LAB SESSION 2: MINI - PROJECT ON GAUSSIAN MIXTURES

Blanchet Alexis

Contents

1 Introduction **1**

2 EM algorithm **2**
2.1 1.Question . . . 2
2.2 2.Question . . . 2
2.3 3.Question . . . 3
2.4 4.Question . . . 5

3 Variational Bayes **5**
3.1 1.Question . . . 5
3.2 2.Question . . . 6
3.3 3.Question . . . 6
3.4 4.Question . . . 6
3.5 5.Question . . . 7
3.6 6.Question . . . 9
3.7 1.Question . . . 11
3.8 2.Question . . . 11
3.9 3.Question . . . 12
3.10 4.Question . . . 12
3.11 5.Question . . . 12
3.12 7.Question . . . 13

4 Predictive distributions versus maximum likelihood distribution **14**
4.1 1.Question . . . 14
4.2 2.Question . . . 14
4.3 3.Question . . . 15
4.4 4.Question . . . 16

1 Introduction

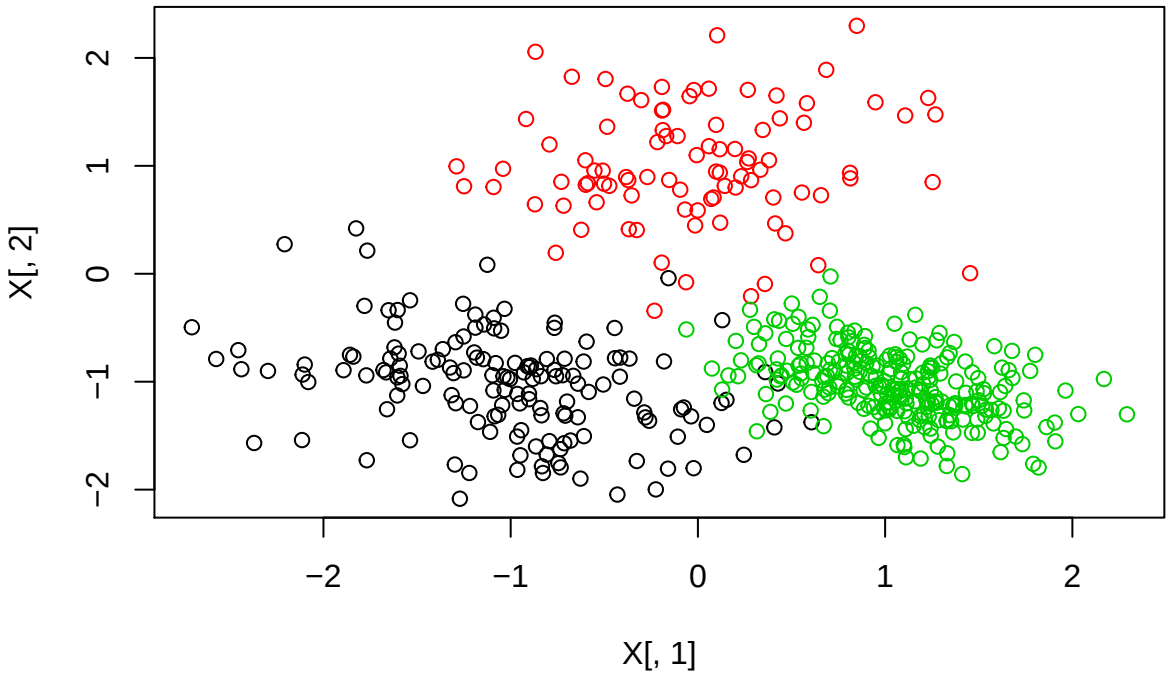
```
#' packages

#install.packages("MASS")
#install.packages("abind")
#install.packages("mnormt")
#install.packages("LaplacesDemon")
#install.packages("coda")

library(MASS)
library(abind)
library(mnormt)
library(LaplacesDemon)
library(coda)
```

```r
source("AlexisBlanchet_functions.R")
```

let's now see the dataset we will use during the entire lab session:



2 EM algorithm

2.1 1.Question

the coupled equations satisfied by the maximum likelihood estimates $(\rho, \mu_{1:k}, \sum_{1:K})$ are:

$$\mathbb{E}[X] = \rho.\mu_{1:k}$$

$$cov[x] = \sum_{k=1}^{K} \rho_k(\mu_k\mu_k^t + diag(\mu_{k,i}(1-\mu_{k,i})_i) - \mathbb{E}[X]\mathbb{E}[X]^t$$

2.2 2.Question

let's now derive from those the EM algorithm that will solve our Gaussian mixture problem. let's have a litlle reminder on the general EM algorithlm: E-Step:

$$\mathcal{L}_t(\theta) = \mathbb{E}[\log p(x, z|\theta)]_{p(z|x,\theta_t)}$$

M-Step:

$$\mathcal{L}_t(\theta) = \underset{\theta}{\text{argmax}}\ \mathcal{L}_t(\theta)$$

In the Gaussian Mixture model, the EM algorithm becomes :

E-Step:

$$\gamma(z_{nk}) = \frac{\pi_k . \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j . \mathcal{N}(x_n; \mu_j, \Sigma_j)}$$

M-Step:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

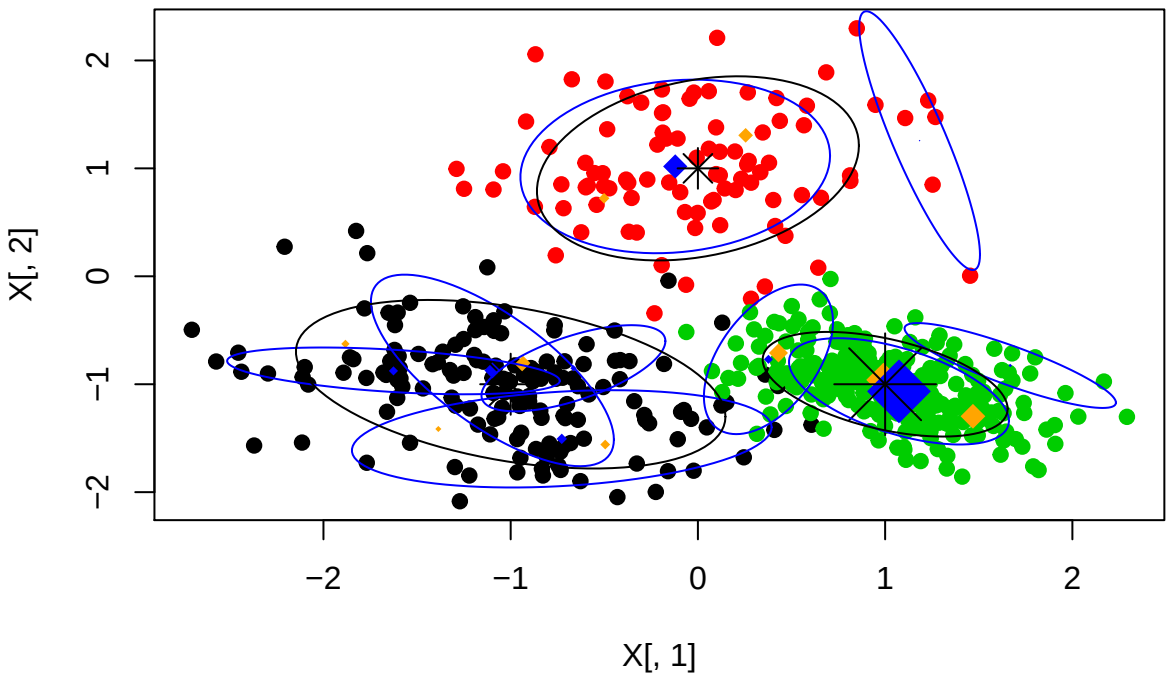
$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

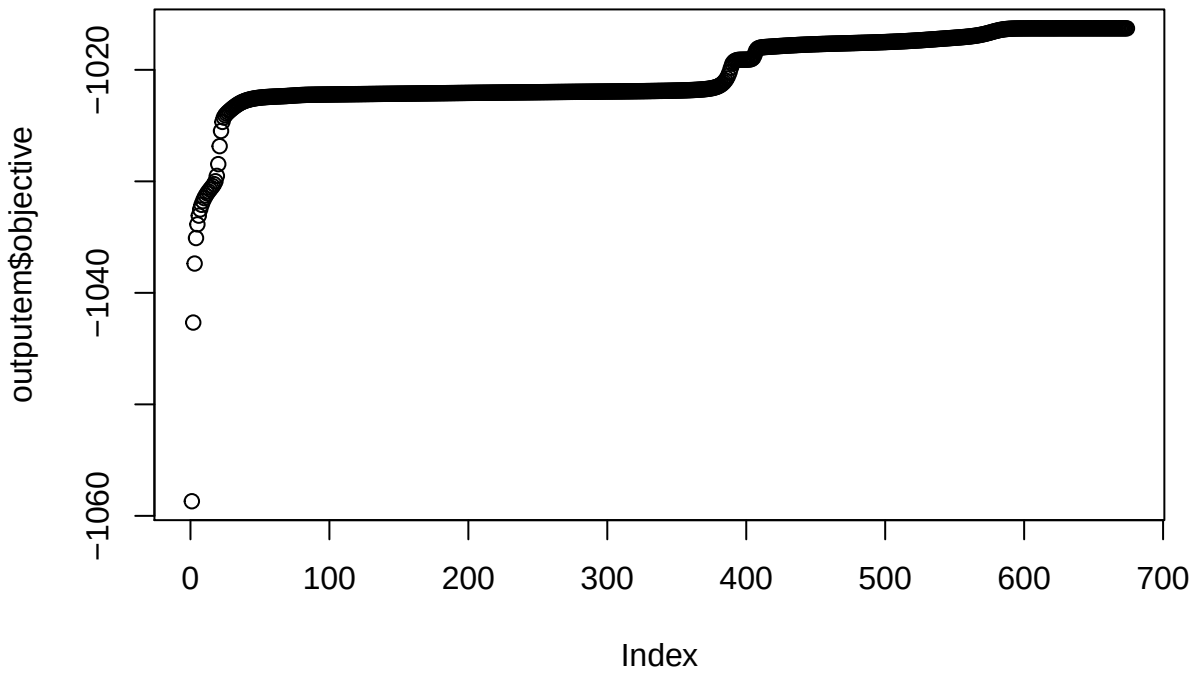
$$\pi_k^{new} = \frac{N_k}{N} \quad where \quad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

we can now implemente the estep,mstep,emalgo functions with those update equations

2.3 3.Question

let's now test this implementation. here we have Kfit = 9 but the real components number is 3 (black ellipses)


X[, 2]
2
1
0
−1
−2
−2
−1
0
1
2
X[, 1]



2.4 4.Question

we see on the plot above that the objective function to be maximized (the likelihood) increases at each iteration.

3 Variational Bayes

3.1 1.Question

let's give the coupled equations satisfied byh $\alpha^*, \nu^*, \beta^*, W^*$:

$$\alpha_k^* = \alpha_0^* + N_k$$

$$\nu_k^* = \nu_0^* + N_k$$

$$\beta_k^* = \beta_0^* + N_k$$

$$W_k^{*-1} = W_k^{*-1} + N_k.S_k + \frac{\beta_0.N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^T$$

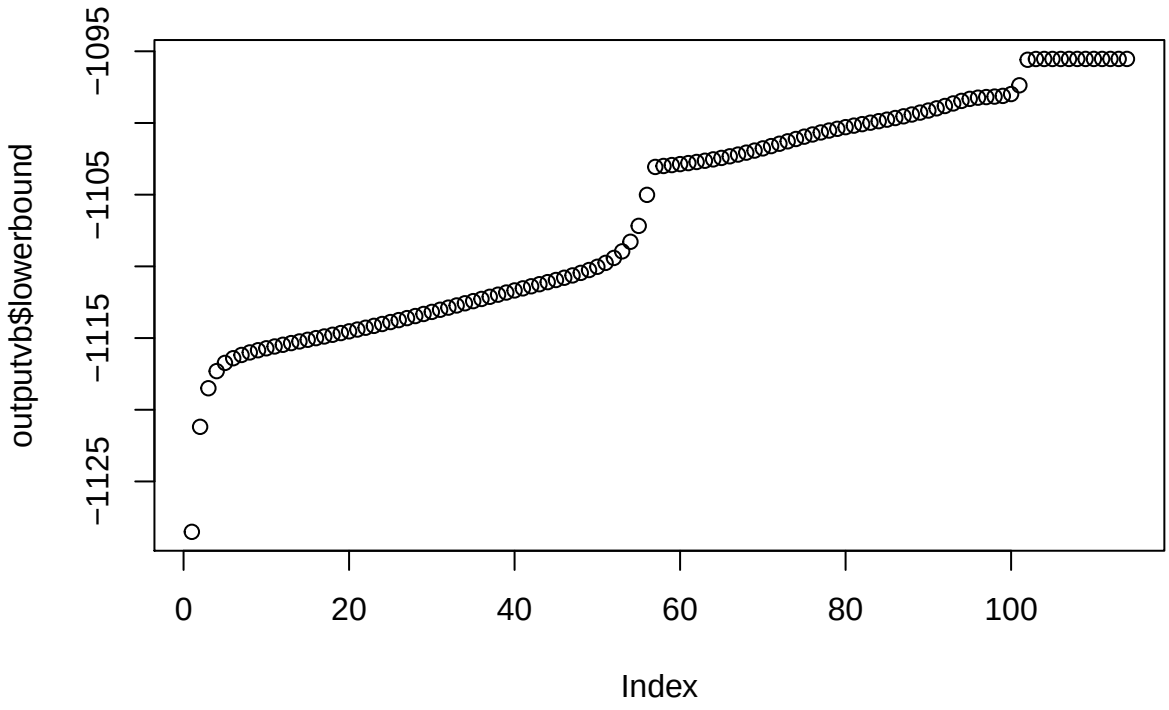
$$Where \;\; N_k = \sum_{n=1}^{N} r_{nk} \;\; and \;\; \mathbb{E}[Z_{nk}] = r_{nk}$$

3.2 2.Question

let's complete the code of functions vbMStep, vbEstep and vbalgo

3.3 3.Question

and now we test our implementation




```
## [1] 262.6737  90.6795   0.1000   0.1000 146.9468
```

```
##               [,1]          [,2]
## [1,]  1.027285e+00 -1.016042e+00
## [2,] -4.395892e-02  1.028034e+00
## [3,]  3.877317e-10 -4.759144e-10
## [4,]  3.877317e-10 -4.759144e-10
## [5,] -1.038370e+00 -1.039631e+00
```

3.4 4.Question

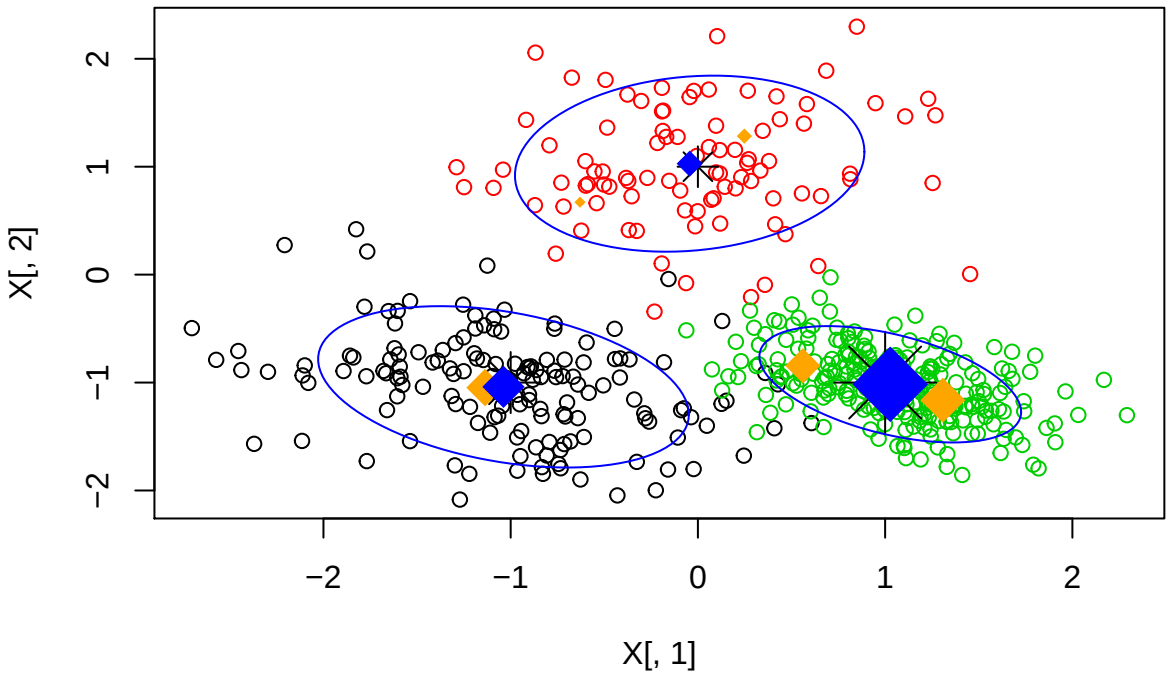
let's give the explicit expressions of the variable:

$$\hat{\rho}_{vb} = \mathbb{E}_{q^*}(\rho) = \mathbb{E}(Dir(\rho|\alpha)) = [\frac{\alpha_k}{\hat{\alpha}}]_k$$

$$\hat{\mu}_{j,vb} = \mathbb{E}_{q^*}(\mu_j) = \mathbb{E}(\mathcal{N}(m_j, ..) = m_j$$

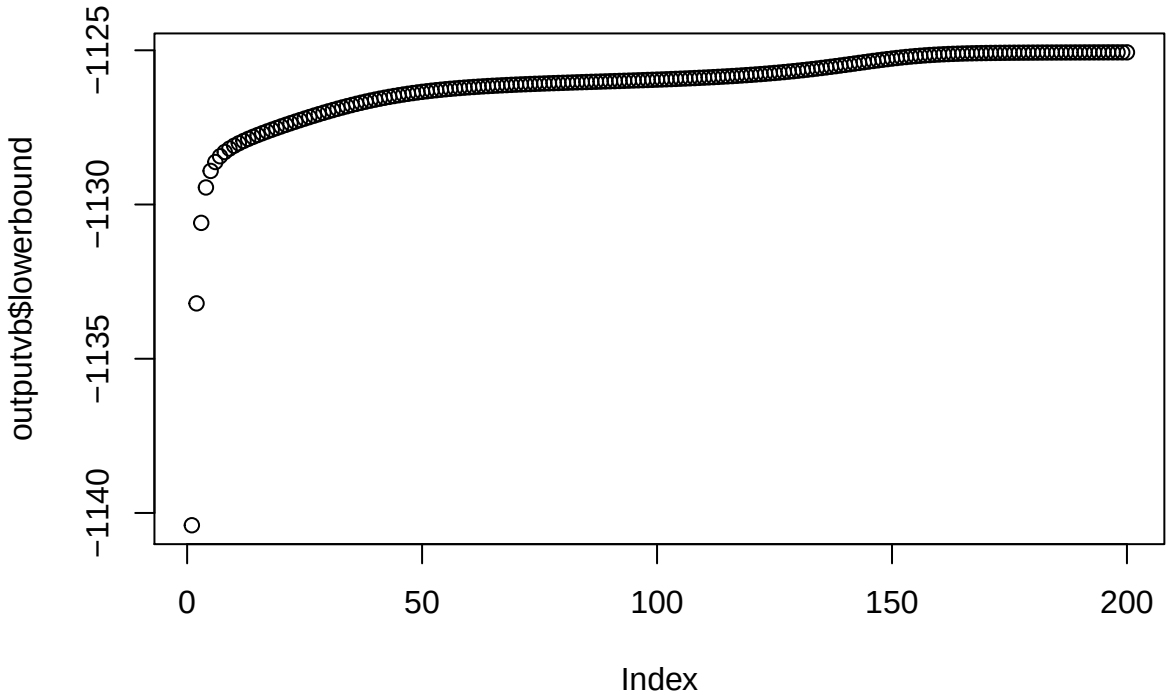
$$\hat{\Sigma}_{j,vb} = (\mathbb{E}_{q^*}(\Lambda_j))^{-1} = \mathbb{E}(\mathcal{W}(W_j, \nu_j))^{-1} = (\nu_j W_j)^{-1}$$

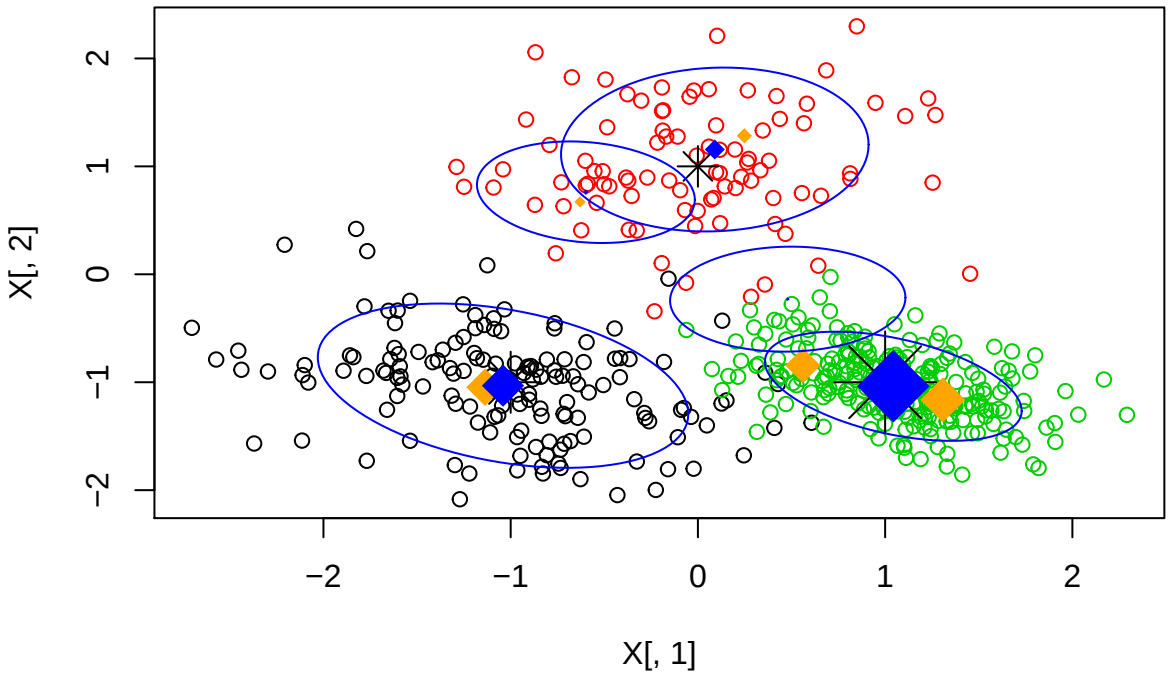
and we now Plot a summary of the corresponding Gaussian mixture



3.5 5.Question

let's now plot use the same code but with different value for α_0

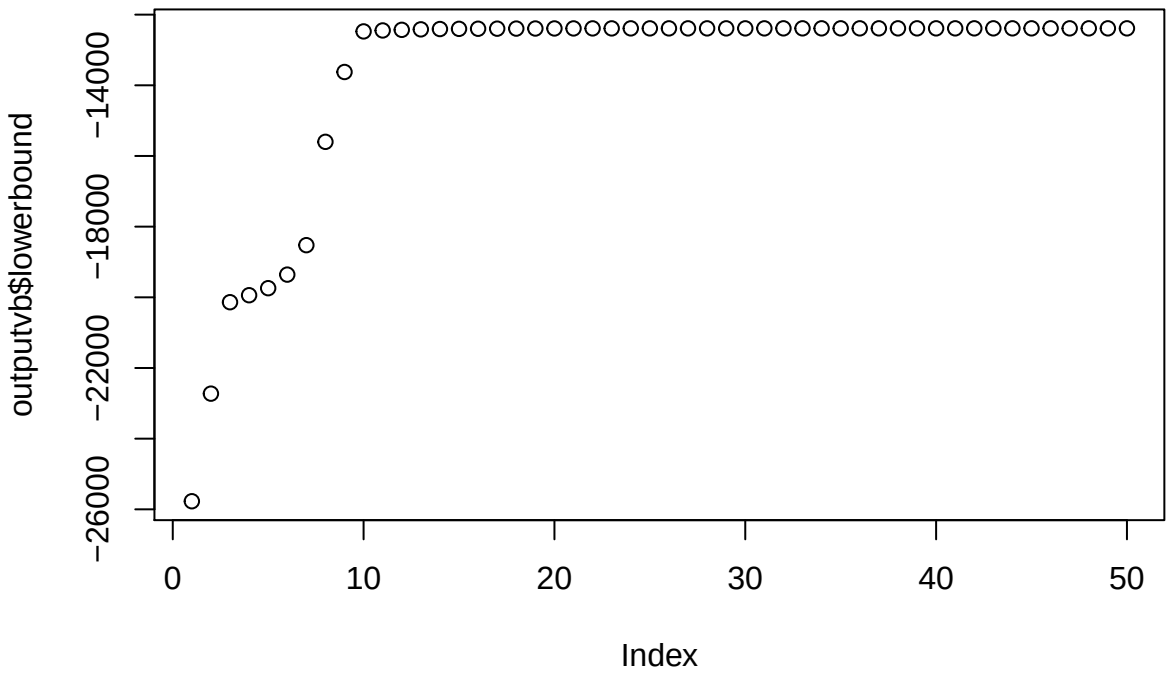

outputvb$lowerbound
−1125
−1130
−1135
−1140
0
50
100
150
200
Index

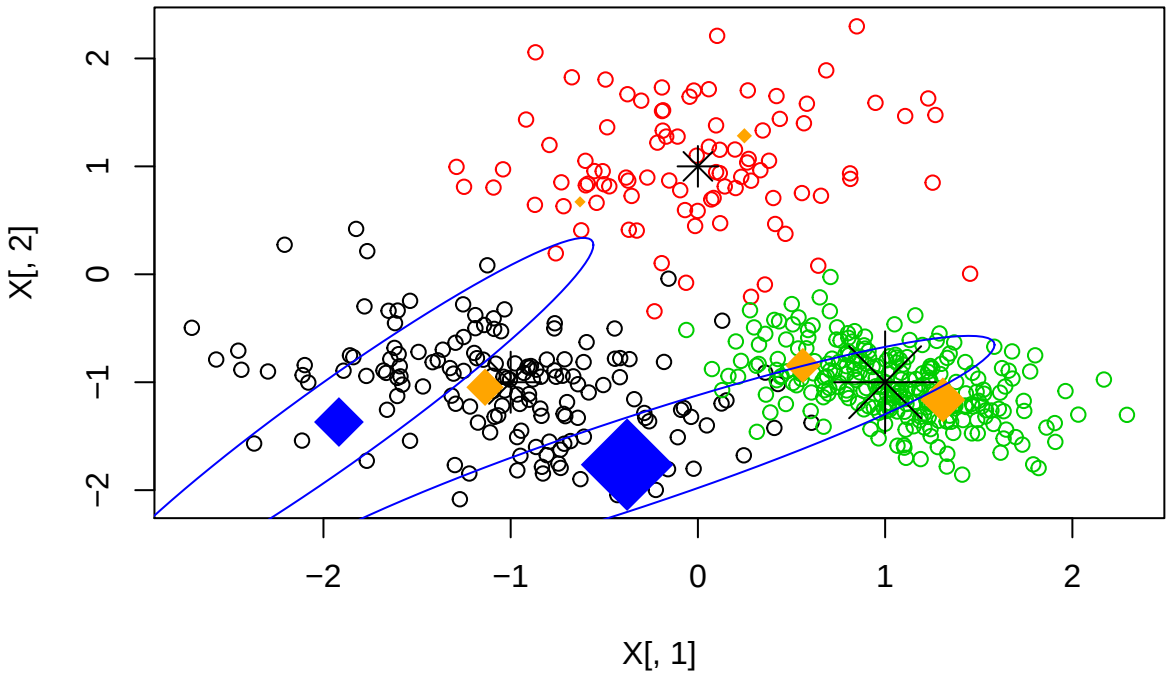


here we can see that for values under 1 (other values were tested but not all can be put here) the length of nonneg wich is diplaying the elipse is corresponding to the true number of components wich is 3 whereas for values over 1 (such as 2), the number of ellipses is not neceraly equal to 3.

3.6 6.Question

let's now test some other values for the other parameters.


outputvb$lowerbound
−26000
−22000
−18000
−14000
0
10
20
30
40
50
Index



m0: changing the inital value m0 is influencing the position of the centroides: shifiting it to the right(positive values) will make the two centroids go up on the right and shifting it to the left(negative values) will make the centroides go to the left.

beta0: this value is changing two things: first higher the value, quicker is the time of computation, higher the value, worse is the result: the ellipses are crossing each other and do not reprensent the actual data set(if we take beta0 equal to 1, 10 or 100 we get some funny results).

and for nu0 if we put it very low all the elipses will be shawn, if we put i high most won't

on the plot above i changed all parameters to show their importance. # Metropolis-Hastings algorithm

3.7 1.Question

We Complete the code for the function rproposal in file AlexisBlanchet_functions.R to generate such aproposal.

3.8 2.Question

We complete the code for the Metropolis-Hastings sampler MHsample. In order to do so we use the following formula from Bishop:

$$\alpha(\theta_t, \theta^*) \; with \; \alpha(s,t) = min(1, \frac{c.\widetilde{\pi}(t).q(t,s)}{c.\widetilde{\pi}(s).q(s,t)})$$

this allows us to get the acceptance ratio $\alpha(\theta_t, \theta^*)$

3.9 3.Question

We test our implemenatation of the algorithm.

```
##    user  system elapsed
##   30.09    0.00   30.09
```

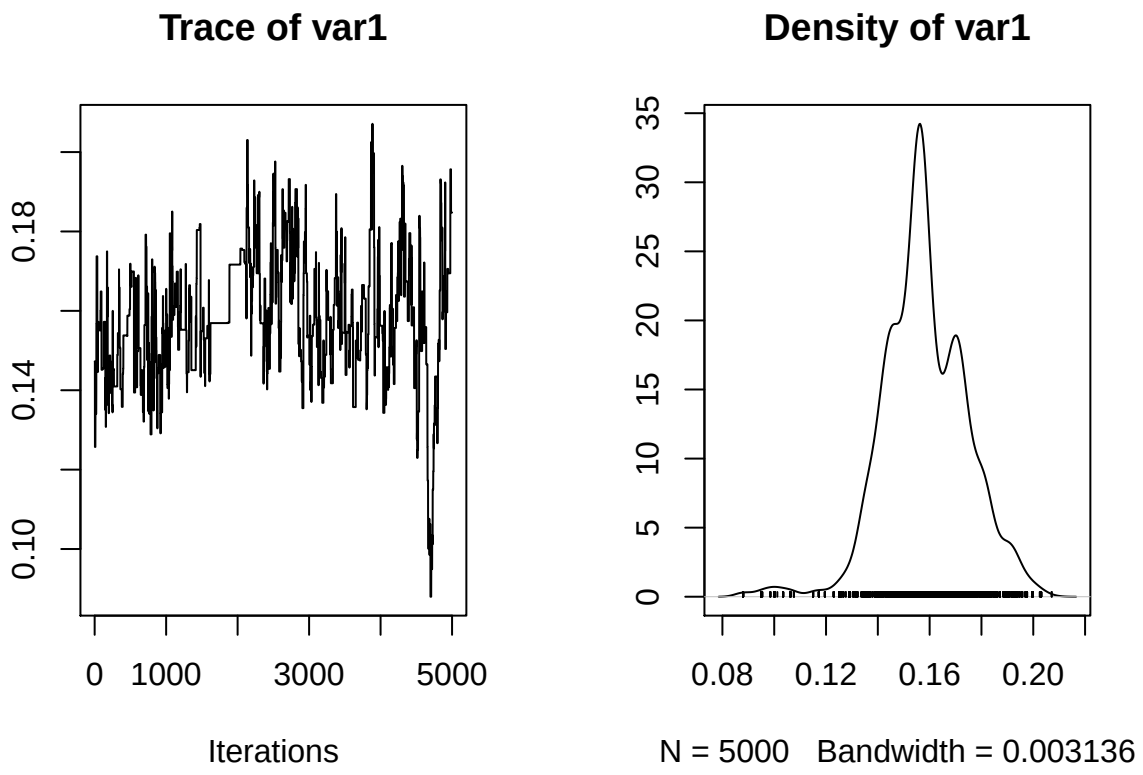
```
## [1] 577
```

3.10 4.Question

To obtain the time series, we complete the code of function cdfTrace from file AlexisBlanchet_functions.R.

3.11 5.Question

```
outputmh_1 <- mcmc(cdfTrace(x=c(-1,1), outputmh))
plot(outputmh_1)
```



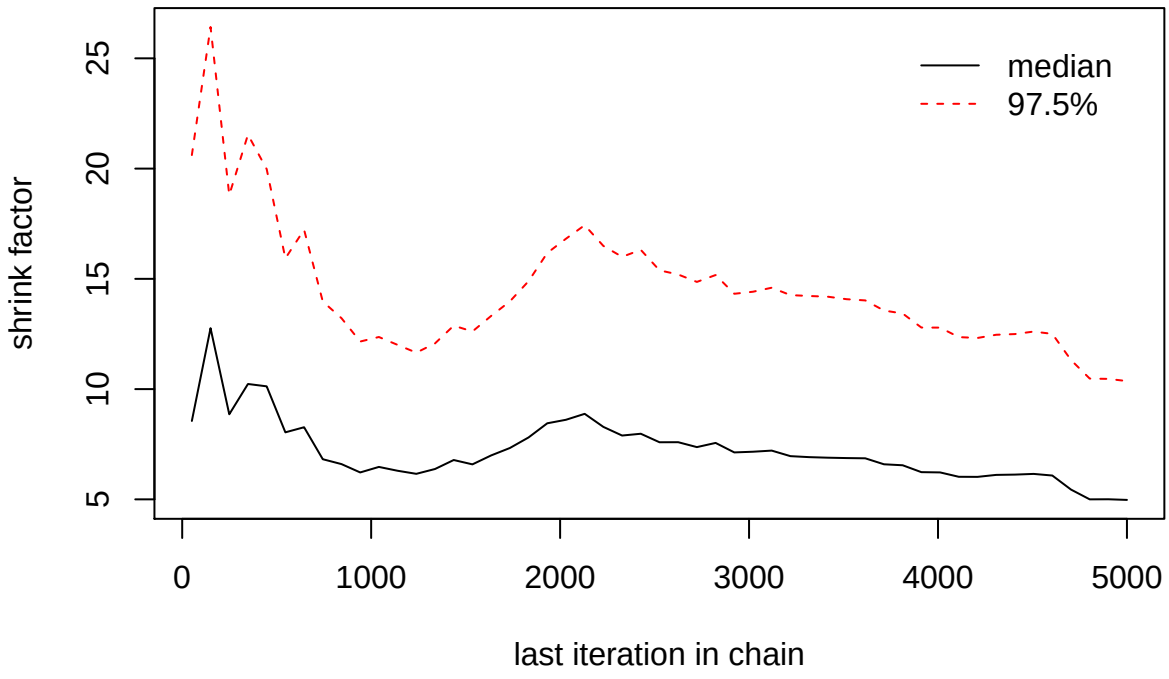

```
heidel.diag(outputmh_1)
```

```
##
##      Stationarity start     p-value
##      test         iteration
## var1 passed       1         0.361
##
```

```
##      Halfwidth Mean  Halfwidth
##      test
## var1 passed    0.158 0.00381
```

from the two results we can say that around 2500 iterations is a good we have already achieve convergence. To do so we burn the 2500 first results and we see if the diagnostique returns a stationary distrubution (wich will mean convergence is achieved) ##6.Question

```
## Potential scale reduction factors:
##
##      Point est. Upper C.I.
## [1,]       4.98       10.4
```

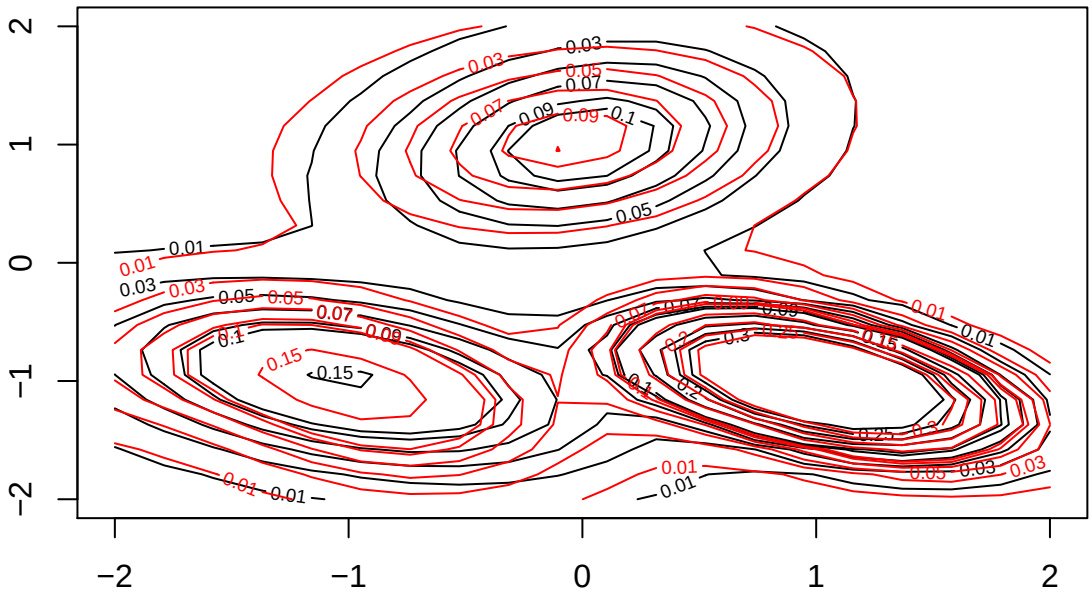


here again we can see that 2500 iterations allows us to achieve convergence.

3.12 7.Question

we complete the code of function MHpredictive, which returns $\hat{f}_{MH}(x)$.

then we plot the desired result together with the true density.(takes most of the time of computation)



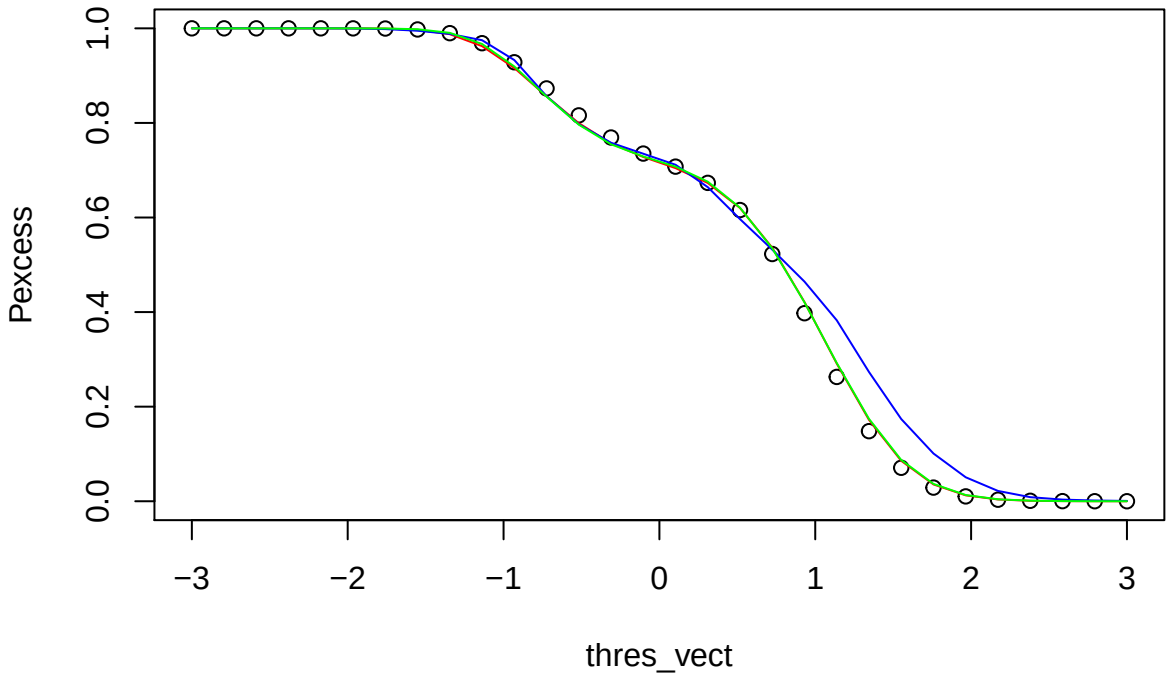
4 Predictive distributions versus maximum likelihood distribution

4.1 1.Question

We complete the code of the function MHpredictiveCdf

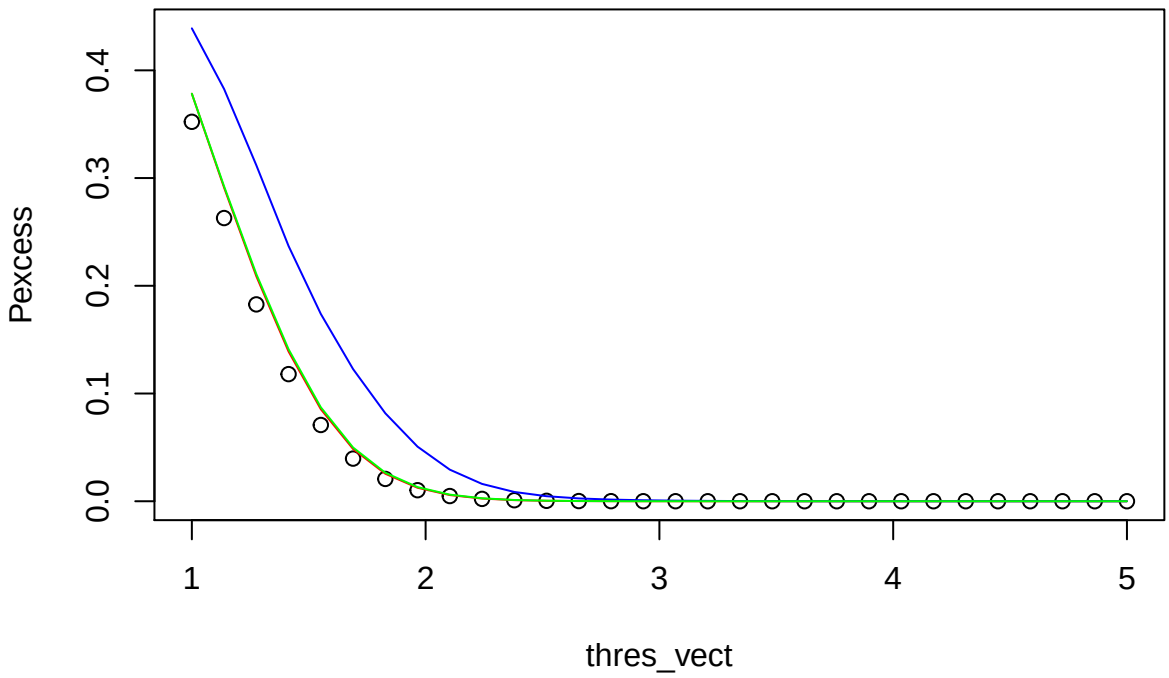
4.2 2.Question

we complete the following code chunk in order to plot on the same graph, as a function of x, ϕ, $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$



4.3 3.Question

we now consider now the tails of the mixture distribution: we replace the third line in the above code chunk



What we can see is that the Variationnal bayes is following the same line as the MH predictive cdf. Both are godd aproximation of the true cdf represented here par black circles.(the EM algorithm does not have a good Kfit and thus does not feat well to the curve). With other parameters we would observe something different. Also, q has a product form in case of variational inference, so it won't be able to approximate well when it is far from the center as most of the weigth will be at the center and thus attract the fitting of the VB algorithm.

4.4 4.Question