Clustering with shot-noise Cox process

Federico Camerlenghi

joint work with Alessandro Carminati, Mario Beraha and Alessandra Guglielmi

January 12–16, 2026

University of Milano – Bicocca

COAUTHORS

Alessandro
Carminati

Mario
Beraha

Alessandra
Guglielmi

Paper available on ArXiv:
"Hierarchical shot-noise Cox process mixtures for clustering across groups"

SETTING: MULTIPLE POPULATIONS

We consider the following setting:

- data are divided into g groups (or populations), where group ℓ contains n_ℓ data points, as $\ell = 1, \dots, g$;
- denote by $Y_{\ell,1}, \dots, Y_{\ell,n_\ell}$ the observations for group ℓ, as $\ell = 1, \dots, g$.
- data are assumed partially exchangeable: exchangeable within groups and conditionally independent across groups.

Setting: multiple populations

We consider the following setting:

- data are divided into g groups (or populations), where group ℓ contains n_ℓ data points, as $\ell = 1, \dots, g$;
- denote by $Y_{\ell,1}, \dots, Y_{\ell,n_\ell}$ the observations for group ℓ, as $\ell = 1, \dots, g$.
- data are assumed partially exchangeable: exchangeable within groups and conditionally independent across groups.

Relevant examples are:

- students' GPAs grouped by university of attendance;
- patients' health data grouped by hospital;
- galaxies grouped by luminosity type.

Standard Bayesian clustering

Bayesian clustering is addressed by specifying a mixture model for each group:

$$Y_{\ell,i} \mid \tilde{p}_\ell \overset{ind}{\sim} \int_{\mathbb{X}} f(\cdot \mid x)\tilde{p}_\ell(\mathrm{d}x) \quad i = 1, \ldots, n_\ell; \quad \ell = 1, \ldots, g$$

where:

- $f(\cdot \mid x)$ is a density with parameter x;
- $(\tilde{p}_1, \ldots, \tilde{p}_g)$ is a vector of mixing measures for all groups.

Standard Bayesian clustering

Bayesian clustering is addressed by specifying a mixture model for each group:

$$Y_{\ell,i} \mid \tilde{p}_\ell \overset{ind}{\sim} \int_{\mathbb{X}} f(\cdot \mid x)\tilde{p}_\ell(\mathrm{d}x) \quad i = 1, \ldots, n_\ell; \quad \ell = 1, \ldots, g$$

where:

- $f(\cdot \mid x)$ is a density with parameter x;
- $(\tilde{p}_1, \ldots, \tilde{p}_g)$ is a vector of mixing measures for all groups.

Several proposals are available to specify $(\tilde{p}_1, \ldots, \tilde{p}_g)$ and achieve across-group clustering:

- dependent Dirichlet process (MacEachern, 1999);
- hierarchical Dirichlet process (Teh et al., 2006) and generalizations;
- additive structures (Lijoi et al., 2014);
- nested Dirichlet process (Rodrǵuez et al., 2008) and generalizations;
- see (Quintana et al., 2022) for a complete review.

Standard Bayesian clustering

By introducing suitable latent variables, the mixture model can be written in a hierarchical fashion:

$$Y_{\ell,i} \mid X_{\ell,i} \overset{\text{ind}}{\sim} f(\cdot \mid X_{\ell,i})$$

$$X_{\ell,i} \mid \tilde{p}_\ell \overset{\text{iid}}{\sim} \tilde{p}_\ell$$

where the random probability measures $\tilde{p}_1, \dots, \tilde{p}_g$ are typically assumed to be almost surely discrete.

Standard Bayesian clustering

By introducing suitable latent variables, the mixture model can be written in a hierarchical fashion:

$$Y_{\ell,i} \mid X_{\ell,i} \overset{\text{ind}}{\sim} f(\cdot \mid X_{\ell,i})$$

$$X_{\ell,i} \mid \tilde{p}_\ell \overset{\text{iid}}{\sim} \tilde{p}_\ell$$

where the random probability measures $\tilde{p}_1, \ldots, \tilde{p}_g$ are typically assumed to be almost surely discrete.

- The discreteness of the $\tilde{p}_\ell$'s induces ties of the latent variables within and across samples
- Ties induce the standard clustering mechanism in mixture models

$$Y_{\ell,i} \text{ and } Y_{\ell,i'} \text{ are in the same cluster} \iff X_{\ell,i} = X_{\ell,i'}$$

- The use of dependent random probability measures allows to borrow information across groups.

Problems with standard clustering

Remarks:

- Two observations are in the same cluster iff they share the same latent variable.
- The standard notion of clustering is based on exact sharing of latent variables.
- This notion of clustering is too rigid when subtle differences across groups matter!

Problems with standard clustering

Remarks:

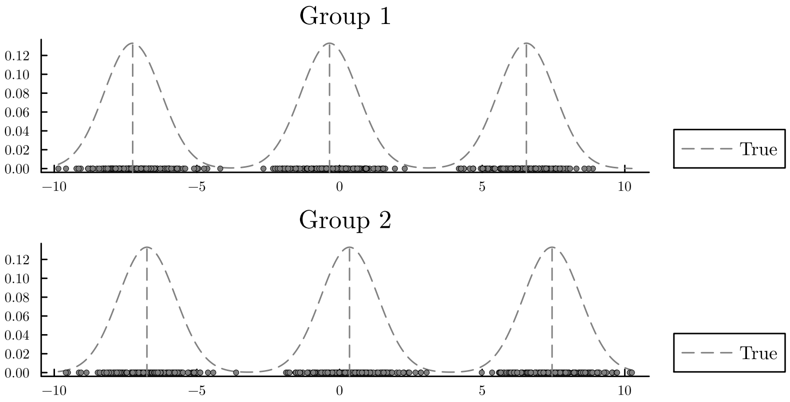
- Two observations are in the same cluster iff they share the same latent variable.
- The standard notion of clustering is based on exact sharing of latent variables.
- This notion of clustering is too rigid when subtle differences across groups matter!

We showcase the rigidity of standard clustering mechanism by considering an illustrative example with $g = 2$ groups:

- Gaussian mixture of three components in each group;
- the components differ slightly across groups $\Rightarrow$ we expect only three clusters.

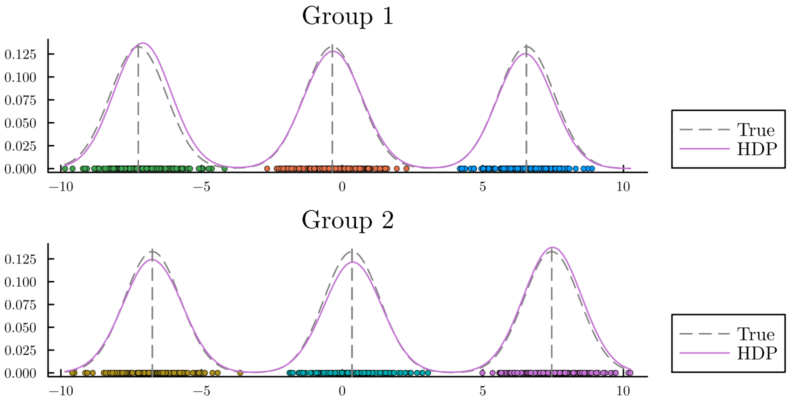
We use the hierarchical Dirichlet process (HDP) to fit the data.

ILLUSTRATIVE EXAMPLE: SHIFTED GAUSSIAN MIXTURES



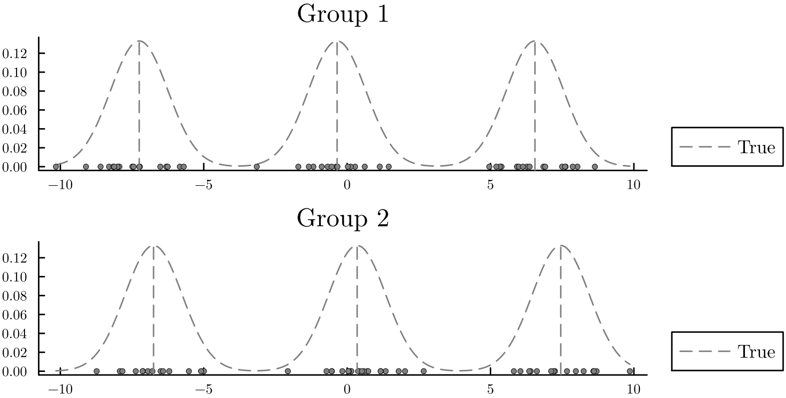
- $n_1 = n_2 = 500$ observations;
- the three components differ slightly across the two groups.

Illustrative example: shifted Gaussian mixtures



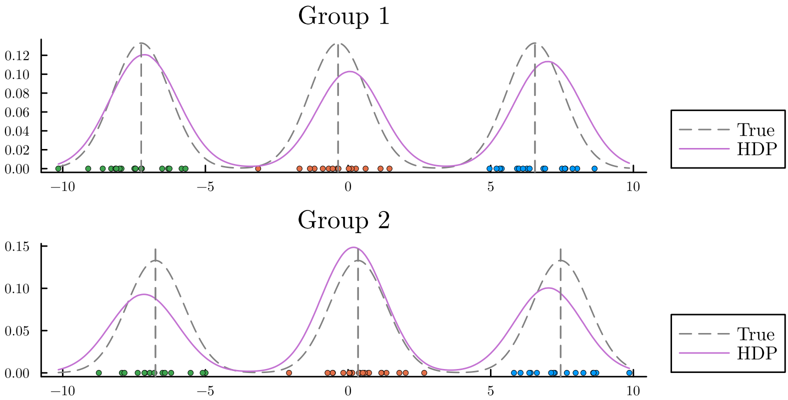
- The HDP recognizes six different cluster!

Illustrative example: shifted Gaussian mixtures



- $n_1 = n_2 = 50$ observations

Illustrative example: shifted Gaussian mixtures



- The HDP recognizes three clusters, but inaccurate density estimates

Our contribution

Summarizing the results:

- large sample size: good density estimations, but bad clustering and information sharing;
- small sample size: bad density estimations, but good clustering and information sharing.

Our contribution

Summarizing the results:

- large sample size: good density estimations, but bad clustering and information sharing;
- small sample size: bad density estimations, but good clustering and information sharing.

Our proposal to overcome the trade-off outlined above:

- we introduce a new model for clustering grouped data;
- we also define a new notion of clustering;
- our proposal is based on the shot-noise Cox process (Møller, 2003).

OUTLINE

Hierarchical shot-noise Cox process

Hierarchical shot-noise Cox process

We introduce the vector of mixing measures

$$(\tilde{p}_1, \ldots, \tilde{p}_g)$$

to define the hierarchical shot-noise Cox process (HSNCP) mixture model.

Hierarchical shot-noise Cox process

We introduce the vector of mixing measures

$$(\tilde{p}_1, \dots, \tilde{p}_g)$$

to define the hierarchical shot-noise Cox process (HSNCP) mixture model.

- We define each $\tilde{p}_\ell$ via normalization

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

where $\tilde{\mu}_\ell$ is a Completely Random Measure (CRM), see (Kingman, 1967).

Hierarchical shot-noise Cox process

We introduce the vector of mixing measures

$$(\tilde{p}_1, \dots, \tilde{p}_g)$$

to define the hierarchical shot-noise Cox process (HSNCP) mixture model.

- We define each $\tilde{p}_\ell$ via normalization

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

  where $\tilde{\mu}_\ell$ is a Completely Random Measure (CRM), see (Kingman, 1967).
- This is a common strategy in the BNP literature: see (Regazzini et al., 2003) and subsequent contributions.

Hierarchical shot-noise Cox process

We introduce the vector of mixing measures

$$(\tilde{p}_1, \dots, \tilde{p}_g)$$

to define the hierarchical shot-noise Cox process (HSNCP) mixture model.

- We define each $\tilde{p}_\ell$ via normalization

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

  where $\tilde{\mu}_\ell$ is a Completely Random Measure (CRM), see (Kingman, 1967).
- This is a common strategy in the BNP literature: see (Regazzini et al., 2003) and subsequent contributions.
- Note that a CRM is a functional of a Poisson process $\tilde{N}_\ell$:

$$\tilde{\mu}_\ell = \sum_{h \geq 1} S_{\ell h} \delta_{\phi_{\ell h}} \iff \tilde{N}_\ell = \sum_{h \geq 1} \delta_{(S_{\ell h}, \phi_{\ell h})}$$

  where $\tilde{N}_\ell$ is a Poisson Process (PP) on $\mathbb{R}_+ \times \mathbb{X}$ with intensity measure given by $\rho_\ell(s)\mathrm{d}s\,\eta(\mathrm{d}x)$.

Hierarchical shot-noise Cox process

We introduce the vector of mixing measures

$$(\tilde{p}_1, \ldots, \tilde{p}_g)$$

to define the hierarchical shot-noise Cox process (HSNCP) mixture model.

- We define each $\tilde{p}_\ell$ via normalization

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

  where $\tilde{\mu}_\ell$ is a Completely Random Measure (CRM), see (Kingman, 1967).

- This is a common strategy in the BNP literature: see (Regazzini et al., 2003) and subsequent contributions.
- Note that a CRM is a functional of a Poisson process $\tilde{N}_\ell$:

$$\tilde{\mu}_\ell = \sum_{h \geq 1} S_{\ell h} \delta_{\phi_{\ell h}} \iff \tilde{N}_\ell = \sum_{h \geq 1} \delta_{(S_{\ell h}, \phi_{\ell h})}$$

  where $\tilde{N}_\ell$ is a Poisson Process (PP) on $\mathbb{R}_+ \times \mathbb{X}$ with intensity measure given by $\rho_\ell(s)\mathrm{d}s\,\eta(\mathrm{d}x)$.

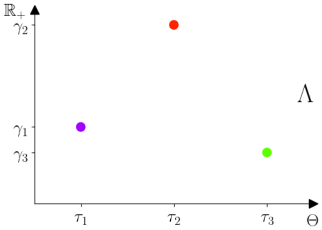
- The measure η is specified to induce dependence across groups.

Hierarchical shot-noise Cox process

- $(\tilde{\mu}_1, \dots, \tilde{\mu}_g)$ is a Hierarchical shot-noise Cox process (HSNCP).

Hierarchical shot-noise Cox process

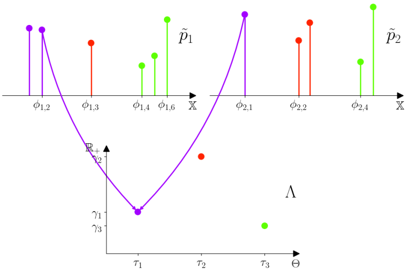
- $(\tilde{\mu}_1, \ldots, \tilde{\mu}_g)$ is a Hierarchical shot-noise Cox process (HSNCP).



- $\Lambda = \sum_{h \geq 1} \delta_{(\gamma_h, \tau_h)}$: Poisson process with intensity measure $\rho_0(\mathrm{d}\gamma) G_0(\mathrm{d}\tau)$

Hierarchical shot-noise Cox process

- $(\tilde{\mu}_1, \ldots, \tilde{\mu}_g)$ is a Hierarchical shot-noise Cox process (HSNCP).

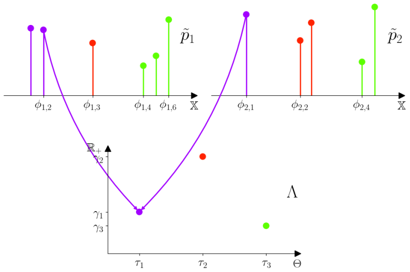


- $\tilde{\mu}_\ell \mid \Lambda = \sum_{h \geq 1} S_{\ell h} \delta_{\phi_{\ell h}}$: CRM with intensity measure $\rho(s) \mathrm{d}s \, \eta_\Lambda(\mathrm{d}x)$

- $\Lambda = \sum_{h \geq 1} \delta_{(\gamma_h, \tau_h)}$: Poisson process with intensity measure $\rho_0(\mathrm{d}\gamma) G_0(\mathrm{d}\tau)$

Hierarchical shot-noise Cox process

- $(\tilde{\mu}_1, \dots, \tilde{\mu}_g)$ is a Hierarchical shot-noise Cox process (HSNCP).



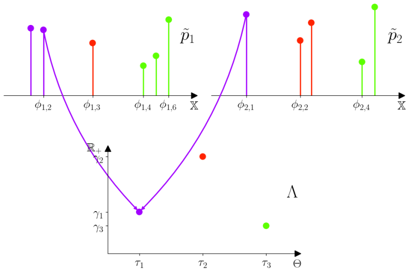
- $\tilde{\mu}_\ell \mid \Lambda = \sum_{h\geq 1} S_{\ell h}\delta_{\phi_{\ell h}}$: CRM with intensity measure $\rho(s)\mathrm{d}s\,\eta_\Lambda(\mathrm{d}x)$

$$\eta_\Lambda(\mathrm{d}x) = \left\{\sum_{j\geq 1} \gamma_j k(x, \tau_j)\right\} \mathrm{d}x$$

- $\Lambda = \sum_{h\geq 1} \delta_{(\gamma_h, \tau_h)}$: Poisson process with intensity measure $\rho_0(\mathrm{d}\gamma)G_0(\mathrm{d}\tau)$

HIERARCHICAL SHOT-NOISE COX PROCESS

- $(\tilde{\mu}_1, \ldots, \tilde{\mu}_g)$ is a Hierarchical shot-noise Cox process (HSNCP).



- $\tilde{p}_\ell = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_\ell(\mathbb{X})}$: random probability of group ℓ
- $\tilde{\mu}_\ell \mid \Lambda = \sum_{h\geq 1} S_{\ell h}\delta_{\phi_{\ell h}}$: CRM with intensity measure $\rho(s)\mathrm{d}s\, \eta_\Lambda(\mathrm{d}x)$

$$\eta_\Lambda(\mathrm{d}x) = \left\{ \sum_{j\geq 1} \gamma_j k(x, \tau_j) \right\} \mathrm{d}x$$

- $\Lambda = \sum_{h\geq 1} \delta_{(\gamma_h, \tau_h)}$: Poisson process with intensity measure $\rho_0(\mathrm{d}\gamma)G_0(\mathrm{d}\tau)$

HIERARCHICAL SHOT-NOISE COX PROCESS MIXTURE MODEL

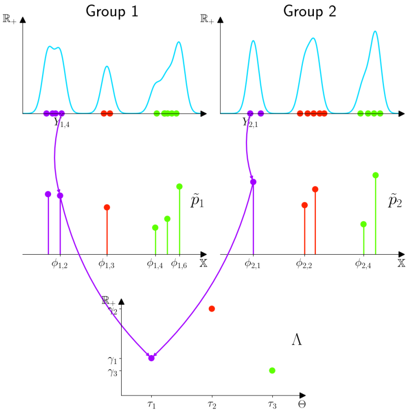
- We use the construction in a mixture setting.

$$Y_{\ell i} \mid X_{\ell i} \overset{ind}{\sim} f(\cdot \mid X_{\ell i})$$

$$X_{\ell i} \mid \tilde{p}_\ell \sim \tilde{p}_\ell = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_\ell(\mathbb{X})}$$

$$\tilde{\mu}_\ell \mid \Lambda = \sum_{h \geq 1} S_{\ell h} \delta_{\phi_{\ell h}} : \ \text{CRM}\ (\rho(s)\mathrm{d}s\, \eta_\Lambda(\mathrm{d}x))$$

$$\Lambda = \sum_{h \geq 1} \delta_{(\gamma_h, \tau_h)} : \ \text{PP}\ (\rho_0(\mathrm{d}\gamma) G_0(\mathrm{d}\tau))$$



$$\eta_\Lambda(\mathrm{d}x) = \left\{ \sum_{j \geq 1} \gamma_j k(x, \tau_j) \right\} \mathrm{d}x$$

A NEW DEFINITION OF CLUSTERING

- We observe that the random measure $\tilde{\mu}_\ell$ can be equivalently written as

$$\tilde{\mu}_\ell = \sum_{j \geq 1} \tilde{\mu}_{\ell j}$$

where

$$\tilde{\mu}_{\ell j} = \sum_{h \geq 1} S_{\ell j h} \delta_{\phi_{\ell j h}} \text{ is a CRM with intensity measure } \gamma_j \rho(s) \mathrm{d}s \, k(x, \tau_j) \mathrm{d}x$$

- The atoms of $\tilde{\mu}_{\ell j}$ are close to the parent atom τ_j's of Λ but not identical.
- $\tilde{\mu}_{\ell j}$: generates similar latent variables $X_{\ell i}$'s, but not identical.

A new definition of clustering

- We observe that the random measure $\tilde{\mu}_\ell$ can be equivalently written as

$$\tilde{\mu}_\ell = \sum_{j \geq 1} \tilde{\mu}_{\ell j}$$

where

$$\tilde{\mu}_{\ell j} = \sum_{h \geq 1} S_{\ell jh} \delta_{\phi_{\ell jh}} \text{ is a CRM with intensity measure } \gamma_j \rho(s) \mathrm{d}s \, k(x, \tau_j) \mathrm{d}x$$

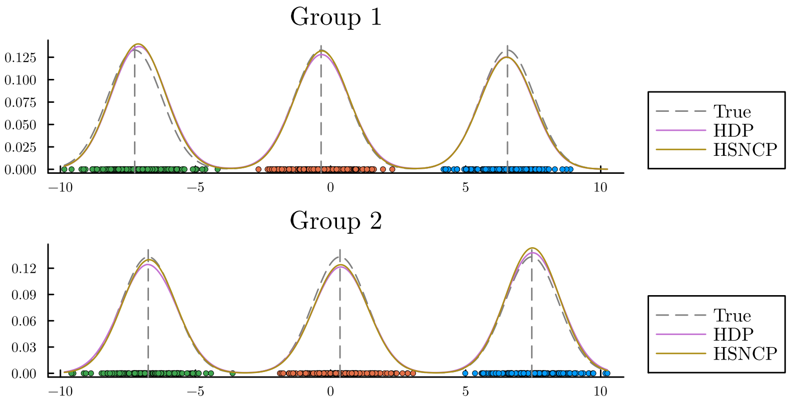
- The atoms of $\tilde{\mu}_{\ell j}$ are close to the parent atom τ_j's of Λ but not identical.
- $\tilde{\mu}_{\ell j}$: generates similar latent variables $X_{\ell i}$'s, but not identical.

This induces a new definition of clustering:

New clustering definition

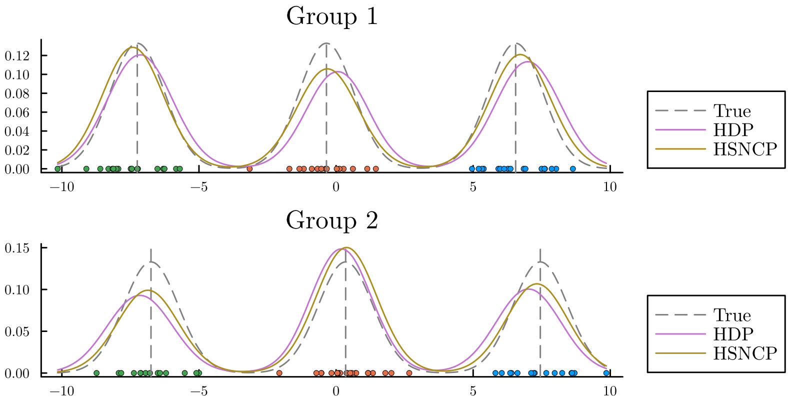
- traditional clustering: observations $Y_{\ell i}$'s are clustered together iff they share the same latent variables $X_{\ell i}$'s;
- clustering with HSNCP: observations $Y_{\ell i}$'s are clustered together iff they have similar ~~same~~ latent variables $X_{\ell i}$'s;

Illustrative example: shifted Gaussian mixtures



- $n_1 = n_2 = 500$ observations.
- The HSNCP borrows information across groups: we recognize only three clusters.

Illustrative example: shifted Gaussian mixtures



- $n_1 = n_2 = 50$ observations.
- The HSNCP provides good density estimates.

Bayesian analysis

Summing up, we now focus on the latent variables in the HSNCP mixture model

$$X_{\ell i} \mid \tilde{p}_\ell \sim \tilde{p}_\ell = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_\ell(\mathbb{X})}$$

$$(\tilde{\mu}_1, \ldots, \tilde{\mu}_g) \sim \text{HSNCP}(\rho, \rho_0, G_0, k(\cdot, \cdot))$$

where the HSNCP has been specified before.

Summing up, we now focus on the latent variables in the HSNCP mixture model

$$X_{\ell i} \mid \tilde{p}_\ell \sim \tilde{p}_\ell = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_\ell(\mathbb{X})}$$

$$(\tilde{\mu}_1, \dots, \tilde{\mu}_g) \sim \text{HSNCP}(\rho, \rho_0, G_0, k(\,\cdot\,,\,\cdot\,))$$

where the HSNCP has been specified before. We have expressions for

- the marginal distribution of the latent parameters $X_{\ell i}$'s;
- the posterior distribution of the HSNCP;
- the predictive distribution of a new latent parameter.

We explain the theoretical properties and induced clustering through a restaurant franchise metaphor.

$X_{\ell i} \mid \tilde{p}_\ell \stackrel{ind}{\sim} \tilde{p}_\ell$

RESTAURANT FRANCHISE METAPHOR

$$X_{\ell i} \mid \tilde{p}_\ell \overset{ind}{\sim} \tilde{p}_\ell$$

Restaurant franchise metaphor:

- Franchise of g restaurants.

Restaurant 1

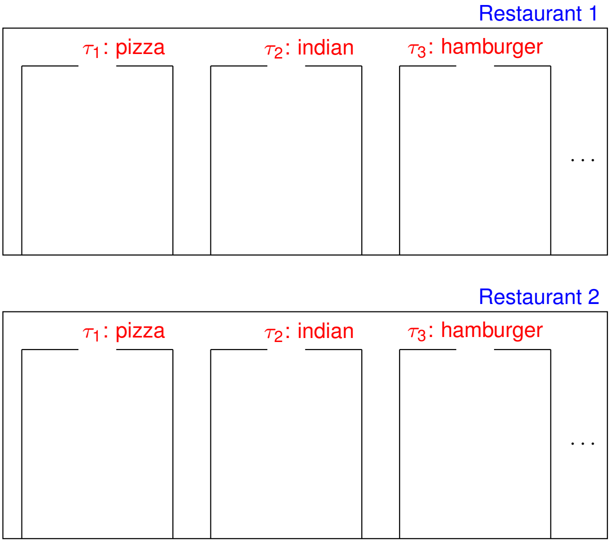
Restaurant 2

Restaurant franchise metaphor

$$X_{\ell i} \mid \tilde{p}_\ell \overset{ind}{\sim} \tilde{p}_\ell$$

Restaurant franchise metaphor:

- Franchise of g restaurants.
- Infinitely many thematic rooms, $(\tau_j)_{j\geq 1}$.

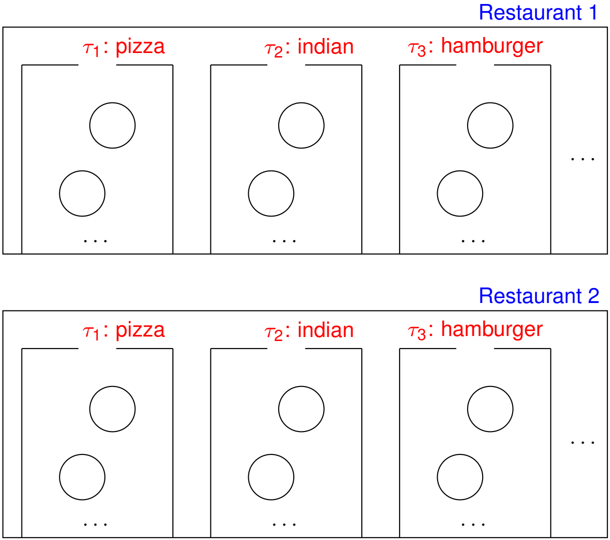


Restaurant franchise metaphor

$$X_{\ell i} \mid \tilde{p}_\ell \overset{ind}{\sim} \tilde{p}_\ell$$

Restaurant franchise metaphor:

- Franchise of g restaurants.
- Infinitely many thematic rooms, $(\tau_j)_{j\geq 1}$.
- Each thematic room contains infinitely many tables.

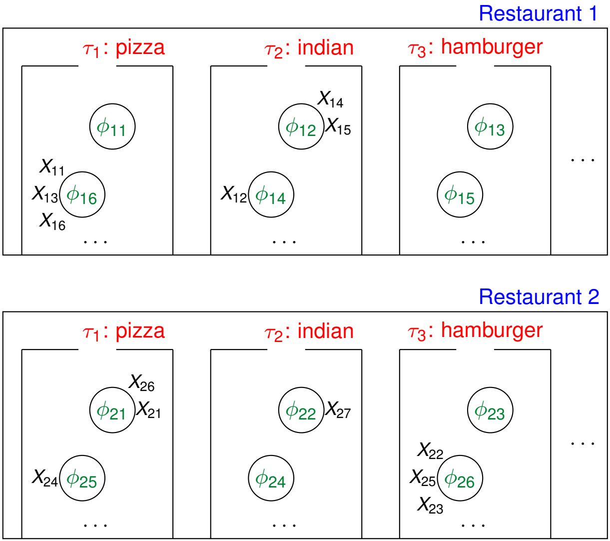


Restaurant franchise metaphor

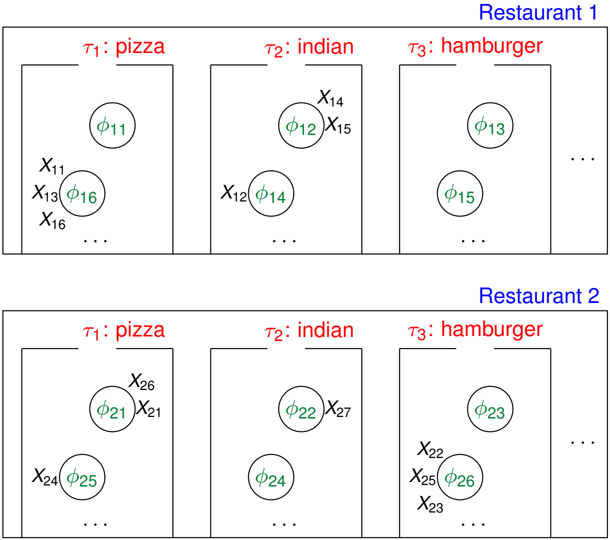
$$X_{\ell i} \mid \tilde{p}_\ell \overset{ind}{\sim} \tilde{p}_\ell$$

Restaurant franchise metaphor:

- Franchise of g restaurants.
- Infinitely many thematic rooms, $(\tau_j)_{j\geq 1}$.
- Each thematic room contains infinitely many tables.
- Customers $X_{\ell i}$s at the same table eat the same dish $\phi_{\ell h}$, related to the room's theme.



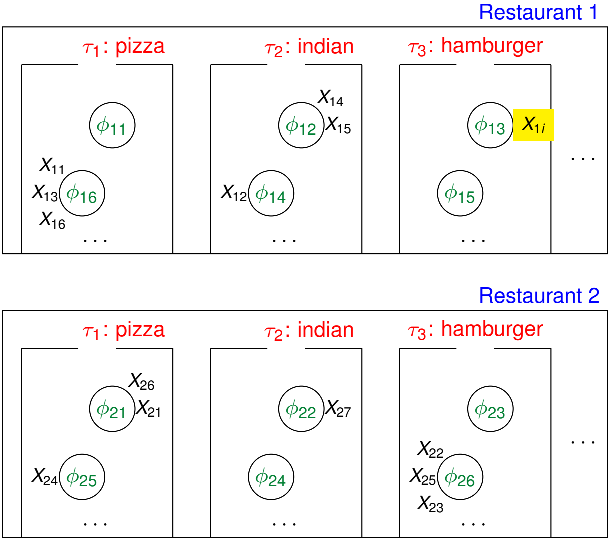
A new customer $X_{\ell i}$ in restaurant ℓ can either:



Predictive distribution: intuition

A new customer $X_{\ell i}$ in restaurant ℓ can either:

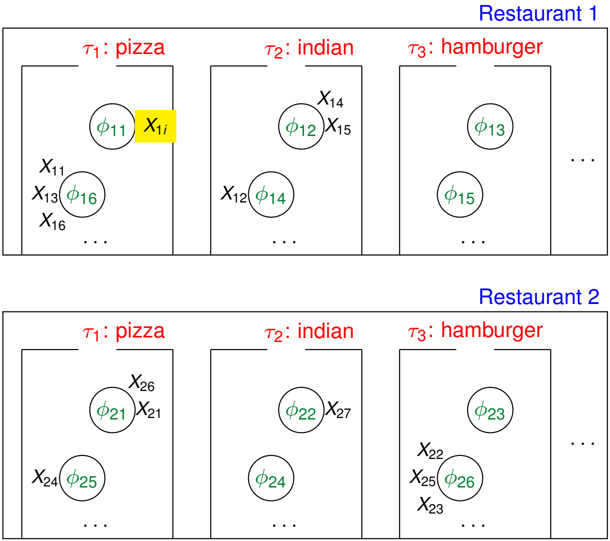
- enter an empty room and sit at an empty table.



Predictive distribution: intuition

A new customer $X_{\ell i}$ in restaurant ℓ can either:

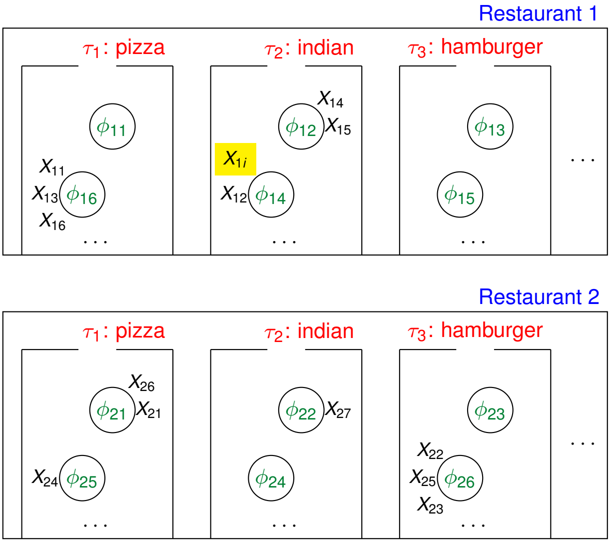
- enter an empty room and sit at an empty table.
- enter an already occupied room and sit at an empty table.



Predictive distribution: intuition

A new customer $X_{\ell i}$ in restaurant ℓ can either:

- enter an empty room and sit at an empty table.
- enter an already occupied room and sit at an empty table.
- enter an already occupied room and sit at an already occupied table.



PREDICTIVE DISTRIBUTION (CARMINATI ET AL., 2026+)

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_g)$ be a sample of size n. Then, for any Borel set A, we have

$$\mathbb{P}(X_{\ell,n_\ell+1} \in A \mid \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}) = \frac{U_\ell}{\Gamma(n_\ell)} \sum_{h=1}^{K_\ell} \frac{\kappa(U_\ell, \xi_{\ell h} + 1)}{\kappa(U_\ell, \xi_{\ell h})} \delta_{X^*_{\ell h}}(A)$$
$$+ \frac{U_\ell}{\Gamma(n_\ell)} \sum_{j=1}^{|\boldsymbol{T}|} \kappa(U_\ell, 1) \frac{\kappa_0(\sum_{\ell=1}^{g} \psi(U_\ell), \zeta_j + 1)}{\kappa_0(\sum_{\ell=1}^{g} \psi(U_\ell), \zeta_j)} \int_A m(dx \mid X^*_{\ell h} : T_{\ell h} = j)$$
$$+ \frac{U_\ell}{\Gamma(n_\ell)} \kappa(U_\ell, 1) \kappa_0 \left(\sum_{\ell=1}^{g} \psi(U_\ell), 1 \right) \int m(dx),$$

$\boldsymbol{U}$ are latent variables similar to those in (James et al., 2009), and $\boldsymbol{T}$ are latent variables describing the room structure.

Predictive distribution: probabilistic structure

Predictive distribution (Carminati et al., 2026+)

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_g)$ be a sample of size n. Then, for any Borel set A, we have

$$
\begin{aligned}
\mathbb{P}(X_{\ell,n_\ell+1} \in A \mid \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}) &= \frac{U_\ell}{\Gamma(n_\ell)} \sum_{h=1}^{K_\ell} \frac{\kappa(U_\ell, \xi_{\ell h}+1)}{\kappa(U_\ell, \xi_{\ell h})} \delta_{X^*_{\ell h}}(A) \\
&\quad + \frac{U_\ell}{\Gamma(n_\ell)} \sum_{j=1}^{|\boldsymbol{T}|} \kappa(U_\ell, 1) \frac{\kappa_0(\sum_{\ell=1}^{g} \psi(U_\ell), \zeta_j + 1)}{\kappa_0(\sum_{\ell=1}^{g} \psi(U_\ell), \zeta_j)} \int_A m(dx \mid X^*_{\ell h} : T_{\ell h} = j) \\
&\quad + \frac{U_\ell}{\Gamma(n_\ell)} \kappa(U_\ell, 1) \kappa_0\left(\sum_{\ell=1}^{g} \psi(U_\ell), 1\right) \int m(dx),
\end{aligned}
$$

$\boldsymbol{U}$ are latent variables similar to those in (James et al., 2009), and $\boldsymbol{T}$ are latent variables describing the room structure.

- first term: the customer chooses an occupied room and table;
- second term: the customer chooses an occupied room and an empty table.
- third term: the customer chooses an empty room and empty table.

POSTERIOR DISTRIBUTION: PROBABILISTIC STRUCTURE

POSTERIOR DISTRIBUTION (CARMINATI ET AL., 2026+)

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_g)$ be a sample of size n. Then the following distributional equality holds true

$$\tilde{\mu}_\ell \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{U} \stackrel{d}{=} \sum_{h=1}^{K_\ell} S_{\ell h}^* \delta_{X_{\ell h}^*} + \sum_{j=1}^{|\boldsymbol{T}|} \tilde{\mu}_{\ell j}^{(p)} + \tilde{\mu}_\ell^{(p)},$$

for any restaurant $\ell = 1, \ldots, g$.

- The $S_{\ell h}^*$'s are independent positive random variables with density $f_{S_{\ell h}^*}(s) \propto e^{-U_\ell s} s^{\xi_{\ell h}} \rho(s)$.
- The $\tilde{\mu}_{\ell j}^{(p)}$'s are independent CRMs, conditionally on latent parameters $(\gamma_j^{(p)}, \tau_j^{(p)})$, with a known density;
- $(\tilde{\mu}_1^{(p)}, \ldots, \tilde{\mu}_g^{(p)})$ is a HSNCP with updated parameters.

APPLICATION

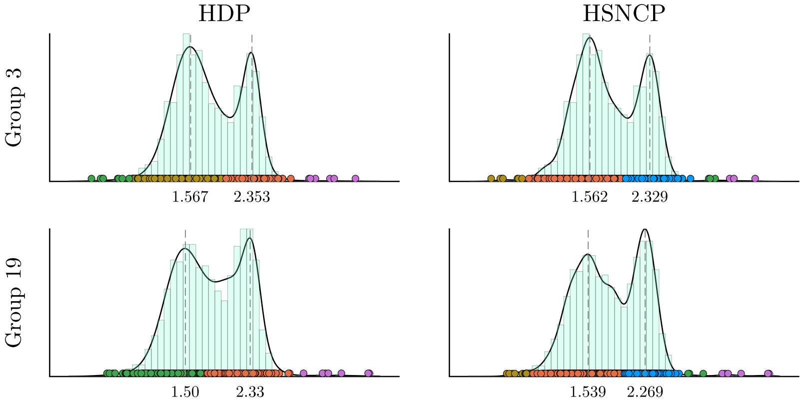
Application: Sloan Digital Sky Survey

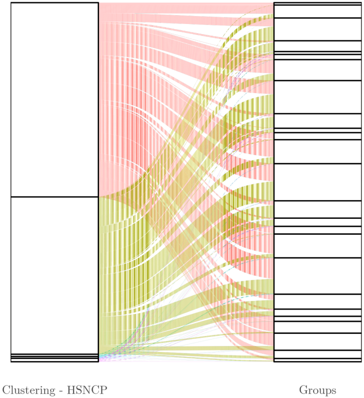
We consider the dataset from the Sloan Digital Sky Survey first data release (Abazajan et al., 2003).

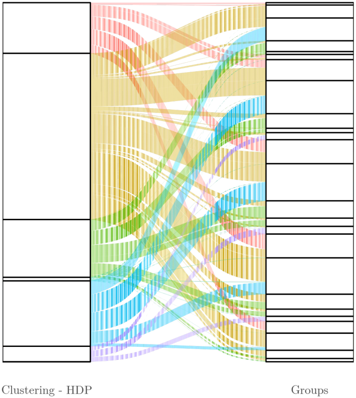
- It contains measurements of the $u - r$ color, that is, the difference between ultraviolet and red color distributions.
- The measurements refer to $\sum_{\ell=1}^{g} n_\ell = 24{,}312$ galaxies.
- Galaxies are divided in $g = 25$ groups according to luminosity type and environment.
- $u - r$ color provides a robust indicator of galaxy type and star formation activity.
- Clustering galaxies according to their $u - r$ color allows us to identify different evolutionary stages.

We run the MCMC algorithm for 50,000 iterations. Running our algorithm on a standard laptop took 3 hours and 19 minutes.

Application: selected groups




Clustering - HSNCP
Groups


Clustering - HDP
Groups

ESSENTIAL BIBLIOGRAPHY

- BERAHA M., CAMERLENGHI F., GHILOTTI L. (2025). Bayesian calculus and predictive characterizations of extended feature allocation models. *Available on ArXiv*.
- DOMBOWSKY A., DUNSON D. B. (2025). Bayesian clustering via fusing of localized densities. *Journal of the American Statistical Association*, in press.
- JAMES L., LIJOI A., PRÜNSTER I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, **36**, 76–97.
- KINGMAN J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.
- LIJOI A., NIPOTI B., PRÜNSTER I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli*, **20**, 1260–91.
- MACEACHERN S. N. (1999). Dependent nonparametric processes. *In ASA proceedings of the section on Bayesian statistical science*, 50–55.
- MALSINER-WALLI G., FRÜHWIRTH-SCHNATTER S., GRÜN B. (2017). Identifying mixtures of mixtures using Bayesian estimation. *Journal of Computational and Graphical Statistics*, **26**, 285–295.
- MØLLER J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, **35**, 614–640.
- QUINTANA F. A., MÜLLER P., JARA A., MACEACHERN S. N. (2022). The dependent Dirichlet process and related models. *Statistical Science*, **37**, 24–41.
- REGAZZINI E., LIJOI A., PRÜNSTER I. (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics*, **31**, 560–585.
- RODRÍGUEZ A., DUNSON D. B., GELFAND A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, **103**, 1131–44.
- TEH Y.W., JORDAN M., BEAL M.J., BLEI D.M. (2006). Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, **101**, 1566–81.

Thank you!