PCA with Matrix 行列でPCA

ryamada

2016年12月24日

1 行列の特徴

正方行列の特徴を思い出そう。

正方行列は回転させても変わらない性質があった。

固有値の和(トレース)と固有値の積(行列式)がそれである。

『行列の特徴』のExercises 4-6, 4-7で見たように、分散共分散行列を回転させると、固有値の和と積は変わらないが、固有値の内訳は変わる。 そして、対角化すると、固有値の大小が最も大きくばらつくようにできることを見た。 最も大きくばらつくとは、最大固有値が、色々な回転の中で最大となり、最小固有値が色々な回転の中で最小になるような内訳の取り方を指す。

2 Exercise 1

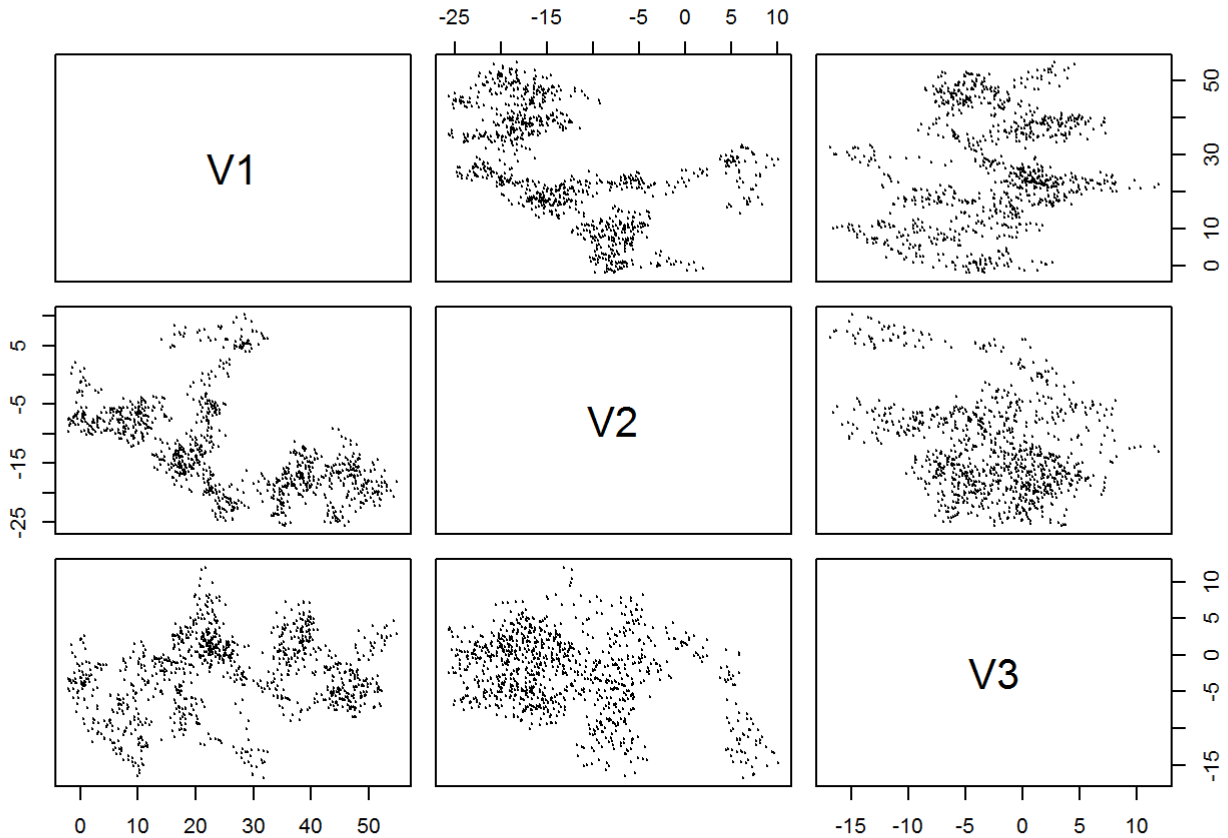
2.1 Exercise 1-1

2.2 Exercise 4

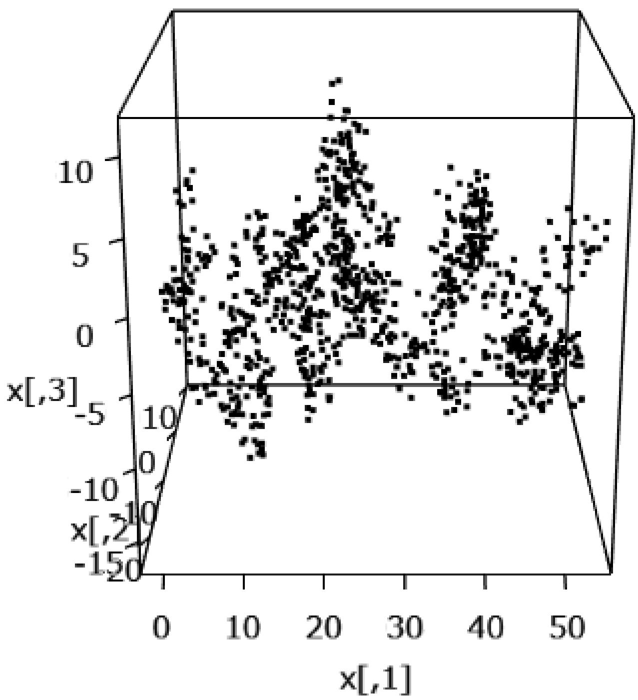
次のようにして生成される、多変量の観察データ x があったとする。

```
d <- 3
n <- 1000
x <- matrix(rnorm(d*n),ncol=d)
x <- apply(x,2,cumsum)

plot(as.data.frame(x),pch=20,cex=0.1)
```


V1
V2
V3
-25 -15 -5 0 5 10
0 10 20 30 40 50
-15 -10 -5 0 5 10

```
plot3d(x)
```


x[,1]
x[,2]
x[,3]
0 10 20 30 40 50
10 5 0 -5 -10 -15

2.3 Exercise 4-1

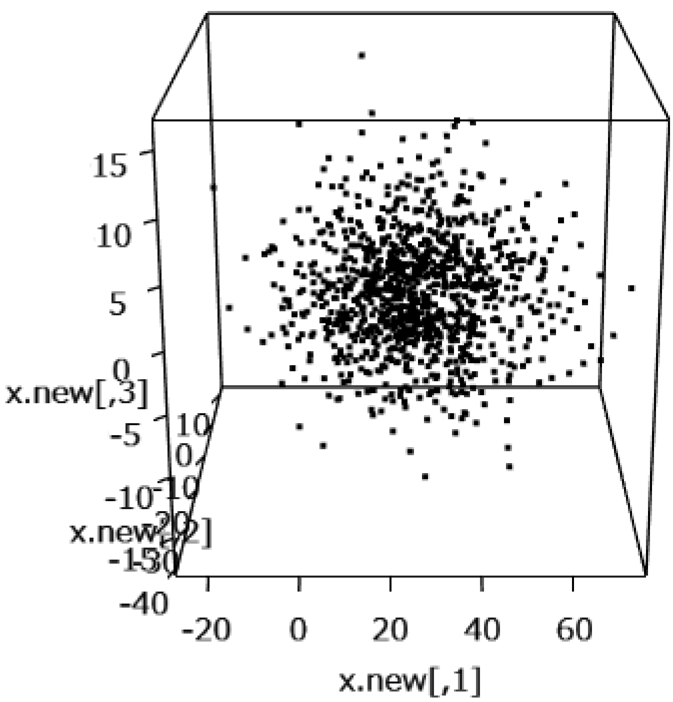
xの各変数の標本平均を求めよ。 xの分散共分散行列を求めよ。

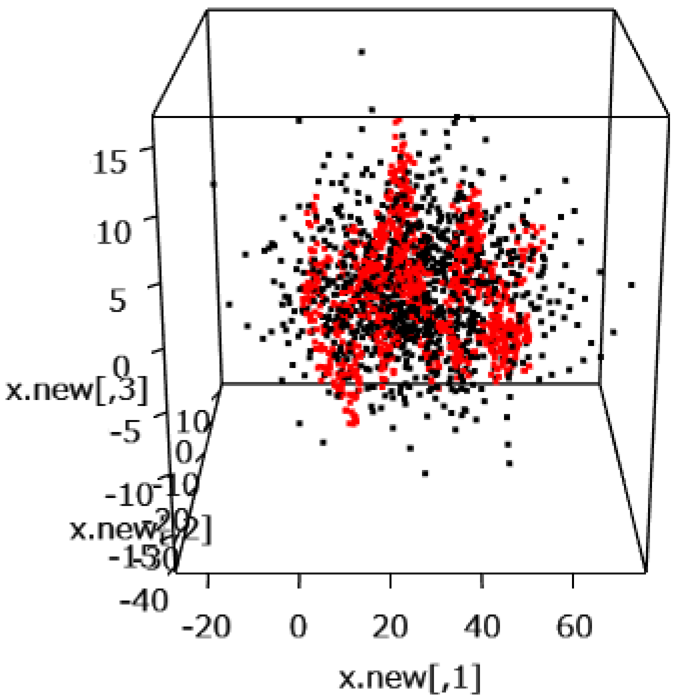
```
## [1]  25.753973 -12.605116  -2.632409
```

```
##            [,1]      [,2]      [,3]
## [1,] 217.80496 -58.46672  14.97610
## [2,] -58.46672  61.06587 -12.83507
## [3,]  14.97610 -12.83507  26.76844
```

2.4 Exercise 4-2

4-1 で求めた平均値ベクトルと分散共分散行列とに基づく、多変量正規乱数を発生し、ペアワイズにプロットせよ。





2.5 Exercise 4-3

回転行列を用いて、xを回転し(軸を取り直し)、その標本平均と分散共分散行列とを求め、4-1のそれと比較せよ。

各変数の平均も分散も異なることを確かめよ。

```
## [1] "mean x"
```

```
## [1]  25.753973 -12.605116  -2.632409
```

```
## [1] "mean x rotated"
```

```
## [1] -14.94189  22.44208  10.10834
```

```
## [1] "cov x"
```

```
##           [,1]      [,2]      [,3]
## [1,] 217.80496 -58.46672  14.97610
## [2,] -58.46672  61.06587 -12.83507
## [3,]  14.97610 -12.83507  26.76844
```

```
## [1] "cov x rotated"
```

```
##            [,1]      [,2]      [,3]
## [1,]  90.50162 -62.32342 -59.61912
## [2,] -62.32342 128.86385  78.68434
## [3,] -59.61912  78.68434  86.27380
```

2.6 Exercise 4-4

分散は異なるが、分散の和は変わっていないことを、分散共分散行列のトレースを計算することで確かめよ。

同様に、分散共分散行列の固有値の和としてトレースを計算することでも確かめよ。

```
## [1] 305.6393
```

```
## [1] 305.6393
```

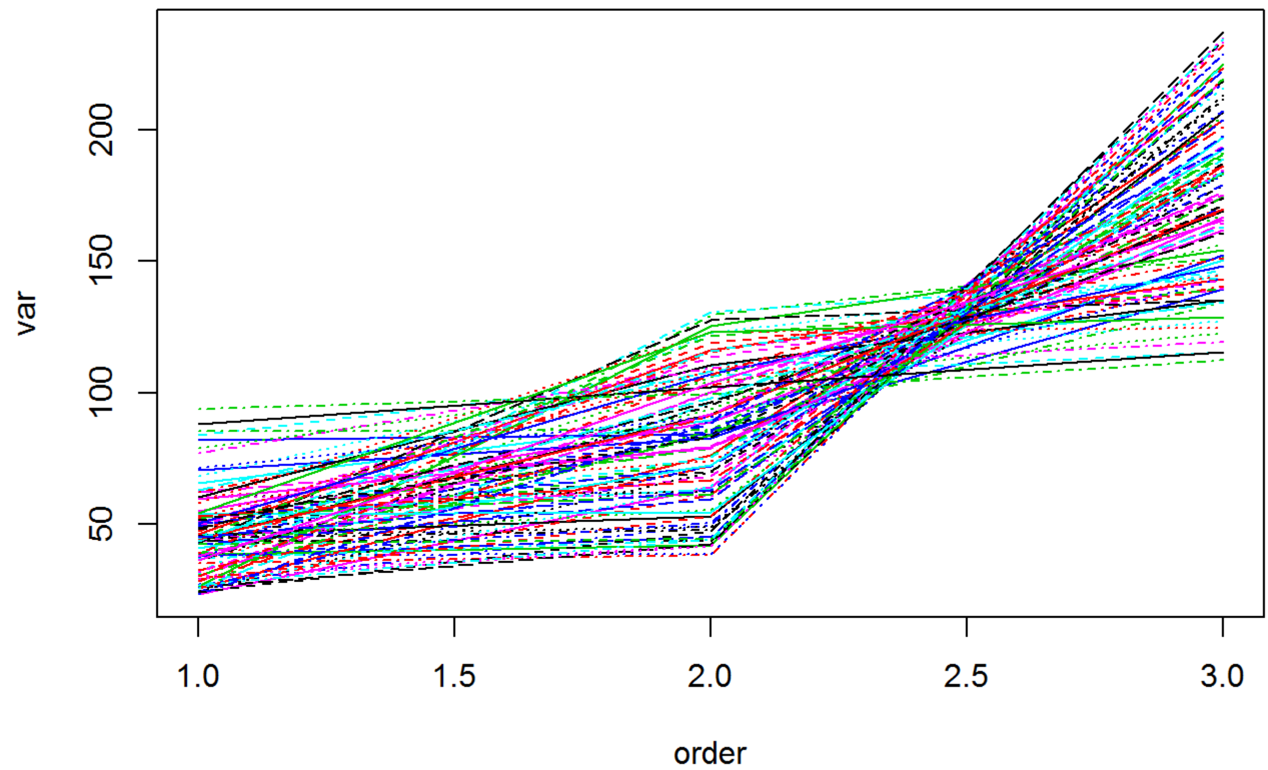
```
## [1] 305.6393
```

```
## [1] 305.6393
```

2.7 Exercise 4-5

データセットxの軸の取り方を回転することによって変えても、各軸の分散の和は変わらない。 分散の和は変わらないが、分散の総和の分配具合は変わる。

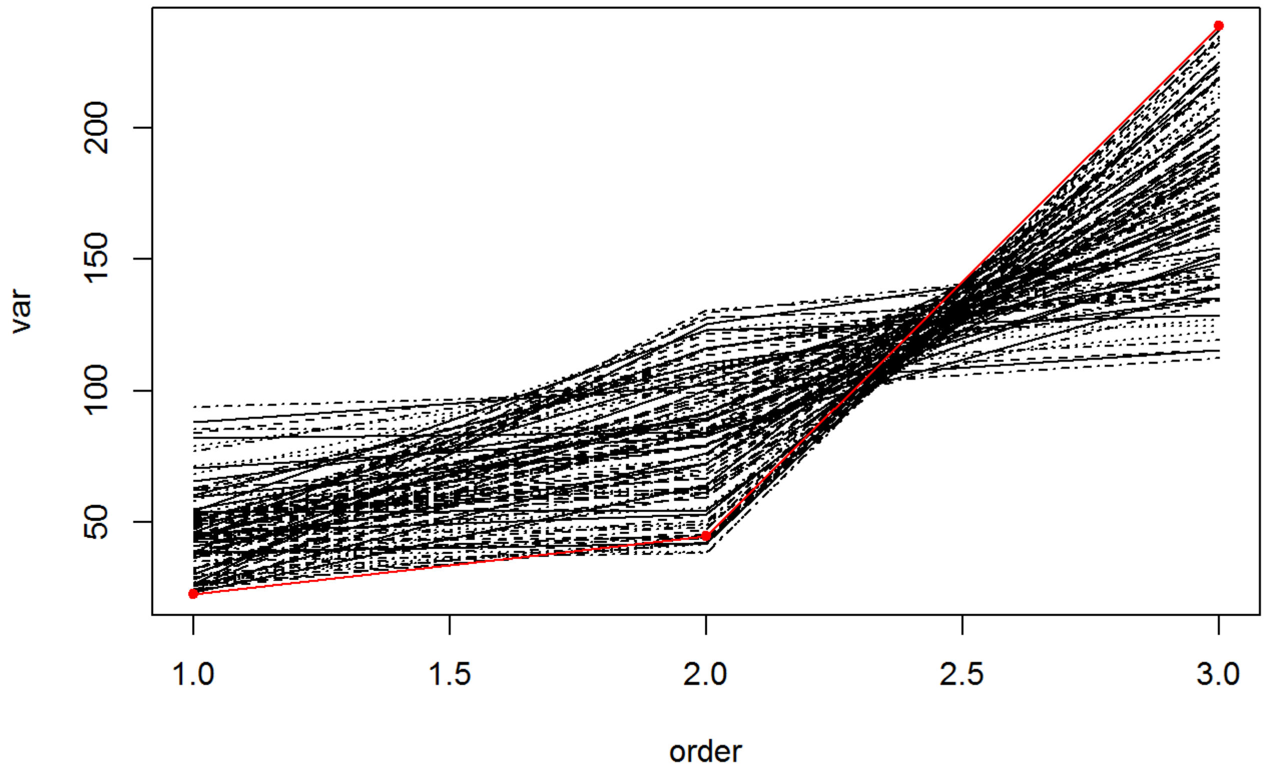
そのことを、多数の回転行列をランダムに発生させ、データセットを回転した上で各軸方向の分散を計算し、その分散の値を大小順序でソートしてプロットし、分配具合が変わることを具体的に確認せよ。



2.8 Exercise 4-6

xの分散共分散行列の固有値を求めることで、分散共分散行列を対角化するような回転をしたときの、各軸の分散を求めることができる。

その分散を4-4のプロットに重ねてプロットせよ。

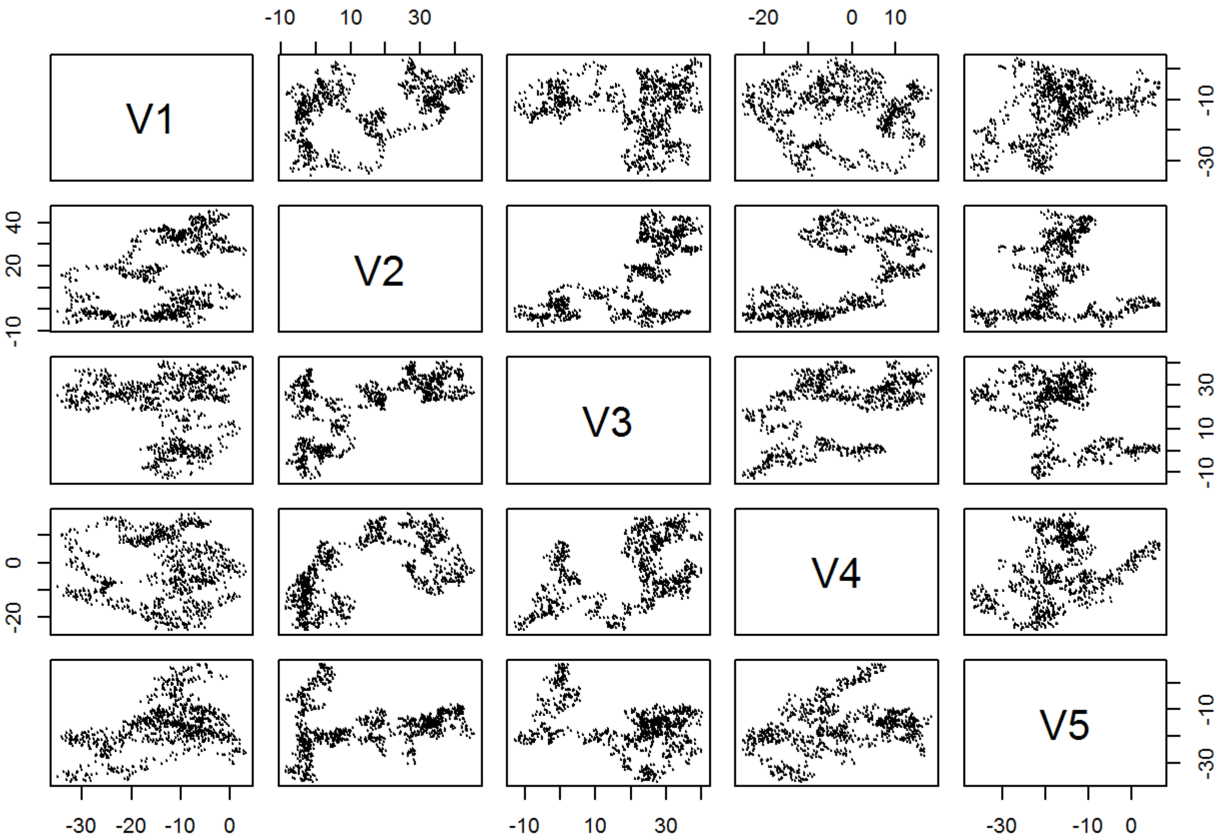


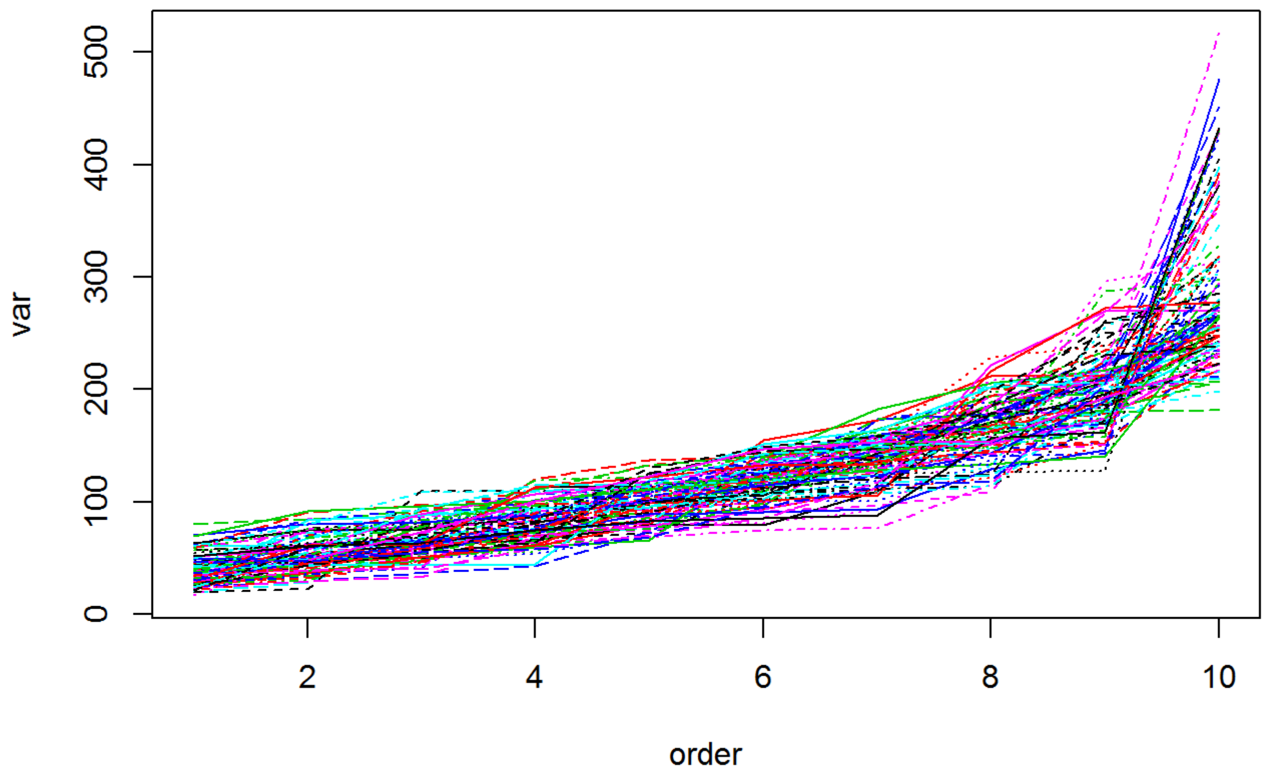
2.9 Exercise 4-7

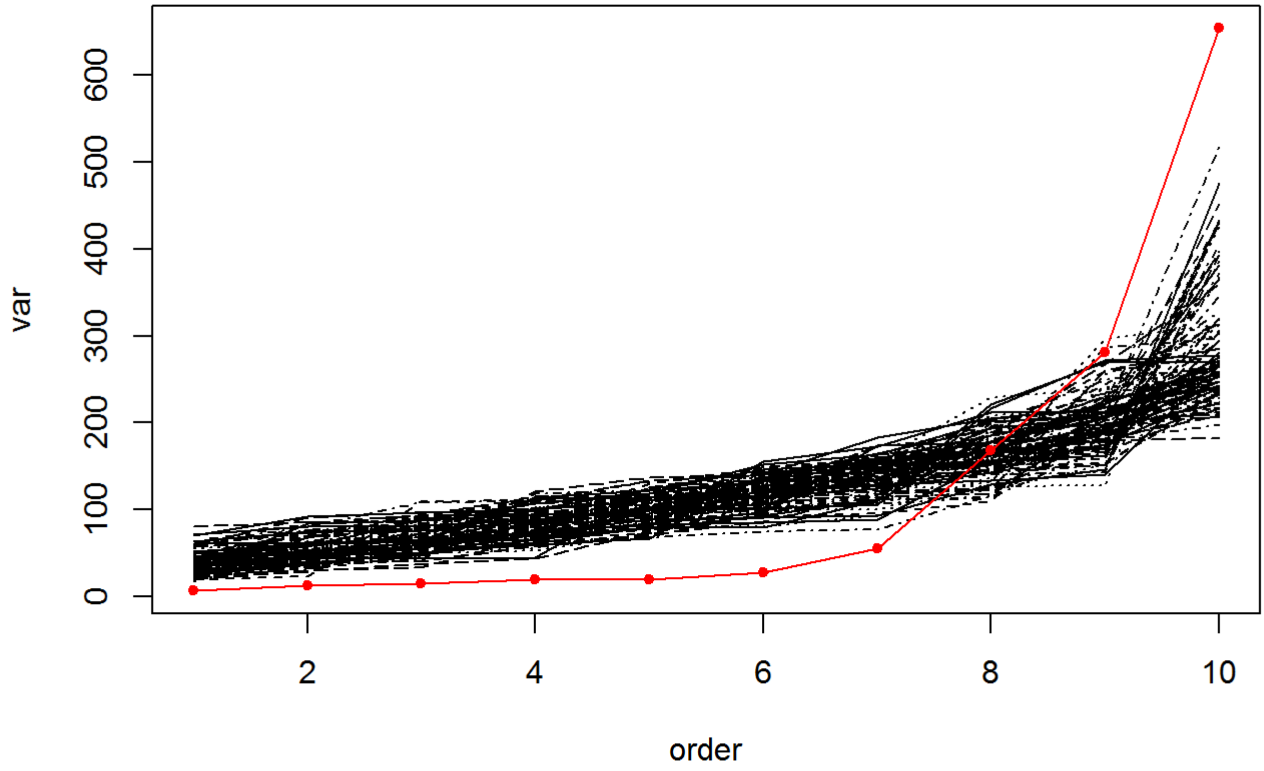
4-4,4-5と同様のプロットを変数の数を増やして実施せよ。

```
d <- 10
n <- 1000
x <- matrix(rnorm(d*n),ncol=d)
x <- apply(x,2,cumsum)

plot(as.data.frame(x[,1:5]),pch=20,cex=0.1)
```


V1
V2
V3
V4
V5
-10
10
30
-20
0
10
-10
-30
40
20
-10
30
10
-10
0
-20
-10
-30
-30
-20
-10
0
-10
10
30
-30
-10
0


var
order
0
100
200
300
400
500
2
4
6
8
10



『行列の特徴』のExercises 4-6, 4-7を再度実施せよ。

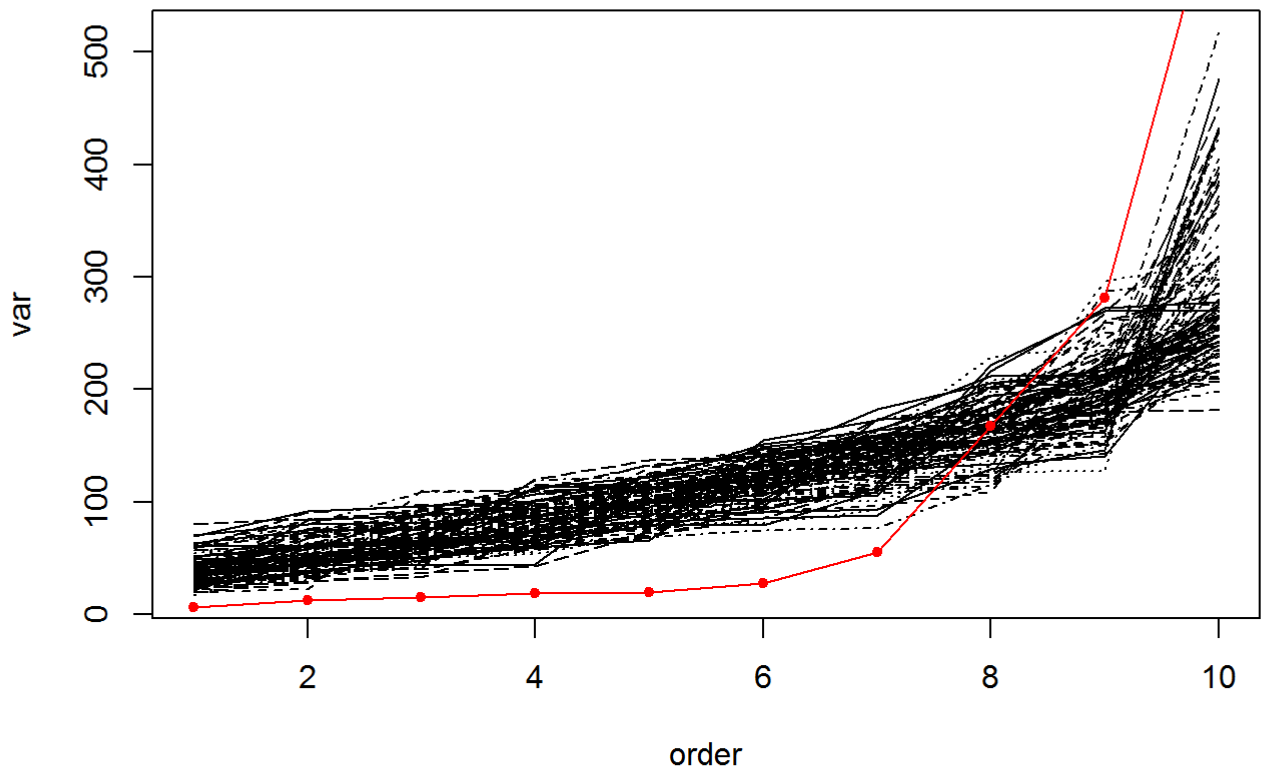
3 PCA

PCAは分散共分散行列を回転させて、固有値のばらつきを大きくして、数少ない固有値で、すべての固有値の和の多くの部分を説明しようとするものである。

3.0.1 Exercise 4-6

xの分散共分散行列の固有値を求めることで、分散共分散行列を対角化するような回転をしたときの、各軸の分散を求めることができる。

その分散を4-4のプロットに重ねてプロットせよ。

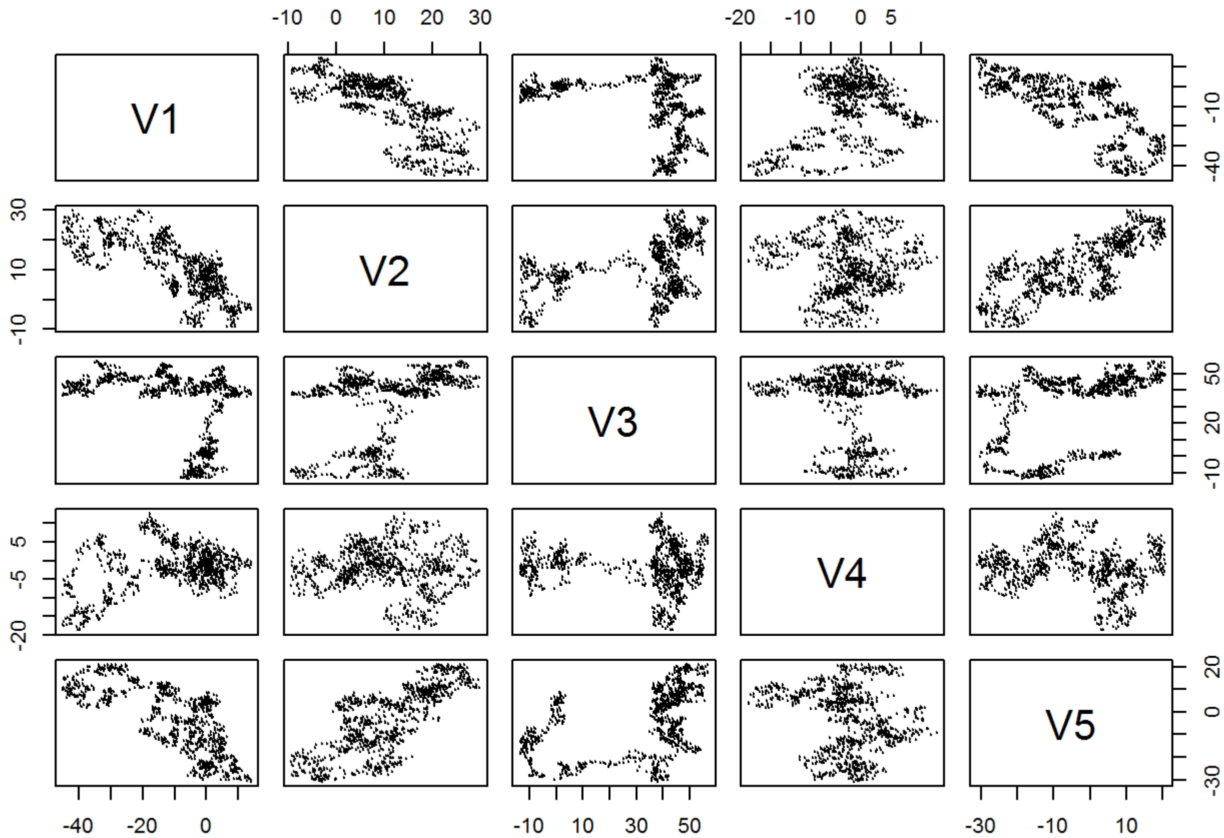


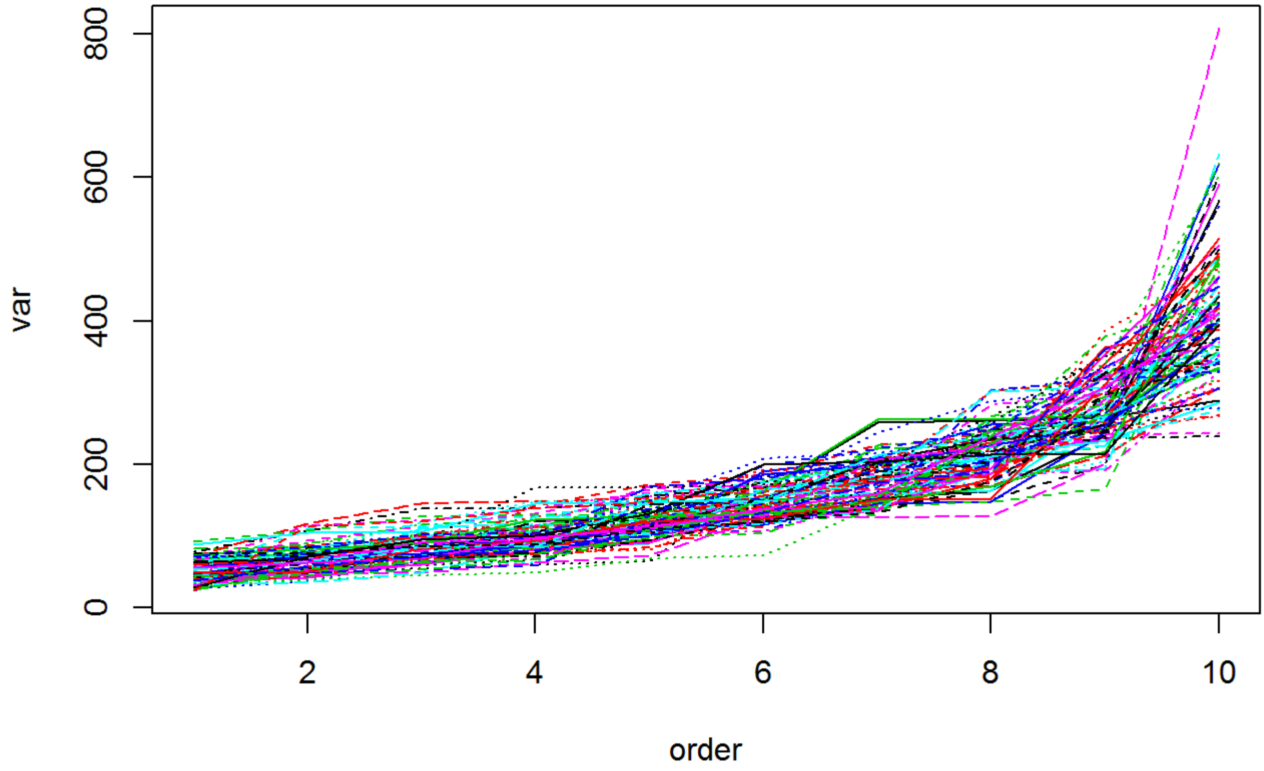
3.0.2 Exercise 4-7

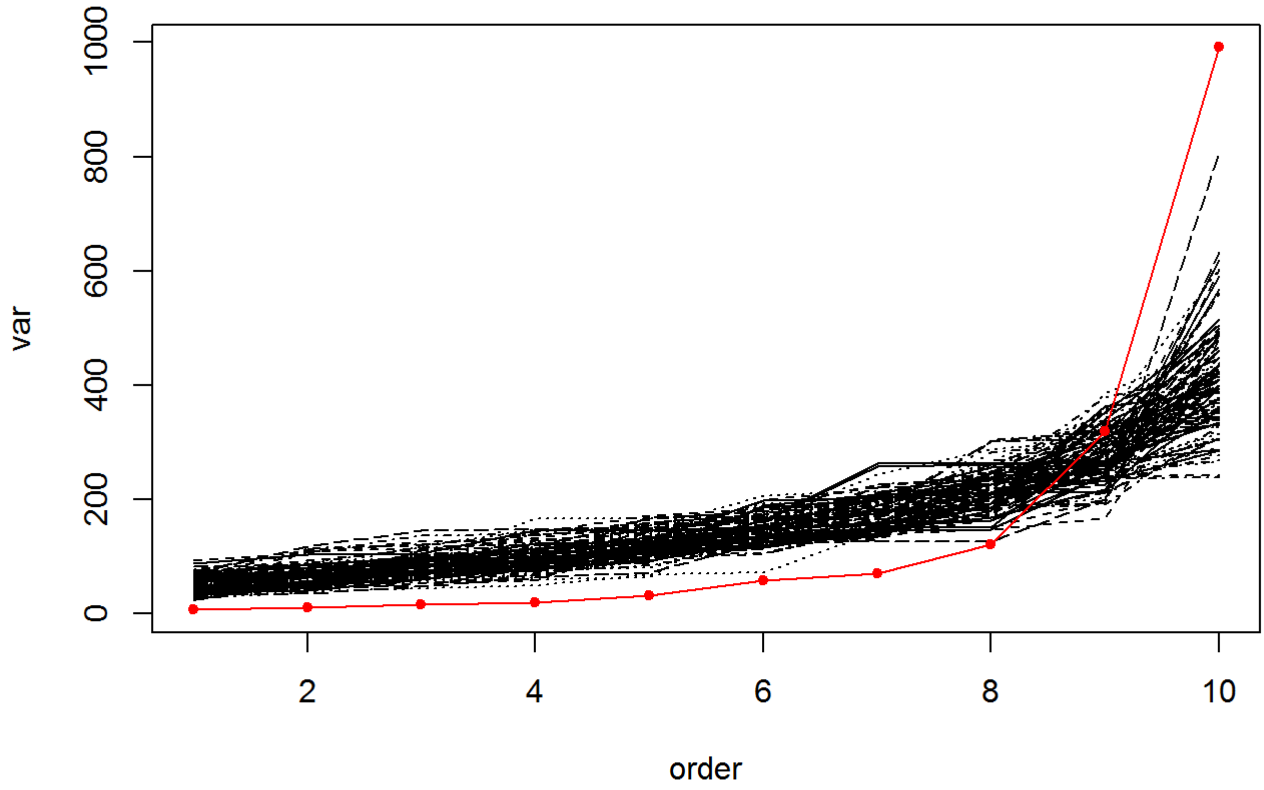
4-4,4-5と同様のプロットを変数の数を増やして実施せよ。

```
d <- 10
n <- 1000
x <- matrix(rnorm(d*n),ncol=d)
x <- apply(x,2,cumsum)

plot(as.data.frame(x[,1:5]),pch=20,cex=0.1)
```


V1
V2
V3
V4
V5


800
600
400
200
0
var
2
4
6
8
10
order


1000
800
600
400
200
0
var
2
4
6
8
10
order