BNPlib for density estimation:

A nonparametric C++ library (part 3)

Bruno Guindani
Elena Zazzetti

February 19th, 2020

POLITECNICO
MILANO 1863

https://github.com/poliprojects/BNPlib

Title

Applications

Cluster estimation

```
unsigned int cluster_estimate();
```

$$\hat{k} = \arg\min_k \left\| D^{(k)} - \bar{D} \right\|_F^2 = \arg\min_k \sum_{i,j} \left(D_{ij}^{(k)} - \bar{D}_{ij} \right)^2$$

Density estimation

```
void eval_density(const std::vector<double> grid);
```

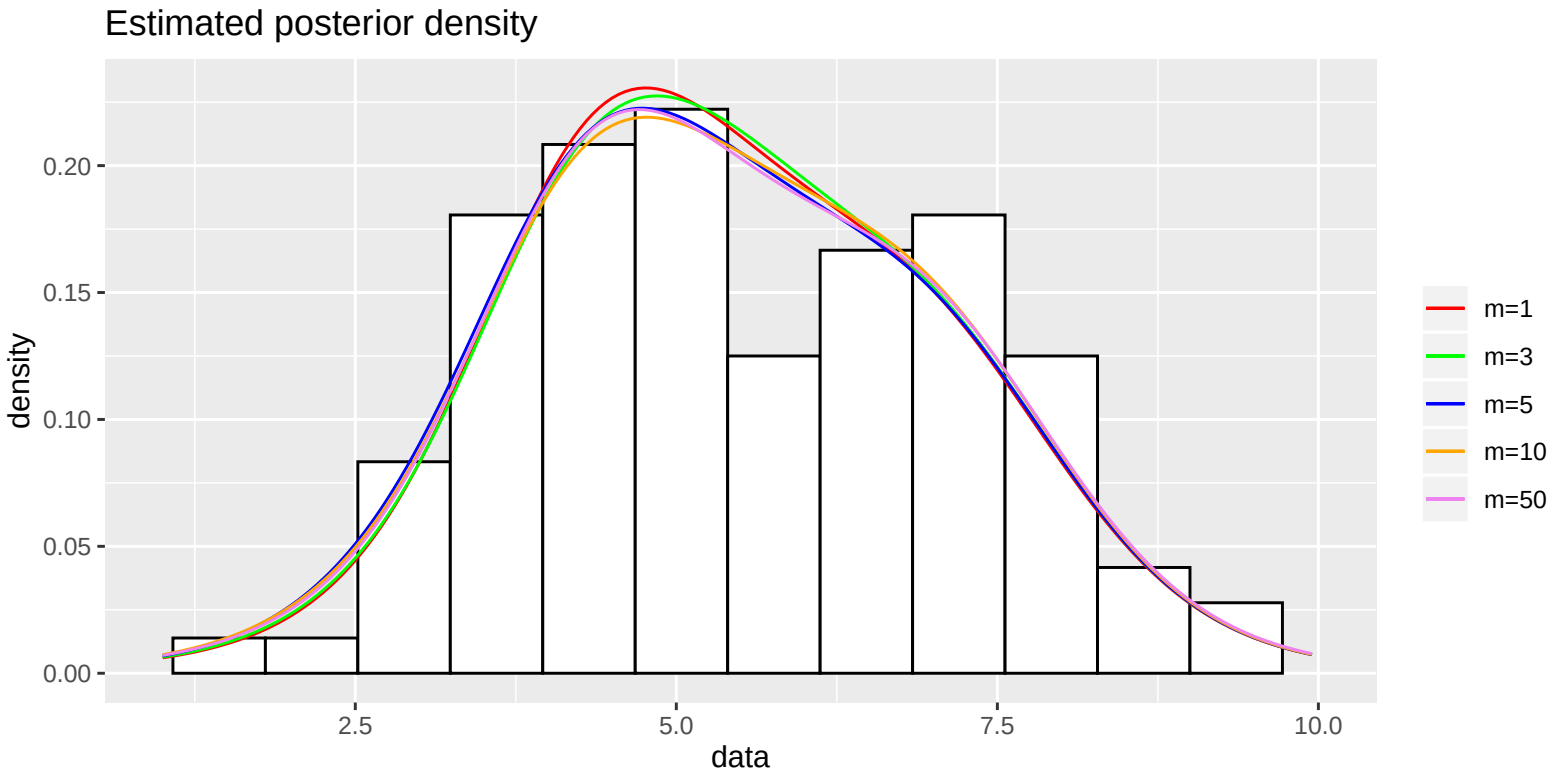
$$\hat{f}^{(k)}(x) = \sum_j \frac{n_j^{(k)}}{M+n} f\left(x|\phi_j^{(k)}\right) + \frac{M}{M+n} m(x)$$

$$\hat{m}(x) = \frac{1}{m} \sum_{h=0}^{m-1} f\left(x|\phi_h\right)$$

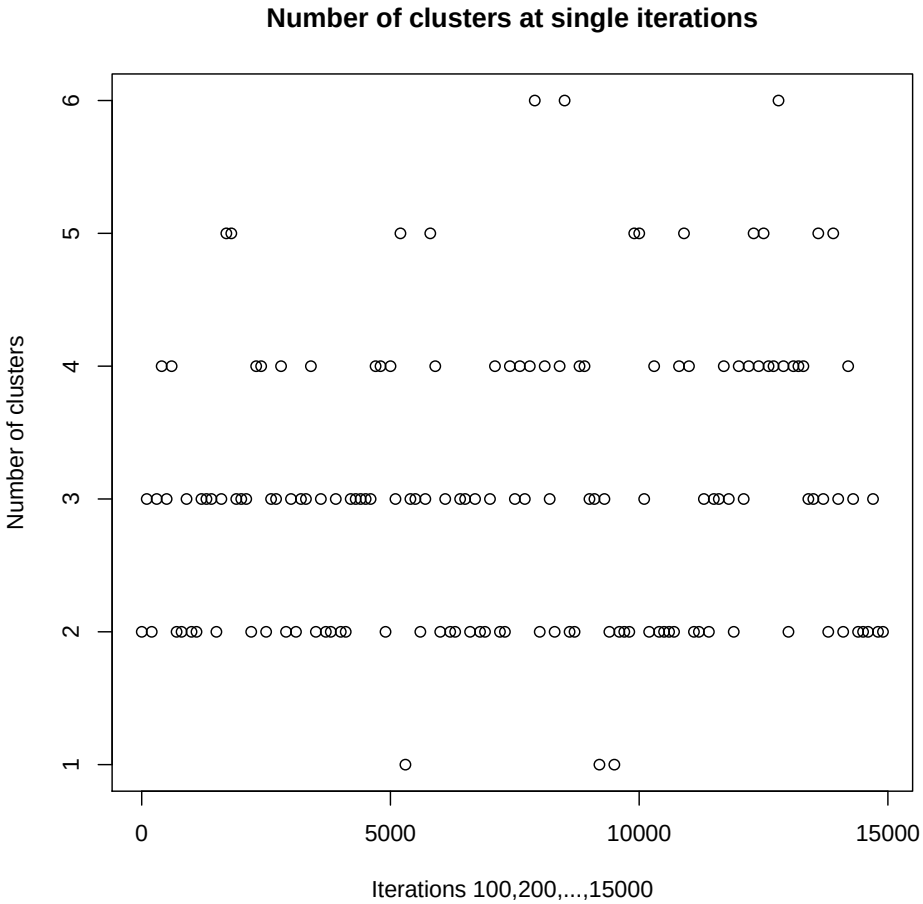
$$\Longrightarrow \hat{f}(x) = \frac{1}{K} \sum_k \hat{f}^{(k)}(x)$$

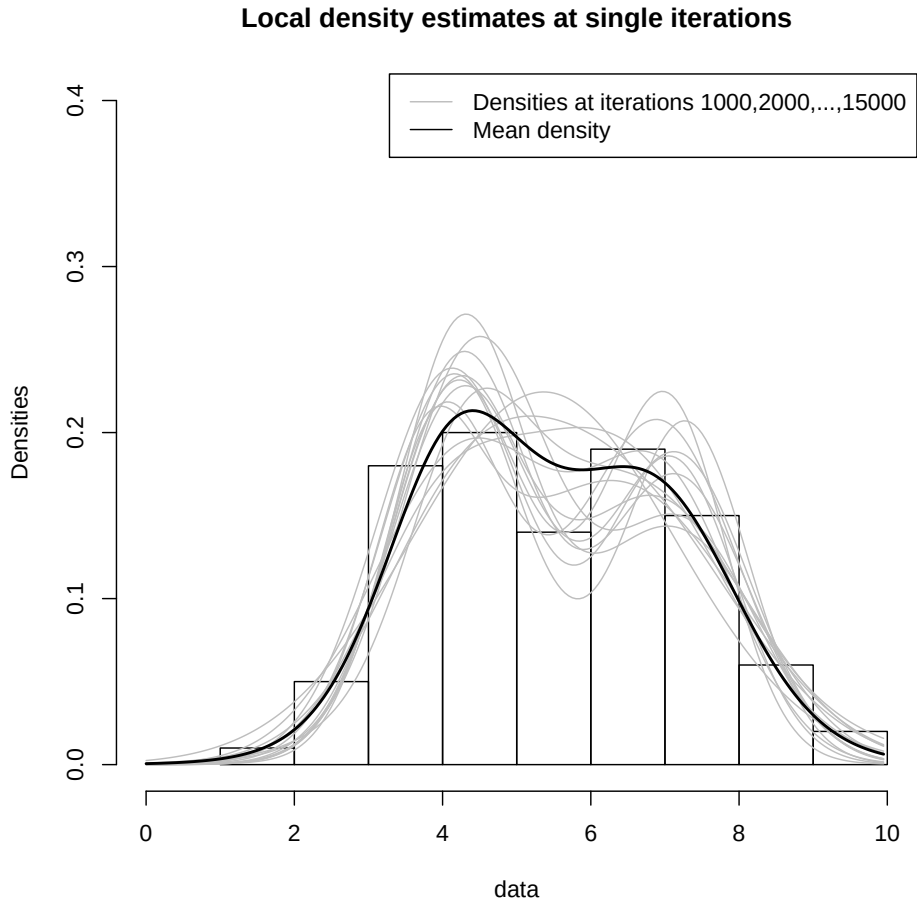
Results

Auxiliary parameters

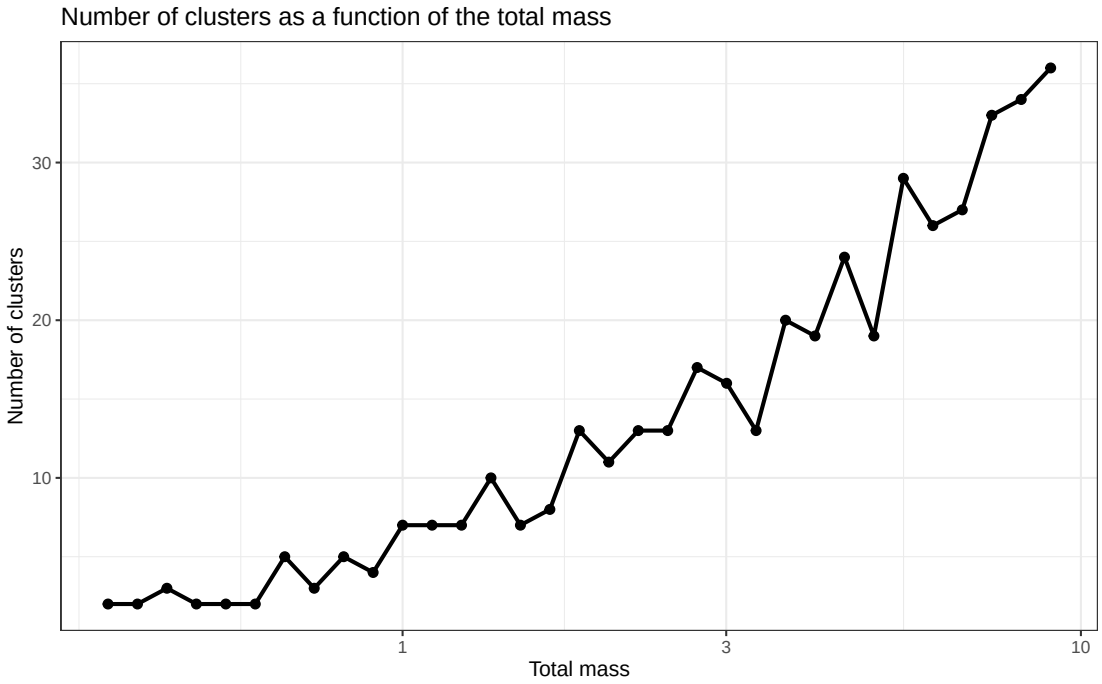


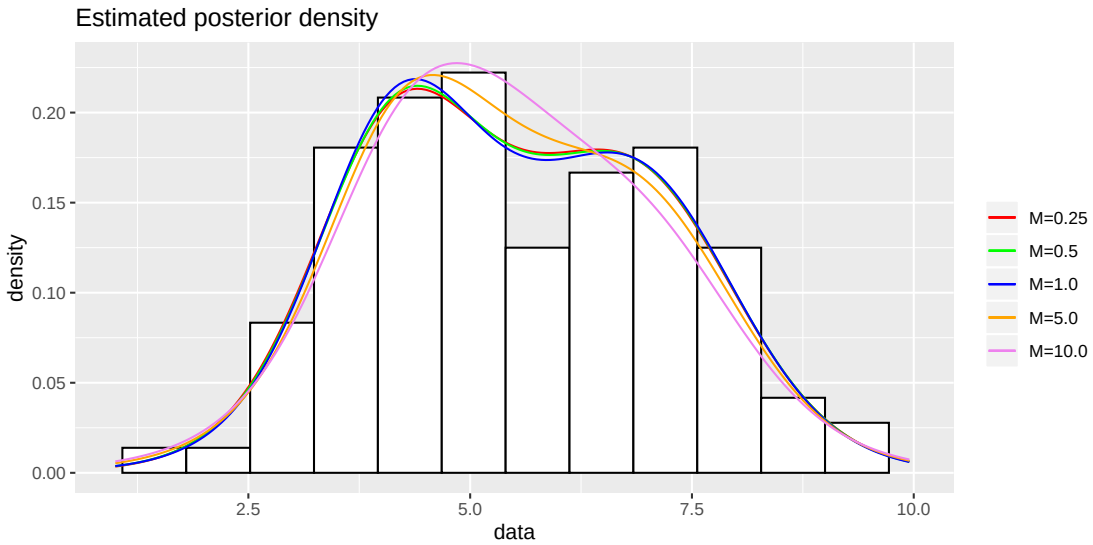
Oscillations



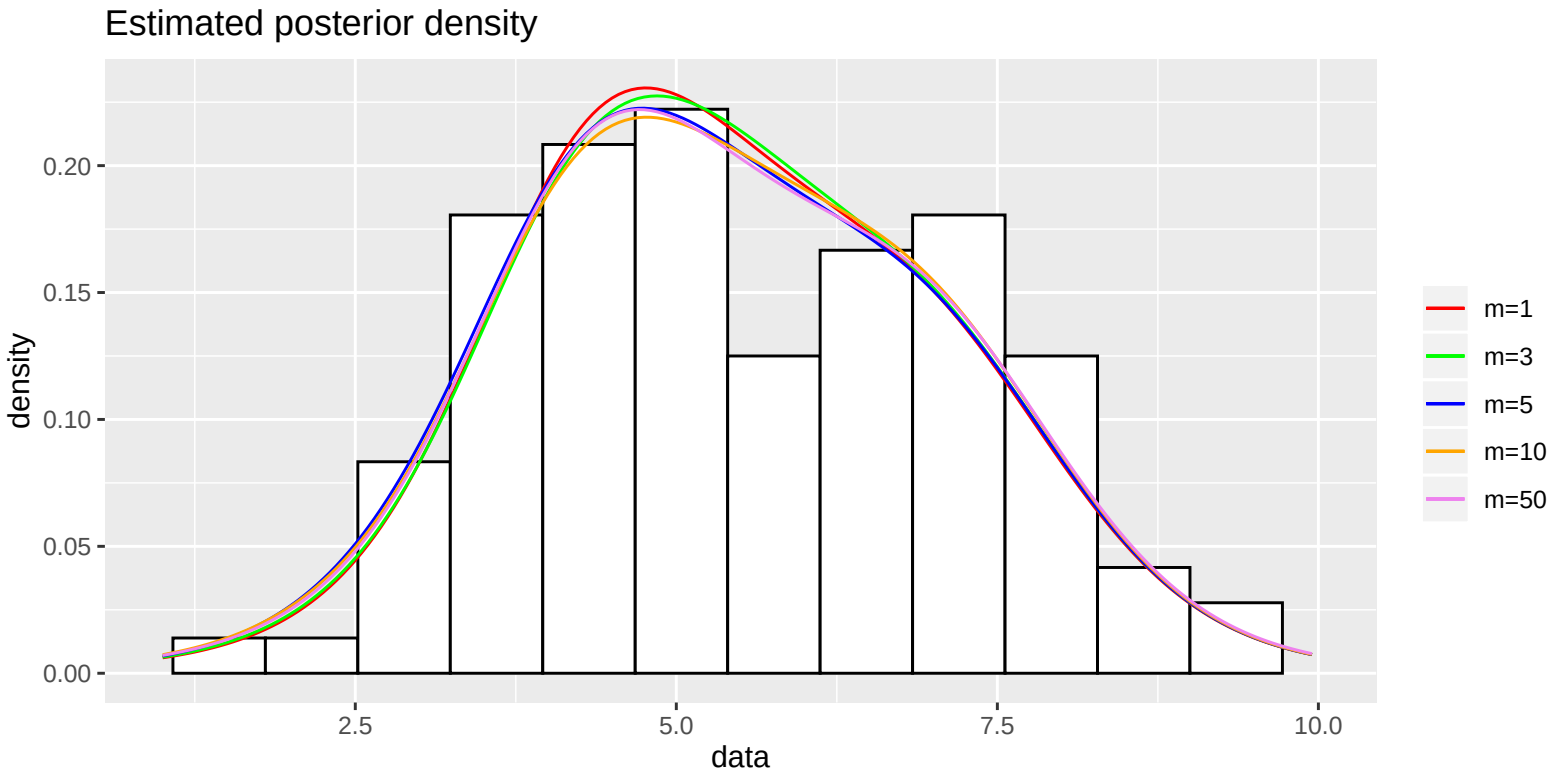


Total mass

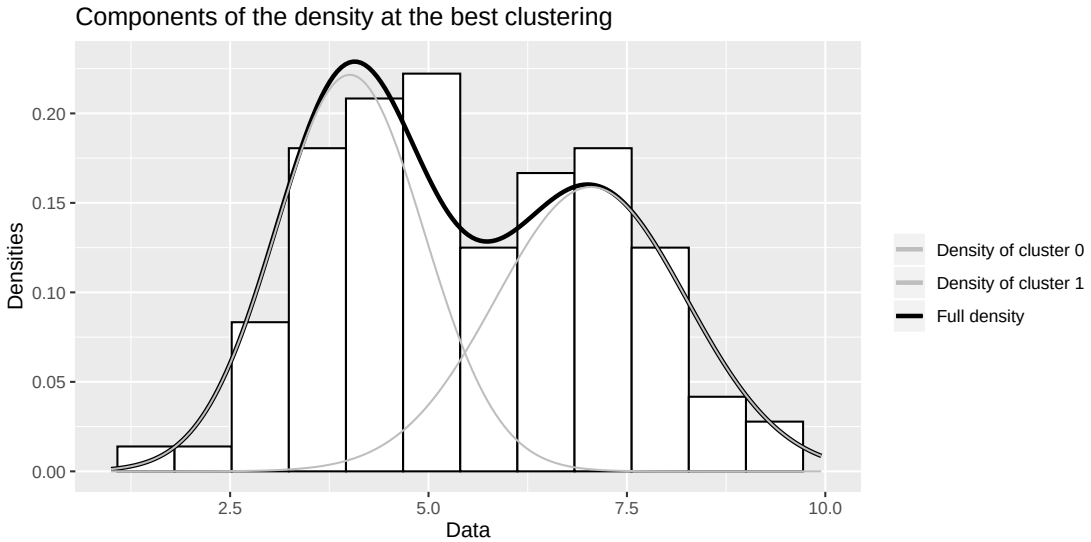

Number of clusters as a function of the total mass
Number of clusters
30
20
10
1
3
10
Total mass

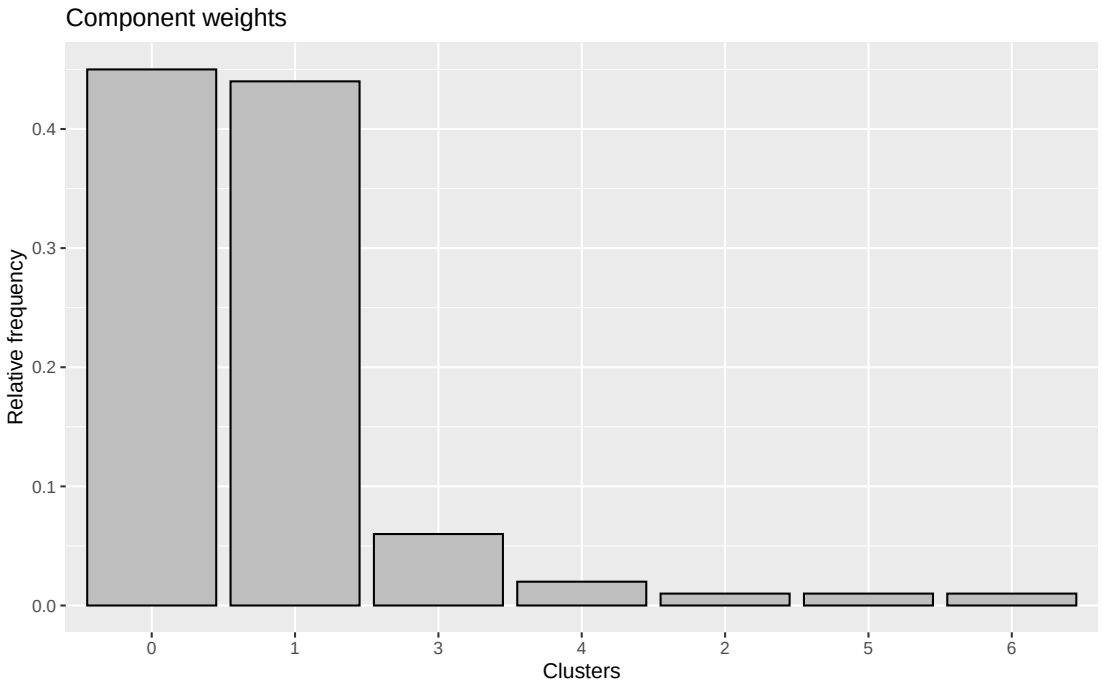

Estimated posterior density
density
0.20
0.15
0.10
0.05
0.00
2.5
5.0
7.5
10.0
data
M=0.25
M=0.5
M=1.0
M=5.0
M=10.0

Auxiliary parameters

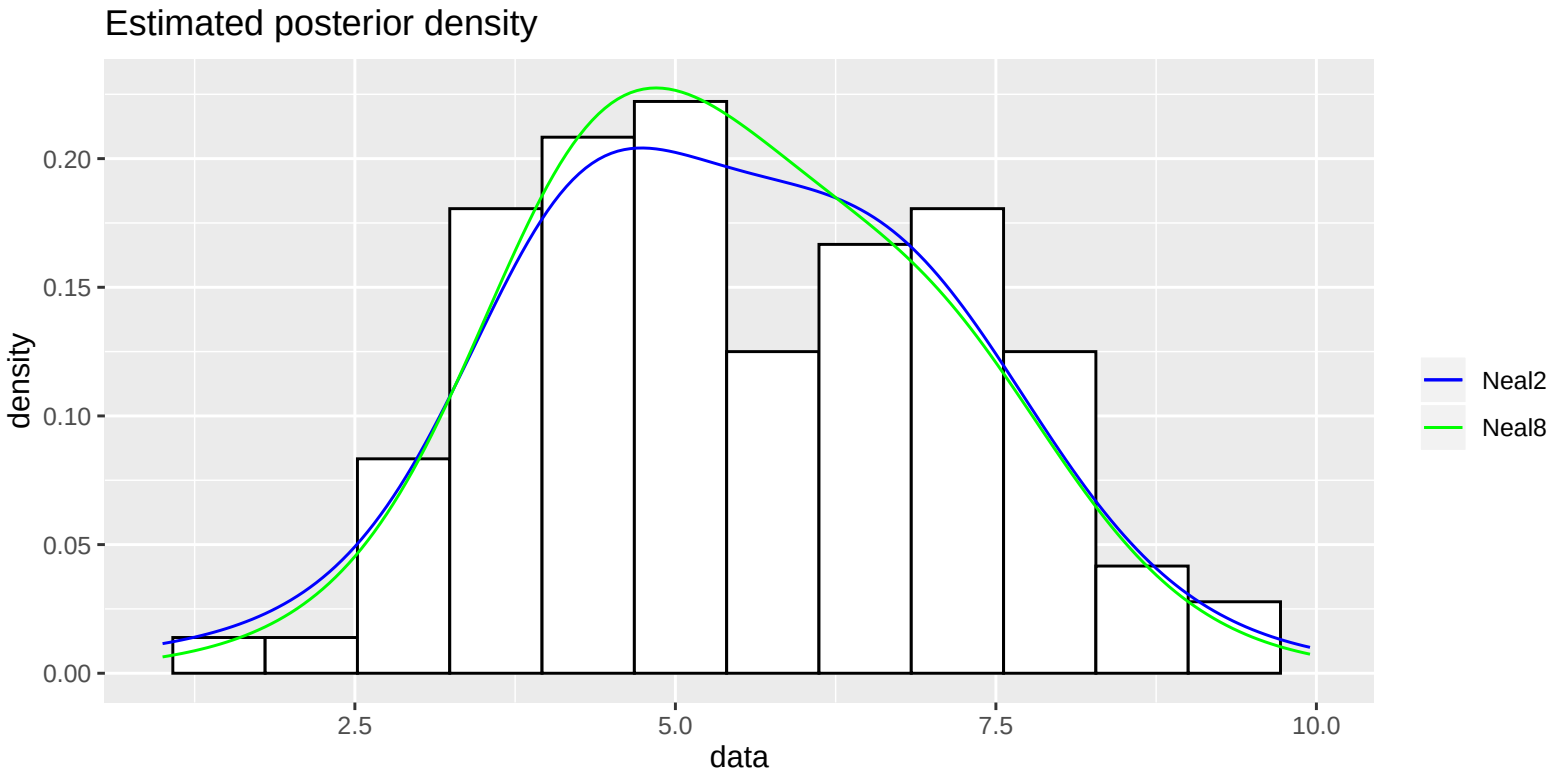


Density components





Neal2 vs Neal8



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*
- Murphy (2007), *Conjugate Bayesian analysis of the Gaussian distribution*
- Protocol Buffers: `https://developers.google.com/protocol-buffers/docs/cpptutorial`
- Stan: `http://mc-stan.org/math`
- Eigen: `https://eigen.tuxfamily.org/dox`
- GitHub codes of Mario Beraha and Riccardo Corradin for similar projects