# Aula 6: Geração da distribuição normal multivariada

Prof. Dr. Eder Angelo Milani

15/05/2023

## Normal multivariada

Um vetor aleatório $X = (X_1, \dots, X_d)$ tem distribuição normal multivariada $d$-dimensional, sendo denotada por $N_d(\mu, \Sigma)$, se a densidade de $X$ é

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{-(1/2)(x-\mu)'\Sigma^{-1}(x-\mu)\}$$

sendo que $x \in R^d$, $\mu = (\mu_1, \dots, \mu_d)'$ é o vetor de médias e $\Sigma$ é uma matriz definida positiva de dimensão $d \times d$, com entradas $\sigma_{ij} = Cov(X_i, X_j)$, sendo que $\Sigma^{-1}$ é a matriz inversa de $\Sigma$.

A distribuição normal bivariada e um caso especial e é denotada por $N_2(\mu, \Sigma)$.

Valores aleatórios da distribuição $N_d(\mu, \Sigma)$ podem ser gerados em dois passos, são eles:

1º Passo - gerar $Z = (Z_1, \dots, Z_d)$, sendo que $Z_1, \dots, Z_d$ são variáveis aleatórias independentes e identicamente disctribuídas normal padrão;

2º Passo - transforme o vetor aleatório $Z$ para que ele tenha o vetor de média desejado $\mu$ e a estrutura de covariância $\Sigma$. A transformação requer fatorar a matriz de covariância $\Sigma$.

**Lembre-se**

Se $Z \sim N_d(\mu, \Sigma)$, então a transformação linear $CZ + b$ é uma normal multivariada com média $C\mu + b$ e covariância $C\Sigma C'$. Então, se $Z \sim N_d(0, I_d)$, logo

$$CZ + b \sim N_d(b, CC').$$

Disto, se $\Sigma$ pode ser fatorado como $CC'$, ou seja, $\Sigma = CC'$, para alguma matriz $C$, então

$$CZ + \mu \sim N_d(\mu, \Sigma),$$

que é a transformação requerida.

A transformação de $\Sigma$ pode ser obtida por decomposição de autovetores, fatoração de Cholesky ou alguma outra.

### Método para gerar amostra da distribuição normal multivariada

Para gerar uma amostra aleatória de tamanho $n$ da distribuição $N_d(\mu, \Sigma)$ siga os passos

1º Passo - gerar uma matriz $Z$, de dimensão $n \times d$, contendo $nd$ valores aleatórios da distribuição $N(0, 1)$;

2º Passo 2 - obter a fatoração $\Sigma = Q'Q$;

3º Passo 3 - aplicar a transformação $X = ZQ + J\mu'$, sendo que $J$ é um vetor coluna unitário de dimensão $n$;

4º Passo 4 - cada linha da matriz $X$ é uma amostra da distribuição $N_d(\mu, \Sigma)$.

## Exemplo 1

Gerar uma amostra de tamanho 1000 para a distribuição $N_2(\mu, \Sigma)$, sendo que $\mu = (1, 2)$ e $\Sigma = (1, 0.9, 0.9, 1)$.

### Solução

A fatoração de Cholesky, pode ser obtida utilizando o comando `chol`, que para esse exemplo é dado por

```
Sigma <- matrix(c(1, 0.9, 0.9, 1), nrow=2, ncol=2, byrow=T)
Sigma
```

```
##      [,1] [,2]
## [1,]  1.0  0.9
## [2,]  0.9  1.0
```

```
Q <- chol(Sigma)
Q
```

```
##      [,1]      [,2]
## [1,]    1 0.9000000
## [2,]    0 0.4358899
```

```
# verificando a fatoracao
t(Q)%*%Q
```

```
##      [,1] [,2]
## [1,]  1.0  0.9
## [2,]  0.9  1.0
```

Seguindo os passos dados acima, obtemos o seguinte código.

```
set.seed(2023)
n <- 1000
d <- 2
mu <- c(1, 2)
Sigma <- matrix(c(1, 0.9, 0.9, 1), nrow=2, ncol=2, byrow=T)

# gerando valores da normal padrao
u <- runif(n)
v <- runif(n)
z1 <- sqrt(-2*log(u))*cos(2*pi*v)
z2 <- sqrt(-2*log(v))*cos(2*pi*u)

# verificando a geracao
summary(z1)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -3.140814 -0.668514  0.020420  0.008785  0.685732  2.988493
```

```r
var(z1)
```

```
## [1] 1.00655
```

```r
summary(z2)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -3.442679 -0.714752 -0.004689 -0.036091  0.697788  3.230314
```

```r
var(z2)
```

```
## [1] 1.043411
```

```r
# criando a matriz z de valores da normal padrao
z <- cbind(z1,z2)

Q <- matrix(c(1, 0.9, 0, sqrt(0.19)), nrow=2, ncol=2, byrow=T)

X <- z%*%Q + matrix(mu, n, d, byrow=T)

head(X)
```

```
##          [,1]     [,2]
## [1,] 1.904563 2.599352
## [2,] 0.460622 1.323237
## [3,] 1.434978 2.548744
## [4,] 2.196746 2.296793
## [5,] 1.720118 3.408747
## [6,] 2.822743 3.766575
```

```r
# verificando
colMeans(X)
```

```
## [1] 1.008785 1.992175
```

```r
var(X[,1])
```

```
## [1] 1.00655
```

```r
var(X[,2])
```

```
## [1] 0.9981939
```

```r
cov(X)
```

```
##          [,1]      [,2]
## [1,] 1.006550 0.8973620
## [2,] 0.897362 0.9981939
```

```r
# scatterplor 3D
library(scatterplot3d)
```

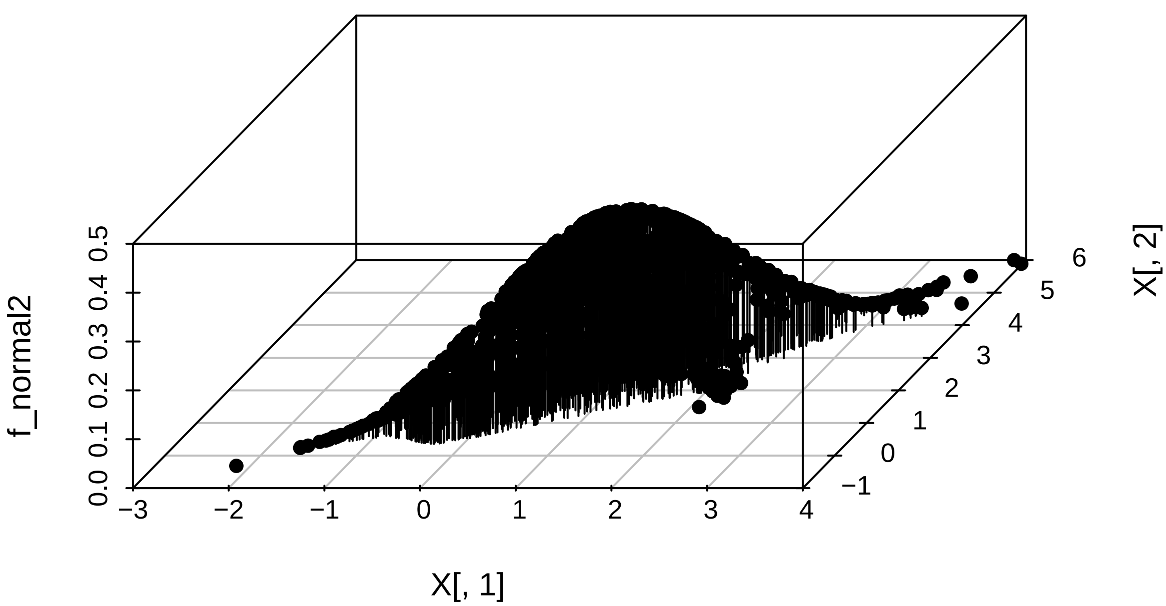
```
## Warning: package 'scatterplot3d' was built under R version 4.2.2
```

```r
# para calcular a densidade da normal multivariada
require(mvtnorm)
```

```
## Carregando pacotes exigidos: mvtnorm
```

```r
f_normal2 <- dmvnorm(X, mean = mu, sigma = Sigma)

scatterplot3d(x=X[,1],
              y=X[,2],
              f_normal2,
              pch=16,
              angle=60,
              type='h')
```




### Exercícios

1. Gerar uma amostra de tamanho 1000 para a distribuição $N_2(\mu, \Sigma)$, sendo que $\mu = (1, 0)$ e $\Sigma = (3, 0.8, 0.8, 1)$. Faça o gráfico em 3D.

2. Gerar uma amostra de tamanho 1000 para a distribuição $N_3(\mu, \Sigma)$, sendo que $\mu = (1, 0, 2)$ e $\Sigma = (3, 0, 0.8, 0, 1, 0.5, 0.8, 0.5, 1)$. Apresente o resumo de cada coordenada e a matriz de variânica-covariância amostral.