

BayesMix: An R package for Bayesian Mixture Modelling

Bettina Grün^{1,2}, Friedrich Leisch¹

¹Institut für Statistik und Wahrscheinlichkeitstheorie, Technische Universität Wien

²supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modeling in Economics and Management Science').

Provided functionality

- Bayesian estimation of finite mixtures of univariate Gaussian distributions with MCMC methods
- accompanying material to the book Frühwirth-Schnatter (2005)
- special purpose interface to JAGS (Just Another Gibbs Sampler; Plummer, 2004)

Model class

Finite Gaussian mixture distributions are defined by

$$\mathbb{P}(y_i) = \sum_{k=1}^K \eta_k \text{Normal}(y_i | \mu_k, \sigma_k^2)$$

where $\eta_k \geq 0$, $\sum_{k=1}^K \eta_k = 1$ and $\text{Normal}(\cdot | \mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2 .

For the prior of η_k it is assumed that

$$\eta_1, \dots, \eta_K \sim \text{Dirichlet}(e_{0,1}, \dots, e_{0,K})$$

For conditionally conjugate priors we have

$$\sigma_k^2 \sim \text{InvGamma}\left(\frac{\nu_{0,k}}{2}, \frac{\nu_{0,k}S_{0,k}}{2}\right) \quad \mu_k | \sigma_k^2 \sim \text{Normal}(b_{0,k}, B_{0,k}\sigma_k^2)$$

For independence priors

$$\sigma_k^2 \sim \text{InvGamma}\left(\frac{\nu_{0,k}}{2}, \frac{\nu_{0,k}S_{0,k}}{2}\right) \quad \mu_k \sim \text{Normal}(b_{0,k}, B_{0,k})$$

For a hierarchical prior on the variances we add

$$S_{0,k} \sim \text{Gamma}\left(\frac{g_{0,k}}{2}, \frac{g_{0,k}G_{0,k}}{2}\right)$$

Implementation

Model specification:

- number of segments
- specific prior distribution and parameter values
- initial values

MCMC sampling:

- number of burn-in and monitored draws
- random seed

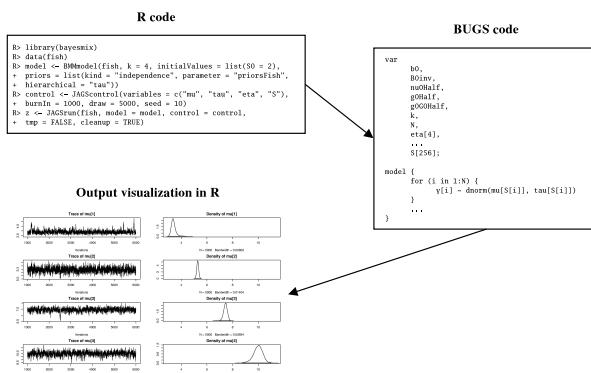


Fig. 1: Flowchart between R and JAGS: R code necessary for creating input for JAGS in BUGS syntax and visualization of results in R

Plot functions for output analysis

- diagnostic plots: for determining a suitable ordering constraint and the number of segments (see Figure 2)
- output visualization: trace and density plots (see Figure 3)
- a-posteriori probabilities (see Figure 4)

Application

"fish" dataset: (Titterington et al., 1985)

- 256 observations of fish length
- *unobserved heterogeneity* because of different age groups present in the population

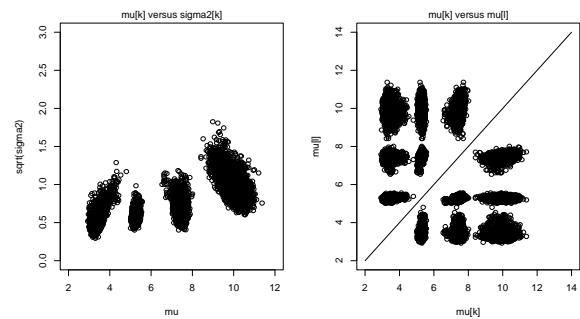


Fig. 2: Diagnostic plots: μ versus σ on the left and μ_k versus μ_l for $k \neq l$ on the right

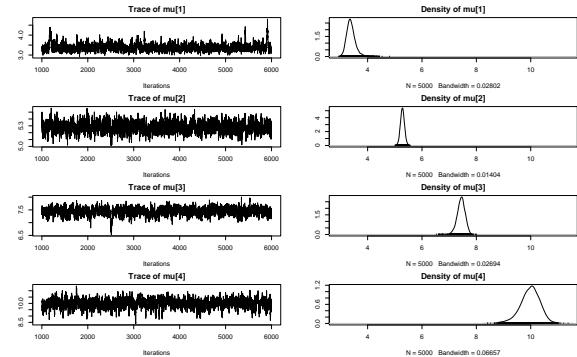


Fig. 3: Traces and density of the MCMC draws for μ under the constraint $\mu_1 < \dots < \mu_4$.

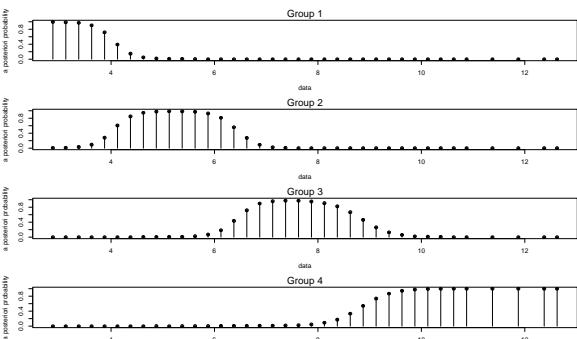


Fig. 4: Posterior classification of the observations where the MCMC draws are sorted with respect to the constraint $\mu_1 < \dots < \mu_4$.

References

- Frühwirth-Schnatter, S. 2005. *Bayesian Mixture Modelling*. Springer.
Forthcoming.
- Plummer, M. 2004. *JAGS Version 0.50 manual*. International Agency for Research on Cancer.
- Titterington, D., Smith, A., & Makov, U. 1985. *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.