模式识别导论上机题1-参数估计

薛犇 1500012752

1. 程序实现说明

本次实验采用Matlab作为编程语言，使用的版本为2016b。

在实验的一开始，利用importdata函数读取A.txt中的数据，保存在向量x中。

```
x = importdata('A.txt');
```

（1）

由作业题中的第二问可以得知：

θ的最大似然估计为：

$$\frac{\sum_{i=1}^{n} lnx_i}{n}$$

σ的最大似然估计为：

$$\sqrt{\frac{\sum_{i=1}^{n}(lnx_i-\theta)^2}{n}}$$

在我的matlab代码中，第5、6行计算了这两个的值，具体代码如下：

```
theta = sum(log(x))/n
sigma = sqrt(sum((log(x)-theta).^2)/n)
```

之后利用题目中给定的概率分布

$$p(x)=\frac{1}{\sigma x\sqrt{2\pi}}exp(-\frac{(lnx-\theta)^2}{2\sigma^2})$$

将计算得到的θ和σ的值代入，得到一个概率p与x的函数（代码第29-31行）

```
function y = p(x, theta, sigma)
    y = exp(-(log(x)-theta).^2/(2*sigma^2)) ./ (sigma*x*(2*pi)^0.5);
end
```

利用这个函数，在已有数据x的range范围之内取一个线性散列x_0，然后计算这个x_0对应的概率分布y_0。

```
x_0 = linspace(min(x), max(x));
y_0 = p(x_0, theta, sigma);
```

之后再plot出来

```
axis([min(x) max(x) 0 0.00005])
plot(x_0, y_0, 'r');
```

（2）

假定分布为正态分布，那么概率分布为：

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

求得μ的最大似然估计为：

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

σ的最大似然估计为：

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$$

同样的方式，求出相应的值（代码第22、23行）：

```
miu = mean(x);
sigma2 = sqrt(sum((x-miu).^2)/n);
```

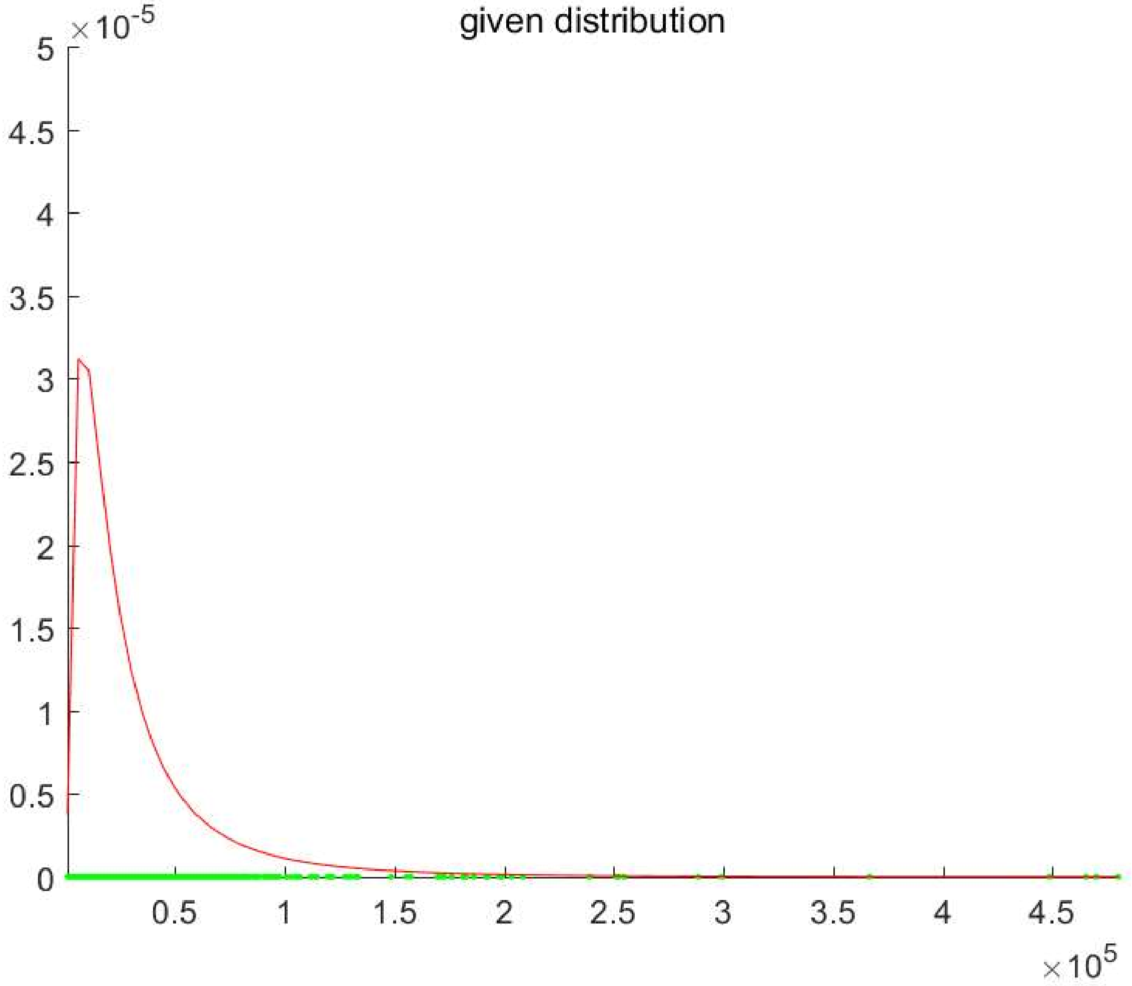
接着实现正态分布相应的概率分布函数（代码第33-35行）：

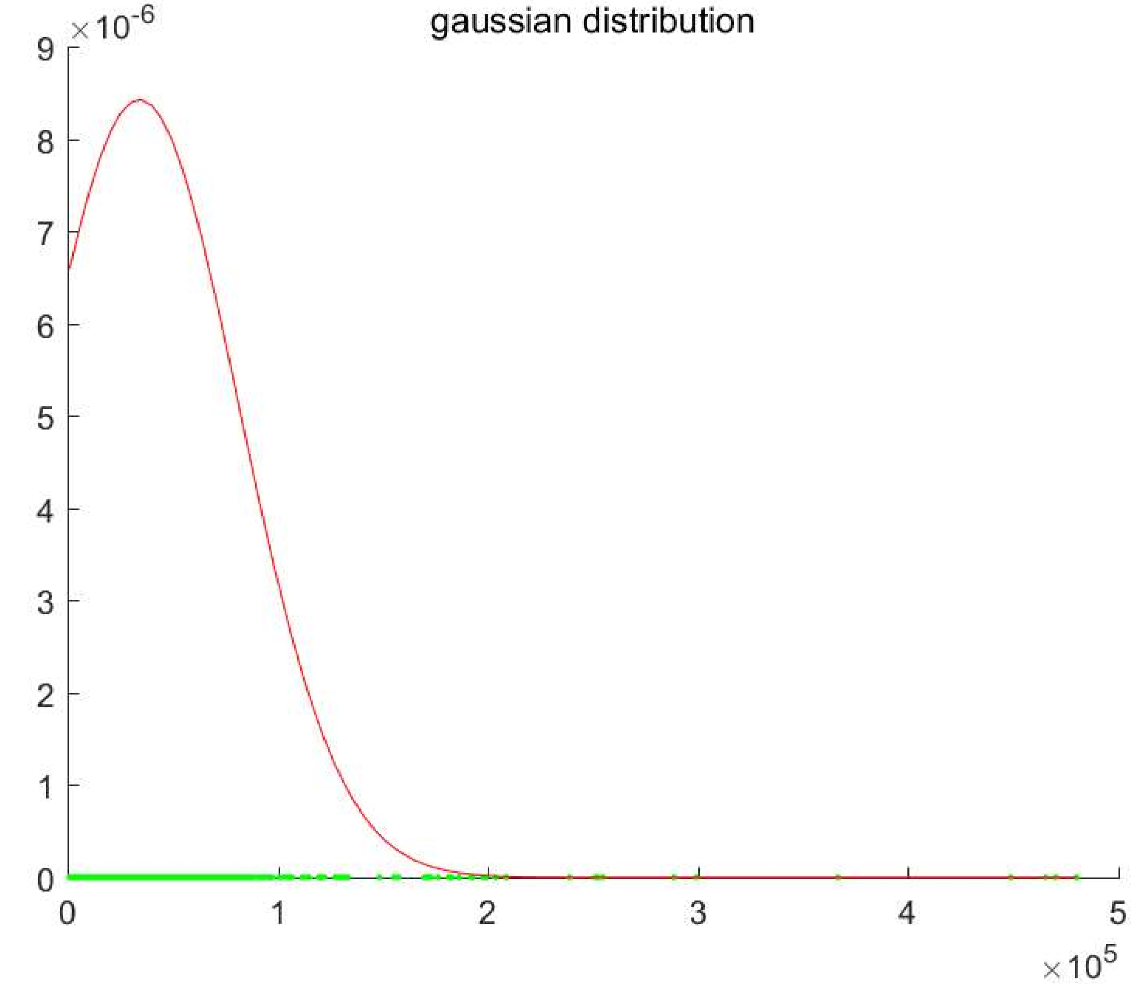
```
function y = p_gaussian(x, miu, sigma)
    y = exp(-(x-miu).^2 / (2*sigma^2)) ./ (sigma*(2*pi)^0.5);
end
```

与第（1）题类似的方法，plot出概率分布图（第25-27行）：

```
x_1 = linspace(min(x), max(x));
y_1 = p_gaussian(x_1, miu, sigma2);
plot(x_1, y_1, 'r')
```

2. 实验结果


given distribution
×10^-5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
×10^5


gaussian distribution
×10^-6
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
×10^5

3. 实验结果分析

本次实验预先将原始数据点plot在了x轴上，可以看到是绿色的点群。

在概率分布较高的地方，可以明显的看到x轴上绿色的点也更加密集，说明实验的结果是合理的。

对比两张概率分布图，可以说基本相似，但是正态分布更加圆滑。但与此同时，正态分布也丢失了原始分布的一些信息。