

### 第三次作业

1220086621 朱金秋

7.3 记  $A_k$  为数据矩阵  $X$  中的任何一行在 bootstrap 矩阵  $X^*$  中正好出现  $k$  次,  $P(A_k)$  为该事件发生的概率

$$\text{那么 } P(A_0) = C_{88}^0 \left(\frac{67}{88}\right)^{88} \approx 0.366$$

$$P(A_1) = C_{88}^1 \left(\frac{1}{88}\right) \left(\frac{87}{88}\right)^{88-1} \approx 0.31$$

$$P(A_2) = C_{88}^2 \left(\frac{1}{88}\right)^2 \left(\frac{87}{88}\right)^{88-2} \approx 0.185$$

$$P(A_3) = C_{88}^3 \left(\frac{1}{88}\right)^3 \left(\frac{87}{88}\right)^{88-3} \approx 0.061$$

## 习题7.6

In [2]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [3]:

```
data = pd.read_csv("D:/Data/mardia.dat", header=None)
f = open("D:/Data/mardia.dat")
L = []
for lines in f.readlines():
    a = lines.split()
    L.append(a)
data = pd.DataFrame(L) #去除分隔符，导入到数据框中

for i in range(0, 5):
    data[i] = pd.to_numeric(data[i]) #转成数值型
```

In [4]:

```
"""
设置随机抽样子函数
输入：抽样数量
输出：抽处的样本
"""
def random_sample(n):
    rd = list(n*np.random.rand(n)) #U(0, n) 随机数
    index = []
    for a in rd:
        index.append(int(a))
        xb = data.iloc[index]
    return xb
def C(cov_data):
    M = np.zeros((5, 5))
    for j in range(5):
        for m in range(5):
            M[j][m] = cov_data[j][m] / np.sqrt(cov_data[j][j]*cov_data[m][m])
    return M
```

In [24]:



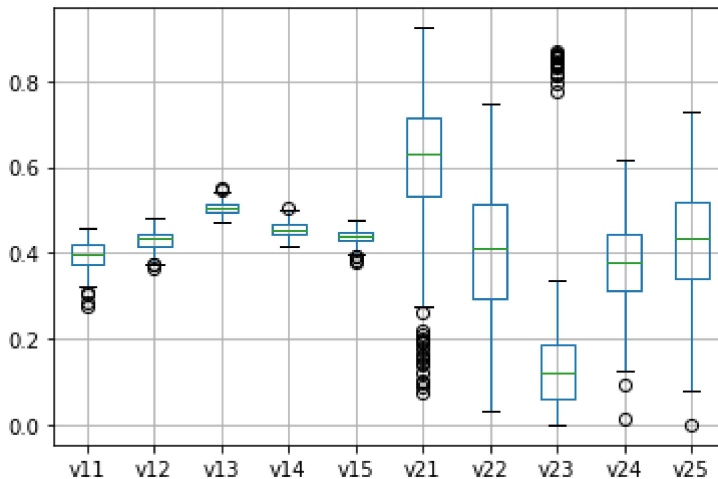
```

"""
主函数
输出结果
"""

if __name__ == '__main__':
    Nr = 200
    n = len(data)
    datavec_Nr = []
    for i in range(Nr): #循环200次
        xb = random_sample(n) #调用子函数
        cov_xb = xb.cov()
        q = C(cov_xb) #根据题意要求G转成C
        xb_val, xb_vec = np.linalg.eig(np.array(q))
        datavec_Nr.append(xb_vec) #将特征向量加入空列表
a = np.abs(datavec_Nr)
#设定第一特征向量V1和V2
v1 = []
for i in range(len(a)):
    v1.append(a[i][:, 0])

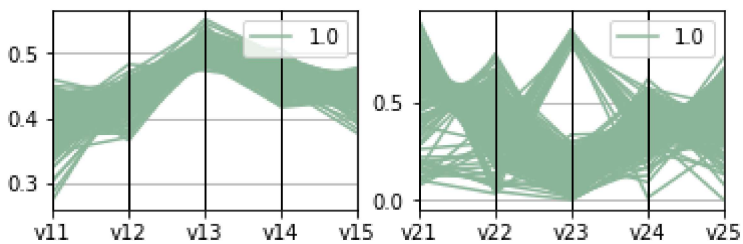
v2 = []
for i in range(len(a)):
    v2.append(a[i][:, 1])
df1 = pd.DataFrame(v1, columns=['v11', 'v12', 'v13', 'v14', 'v15'])
df2 = pd.DataFrame(v2, columns=['v21', 'v22', 'v23', 'v24', 'v25'])
df = pd.concat([df1, df2])
df.boxplot() #绘制箱线图
plt.show()

```



In [25]:

```
a = df1
a['label'] = np.ones(len(a))
b = df2
b['label'] = np.ones(len(b))
#绘制折线图
from pandas.plotting import parallel_coordinates
%matplotlib inline
plt.figure()
plt.subplot(2,2,1)
parallel_coordinates(a, 'label')
plt.subplot(2,2,2)
parallel_coordinates(b, 'label')
plt.show()
```



通过观察两组特征向量（第一特征向量组、第二特征向量组箱型图和折线图），可以发现第二特征向量比第一特征向量离散，第一特征向量比第二特征向量更准确，与书上结论类似

In [ ]: