Title

Subtitle

Bruno Guindani
Elena Zazzetti

November 22, 2019

https://github.com/poliprojects/BNPlib

Non-Parametric statistics

- Goal: density estimation
- **Infinite-dimensional** parameters, e.g. functions

Non-Parametric statistics (the Bayesian way)

- Goal: density estimation
- **Infinite-dimensional** parameters, e.g. functions
- Model:

$$
\begin{aligned}
y_i | G &\overset{\text{iid}}{\sim} G, \quad i = 1, \ldots, n \\
G &\sim \mathcal{P}
\end{aligned}
$$

Non-Parametric statistics (the Bayesian way)

- Goal: density estimation
- **Infinite-dimensional** parameters, e.g. functions
- Model:

$$
\begin{aligned}
y_i | G &\overset{\text{iid}}{\sim} G, \quad i = 1, \dots, n \\
G &\sim \mathcal{P}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{P} : \Omega &\to M(S) \ \ \text{fixed} \\
\big[\ \omega &\mapsto G(\cdot)\ \big]
\end{aligned}
$$

- Model name: **BNP model**

Dirichlet Process prior

$$
\begin{aligned}
y_i|G &\overset{\text{iid}}{\sim} G \\
G &\sim \mathcal{P} = DP(MG_0)
\end{aligned}
$$

- Parameters: $M > 0, \; G_0 \in M(S)$
- Defining property: $\forall\{B_{1:k}\}$ partition of S,

$$[G(B_1), \dots, G(B_k)] \sim \mathrm{Dir}\left(MG_0(B_1), \dots, MG_0(B_k)\right)$$

Dirichlet Process prior

$$y_i|G \overset{\text{iid}}{\sim} G$$
$$G \sim \mathcal{P} = DP(MG_0)$$

- Parameters: $M > 0,\ G_0 \in M(S)$
- Defining property: $\forall\{B_{1:k}\}$ partition of S,

$$[G(B_1), \ldots, G(B_k)] \sim \mathrm{Dir}\left(MG_0(B_1), \ldots, MG_0(B_k)\right)$$

- **Discreteness** (stick-breaking): $G(\cdot) = \sum_{k=1}^{+\infty} w_h \delta_{m_h}(\cdot)$
- **Conjugacy**: $G|\mathbf{y} \sim DP(MG_0 + \sum_i \delta_{y_i}) \quad \Longrightarrow \quad$ density estimation

Continuous density estimation

- **Mixtures** (kernel f + mixing distribution G):

$$y_i | G \sim f_G(y) = \int f_\vartheta(y) \, \mathrm{d}G(\vartheta)$$

Continuous density estimation

- **Mixtures** (kernel f + mixing distribution G):

$$y_i|G \sim f_G(y) = \int f_\vartheta(y)\, \mathrm{d}G(\vartheta)$$
$$G \sim DP(MG_0)$$

- Model name: **DPM model**

Continuous density estimation

- **Mixtures** (kernel f + mixing distribution G):

$$y_i|G \sim f_G(y) = \int f_\vartheta(y)\,\mathrm{d}G(\vartheta)$$
$$G \sim DP(MG_0)$$

- Model name: **DPM model**
- Equivalent to:

$$y_i|\vartheta_i \stackrel{\perp\!\!\!\perp}{\sim} f_{\vartheta_i}$$
$$\vartheta_i|G \stackrel{\text{iid}}{\sim} G$$
$$G \sim DP(MG_0)$$

- ϑ_i "latent variables" $\forall i = 1, \ldots, n$

Clustering in the DPM

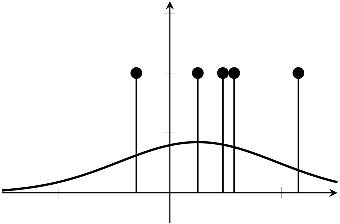
- Discreteness: the ϑ_i have one of the k **unique values** $\boldsymbol{\phi_j}$ $_{(j\,=\,1,\ldots,k)}$
- $k \simeq M \log(n) \ll n$
- All i s.t. $\vartheta_i = \phi_j$ belong to cluster S_j $_{(j\,=\,1,\ldots,k)}$, and $n_j = |S_j|$

Clustering in the DPM

- Discreteness: the ϑ_i have one of the k **unique values** $\boldsymbol{\phi_j}$ $_{(j\,=\,1,\,...\,,\,k)}$
- $k \simeq M \log(n) \ll n$
- All i s.t. $\vartheta_i = \phi_j$ belong to cluster S_j $_{(j\,=\,1,\,...\,,\,k)}$, and $n_j = |S_j|$
- Conditional prior for ϑ_i:

$$\mathcal{L}(\vartheta_i|\boldsymbol{\vartheta}_{-i}) \propto \sum_{j=1}^{k^-} n_j^- \delta_{\phi_j^-}(\vartheta_i) + M G_0(\vartheta_i)$$

$$\uparrow \qquad\qquad\qquad \uparrow$$

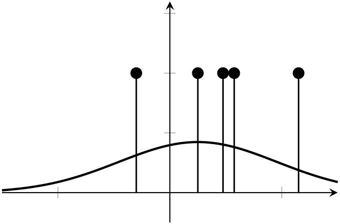


Clustering in the DPM

- Discreteness: the ϑ_i have one of the k **unique values** $\boldsymbol{\phi_j}$ $_{(j\,=\,1,\ldots,k)}$
- $k \simeq M\log(n) \ll n$
- All i s.t. $\vartheta_i = \phi_j$ belong to cluster S_j $_{(j\,=\,1,\ldots,k)}$, and $n_j = |S_j|$
- Conditional prior for ϑ_i:

$$\mathcal{L}(\vartheta_i|\boldsymbol{\vartheta}_{-i}) \propto \sum_{j=1}^{k^-} n_j^- \delta_{\phi_j^-}(\vartheta_i) + MG_0(\vartheta_i)$$

$$\uparrow \qquad\qquad\qquad \uparrow$$



- Conditional posterior for ϑ_i:

$$\mathcal{L}(\vartheta_i|\boldsymbol{\vartheta}_{-i}, y_i) \propto \sum_{j=1}^{k^-} f_\vartheta(y_i)\delta_{\phi_j^-}(\vartheta_i) + M\, r_i\, G_0(\vartheta_i|y_i)$$

Title

Stuff

Bibliography

- Muller, Quintana, *Bayesian Nonparametric Data Analysis*
- Neal (2000), *Markov Chain Sampling Methods for Dirichlet Process Mixture Models*
- Ishwaran, James (2001), *Gibbs Sampling Methods for Stick-Breaking Priors*