國立交通大學

統計學研究所

碩士論文

利用混合模型的方法對正子電腦斷層掃描影像做影像分割
並與 K-means 及常態混合模型的方法比較

Segmentation of PET/CT images by Flexible Mixture Models and
Comparison with K-means and Gaussian Mixture Models

研 究 生：葉孟樵

指導教授：盧鴻興　教授

中 華 民 國 九 十 八 年 六 月

利用混合模型的方法對正子電腦斷層掃描影像做影像分割
並與 K-means 及常態混合模型的方法比較

Segmentation of PET/CT images by Flexible Mixture Models and
Comparison with K-means and Gaussian Mixture Models

研 究 生：葉孟樵　　　　　　Student：Meng-Ciao Ye

指導教授： 盧鴻興 博士　　　　Advisor：Dr. Henry Horng-Shing Lu

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2009

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 八 年 六 月

利用混合模型的方法對正子電腦斷層掃描影像做影像分割並與K-means及常態混合模型的方法比較

學生：葉孟樵　　　　　　　　　　　　　指導教授：盧鴻興　教授

國立交通大學統計學研究所　碩士班

<u>摘　　要</u>

　　正子電腦斷層掃描影像能協助醫師判定異常部位,在 PET 影像上特異的亮點或暗點則表示這些異常部位可能發生的位置,因此 PET 影像的分割是非常重要。我們使用了以下方法去進行影像的分割以圈選出我們所感興趣的區域,包含了 K-means, Gaussian mixture model (GMM)以及 Flexible mixture model (FMM)。FMM 與 GMM 最大的差異在於兩者所使用的混合分配,FMM 此混合模型多考慮了 PET 影像的結構的特性,使用右偏分配估計背景的部份,另外以常態分配的混合估計非背影之影像。而 FMM 之分群結果也比 K-means 與 GMM 較佳。

i

Segmentation of PET/CT images by Flexible Mixture Models and Comparison with K-means and Gaussian Mixture Models

Student: Meng-Ciao Ye   Advisors: Dr. Henry Horng-Shing Lu

Institute of Statistics
National Chiao Tung University

ABSTRACT

Positron Emission Tomography (PET) helps doctors determine the abnormal regions. The specific brightened regions in PET images show the location of abnormal region. Hence the segmentation of the data form PET images is very important. There are three methods to classify the data from PET image to obtain the region of interest, K-means with KDE, Gaussian mixture model (GMM) with KDE and flexible mixture model (FMM) with KDE. The main difference between GMM and FMM is that, GMM uses several normal distributions to fit the original PET data, while FMM does not. FMM considers the property and structure of PET image data. It uses a right-skewed distribution to fit the background images. The mixture normal distribution is used to fit other regions. Finally, the result of FMM with KDE is better than the result of K-means with KDE and GMM with KDE.

Contents

Chapter 1.   Introduction

Positron Emission Tomography (PET) is a very important medical tool. It helps doctors determine the abnormal **region in the human body**. Multi-functional medical scanning is developed in recent years such that nuclear medical imagining and other traditional imaging functions can be combined. It can provide exact dissected locations of nuclear medical imagining and attenuation correction, such as, PET and CT, PET and MRI, SPECT and CT, SPECT and MRI, SPECT and CT and X-ray and MRI… etc. For tumor scanning, because glucose absorption ratios between normal cells and cancer cells are different for nuclear medicine in PET images, high contrast images can be provide by PET images. The region which has high absorption ratio shows high intensity and this region is most likely the region of primary tumor or high-activity tumor. The methods like K-means, Gaussian mixture model (GMM) and flexible mixture model (FMM) are used to classify the data for finding the regions of interest.

The concept of FMM is very similar to GMM. They both use EM algorithm to estimate the unknown parameters in their models to classify the data. The main difference between the two methods is that, GMM uses several normal distributions to fit the original PET data, while FMM does not. The FMM method considers the property and structure of the data so that the model can be consistent with the original data. It can be detected by observing the data of PET images, that the proportion of small pixel value is very large so the distribution is skewed to the right in this part. Therefore in FMM, the right-skewed distribution will be taken to estimate such part. The mixture normal distribution is used to estimate other parts. Before classifying data, we have to consider how to select the number of groups, how to give initial values for GMM and how to give the initial seeds for K-means. Kernel density estimation (KDE) is used for solving these issues. It provides the initial values for GMM and FMM and the initial seeds for K-means. The number of groups can also be provided by KDE. Then we

can get the several sets of estimations and number of groups. Finally, the issue of model selection is then solved by Akaike information criterion (AIC) or Bayesian information criterion (BIC).

K-means with KDE, GMM with KDE and FMM with KDE, are methods used for clustering. The PET phantom experiment is designed for comparing these methods. There are 63 slices where size is 256×256. Hence all data size is $63 \times 256 \times 256$. The exact locations of every kind of activity regions is known, so we can compare the accuracy of different methods. Additionally, a computerized tomography (CT) image is classified by these methods. The CT image is a sagittal plane of a human body and its data size is 512×373. The result of classifying the CT image can be compared to the original CT image because of good outlining effect. The FMM with KDE method is better than K-means and GMM in the PET phantom experiment and the CT scanning.

Chapter 2. Methodologies

2.1 K-means Clustering

K-means is a typical method for clustering. The goal is to decrease the square error for each observation belonging to the cluster center. Suppose that data size is N, and we want to partition the data into K clusters. After K-means algorithm, the kth group is expressed as G_k. The set G_k whose centroid is μ_k contains n_k data points. Then the square error of set G_k can be defined as (2.1.1)

$$e_k = \sum_{x_i \in G_i} (x_i - \mu_k)^2, k = 1, \cdots, K ;$$

(2.1.1)

$$E = \sum_{k=1}^{K} e_k .$$

(2.1.2)

In equation (2.1.2), E is the sum of the square error, or the overall clustering error. K-means clustering method is designed to minimize E. The algorithm will be described in detail.
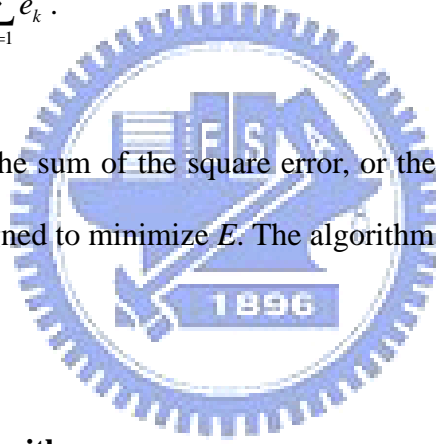
K-means Clustering Algorithm

Step 1. Input data $X = \{x_1, x_2, \cdots, x_N\}$

Step 2. Decide the number of groups K, and randomly selected seeds in K to be the centroids of each group. The centroids are $\mu_1, \mu_2, \cdots, \mu_K$

Step 3. Calculate the distance between each data points and its group centroids according K centroids from last step. Join the data point to the group with the nearest centroid. The kth group set is shown as follows

$$G_k = \left\{ x_i \mid \|x_i - \mu_k\| \le \|x_i - \mu_j\| \quad \forall j = 1, \cdots, K \right\} , k = 1, \cdots, K .$$

(2.1.3)

Step 4. Recalculate K centroids as follows

$$\mu_k = \frac{\sum\limits_{x_i \in G_k} x_i}{n_k}, k = 1, \cdots, K.$$ (2.1.4)

Step 5.   If the differences of centroids in old and new iterations are smaller than a tolerance level, then the iteration process stops. Else, the iteration process will go to Step2.

2.2 Gaussian Mixture Model (GMM)

The locations of high pixel value in PET image are likely to be the abnormal regions, we want to find an appropriate model to fit the true data $X = \{x_1, x_2, \cdots, x_N\}$. We further segment X by this model. If distribution of data is shown to be a single elliptical curve, we can fit it with one distribution. If it is shown to be a complex curve, then using a single distribution is not suitable to describe the probability density function of the data. Therefore, we consider Gaussian mixture model (GMM) [7]. This model consists of K different normal distributions corresponding to K different weights $\pi_1, \pi_2, \cdots, \pi_K$. The model is expressed as

$$f(x_i | \Phi) = \sum_{k=1}^{K} \pi_k f_k(x_i | \theta_k) \text{ with } \sum_{k=1}^{K} \pi_k = 1, \ 0 \le \pi_k \le 1,$$
$$\text{and} \ \ \Phi = (\theta_1, \theta_2, \cdots, \theta_K, \pi_1, \pi_2, \cdots, \pi_K),$$ (2.2.1)

where $f_k(x | \theta_k)$ represents probability density function of the kth normal distribution, and its parameters denote $\theta_k = (\mu_k, \sigma_k)$. Next, we use maximum likelihood estimator (MLE) to estimate the parameters of Gaussian mixture model,

$$\log L(\Phi) = \log \prod_{i=1}^{N} f(x_i | \Phi)$$ (2.2.2)

$$= \log \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k f_k(x_i | \theta_k) \right)$$ (2.2.3)

$$= \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k f_k(x_i | \theta_k) \right).$$ (2.2.4)

In order to find $\Phi = (\theta_1, \theta_2, \cdots, \theta_K, \pi_1, \pi_2, \cdots, \pi_K)$ such that log likelihood function is maximized, we perform partial differentiation,

Let $\dfrac{\partial \log L(\Phi)}{\partial \Phi} = 0$ for obtaining $\hat{\Phi}_{MLE} = \arg\max_{\Phi} L(\Phi)$. 　　(2.2.5)

It is very difficult to obtain Φ from equation (2.2.5). Therefore the EM algorithm is used to estimate the parameters. First the indicator function Y_{ik} is defined as

$$Y_{ik} = \begin{cases} 1, & \text{if } x_i \text{ comes from kth normal distribution;} \\ 0, & \text{otherwise.} \end{cases} \qquad (2.2.6)$$

Let Y_i to be $(Y_{i1,} \cdots, Y_{iK})$. Then, the joint probability density function of x_i and y_i is

$$f(x_i, y_i) = \sum_{k=1}^{K} y_{ik} \pi_k f(x_i \mid \theta_k). \qquad (2.2.7)$$

The log likelihood function of joint density probability function is

$$\log L(\Phi) = \log \prod_{i=1}^{N} f(x_i, y_i) \qquad (2.2.8)$$

$$= \sum_{i=1}^{N} \log f(x_i, y_i) \qquad (2.2.9)$$

$$= \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} y_{ik} \pi_k f_k(x_i \mid \theta_k) \right) \qquad (2.2.10)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log \left(\pi_k f_k(x_i \mid \theta_k) \right) \qquad (2.2.11)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log(\pi_k) + \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log(f_k(x_i \mid \theta_k)). \qquad (2.2.12)$$

Given $\Phi^{(old)}$ and X then the Expectation step in EM algorithm starts.

$$Q(\Phi; \Phi^{(old)}) = E(\log L(\Phi) \mid X, \Phi^{(old)}) \qquad (2.2.13)$$

$$= E\left(\sum_{i=1}^{N} \sum_{k=1}^{K} Y_{ik} \log(\pi_k) + \sum_{i=1}^{N} \sum_{k=1}^{K} Y_{ik} \log(f_k(x_i \mid \theta_k)) \Big| x, \Phi^{(old)} \right) \qquad (2.2.14)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} E(Y_{ik} \mid x, \Phi^{(old)}) \log \pi_k + \sum_{i=1}^{N} \sum_{k=1}^{K} E(Y_{ik} \mid x, \Phi^{(old)}) \log(f_k(x_i \mid \theta_k)). \qquad (2.2.15)$$

Equation (22.15) of $E(Y_{ik} \mid x, \Phi^{(old)})$ is

$$E(Y_{ik} \mid x, \Phi^{(old)}) = 1 \cdot f_{y_{ik}}(1 \mid x, \Phi^{(old)}) + 0 \cdot f_{y_{ik}}(1 \mid x, \Phi^{(old)}) \qquad (2.2.16)$$

$$= \frac{f(x_i, Y_{ik} = 1 \mid \Phi^{(old)})}{f(x_i \mid \Phi^{(old)})} \qquad (2.2.17)$$

$$= \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} . \tag{2.2.18}$$

So, $Q(\Phi; \Phi^{(old)})$ is obtained with combination of equation (2.2.15) and equation (2.2.18) as follows:

$$Q(\Phi; \Phi^{(old)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log \pi_k + \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_k(x_i \mid \theta_k)) . \tag{2.2.19}$$

Equation (2.2.19) is divided into two parts for convenience of maximization step in the EM algorithm. The first part is

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log \pi_k . \tag{2.2.20}$$

The second part is

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_k(x_i \mid \theta_k)) . \tag{2.2.21}$$

Formula (2.2.20) is only related with the parameter π_k, and formula is only related to the parameter θ_k. In other words, the issue of choosing the parameters that maximizes $Q(\Phi; \Phi^{(old)})$ is now simplified to choosing the parameter π_k that maximizes formula (2.2.20) and choosing the parameter θ_k that maximizes formula (2.2.21). The first formula (2.2.20) is calculated as follows.

let $\lambda \in \mathbb{R}$, and

$$\frac{\partial}{\partial \pi_k} \left(\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum\limits_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log \pi_k + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \right) = 0, \quad k = 1, \dots, K \tag{2.2.22}$$

$$\Rightarrow \quad \sum_{i=1}^{N} \frac{1}{\pi_k} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} + \lambda = 0, \quad k = 1, ..., K \qquad (2.2.23)$$

$$\Rightarrow \quad \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} + \pi_k \lambda = 0, \quad k = 1, ..., K \qquad (2.2.24)$$
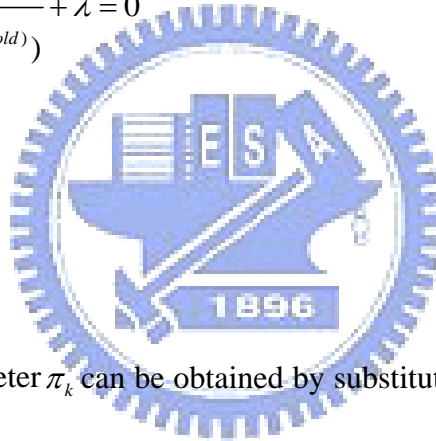
$$\Rightarrow \quad \sum_{k=1}^{K} \left(\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} + \pi_k \lambda \right) = 0 \qquad (2.2.25)$$

$$\Rightarrow \quad \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} + \lambda \sum_{k=1}^{K} \pi_k = 0 \qquad (2.2.26)$$

$$\Rightarrow \quad \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} + \lambda = 0 \qquad (2.2.27)$$

$$\Rightarrow \quad \sum_{i=1}^{N} 1 + \lambda = 0 \qquad (2.2.28)$$

$$\Rightarrow \quad \lambda = -N . \qquad (2.2.29)$$

The estimation for parameter π_k can be obtained by substituting the result of equation (2.2.29)

into equation (2.2.24).

$$\pi_k^{(new)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} . \qquad (2.2.30)$$

And because of

$$\frac{\partial^2}{\partial \pi_k^2} \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log \pi_k < 0 . \qquad (2.2.31)$$

The result of (2.2.31) means

$$\pi_k^{(new)} = \arg\max_{\pi_k} \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log \pi_k \ . \tag{2.2.32}$$

In other words, the estimation $\pi_k^{(new)}$ maximizes formula (2.2.20). Next, the second formula (2.2.21) is expanded. The parameter set θ_k in formula (2.2.21) contains μ_k and σ_k

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_k(x_i \mid \theta_k))$$

$$= \sum_{i=1}^{N} \left(\frac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_1(x_i \mid \theta_1)) + \cdots + \frac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_K(x_i \mid \theta_K)) \right) \tag{2.2.33}$$

$$= \sum_{i=1}^{N} \left(\begin{array}{l} \dfrac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_1^2 - \dfrac{(x_i - \mu_1)^2}{2\sigma_1^2} \right) \\[4mm] + \cdots + \dfrac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_K^2 - \dfrac{(x_i - \mu_K)^2}{2\sigma_K^2} \right) \end{array} \right) . \tag{2.2.34}$$

In order to find θ_k such that formula (2.2.21) is maximized, we partially differentiate with μ_k,

Let $\dfrac{\partial}{\partial \mu_k} \displaystyle\sum_{i=1}^{N} \left(\begin{array}{l} \dfrac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_1^2 - \dfrac{(x_i - \mu_1)^2}{2\sigma_1^2} \right) \\ + \cdots + \dfrac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_K^2 - \dfrac{(x_i - \mu_K)^2}{2\sigma_K^2} \right) \end{array} \right) = 0 \qquad (2.2.35)$

$$\Rightarrow \quad \sum_{i=1}^{N} \left(\frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(\frac{x_i - \mu_k}{\sigma_k^2} \right) \right) = 0 \qquad (2.2.36)$$

$$\Rightarrow \quad \sum_{i=1}^{N} x_i \, \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} - \mu_k \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} = 0 \qquad (2.2.37)$$

$$\Rightarrow \quad \mu_k^{(new)} = \frac{\displaystyle\sum_{i=1}^{N} x_i \, \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}. \qquad (2.2.38)$$

In addition, σ_k must be considered for obtaining the parameter set θ_k such that formula

(2.2.21) is maximized. We partially differentiate with σ_k,

Let $\dfrac{\partial}{\partial \sigma_k^2} \displaystyle\sum_{i=1}^{N} \left(\begin{array}{l} \dfrac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_1^2 - \dfrac{(x_i - \mu_1)^2}{2\sigma_1^2} \right) \\ + \cdots + \dfrac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_K^2 - \dfrac{(x_i - \mu_K)^2}{2\sigma_K^2} \right) \end{array} \right) = 0$

$$\Rightarrow \quad \sum_{i=1}^{N} \left(\frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\frac{1}{2\sigma_k^2} + \frac{(x_i - \mu_k)^2}{2(\sigma_k^2)^2} \right) \right) = 0 \qquad (2.2.39)$$

$$\Rightarrow \quad \sigma_k^{2(new)} = \frac{\displaystyle\sum_{i=1}^{N} (x_i - \mu_k)^2 \, \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}. \qquad (2.2.40)$$

Partial derivatives of formula (2.2.21) with respect to μ_k and σ_k are less than zero. It means the result $\mu_k^{(new)}$ of equation (2.2.38) and $\sigma_k^{2(new)}$ of equation (2.2.40) maximizes formula (2.2.21). So $Q(\Phi;\Phi^{(old)})$ is maximized by the results of equations (2.2.30), (2.2.38) and (2.2.40). Next we consider how to use the estimations to segment data. First let $x_1, x_2, \cdots, x_N$ be from K normal distributions with weight $\pi_1, \pi_2, \cdots, \pi_N$. The data point x_i is classified to No. z_i group. Therefore,

$$f(z_i = j \mid \Phi) = \pi_j \quad , \ i = 1, \cdots, N \ \text{ and } \ j = 1, \cdots, K \qquad (2.2.41)$$

The conditional probability density function of z_i given data point x_i and the estimations of parameters Φ is

$$f(z_i \mid x_i, \Phi) = \frac{f(x_i, z_i, \Phi)}{f(x_i, \Phi)} \qquad (2.2.42)$$

$$= \frac{f(x_i \mid z_i, \Phi) f(z_i, \Phi)}{f(x_i, \Phi)} \qquad (2.2.43)$$

$$= \frac{f(x_i \mid z_i, \Phi) f(z_i \mid \Phi) f(\Phi)}{f(x_i \mid \Phi) f(\Phi)} \qquad (2.2.44)$$

$$= \frac{f(x_i \mid z_i, \Phi) f(z_i \mid \Phi)}{f(x_i \mid \Phi)} \qquad (2.2.45)$$

$$= \frac{f_{z_i}(x_i \mid \theta_{z_i}) \pi_{z_i}}{\sum_{j=1}^{K} \pi_j f_j(x_i \mid \theta_j)} \quad , z_i \in \{1, 2, \cdots, K\}. \qquad (2.2.46)$$

Thus the probability that the data point x_i comes from the kth normal distribution is

$$f(z_i = k \mid x_i, \Phi) = \frac{f_k(x_i \mid \theta_k) \pi_k}{\sum_{j=1}^{K} \pi_j f_j(x_i \mid \theta_j)} \quad . \qquad (2.2.47)$$

Take equation (2.2.47) to be the standard for data segmentation. Then the set G_m where the mth group is as follows

$$x_i \in G_m \quad ,where \quad m = \arg\max_k \left(\frac{f_k(x_i \mid \theta_k)\pi_k}{\sum\limits_{j=1}^{K} \pi_j f_j(x_i \mid \theta_j)} \right). \tag{2.2.48}$$

The estimations for the parameters are updated by the EM algorithm. After iterating with newly updated estimates, it can be ensured that likelihood is increasing. The detailed algorithm is

GMM Clustering Algorithm

Step 1.   Input all data $X = \{x_1, x_2, \cdots, x_N\}$

Step 2.   Decide the number of groups K, and give initial value $\Phi^{(old)}$.

Step 3.   Update the estimations by s (2.2.30), (2.2.37) and (2.2.39) to obtain the new estimations $\Phi^{(new)}$.

Step 4.   If $\log L(\Phi^{(new)}) - \log L(\Phi^{(old)}) < tolerance$, stop iteration with newly updated estimates, else replace $\Phi^{(old)}$ by new estimations and repeat step3.

Step 5.   After iterations finishes, use equation (2.2.48) to classify data points.

2.3 Flexible Mixture Model (FMM)

The data is modeled by several normal distributions in section 2.2. However, it is observed that the proportion of small pixel value is very large in all PET images. The main reason for this phenomenon is that luminosity of non-living parts in PET images is very weak but occupies large area.   The region is called the background image. The following is an example of mouse brain. It shows that the histogram for pixels of a mouse brain.
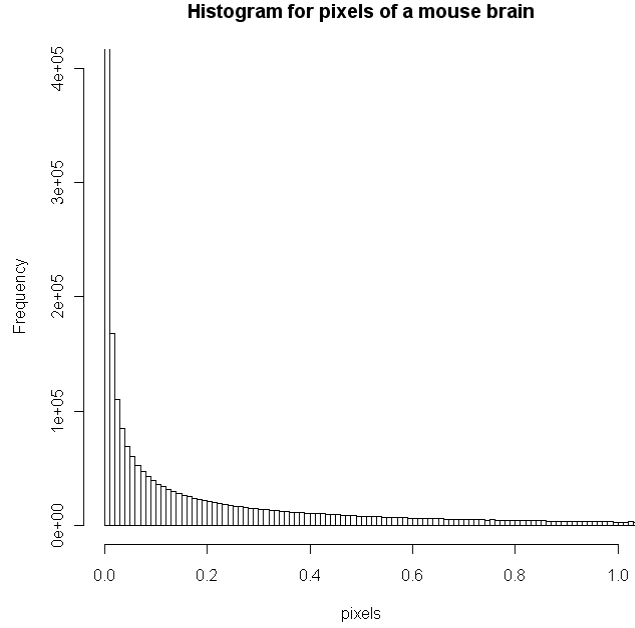
Figure 2.3.1: It is the histogram for small pixels of a mouse brain in MicroPET image

The distribution of all background pixels tends to accumulate in the left in figure 2.3.1, and the peak is found when pixel value is zero. Therefore, the right-skewed distribution is considered to be the model of small pixels in PET image. Otherwise, we use mixed normal distribution to model non-background data. The model called flexible mixture model (FMM) is defined by

$$f(x_i \mid \Phi) = \sum_{k=1}^{K} \pi_k f_k(x_i \mid \theta_k) \quad \text{with} \quad \sum_{k=1}^{K} \pi_k = 1 \ , \ 0 \le \pi_k \le 1,$$

$$\text{and} \quad \Phi = (\theta_1, \theta_2, \cdots, \theta_K, \pi_1, \pi_2, \cdots, \pi_K),$$

(2.3.1)

where $f_k(\mathrm{x} \mid \theta_k)$ means the kth distribution. When $k=1$, the probability density function $f_{k=1}(\mathrm{x} \mid \theta_{k=1})$ is skewed to right and it can be used to model the background image. When $k=2,\ldots,K$, $f_k(\mathrm{x} \mid \theta_k)$ is the kth normal distribution where parameters are denoted by $\theta_k = (\mu_k, \sigma_k)$ and these mixture normal distribution can be used to model the non-background image. First the indicator function Y_{ik} is defined as

$$Y_{ik} = \begin{cases} 1 & \text{,if } x_i \text{ comes from kth distribution.} \\ 0 & \text{,otherwise.} \end{cases} \tag{2.3.2}$$

The concept using the EM algorithm in FMM is similar to GMM. The main difference is $f_{k=1}(\mathrm{x} \mid \theta_{k=1})$ of the two models. The first let $Y_i = (Y_{i1}, \cdots, Y_{iK})$, then joint probability density function of x_i and y_i is

$$f(x_i, y_i) = \sum_{k=1}^{K} y_{ik} \pi_k f(x_i \mid \theta_k). \tag{2.3.3}$$

The following derivation is similar to (2.2.8)~(2.2.30) for finding $\Phi = (\theta_1, \theta_2, \cdots, \theta_K$ $, \pi_1, \pi_2, \cdots, \pi_K)$ such that $Q(\Phi; \Phi^{(old)})$ is maximized. And because $Q(\Phi; \Phi^{(old)})$ can be divided into two parts (2.2.20) and (2.2.21). First formula (2.2.20) is considered and $\pi_k^{(new)}$ maximize formula (2.2.20) in the result of (2.2.29).

$$\pi_k^{(new)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}. \tag{2.3.4}$$

At another part of formula (2.2.21), $f_{k=1}(\mathrm{x} \mid \theta_{k=1})$ skewing to the right is different from $f_{k=1}(\mathrm{x} \mid \theta_{k=1})$ in section 2.2. Here let $f_{k=1}(\mathrm{x} \mid \theta_{k=1})$ to be exponential distribution first.

$$f_{k=1}(x \mid \theta_{k=1}) = \frac{1}{\beta} \exp(-\beta^{-1} x) \qquad , x > 0, \beta > 0 \tag{2.3.5}$$

Formula (2.2.21) is simplified for obtaining the estimations of θ_k

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_k(x_i \mid \theta_k)) \tag{2.3.6}$$

$$
= \sum_{i=1}^{N} \left(
\begin{array}{l}
\dfrac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\log \beta - \beta^{-1} x_i \right) \\[2em]
+ \dfrac{\pi_2^{(old)} f_2(x_i \mid \theta_2^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_2^2 - \dfrac{(x_i - \mu_2)^2}{2\sigma_2^2} \right) \\[2em]
+ \cdots + \dfrac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log \sigma_K^2 - \dfrac{(x_i - \mu_K)^2}{2\sigma_K^2} \right)
\end{array}
\right). \tag{2.3.7}
$$

The three parameters β, μ_k and σ_k contained in the parameter set θ_k are considered for maximizing of formula (2.3.7). First, we partially differentiate with respect to β,

$$
\text{Let} \quad \sum_{i=1}^{N} \left(\frac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\frac{1}{\beta} + \beta^{-2} x_i \right) \right) = 0 \tag{2.3.8}
$$

$$
\Rightarrow \quad \sum_{i=1}^{N} \left(\frac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(\frac{-\beta + x_i}{\beta^2} \right) \right) = 0 \tag{2.3.9}
$$

$$
\Rightarrow \quad \beta^{(new)} = \frac{\displaystyle\sum_{i=1}^{N} x_i \, \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}. \tag{2.3.10}
$$

In addition, μ_k must be considered for obtaining the parameter set θ_k such that formula (2.3.7) is maximized. Next, we partially differentiate with respect to μ_k,

$$
\text{Let} \quad \sum_{i=1}^{N} \left(\frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(\frac{x_i - \mu_k}{\sigma_k^2} \right) \right) = 0 \tag{2.3.11}
$$

$$
\Rightarrow \quad \sum_{i=1}^{N} x_i \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} - \mu_k \sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} = 0 \tag{2.3.12}
$$

$$\Rightarrow \quad \mu_k^{(new)} = \frac{\displaystyle\sum_{i=1}^{N} x_i \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}} . \tag{2.3.13}$$

Finally, we partially differentiate with respect to μ_k,

$$\text{Let} \quad \sum_{i=1}^{N} \left(\frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\frac{1}{2\sigma_k^2} + \frac{(x_i - \mu_k)^2}{2(\sigma_k^2)^2} \right) \right) = 0 \tag{2.3.14}$$

$$\Rightarrow \quad \sigma_k^{2(new)} = \frac{\displaystyle\sum_{i=1}^{N} (x_i - \mu_k)^2 \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}} . \tag{2.3.15}$$

Partial derivatives of formula (2.3.6) with respect to β, μ_k and σ_k are less than zero. It means

the result $\beta^{(new)}$ of equation (2.3.10), $\mu_k^{(new)}$ of equation (2.3.13) and $\sigma_k^{2(new)}$ of equation (2.3.15)

maximizes formula (2.3.6). So $Q(\Phi; \Phi^{(old)})$ is maximized by the result of equations (2.3.4),

(2.3.10), (2.3.13) and (2.3.15). After getting the estimations, use (2.2.48) as classification rule

to determine how data points should be classified.

$$x_i \in G_m \quad ,where \quad m = \arg\max_k \left(\frac{f_k(x_i \mid \theta_k)\pi_k}{\displaystyle\sum_{j=1}^{K} \pi_j f_j(x_i \mid \theta_j)} \right) . \tag{2.3.16}$$

FMM Clustering Algorithm (Exponential distribution and normal distribution mixture)

Step 1.   Input data $X = \{x_1, x_2, \cdots, x_N\}$

Step 2.   Decide the number of groups K, and give initial value $\Phi^{(old)}$.

Step 3.   Update the estimations by equations (2.3.4), (2.3.10), (2.3.13) and (2.3.15) to obtain

the new estimations $\Phi^{(new)}$.

Step 4.   If $\log L(\Phi^{(new)}) - \log L(\Phi^{(old)}) < tolerance$, stop iterating with newly updated

estimates, else replace $\Phi^{(old)}$ by new estimations and repeat step3.

Step 5.   After iterations finish, use equation (2.3.16) to classify data points.

Gamma distribution with $\alpha = 1$ is an exponential distribution, $f_{k=1}(x \mid \theta_{k=1})$ is replaced by

Gamma distribution. The derivation is similar to the derivation of using exponential

distribution and normal distributions mixture model above.

$$f_{k=1}(x \mid \theta_{k=1}) = x^{\alpha-1} \frac{\exp(-\beta^{-1}x)}{\beta^{\alpha}\Gamma(\alpha)} \qquad , x > 0, \alpha > 0, \beta > 0 \qquad (2.3.17)$$

The estimation $\pi_k^{(new)}$ is obtained by result of equation (2.2.29). That is

$$\pi_k^{(new)} = \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K}\pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} . \qquad (2.2.18)$$

The estimation $\theta_k^{(new)}$ is still obtained by (2.2.21) and it is simplified as follows

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\sum_{j=1}^{K}\pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \log(f_k(x_i \mid \theta_k)) \qquad (2.3.19)$$

$$= \sum_{i=1}^{N} \left(\begin{array}{l} \dfrac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\sum_{j=1}^{K}\pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}\left(-\alpha\log\beta - \log\Gamma(\alpha) + (\alpha-1)\log x_i - \beta^{-1}x_i\right) \\[4ex] + \dfrac{\pi_2^{(old)} f_2(x_i \mid \theta_2^{(old)})}{\sum_{j=1}^{K}\pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}\left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log\sigma_2^2 - \dfrac{(x_i-\mu_2)^2}{2\sigma_2^2}\right) \\[4ex] + \cdots + \dfrac{\pi_K^{(old)} f_K(x_i \mid \theta_K^{(old)})}{\sum_{j=1}^{K}\pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}\left(-\dfrac{1}{2}\log 2\pi - \dfrac{1}{2}\log\sigma_K^2 - \dfrac{(x_i-\mu_K)^2}{2\sigma_K^2}\right) \end{array} \right) . \qquad (2.3.20)$$

Where θ_k contains the parameters α, β, μ_k and σ_k. And the estimations $\beta^{(new)}$, $\mu_k^{(new)}$ and $\sigma_k^{(new)}$ are obtained by (2.3.10), (2.3.13), (2.3.15). Those are

$$\beta^{(new)} = \frac{\displaystyle\sum_{i=1}^{N} x_i \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}} \cdot \tag{2.2.21}$$

$$\mu_k^{(new)} = \frac{\displaystyle\sum_{i=1}^{N} x_i \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}} \cdot \tag{2.2.22}$$

$$\sigma_k^{2(new)} = \frac{\displaystyle\sum_{i=1}^{N} (x_i - \mu_k)^2 \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}}{\displaystyle\sum_{i=1}^{N} \frac{\pi_k^{(old)} f_k(x_i \mid \theta_k^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})}} \cdot \tag{2.2.23}$$

The estimation $\alpha^{(new)}$ is obtained as follows. First we partially differentiate formula (2.3.20) with respect to α.

$$\text{Let } \sum_{i=1}^{N} \left(\frac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\log \beta - \frac{\dfrac{d\Gamma(\alpha)}{d\alpha}}{\Gamma(\alpha)} + \log x_i \right) \right) = 0 \tag{2.3.24}$$

(The $n-$th derivative of the Gamma function is : $\dfrac{d^n}{(dx)^n}\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} (\ln t)^n \, dt$)

$$\Rightarrow \quad \sum_{i=1}^{N} \left(\frac{\pi_1^{(old)} f_1(x_i \mid \theta_1^{(old)})}{\displaystyle\sum_{j=1}^{K} \pi_j^{(old)} f_j(x_i \mid \theta_j^{(old)})} \left(-\log \beta - \frac{\displaystyle\int_0^{\infty} t^{\alpha-1} e^{-t} \ln t \, dt}{\Gamma(\alpha)} + \log x_i \right) \right) = 0 \tag{2.3.25}$$

It is difficult to solve the equation (2.3.25) for obtaining the estimation $\alpha_k^{(new)}$. Therefore, the concept of maximization step in generalized EM (GEM) algorithm and method of moment estimator (MME) are used to find estimation $\alpha_k^{(new)}$. First let the estimation by MME denote by $\hat{\alpha}_{k(MME)}$. The estimation $\alpha_k^{(new)}$ is found to have the maximum value of incomplete space log-likelihood for λ in a range of value and

$$\alpha_k^{(new)} = \alpha_k^{(old)} + \lambda(\hat{\alpha}_{k(MME)} - \alpha_k^{(old)}) . \qquad (2.3.26)$$

In others words, the estimations $\pi_k^{(new)}, \beta^{(new)}, \mu_k^{(new)}, \sigma_k^{(new)}, \alpha_k^{(new)}$ by (2.3.18), (2.3.21), (2.3.22), (2.3.23) and (2.3.26) increase the incomplete space log-likelihood as much as possible. The algorithm is as follows.

FMM Clustering Algorithm (Gamma distribution and normal distribution mixture)

Step 1. Input data $X = \{x_1, x_2, \cdots, x_N\}$

Step 2. Decide the number of groups K, and give initial value $\Phi^{(old)}$.

Step 3. Updated the estimations by equation (2.3.18), (2.3.21), (2.3.22), (2.3.23) and (2.2.26) to obtain the new estimations $\Phi^{(new)}$.

Step 4. If $\log L(\Phi^{(new)}) - \log L(\Phi^{(old)}) < tolerance$, stop iteration with newly updated estimates, else replace $\Phi^{(old)}$ by new estimations and repeat step3.

Step 5. After iterations finish, use equation (2.3.16) to classify data points.

2.4 Kernel Density Estimation (KDE)

The issue of selecting initial values in GMM or FMM is solved by kernel density estimation (KDE) method [7]. The high and low peaks in kernel density curve are used for giving the initial values in KDE, and its bandwidth is determined by the program code of "density" in software R. Number of the high peaks is also used to select cluster size K. Therefore KDE method is considered before classifying the data. The concept of KDE is when data contains different structure from different distributions, the mixture distributions are used. First the kernel density curve is made by data and K high peaks are determined by kernel density. The definition between finding K different structures in data is the same. We give a basic example as follows
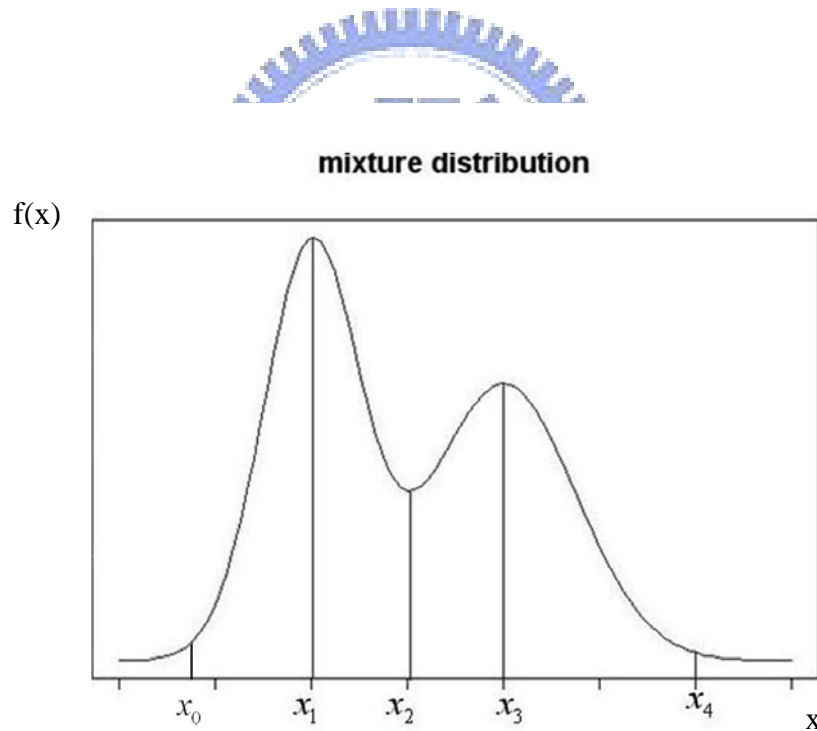
Figure 2.4.1: The mixture distribution is composed of two normal mixture distributions, the mean is μ_1 and the standard deviation is σ_1 in one of normal mixture distributions, mean is μ_2 and the standard deviation is σ_2 in another. It is observed that data contains two structures by kernel density. The data is actually generated from two normal distributions.

The model is normal mixture model, the high peaks can be estimated from the means of two normal distributions. That is $\hat{\mu}_1 = x_1$, $\hat{\mu}_2 = x_3$. And because 99% data of one structure comes

19

from normal distribution is contained in 6-standard deviation. In other words, maximum and minimum of 99% of the data from one structure between 6-standaed deviations are in the lower peaks. So the estimations of standard deviation are obtained by the above concept. That is $\hat{\sigma}_1 = \dfrac{x_1 - x_0}{3}$, $\hat{\sigma}_2 = \dfrac{x_4 - x_3}{3}$. But this method can not be used in FMM. We give a basic example as follows

mixture distribution
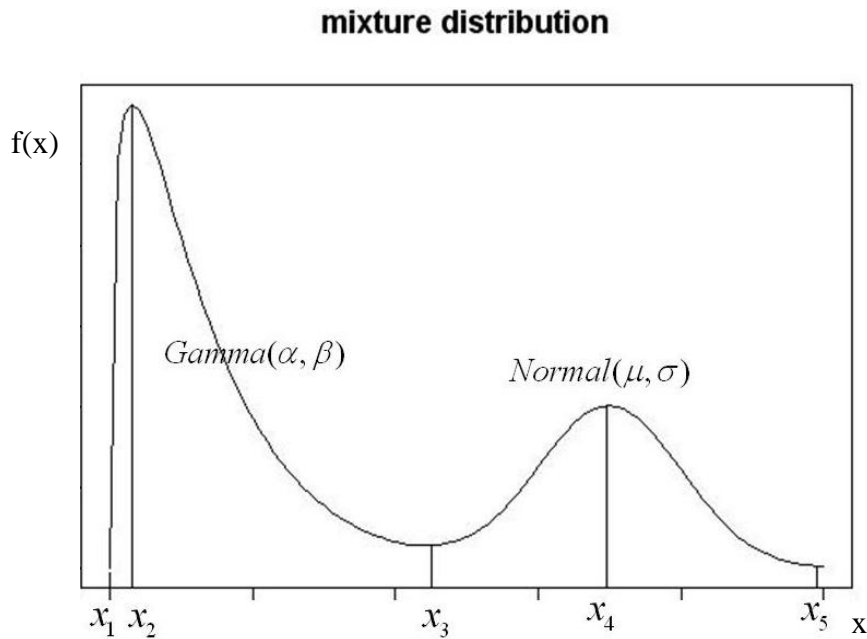
Figure 2.4.2: The mixture distribution is composed of $Gamma(\alpha, \beta)$ and $Normal(\mu, \sigma)$.

We want to estimate α and β of Gamma distribution in Figure 2.4.2. Since most of the same structure is among the low peaks, we first find low peaks from kernel density curve such that data is separated as several structures. Then the low peaks x_1 and x_3 can be found in the structure of $Gamma(\alpha, \beta)$. The initial values can be obtained by method of moment estimator (MME) for α and β from the low peaks x_1 and x_3 can be found in the structure of $Gamma(\alpha, \beta)$. Finally, put the initial values into FMM to classify the data.

2.5 Model Selection

First the number of groups K must be decided before using K-means or GMM or FMM for further research. The choice of the number of groups is very important. The grouping results are always very different when we choose different numbers of groups. Model selection criterion like Akaike information criterion (AIC) or Bayesian information criterion (BIC) is useful to select the cluster size K [7]. They are defined as follows

$$AIC = -2\log(L) + 2p \qquad\qquad (2.5.1)$$

$$BIC = -2\log(L) + p\log(N) \qquad\qquad (2.5.2)$$

p means the number of free parameters. L is maximum likelihood of the model estimated. N means the number of all data. The goal is selecting the number of groups K by finding p such that AIC or BIC minimizes. The value AIC and BIC are the same when the sample size N is 7.389, and the penalty term in BIC is bigger than in AIC when the sample size N is bigger than 7.389. Therefore the BIC is preferred because pixel size in PET images are larger than 7.389. In other words, a simpler model can be selected with smaller numbers of groups by BIC.
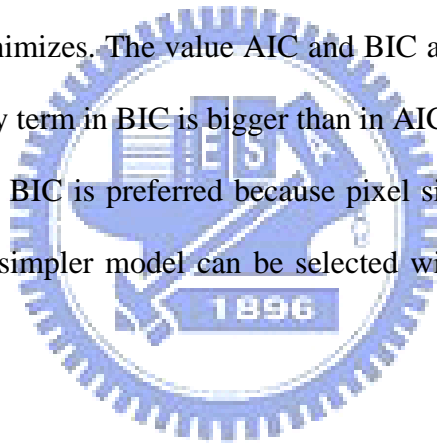
2.6 Procedure

The procedures of K-means with KDE, GMM with KDE and FMM with KDE are shown as follows.
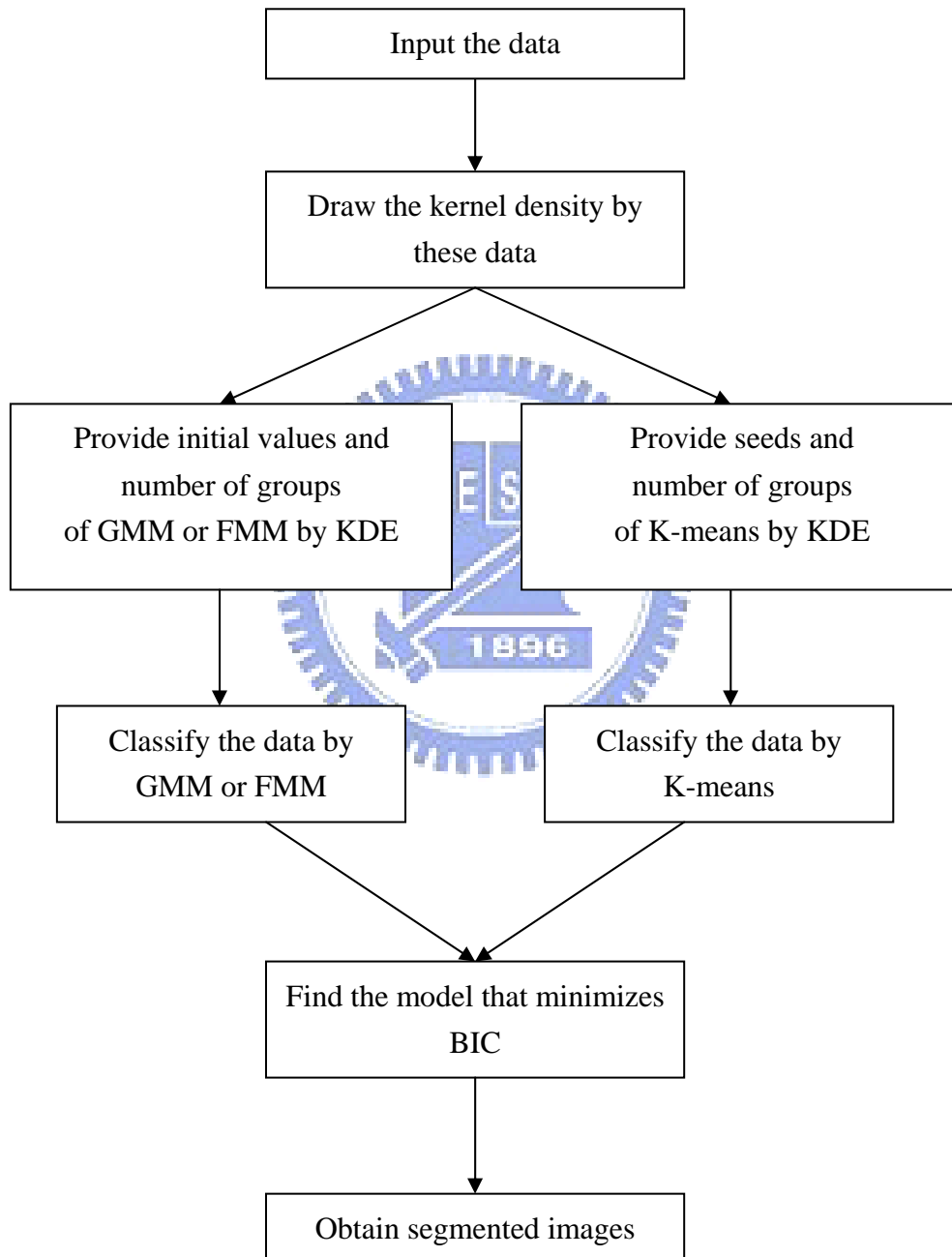
Figure 2.6.1: The procedure for segmentation of data by K-means with KDE, GMM with KDE and FMM with KDE

Chapter 3.   Application

3.1 The PET Scanner

Positron emission tomography (PET) is a non-invasive nuclear medical scanning method. The effect of mutual destruction between positron and electron is used. The gamma ray which is generated by the above effect is measured. The image is composed by computers from the distributions of positron isotope in human organs. Then the metabolic rate per unit is calculated. The medicine ^{18}FDG (^{18}F-labelled deoxyglucose) is used before PET scanning because ^{18}FDG is similar to the glucose which is absorbed by malignant tumor. If there is specific light region in PET scanning, it means abnormal situation may occur at this region. Therefore, the PET scanning helps doctors to determine the abnormal regions in the human body. It can also help doctors assess the effects of tumor treatment. This is one of the medical tools for finding tumor in the human body in recent years.

Furthermore, MicroPET is used to research on small animals like mouse. The main reason is that human disease behavior can often be modeled using small animals. Therapies of human diseases are based on analyzing physiological change. The slice of mouse brain in MicroPET images is shown as follows.

Figure 3.1.1: The 47th slice of the mouse brain MicroPET scanning

3.2 Phantom Study

PET images can be used to find the abnormal regions in organisms as described in section 3.1. However, the region we discovered by clustering methods in PET images can not determine the abnormal structure of organisms. Hence the phantom experiment is designed for comparing the accuracies of these clustering methods. We use the cylindrical container with radius 40mm and height 110mm for the phantom experiment. The width of the clapboard which divided the container into four blocks evenly is 5mm.
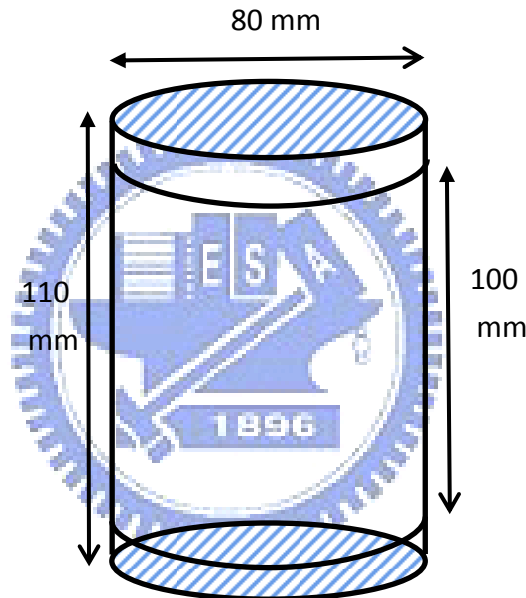

Figure 3.2.1: The geometric information of the cylindrical container

Cylindrical container is subdivided into four compartments, each containing substance of different concentration. The four regions show different brightness in the PET image of the phantom experiment. There are 63 slices where pixel size is 256×256. Therefore the data size is $63 \times 256 \times 256$.
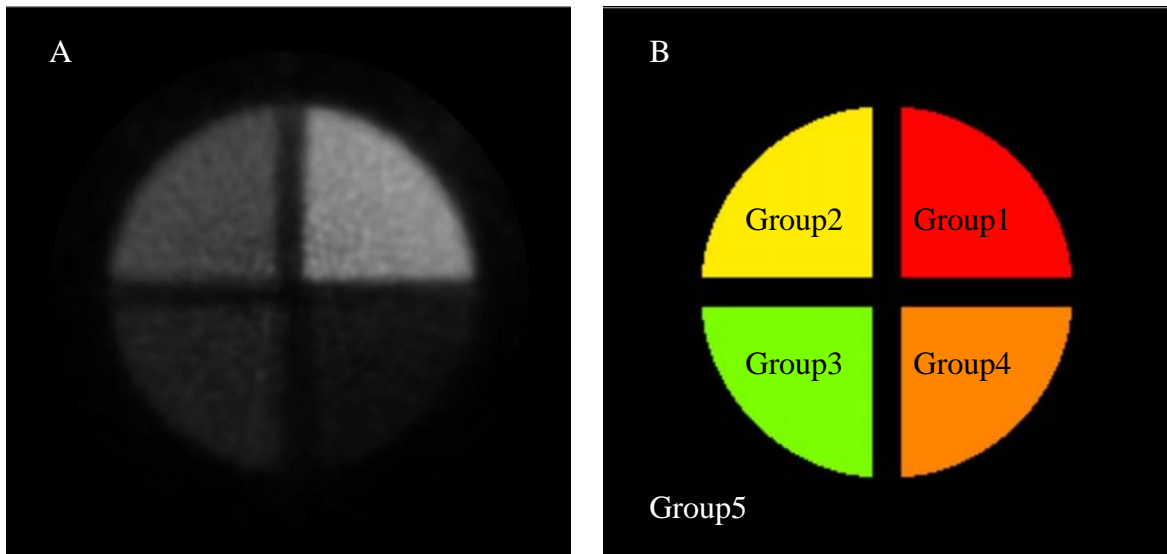
Figure 3.2.2: (A) The 13th slice image of original MicroPET data in the phantom experiment. It is divided into four regions by the clapboard. Individual region shows different brightness. (B) The exact location of four different activities is known by the geometric information of the cylindrical container. The background image region is called Group5.
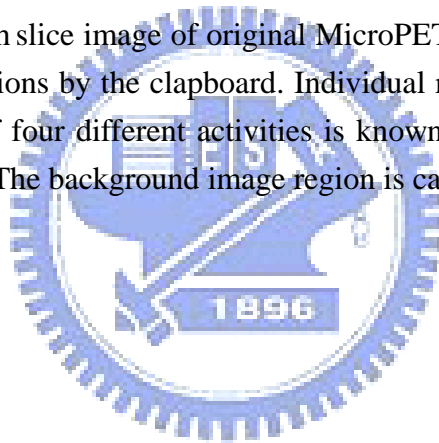
The highest activity region is defined as Group1, and the second-highest activity region is defined as Group2, and third highest activity region is defined as Group3, and fourth highest activity region is defined as Group4. Final the lowest activity region called background image is defined as Group5. We will use K-means with KDE, GMM with KDE and FMM with KDE to classify the original PET image data. The following figure is made by PET image data from the phantom experiment.
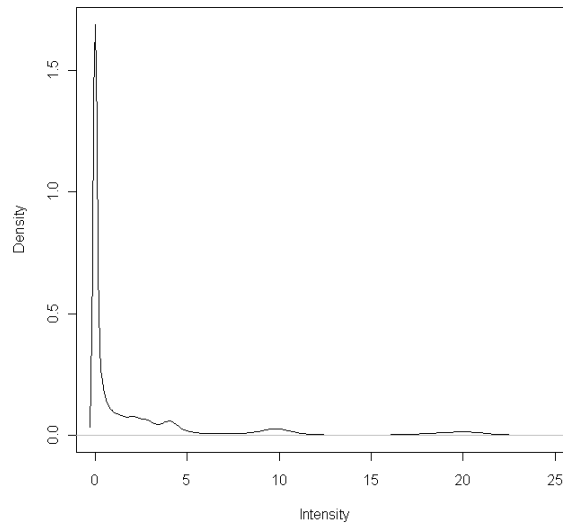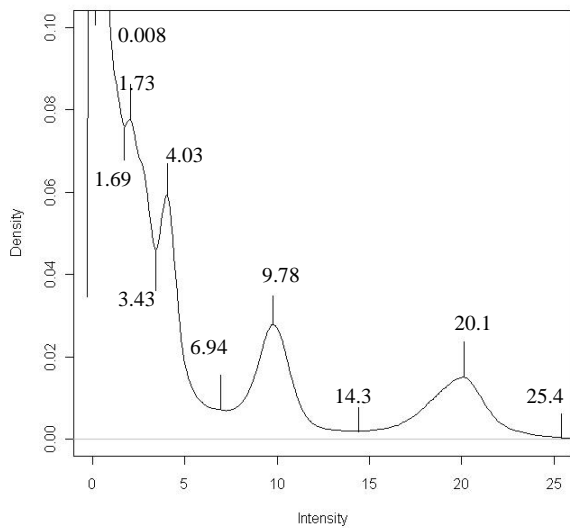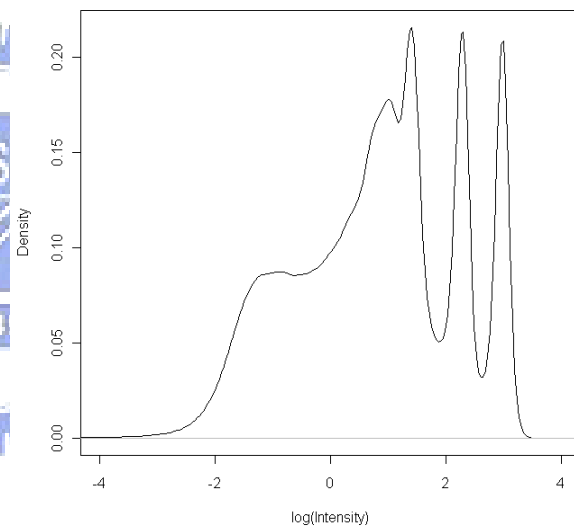
Figure 3.2.3: (A) The kernel density is made by the PET image data from the phantom experiment. (B) It is made by adjusting appropriate scope in x-axis and y-axis. The goal is to conveniently observe the peaks in kernel density and estimating the model parameters.
(C) Take log of data and then graph kernel density for more obvious peaks. These peaks are used because they are more obvious.

After selecting the initial values by KDE, the data is classified by the clustering methods in chapter 2. Since the grouping results are always very different because of differences about the number of groups, BIC is used to decide the numbers of groups.
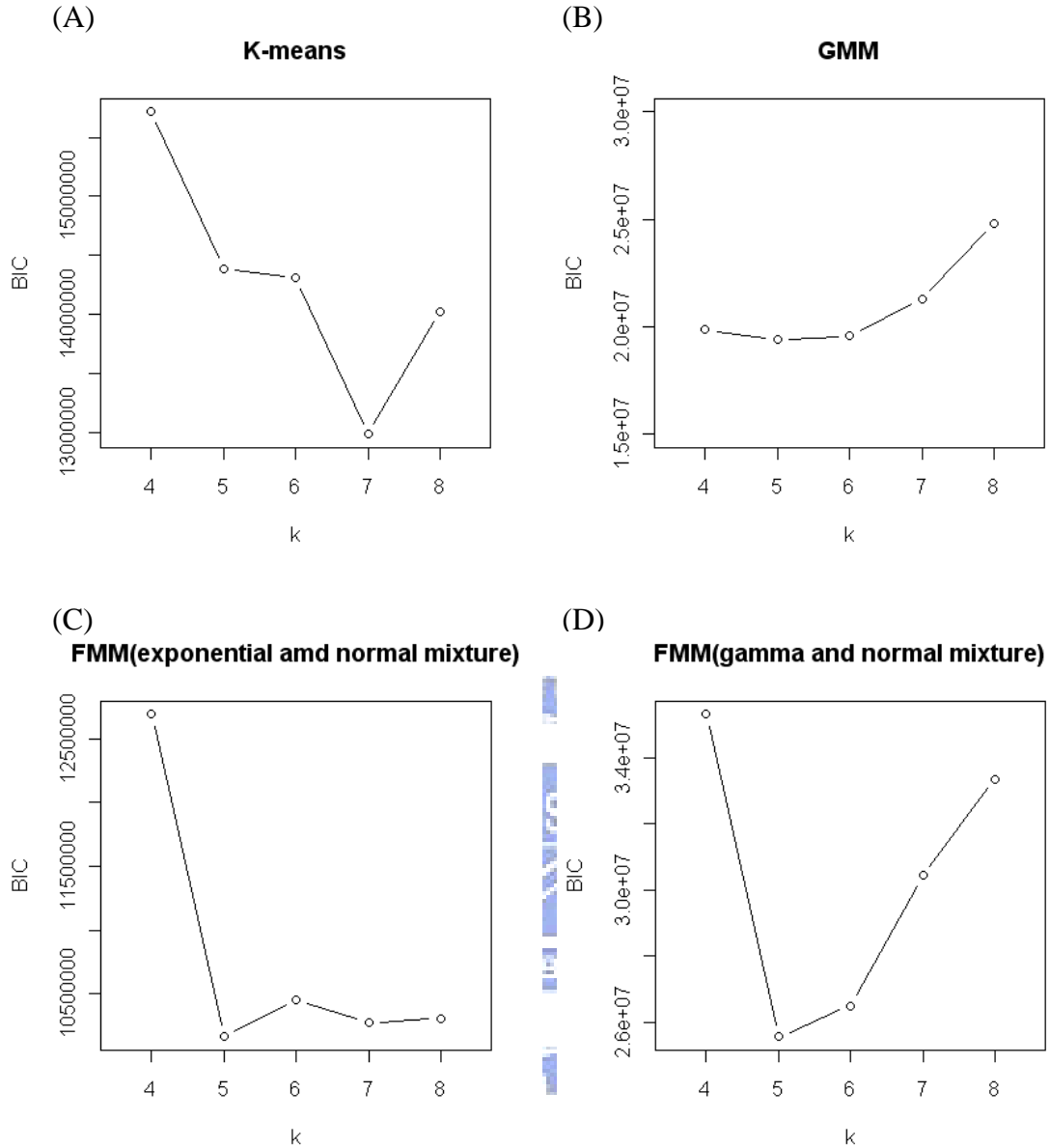
(A) K-means

(B) GMM

(C) FMM(exponential amd normal mixture)

(D) FMM(gamma and normal mixture)

Figure 3.2.4: The BIC values are calculated from different models, k is the number of the groups. (A) K-means + KDE, when k=7 BIC is the minimum. (B)GMM + KDE, when k=5 BIC is the minimum (C) FMM (exponential and normal mixture) + KDE, when k=5 BIC is the minimum. (D) FMM (Gamma and normal mixture)+KDE, when k=5 BIC is the minimum.

It can be decided how to select the number of the groups by the result of Figure 3.2.4. Therefore, the following results are decided in the phantom experiment to classify the PET image data. Select $k = 7$ when K-means with KDE is used, $k = 5$ when GMM with KDE is used, $k = 5$ when FMM (exponential and normal mixture) with KDE is used and $k = 5$

when FMM (Gamma and normal mixture) with KDE is used. The images after different clustering methods show as follows.
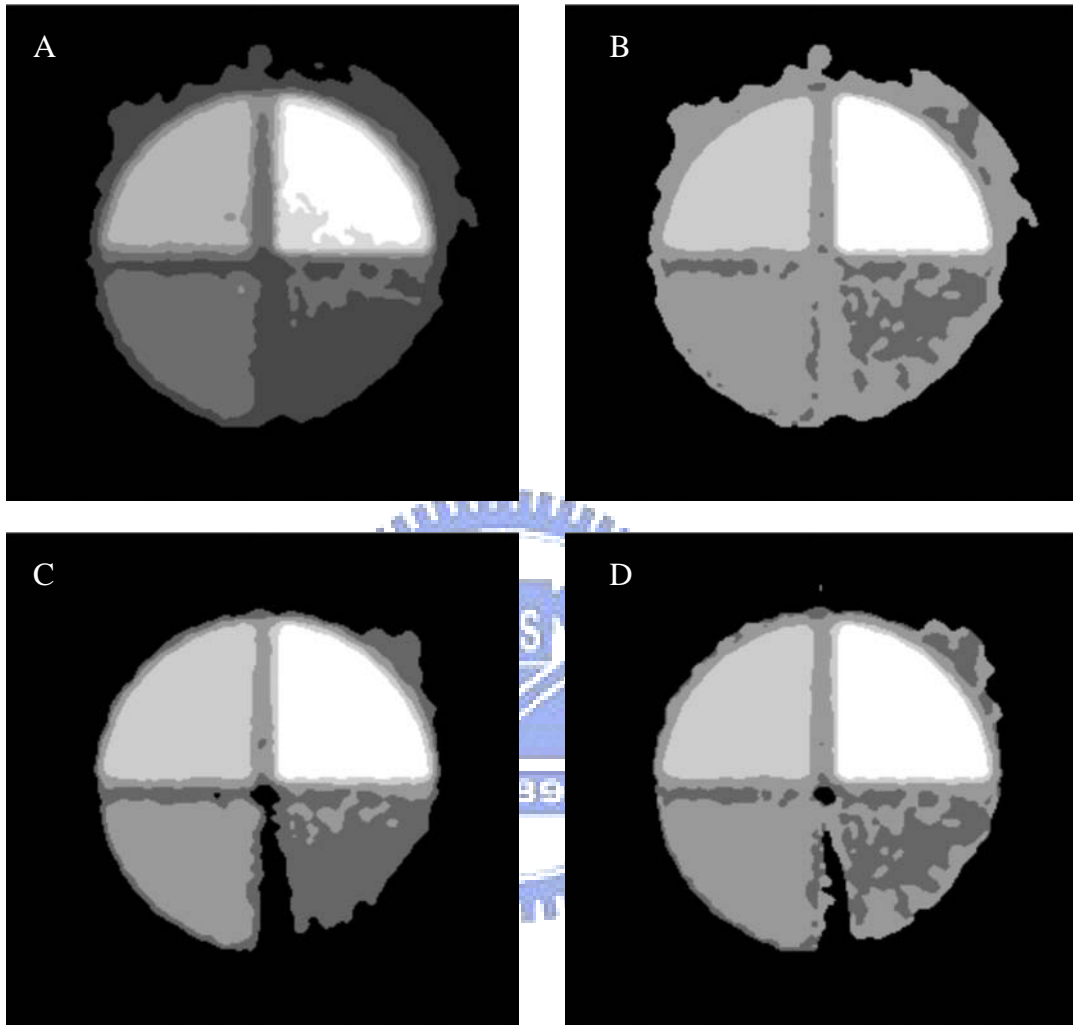

Figure 3.2.5: (A) $k = 7$, the 13th slice image by K-means with KDE (B) $k = 5$, the 13th slice image by GMM with KDE (C) $k = 5$, the 13th slice image by FMM (exponential and normal mixture) with KDE (D) $k = 5$, the 13th slice image by FMM (Gamma and normal mixture) with KDE

Figure 3.2.5 shows the 13th slice images with different clustering methods. It is observed that the clustering regions of Group1 and Group2 are a little different in (B), (C) and (D). However, since we are using different distributions for estimating the dark regions in PET image, the regions in Group5 are very different in (B), (C) and (D). The following table shows

three kinds of accuracy in Group1, Group2 and Group5.

Table 3.2.1: Compare the accuracy with four different methods from the result of Figure 3.2.5. Those methods are individual K-means with KDE, GMM with KDE, FMM(1) with KDE and FMM(2) with KDE. True positive rate shows in (A), negative rate shows in (B) and overall accuracy shows in (C).

Note: FMM(1) means exponential distribution and normal distribution mixture. FMM(2) means Gamma distribution and normal distribution mixture.

(A)

	True positive rate			
	K-means+KDE	GMM+KDE	FMM(1)+KDE	FMM(2)+KDE
Group1	0.580477	0.906258	0.945634	0.946009
Group2	0.791660	0.922194	0.942516	0.930748
Group5	0.830094	0.786272	0.902402	0.882827

(B)

	True negative rate			
	K-means+KDE	GMM+KDE	FMM(1)+KDE	FMM(2)+KDE
Group1	0.999992	0.999549	0.998441	0.998408
Group2	0.990613	0.985185	0.985267	0.986665
Group5	0.985533	0.995245	0.910743	0.948743

(C)

	Overall accuracy			
	K-means+KDE	GMM+KDE	FMM(1)+KDE	FMM(2)+KDE
Group1	0.971257	0.993159	0.994824	0.994819
Group2	0.976985	0.979835	0.982338	0.982835
Group5	0.874153	0.845675	0.904766	0.901510

The accuracy of K-means with KDE is low in Group1, Group2 and Group5. The accuracy of GMM with KDE in Group1 and Group2 is a little better but it is still low in Group5. FMM(1) and FMM(2) are common in using the right-skewed distribution to be the model of background images. Hence the accuracy of FMM(1) or FMM(2) in Group1, Group2 and Group5 is significantly better than K-means or GMM. For Group3, FMM(1) is better than FMM(2). It is because the profile in FMM(1) can be observed clearly from Figure 3.2.5, but that in FMM(2) can not.

3.3 CT Study

In addition to PET images, we would like to apply these clustering methods to the other kind of data. Here we consider computerized tomography (CT) images. CT scanning is also one of the medical diagnosis tools using X-ray. It shows clearly the organization of organisms through a slice of image by X-ray absorption. The region of higher pixel value is lighter at the parts of high absorption, and the region of lower pixel value is darker at the parts of lower absorption. Comparing CT images and PET images, it is observed that it is easier to find the location of abnormal regions in the human body by PET images, and it is easier to outline by CT images. They are shown as the following figures.
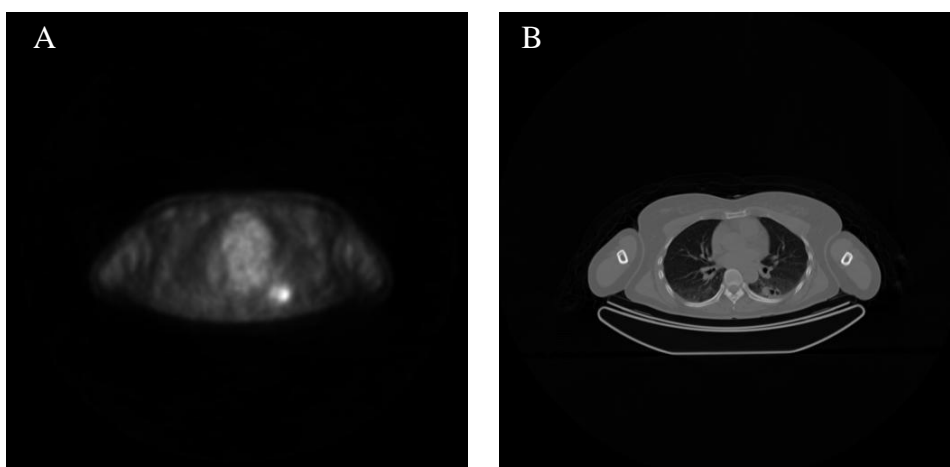
Figure 3.3.1: The 2D CT image is shown and there are 390 slices.
(A) The 161 th slice PET image   (B) The 161 th slice CT image

The profile and organ inside the body is very clearly from the CT image by Figure 3.3.1. Hence the result of segmentation for CT image is compared to original data by this feature. The other CT image whose size is 512×373 is used. Then the kernel density is made by this data for finding initial values.

A

(B)

Kernel density for the CT image(A)

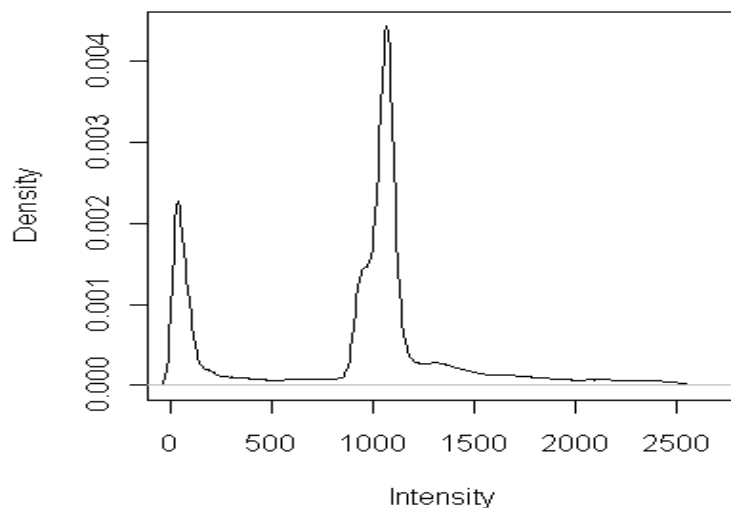Figure 3.3.2: (A) The sagittal plane in the CT image contains teeth, spine, trachea, etc. The data size is 512×373. (B) The kernel density is made by the pixel data in (A).

Next the BIC is used to select the numbers of groups *k*. It is shown the relationship between BIC and the numbers of groups *k* is as follows
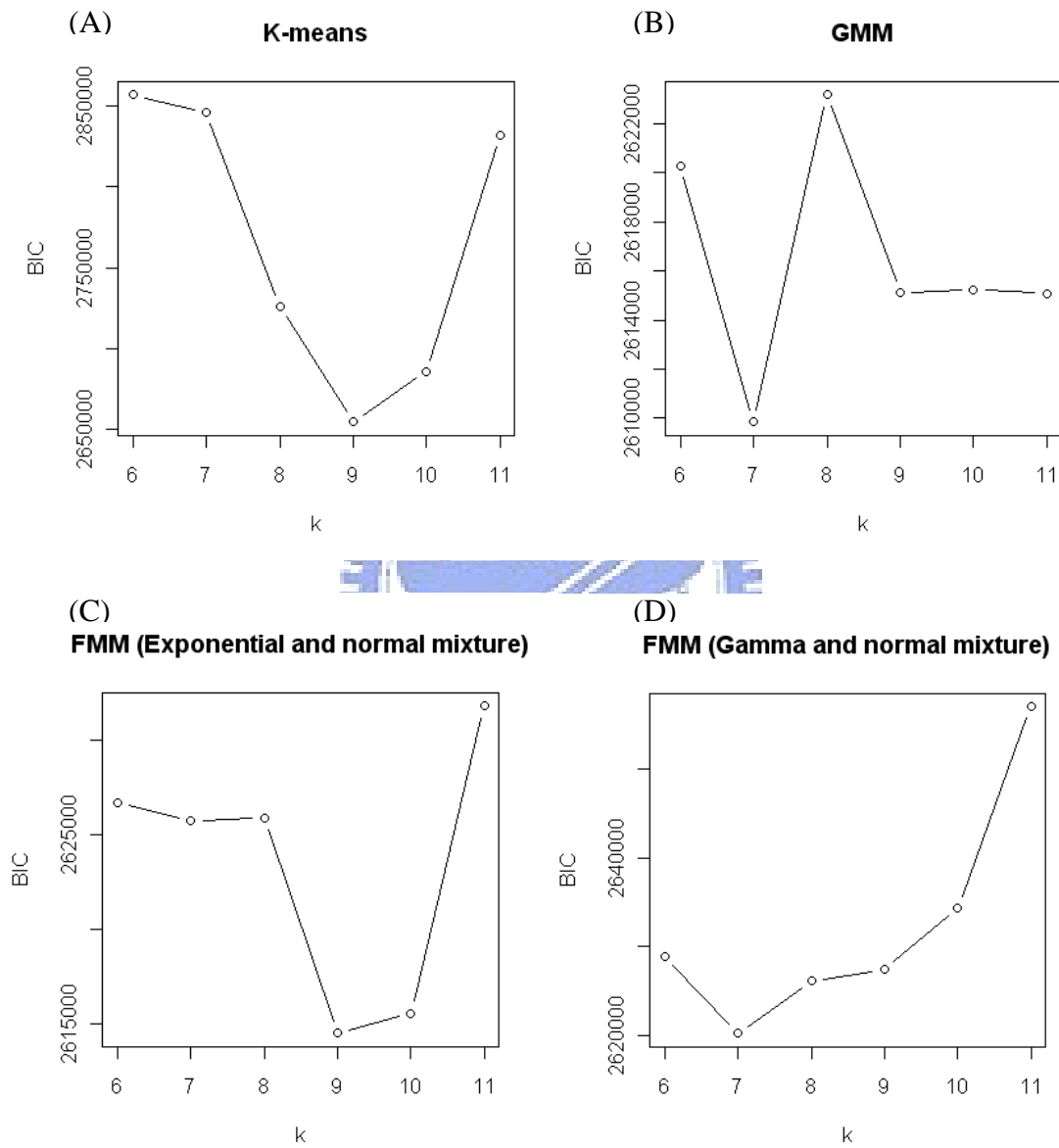

Figure 3.3.3: The BIC values are calculated from different models; k is the number of groups. (A) K-means + KDE, when k=9 BIC is the minimum. (B)GMM + KDE, when k=   7 BIC is the minimum (C) FMM (exponential and normal mixture) + KDE, when k=9 BIC is the minimum. (D) FMM (Gamma and normal mixture) + KDE, when k=7 BIC is the minimum.

It can be decided how to select the number of the groups by using the result of Figure

3.3.3. Therefore, select $k = 9$ when K-means with KDE is used, $k = 7$ when GMM with KDE is used, $k = 9$ when FMM (exponential and normal mixture) with KDE is used and $k = 7$ when FMM (Gamma and normal mixture) with KDE is used. The images after different clustering methods are,
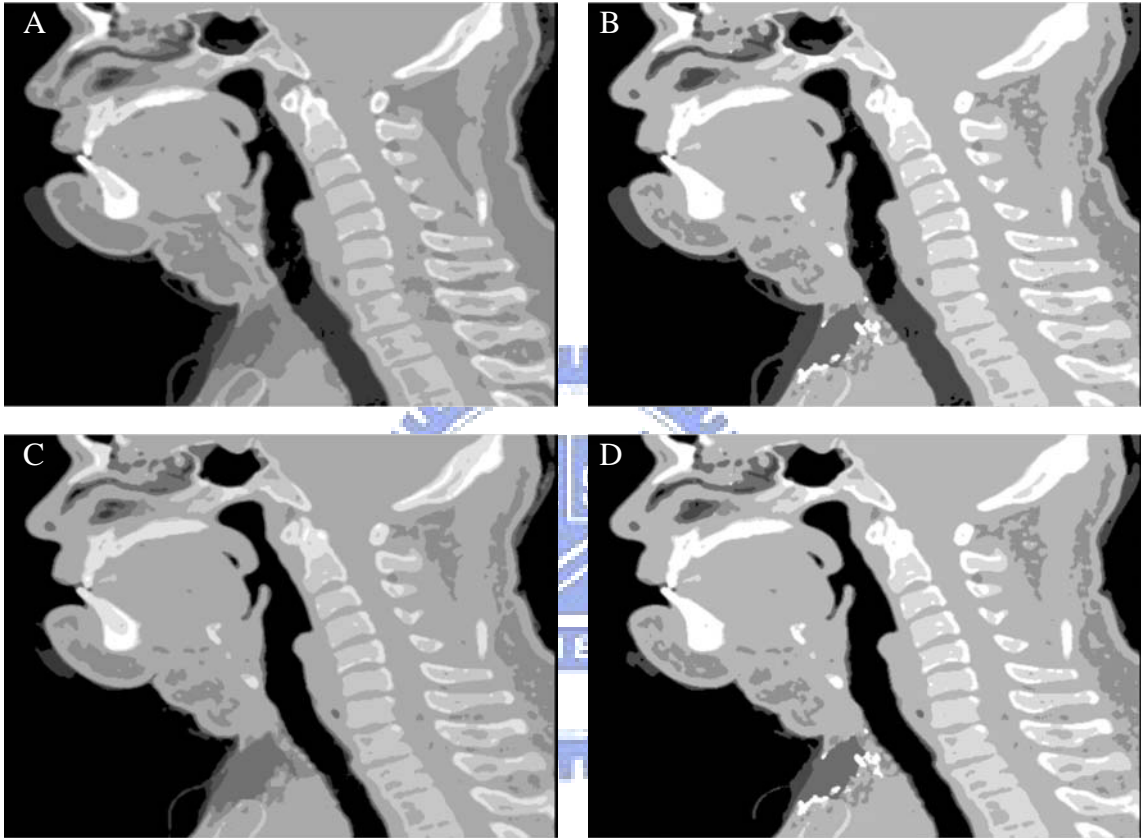

Figure 3.3.4: (A) $k = 9$, the image after using K-means with KDE (B) $k = 7$, the image after using GMM with KDE (C) $k = 9$, the image after using FMM (exponential and normal mixture) with KDE (D) $k = 7$, the image after using FMM (Gamma and normal mixture) with KDE

The trachea in the segmentation of the CT image is clear in Figure 3.3.4 (C) and Figure (D). They both are results of using right-skewed distribution and normal distribution mixture model. That is the FMM method. The K-means result in Figure 3.3.4 (A) is poor in the region of trachea. There are two groups in trachea. GMM also has two groups in trachea, as shown in Figure 3.3.4 (B).

Chapter 4.   Conclusion

The methods of K-means with KDE, GMM with KDE and FMM with KDE are used to classify the data of PET images and CT images. It is observed that the FMM method has high accuracy in the high-activity region in PET images from Chapter 3. The accuracy of low-activity region from FMM is higher than that from K-means or GMM in PET images. The CT image is also clustered by these methods. The result of FMM is better than that of K-means or GMM in the lower pixels of data. After comparing these methods, the data of the PET images or CT images can be classified by FMM because the high accuracy of segmentation obtained by FMM in high and lower activity regions.

The FMM method considers the property and structure of data. It constructs a model to fit the original data more closely. Therefore, the result we obtained from FMM is better than GMM. We used exponential distribution with normal distribution mixture model and Gamma distribution with normal distribution mixture model. For further work, one can test and attempt to find more suitable distributions to increase the accuracy of data classification. Otherwise, the block of pixel values can be used to make pixel data to have more information, like the pixel location. So we can take multivariate normal distribution to model the original data, and compare the results by block method with one pixel point.

References

[1] Akaike H. Anew look at the statistical model identification. IEEE Trans. Automat. Ontr., 1974, AC-19, 716-723

[2] David W. Townsend, Jonathan P.J. Carney, Jerry T. Yap, Nathan C. Hall. PET/CT Today and Tomorrow. *The Journal of Nuclear Medicine*, Vol. 45, pp. 4S-14S, 2004.

[3] Hsiao IT, Rangarajan A, Gindi G. Joint-Map Reconstruction/Segmentation for transmission Tomography Using Mixture-Model as Priors. Proc. IEEE Nuclear Science Symposium and Medical Image Conference, 1998, 3, 1689-1693.

[4] Hung-Yuan Chiang. Segmentation of Dynamic MicroPET Images by K-mean and Mixture Methods. Master's thesis, institute of Statistics National Chiao Tung University, 2004.

[5] McLachlan GJ, Bean RW, Peal D. A mixture model -based approach to the clustering of microarray expression data. *Bioinformatics*, 2000, 18, 3:413-22.

[6] Olinger JM, Fessler JA. Positron-emission tomography. *EEE Signal Processing Magazine*, 1997, 41, 1, 43-55.

[7] Tai Been Chen. Statistical Applications of Maximized Likelihood Estimates with the Expectation-Maximization Algorithm for Reconstruction and Segmentation of MicroPET and Spotted Microarray Images. PhD's thesis, institute of Statistics National Chiao Tung University, 2007.

[8] Vardi Y, Shepp La, Kaufman L. A statistical model for positron emission tomography. *Journal of the American Statistics Association*; 1985, 80, 8-20.