

MATH 680: EXERCISE 4

THE EM ALGORITHM

The galaxies data

Task

1. See the script file

▸ <http://www.math.mcgill.ca/dstephens/680/R/Ex4-2015.R>

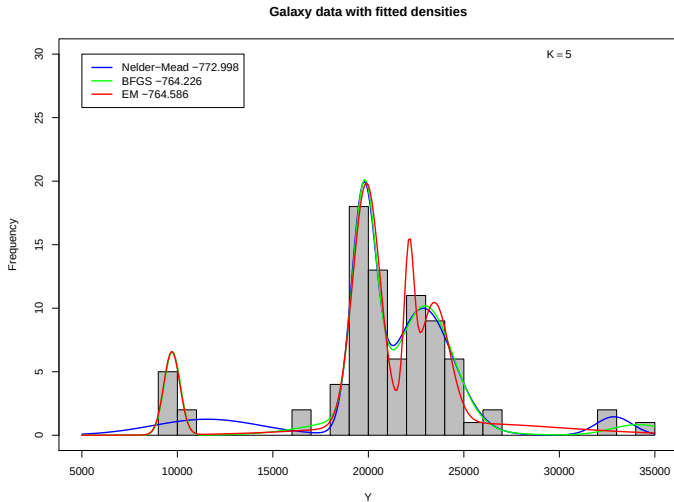
2. We have

K	BIC
3	1592.286
4	1602.333
5	1590.867

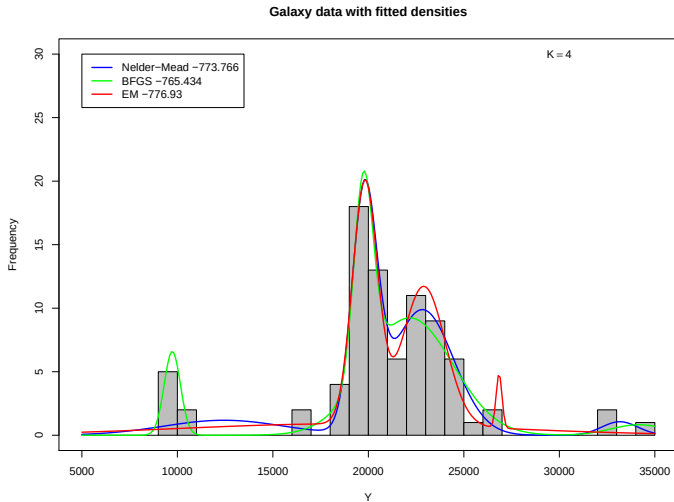
so $K = 5$ is preferred using EM.

There is some dependence on starting values, and direct optimization of the incomplete data likelihood using the BFGS Quasi-Newton algorithm can work well.

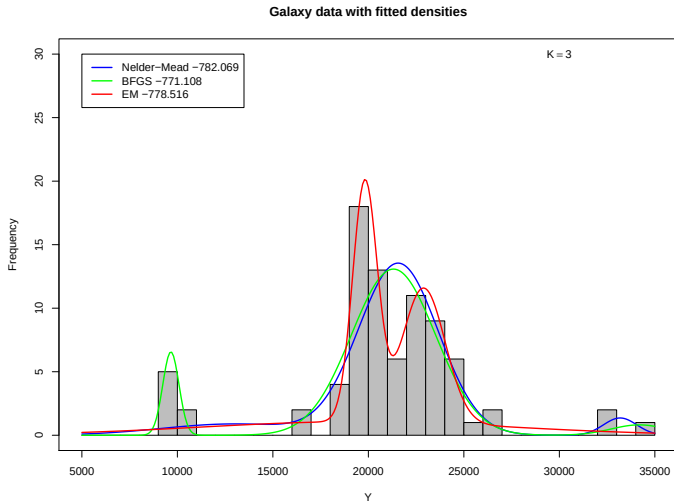
The galaxies data $K = 5$



The galaxies data $K = 4$



The galaxies data $K = 3$



1. Suppressing the dependence on (θ, σ) for the moment, we have

$$f_Z(z|\nu) = \int_0^\infty f_{Z|X}(z|x)f_X(x|\nu) \, dx$$

with

- ▶ $Z|X = x \sim \text{Normal}(0, 1/x)$,

$$f_{Z|X}(z|x) = \left(\frac{x}{2\pi}\right)^{1/2} \exp\left\{-\frac{x}{2}z^2\right\} \quad z \in \mathbb{R}, x > 0$$

- ▶ $X \sim \text{Gamma}(\nu/2, \nu/2)$

$$f_X(x|\nu) = \frac{(\nu/2)^{(\nu/2)}}{\Gamma\left(\frac{\nu}{2}\right)} x^{\nu/2-1} \exp\left\{-\frac{\nu}{2}x\right\} \quad x > 0.$$

Therefore a term in the complete data log likelihood is

$$\begin{aligned}\log f_{X,Z}(x, z; \nu) &= \log f_{Z|X}(z|x) + \log f_X(x; \nu) \\ &= \left\{ \frac{1}{2} \log x - \frac{z^2}{2}x \right\} + \left\{ \frac{\nu}{2} \log x - \frac{\nu}{2}x \right\} + \text{constant} \\ &= \frac{(\nu + 1)}{2} \log x - \frac{(\nu + z^2)}{2}x + \text{constant}\end{aligned}$$

The galaxies data: Task II (cont.)

Now we must compute the conditional density $f_{X|Z}(x|z, \nu)$; we have by Bayes Theorem

$$\begin{aligned} f_{X|Z}(x|z, \nu) &\propto f_{Z|X}(z|x) f_X(x) \\ &\propto \left\{ x^{1/2-1} \exp \left\{ -\frac{z^2}{2} x \right\} \right\} \left\{ x^{\nu/2-1} \exp \left\{ -\frac{\nu}{2} x \right\} \right\} \\ &= \left\{ x^{(\nu+1)/2-1} \exp \left\{ -\frac{(\nu + z^2)}{2} x \right\} \right\} \end{aligned}$$

so that

$$f_{X|Z}(x|z, \nu) \equiv \text{Gamma} \left(\frac{\nu + 1}{2}, \frac{\nu + z^2}{2} \right)$$

The galaxies data: Task II (cont.)

If $U \sim \text{Gamma}(\alpha, \beta)$, then $\mathbb{E}_U[U] = \alpha/\beta$, and

$$\begin{aligned}\mathbb{E}_U[\log U] &= \int_0^\infty \log u \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\beta u} du \\&= \int_0^\infty \log t \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-t} dt - \log \beta \\&= \frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{d}{d\alpha} \{t^{\alpha-1} e^{-t}\} dt - \log \beta \\&= \frac{1}{\Gamma(\alpha)} \frac{d}{d\alpha} \left\{ \int_0^\infty t^{\alpha-1} e^{-t} dt \right\} - \log \beta \\&= \frac{\dot{\Gamma}(\alpha)}{\Gamma(\alpha)} - \log \beta = \psi(\alpha) - \log \beta.\end{aligned}$$

where $\psi(\alpha) = \dot{\Gamma}(\alpha)/\Gamma(\alpha)$ is known as the *digamma* function.

The galaxies data: Task II (cont.)

We now consider the estimation of location parameter θ with $\sigma = 1$ for data $y_1, \dots, y_n$. Here we let

$$Z = Y - \theta.$$

The expected complete data log-likelihood at step t , required for the E-step, has terms

$$\mathbb{E}_{X|Z}[\log f_{X,Z}(X, Y; \theta, \nu) | y, \theta^{(t)}]$$

which equate to

$$\begin{aligned} \frac{(\nu + 1)}{2} \left[\psi((\nu + 1)/2) - \log \left(\frac{(\nu + (y - \theta^{(t)})^2)}{2} \right) \right] \\ - \frac{(\nu + (y - \theta)^2)}{2} \frac{(\nu + 1)}{\nu + (y - \theta^{(t)})^2}. \end{aligned}$$

The galaxies data: Task II (cont.)

Thus the expected complete data log-likelihood is, up to an additive constant which does not depend on θ ,

$$\begin{aligned}\mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\log f_{\mathbf{X},\mathbf{Y}}(\mathbf{X}, \mathbf{Y}; \theta, \nu) | \mathbf{y}, \theta^{(t)}] \\ = -\frac{(\nu + 1)}{2} \sum_{i=1}^n \frac{1}{\nu + (y_i - \theta^{(t)})^2} (y_i - \theta)^2.\end{aligned}$$

for which the maximizer for θ is available in closed form as

$$\theta^{(t+1)} = \frac{\sum_{i=1}^n \frac{y_i}{\nu + (y_i - \theta^{(t)})^2}}{\sum_{i=1}^n \frac{1}{\nu + (y_i - \theta^{(t)})^2}}$$

The galaxies data: Task II (cont.)

Allowing for the scale parameter σ to be estimated, we consider

$$Z = \frac{Y - \theta}{\sigma}$$

A term in the complete data log likelihood is

$$\begin{aligned}\log f_{X,Y}(x,y;\theta,\sigma,\nu) &= \log f_{Y|X}(y|x;\theta,\sigma) + \log f_X(x;\nu) \\ &= \left\{ \frac{1}{2} \log x - \frac{(y-\theta)^2}{2\sigma^2} x - \log \sigma \right\} \\ &\quad + \left\{ \frac{\nu}{2} \log x - \frac{\nu}{2} x \right\} + \text{constant} \\ &= \frac{(\nu+1)}{2} \log x - \frac{(\nu\sigma^2 + (y-\theta)^2)}{2\sigma^2} x - \log \sigma \\ &\quad + \text{constant}\end{aligned}$$

The galaxies data: Task II (cont.)

Letting

$$z_i^{(t)} = \frac{y_i - \theta^{(t)}}{\sigma^{(t)}}$$

the expected complete data log-likelihood at step t , required for the E-step, has terms

$$\mathbb{E}_{X|Z}[\log f_{X,Z}(X, Y; \theta, \sigma, \nu) | y, \theta^{(t)}, \sigma^{(t)}]$$

which equate to

$$\begin{aligned} \frac{(\nu + 1)}{2} \left[\psi((\nu + 1)/2) - \log \left(\frac{(\nu + z^{(t)2})}{2} \right) \right] \\ - \frac{(\nu \sigma^2 + (y - \theta)^2)}{2\sigma^2} \frac{(\nu + 1)}{\nu + z^{(t)2}} - \log \sigma. \end{aligned}$$

The galaxies data: Task II (cont.)

Thus the expected complete data log-likelihood is, up to an additive constant which does not depend on θ or σ ,

$$\begin{aligned}\mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\log f_{\mathbf{X},\mathbf{Y}}(\mathbf{X}, \mathbf{Y}; \theta, \sigma, \nu) | \mathbf{y}, \theta^{(t)}, \sigma^{(t)}] \\ = -n \log \sigma - \frac{(\nu + 1)}{2\sigma^2} \sum_{i=1}^n \frac{[\nu\sigma^2 + (y_i - \theta)^2]}{\nu + z_i^{(t)2}}\end{aligned}$$

for which as before the maximizer for θ is available in closed form as

$$\theta^{(t+1)} = \frac{\sum_{i=1}^n \frac{y_i}{\nu + z_i^{(t)2}}}{\sum_{i=1}^n \frac{1}{\nu + z_i^{(t)2}}}$$

For σ , we have

$$\sigma^{(t+1)} = \sqrt{\frac{\nu + 1}{n} \sum_{i=1}^n \frac{(y_i - \theta^{(t+1)})^2}{\nu + z_i^{(t)^2}}$$

The galaxies data: Task II (cont.)

2. If ν also needs to be estimated, then there is an extra term in the expected complete data log-likelihood that enters due to the normalizing constant in the Gamma model $f_X(x; \nu)$:

$$\begin{aligned}\mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\log f_{\mathbf{X},\mathbf{Y}}(\mathbf{X}, \mathbf{Y}; \theta, \sigma, \nu) | \mathbf{y}, \theta^{(t)}, \sigma^{(t)}] \\ = -n \log \sigma - \frac{(\nu + 1)}{2\sigma^2} \sum_{i=1}^n \frac{[\nu\sigma^2 + (y_i - \theta)^2]}{\nu + z_i^{(t)^2}} + g(\nu; \mathbf{z}^{(t)})\end{aligned}$$

where

$$\begin{aligned}g(\nu; \mathbf{z}^{(t)}) = (n\nu/2) \log(\nu/2) - n \log \Gamma\left(\frac{\nu}{2}\right) \\ + \frac{(\nu + 1)}{2} \left[n\psi((\nu + 1)/2) - \sum_{i=1}^n \log \left(\frac{(\nu + z_i^{(t)^2})}{2} \right) \right]\end{aligned}$$

We can maximize as before for (θ, σ) as a function of ν yielding

$$\theta^{(t+1)} = \frac{\sum_{i=1}^n \frac{y_i}{\nu^{(t+1)} + z_i^{(t)^2}}}{\sum_{i=1}^n \frac{1}{\nu^{(t+1)} + z_i^{(t)^2}}}$$

and

$$\sigma^{(t+1)} = \sqrt{\frac{\nu^{(t+1)} + 1}{n} \sum_{i=1}^n \frac{(y_i - \theta^{(t+1)})^2}{\nu^{(t+1)} + z_i^{(t)^2}}$$

but the maximization over ν of the profile likelihood must be achieved numerically.