Tutorials on Multivariate Normal Distribution

Module I (Lecture 5)

David Raj Micheal

August 2018

Contents

1 Solve all the following problems

1. Given the data matrix

$$X = \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$$

(a) Graph the scatter plot in $p = 2$ dimensions. Locate the sample mean on your diagram

(b) Sketch the $n = 3$dimensional representation of the data, and plot the centered vectors. That is plot the vectors from $X - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} \overline{x_1} & \overline{x_2} \end{pmatrix}$ where e is the vector of ones. These vectors are also called deviation vectors.

2. Given the data matrix

$$X = \begin{bmatrix} 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{bmatrix}$$

(a) Graph the scatter plot in $p = 2$ dimensions. Locate the sample mean on your diagram.
(b) Sketch the $n = 3$dimensional representation of the data, and plot the centered vectors. That is plot the vectors from $X - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} \overline{x_1} & \overline{x_2} \end{pmatrix}$ where e is the vector of ones. These vectors are also called deviation vectors.

3. Consider the data matrix $X == \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$. We have $n = 3$ observations on $p = 2$ variables x_1 and $x_24.

Form the linear combinations

$$c^T x = \begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -x_1 + 2x_2$$

and

$$b^T x = \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1 + 3x_2.$$

(a) What the mean and variance of the random variable $b^T x$?

(b) What the mean and variance of the random variable $c^T x$?

(c) What is the covariance between $b^T x$ and $c^T x$?

4. Consider the data matrix $X = \begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline & 12 & 2 & 5 \\ & 4 & 1 & 6 \\ & 4 & 0 & 4 \end{array}$.

(a) What the mean and variance of the random variable $2x_1 + 2x_2 - x_3$?

(b) What the mean and variance of the random variable $x_1 - x_2 - 3x_3$?

(c) What is the covariance between $2x_1 + 2x_2 - x_3$ and $x_1 - x_2 - 3x_3$?

5. Calculate the generalized variance and total variance of the following data matrices.

(a) $X = \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$

(b) $X = \begin{bmatrix} 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{bmatrix}$

6. Sketh the solid ellipsoids $(x - \overline{x})S^{-1}(x - \overline{x}) \leq 1$ for the three matrices

$$S = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, \quad S = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}, \quad S = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}.$$

(Note that these matrices have the same generalized variance.)

2 Plotting Confidence ellipses using R

```
#Generating  a bivariate data  of 1000 observation
# with given mean vector and covariance matrix using MASS package.
library(MASS)
m = c(2,3)
S = matrix(c(2,1,1,2), ncol = 2, byrow = T)
X =mvrnorm(n = 1000,
           mu = m,
           Sigma = S,
           tol = 1e-6,
           empirical = TRUE)
```

```
library('car')
```

```
## Loading required package: carData
```

```
#plotting the scatter plot
plot(X, xlab = 'x', ylab = 'y', asp = 1)
#Adding the 90% confidence region
ellipse(center = colMeans(X),
        shape = cov(X),
        radius = sqrt(qchisq(.90, df=2)),
        col = 'red')
#Adding the 95% confidence region
ellipse(center = colMeans(X),
        shape = cov(X),
        radius = sqrt(qchisq(.95, df=2)),
```

```r
        col = 'darkgreen')
#Adding the 99% confidence region
ellipse(center = colMeans(X),
        shape = cov(X),
        radius = sqrt(qchisq(.99, df=2)),
        col = 'brown4')
#Adding the legend to the plot
legend('topleft',
       title="Confidence Region",
       legend=c("90%", "95%", "99%"),
       col=c("red", "green","brown4"),
       lty="solid",
       cex=0.8)
```