It might be the case that we want to draw samples from an aribitrary posterior pdf having more than one mode. In such cases, normal approximation might be too bad an approximation. We consider modeling this arbirary pdf as a mixture of normal densities. Further, the number of modes (which in turn dictate the number of components in the mixute) in the posterior is also assumed unknown.

The mixture model considered in [1] tries to estimate the model parameters based on data. However, in the present context, we have the pdf in semi-closed form and this becomes a special case of [1].

Let $f(\theta)$ be the known arbitrary pdf which we want to approximate. For illustration purposes, we assume a bivarite density ($\theta = [\theta_1, \theta_2]^T$) and discretize θ on a uniform $N \times M$ grid. Further let θ_{mn} represent the $(\theta_1(n), \theta_2(m))$ sample in the grid. Also assume that the discretized pdf ($f(\theta_{mn})$)is appropriately normalized.

$$f(\theta_{mn}; \Phi) = \sum_{k=1}^{K} \lambda_k f_k(\theta_{mn}; \phi_k) \tag{1}$$

where

- K is the total number of densities (unknown)
- $\boldsymbol{\phi}_k = [\mu_k, \Sigma_k]$ represent the mean vector and covariance matrix of the kth density
- $f_k \sim \mathcal{N}(\mu_k, \Sigma_k)$
- $\{\lambda_k\}$ represent the mixing weights with constraint $\sum_{k=1}^{K} \lambda_k = 1$
- $\Phi = [\phi_1 \dots \phi_K, \lambda_1 \dots \lambda_K]$

Then we obtain the estimates for the mixture model parameters as:

$$\hat{\lambda}_{k,new} = \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi}) \tag{2}$$

$$\hat{\mu}_{k,new} = \frac{1}{\hat{\lambda}_{k,new}} \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi}) \theta_{mn} \tag{3}$$

$$\hat{\Sigma}_{k,new} = \frac{1}{\hat{\lambda}_{k,new}} \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_k(\theta_{mn}; \hat{\Phi}) (\theta_{mn} - \hat{\mu}_k)^T (\theta_{mn} - \hat{\mu}_k) \tag{4}$$

where

$$\gamma_k(\theta_{mn}; \hat{\Phi}) = \frac{\hat{\lambda}_k f_k(\theta_{mn}; \hat{\theta}_k) f(\theta_{mn}; \Phi)}{f(\theta_{mn}; \hat{\Phi})} \tag{5}$$

However, we need to know the number of components K in the above equations. In order to circumvent this difficulty, we overestimate the number of components

and estimate the model components. Later, we repeatedly: 1) merge any two distributions and 2) replace them with a new merged distribution (still a normal distribution) such that the distance between the newly formed distribution and the two distributions which got merged is minimized and this distance is above a certain threshold. Essentially, we overfit the posterior distribution and later prune the mixture until no further merging is possible. Based on the distance measure chosen, there could be different approaches. For a discussion please see [2, 3].

Now it becomes a trivial task to generate samples from the posterior. Instead of drawing the samples from $p(\theta)$, we draw from the mixture. The algorithm is outlined below:

- $u \sim \mathcal{U}[0, 1]$
- choose kth component if $\sum_{i=0}^{k-1} \lambda_i \leq u < \sum_{i=0}^{k} \lambda_i$
- generate a multivariate random number from $f_k(\mu_k, \Sigma_k)$

Essentially, we have K number of states and each state has a certain pdf. Based on this relationship, we can draw random numbers quite efficiently.

References

[1] Soma Sekhar Dhavala, "EM algorithm for Gaussian Mixture modeling," report , ECpE Dept., Iowa State University, url: www.asankhya.org/me/research/stat544/gmm.pdf.

[2] Mark Coates, *Time-frequency modeling*, Ph. D, Univ. of Cambridge, Signal Procesing and Communications Laboratory, Dept. of Engineering, August 1998.

[3] M̃. Figueiredo J. Leitao and A. K. Jain, "On fitting mixture models," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 1654, pp. 54–69. Springer-Verlag, 1999.