Module 1 (Lecture 6)

Properities & MLE of Normal Distribution

David Raj Micheal

Contents

1 Properties of Normal Distribution **1**

2 Estimation in Multivariate Normal- MLE **2**

3 The sampling distribution of sample mean vector $\overline{x}$ and S **2**

3.1 Sampling distribution of sample mean 3

3.2 Sampling distribution of sample covariance matrix 3

3.3 Multivariate central limit theorem 3

1 Properties of Normal Distribution

Let $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ be a random vector and $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$ a constant vector. Then the following properties are very easy to prove and left as excercise. Proving the normallity, wherever required, involves matrix algebra techniques which if out the scope of this course. So, readers are expected to drive only the mean and variance-covariance matrices, wherever applicable.

1. If $x \sim N_p(\mu, \Sigma)$ then
$$a^T x \sim N(a^T \mu, a^T \Sigma a).$$

2. If $a^T x \sim N(a^T \mu, a^T \Sigma a)$ for every a then
$$x \sim N_p(\mu, \Sigma).$$

3. If $x \sim N_p(\mu, \Sigma)$ then the q linear combinations
$$A_{q\times p} x \sim N_q(A\mu, A\Sigma A^T).$$

4. If $x \sim N_p(\mu, \Sigma)$ then
$$x + a \sim N_p(\mu + a, \Sigma).$$

5. All subsets of x are normally distributed. That is, if $x = \left(\begin{array}{c} x_1 \\ \vdots \\ x_q \\ \hline x_{q+1} \\ \vdots \\ x_p \end{array}\right)$, $\mu = \left(\begin{array}{c} \mu_1 \\ \vdots \\ \mu_q \\ \hline \mu_{q+1} \\ \vdots \\ \mu_p \end{array}\right)$ and $\Sigma = \left(\begin{array}{c|c} \Sigma_{11_{q\times q}} & \Sigma_{12_{q\times (p-q)}} \\ \hline \Sigma_{21_{(p-q)\times q}} & \Sigma_{22_{(p-q)\times (p-q)}} \end{array}\right)$, then

$$\begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} \sim N_q \left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_q \end{pmatrix}, \Sigma_{11_{q\times q}} \right)$$

and

$$\begin{pmatrix} x_{q+1} \\ \vdots \\ x_p \end{pmatrix} \sim N_{p-q} \left(\begin{pmatrix} \mu_{q+1} \\ \vdots \\ \mu_p \end{pmatrix}, \Sigma_{22_{(p-q)\times(p-q)}} \right).$$

6. If $x_{p\times 1}$ and $y_{q\times 1}$ are independent then

$$\mathrm{cov}(x, y) = 0, \text{ a } p \times q \text{ matrix of zeroes.}$$

7. If $\begin{pmatrix} x_{p\times 1} \\ y_{q\times 1} \end{pmatrix} \sim N_{p+q} \left(\begin{pmatrix} \mu_{p\times 1} \\ \mu_{q\times 1} \end{pmatrix}, \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right) \right)$ then $x_{p\times 1}$ and $y_{q\times 1}$ are independent if and only if $\Sigma_{12} = 0$.
8. If $\begin{pmatrix} x_{p\times 1} \\ y_{q\times 1} \end{pmatrix} \sim N_{p+q} \left(\begin{pmatrix} \mu_{p\times 1} \\ \mu_{q\times 1} \end{pmatrix}, \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right) \right)$, and $|\Sigma_{22}| > 0$, then the conditional distribution of $x_{p\times 1}$ given that $y_{q\times 1}$ is normal and has

$$\text{Mean } = \mu_{p\times 1} + \Sigma_{12}\Sigma_{22}^{-1}(y_{q\times 1} - \mu_{q\times 1})$$

and

$$\text{Variance-Covariance } = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

2 Estimation in Multivariate Normal- MLE

When a distribution such as the multivariate normal is assumed to hold for a population, estimates of the parameters are often found by the method of maximum likelihood. This technique is conceptually simple: The observation vectors $y_1, y_2, \ldots, y_n$ where

$$y_i = \begin{pmatrix} x_{i1} & x_{i2} & \ldots & x_{ip} \end{pmatrix}$$

are considered to be known and values of μ and Σ are sought that maximize the joint density of the y's, called the *likelihood function*. For the multivariate normal, the maximum likelihood estimates of μ and Σ are

$$\begin{aligned} \hat{\mu} &= \overline{x}, \\ \hat{\Sigma} &= \frac{1}{n} Y^T Y, \ Y \text{ is the matrix of centered observations} \\ &= \frac{n-1}{n} S \end{aligned}$$

where S is the sample variance-covariance matrix.

Remark Since $\hat{\Sigma}$ has divisor n instead of $n-1$, it is a biased estimate.

3 The sampling distribution of sample mean vector $\overline{x}$ and S

The tentative assumption that $y_1, y_2, \ldots, y_n$ where

$$y_i = \begin{pmatrix} x_{i1} & x_{i2} & \ldots & x_{ip} \end{pmatrix}$$

constitute a random sample from a multivariate normal population with mean μ and Σ completely determines the sampling distributions of $\overline{x}$ adn S. Here we present the results on the sampling distributions of $\overline{x}$ adn S by drawing a parallel with the familiar univariate conclusions.

3.1 Sampling distribution of sample mean

In the univariate case ($p = 1$), we know that $\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ where μ, σ^2, n are population mean, population variance and sample size respectively.

The result for the multivariate case ($p \geq 2$) is analogous in that

$$\overline{x} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right)$$

where μ, Σ, n are population mean vector, population variance-covariance matrix and sample size respectively.

3.2 Sampling distribution of sample covariance matrix

In the univariate case ($p = 1$), for the sample variance

$$(n-1)s^2 \sim \sigma^2\chi^2_{n-1}.$$

In turn, this chi-square is the distribution of a sum of squares of independent standard noramal random variables. That is , $(n-1)s^2$ is distributed as $\sigma^2(z_1^2 + \cdots + z_{n-1}^2) = (\sigma z_1)^2 + \ldots (\sigma z_{n-1})^2$. The individual terms σz_i are independently distributed as $N(0, \sigma^2)$. This discussion can be suitably generalized to the basic sampling distribution for the sample covariance matrix.

Note that in the sample covariance matrix there are p variances and $\frac{p(p-1)}{2}$ covariances, for a total of

$$p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$$

distinct entries. The joint distribution of these $\frac{p(p+1)}{2}$ distinct variables in $W = (n-1)S$ is called the *Wishart distribution*, denoted by

$$W_p(n-1, \Sigma)$$

where $n-1$ is the degrees of freedom. Specifically,

$$\begin{aligned} W_p(n-1, \Sigma) &= \text{Wishart distribution with } n-1 \text{ degrees of freedom} \\ &= \text{ distribution of } \sum_{i=1}^{n-1} z_i z_i^T \end{aligned}$$

where the z_i are each independently distributed as $N_p(0, \Sigma)$.

Hence, when $\overline{x}$ is considered instead of μ then

$$(n-1)S \sim W_p(n-1, \Sigma).$$

Remark: When sampling from a multivariate normal distribution, $\overline{x}$ and S are independent.

3.3 Multivariate central limit theorem

Let $y_1, \ldots, y_n$ be independent observations from any population with mean μ and finite variance-covariance matrix Σ. Then

$$\sqrt{n}(\overline{x} - \mu) \text{ has an approximate } N_p(0, \Sigma)$$

for large sample sizes. Here n should also be large relative to p.