# Statistical learning in Finance

## EM ALGORITHM METHODOLOGY

In Gaussian Mixture Models, EM Algorithm by Dempster et al. (1977) is used to determine the data set values belongs to which observation. Firstly in E-Step, we evaluate the probability of each point in data set that belongs to which observation. Then in M-Step, the sample mean and covariance is computed based on the probability derived in earlier step.

The process is initialized by developing a dataset based on samples' initial mean, probability and covariance values, using mvnrnd command. A multivariate normal distribution density function is used on the data set that generate a matrix with sample values in rows and columns denotes the sample.

$$f_x(x_1, \ldots, x_k) = \frac{\exp\left({}^1\!/_2\,(x-\mu)\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^k}|\Sigma|}$$

Whereas, probability of the dataset value that belongs to which observation is calculated as follows:

$$wp_j^i = \frac{f_j(x)\alpha_j}{\sum_{p=1}^{K} f_p(x)\alpha_p}$$

Then in M-step, we figured out a weighted mean of the dataset values and weighted probabilities.

$$wAvg = \frac{\sum_{i=1}^{m}(wp_i hX_i)}{\sum_{i=1}^{m}(wp_i)}$$

Hence, applied M-Step technique as given below:

$$\alpha_j = \frac{1}{m}\sum_{1}^{m} wp_j^i$$

$$\widehat{\mu_J} = \frac{\sum_{i=1}^{m} wp_j^i hX_i}{\sum_{i=1}^{m} wp_j^i}$$

$$\Sigma_j = \frac{\sum_1^m wp_j^i\left(hX^i - \widehat{\mu_J}\right)\left(hX^i - \widehat{\mu_J}\right)^T}{\sum_1^m wp_j^i}$$

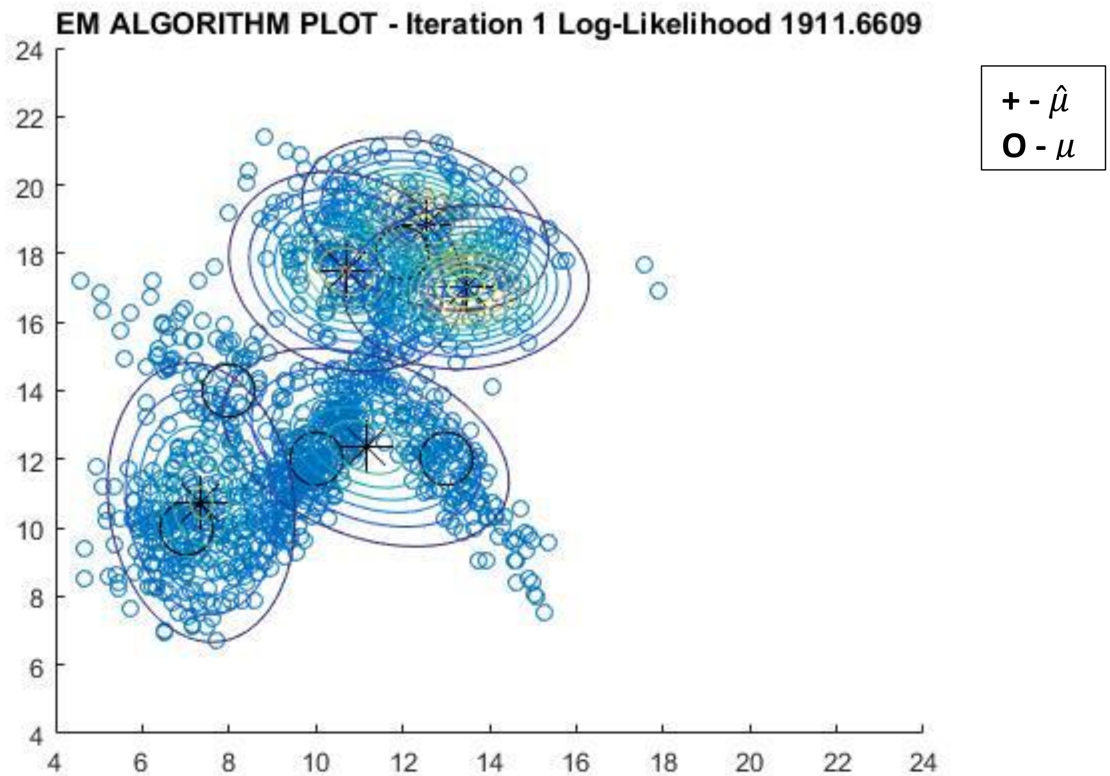
The values of the estimated parameters and their true values are reasonably close. The 2D plots show estimated probability density functions by contour plots.

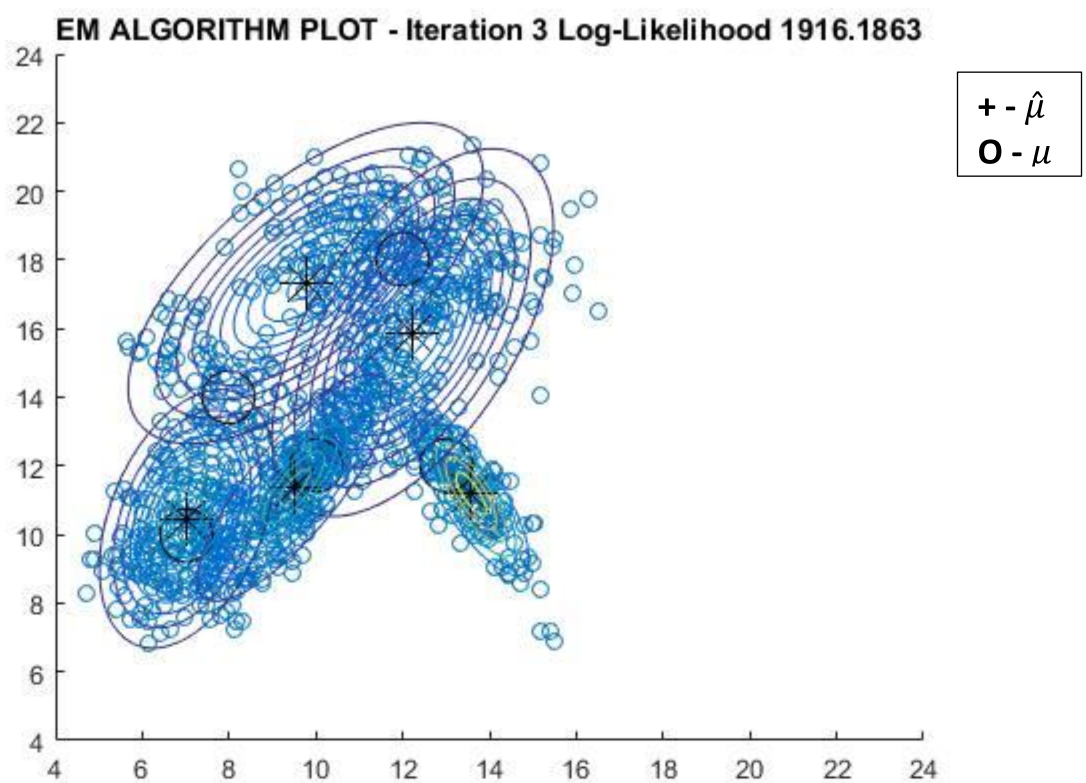
$$\hat{\mu} = \left[\frac{7.9}{14.3}, \frac{6.99}{10.1}, \frac{13}{12.1}, \frac{12.1}{17.8}, \frac{10.1}{12}\right]$$

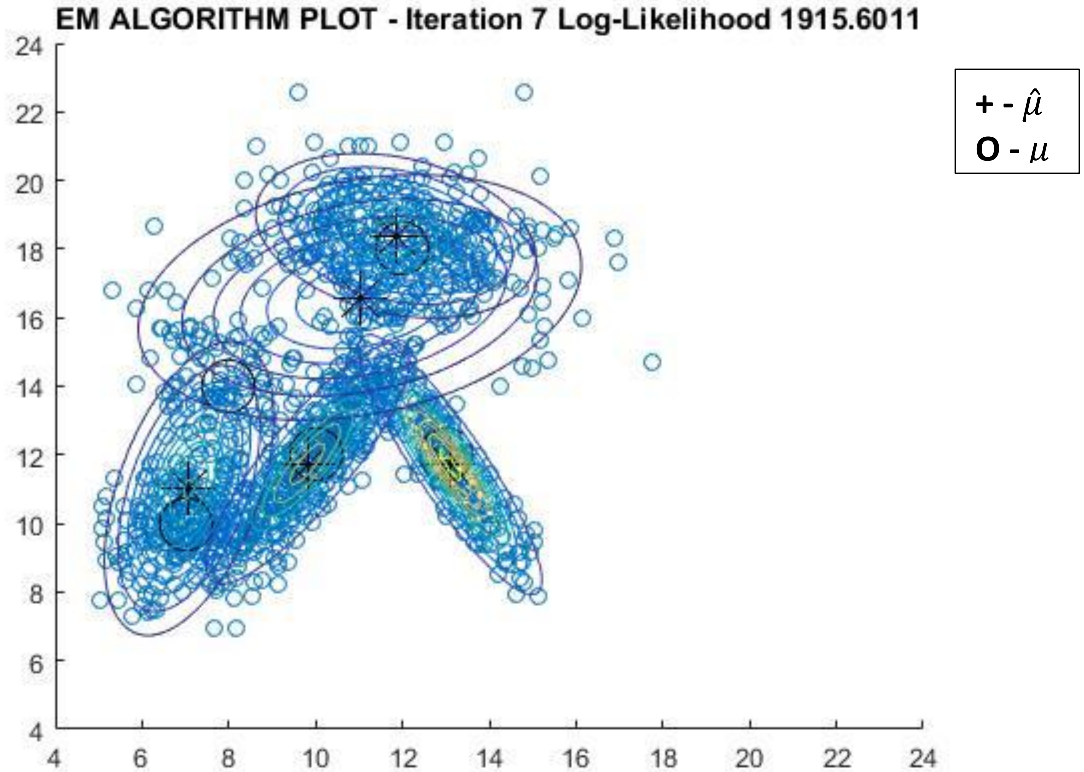
$$\hat{\alpha} = [0.1, 0.208, 0.14, 0.30, 0.27]$$

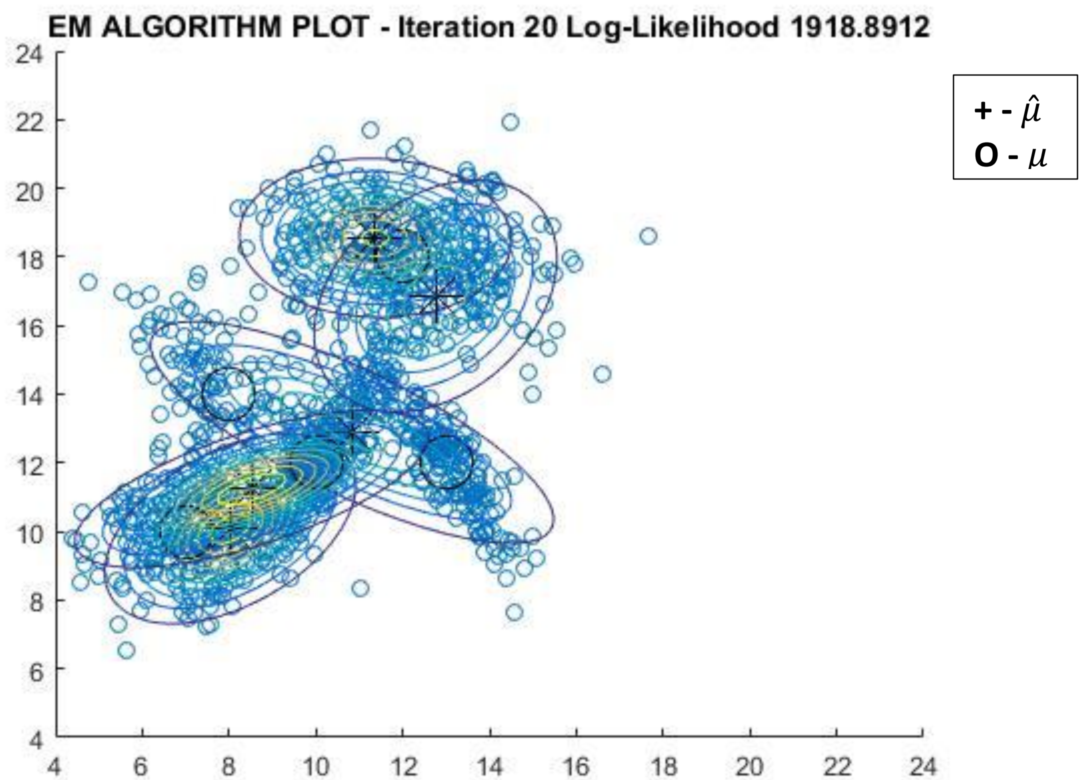
Below are the five plots, each plot contains

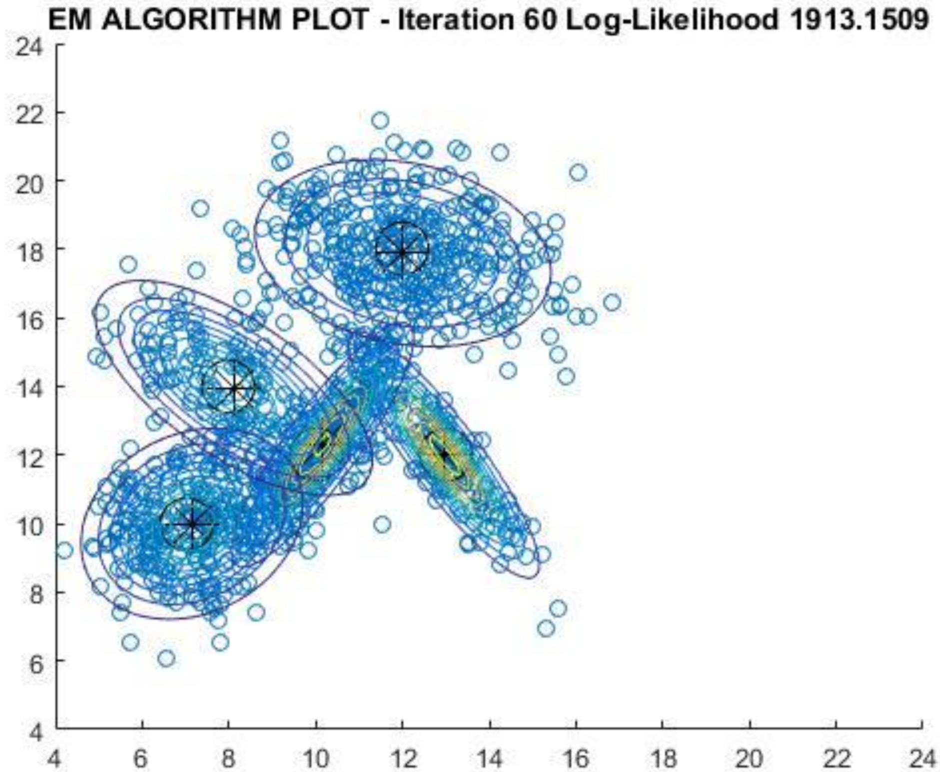
- The iteration number and the log-likelihood value.
- The generated sample points.
- The true values for $\mu$j , denoted by **o**
- The estimated values for $\hat{\mu}$j denoted by **+**
- The level curves of the estimated density functions.


EM ALGORITHM PLOT - Iteration 1 Log-Likelihood 1911.6609
+ - $\hat{\mu}$
O - $\mu$
24
22
20
18
16
14
12
10
8
6
4
4
6
8
10
12
14
16
18
20
22
24


EM ALGORITHM PLOT - Iteration 3 Log-Likelihood 1916.1863
+ - $\hat{\mu}$
O - $\mu$
24
22
20
18
16
14
12
10
8
6
4
4
6
8
10
12
14
16
18
20
22
24


EM ALGORITHM PLOT - Iteration 7 Log-Likelihood 1915.6011
24 22 20 18 16 14 12 10 8 6 4
4 6 8 10 12 14 16 18 20 22 24
+ - $\hat{\mu}$
O - $\mu$


EM ALGORITHM PLOT - Iteration 20 Log-Likelihood 1918.8912
24 22 20 18 16 14 12 10 8 6 4
4 6 8 10 12 14 16 18 20 22 24
+ - $\hat{\mu}$
O - $\mu$


EM ALGORITHM PLOT - Iteration 60 Log-Likelihood 1913.1509
24
22
20
18
16
14
12
10
8
6
4
4
6
8
10
12
14
16
18
20
22
24

**+ -** $\hat{\mu}$
**O -** $\mu$