Random Effect Modelling with Mode Trees

Jochen Einbeck and John Hinde

National University of Ireland, Galway

{jochen.einbeck, john.hinde}@nuigalway.ie

Mikulov, 19th July 2005

Work supported by science foundation ireland
fondúireacht eolaíochta éireann

Framework: Generalized linear model with random effect

$$\mu_i \equiv E(y_i|z_i,\beta) = h(\eta_i) \equiv h(x_i'\beta + z_i).$$

The marginal likelihood can be approximated by a finite mixture (Laird, 1978)

$$L(\beta, g(z)) = \prod_{i=1}^{n} \int f(y_i|z_i,\beta) g(z_i)\, dz_i \approx \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} f(y_i|z_k,\beta)\pi_k \right\}$$

with mass points z_k and masses π_k.

Two well-known approaches:

- $Z \sim N(0,\sigma^2) \longrightarrow$ Gaussian Quadrature (Hinde, 1982)
- No distributional assumption about $Z \longrightarrow$ Nonparametric maximum likelihood (NPML, Aitkin, 1996).

A special case: Fitting finite Gaussian mixtures

Given: A set of observations $y_1, \ldots, y_n$ with

- $y_i \sim N(z_i, \sigma_i^2), i = 1, \ldots, n$
- $z_i \sim Z$, where Z is left unspecified.

Random effect model without fixed term:

$$\mu_i = E(y_i|z_i) = z_i$$

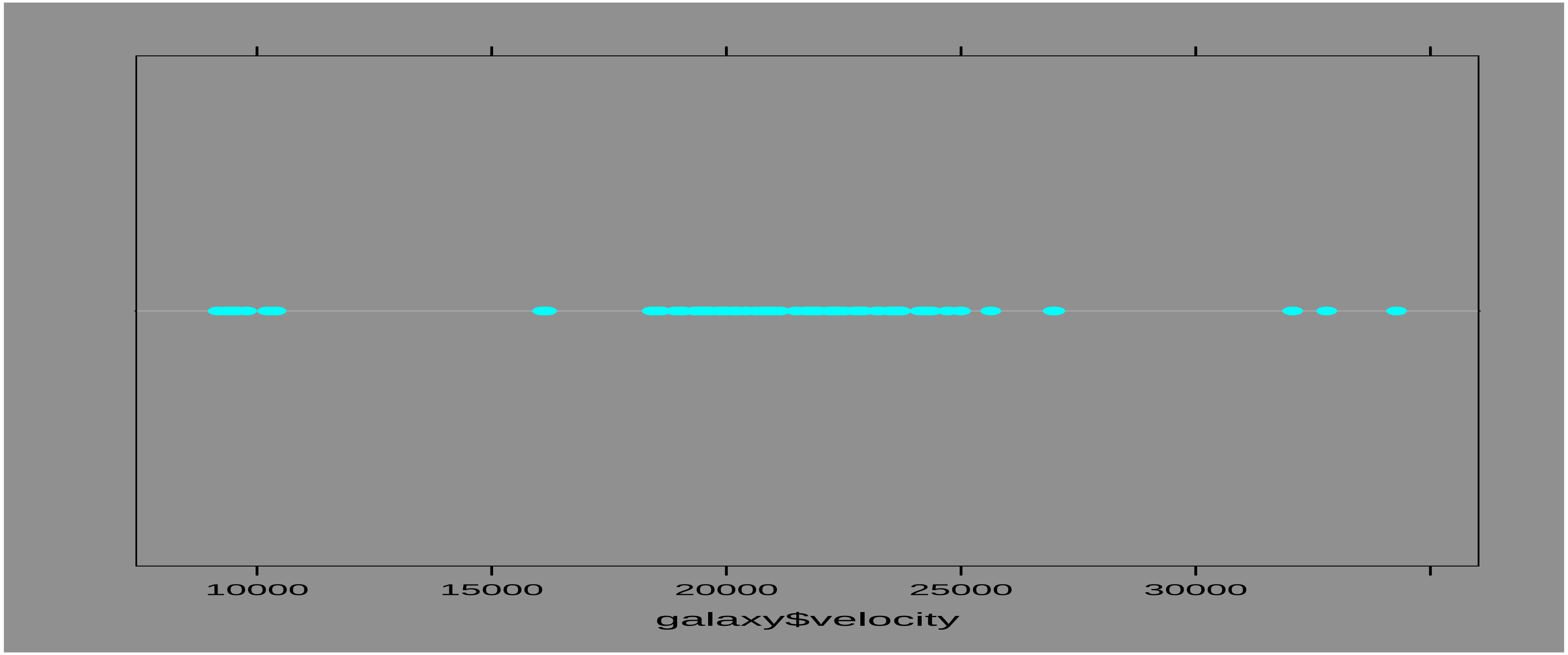
Mixture density:

$$f(y|(z_k, \pi_k, \sigma_k)_{k=1,\ldots,K}) = \sum_{k=1}^{K} \pi_k f(y|z_k, \sigma_k^2)$$

where $f(y|z_k, \sigma_k^2)$ is a normal density with mean z_k and standard deviation σ_k.

Example: Galaxy Data

Recession velocities (in km/s) of 82 galaxies.



Aim: Estimate the mass points z_k, the variances σ_k^2, the masses π_k, and the number of components K.

NPML Estimation

For fixed K, consider the log-likelihood

$$\ell = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k f(y_i | z_k, \sigma_k^2) \right\},$$

and calculate the score equations

$$\frac{\partial \ell}{\partial z_k} = 0, \quad \frac{\partial \ell - \lambda(\sum \pi_k - 1)}{\partial \pi_k} = 0, \quad \frac{\partial \ell}{\partial \sigma_k} = 0,$$

which turn out to be weighted versions of the single-distribution score equations.

$\Longrightarrow$ can be solved by standard EM algorithm:

Starting points Select starting values z_k^0, π_k^0, and σ_k^0, $k = 1, \ldots, K$.

E-Step Adjust weights given current parameter estimates.

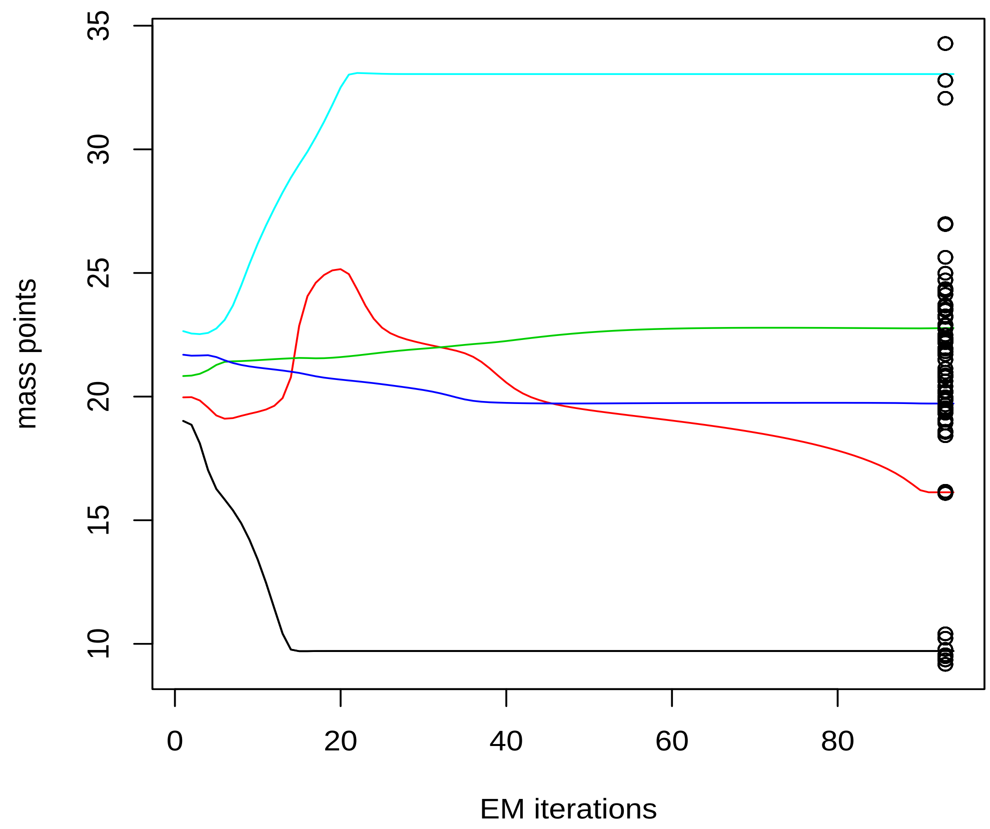
M-Step Update parameter estimates.

Application on galaxy data

Set e.g. K=5:

```
Coefficients:
 MASS1  MASS2  MASS3  MASS4  MASS5

  9.71  16.13  22.78  19.72  33.04

Mixture proportions:

 0.085  0.024  0.512  0.342  0.037

Standard deviations:

 0.423  0.043  1.721  0.626  0.922


-2 log L:     380.9
```

EM Trajectories:



Properties of current NPML implementations (as in GLIM 4)

- The algorithm is simple, converges in every case, and is *"impressively stable"* (Aitkin, 1996).

- Results depend heavily on the choice of starting points z_k^0, usually defined as

$$z_k^0 = \bar{y} + \texttt{tol} * \hat{\sigma} * g_k$$

  where `tol`: scaling parameter, g_k: Gauss-Hermite mass points, $\hat{\sigma} = \frac{1}{n}\sum(y_i - \bar{y})^2$.

  Finding the optimal solution requires a tedious grid search for `tol`.

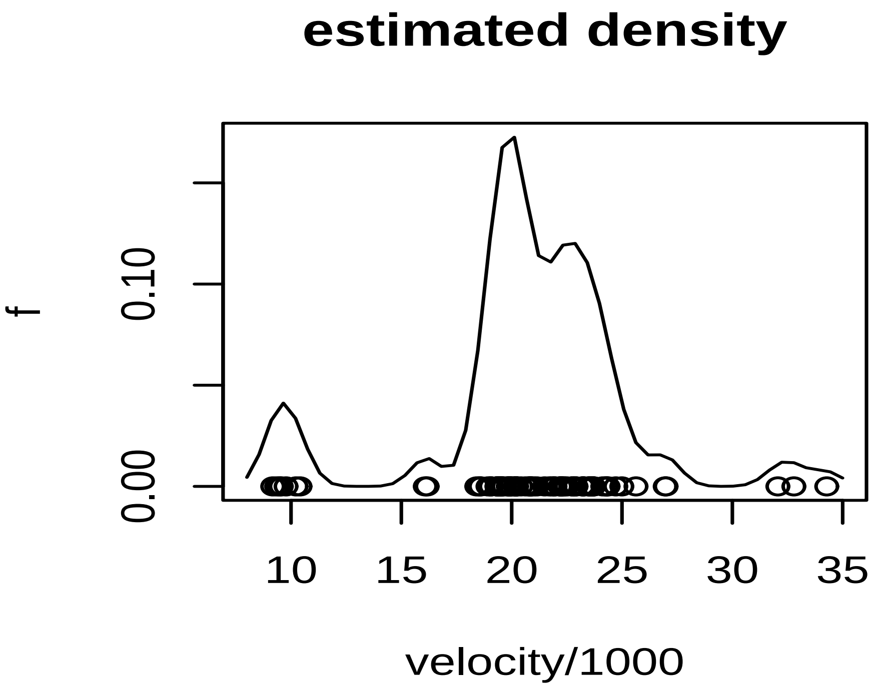
- The EM trajectories behave quite erratically in the first cycles, and tend to cross.

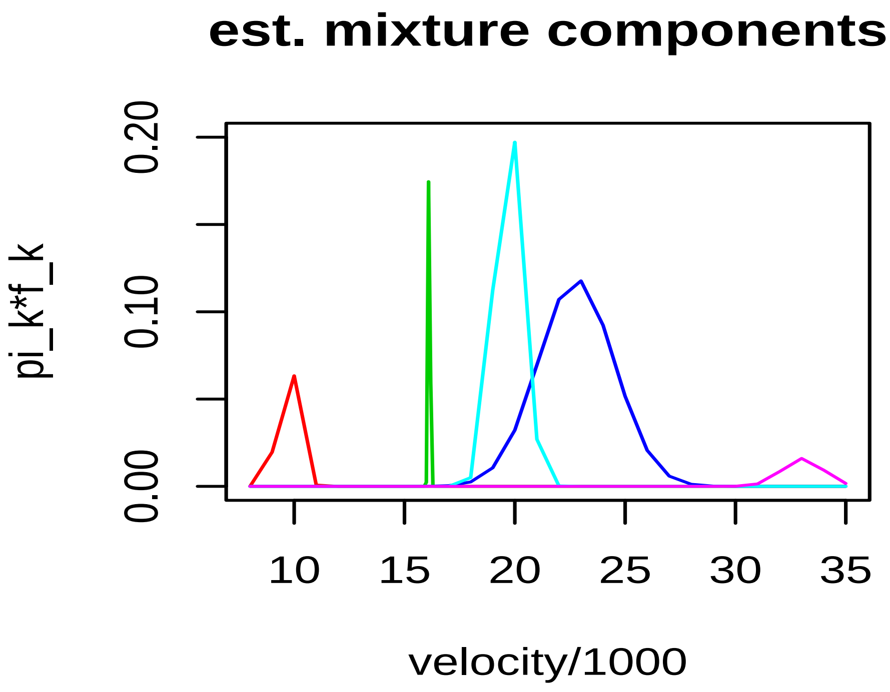
- There does not exist any automatic routine to select K. Richardson & Green, 1997:

  "one of the things you do not know is the number of things that you do not know"

Exploiting the multimodal structure

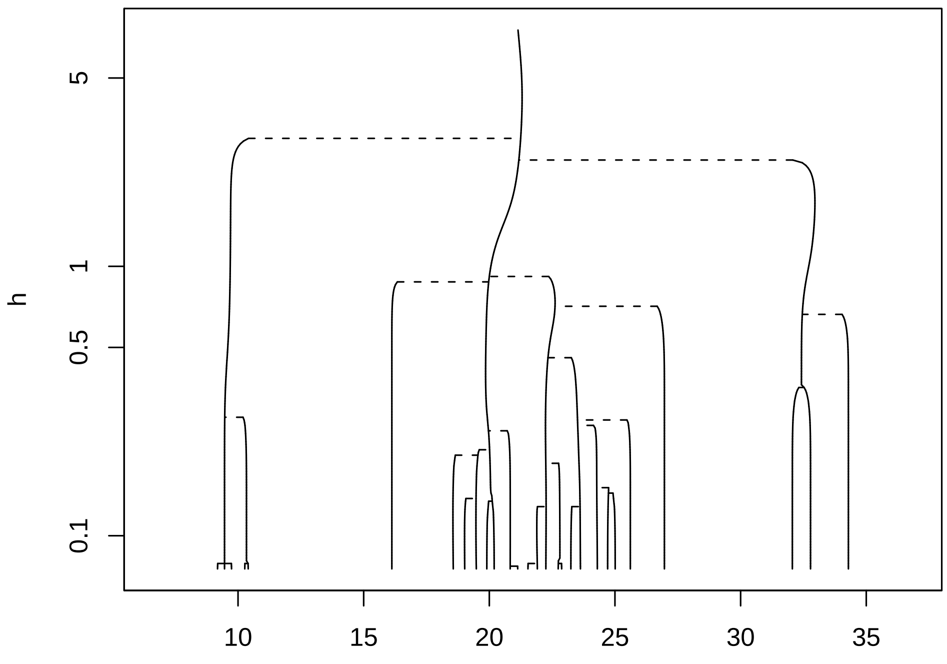
Idea: Consider density estimate $\hat{f}(y,h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right)$.





Carreira-Perpiñan & Williams, 2003: The number of modes is a *lower bound* for the number of components of a Gaussian mixture. But: Number of modes depends on bandwidth h.

The mode tree (Minnotte & Scott, 1993)



y_i (velocity/1000)

$\longleftarrow$ fixed effect model $(K = 1)$

⋮

$\longleftarrow$ random effect models $(1 < K < 82)$

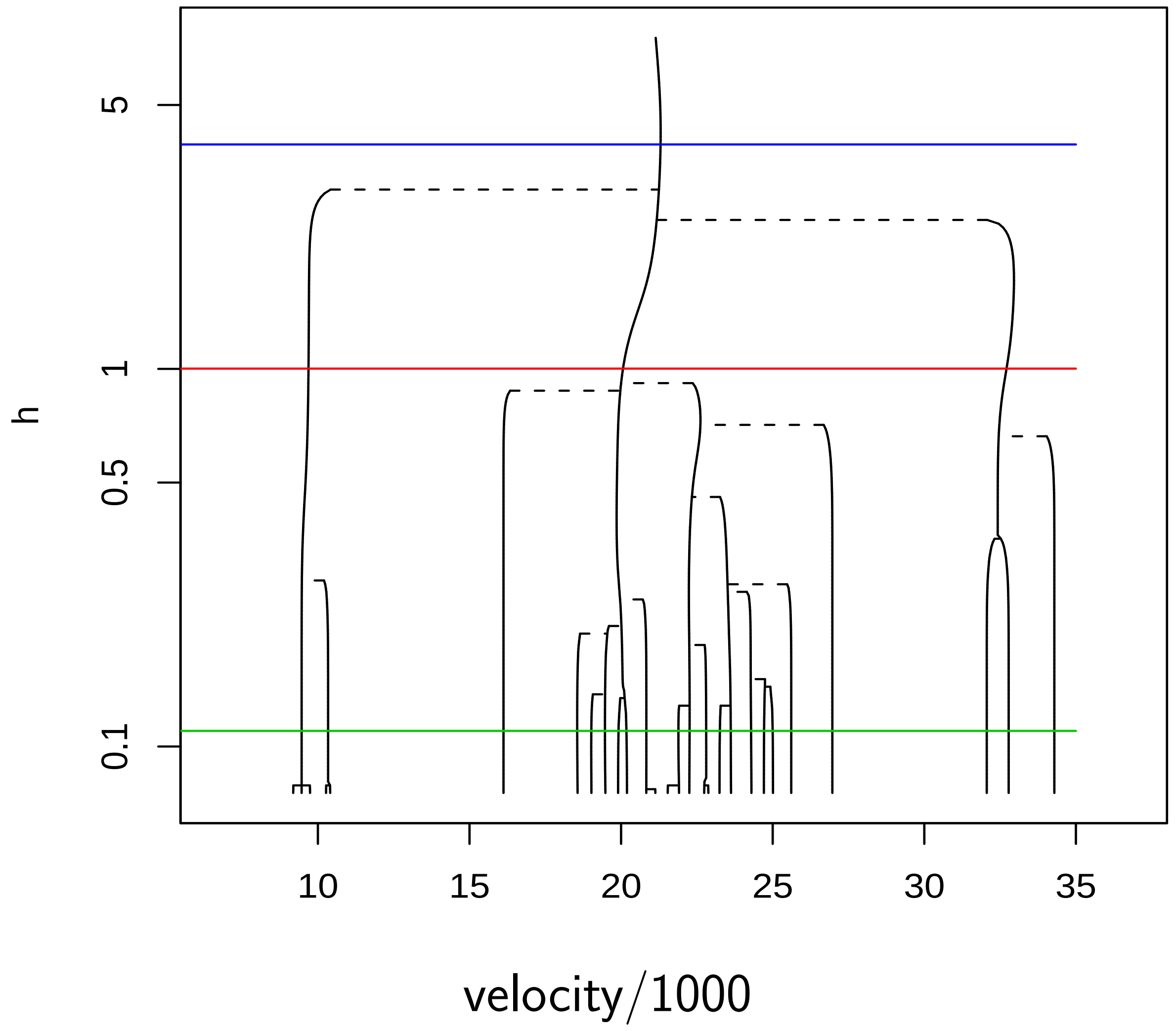
⋮

$\swarrow$ saturated model $(K = 82)$

More generally, applied on the 'residuals' $h^{-1}(y_i) - x_i'\hat{\beta}$ of a GLM:

"zoom into the random effect distribution"

Examples for bandwidth selection



Bandwidth selectors:

BCV (1 mode)

Silverman (3 modes)

AIC (22 modes)

Bandwidth selection in 2 steps

- Calculate Silverman's optimal bandwidth

$$h_{opt} = 0.9An^{-1/2},$$

where $A = \min\{\hat{\sigma}, IQR/1.34\}$

- From that bandwidth, climb down the mode tree until the next critical bandwidth

(Silverman, 81)

$$h_{crit} = \inf\{h, \hat{f}(\cdot, h) \text{ has at most } k \text{ modes}\}$$

is reached.

Using h_{crit}, the mode tree gives

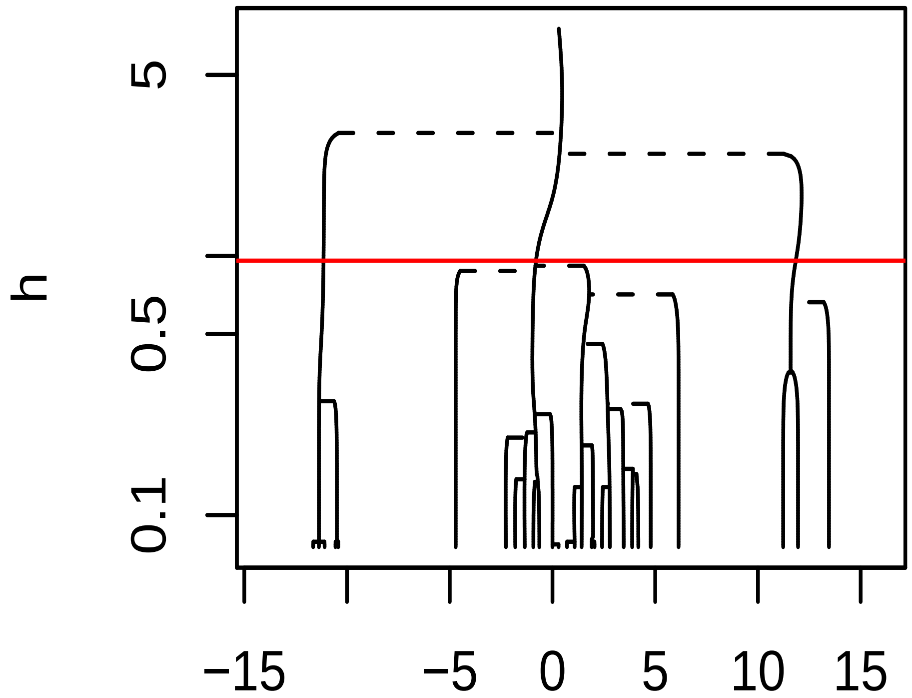
- an estimate for the number of modes, and thus for the number of components
- a very accurate estimate for the location of the mass points, which can be either used directly as mass point estimates $\hat{z}_k$, or as starting points z_k^0 for the EM algorithm. In the latter case, one has to damp the standard deviation in the initial loops of the EM algorithm via

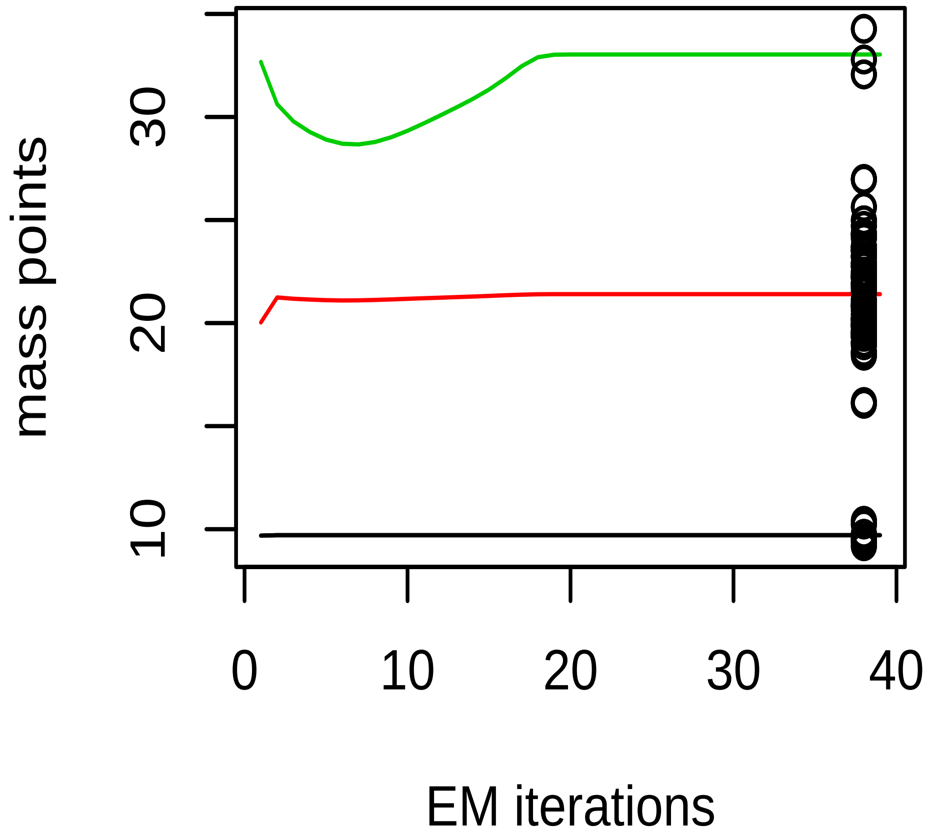
$$d_j = 1 - (1 - \texttt{tol})^j, \qquad (0 < \texttt{tol} \leq 1)$$

as otherwise the EM trajectories might be kicked out off the optimal solution immediately.

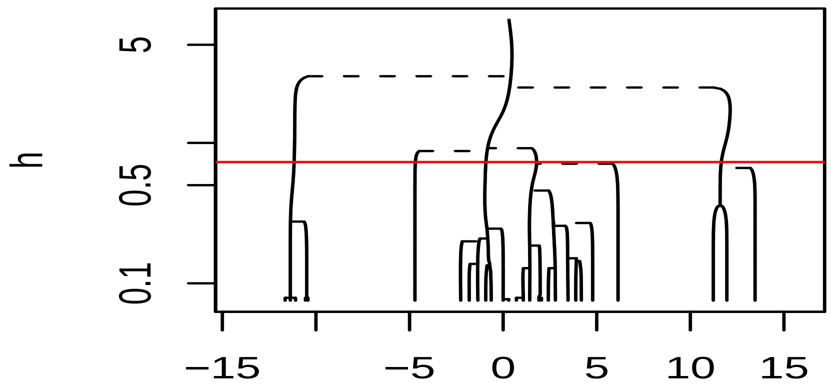
Application on galaxy data

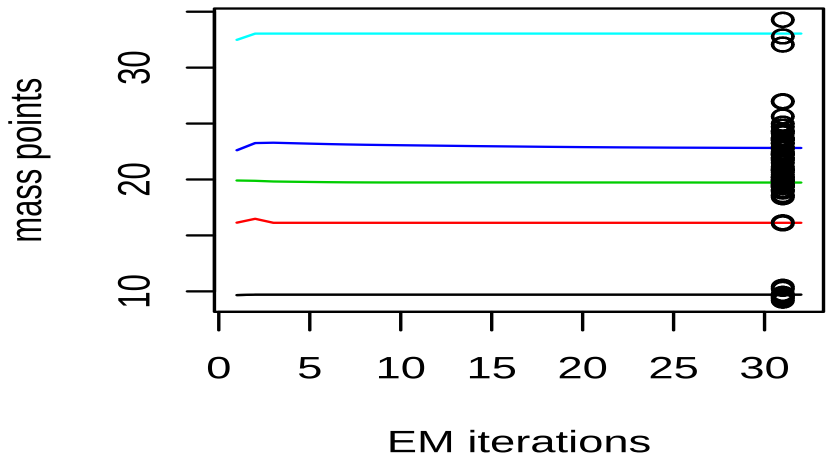
Mode tree, critical bandwidth, and EM trajectories:

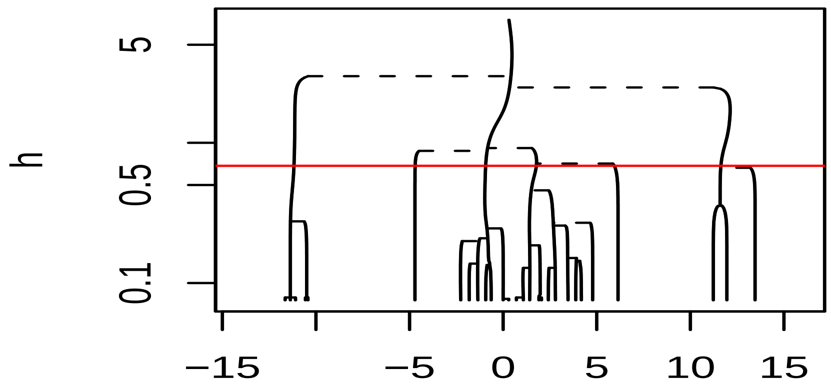


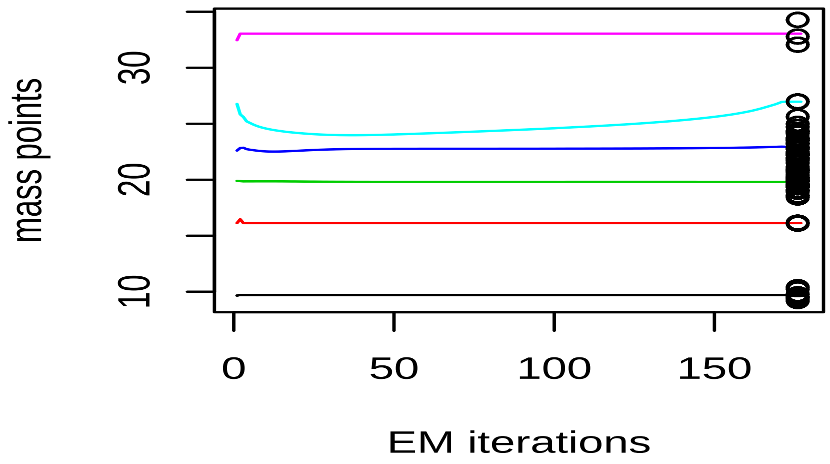


Climbing down the tree



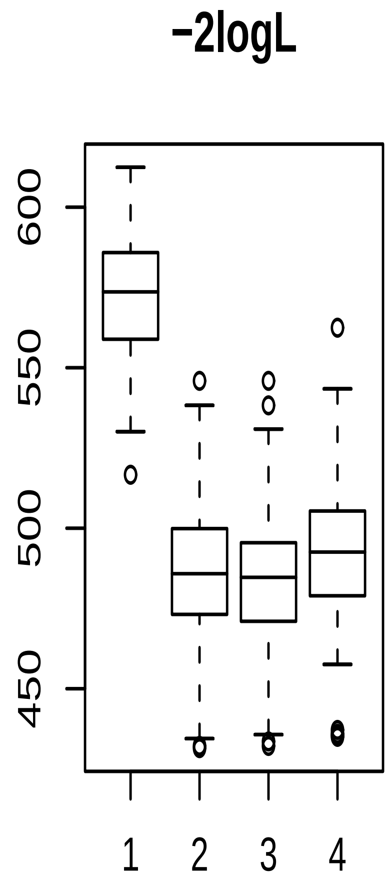


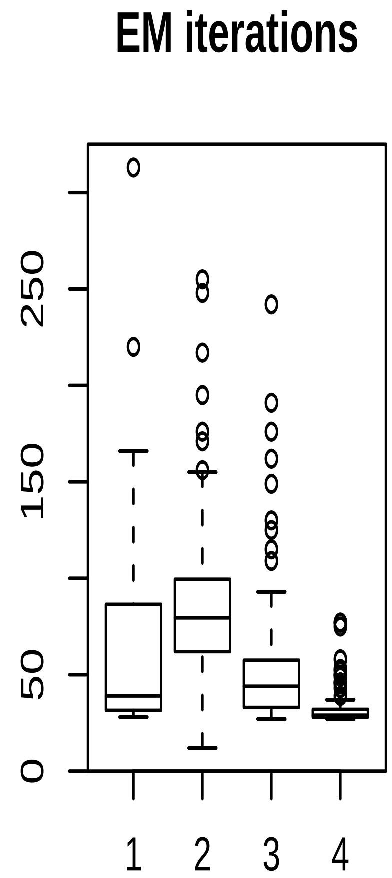




Simulation study: 100 samples from fitted mixture distribution $(\sigma = \sigma_k = const)$.

Starting points z_k^0	Gauss-Hermite		Density Modes	
	↓	↓	↓	↓
Mass Points z_k	$= z_k^0$	via EM algorithm		$= z_k^0$
Estimation Method	GQ	NPML	NPML	Mode Trees
	(1)	(2)	(3)	(4)





Summary

- The mode tree is a useful instrument to assess visually the number of components, even if the multimodal structure cannot be seen in the data cloud itself.
- Mode trees together with a suitable bandwidth selector give a useful *recommendation* for the choice of K. However, this is no reliable automatic routine, as small variations in the bandwidth may change drastically the number of detected modes.
- Given K, mode trees give a set of estimated mass points, which are in many cases so accurate that one hardly needs EM at all!
- When using mode trees together with the proposed damping procedure, the sensitivity to tuning parameters (in particular `tol`) is drastically reduced.

Everything more general....

- Replace Gaussian by another exponential family distibution
- Set an appropriate link function
- Include explanatory variables
- Random coefficient models
- Variance component models

..... is being implemented in an R package {npml} (Einbeck, Darnell, & Hinde), see

`www.nuigalway.ie/maths/je/npml.html`

References

AITKIN, M. (1996): A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6**, 251–262.

HINDE, J. (1982): Compound Poisson regression models. *Lecture Notes in Statistics* **14** ,109-121.

CARREIRA-PERPIÑAN, M. A. and WILLIAMS, C.K.I. (2003): On the number of modes of a Gaussian mixture. *Lecture Notes in Computer Science* **2695**, 625–640.

LAIRD, N. M. (1978): Nonparametric maximum likelhood estimation of a mixing distribution. *JASA*, **73**, 805–811.

MINNOTTE, M. C. and SCOTT, D. W. (1993): The mode tree: A tool for visualization of nonparametric density features. *JCGS* **2**, 51-68.

RICHARDSON, S. and GREEN, P. (1997): On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion) *JRSSB* **59**, 731–792.

SILVERMAN. (1981): Using kernel density estimates to investigate multimodal regression. *JRSSB* **43**, 97–99.