Playing with dirichletprocess

```
## Loading required package: dirichletprocess
## Loading required package: mvtnorm
## Loading required package: foreach
```

Questions on `clusterParameters`

```
## Generate 10 datapoints on 2D data space from two multinormals:
## one centred at (-1,-1), the other at (1,1),
## with sd = 0.1
datasize <- 10
y <- matrix(rnorm(n=2*datasize, mean=c(-1,-1,1,1), sd=0.1), nrow=datasize, ncol=2, byrow=TRUE)
y
```

```
##         [,1]   [,2]
##  [1,] -0.851 -1.000
##  [2,]  1.138  0.962
##  [3,] -0.982 -1.025
##  [4,]  0.878  1.156
##  [5,] -0.957 -1.120
##  [6,]  1.105  0.869
##  [7,] -1.069 -0.940
##  [8,]  0.980  0.881
##  [9,] -1.201 -0.999
## [10,]  1.052  0.925
```

```
## Create Dirichlet-process object, multinormal mixture:
dp <- DirichletProcessMvnormal(y)

## This object has clusterParameters, with specific values:
dp$clusterParameters
```

```
## $mu
## , , 1
##
##       [,1]    [,2]
## [1,] 0.093 -0.0895
##
##
## $sig
## , , 1
##
##      [,1] [,2]
## [1,] 1.17 1.06
## [2,] 1.06 1.07
```

Question: where do the `clusterParameters` in the initial Dirichlet-process object come from? Can they be

considered prior samples of the θs, rather than posterior?

```
## Fit the first Dirichlet-process, save the result under a new name:
fitdp <- Fit(dp, its=1000, progressBar=FALSE)
```

The parameters in the first sample in the fitted object seem to be equal to the `clusterParameters` in the initial object:

```
fitdp$clusterParametersChain[[1]]
```

```
## $mu
## , , 1
##
##       [,1]    [,2]
## [1,] 0.093 -0.0895
##
##
## $sig
## , , 1
##
##      [,1] [,2]
## [1,] 1.17 1.06
## [2,] 1.06 1.07
```

Question: in general, the `clusterParameters` of a fitted object are effectively the last samples of the Monte Carlo chain. Is this correct?

LikelihoodDP

This actually gives the matrix of probability densities of the data conditional on the `clusterParameters` (or equivalently the likelihood of the `clusterParameters` in view of the data)

$$\big(\mathrm{p}(y_i|\theta_j, \text{hyperparameters})\big)_{ij}$$

as can be seen by an explicit calculation with `dmvnorm`:

```
## with LikelihoodDP
LikelihoodDP(fitdp)
```

```
##          [,1]   [,2]   [,3]   [,4]   [,5]     [,6]   [,7]   [,8]   [,9]  [,10]
##  [1,] 0.2539 0.2539 0.2539 0.2539 0.2539 1.00e-05 0.2539 0.2539 0.2539 0.0533
##  [2,] 0.1607 0.1607 0.1607 0.1607 0.1607 1.56e-02 0.1607 0.1607 0.1607 0.0718
##  [3,] 0.2731 0.2731 0.2731 0.2731 0.2731 1.79e-05 0.2731 0.2731 0.2731 0.0336
##  [4,] 0.0882 0.0882 0.0882 0.0882 0.0882 1.32e-01 0.0882 0.0882 0.0882 0.1415
##  [5,] 0.1912 0.1912 0.1912 0.1912 0.1912 3.99e-06 0.1912 0.1912 0.1912 0.0365
##  [6,] 0.1753 0.1753 0.1753 0.1753 0.1753 9.98e-03 0.1753 0.1753 0.1753 0.0780
##  [7,] 0.3175 0.3175 0.3175 0.3175 0.3175 9.25e-05 0.3175 0.3175 0.3175 0.0233
##  [8,] 0.2189 0.2189 0.2189 0.2189 0.2189 2.18e-02 0.2189 0.2189 0.2189 0.1141
##  [9,] 0.2586 0.2586 0.2586 0.2586 0.2586 9.65e-05 0.2586 0.2586 0.2586 0.0129
## [10,] 0.1906 0.1906 0.1906 0.1906 0.1906 1.97e-02 0.1906 0.1906 0.1906 0.0936
```

```
## with dmvnorm
unname(foreach(i=1:datasize, .combine='rbind') %:% # iterate over data
       foreach(j=1:datasize, .combine='cbind') %do% { #iterate over params
           dmvnorm(x=y[i,],
                   mean=fitdp$clusterParameters$mu[,,fitdp$clusterLabels[j]],
                   sigma=fitdp$clusterParameters$sig[,,fitdp$clusterLabels[j]],
```

```
                  checkSymmetry=FALSE, log=FALSE)
      })
```

```
##          [,1]   [,2]   [,3]   [,4]   [,5]     [,6]   [,7]   [,8]   [,9]  [,10]
##  [1,] 0.2539 0.2539 0.2539 0.2539 0.2539 1.00e-05 0.2539 0.2539 0.2539 0.0533
##  [2,] 0.1607 0.1607 0.1607 0.1607 0.1607 1.56e-02 0.1607 0.1607 0.1607 0.0718
##  [3,] 0.2731 0.2731 0.2731 0.2731 0.2731 1.79e-05 0.2731 0.2731 0.2731 0.0336
##  [4,] 0.0882 0.0882 0.0882 0.0882 0.0882 1.32e-01 0.0882 0.0882 0.0882 0.1415
##  [5,] 0.1912 0.1912 0.1912 0.1912 0.1912 3.99e-06 0.1912 0.1912 0.1912 0.0365
##  [6,] 0.1753 0.1753 0.1753 0.1753 0.1753 9.98e-03 0.1753 0.1753 0.1753 0.0780
##  [7,] 0.3175 0.3175 0.3175 0.3175 0.3175 9.25e-05 0.3175 0.3175 0.3175 0.0233
##  [8,] 0.2189 0.2189 0.2189 0.2189 0.2189 2.18e-02 0.2189 0.2189 0.2189 0.1141
##  [9,] 0.2586 0.2586 0.2586 0.2586 0.2586 9.65e-05 0.2586 0.2586 0.2586 0.0129
## [10,] 0.1906 0.1906 0.1906 0.1906 0.1906 1.97e-02 0.1906 0.1906 0.1906 0.0936
```

Posterior predictive in conjugate multinormal case

question on `PosteriorClusters`: where are they drawn from? can they be used to draw from the prior? They also work in case of unfitted object: where are they drawn from?

What does `ClusterLabelPredict` do?

How do I draw prior probabilities from the process?