BNPlib for density estimation:

A nonparametric C++ library (part 2)

Bruno Guindani
Elena Zazzetti

January 8th, 2020

POLITECNICO
MILANO 1863

https://github.com/poliprojects/BNPlib

Model

$$
\begin{aligned}
(Y_i|\vartheta_i) &\sim F(\cdot, \vartheta_i) \\
(\vartheta_i|G) &\sim G \\
G &\sim DP(M, G_0)
\end{aligned}
$$

(hierarchical model)

$$\overset{K\to\infty}{\Longleftrightarrow}$$

$$
\begin{aligned}
(Y_i|\phi, c_i) &\sim F(\cdot, \phi_{c_i}) \\
(c_i|\mathbf{p}) &\sim \sum_{k=1}^{K} p_k \delta_k(\cdot) \\
\phi_c &\sim G_0 \\
\mathbf{p} &\sim \mathrm{Dir}(M/K, \ldots, M/K)
\end{aligned}
$$

(K-discrete model)

with $\quad \boldsymbol{\vartheta} \leftrightsquigarrow (\boldsymbol{\phi}, \mathbf{c})$

Algorithms

- Neal2, Neal8, blocked Gibbs
- **Gibbs sampling** procedures
- General structure:
  - sample **allocations** c from some conditional distribution
  - sample **unique values** ϕ from some conditional distribution
  - (sample **weighGibbs samplingts** p of the unique values deltas)

Title

- Text

Impending extensions

- **Hyperpriors**: objects of class `Hypers` store pointers to objects of class `HypersFixed`
- **Non-conjugacy**: via Stan's HMC sampler
- **R interface**: via **protocol buffers**

Protocol Buffers

- API developed by Google
- Data is saved in XML-like structures, called **messages**, that are defined in `.proto` files
- Each message corrresponds to a class in C++
- The `protoc` compiler produces the C++ files that make up the API
- RProtoBuf
- Compromise between efficiency and human-readibility

A general library?

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*
- https://developers.google.com/protocol-buffers/docs/cpptutorial