A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data
SUPPLEMENT

Arnošt Komárek [1,2]

Dept. of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, CZ–186 75 Praha 8, Czech Republic

---

This document supplements the paper Komárek (2009) by some technical details which are not included in the paper. Sections in this document are numbered in the same way as in the paper. Equations from the main paper are referred with the same numbers in this document.

3 Markov chain Monte Carlo

The sample $\{\boldsymbol{\psi}^{(t)}, \boldsymbol{\theta}^{(t)}, K^{(t)} : t = 1, \ldots, T\}$ from the posterior distribution is obtained using a (reversible jump) Markov chain Monte Carlo simulation in which one iteration consists of the following move types depending on whether the number of mixture components K is prespecified or not. With K prespecified, the algorithm iterates between

(1) updating the latent (censored) observations, see subsection 3.1;
(2) updating the mixture related parameters, see subsection 3.2.

With K random (allowed currently only when $p = 1$), we iterate between

(1) updating the latent (censored) observations, see subsection 3.1;

Email address: `arnost.komarek@mff.cuni.cz` (Arnošt Komárek).
[1] Supported by a grant GAČR 201/09/P077 (Czech Science Foundation).
[2] Supported by a grant MSM 0021620839 (Ministry of Education, Youth and Sports of the Czech Republic).

(2) update of mixture related parameters as with fixed K as described in subsection 3.2;
(3) split-combine move, see subsection 3.3;
(4) birth-death move, see subsection 3.3.

Improvement in the mixing of the chains can be achieved if within one MCMC sweep, only one of the steps (2)–(4) is performed, each with a given probability π_a^{mix}, π_b^{mix} or π_c^{mix}, respectively ($\pi_a^{mix}+\pi_b^{mix}+\pi_c^{mix}=1$). This is also the user's option in the package `mixAK`.

3.1 Update of latent (censored) observations

For each i, when $\boldsymbol{y}_i^*$ contains some censored components, it is updated by sampling from the full conditional distribution which is a (multivariate) normal distribution $\mathcal{N}_p(\boldsymbol{\mu}_{r_i},\,\boldsymbol{\Sigma}_{r_i})$ constrained by the limits of the observed intervals $\boldsymbol{l}_i^*$ and $\boldsymbol{u}_i^*$. In a univariate case, this is done by the inverse cdf sampling, in a multivariate case a method described by Geweke (1991) is used.

3.2 Update of mixture related parameters without a change of K

The moves where the mixture parameters are updated, however when the number of mixture components K remains unaltered, follow largely the proposal of Diebolt and Robert (1994). In fact, all parameters are updated in blocks using a Gibbs kernel by sampling from the full conditional distribution. In the following, let $f(\cdot|\cdots)$ denote a density of the full conditional distribution. In summary we perform the following sub-moves:

Update of component allocations: sampling from

$$f(r_i\,|\,\cdots)\equiv \mathrm{P}(r_i=k\,|\,\cdots)\propto w_k\,\varphi(\boldsymbol{y}_i^*\,|\,\boldsymbol{\mu}_k,\,\boldsymbol{\Sigma}_k)\quad (k=1,\ldots,K)$$

sequentially for $i=1,\ldots,n$.

Update of mixture weights: sampling from the Dirichlet distribution $\mathrm{D}(\delta+N_1,\,\ldots,\,\delta+N_k)$, where

$$N_k=\sum_{i=1}^{n} I[r_i=k]\qquad (k=1,\ldots,K).$$

Update of mixture precision matrices: sampling from the Wishart distribution $\mathrm{W}_p(\tilde{\zeta}_k,\,\tilde{\boldsymbol{\Xi}}_k)$ sequentially for $k=1,\ldots,K$ to update $\boldsymbol{Q}_k$. For both priors (7) and (8), degrees of freedom of the k-th full conditional Wishart distribution are equal $\tilde{\zeta}_k=\zeta+N_k$. In the case of semiconjugate independent prior (7), the scale matrix of the k-th full conditional Wishart distribution is

$$\tilde{\boldsymbol{\Xi}}_k = \Big\{\boldsymbol{\Xi}^{-1} + \sum_{i:r_i=k} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_k)(\boldsymbol{y}_i^* - \boldsymbol{\mu}_k)'\Big\}^{-1}.$$

In the case of natural-conjugate prior (8), the scale matrix of the k-th full conditional Wishart distribution is

$$\tilde{\boldsymbol{\Xi}}_k = \Big\{\boldsymbol{\Xi}^{-1} + \sum_{i:r_i=k} (\boldsymbol{y}_i^* - \bar{\boldsymbol{y}}_k^*)(\boldsymbol{y}_i^* - \bar{\boldsymbol{y}}_k^*)' + \frac{N_k\, c_k}{N_k + c_k}(\bar{\boldsymbol{y}}_k^* - \boldsymbol{\xi}_k)(\bar{\boldsymbol{y}}_k^* - \boldsymbol{\xi}_k)'\Big\}^{-1},$$

where $\bar{\boldsymbol{y}}_k^* = N_k^{-1} \sum_{i:r_i=k} \boldsymbol{y}_i^*$.

Update of mixture means: sampling from the normal distribution $\mathcal{N}_p(\boldsymbol{m}_k,\, \boldsymbol{S}_k)$ sequentially for $k = 1, \ldots, K$ to update $\boldsymbol{\mu}_k$. For semiconjugate independent prior (7), parameters of the k-th full conditional normal distribution are

$$\boldsymbol{S}_k = \Big(N_k \boldsymbol{Q}_k + \boldsymbol{D}_k^{-1}\Big)^{-1}, \qquad \boldsymbol{m}_k = \boldsymbol{S}_k\Big(N_k \boldsymbol{Q}_k \bar{\boldsymbol{y}}_k^* + \boldsymbol{D}_k^{-1}\boldsymbol{\xi}_k\Big).$$

For natural-conjugate prior (8), parameters of the k-th full conditional normal distribution are

$$\boldsymbol{S}_k = \Big\{(N_k + c_k)\boldsymbol{Q}_k\Big\}^{-1}, \qquad \boldsymbol{m}_k = (N_k + c_k)^{-1}(N_k \bar{\boldsymbol{y}}_k^* + c_k \boldsymbol{\xi}_k).$$

Update of variance hyperparameter: sampling from gamma distributions $\mathrm{Gamma}(\tilde{g}_j,\, \tilde{h}_j)$ sequentially for $j = 1, \ldots, p$ to update γ_j. Parameters of the j-th full conditional gamma distribution are

$$\tilde{g}_j = g_j + \frac{K\zeta}{2}, \qquad \tilde{h}_j = h_j + \frac{1}{2}\sum_{k=1}^{K} \boldsymbol{Q}_k(j,\, j),$$

where $\boldsymbol{Q}_k(j,\, j)$ denotes the j-th diagonal element of matrix $\boldsymbol{Q}_k$.

3.3 Moves allowing a change of the number of mixture components

For univariate data ($p = 1$), moves allowing a change of the number of mixture components follow the RJ-MCMC approach taken in Richardson and Green (1997). Let $\mu_1, \ldots, \mu_K$ be the (univariate) mixture means and $\sigma_1^2, \ldots, \sigma_K^2$ be the (univariate) mixture variances. Two move types are proposed.

Split–combine move consists of either splitting an existing component into two new components or combining two existing components into a new one. Firstly a random choice is made whether to perform the split or combine move. Namely, given K, the probability of attempting the split move is π_K^{split} and the probability of attempting the combine move is $\pi_K^{combine} = 1 - \pi_K^{split}$.

Default values in the package `mixAK` are $\pi_1^{split} = 1$, $\pi_2^{split} = \cdots = \pi_{K_{max}-1}^{split} = 0.5$ and $\pi_{K_{max}}^{split} = 0$ and can possibly be modified by the user.

When a *combine* move is attempted, a pair of neighboring components is chosen, let say k_1 and k_2 such that $\mu_{k_1} < \mu_{k_2}$ and there is no other mixture mean in the interval $[\mu_{k_1},\, \mu_{k_2}]$. The two components are combined such that a new component, let say k^*, has the same 0th, 1st and 2nd moment as the original two components. That is, the weight w_{k^*}, the mean μ_{k^*} and the variance $\sigma_{k^*}^2$ of the new component are given by

$$w_{k^*} = w_{k_1} + w_{k_2}, \qquad \mu_{k^*} = \frac{w_{k_1}\mu_{k_1} + w_{k_2}\mu_{k_2}}{w_{k^*}},$$
$$\sigma_{k^*}^2 = \frac{w_{k_1}(\mu_{k_1}^2 + \sigma_{k_1}^2) \, + \, w_{k_2}(\mu_{k_2}^2 + \sigma_{k_2}^2)}{w_{k^*}} \, - \, \mu_{k^*}^2.$$

The move into a new state with components k_1 and k_2 replaced by the component k^* and the number of mixture components K decreased by 1 is then either accepted or rejected with a certain Metropolis-Hastings probability, see Richardson and Green (1997) for more details. In the case of acceptance, (latent) observations allocated originally into the components k_1 and k_2 are re-allocated in the new component k^*.

In the *split* move, a component, let say k^* is chosen at random from existing K components and two new components, let say k_1 and k_2 are proposed such that the split move is reversible to the combine move is a sense described generally in Green (1995). The weights w_{k_1}, w_{k_2}, the means μ_{k_1}, μ_{k_2}, and the variances $\sigma_{k_1}^2$, $\sigma_{k_2}^2$ of the two new components are given by

$$\begin{aligned}
w_{k_1} &= w_{k^*} u_1, & w_{k_2} &= w_{k^*}(1 - u_1),\\
\mu_{k_1} &= \mu_{k^*} - u_2 \sigma_{k^*} \sqrt{\frac{w_{k_2}}{w_{k_1}}}, & \mu_{k_2} &= \mu_{k^*} + u_2 \sigma_{k^*} \sqrt{\frac{w_{k_1}}{w_{k_2}}},\\
\sigma_{k_1}^2 &= u_3(1 - u_2^2)\sigma_{k^*}^2 \frac{w_{k^*}}{w_{k_1}}, & \sigma_{k_2}^2 &= (1 - u_3)(1 - u_2^2)\sigma_{k^*}^2 \frac{w_{k^*}}{w_{k_2}},
\end{aligned}$$

where $\boldsymbol{u} = (u_1,\, u_2,\, u_3)'$ is an auxiliary random vector whose components are generated randomly independently, u_l from Beta$(a_l,\, b_l)$ $(l = 1, 2, 3)$. Default values in package `mixAK` are $a_1 = b_1 = 2$, $a_2 = 1$, $b_2 = 2$, $a_3 = b_3 = 1$. Additionally, for (latent) observations allocated in the component k^* a new allocation, either k_1 or k_2 is randomly proposed. The whole proposal is accepted with a certain Metropolis-Hastings probability as described in detail by Richardson and Green (1997). Note that the proposal is rejected with probability one if there is other mixture mean in the interval $[\mu_{k_1},\, \mu_{k_2}]$ to ensure reversibility with the combine move.

Birth–death move consists of either proposing a new component (birth) or deleting one of the empty components (death). Similarly to the split–combine move, a random choice is made whether to perform the birth or the death move. Namely, given K, the probability of attempting the birth move is π_K^{birth} and the probability of attempting the death move is $\pi_K^{birth} = 1 - \pi_K^{death}$. Default values in the package `mixAK` are $\pi_1^{birth} = 1$, $\pi_2^{birth} = \cdots = \pi_{K_{max}-1}^{birth} = 0.5$ and $\pi_{K_{max}}^{birth} = 0$ and can possibly be modified by the user.

When a *birth* move is attempted a new component weight w_{k^*}, mean μ_{k^*} and variance $\sigma_{k^*}^2$ are sampled from the prior distribution, the weights of existing components are rescaled to satisfy the sum-up-to-one constraint and the whole proposal is accepted with a certain Metropolis-Hastings probability computed along the lines described in Richardson and Green (1997).

When a *death* move is attempted one of empty components, i.e. components into which currently there are no observations allocated and hence $N_k = 0$, is chosen at random. The proposal consists of deleting this component and rescaling the remaining weights to satisfy the sum-up-to-one constraint. Analogously to the previous cases, the whole proposal is accepted with a certain Metropolis-Hastings probability, see Richardson and Green (1997) for details.

References

Diebolt, J., Robert, C. P., 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society, Series B 56, 363–375.

Geweke, J., 1991. Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. Computer Sciences and Statistics 23, 571–578.

Green, P. J., 1995. Reversible jump Markov chain computation and Bayesian model determination. Biometrika 82, 711–732.

Komárek, A., 2009. A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data. Computational Statistics and Data Analysis 53 (12), 3932–3947.

Richardson, S., Green, P. J., 1997. On Bayesian analysis of mixtures with unknown number of components (with Discussion). Journal of the Royal Statistical Society, Series B 59, 731–792.