一、高斯分布

（一）简单正态分布

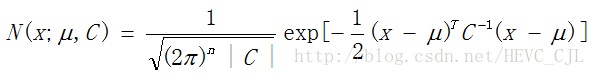
高斯分布就是正态分布，若随机变量X服从一个数学期望为μ、标准方差为σ2的高斯分布，记为：X∼N(μ,σ2),则其概率密度函数为。

注：（高斯分布的特点）

1. 正态分布的期望值决定了其位置，其标准差决定了分布的幅度。
2. 标准正态分布是的正态分布。
3. ,则。
4. 和是独立正态随机变量，则:



## （二）单高斯模型（Single GaussianModel, SGM）

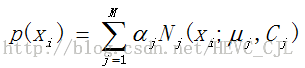
（1）

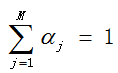
故μ通常由训练样本均值代替，由样本方差代替C。

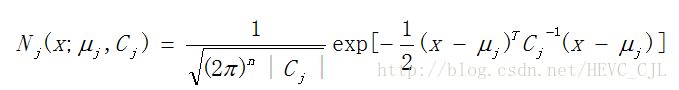
## （三）高斯混合模型（GaussianMixture Model，GMM）

1.介绍：

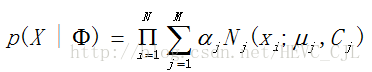
设有一批观察数据，数据个数为n，在d 维空间中的分布不是椭球状，那么就不适合以一个单一的高斯密度函数来描述这些数据点的概率密度函数。此时我们采用一个变通方案，假设每个点均由一个单高斯分布生成，而这一批数据共由M（明确，基本为3到5个）个单高斯模型生成，具体某个数据属于哪个单高斯模型未知，且每个单高斯模型在混合模型中占的比例未知，将所有来自不同分布的数据点混在一起，该分布称为高斯混合分布。

从数学上讲，我们认为这些数据的概率分布密度函数可以通过加权函数表示：



其中表示第j个SGM的PDF。

令，GMM共有M个SGM，现在，我们就需要通过样本集X来估计GMM的所有参数：，样本X的概率公式为：



       通常用EM（Expectation Maximum）算法(该算法可行性大门,稳定性高)对GMM参数进行估计。

**2.**EM（Expectation Maximum）算法对GMM参数估计的**流程：**

（1）初始化

       方案1：协方差矩阵设为单位矩阵，每个模型比例的先验概率；均值设为随机数。

## 方案2：由k均值（k-means）聚类算法对样本进行聚类，利用各类的均值作为，并计算，取各类样本占样本总数的比例。

k-means算法是输入[聚类](http://baike.baidu.com/view/31801.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)个数k，以及包含 n个[数据对象](http://baike.baidu.com/view/178571.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)的数据库，输出满足方差最小标准的k个聚类。同一聚类中的对象相似度较高；而不同聚类中的对象相似度较小。[聚类](http://baike.baidu.com/view/31801.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)相似度是利用各聚类中对象的[均值](http://baike.baidu.com/view/1052684.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)所获得一个“中心对象”（引力中心）来进行计算的。

k-means算法的基本步骤：

（1）从 n个[数据对象](http://baike.baidu.com/view/178571.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)任意选择k个对象作为初始聚类中心；

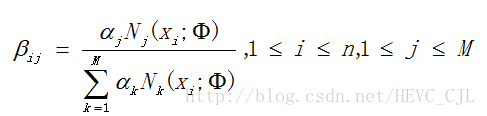
（2）根据每个[聚类](http://baike.baidu.com/view/31801.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)对象的[均值](http://baike.baidu.com/view/1052684.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)（中心对象），计算每个对象与这些中心对象的距离；并根据最小距离重新对相应对象进行划分；

（3）重新计算每个（有变化）[聚类](http://baike.baidu.com/view/31801.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)的[均值](http://baike.baidu.com/view/1052684.htm" \t "http://blog.csdn.net/hevc_cjl/article/details/_blank)（中心对象）；

（4）计算标准测度函数，当满足一定条件，如函数收敛时，则算法终止；如果条件不满足则回到步骤（2）。

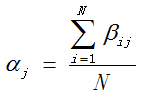
（2）估计步骤（E-step）

令的后验概率为

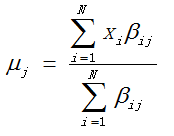


（3）最大化步骤（M-step）

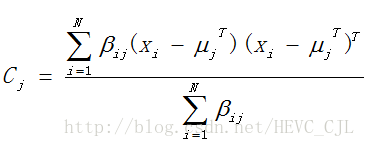
更新权值：



       更新均值：



       更新方差矩阵：



（4）收敛条件

  不断地迭代步骤（2）和（3），重复更新上面三个值，直到，其中为更新参数后计算的值，即前后两次迭代得到的结果变化小于一定程度则终止迭代，通常。

1. 应用

在DBAB中，主要应用在对个体加速度的分析。

1. von. Mises 分布
2. 定义：

冯·米塞斯分布（von Mises distribution）指一种圆上连续概率分布模型，在概率论和定向统计中，有些观点把它认为是缠绕正态分布（wrapped normal distribution）的一种近似，因其是正态分布的循环模拟。冯·米赛斯分布比缠绕正态分布拥有更好的数学可控性，在很多应用中得到青睐。冯·米赛斯分布应用于定向统计中，描绘了来自于相互独立的角度偏差小样本之和的角度分布，样本空间比如有目标感知，或颗粒材料中的颗粒方向。

设x为角度的随机变量，那么在冯·米赛斯分布中，使用复数z=exp(ix)作为分布的随机变量会比实数x更有用处。

角度x的冯·米赛斯分布概率密度函数：



其中是0阶修正贝塞尔函数。

1. 性质：

1. 。

2.参数μ和是μ和（对应正态分布中的均值和方差）的模拟量：μ是位置的度量（分布将围绕μ成簇），κ是集中度的度量（分散度的倒数，所以是的模拟量）。

3.如果κ为0，分布是均匀分布，对于κ很小的情形，分布近似均匀分布。如果κ很大，分布紧紧围绕μ集中分布。

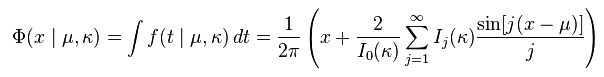
4.实际上，随着κ增加，分布将趋于x以μ为均值为方差的正态分布。

5.它的概率密度函数也可以用一系列贝塞尔函数的级数来表示：



其中是j阶的修正贝塞尔函数。

1. 密度函数的不定积分为：



可写为：

IMG_256

1. 混合von.Mises 模型

1.定义：



其中

为参数。



混合von Mises 模型的分布函数为：

 （1）

1. 估计混合von Mises的参数：

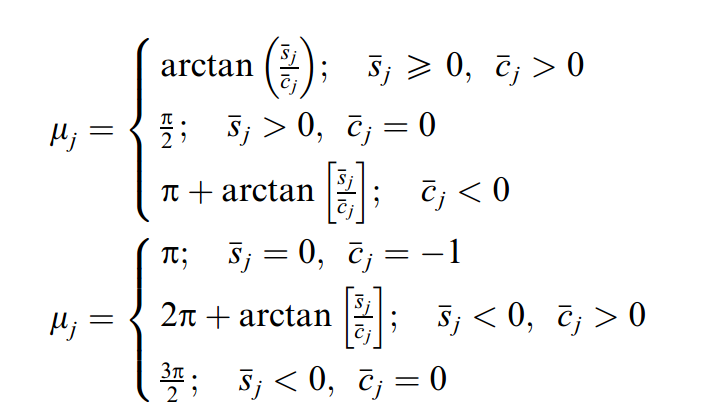
我们利用的是LS（least Squares）算法。

我们需要估计3N个未知参数。

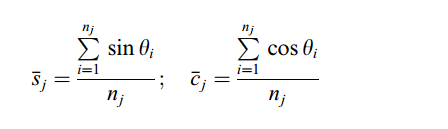
设P是包含从观测样本中获得的试验累计相对频率函数的向量。

若我们将角度分成N部分有thita,发生的相对频率为F

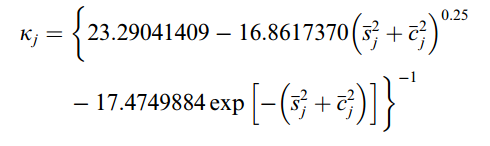
则参数的估计为：



其中：

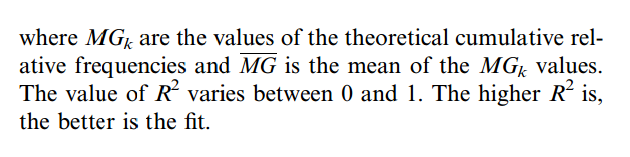


参数k的估计为：



我们可以用以下公式来判断拟合的好坏:





基于EM算法的高斯混合模型参数估计（主要提到了EM算法的用法和例子结果，EM算法和k-means算法连用的好处和理论根据）

<http://wenku.baidu.com/link?url=mmjbpzgx-O63Jn3Y6H2jq3vi0VXA9o-JnQvGxop9xxb5zNRUcvCNqCCDfVXn6Huu46FvGUUWy9bktqsk4h8CM5gMDSEZ6LDvVPSSHBq2_Gy>

PDF：利用高斯混合模型实现概率密度函数逼近

混合von Mises模型的参数估计（主要是讲了用惩罚最大似然估计来估计2个单模型混合的参数）