RMSC4002 Tutorial 2

Multivariate Normal Distribution

September 18, 2017

1 The Normal Assumption

1.1 qqplot

```
> d<-read.csv("stock.csv")# read in data file
> names(d)                       # display names in d
[1] "HSBC" "CLP"  "CK"
> t1<-as.ts(d$HSBC)              # save as time series
> t2<-as.ts(d$CLP)
> t3<-as.ts(d$CK)
> u1<-(lag(t1)-t1)/t1            # compute daily percentage return
> u2<-(lag(t2)-t2)/t2
> u3<-(lag(t3)-t3)/t3

> par(mfrow=c(3,1))              # define multi-frame for ploting
> plot(u1)                       # plot u
> plot(u2)
> plot(u3)

> par(mfrow=c(3,2))
> hist(u1)             # histogram
> qqnorm(u1)           # qq-normal plot
> qqline(u1)           # add a line for reference
> hist(u2)             # if the dist is normal, the plot should
> qqnorm(u2)           # close to this line
> qqline(u2)
> hist(u3)
> qqnorm(u3)
> qqline(u3)
```

1.2 Kolmogorov-Smirnov Test

```
> ks.test(u1,pnorm)                  # KS-test for normality on u1, u2 and u3
D = 0.4773, p-value < 2.2e-16
> ks.test(u2,pnorm)
D = 0.4795, p-value < 2.2e-16
> ks.test(u3,pnorm)
D = 0.4721, p-value < 2.2e-16
```

1.3 Jarque-Bera Test

```
JB.test<-function(u) {                          # function for JB-test
  n<-length(u)                                  # sample size
  s<-sd(u)                                      # compute sd
  sk<-sum(u^3)/(n*s^3)                          # compute skewness
  ku<-sum(u^4)/(n*s^4)-3                        # excess kurtosis
  JB<-n*(sk^2/6+ku^2/24)                        # JB test stat
  p<-1-pchisq(JB,2)                             # p-value
  cat("JB-stat:",JB," p-value:",p,"\n")         # output
}
> JB.test(u1)
JB-stat: 317.7214  p-value: 0
> JB.test(u2)
JB-stat: 981.1252  p-value: 0
> JB.test(u3)
JB-stat: 136.4649  p-value: 0
```

2 Student's t(v) distribution

2.1 qqplot

```
QQt.plot<-function(u) {            # function for QQ-t plot
  su<-sort(u)                      # sort u
  n<-length(u)                     # sample size
  s<-sd(u)                         # sd
  ku<-sum(u^4)/(n*s^4)-3           # excess kurtosis
  v<-round(6/ku+4)                 # estimate df, round to the nearest integer
  i<-((1:n)-0.5)/n                 # create a vector of percentile
  q<-qt(i,v)                       # percentile point from t(v)

  hist(u)                          # histogram of u
  plot(q,su,main="qq-t plot")      # plot(q,su)
  abline(lsfit(q,su))              # add reference line
  v                                # output degree of freedom
}
v1<-QQt.plot(u1)                   # QQ-t plot for u1, u2, u3
v2<-QQt.plot(u2)                   # and save degree of freedom to v1, v2, v3
v3<-QQt.plot(u3)
```

2.2 Kolmogorov-Smirnov Test

```
> ks.test(u1,pt,v1)                    # ks-t test for u1, u2 and u3
D = 0.478, p-value < 2.2e-16
> ks.test(u2,pt,v2)
D = 0.4803, p-value < 2.2e-16
> ks.test(u3,pt,v3)
D = 0.4729, p-value < 2.2e-16
```

3 Generating Correlated Random Vectors

In univariate case, we can generate $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. However, if we want to find the σ in multivariate case, the most common way is using the Cholesky decomposition such that $\mathbf{C}'\mathbf{C} = \Sigma$ where Σ is the covariance matrix, where C is an upper triangular matrix.

- Generate $\mathbf{z}' = (z_1, ..., z_p)$ where z_i is iid $N(0, 1)$.
- Transform $\mathbf{z}$ as $\mathbf{x} = \mu + \mathbf{C}\mathbf{z}$, then $\mathbf{x}$ will be p-variate vector with covariance Σ.

4 Cholesky Decomposition

Decompose

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

using Cholesky decomposition.

Exercise (2015-16 final Q1.b)

$$R = \begin{bmatrix} 1 & 0.2599 & 0.4981 \\ 0.2599 & 1 & 0.2550 \\ 0.4981 & 0.2550 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & c_{12} & c_{13} \\ 0 & 0.9656 & 0.1300 \\ 0 & 0 & 0.8574 \end{bmatrix}$$

C is the Cholesky decomposition of R. Compute c_{12} and c_{13}.

5 Cholesky Decomposition on Covariance/Correlation

Suppose for random vector $\mathbf{X}$, the covariance matrix is Σ, the correlation matrix is $\mathbf{R}$, **variance** vector is σ. The Cholesky decomposition of Σ and $\mathbf{R}$ are $\mathbf{V}$ and $\mathbf{C}$ respectively. Define $\mathbf{D} = diag(\sigma)$ a diagonal matrix. Then we have

$$\mathbf{R} = \mathbf{C}'\mathbf{C} = \mathbf{D}^{-1/2}\Sigma\mathbf{D}^{-1/2} = (\mathbf{D}^{-1/2}\mathbf{V}')(\mathbf{V}\mathbf{D}^{-1/2})$$

Exercise (2014-15 final Q1) Let $u = (u_1, u_2, u_3)'$ be the daily relative return in percentage of three stocks: A, B, and C. Suppose the sample mean of u is $\bar{u} = (0.1314, 0.0395, 0.1031)'$ and the last value of u is $u_n = (-0.2801, -0.6536, 1.7301)'$ and the last stock price is $(89.25, 30.6, 72.25)'$. Let S be the covariance matrix of u and C be the Cholesky decomposition of S, $C'C = S$.

$$S = \begin{bmatrix} 2.5404 & 0.3257 & 1.6883 \\ 0.3257 & 1.4266 & 0.5675 \\ 1.6883 & 0.5675 & 4.1031 \end{bmatrix} \qquad C = \begin{bmatrix} 1.594 & 0.2043 & 1.0592 \\ 0 & 1.1768 & 0.2984 \\ 0 & 0 & 1.7006 \end{bmatrix}$$

- Compute the correlation matrix of u, denoted by R.
- Compute the Cholesky decomposition of R.
- A random vector $z = (-0.021, -0.312, 0.958)'$ is generated from $N_3(0, I_3)$. Based on this, compute the next simulated $u_{n+1} \sim N_2(\bar{u}, S)$ and hence compute the next simulated stock price of A, B and C respectively.