Rmd_Clustering_part1

Alessandro Colombi

20/5/2022

Inizio a fare dei test sul clustering.

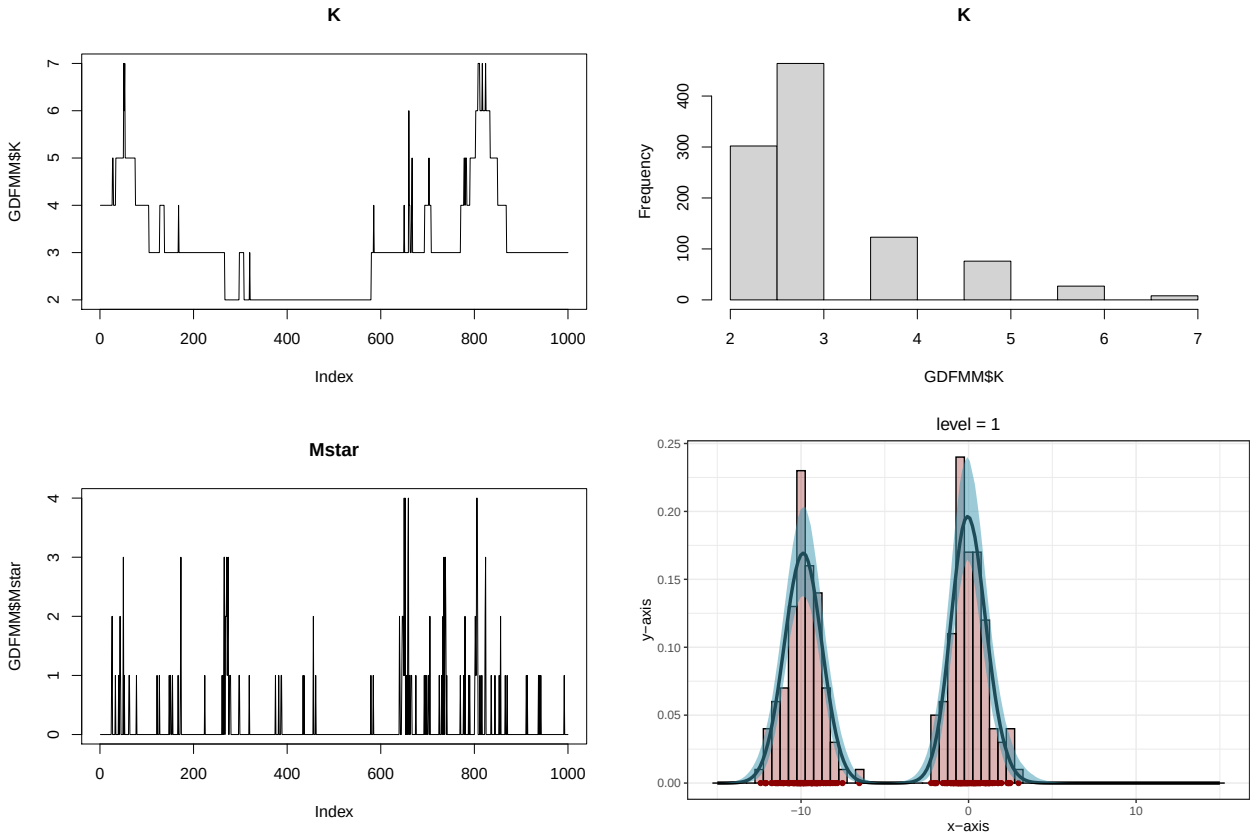
d=1

Primo caso, K = 2

Caso non simmetrico.

```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
suppressWarnings(suppressPackageStartupMessages(library(salso)))
suppressWarnings(suppressPackageStartupMessages(library(wesanderson)))
suppressWarnings(suppressPackageStartupMessages(library(mcclust.ext)))

# color palette ------------------------------------------------------------
mycol = hcl.colors(n=3,palette = "Zissou1")
# data generation ----------------------------------------------------------

d = 1                # number of groups
K = 2                # number of global clusters
mu = c(-10,0)    # vectors of means
sd = c(1,1)       # vector of sd
n_j = rep(200, d)  # set cardinality of the groups
seed = 1243
mycol_cluster = brewer.pal(n=K, name = "Dark2")
#> Warning in brewer.pal(n = K, name = "Dark2"): minimal value for n is 3, returning requested palette

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition

# Run  ---------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
```

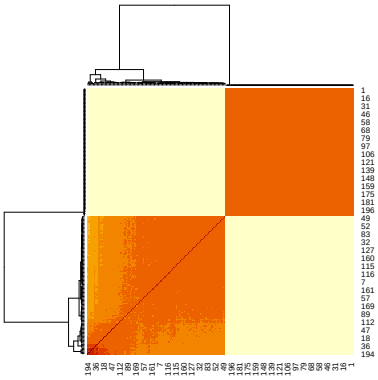
```
              "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
              "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
              "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```



Controllo la qualità del clustering

- `part_matrix`: This is a `(n_iter x n)` matrix, the j-th column contains all the labels assign to the j-th data point for all saved iterations.
- `sim_matrix`: This is a `(n x n)` matrix. Element (i,j) contains the probability that data points i and j are clustered together.
- `estimate_partition`: this is a vector of length `n` containing the estimated partition.

```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION ------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```

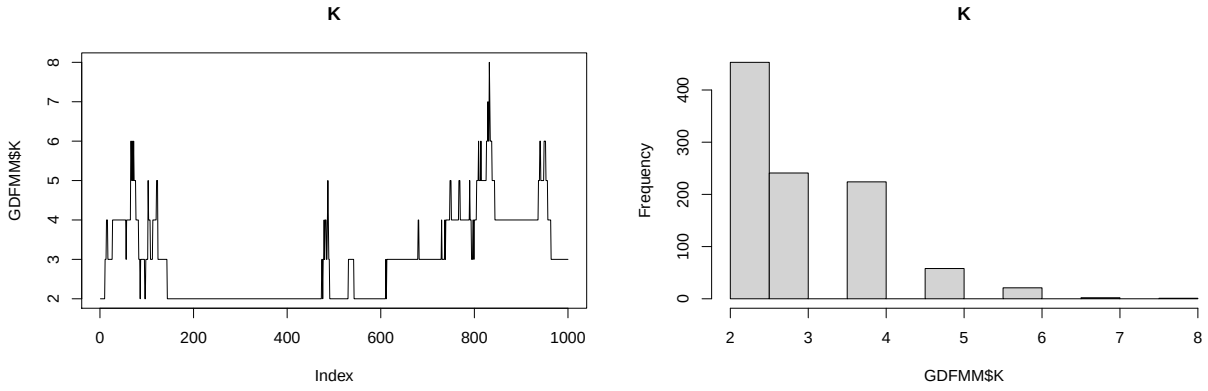


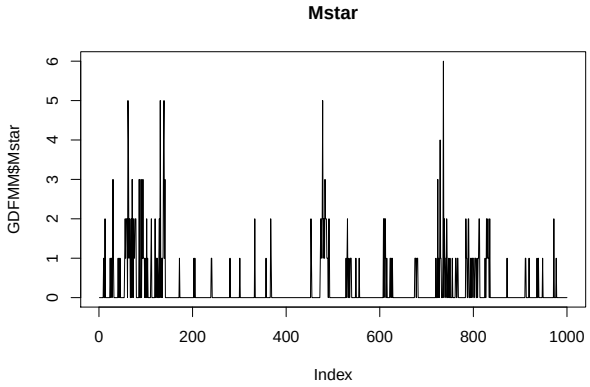
```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2
#>  96 104
table(VI_sara$cl)
#>
#>   1   2
#>  96 104
table(real_partition)
#> real_partition
#>   1   2
#>  96 104

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

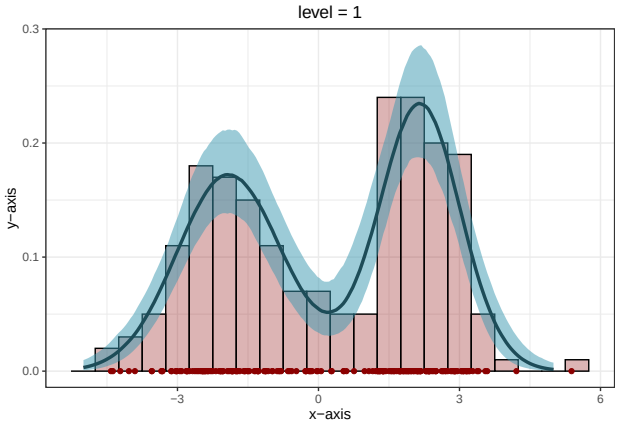
Caso simmetrico

```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
```

```r
# color palette ------------------------------------------------------------
mycol = hcl.colors(n=3,palette = "Zissou1")
# data generation ----------------------------------------------------------

d = 1                # number of groups
K = 2                # number of global clusters
mu = c(-2,2)    # vectors of means
sd = c(1,1)       # vector of sd
n_j = rep(200, d)  # set cardinality of the groups
seed = 32156
mycol_cluster = brewer.pal(n=K, name = "Dark2")
#> Warning in brewer.pal(n = K, name = "Dark2"): minimal value for n is 3, returning requested palette

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition
# Run  ---------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```





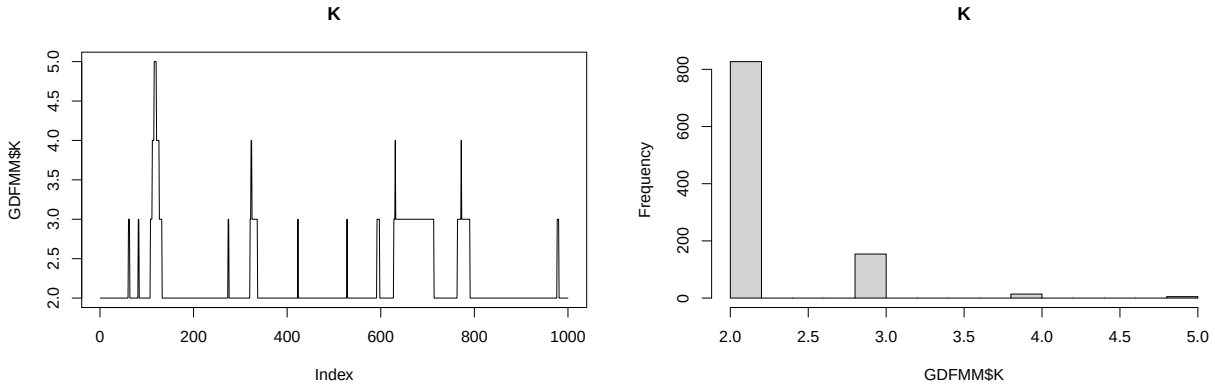


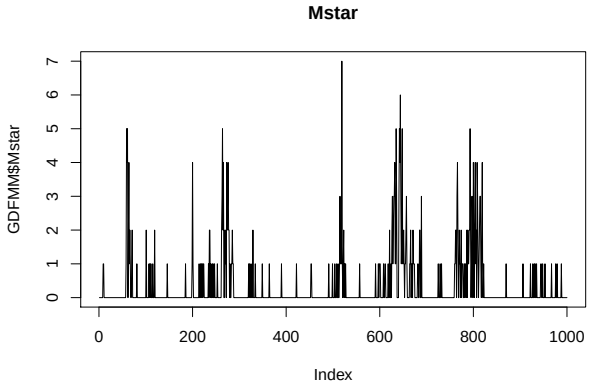

```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION --------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
estimate_partition = as.vector(binder_dahl)

# Get quality indicies
Kest = length(unique(estimate_partition))
cat('\n Estimated number of clusters \n')
#>
#>  Estimated number of clusters
Kest
#> [1] 4
Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
cat('\n Binder loss function \n')
#>
#>  Binder loss function
Binder_loss
#> [1] 0.10075
Tab = table(real_partition,estimate_partition)
cat('\n Miss classification table \n')
#>
#>  Miss classification table
Tab
#>                estimate_partition
#> real_partition  0  1  2  3
#>              1 95  0  0  0
#>              2  4 87 13  1
```

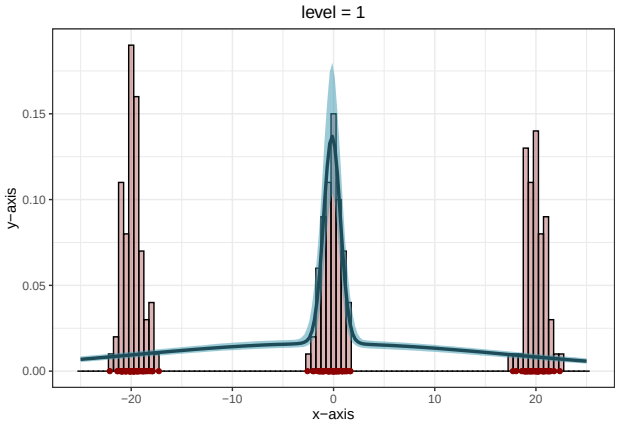
Secondo caso, K = 3

Caso simmetrico.

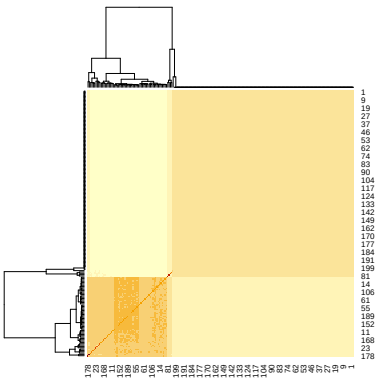
```r
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
# data generation ------------------------------------------------------------

d = 1                # number of groups
K = 3                # number of global clusters
mu = c(-20,0, 20)    # vectors of means
sd = c(1,1,1)       # vector of sd
n_j = rep(200, d)   # set cardinality of the groups
seed = 124123
mycol_cluster = brewer.pal(n=K, name = "Dark2")

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition

# Run  -----------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```








```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION --------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```



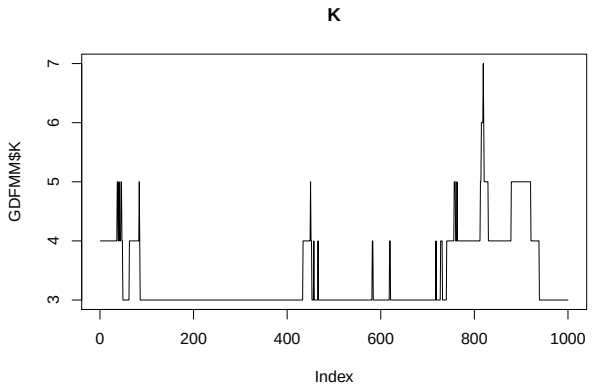
```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2
#> 140  60
table(VI_sara$cl)
#>
#>   1   2
#> 140  60
table(real_partition)
#> real_partition
```

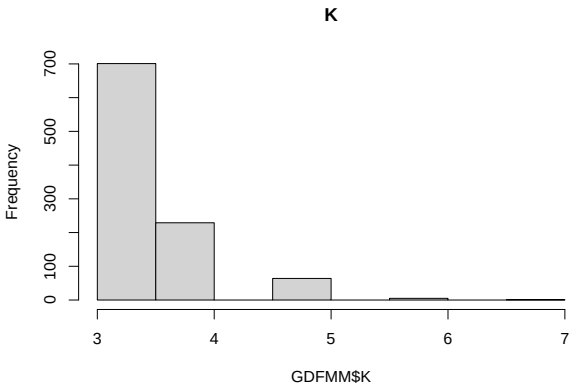
```
#>  1  2  3
#> 72 65 63

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

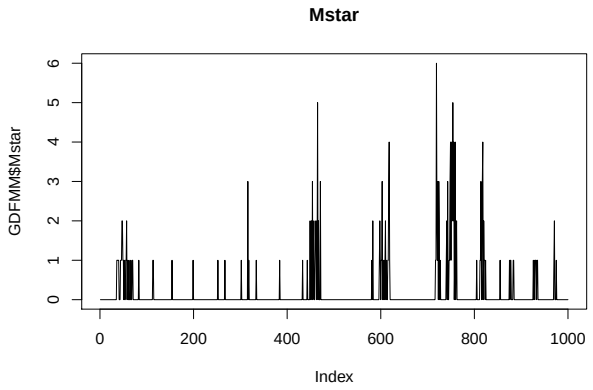
Caso NON simmetrico.

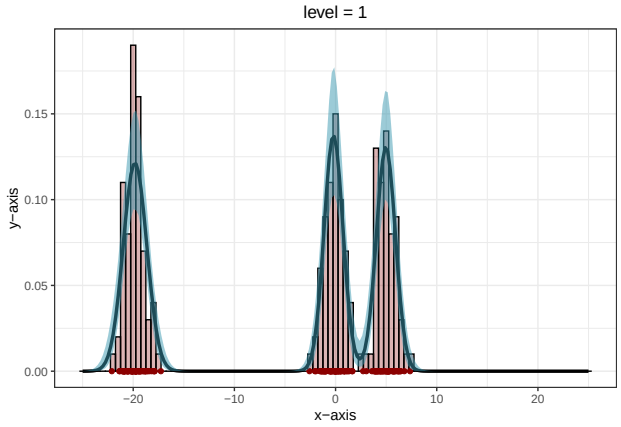
```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
# data generation ------------------------------------------------------------

d = 1               # number of groups
K = 3               # number of global clusters
mu = c(-20,0, 5)   # vectors of means
sd = c(1,1,1)      # vector of sd
n_j = rep(200, d)  # set cardinality of the groups
seed = 124123
mycol_cluster = brewer.pal(n=K, name = "Dark2")

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition

# Run  -----------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)
```

```
#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```

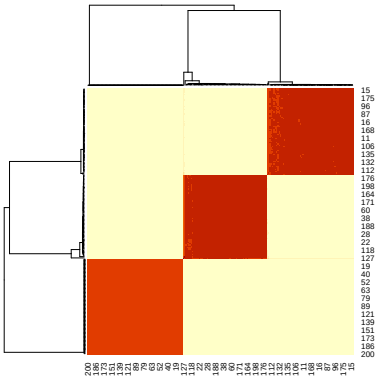









```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION ------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```



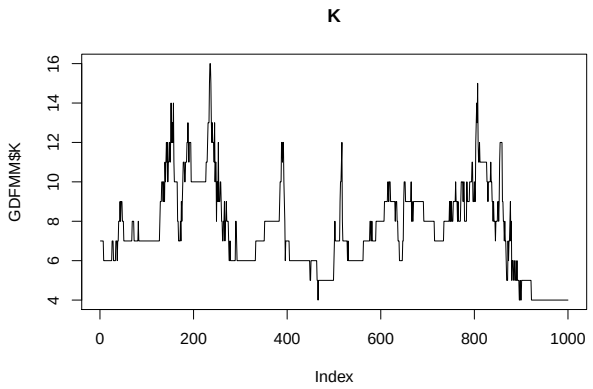
```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>  1  2  3
#> 63 72 65
table(VI_sara$cl)
#>
#>  1  2  3
#> 63 72 65
table(real_partition)
#> real_partition
#>  1  2  3
#> 72 65 63

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

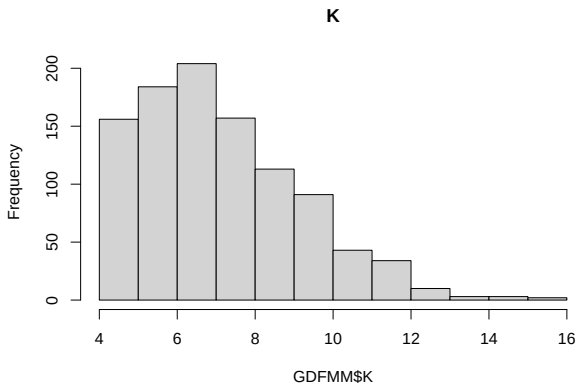
Caso d = 2

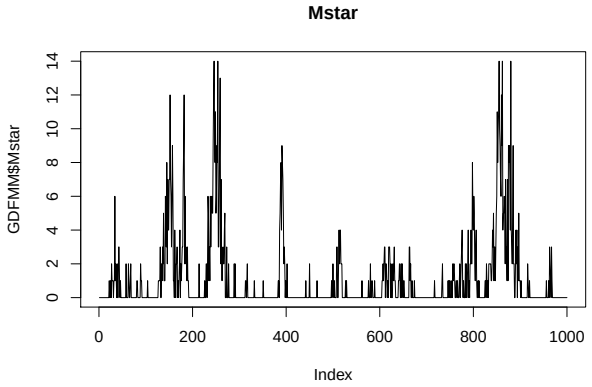
Fisso K = 4 Caso NON simmetrico.

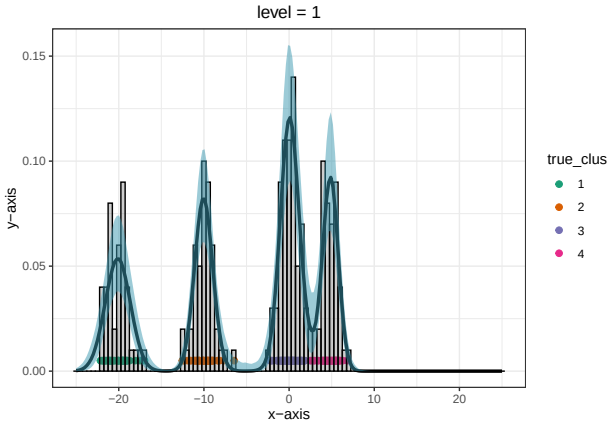
```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
```

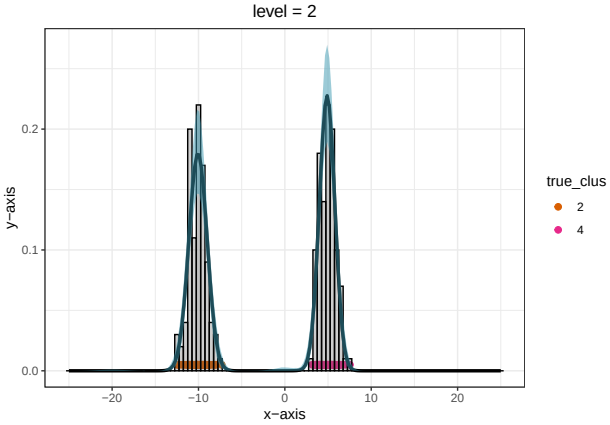
```r
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))

# color palette ------------------------------------------------------------
mycol = hcl.colors(n=3,palette = "Zissou1")
# data generation ----------------------------------------------------------

d = 2                # number of groups
K = 4                # number of global clusters
mu = c(-20,-10,0, 5)    # vectors of means
sd = c(1,1,1,1)       # vector of sd
n_j = rep(200, d)  # set cardinality of the groups
seed = 20051131
mycol_cluster = brewer.pal(n=K, name = "Dark2")

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition

# Run  ---------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```



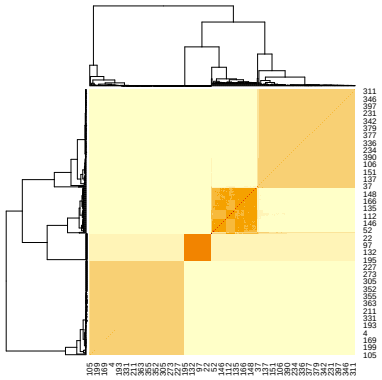









```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION -------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```



```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2   3   4   5   6   7   8
#> 142  41  52 147   1  15   1   1
table(VI_sara$cl)
#>
#>   1   2   3   4
#> 142  41  69 148
table(real_partition)
#> real_partition
#>   1   2   3   4
#>  41 142  68 149

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

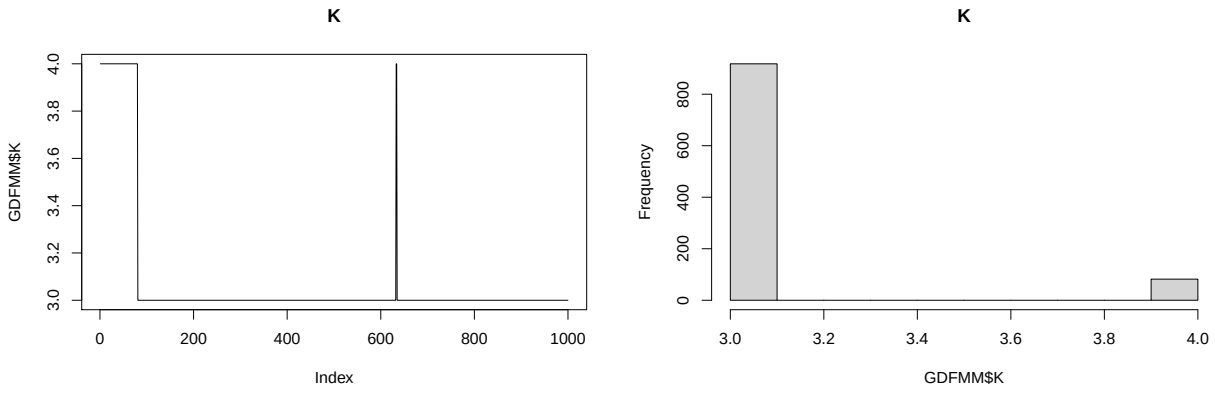
d=3

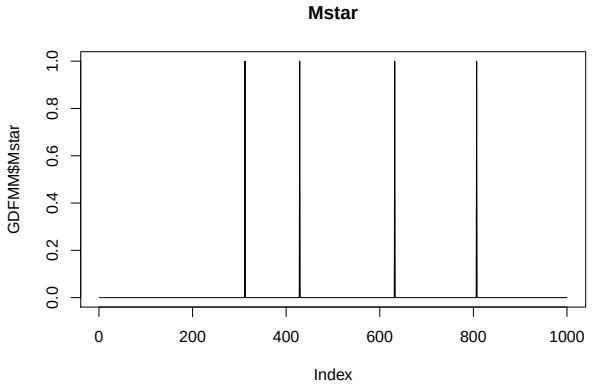
K = 4

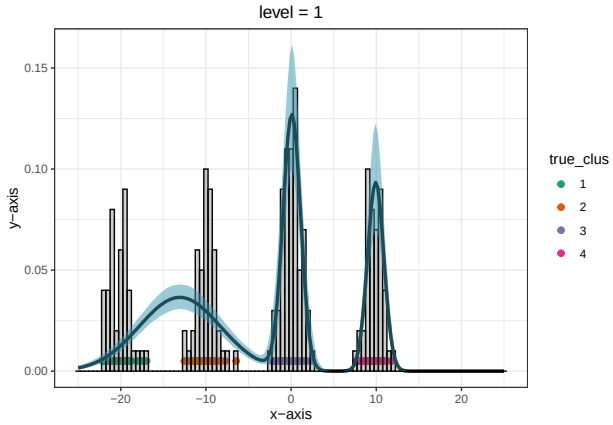
Primo caso, default come gli altri.

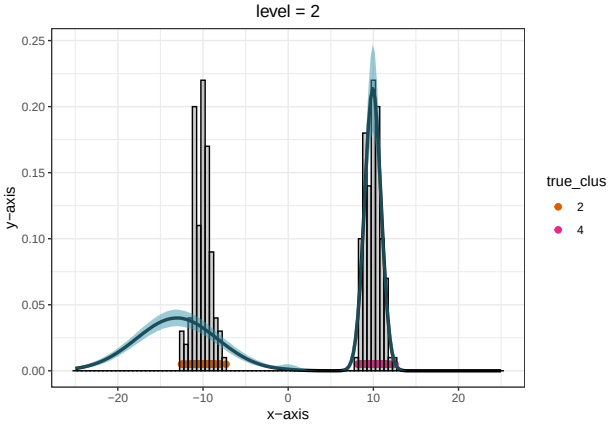
```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
suppressWarnings(suppressPackageStartupMessages(library(salso)))
suppressWarnings(suppressPackageStartupMessages(library(wesanderson)))



# color palette ----------------------------------------------------------
mycol = hcl.colors(n=3,palette = "Zissou1")

# data generation --------------------------------------------------------

d = 3               # number of groups
```

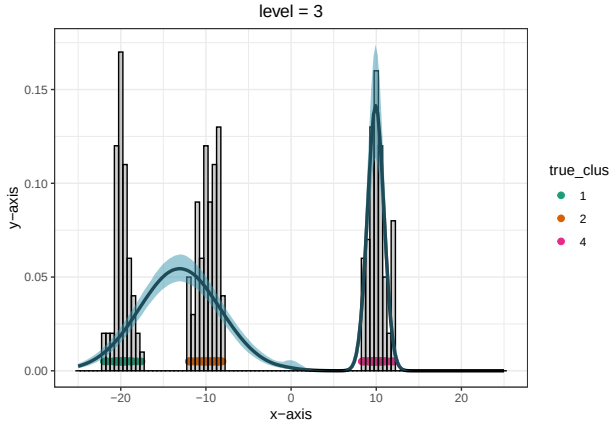
```r
K = 4                # number of global clusters
mu = c(-20,-10,0, 10)   # vectors of means
sd = c(1,1,1,1)       # vector of sd
n_j = rep(200, d)  # set cardinality of the groups
seed = 20051131

genD = generate_data(d=d, K=K, mu = mu, sd = sd, n_j = n_j, seed = seed)
data = genD$data
real_partition = genD$real_partition


mycol_cluster = brewer.pal(n=K, name = "Dark2")
# Run  ---------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 2, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```



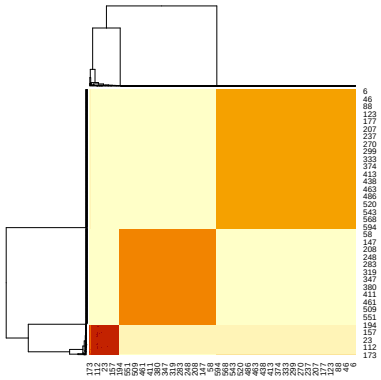








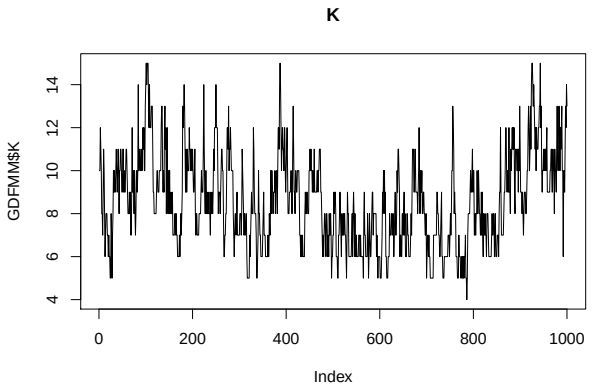

```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION ------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```

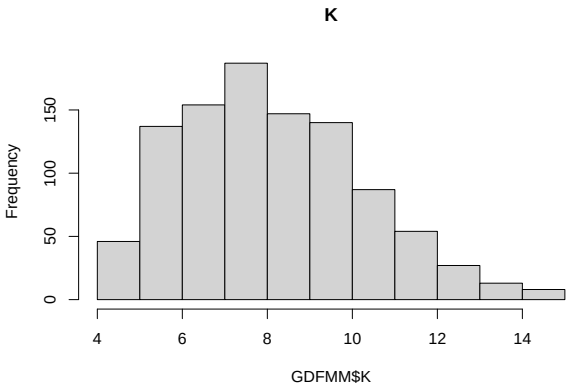


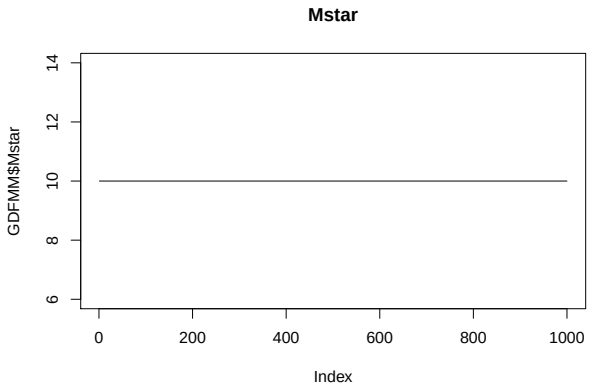
```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2   3
#> 314  68 218
table(VI_sara$cl)
#>
#>   1   2   3
#> 314  68 218
table(real_partition)
#> real_partition
#>   1   2   3   4
#> 100 214  68 218

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

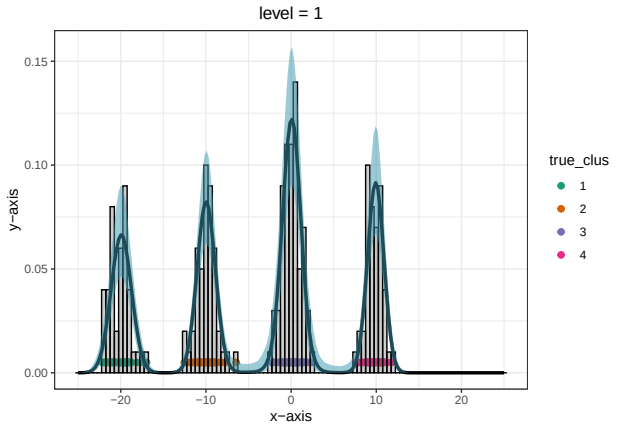
Secondo caso, fisso `Mstar = 10`. Mixxa molto meglio.

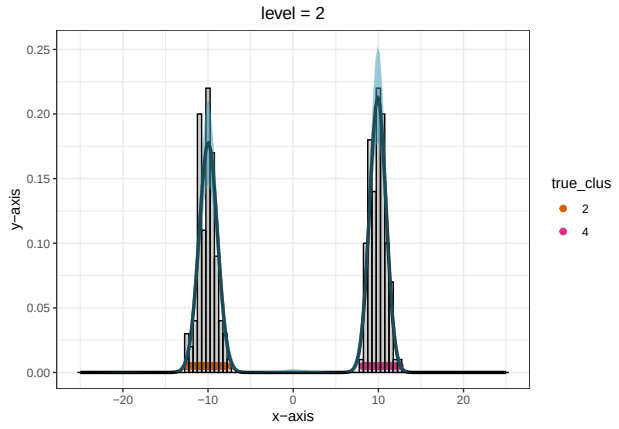
```
# Run  ----------------------------------------------------------------------


niter  <- 1000
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 10, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
              "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
              "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
              "beta_gamma"=1, "alpha_lambda"=1, "beta_lambda"=1,
              "UpdateU" = T, "UpdateM" = F, "UpdateGamma" = T, "UpdateS" = T,
              "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```

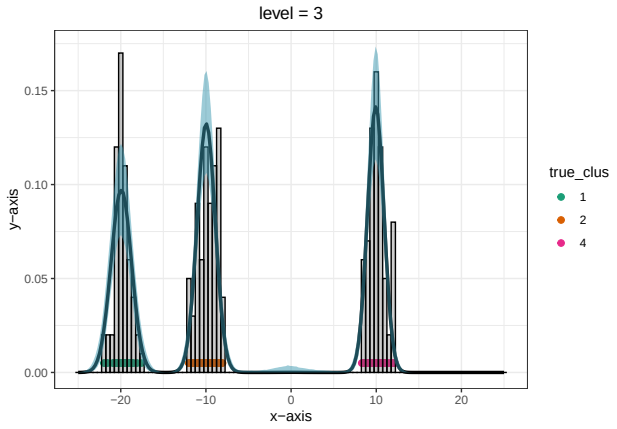




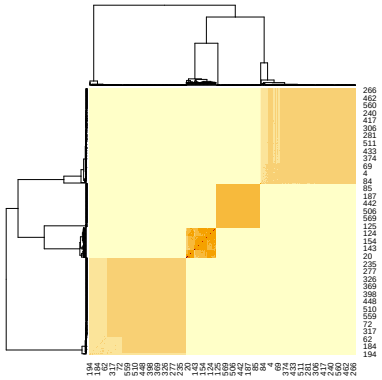









```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION --------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```



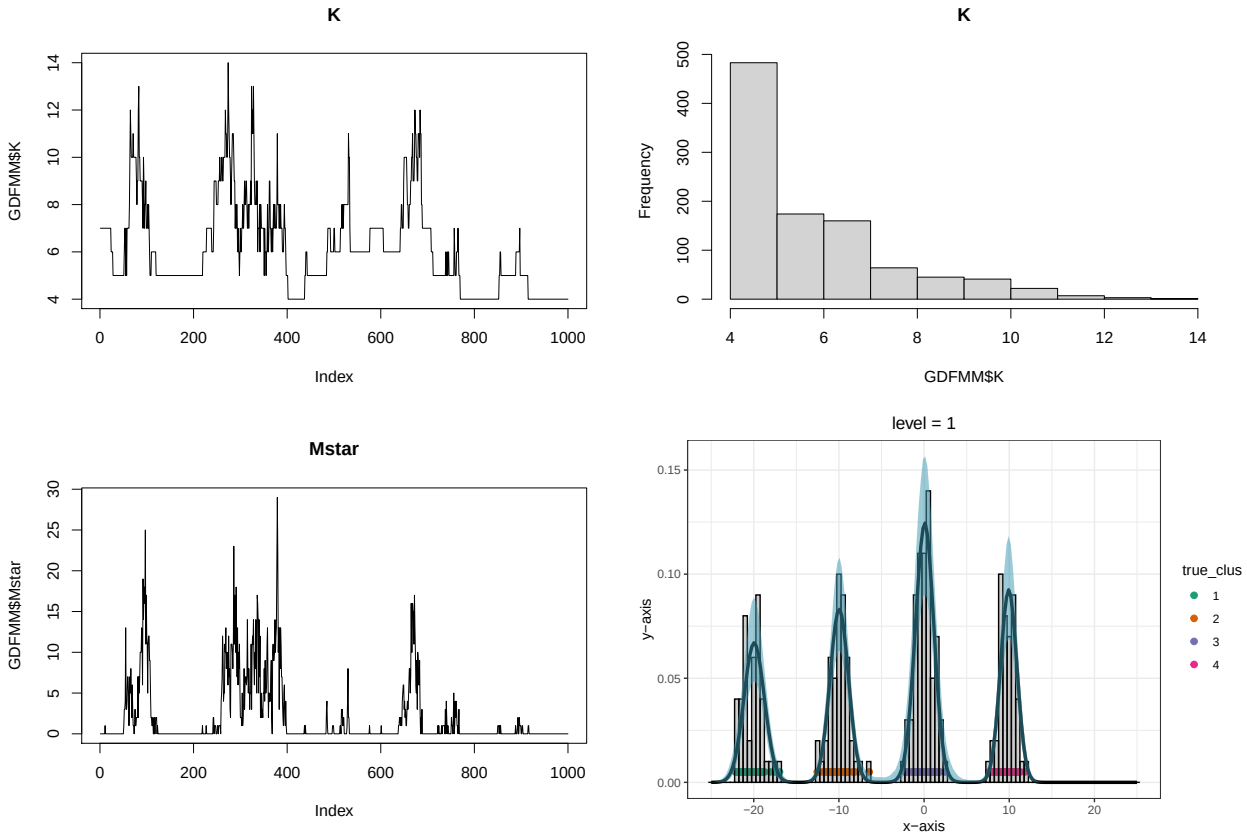
```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2   3   4   5   6   7   8   9
#> 214 100  30  18 218   4  14   1   1
table(VI_sara$cl)
#>
#>   1   2   3   4
#> 214 100  68 218
table(real_partition)
#> real_partition
#>   1   2   3   4
#> 100 214  68 218

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```

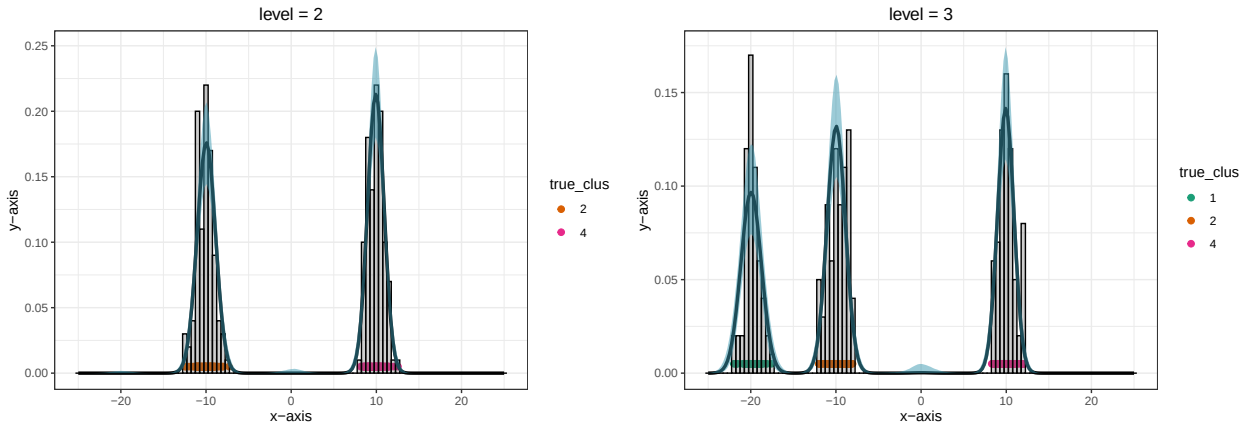
Terzo caso, `Mstar` random ma aumento prior su numbero di componenti.

Visto che $\Lambda \sim Gamma(a_\Lambda, b_\Lambda)$, metto $a_\Lambda = 15$.

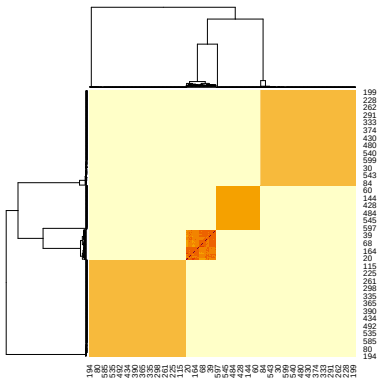
```
# Run  ----------------------------------------------------------------------


niter  <- 1000
```

```r
burnin <- 1000
thin   <- 1

option<-set_options("nu" = 1, "Mstar0" = 10, "Lambda0" = 3, "mu0" = 0,"sigma0"= 1, "gamma0" = 1,
             "Adapt_MH_hyp1"= 0.7,"Adapt_MH_hyp2"= 0.234, "Adapt_MH_power_lim"=10, "Adapt_MH_var0"=1,
             "k0"= 1/10, "nu0"=10, "alpha_gamma"=1,
             "beta_gamma"=1, "alpha_lambda"=15, "beta_lambda"=1,
             "UpdateU" = T, "UpdateM" = T, "UpdateGamma" = T, "UpdateS" = T,
             "UpdateTau" = T, "UpdateLambda" = T
)

#GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, option = option)
GDFMM = GDFMM_sampler(data, niter, burnin, thin, seed = 123, FixPartition = F, option = option)
```






```
# COMPUTE BINDER LOSS FUNCTION TO SELECT BEST PARTITION --------------------

# Get labels for each iterations for each data point
part_matrix <- GDFMM$Partition

# Compute similarity matrix
sim_matrix <- psm(part_matrix)


heatmap(sim_matrix)
```



```
binder_sara = minbinder(sim_matrix)
VI_sara = minVI(sim_matrix)

table(binder_sara$cl)
#>
#>   1   2   3   4
#> 214 100  68 218
table(VI_sara$cl)
#>
#>   1   2   3   4
#> 214 100  68 218
table(real_partition)
#> real_partition
```

```
#>   1   2   3   4
#> 100 214  68 218

# Get estimated partition according to binder (or VI) loss functions.
# VI_dahl <- dlso(matr, loss = 'VI', estimate=NULL)
# binder_dahl <- dlso(part_matrix, loss = 'binder', estimate = sim_matrix)
# estimate_partition = as.vector(binder_dahl)
#
# # Get quality indicies
# Kest = length(unique(estimate_partition))
# cat('\n Estimated number of clusters \n')
# Kest
# Binder_loss = salso::binder(truth = real_partition, estimate = estimate_partition)
# cat('\n Binder loss function \n')
# Binder_loss
# Tab = table(real_partition,estimate_partition)
# cat('\n Miss classification table \n')
# Tab
```