多元统计分析作业1

辛柏嬴 2020111753

2023-03-10

### **1) 随机生成一个5维多元正态分布，其中：均值向量：; 协方差矩阵：**

解：思路如下

首先生成, 通过Cholesky分解求出矩阵使得,设置向量使得, 最终令 即为所求

*引理【Cholesky分解】：对于一个正定矩阵，存在一个对角元全为正数的下三角矩阵使得成立.*

set.seed(123) #生成随机数种子  
Z <- rnorm(5,0,1) #初始化5\*1的标准正态分布向量  
sigma <- matrix(0,5,5) #sigma是协方差矩阵，这里首先进行初始化  
for (i in 0:5){  
 for (j in 0:5){  
 sigma[i,j] <- 0.7^(abs(i-j)) #得到sigma矩阵  
 }  
}  
A <- chol(sigma) #Cholesky分解  
mu <- c(1,2,3,4,5) #设置均值  
X <- A%\*%Z+mu #X = AX+\mu  
print(X)

## [,1]  
## [1,] 1.097394  
## [2,] 2.671161  
## [3,] 4.193629  
## [4,] 4.114984  
## [5,] 5.092330

此外，还可以通过调用MASS包直接生成：

# install.packages("MASS")  
library(MASS)  
X\_mass <- mvrnorm(1,mu,sigma) #通过mvrnorm可以直接生成  
print(X\_mass)

## [1] 0.6312591 0.5540929 0.8719170 2.5758327 3.8113588

### **2）对上述正态分布的协方差矩阵：**

#### **a）进行谱分解、奇异值分解**

#谱分解  
eig <- eigen(sigma) #求sigma的特征值及对应特征向量  
V <- eig$vectors #V是特征向量 （组成的矩阵）  
lam <- eig$values #lam是特征值（组成的向量）  
#输出谱分解结果  
print("V:")  
print(V)  
print("lambda:")  
print(lam)  
print("V inverse:")  
print(solve(V))

## [1] "V:"  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.3940662 5.767041e-01 -0.5453050 -4.091606e-01 0.2176104  
## [2,] -0.4704256 4.091606e-01 0.1364390 5.767041e-01 -0.5099845  
## [3,] -0.4968131 -2.775558e-16 0.6066743 -6.106227e-16 0.6205828  
## [4,] -0.4704256 -4.091606e-01 0.1364390 -5.767041e-01 -0.5099845  
## [5,] -0.3940662 -5.767041e-01 -0.5453050 4.091606e-01 0.2176104  
## [1] "lambda:"  
## [1] 3.1029654 1.0131847 0.4339891 0.2567153 0.1931455  
## [1] "V inverse:"  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.3940662 -0.4704256 -4.968131e-01 -0.4704256 -0.3940662  
## [2,] 0.5767041 0.4091606 -1.925117e-16 -0.4091606 -0.5767041  
## [3,] -0.5453050 0.1364390 6.066743e-01 0.1364390 -0.5453050  
## [4,] -0.4091606 0.5767041 -3.841621e-16 -0.5767041 0.4091606  
## [5,] 0.2176104 -0.5099845 6.205828e-01 -0.5099845 0.2176104

故谱分解：，具体系数见上方输出。

#奇异值分解  
svd <- svd(sigma)  
print(svd)  
print("奇异值分解：sigma = U %\*% D %\*% transpose(V)") #sigma = UDV^{T}

## $d  
## [1] 3.1029654 1.0131847 0.4339891 0.2567153 0.1931455  
##   
## $u  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.3940662 0.5767041 0.5453050 -4.091606e-01 0.2176104  
## [2,] -0.4704256 0.4091606 -0.1364390 5.767041e-01 -0.5099845  
## [3,] -0.4968131 0.0000000 -0.6066743 -1.151856e-15 0.6205828  
## [4,] -0.4704256 -0.4091606 -0.1364390 -5.767041e-01 -0.5099845  
## [5,] -0.3940662 -0.5767041 0.5453050 4.091606e-01 0.2176104  
##   
## $v  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.3940662 5.767041e-01 0.5453050 -4.091606e-01 0.2176104  
## [2,] -0.4704256 4.091606e-01 -0.1364390 5.767041e-01 -0.5099845  
## [3,] -0.4968131 -8.326673e-17 -0.6066743 -1.221245e-15 0.6205828  
## [4,] -0.4704256 -4.091606e-01 -0.1364390 -5.767041e-01 -0.5099845  
## [5,] -0.3940662 -5.767041e-01 0.5453050 4.091606e-01 0.2176104  
##   
## [1] "奇异值分解：sigma = U %\*% D %\*% transpose(V)"

故奇异值分解：,具体系数见上方输出。

#### **b）求其特征值、特征向量**

eigen(sigma)

## eigen() decomposition  
## $values  
## [1] 3.1029654 1.0131847 0.4339891 0.2567153 0.1931455  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.3940662 5.767041e-01 -0.5453050 -4.091606e-01 0.2176104  
## [2,] -0.4704256 4.091606e-01 0.1364390 5.767041e-01 -0.5099845  
## [3,] -0.4968131 -2.775558e-16 0.6066743 -6.106227e-16 0.6205828  
## [4,] -0.4704256 -4.091606e-01 0.1364390 -5.767041e-01 -0.5099845  
## [5,] -0.3940662 -5.767041e-01 -0.5453050 4.091606e-01 0.2176104

见上方输出，其中values项为特征值，vectors矩阵的每一列为相应特征值对应的特征向量。

### **3) 随机生成一个5维满秩的方阵A， 求Y=AX的分布**

首先可知服从正态分布，下求解具体系数

A <- matrix(rnorm(25),nrow=5,ncol=5) #随机生成矩阵  
qr(A)$rank #这里检验是否是满秩的(rank=5)  
Y <- A%\*%X #按照要求生成Y  
print(Y)  
#下面计算理论值：  
mu\_Y <- A%\*%mu  
var\_Y <- A%\*%sigma%\*%t(A)  
print("Y分布均值")  
print(mu\_Y)  
print("Y协方差矩阵")  
print(var\_Y)

## [1] 5  
## [,1]  
## [1,] -3.130644  
## [2,] 2.755469  
## [3,] -3.926626  
## [4,] -1.273472  
## [5,] 4.849129  
## [1] "Y分布均值"  
## [,1]  
## [1,] -3.0200152  
## [2,] 2.5773808  
## [3,] -1.5213554  
## [4,] -0.8351595  
## [5,] 5.7466233  
## [1] "Y协方差矩阵"  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] 5.4400425 0.58298850 -2.0670493 1.7847065 -2.63501926  
## [2,] 0.5829885 1.16085191 -1.6977988 -0.3944269 0.07245331  
## [3,] -2.0670493 -1.69779878 5.0072108 0.4589221 1.64394018  
## [4,] 1.7847065 -0.39442687 0.4589221 1.1512414 -0.52256349  
## [5,] -2.6350193 0.07245331 1.6439402 -0.5225635 2.58928343

随机向量Y的取值如上面输出所示，其服从正态分布，具体均值、方差输出见上。

### **4) 求给定（X3,X4,X5）时 （X1, X2）的条件分布**

对按照题设进行分块，记为，相应的其均值为， 其协方差矩阵记为

#分块后的协方差矩阵  
sig11 <- sigma[1:2,1:2]  
sig12 <- sigma[1:2,3:5]  
sig21 <- sigma[3:5,1:2]  
sig22 <- sigma[3:5,3:5]  
  
#由条件期望公式可以计算：  
k <- sig12%\*%solve(sig22)  
mu1 <- c(1,2)  
mu2 <- c(3,4,5)  
print("mu\_1.2=mu1+sig12 inv(sig22) (x2-mu2)，其中：")  
print("mu1")  
print(mu1)  
print("sig12 inv(sig22):")  
print(k)  
print("mu2")  
print(mu2)  
  
  
#由条件方差公式可以计算：  
sig11\_2 <- sig11-sig12%\*%solve(sig22)%\*%sig21  
print("sig11\_2")  
print(sig11\_2)

## [1] "mu\_1.2=mu1+sig12 inv(sig22) (x2-mu2)，其中："  
## [1] "mu1"  
## [1] 1 2  
## [1] "sig12 inv(sig22):"  
## [,1] [,2] [,3]  
## [1,] 0.49 -1.665335e-16 5.551115e-17  
## [2,] 0.70 -1.665335e-16 0.000000e+00  
## [1] "mu2"  
## [1] 3 4 5  
## [1] "sig11\_2"  
## [,1] [,2]  
## [1,] 0.7599 0.357  
## [2,] 0.3570 0.510

条件分布仍是正态分布，其均值， 协方差矩阵 其具体取值由于没有给出具体的取值，故无法进行进一步整理，算式中其余各元素取值见上方输出。

### **5) 求Y1 与 （Y2,Y3,Y4,Y5）的复相关系数**

复相关系数计算：

# 为了方便计算，这里按照规定分布生成具有100行观测的Y  
set.seed(123) #设置种子  
Y\_expand <- mvrnorm(n=100,mu\_Y,var\_Y)  
R <- cor(Y\_expand) # R是Y的相关系数矩阵  
r\_xy <- R[1,2:5] #提取分块矩阵  
R\_xx <- R[2:5,2:5] #提取分块矩阵元素   
rho\_y.x <- sqrt(t(r\_xy)%\*%solve(R\_xx)%\*%r\_xy) #根据复相关系数公式计算  
print(rho\_y.x)

## [,1]  
## [1,] 0.9999892

复相关系数计算结果如上所示。

### **6) 求（Y2,Y3,Y4,Y5）对（Y1）线性回归的R2，并对比5）中求的复相关系数**

#提取各Y列向量  
Y1 <- Y\_expand[,1]  
Y2 <- Y\_expand[,2]  
Y3 <- Y\_expand[,3]  
Y4 <- Y\_expand[,4]  
Y5 <- Y\_expand[,5]  
#线性回归  
lm <- lm(Y1~Y2+Y3+Y4+Y5) #设置回归模型  
summary(lm) #展示回归报告

##   
## Call:  
## lm(formula = Y1 ~ Y2 + Y3 + Y4 + Y5)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0256991 -0.0060208 0.0008331 0.0071306 0.0256328   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0255548 0.0049582 5.154 1.38e-06 \*\*\*  
## Y2 1.3597182 0.0016092 844.988 < 2e-16 \*\*\*  
## Y3 0.1912730 0.0008772 218.047 < 2e-16 \*\*\*  
## Y4 1.5476887 0.0012054 1283.921 < 2e-16 \*\*\*  
## Y5 -0.8640382 0.0009075 -952.088 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01019 on 95 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.104e+06 on 4 and 95 DF, p-value: < 2.2e-16

由上方报告可见，在回归分析中的为报告输出倒数第二行的 Multiple R-squared 项，其数值要略小于其复相关系数，但较为接近。

### **7) 求给定（Y3,Y4,Y5）时 （Y1, Y2）的偏相关系数矩阵，并与（Y1，Y2）的相关系数矩阵进行对比**

由公式：可求片协方差矩阵. 再由 即可求得偏相关系数

Sig\_Y\_e <- cov(Y\_expand) #求出协方差矩阵  
#提取各分块  
Sig\_Y11 <- Sig\_Y\_e[1:2,1:2]  
Sig\_Y12 <- Sig\_Y\_e[1:2,3:5]  
Sig\_Y21 <- Sig\_Y\_e[3:5,1:2]  
Sig\_Y22 <- Sig\_Y\_e[3:5,3:5]  
#根据公式计算偏相关系数矩阵：  
Sig\_Y\_11.2 <- Sig\_Y11-Sig\_Y12%\*%solve(Sig\_Y22)%\*%Sig\_Y21 #偏协方差  
diag\_Y\_11.2 <- sqrt(diag(Sig\_Y\_11.2)) #得到对角线上的分量标准差  
sd\_Y\_11.2 <- matrix(c(diag\_Y\_11.2[1]\*diag\_Y\_11.2[1],diag\_Y\_11.2[1]\*diag\_Y\_11.2[2],diag\_Y\_11.2[2]\*diag\_Y\_11.2[1],diag\_Y\_11.2[2]\*diag\_Y\_11.2[2]),2,2) #初始化相关系数公式中的分母部分  
corr\_Y\_11.2 <- Sig\_Y\_11.2 / sd\_Y\_11.2 # rho = cov(x,y)/sqrt(var(x)\*var(y))  
#计算相关系数矩阵：  
corr\_Y\_12 <- cor(Y\_expand[,1:2])  
#输出结果  
print("偏相关系数矩阵：")  
print(corr\_Y\_11.2)  
print("相关系数矩阵：")  
print(corr\_Y\_12)

## [1] "偏相关系数矩阵："  
## [,1] [,2]  
## [1,] 1.0000000 0.9999335  
## [2,] 0.9999335 1.0000000  
## [1] "相关系数矩阵："  
## [,1] [,2]  
## [1,] 1.0000000 0.1358897  
## [2,] 0.1358897 1.0000000

### **8) 基于1）的总体分布，随机产生一组样本容量为100的样本，a)计算样本均值、样本协方差矩阵；b)验证样本协方差阵的正定性**

sample <- mvrnorm(100,mu,sigma) #生成样本  
sample\_mean <- mean(sample) #计算样本均值  
sample\_cov <- cov(sample) #计算样本方差  
sample\_cov\_eign <- eigen(sample\_cov) #计算样本协方差矩阵特征值  
print("样本均值：")  
print(sample\_mean)  
print("样本协方差：")  
print(sample\_cov)  
print("样本协方差矩阵特征值：")  
print(sample\_cov\_eign)

## [1] "样本均值："  
## [1] 3.030165  
## [1] "样本协方差："  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] 1.1189391 0.6928712 0.4675340 0.3418906 0.1743994  
## [2,] 0.6928712 0.8968692 0.5738899 0.4022014 0.2561689  
## [3,] 0.4675340 0.5738899 0.8697508 0.5752501 0.3812191  
## [4,] 0.3418906 0.4022014 0.5752501 0.8684715 0.6012226  
## [5,] 0.1743994 0.2561689 0.3812191 0.6012226 0.9793495  
## [1] "样本协方差矩阵特征值："  
## eigen() decomposition  
## $values  
## [1] 2.7484542 1.0697058 0.4383362 0.2767548 0.2001290  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.4687556 0.5717940 -0.5174379 -0.37502638 0.2119739  
## [2,] -0.4700363 0.3334439 0.1254122 0.65920479 -0.4664781  
## [3,] -0.4721632 -0.0343484 0.6493464 -0.02195075 0.5947685  
## [4,] -0.4480616 -0.3828681 0.1536140 -0.55558303 -0.5660239  
## [5,] -0.3682890 -0.6435052 -0.5208472 0.34007341 0.2516596

样本的均值、协方差矩阵如图所示。由正定矩阵的性质：对于阶实值对称矩阵，是正定矩阵等价于的所有特征值均为正。该命题再上述输出中得到数值验证。