

Q1. 习题 4 第 16 题, 请说明正确或不正确的理由并改正错误。

从数据文件 prob\_data416.mat 得到随机数组 R, 下面有一段求取随机数组全部数据最大值、均值和标准差的程序。Mx=max(max(R)), Me=mean(mean(R)), St=std(std(R)), 试问该程序所得的结果都正确吗? 假如不正确, 请写出正确的程序。(提示: load; R(:).)

答: (1) 首先提供本题代码的模拟过程与输出结果

```
load prob_data416
Mx=max(max(R))
Me=mean(mean(R))
S=std(std(R))
Mx = 0.9997
Me = 0.5052
S = 0.0142
```

(2) 按照题意的理解, 所有元素的最大值、均值和标准差, 应当先拉成向量再计算, 即

```
Mx1=max(R(:))
Me1=mean(R(:))
S1=std(R(:))
Mx1 = 0.9997
Me1 = 0.5052
S1 = 0.2919
```

(3) 结果分析: 两层调用max、min或std函数, 含义为先对每一列计算最大值、均值与标准差, 再对结果行向量分别计算一次最大值、均值与标准差。按列选最大, 再选最大与所有元素的最大值是等价的。均值同理。但矩阵所有元素的标准差, 和列标准差再取标准差显然是不同的数学概念。如 $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ , 其两列的样本标准差显然均为 $\frac{\sqrt{2}}{2}$ , 但行向量 $[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$ 的标准差就等于0。另一方面, A所有元素并不完全相同, 样本标准差必然大于0, 事实上这个标准差等于 $\frac{\sqrt{3}}{3} \neq 0$ 。

Q2. 假设 $1000 \times 1000$ 的随机数值矩阵A各元素总体服从均匀分布 $U(0, \delta)$ 。每个元素可以视为一个样本。[2016级期末]

(1)根据A的数据信息, 参数 $\delta$ 的一阶矩估计是什么, 对应MATLAB代码是什么?

(2)参数 $\delta$ 的最大似然估计是什么, 对应MATLAB代码是什么?

答: (1)一阶矩估计, 即使用样本的一阶矩 $\bar{X}$ 作为总体期望 $E(X)$ 来进行参数估计, 因为均匀分布的总体期望 $E(X) = \frac{\delta}{2}$ , 令 $\bar{X} = \frac{\delta}{2}$ , 有 $\hat{\delta} = 2\bar{X}$

对应MATLAB代码为: `delta_1 = 2 * mean(A(:));` %注意: 按列取平均是不对的

(2)根据总体X的概率密度函数 $f(x) = \begin{cases} \frac{1}{\delta}, & 0 < x < \delta \\ 0, & \text{其他} \end{cases}$ , 因此可以定义似然函数:

$$L(X_1, X_2, \dots, X_{1 \times 10^6}, \delta) = \begin{cases} \delta^{-1 \times 10^6}, & 0 < X_1, X_2, \dots, X_{1 \times 10^6} < \delta \\ 0, & \text{其他} \end{cases}$$

注意到似然函数随 $\delta$ 单调递减。因此 $\delta$ 应尽量小，但需确保 $\delta > \max\{X_1, X_2, \dots, X_{1 \times 10^6}\}$ ，因

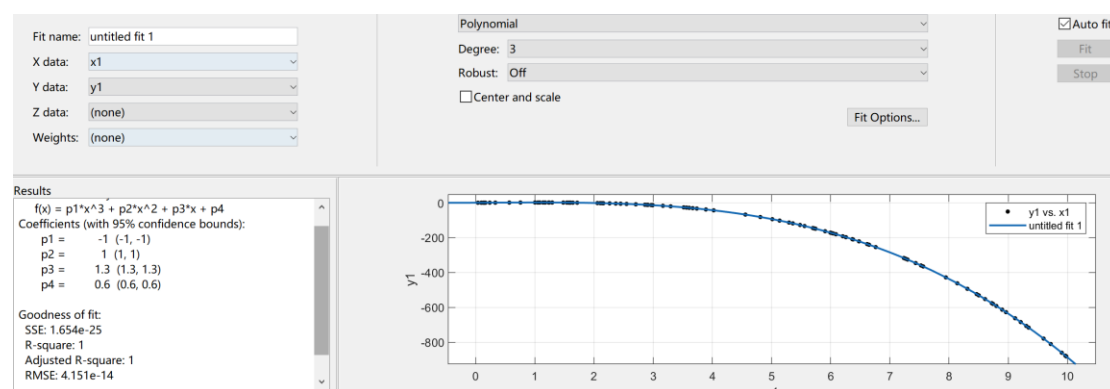
此我们可取其最大似然估计为 $\hat{\delta}_{MLE} = \max\{X_1, X_2, \dots, X_{1 \times 10^6}\}$

对应MATLAB代码为：`delta_2 = max(A(:));` %注意：按列取最大是不对的

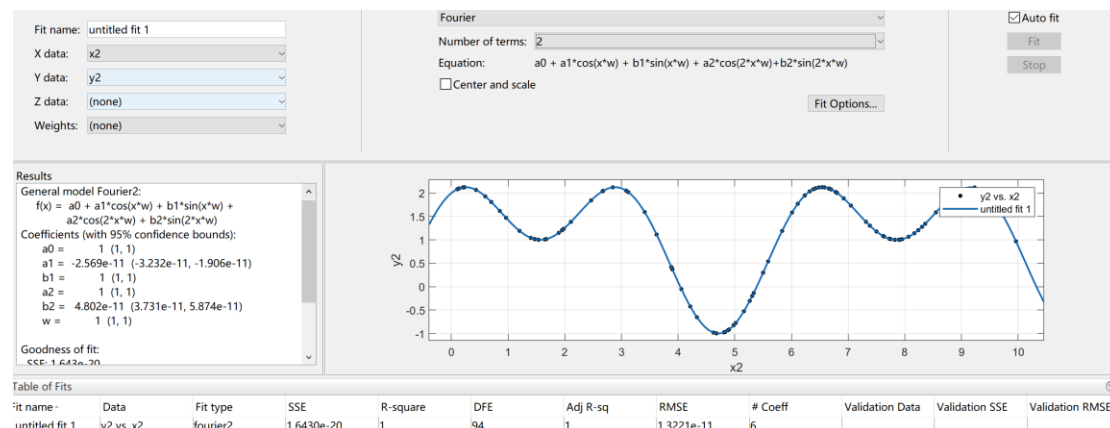
Q3.读取 W13Q3.mat 文件，里面包含了  $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$  共四组函数采样值，它们分别对应了四种不同的简单初等函数的定义（ $y_1, y_2, y_3$  为真实值， $y_4$  有冲击噪声），尝试通过今天学的 `cftool` 工具，分析和预测四个简单初等函数的具体定义。（简述你的拟合方法与结果）

答：对于前三组数据，因为老师有这些函数的真实值，所以仅提供真实函数与拟合结果图

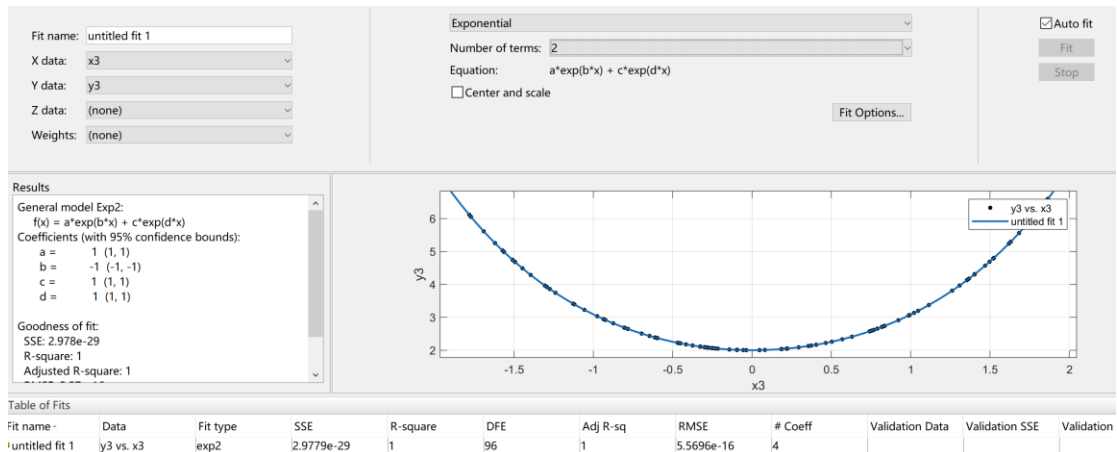
$$y_1(x) = -x^3 + x^2 + 1.3x - 0.6, 0 \leq x \leq 10, \text{3次多项式}$$



$$y_2(x) = 1 + \cos 2x + \sin x, 0 \leq x \leq 10, \text{三角函数}$$



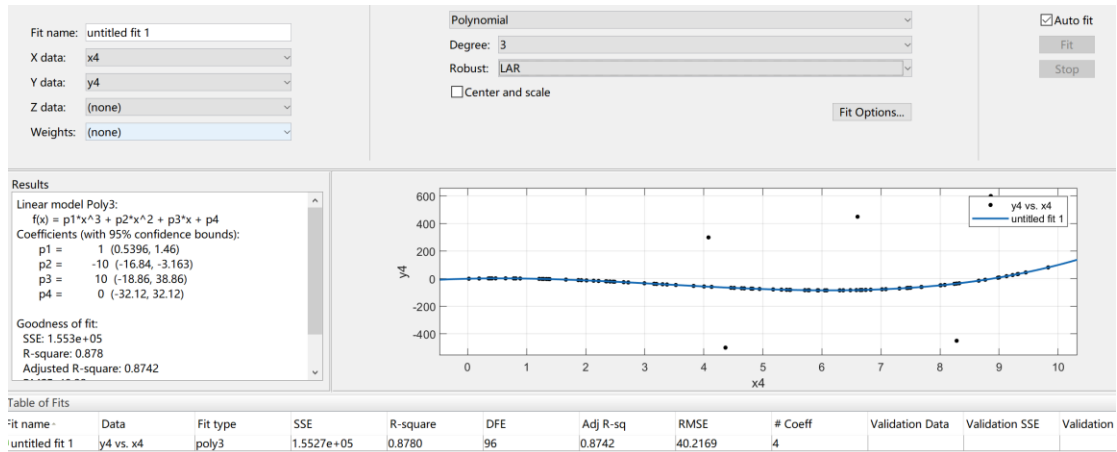
$$y_3(x) = e^x + e^{-x} = 2 \cosh x, -2 \leq x \leq 2, \text{指数型 (悬链线)}$$



对于最后一组拟合数据，因冲击噪声影响，在中间有5个离群值点，函数定义：

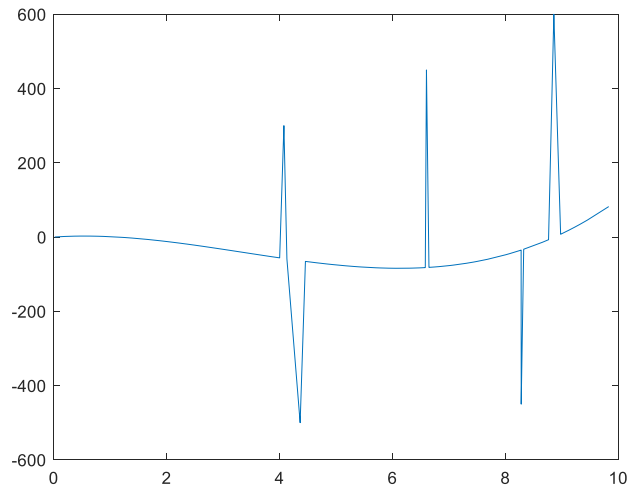
$$y_4(x) = x^3 - 10x^2 + 10x + E, 0 \leq x \leq 10,$$

其中， $E$ 为离群值，此时，使用 LAR（最小绝对值残差）拟合即可得到较好的拟合结果如图：



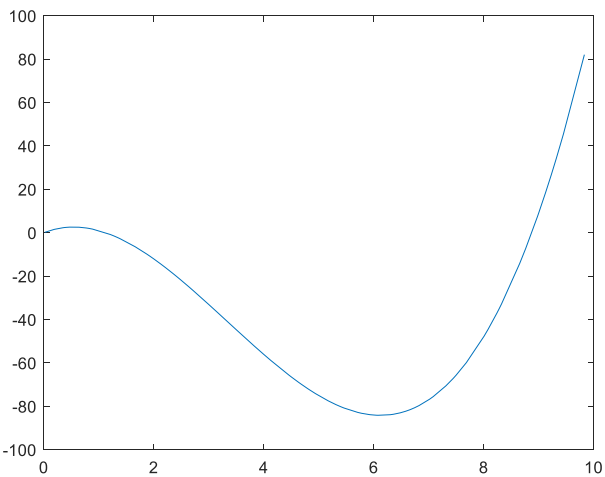
下面再提供“抠点”法的介绍。首先对此函数的自变量进行排序处理后绘图显示：

```
clear
load W13Q3.mat
[x4s,ind] = sort(x4);
y4s = y4(ind);
plot(x4s,y4s);
```



一共 5 个受到冲击噪声影响的点清晰可见，且对应的函数值已经完全不具有任何分析和讨论的价值。在各种方法中，推荐使用人工提取并去除冲击噪点的方法，这里推荐一种较为准确的冲击噪点提取方法。

```
Y4sf = medfilt1(y4s,5); %五点中值滤波
med_error = abs(y4s-y4sf); %计算滤波前后误差
 [~,diffind] = sort(med_error); %误差排序
impulse_ind = diffind(end-4:end); %最大五点误差位置
%注意：有些实际问题为了保证健壮性，冲击噪点的筛选往往通过阈值处理而不是排序取
%最大中值滤波误差的若干个采样点
X4s(impulse_ind) = [];
Y4s(impulse_ind) = []; %删除对应采样点的信息
figure, plot(x4s,y4s); %重新绘图后，曲线已平滑
```



随后的 **cftool** 工具只需采用**三次多项式拟合**即可获得正确的拟合结果：

