested structure
If $Q_0$ is a probability measure on $P_{P_X}$

\[ \tilde{\rho}_1, \tilde{\rho}_2 \mid \tilde{\rho}_0 \overset{\text{iid}}{\sim} \tilde{\rho}_0, \]

\[ \tilde{\rho}_0 \sim Q_0 \]
Nested random probabilities

If $Q_0$ is a probability measure on $P_{P_{\mathcal{X}}}$

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

$$\tilde{p}_0 \sim Q_0$$

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 $P_{\tilde{X}}$

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

Discrete case

\[
\tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \\
G_j = \sum_{h \geq 1} \gamma_{h,j} \delta_{\theta_{h,j}} \\
\theta_{h,j} \overset{\text{iid}}{\sim} P_0
\]
Nested random probabilities

If $Q_0$ is a probability measure on $P_{P_X}$

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

Discrete case

\[ \tilde{p}_0 = \sum_{j \geq 1} \omega_j \delta_{G_j} \quad G_j = \sum_{h \geq 1} \gamma_{h,j} \delta_{\theta_{h,j}} \quad \theta_{h,j} \overset{\text{iid}}{\sim} P_0 \]
Nested structure

**Nested random probabilities**

If $Q_0$ is a probability measure on $P_{P_X}$

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

**Discrete case**

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

- $(\gamma_{h,1})_{h \geq 1}, (\gamma_{h,2})_{h \geq 1}, (\gamma_{h,3})_{h \geq 1}, \ldots$
  - independent sequences of non-negative rv's such that $\sum_{h \geq 1} \gamma_{h,j} = 1$
If $Q_0$ is a probability measure on $P_{\mathcal{X}}$

$$\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \sim \tilde{p}_0,$$

$$\tilde{p}_0 \sim Q_0$$

**Discrete case**

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

- $(\gamma_{h,1})_{h \geq 1}, (\gamma_{h,2})_{h \geq 1}, (\gamma_{h,3})_{h \geq 1}, \ldots$
  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 $\sum_j \omega_j = 1$
Nested structure

Nested random probabilities

If $Q_0$ is a probability measure on $P_{P_X}$

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

$$\tilde{p}_0 \sim Q_0$$

Discrete case

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

- $(\gamma_{h,1})_{h \geq 1}, (\gamma_{h,2})_{h \geq 1}, (\gamma_{h,3})_{h \geq 1}, \ldots$ 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 $P_{P_X}$

\[
\tilde{p}_1, \tilde{p}_2 \mid \tilde{p}_0 \overset{iid}{\sim} \tilde{p}_0, \\
\tilde{p}_0 \sim Q_0
\]

**Discrete case**

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

- $(\gamma_{h,1})_{h \geq 1}, (\gamma_{h,2})_{h \geq 1}, (\gamma_{h,3})_{h \geq 1}, \ldots$
  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

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

Work with the set $M_X$ of boundedly finite measures on $X$. 

Nested component: discrete random probability measure on $M_X$

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

$\triangleright m_j = \sum_{h \geq 1} J_{h,j} \delta_{\theta_{h,j}} \in M_X$

Shared component: discrete random probability measure on $X$

$\mu_S = \sum_{j \geq 1} J_{S,j} \delta_{\theta_{S,j}} \in M_X$

with $\theta_{S,j} \text{iid} \sim P_0$
Latent nested construction

Work with the set $\mathcal{M}_X$ of boundedly finite measures on $X$.

**Latent nested processes**

(1) **Nested component**: discrete random probability measure on $\mathcal{M}_X$

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

- $m_j = \sum_{h \geq 1} J_{h,j} \delta_{\theta_{h,j}} \in \mathcal{M}_X$
- $\theta_{h,j} \overset{iid}{\sim} P_0$

(2) **Shared component**: discrete random probability measure on $X$

$$\mu_S = \sum_{j \geq 1} J_j^{S} \delta_{\theta_j^S} \in \mathcal{M}_X \quad \text{with} \quad \theta_j^S \overset{iid}{\sim} P_0$$
Latent nested construction (ctd)

$$(\mu_1, \mu_2) \mid \tilde{q} \sim \tilde{q} \iff \text{Prob}[\{\mu_1 = m_h\} \cap \{\mu_2 = m_\kappa\} \mid \tilde{q}] = \omega_h \omega_\kappa$$
Latent nested construction (ctd)

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

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

- analytical tractability
- modeling flexibility
Latent nested construction (ctd)

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

How to choose \((\omega_j)_{j \geq 1}, (J_{h,j})_{h,j \geq 1}, (J^S_j)_{j \geq 1}\) to achieve

- analytical tractability
- modeling flexibility

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

\( \mathcal{P} = \{(J_j, m_j) : j \geq 1\} \) sequence of points in \( \mathbb{R}^+ \times M_X \) such that

\[
\text{card}(\mathcal{P} \cap A) \sim \text{Po}(\nu(A)) \quad \nu(A) = \int_A \rho(s) \, ds \, c \, H(dm)
\]

and \( \omega_j = J_j / \sum_j J_j \)
The role of CRMs

\[ \mathcal{P} = \{(J_j, m_j) : j \geq 1\} \text{ sequence of points in } \mathbb{R}^+ \times M_X \text{ such that } \]
\[ \text{card}(\mathcal{P} \cap A) \sim \text{Po}(\nu(A)) \quad \nu(A) = \int_A \rho(s) \, ds \, c \, H(dm) \]
and \( \omega_j = J_j / \sum_j J_j \)

\[ \mathcal{P}_j = \{(J_{h,j}, \theta_{h,j}) : h \geq 1\} \text{ sequence of points in } \mathbb{R}^+ \times X \text{ such that } \]
\[ \text{card}(\mathcal{P}_j \cap A) \sim \text{Po}(\nu_0(A)) \quad \nu_0(A) = \int_A \rho_0(s) \, ds \, c_0 \, P_0(dx) \]
The role of CRMs

- $\mathcal{P} = \{(J_j, m_j) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times M_X$ such that
  
  \[ \text{card}(\mathcal{P} \cap A) \sim \text{Po}(\nu(A)) \quad \nu(A) = \int_A \rho(s) \, ds \, c \, H(dm) \]

  and $\omega_j = J_j / \sum_j J_j$

- $\mathcal{P}_j = \{(J_{h,j}, \theta_{h,j}) : h \geq 1\}$ sequence of points in $\mathbb{R}^+ \times X$ such that
  
  \[ \text{card}(\mathcal{P}_j \cap A) \sim \text{Po}(\nu_0(A)) \quad \nu_0(A) = \int_A \rho_0(s) \, ds \, c_0 \, P_0(dx) \]

- $\mathcal{P}_S = \{(J_j^S, \theta_j^S) : j \geq 1\}$ sequence of points in $\mathbb{R}^+ \times X$ such that
  
  \[ \text{card}(\mathcal{P}_S \cap A) \sim \text{Po}(\nu_S(A)) \quad \nu_S(A) = \int_A \rho_0(s) \, ds \, \gamma c_0 \, P_0(dx) \]
Latent nested processes

**Homogeneous CRMs:** the measures

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

are *completely random* and we use the notation

\[
\mu \sim \text{CRM}(\rho, c; H) \quad m_j \overset{\text{iid}}{\sim} \text{CRM}(\rho_0, c_0; P_0) \quad \mu_S \sim \text{CRM}(\rho_0, \gamma c_0; P_0)
\]
Latent nested processes

**Homogeneous CRMs:** the measures

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

are completely random and we use the notation

\[
\mu \sim \text{CRM}(\rho, c; H) \quad m_j \overset{iid}{\sim} \text{CRM}(\rho_0, c_0; P_0) \quad \mu_S \sim \text{CRM}(\rho_0, \gamma c_0; P_0)
\]

**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

\[
\tilde{p}_1 = \frac{\mu_1 + \mu_S}{\mu_1(X) + \mu_S(X)} \quad \tilde{p}_2 = \frac{\mu_2 + \mu_S}{\mu_2(X) + \mu_S(X)}
\]

is a latent nested process and denoted as \((\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)\)
Some special cases

\[ \gamma = 0 \implies \mu_S \text{ degenerates on the null measure (no shared component)} \]
Some special cases

\[ \gamma = 0 \implies \mu_S \text{ degenerates on the null measure (no shared component)} \]

\[ \implies (\tilde{p}_1, \tilde{p}_2) = \text{the standard nested process} \]
**Some special cases**

\[
\gamma = 0 \implies \mu_S \text{ degenerates on the null measure (no shared component)}
\]

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

**Nested Dirichlet process** (Rodríguez et al., 2008)

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

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

**Nested Dirichlet process** (Rodríguez et al., 2008)

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

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

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

For \( \sigma \) and \( \sigma_0 \) 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 \( (\tilde{\rho}_1, \tilde{\rho}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho) \), then

\( (\tilde{\rho}_1, \tilde{\rho}_2) = \) latent nested stable process
Discrete random probabilities

μ₁ and μ₂ are independently drawn from

\[ \tilde{q} = \sum_{j \geq 1} \omega_j \delta_{m_j} \]
Discrete random probabilities

- $\mu_1$ and $\mu_2$ 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
  \[ \tilde{p}_1 = \sum_{h \geq 1} \omega_{h,i_1} \delta_{\theta_{h,i_1}} + \sum_{j \geq 1} \omega^S_j \delta_{\theta^S_j} \]
  \[ \tilde{p}_2 = \sum_{h \geq 1} \omega_{h,i_2} \delta_{\theta_{h,i_2}} + \sum_{j \geq 1} \omega^S_j \delta_{\theta^S_j} \]
Discrete random probabilities

- \( \mu_1 \) and \( \mu_2 \) 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

- $\mu_1$ and $\mu_2$ 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

\[
(X_{i,1}, X_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) \overset{\text{ind}}{\sim} \tilde{p}_1 \times \tilde{p}_2
\]

\[
(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)
\]

- Two samples \(X_1 = \{X_{1,1}, \ldots, X_{N_1,1}\}\) and \(X_2 = \{X_{1,2}, \ldots, X_{N_2,2}\}\)

- Random partition \(\psi_N\) of \([N] = \{1, \ldots, N_1 + N_2\}\) induced by \(X_1\) and \(X_2\) into
  - \(k_1\) clusters \(C_{1,1}, \ldots, C_{k_1,1}\) specific to \(X_1\) with \(\text{card}(C_{j,1}) = n_{j,1}\)
  - \(k_2\) clusters \(C_{1,2}, \ldots, C_{k_2,2}\) specific to \(X_2\) with \(\text{card}(C_{j,2}) = n_{j,2}\)
  - \(k_0\) shared clusters \(C_1, \ldots, C_{k_0}\) with \(\text{card}(C_j) = q_{j,1} + q_{j,2}\)
Partially exchangeable samples

\[(X_{i,1}, X_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) \overset{\text{ind}}{\sim} \tilde{p}_1 \times \tilde{p}_2\]

\[(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)\]

- Two samples \(X_1 = \{X_{1,1}, \ldots, X_{N_1,1}\}\) and \(X_2 = \{X_{1,2}, \ldots, X_{N_2,2}\}\)
- Random partition \(\psi_N\) of \([N] = \{1, \ldots, N_1 + N_2\}\) induced by \(X_1\) and \(X_2\) into
  - \(k_1\) clusters \(C_{1,1}, \ldots, C_{k_1,1}\) specific to \(X_1\) with \(\text{card}(C_{j,1}) = n_{j,1}\)
  - \(k_2\) clusters \(C_{1,2}, \ldots, C_{k_2,2}\) specific to \(X_2\) with \(\text{card}(C_{j,2}) = n_{j,2}\)
  - \(k_0\) shared clusters \(C_1, \ldots, C_{k_0}\) with \(\text{card}(C_j) = q_{j,1} + q_{j,2}\)

Partially exchangeable partition probability function (pEPPF)

\[\Pi_k^{(N)}(n, n_2, q_1, q_2) = \text{Prob}(\psi_N = C)\]
The partition probability function corresponding to exchangeability ($\tilde{\rho}_1 = \tilde{\rho}_2$)

\[
\Phi_{k, \gamma}(n_1, n_2, q_1, q_2) = \text{function of } (c_0, \rho_0)
\]
The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi^{(N)}_{k,\gamma}(n_1, n_2, q_1, q_2) = \text{function of } (c_0, \rho_0)$$

---

**pEPPF** (Camerlenghi et al., 2018)

If $(X_1, X_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi^{(N)}_k(n_1, n_2, q_1, q_2) = \pi_1^* \Phi^{(N)}_{k,\gamma}(n_1, n_2, q_1, q_2) + (1-\pi_1^*) \sum l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*)$$

where the sum $(\ast)$ runs over all vectors $\zeta^* \in \{0, 1\}^{k_1+k_2}$. 
The partition probability function corresponding to exchangeability \((\tilde{p}_1 = \tilde{p}_2)\)

\[ \Phi_{k,\gamma}^{(N)}(n_1, n_2, q_1, q_2) = \text{function of } (c_0, \rho_0) \]

**pEPPF** (Camerlenghi et al., 2018)

If \((X_1, X_2)\) is partially exchangeable from \((\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)\), then

\[ \Pi_k^{(N)}(n_1, n_2, q_1, q_2) = \pi_1^* \Phi_{k,\gamma}^{(N)}(n_1, n_2, q_1, q_2) + (1-\pi_1^*) \sum_{(\cdot)} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*) \]

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

- The probability \(\pi_1^*\) is determined by \((c, \rho)\)
The **partition probability function** corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi^{(N)}(n_1, n_2, q_1, q_2) = \text{function of } (c_0, \rho_0)$$

---

**pEPPF** (Camerlenghi et al., 2018)

If $(X_1, X_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$
\Pi^{(N)}_k(n_1, n_2, q_1, q_2) = \pi^*_1 \Phi^{(N)}_{k, \gamma}(n_1, n_2, q_1, q_2) + (1 - \pi^*_1) \sum_{(\cdot)} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*)
$$

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

- The probability $\pi_1^*$ is determined by $(c, \rho)$
- The function $l_\gamma$ is determined by $(c_0, \rho_0)$
The partition probability function corresponding to exchangeability ($\tilde{p}_1 = \tilde{p}_2$)

$$\Phi^{(N)}_{k,\gamma}(n_1, n_2, q_1, q_2) = \text{function of } (c_0, \rho_0)$$

### pEPPF (Camerlenghi et al., 2018)

If $(X_1, X_2)$ is partially exchangeable from $(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)$, then

$$\Pi_{k}^{(N)}(n_1, n_2, q_1, q_2) = \pi_1^* \Phi^{(N)}_{k,\gamma}(n_1, n_2, q_1, q_2) + (1 - \pi_1^*) \sum_{(*)} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*)$$

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

- The probability $\pi_1^*$ is determined by $(c, \rho)$
- The function $l_{\gamma}$ is determined by $(c_0, \rho_0)$
- The sum $(*)$ can be evaluated in some cases, otherwise one may sample $\zeta^*$
Testing distributional homogeneity
Posterior probability of exchangeability across samples

\[
\text{Prob}[\tilde{\rho}_1 = \tilde{\rho}_2 \mid \text{data}] = \frac{\pi^*_1 \Phi_{k,\gamma}(n_1, n_2, q_1, q_2)}{\Pi_k^{(N)}(n_1, n_2, q_1, q_2)}
\]
Posterior probability of exchangeability across samples

\[
\text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{\pi_1^* \Phi_{k, \gamma}(n_1, n_2, q_1, q_2)}{\Pi_k^{(N)}(n_1, n_2, q_1, q_2)}
\]

⇒ The standard nested process case is recovered with \( \gamma \to 0 \) and

\[
\lim_{\gamma \to 0} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*) > 0 \iff k_0 = 0
\]
Posterior probability of exchangeability across samples

\[
\text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{\pi_1^* \Phi_{k, \gamma}(n_1, n_2, q_1, q_2)}{\prod_k^{(N)}(n_1, n_2, q_1, q_2)}
\]

⇒ The standard nested process case is recovered with \( \gamma \to 0 \) and

\[
\lim_{\gamma \to 0} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*) > 0 \iff k_0 = 0
\]

⇒ Even with a single shared cluster across samples, i.e. \( k_0 > 0 \), one has

\[
\lim_{\gamma \to 0} \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = 1
\]

i.e. the posterior degenerates on exchangeability across samples.
Posterior probability of exchangeability across samples

\[
\text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{\pi_1^* \Phi_{k,\gamma}(n_1, n_2, q_1, q_2)}{\Pi_k^{(N)}(n_1, n_2, q_1, q_2)}
\]

⇒ The standard nested process case is recovered with \( \gamma \to 0 \) and

\[
\lim_{\gamma \to 0} l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*) > 0 \iff k_0 = 0
\]

⇒ Even with a single shared cluster across samples, i.e. \( k_0 > 0 \), one has

\[
\lim_{\gamma \to 0} \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = 1
\]

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

⇒ For any \( \gamma > 0 \), \( \zeta^* \) and \((k_1, k_2, k_0)\) one can show that

\[
l_{\gamma}(n_1, n_2, q_1, q_2; \zeta^*) > 0
\]

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

\[
\text{Prob}[\tilde{p}_1 \neq \tilde{p}_2 \mid \text{data}] > 0 \quad \forall k_0 \geq 0
\]
Latent nested $\sigma$–stable process

$\sigma$–stable partition probability function

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

$$
\Pi_k^{(N)}(n, m, q_1, q_2) = \frac{\sigma^{-1}_0 \Gamma(k)}{\Gamma(N)} \xi_{\sigma_0} \left\{ 1 - \sigma + \sigma \gamma^0 \frac{B(k_1 \sigma_0, k_2 \sigma_0)}{B(N_1, N_2)} \right\}
\times \int_0^1 \prod_{i=1}^{k_1} (1 + \gamma w^{n_i - \sigma_0}) \prod_{j=1}^{k_2} (1 + \gamma (1 - w)^{m_j - \sigma_0}) \left\{ \gamma + w^{\sigma_0} + (1 - w)^{\sigma_0} \right\}^k
\times \text{Beta}(dw; k_1 \sigma_0, k_2 \sigma_0)
$$

where $\xi_{\sigma_0} = \prod (1 - \sigma_0)_{\text{card}(c_j) - 1}$
If \((\tilde{p}_1, \tilde{p}_2)\) is a latent nested \(\sigma\)–stable process, then

\[
\Pi^{(N)}_k(n, m, q_1, q_2) = \frac{\sigma_0^{k-1} \Gamma(k)}{\Gamma(N)} \xi_{\sigma_0} \left\{ 1 - \sigma + \sigma \gamma^k B(k_1 \sigma_0, k_2 \sigma_0) \right\}
\]

\[
\times \int_0^1 \frac{\prod_{i=1}^{k_1} (1 + \gamma w^{n_i - \sigma_0}) \prod_{j=1}^{k_2} (1 + \gamma (1 - w)^{m_j - \sigma_0})}{\Gamma(N)} \left\{ \gamma + w^{\sigma_0} + (1 - w)^{\sigma_0} \right\}^k \text{Beta}(dw; k_1 \sigma_0, k_2 \sigma_0)
\]

where \(\xi_{\sigma_0} = \prod (1 - \sigma_0) \text{card}(c_j) - 1\)

\[\Rightarrow \text{Prior probability of full exchangeability: } \pi_*^* = \Pr(\tilde{p}_1 = \tilde{p}_2) = 1 - \sigma\]
Latent nested $\sigma$–stable process

$\sigma$–stable partition probability function

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

$$
P_{(N)}^k(n, m, q_1, q_2) = \frac{\sigma_0^{k-1} \Gamma(k)}{\Gamma(N)} \xi_{\sigma_0} \left\{ 1 - \sigma + \sigma \gamma^k \frac{B(k_1 \sigma_0, k_2 \sigma_0)}{B(N_1, N_2)} \right\}$$

$$\times \int_0^1 \prod_{i=1}^{k_1} \frac{1 + \gamma w^{n_i - \sigma_0}}{1 + \gamma (1 - w)^{m_j - \sigma_0}} \prod_{j=1}^{k_2} \left( \frac{1 + \gamma (1 - w)^{m_j - \sigma_0}}{1 + \gamma w^{n_i - \sigma_0}} \right) \left\{ \gamma + w^{\sigma_0} + (1 - w)^{\sigma_0} \right\}^k}$$

$$\times \left\{ \frac{\Gamma(k)}{\Gamma(N)} \right\} \xi_{\sigma_0} = \prod (1 - \sigma_0) \text{card}(c_j) - 1$$

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

$\Rightarrow$

$$\frac{\sigma_0^{k-1} \Gamma(k)}{\Gamma(N)} \xi_{\sigma_0} = \text{PPF in the fully exchangeable case}$$
Posterior probability of full exchangeability:

\[ \pi_1^* (x_1, x_2) := \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{1 - \sigma}{1 - \sigma + \sigma f(\gamma, k_0)} \]

\( f \) is an increasing function of \( \gamma \), for any \( k_0 \geq 0 \) and

\[ f(0, k_0) = 0 \quad \text{if} \quad k_0 > 0 \]
Posterior probability of full exchangeability:

\[ \pi^*_1(x_1, x_2) := \text{Prob}[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}] = \frac{1 - \sigma}{1 - \sigma + \sigma f(\gamma, k_0)} \]

\( f \) is an increasing function of \( \gamma \), for any \( k_0 \geq 0 \) and

\[ f(0, k_0) = 0 \quad \text{if} \quad k_0 > 0 \]

\( \gamma > 0 \), for any \( k_0 \geq 0 \)

\[ \pi^*_1(x_1, x_2) < 1 \]
Posterior probability of full exchangeability:

\[
\pi_1^*(x_1, x_2) := \text{Prob}[\tilde{p}_1 = \tilde{p}_2 | \text{data}] = \frac{1 - \sigma}{1 - \sigma + \sigma f(\gamma, k_0)}
\]

\(f\) is an increasing function of \(\gamma\), for any \(k_0 \geq 0\) and

\[f(0, k_0) = 0 \quad \text{if} \quad k_0 > 0\]

\(\gamma > 0\), for any \(k_0 \geq 0\)

\[
\pi_1^*(x_1, x_2) < 1
\]

\(\gamma = 0\)

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

\[
\pi_1^*(x_1, x_2) < 1
\]

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

\[
\pi_1^*(x_1, x_2) = 1
\]
Illustrations
Mixture model for density estimation

- $\theta \in \Theta \subset \mathbb{R}^d$ latent variables
Mixture model for density estimation

- \( \theta \in \Theta \subset \mathbb{R}^d \) latent variables
- \( x \mapsto h(x; \theta) \) density on \( X \), i.e. for any \( \theta \)
  \[
  h(\cdot; \theta) > 0 \quad \int_X h(x; \theta) \, dx = 1
  \]
Mixture model for density estimation

- $\theta \in \Theta \subset \mathbb{R}^d$ latent variables
- $x \mapsto h(x; \theta)$ density on $\mathbb{X}$, i.e. for any $\theta$
  \[
  h(\cdot ; \theta) > 0 \quad \int_{\mathbb{X}} h(x; \theta) \, dx = 1
  \]

Mixture model

\[
(X_{i,1}, X_{j,2}) \mid (\theta_{i,1}, \theta_{j,2}) \overset{\text{ind}}{\sim} h(\cdot ; \theta_{i,1}) \times h(\cdot ; \theta_{j,2})
\]

\[
(\theta_{i,1}, \theta_{j,2}) \mid (\tilde{p}_1, \tilde{p}_2) \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2
\]

\[
(\tilde{p}_1, \tilde{p}_2) \sim \text{LNP}(\gamma, c_0, \rho_0, c, \rho)
\]
Mixture model for density estimation

- \( \theta \in \Theta \subset \mathbb{R}^d \) latent variables
- \( x \mapsto h(x; \theta) \) density on \( X \), i.e. for any \( \theta \)
  \[
  h(\cdot; \theta) > 0 \quad \int_X h(x; \theta) \, dx = 1
  \]

**Goals:**

- Estimate the individual population densities \( \tilde{f}_\ell(x) = \int_\Theta h(x; \theta) \, \tilde{p}_\ell(d\theta) \)
- Infer the clustering of the data, i.e. \#\{clusters\} or \#\{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

$$P_0(dM, dV) = P_{0,1}(dV) P_{0,2}(dM | V)$$

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

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

\[
P_0(dM, dV) = P_{0,1}(dV) P_{0,2}(dM | V)
\]

\[
P_{0,1} = \text{Inv–Ga}(s_0, S_0) \quad P_{0,2} = \text{N}(m, \tau V)
\]

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

Illustrations with synthetic data

- Three simulation scenarios for $X_1$ and $X_2$
  
  (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

\[ X_1 \sim \frac{1}{2} \mathcal{N}(5, 0.6) + \frac{1}{2} \mathcal{N}(10, 0.6) \]
\[ X_2 \sim \frac{1}{2} \mathcal{N}(5, 0.6) + \frac{1}{2} \mathcal{N}(0, 0.6). \]
Simulated data

(I) First simulation scenario

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

Comparison between

- LNP(0, c_0, \rho_0, c, \rho)
  
  \[ \Rightarrow \mu_S = 0 \text{ almost surely} \]
  \[ \Rightarrow \tilde{p}_\ell = \frac{\mu_\ell}{\mu_\ell(X)} \text{ for each } \ell = 1, 2, \text{ where} \]

\[ (\mu_1, \mu_2) \mid \tilde{q} \overset{\text{iid}}{\sim} \tilde{q} \quad \& \quad \tilde{q} = \tilde{\mu} / \tilde{\mu}(M_X) \]

\[ \Rightarrow \text{If there are shared values among} \]

\[ \theta_1 = \{(M_{i,1}, V_{i,1}) : i = 1, \ldots, N_1\} \quad \text{and} \quad \theta_2 = \{(M_{j,2}, V_{j,2}) : j = 1, \ldots, N_2\} \]

then the posterior degenerates on \( \{(p_1, p_2) \in P_X : p_1 = p_2\} \)
Simulated data

(I) First simulation scenario

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

Comparison between

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

\[ \Rightarrow \mu_S = 0 \text{ almost surely} \]
\[ \Rightarrow \bar{\mu}_\ell = \mu_\ell / \mu_\ell(X) \text{ for each } \ell = 1, 2, \text{ where} \]

\[ (\mu_1, \mu_2) \mid \tilde{q} \sim \tilde{q} \quad \& \quad \tilde{q} = \bar{\mu} / \bar{\mu}(M_X) \]

\[ \Rightarrow \text{If there are shared values among} \]
\[ \theta_1 = \{(M_{i,1}, V_{i,1}) : i = 1, \ldots, N_1\} \quad \text{and} \quad \theta_2 = \{(M_{j,2}, V_{j,2}) : j = 1, \ldots, N_2\} \]

then the posterior degenerates on \( \{(p_1, p_2) \in P_X : p_1 = p_2\} \)

- LNP(\gamma, c_0, \rho_0, c, \rho)
Posterior of the number of clusters with $N_1 = N_2 = 50$

\[ LNP(0, c_0, \rho_0, c, \rho) \]

\[ LNP(\gamma, c_0, \rho_0, c, \rho) \]
Density estimates with $N_1 = N_2 = 50$

$\text{True density and estimated density}$

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

$LNP(\gamma, c_0, \rho_0, c, \rho)$
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 $N_1 = N_2 = 100$

True density and estimated density:

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

$LNP(\gamma, c_0, \rho_0, c, \rho)$
(II) Second simulation scenario

\[ X_1 \overset{d}{=} X_2 \sim \frac{1}{2} \text{N}(0, 1) + \frac{1}{2} \text{N}(5, 1). \]
Simulated data

(II) Second simulation scenario

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

Clustering:  \( K_1 = \#\{\text{clusters specific to } X_1\} \quad K_2 = \#\{\text{clusters specific to } X_2\} \)
\( K_{1,2} = \#\{\text{clusters shared by } X_1 \text{ and } X_2\} \)

<table>
<thead>
<tr>
<th># comp.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>≥ 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>0</td>
<td>0</td>
<td>0.638</td>
<td>0.232</td>
<td>0.079</td>
<td>0.029</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>0</td>
<td>0</td>
<td>0.635</td>
<td>0.235</td>
<td>0.083</td>
<td>0.029</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>( K_{1,2} )</td>
<td>0</td>
<td>0</td>
<td>0.754</td>
<td>0.187</td>
<td>0.045</td>
<td>0.012</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Testing: Bayes factor for testing \( H_0: p_1 = p_2 \) vs \( H_0: p_1 \neq p_2 \)
\[ BF = \frac{P[\tilde{p}_1 = \tilde{p}_2 | \text{data}]}{P[\tilde{p}_1 \neq \tilde{p}_2 | \text{data}]} \approx 5.85. \]
Simulated data

(II) Second simulation scenario

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

Clustering: \( K_1 = \#\{\text{clusters specific to } X_1\} \quad K_2 = \#\{\text{clusters specific to } X_2\} \)

\[ K_{1,2} = \#\{\text{clusters shared by } X_1 \text{ and } X_2\} \]

<table>
<thead>
<tr>
<th># comp.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>≥ 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>0</td>
<td>0</td>
<td>0.638</td>
<td>0.232</td>
<td>0.079</td>
<td>0.029</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>0</td>
<td>0</td>
<td>0.635</td>
<td>0.235</td>
<td>0.083</td>
<td>0.029</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>( K_{1,2} )</td>
<td>0</td>
<td>0</td>
<td>0.754</td>
<td>0.187</td>
<td>0.045</td>
<td>0.012</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

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

\[ BF = \frac{\Pr[\tilde{p}_1 = \tilde{p}_2 \mid \text{data}]}{\Pr[\tilde{p}_1 \neq \tilde{p}_2 \mid \text{data}]} \frac{1 - \pi_1^*}{\pi_1^*} \approx 5.85 \]
(III) Third simulation scenario

\[ X_1 \sim 0.9 \mathcal{N}(5, 0.6) + 0.1 \mathcal{N}(10, 0.6) \quad X_2 \sim 0.1 \mathcal{N}(5, 0.6) + 0.9 \mathcal{N}(0, 0.6). \]
Simulated data

(III) Third simulation scenario

\[ X_1 \sim 0.9 \mathcal{N}(5, 0.6) + 0.1 \mathcal{N}(10, 0.6) \quad X_2 \sim 0.1 \mathcal{N}(5, 0.6) + 0.9 \mathcal{N}(0, 0.6). \]

Clustering:

<table>
<thead>
<tr>
<th># comp.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>≥ 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>0</td>
<td>0</td>
<td>0.679</td>
<td>0.232</td>
<td>0.065</td>
<td>0.018</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>0</td>
<td>0</td>
<td>0.778</td>
<td>0.185</td>
<td>0.032</td>
<td>0.004</td>
<td>0.001</td>
<td>0</td>
</tr>
<tr>
<td>( K_{1,2} )</td>
<td>0</td>
<td>0.965</td>
<td>0.034</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Simulated data

(III) Third simulation scenario

\[ X_1 \sim 0.9 N(5, 0.6) + 0.1 N(10, 0.6) \quad X_2 \sim 0.1 N(5, 0.6) + 0.9 N(0, 0.6). \]

Clustering:

<table>
<thead>
<tr>
<th># comp.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>≥ 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_1)</td>
<td>0</td>
<td>0</td>
<td>0.679</td>
<td>0.232</td>
<td>0.065</td>
<td>0.018</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>(K_2)</td>
<td>0</td>
<td>0</td>
<td>0.778</td>
<td>0.185</td>
<td>0.032</td>
<td>0.004</td>
<td>0.001</td>
<td>0</td>
</tr>
<tr>
<td>(K_{1,2})</td>
<td>0</td>
<td>0.965</td>
<td>0.034</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Testing: Bayes factor for testing \(H_0: p_1 = p_2\)

\[
BF = \frac{\mathbb{P}[^{\tilde{p}_1} = ^{\tilde{p}_2} \mid \text{data}]}{\mathbb{P}[^{\tilde{p}_1} \neq ^{\tilde{p}_2} \mid \text{data}]} \frac{1 - \pi_1^*}{\pi_1^*} \approx 0.00022
\]
Density estimates

Estimated (blue) and true densities (red) for $X_1$ in Panel (a) and $X_2$ 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)

Reference

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


Thank you!