

# 模式识别实验三

实验人：叶平

实验内容：

## P409, Prob. 2

1、通过仿真，举例说明偏差-方差分解以及回归中的偏差-方差两难问题。假设目标函数  $F(x)=x^2$ ，高斯噪声的方差是 0.1。首先通过选择在  $-1 \leq x \leq 1$  上均匀分布的  $x$  的值，并将在  $F(x)$  上附加噪声，任意产生 100 个数据集，每个集的大小  $n=10$ 。训练(a)~(d)中的每个回归函数的任意一个自由参数  $a_i$ （用最小平方差准则），每次只训练一组数据。作出式(11)中的直方图。对每个模型，利用你得到的结果去估计偏差和方差。

(a)  $g(x)=0.5$

(b)  $g(x)=1.0$

(c)  $g(x)=a_0+a_1*x$

(d)  $g(x)=a_0+a_1*x+a_2*x^2+a_3*x^3$

2、对 100 个数据集，其大小  $n=100$  重复问题(a)~(d)。

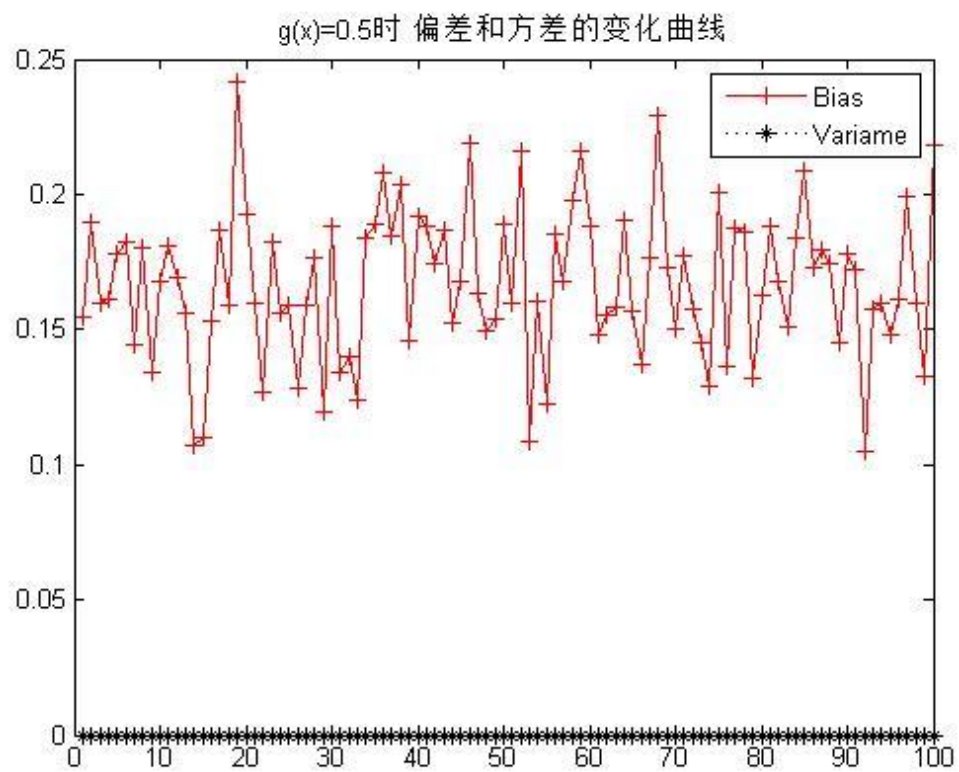
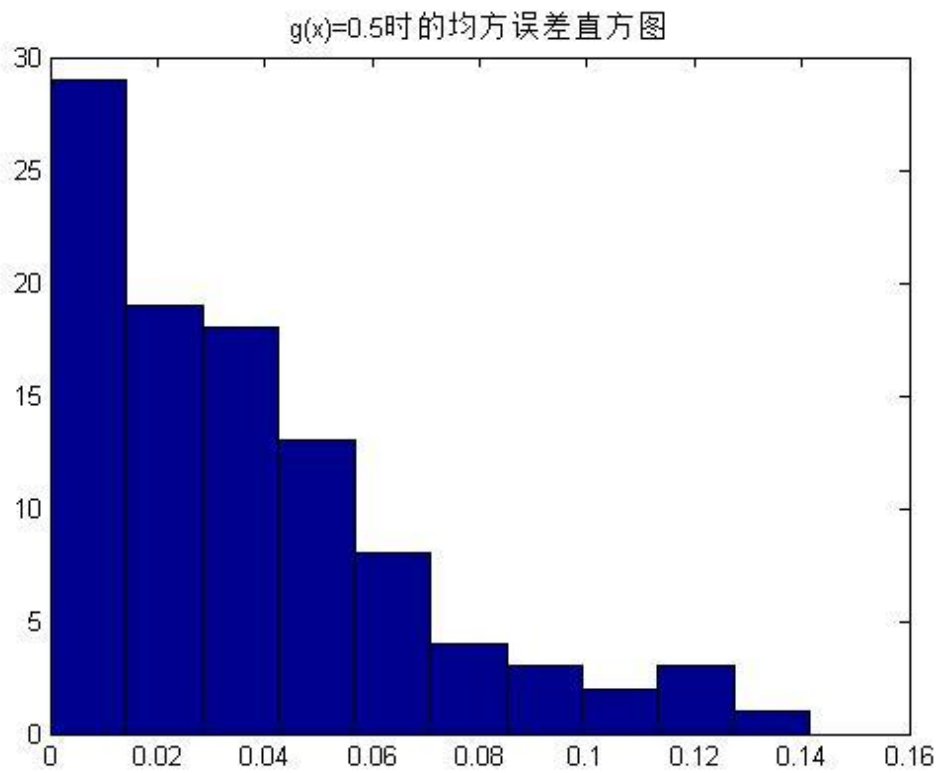
实验结果：

在 Command Window 下运行试验三文件夹下 Res.m 文件，可以得到实验结果。Res(10)表示， $n$  取 10。Res(100)表示， $n$  取 100。如下图：

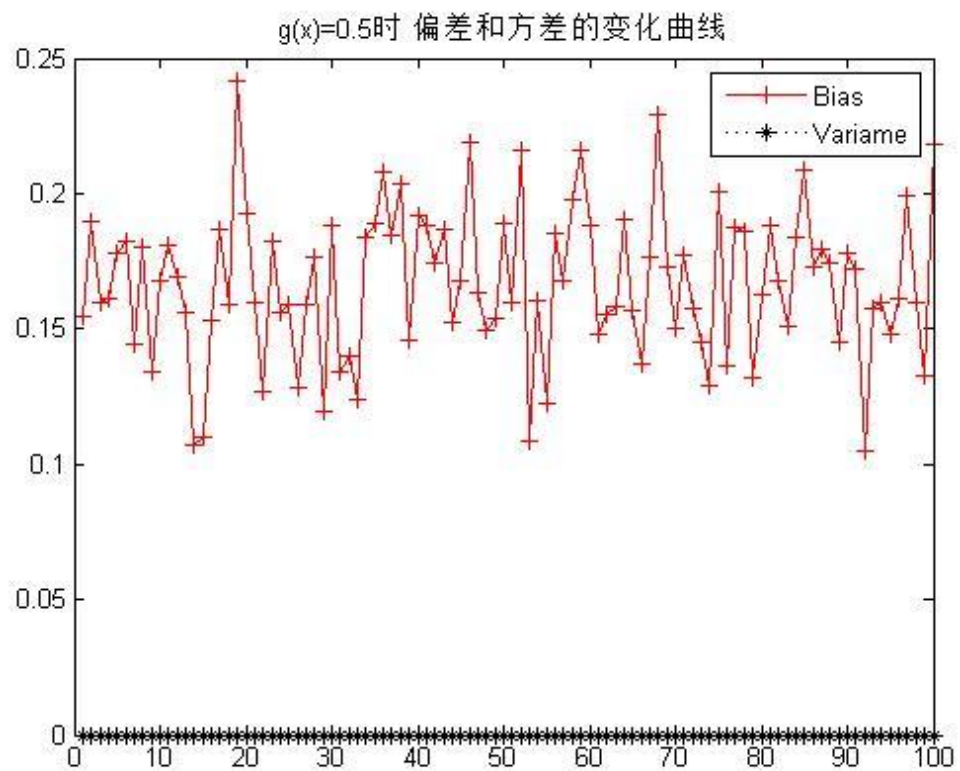
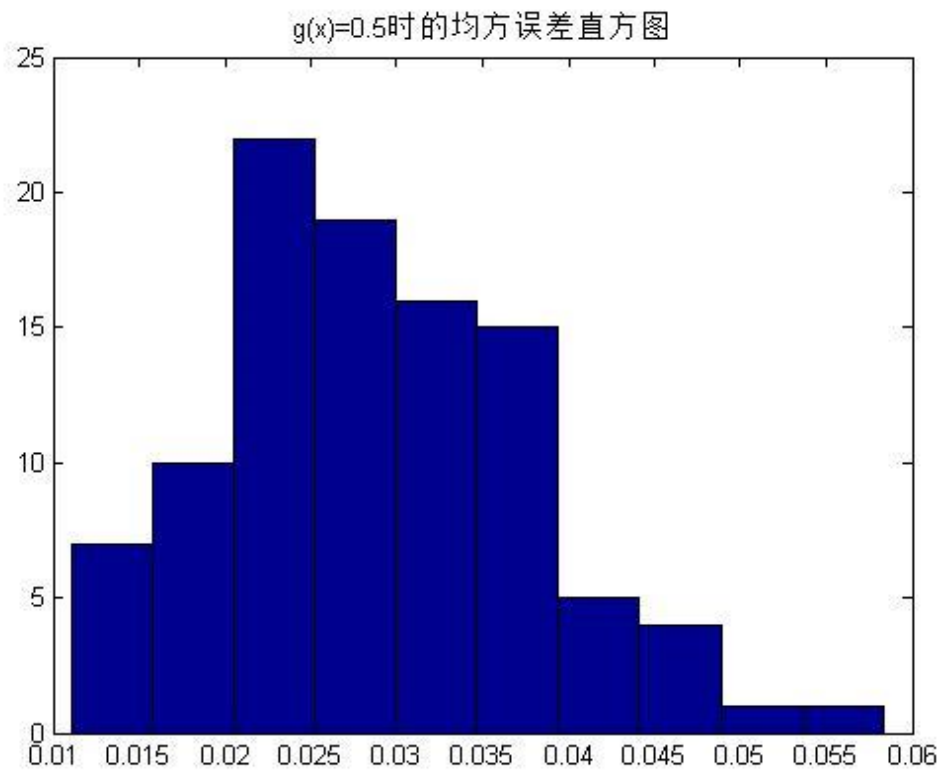
```
>> Res(10)
n=10, g(x)=0.5时, 偏差均值为: 0.171896 方差均值为: 0.000000
n=10, g(x)=1.0时, 偏差均值为: 0.671896 方差均值为: 0.000000
n=10, g(x)=a0 + a1*x时, 偏差均值为: 0.160468 方差均值为: 0.011810
n=10, g(x)=a0+a1*x+a2*x^2+a3*x^3时, 偏差均值为: 0.160468 方差均值为: 0.084234
>> Res(100)
n=100, g(x)=0.5时, 偏差均值为: 0.166820 方差均值为: 0.000000
n=100, g(x)=1.0时, 偏差均值为: 0.666820 方差均值为: 0.000000
n=100, g(x)=a0 + a1*x时, 偏差均值为: 0.159056 方差均值为: 0.001533
n=100, g(x)=a0+a1*x+a2*x^2+a3*x^3时, 偏差均值为: 0.159056 方差均值为: 0.088664
fx >>
```

当  $g(x)=0.5$  时:

N 取 10, 运行的详细结果如下:

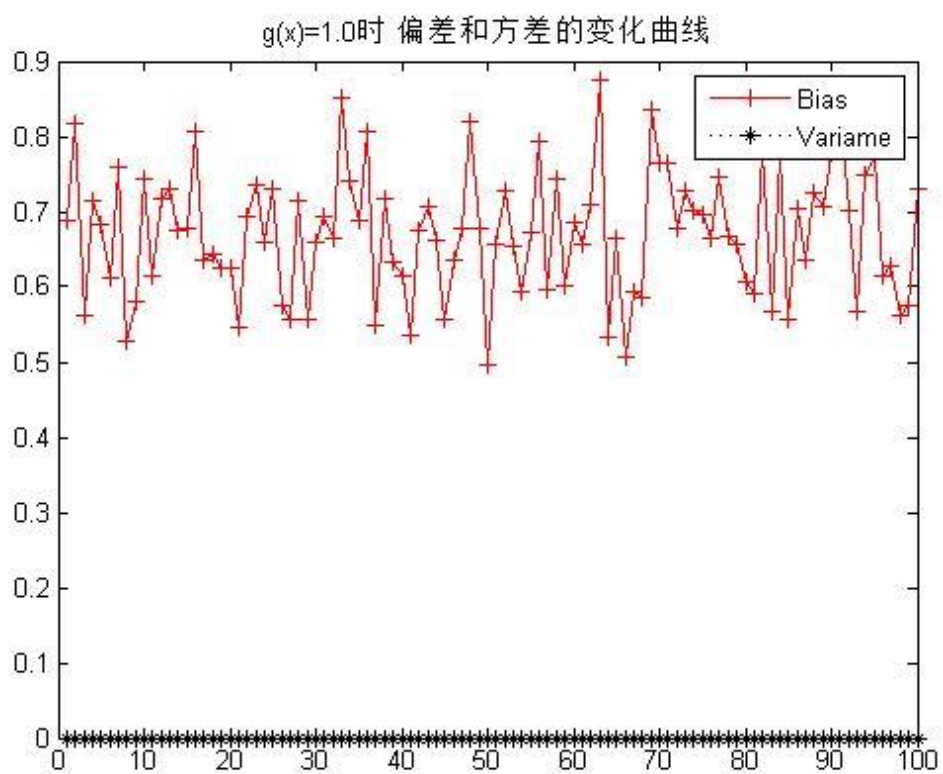
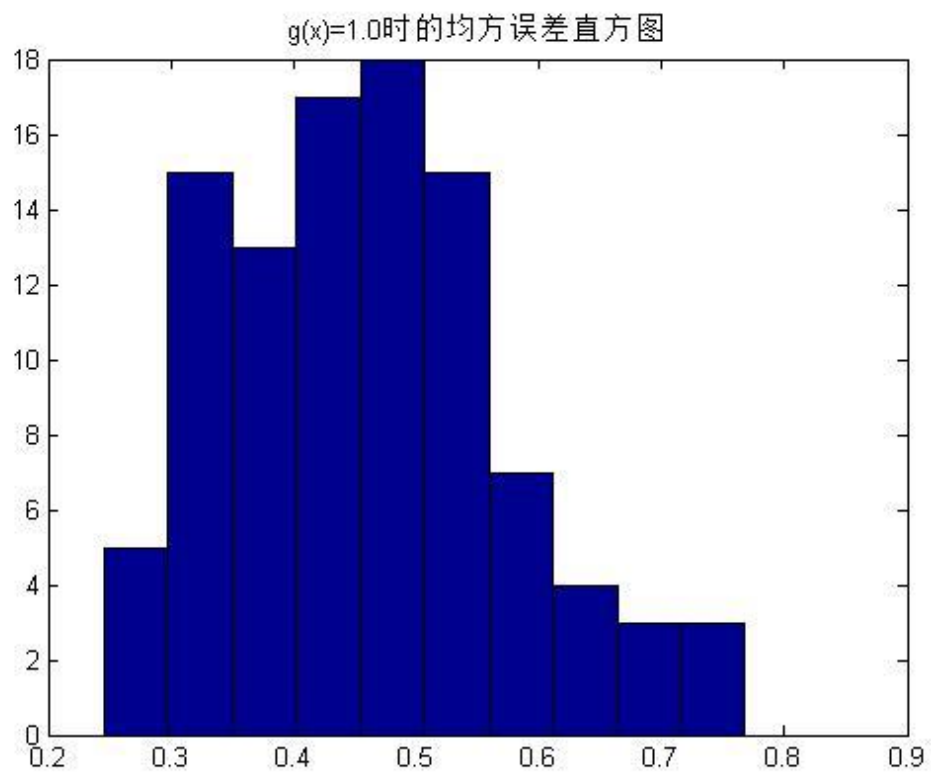


N 取 100, 运行的详细结果如下:

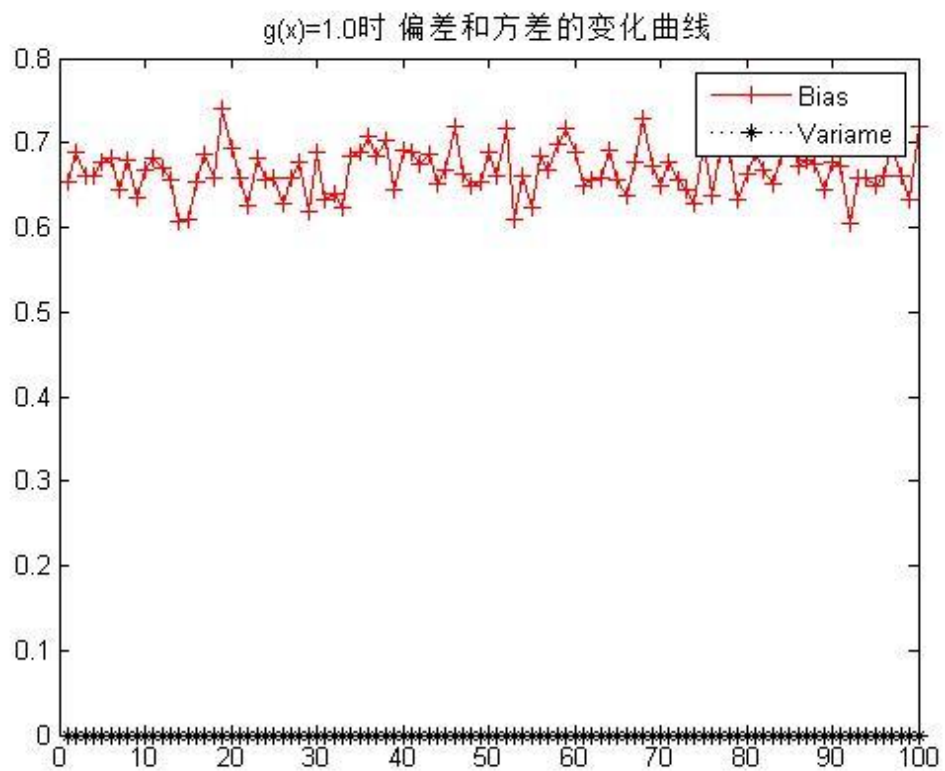
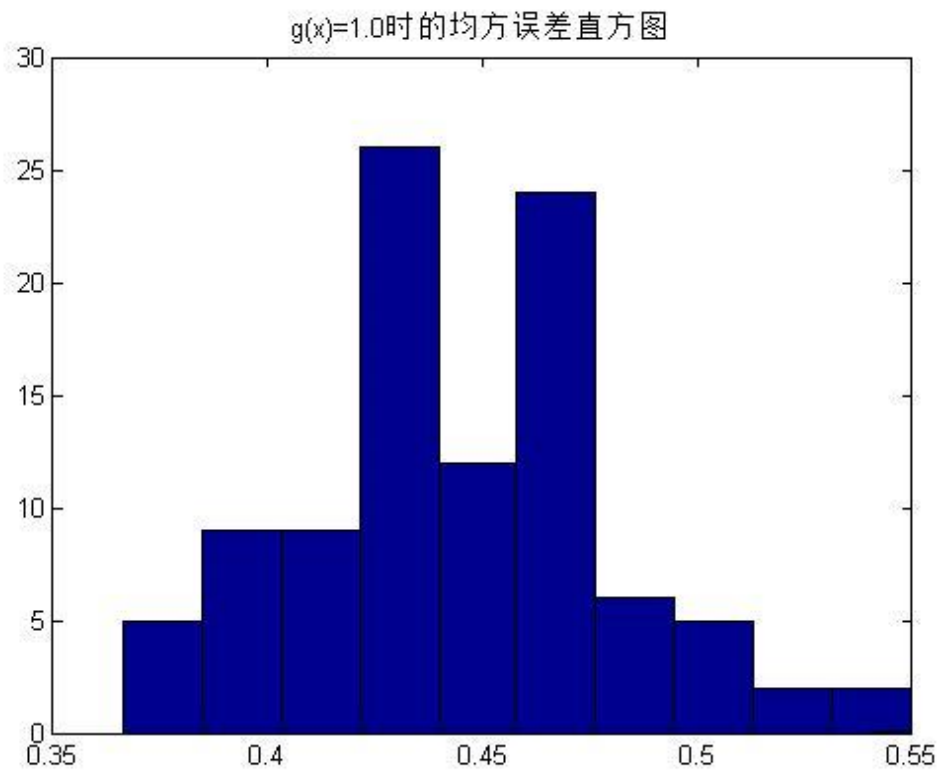


当  $g(x)=1.0$  时:

N 取 10, 运行的详细结果如下:

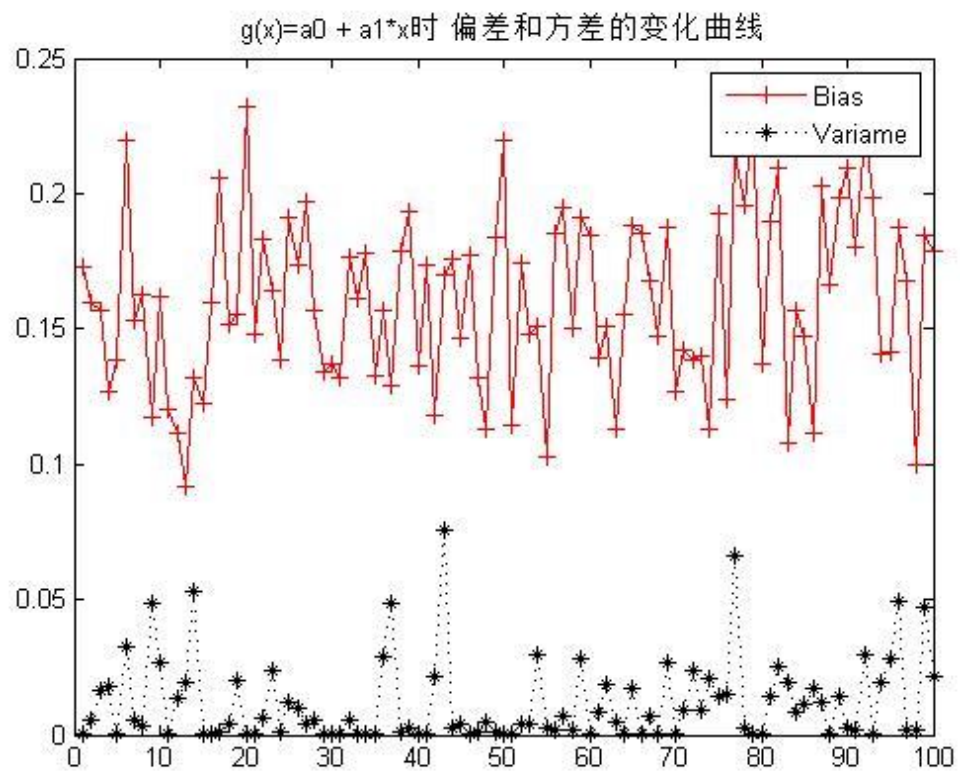
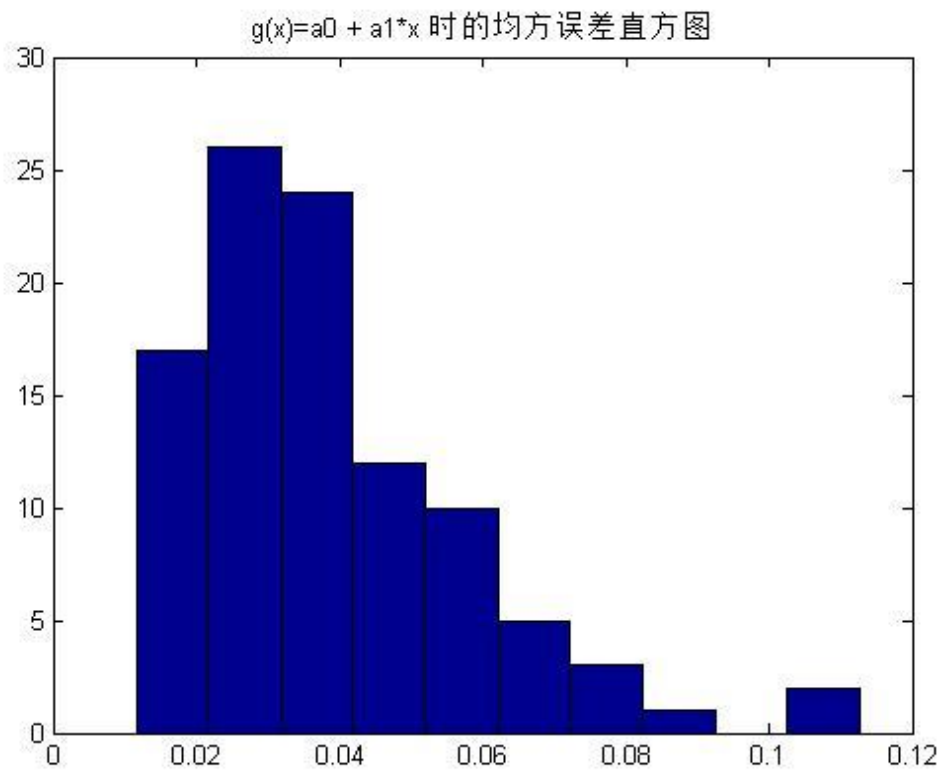


N 取 100，运行的详细结果如下：

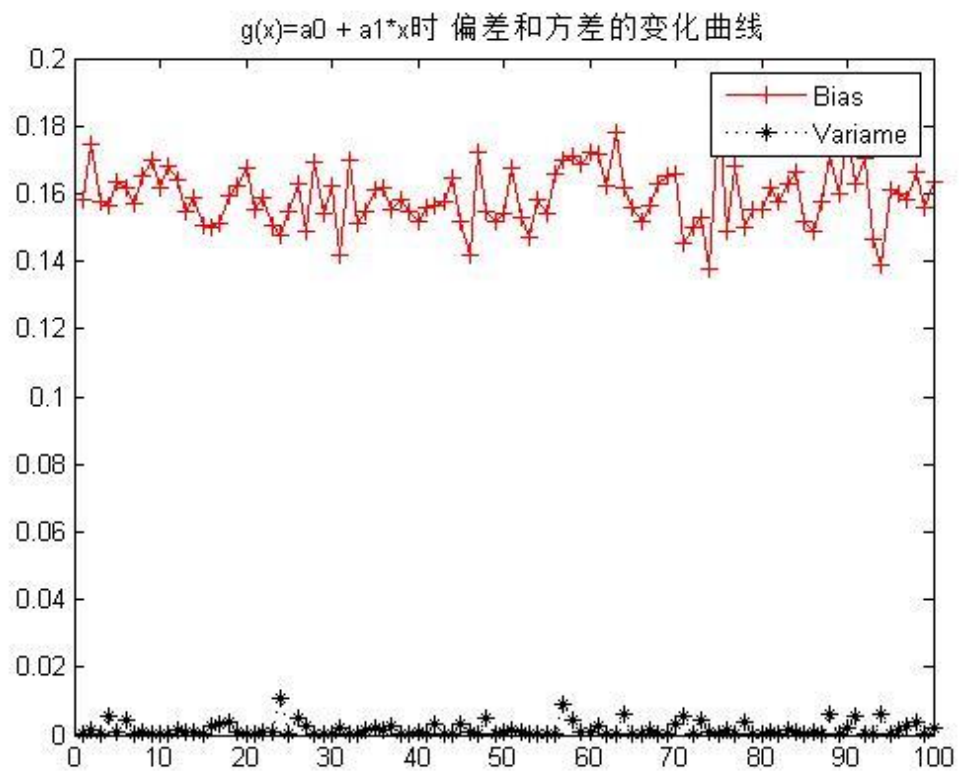
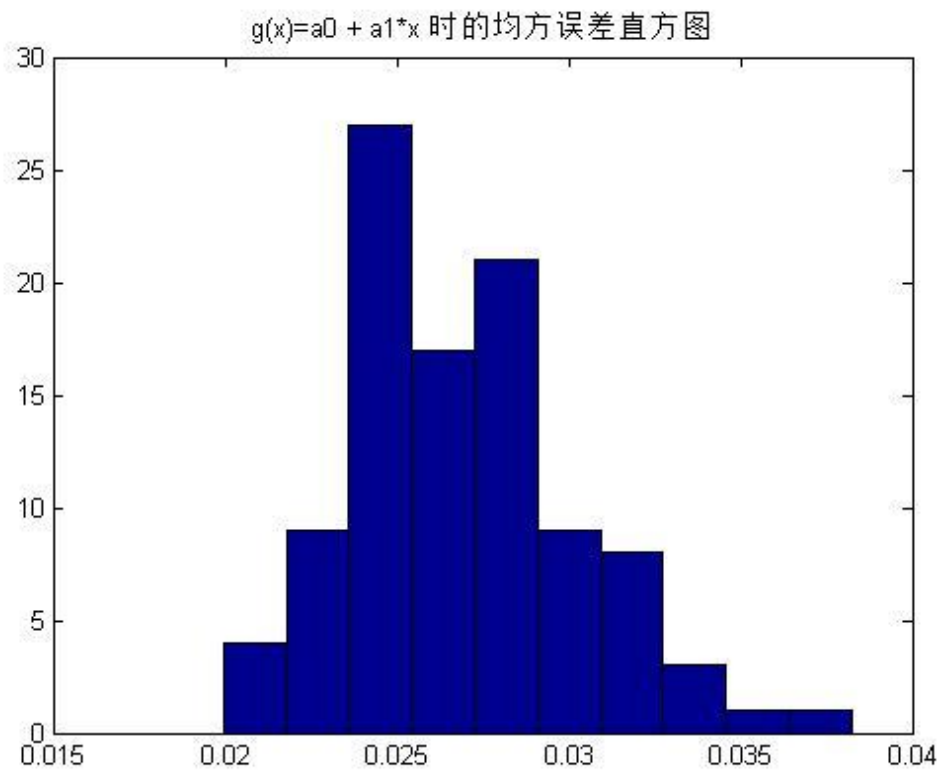


当  $g(x)=a_0 + a_1 \cdot x$  时

N 取 10，运行的详细结果如下：



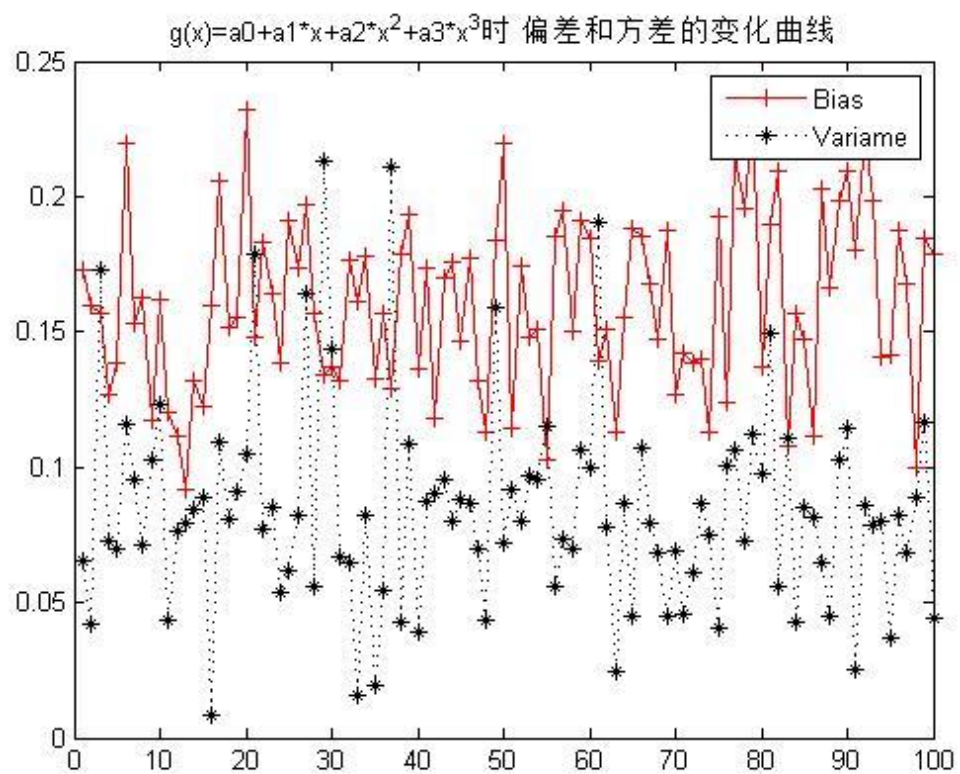
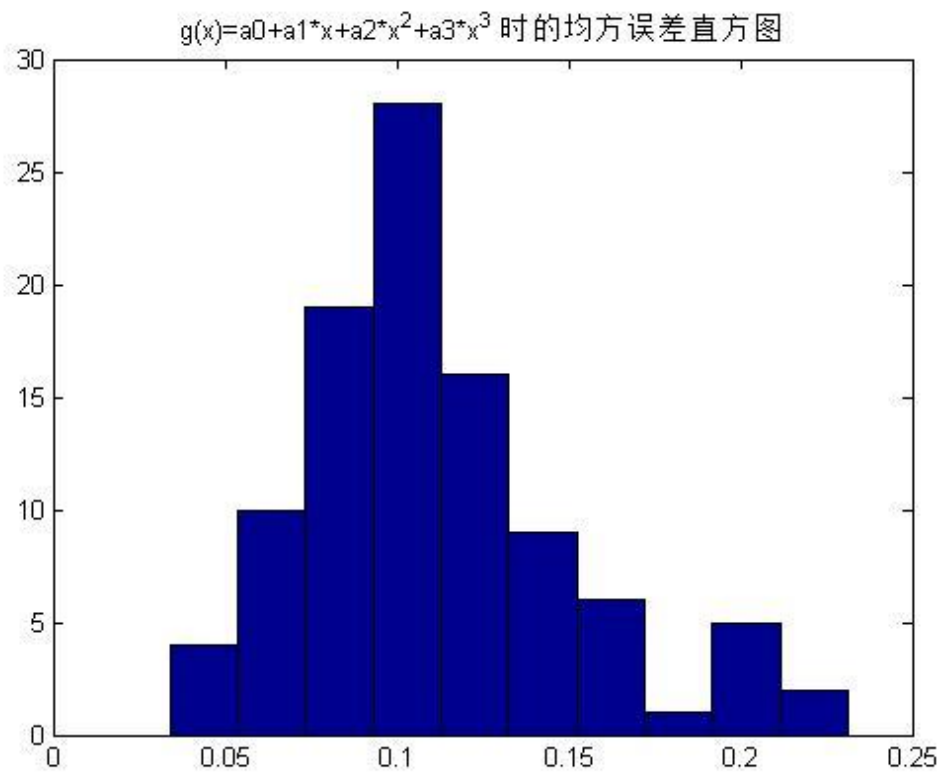
N 取 100，运行的详细结果如下：



当  $g(x)=a_0+a_1x+a_2x^2+a_3x^3$  时:

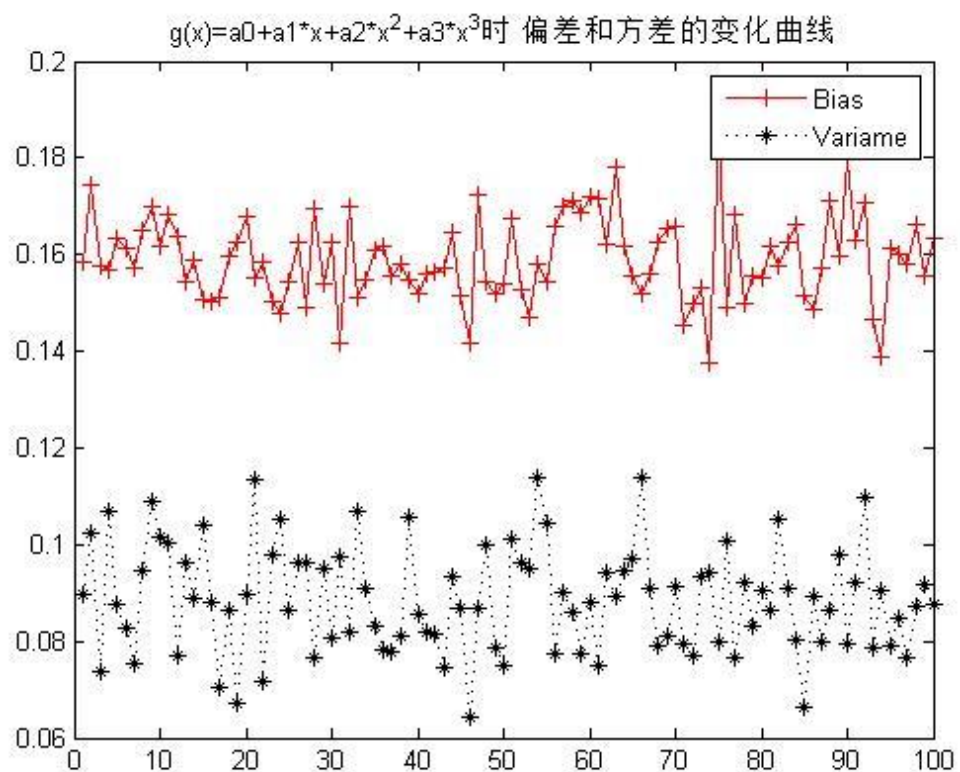
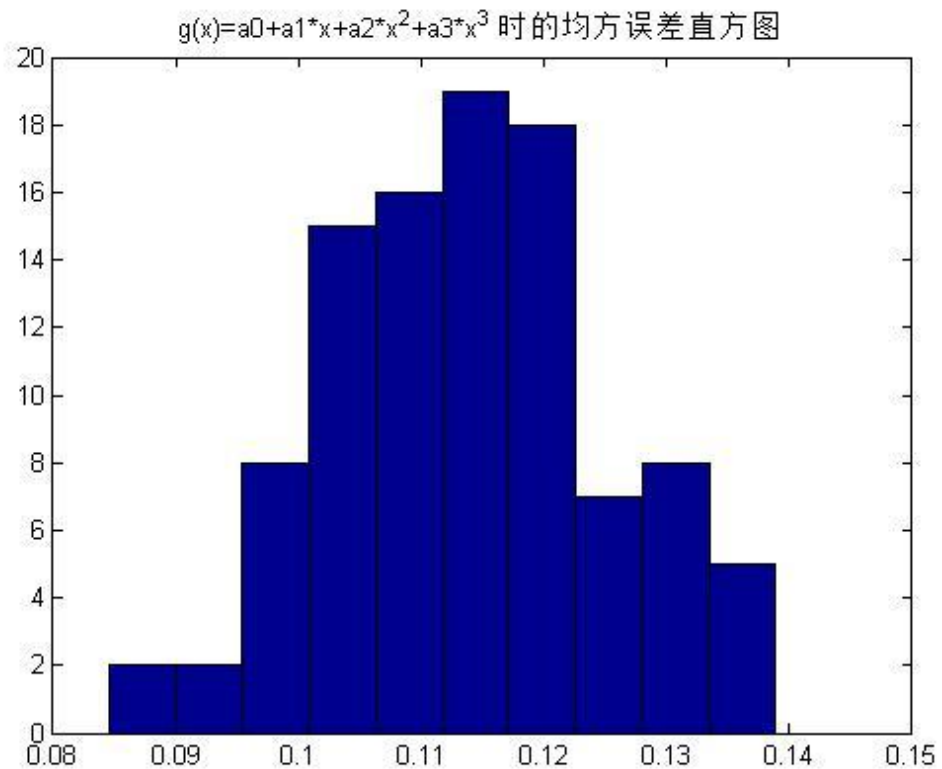
N 取 10, 运行的详细结果如下:





N 取 100，运行的详细结果如下：





由上述图表，当假设模型是常数时，方差为 0，但是偏差往往很大。当假设模型是线性函数时，由偏差和方差的变化曲线可以看出，当 Bias 变小时，Variance 变大。可以看出偏差方差两难问题。此外，n=100 时，B/V 的上下波动比 n=10 的时候小很多，表明随着数据的增多，B/V 趋于稳定。