

**随机过程实验**

**实验题目 泊松过程和高斯混合模型**

**学 号 1180300922**

**姓 名 王志泓**

**指导教师 范晓鹏**

**日 期 2020.11.14**

1. 实验目的

利用计算机模拟生灭过程及泊松过程，对这两种随机过程的特性更加了解。学习GMM-EM算法，能够应用高斯混合模型解决实际问题，例如水果的聚类问题。

1. 实验内容

**问题一：**

生成一个的生灭过程，画出随变化曲线。

**问题二：**

1.生成一个的泊松过程，画出随变化曲线。

2.统计时间内的区间内事件发生一次的概率

3.统计的区间内事件发生次及以上的概率

**问题三：**

下表中有条水果数据，包括密度和含糖率，根据其特征将水果分类，求其高斯混合模型（类别数）。

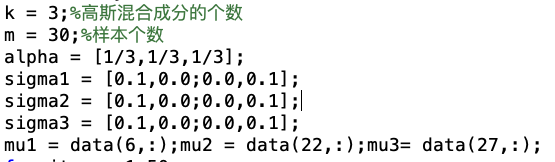


1. 实验过程

**对于问题一**，由于我们知道生灭过程的间隔分布符合指数分布，所以可以利用MATALB中的函数生成符合指数分布的随机数和，其中表示生的时间间隔，表示灭的时间间隔。比较和，如果，则表示状态为生，反之则为灭，然后重新生成随机数，重复以上步骤即可。

**对于问题二**，第一问与问题一类似，生成随机数，且服从指数分布，重复该过程即可。对于第二问及第三问，通过计算个间隔为的区间中发生事件的次数，即可得到事件发生次或者次以上的概率。通过多次实验，事件发生一次的概率趋近于，发生两次以上的概率趋近于。

**对于问题三**，按照实验指导书，首先随机初始化参数如下

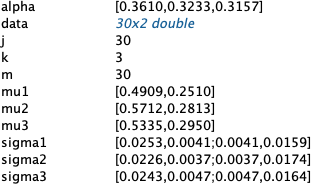


然后利用MATLAB中的函数计算多维高斯分布的概率密度

根据以下公式计算后验概率：

根据以下公式更新模型参数：

得到新参数如下：



与实验指导书一致，说明计算过程没有问题，然后迭代50次即可。

1. 实验结果

**问题一代码如下：**

**﻿**clear;

birth = 0.05;

death = 0.03;

x = [];

x(1) = 0;

t = [];

t(1) = 0;

i = 0;

num = 0;

j = 2;

while i < 1000

t1 = exprnd(1/birth);

t2 = exprnd(1/death);

if t1 < t2

num = num + 1;

i = i + t1;

x(j) = num;

t(j) = i;

else

num = num - 1;

i = i + t2;

x(j) = num;

t(j) = i;

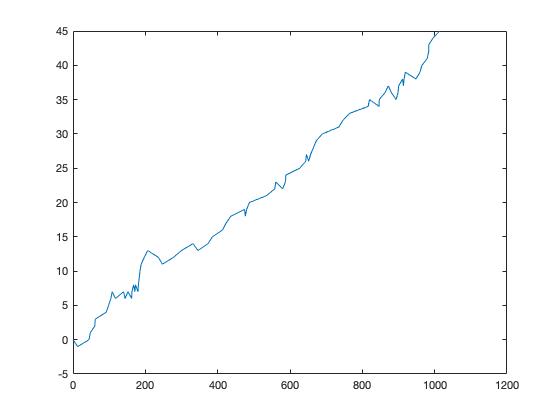
end

j = j + 1;

end

plot(t,x)

**问题一结果如下：**

****

**问题二代码如下：**

clear;

possion = 0.05;

x = [];

x(1) = 0;

t = [];

t(1) = 0;

i = 0;

num = 0;

j = 2;

while i < 10000

t1 = exprnd(1/possion);

num = num + 1;

i = i + t1;

x(j) = num;

t(j) = i;

j = j + 1;

end

event1 = 0;

event2 = 0;

fre = 0;

m = 0;

dt = 2;

total = 10000/dt;

for k = 0:total-1

for l = 1:num-m

if t(l) > dt\*(k+1)

break;

elseif t(l) >= dt\*k && t(l) <= dt\*(k+1)

t(l) = [];

m = m + 1;

fre = fre + 1;

else

t(l) = [];

m = m + 1;

end

end

if fre == 0

continue;

elseif fre == 1

event1 = event1 + 1;

else

event2 = event2 + 1;

end

fre = 0;

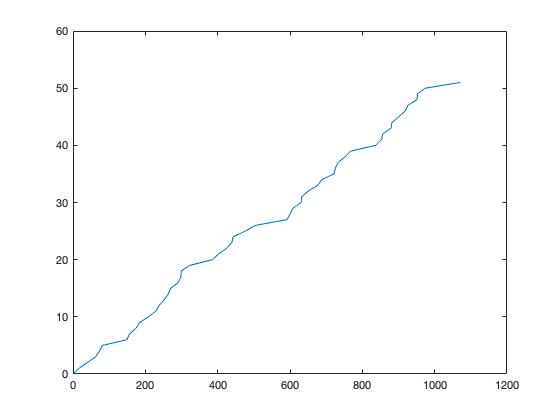
end

p1 = event1/total

p2 = event2/total

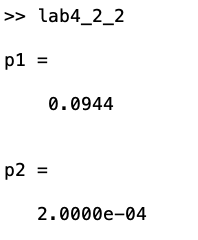
**问题二结果如下：**

**2.1 变化曲线：**

****

**2.2 事件发生一次的概率**

P1 = 0.0944（经过多次实验，概率趋近于0.1）

****

某一次实验的结果

**2.3 事件发生两次以上的概率**

P2 = 2.0000e-04（经过多次实验，概率趋近于0）

**问题三代码如下：**

**﻿**clear;

x = [0.697,0.774,0.634,0.608,0.556,0.403,0.481,0.437,0.666,0.243,0.245,0.343,0.639,0.657,0.360,0.593,0.719,0.359,0.339,0.282,0.748,0.714,0.483,0.478,0.525,0.751,0.532,0.473,0.725,0.446];

y = [0.460,0.376,0.264,0.318,0.215,0.237,0.149,0.211,0.091,0.267,0.057,0.099,0.161,0.198,0.370,0.042,0.103,0.188,0.241,0.257,0.232,0.346,0.312,0.437,0.369,0.489,0.472,0.376,0.445,0.459];

data = [x;y]';

k = 3;

m = 30;

alpha = [1/3,1/3,1/3];

sigma1 = [0.1,0.0;0.0,0.1];

sigma2 = [0.1,0.0;0.0,0.1];

sigma3 = [0.1,0.0;0.0,0.1];

mu1 = data(6,:);mu2 = data(22,:);mu3= data(27,:);

for iter = 1:50

gamma1 = [];gamma2 = [];gamma3 =[];

for j = 1:m

gamma1(j) = alpha(1)\*mvnpdf(data(j,:),mu1,sigma1)/(alpha(1)\*mvnpdf(data(j,:),mu1,sigma1)+alpha(2)\*mvnpdf(data(j,:),mu2,sigma2)+alpha(3)\*mvnpdf(data(j,:),mu3,sigma3));

gamma2(j) = alpha(2)\*mvnpdf(data(j,:),mu2,sigma2)/(alpha(1)\*mvnpdf(data(j,:),mu1,sigma1)+alpha(2)\*mvnpdf(data(j,:),mu2,sigma2)+alpha(3)\*mvnpdf(data(j,:),mu3,sigma3));

gamma3(j) = alpha(3)\*mvnpdf(data(j,:),mu3,sigma3)/(alpha(1)\*mvnpdf(data(j,:),mu1,sigma1)+alpha(2)\*mvnpdf(data(j,:),mu2,sigma2)+alpha(3)\*mvnpdf(data(j,:),mu3,sigma3));

end

total1 = [0,0];total2 = [0,0];total3 = [0,0];

total21 = [0,0;0,0];total22 = [0,0;0,0];total23 = [0,0;0,0];

alpha(1) = sum(gamma1)/m;alpha(2) = sum(gamma2)/m;alpha(3) = sum(gamma3)/m;

for j = 1:m

total1 = total1 + gamma1(j).\*data(j,:);

total2 = total2 + gamma2(j).\*data(j,:);

total3 = total3 + gamma3(j).\*data(j,:);

end

mu1 = total1/sum(gamma1);mu2 = total2/sum(gamma2);mu3 = total3/sum(gamma3);

for j = 1:m

total21 = total21 + gamma1(j).\*((data(j,:)-mu1)'\*(data(j,:)-mu1));

total22 = total22 + gamma2(j).\*((data(j,:)-mu2)'\*(data(j,:)-mu2));

total23 = total23 + gamma3(j).\*((data(j,:)-mu3)'\*(data(j,:)-mu3));

end

sigma1 = total21/sum(gamma1);sigma2 = total22/sum(gamma2);sigma3 = total23/sum(gamma3);

end

groupx1 = [];groupx2 = [];groupx3 = [];

groupy1 = [];groupy2 = [];groupy3 = [];

for n = 1:m

A = [gamma1(n),gamma2(n),gamma3(n)];

if gamma1(n) == max(A)

groupx1 = [groupx1,data(n,1)];

groupy1 = [groupy1,data(n,2)];

elseif gamma2(n) == max(A)

groupx2 = [groupx2,data(n,1)];

groupy2 = [groupy2,data(n,2)];

else

groupx3 = [groupx3,data(n,1)];

groupy3 = [groupy3,data(n,2)];

end

end

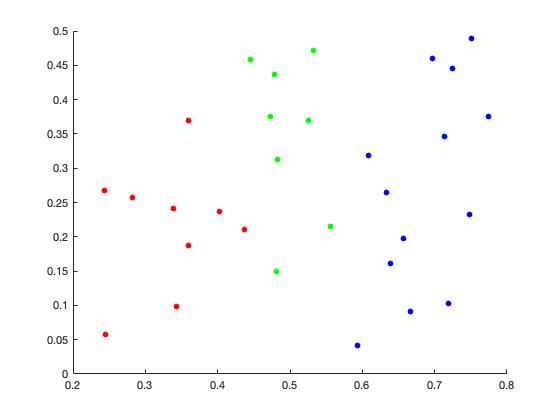
scatter(groupx1,groupy1,'r','filled')

hold on

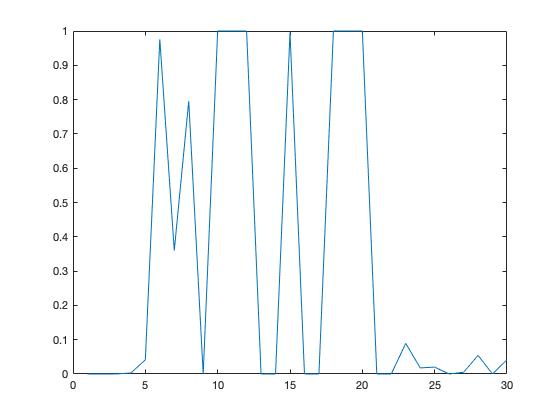
scatter(groupx2,groupy2,'b','filled')

hold on

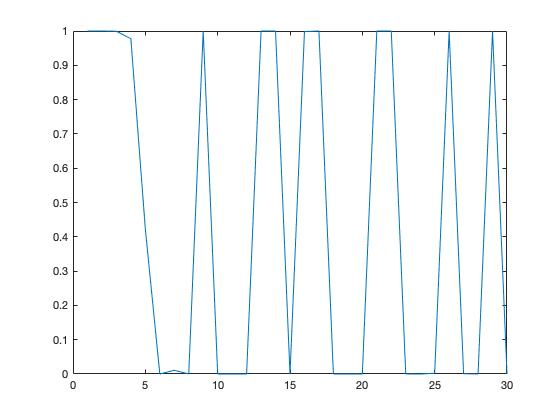
scatter(groupx3,groupy3,'g','filled')

**问题三结果如下：**

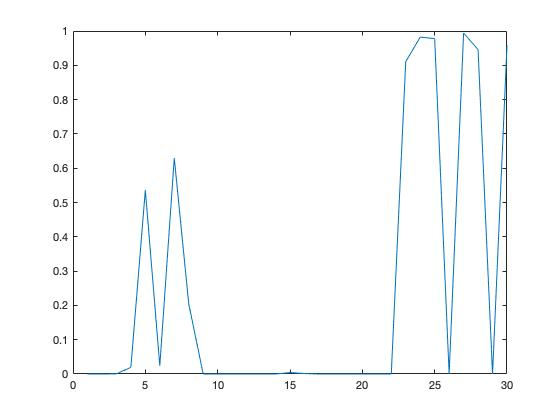
经过50次迭代后的数组内容如下：



图像



图像



图像

1. 心得体会

生灭过程和泊松过称的模拟可以借助其数学性质——间隔分布符合指数分布完成。高斯混合模型对于聚类问题的解决非常有效，而GMM-EM算法可以有效的解决高斯混合分布的求解。