Econometrics I
TA Session 3

Jukina HATAKEYAMA *

May 2, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 Multivariate Normal Distribution

We provide important reminders of definitions and properties (e.g. the multivariate normal distribution) for the Econometrics class. Before explaining the multivariate normal distribute function, we introduce the moment generating function(: mgf), whose property is useful for the analysis of the multivariate normal distribute function.

1.1 Moment Generating Function

Suppose that a random variable X follows a normal distribution $X \sim N_{\mathbb{R}}(\mu, \sigma^2)$. In this case, the probability density function of X is given as follows.

Definition 1.1. We can say that a random variable X follows a normal distribution if its probability density function (: pdf) takes a form below:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2). \tag{1}$$

A useful property of a normal distribution is its preservation under linear transformation.

Theorem 1.2. Assume that X is a normal distributed with mean μ and variance σ^2. Then, the linear transformation of X, $Y = aX + b$, follows a distribution such that $Y \sim N(a\mu + b, a^2\sigma^2)$.

We can confirm this fact by Change of Valiables. If these two random variables follow the same distribution, i.e. $F_x(z) = F_y(z)$ for all z, then their moment generation function are certainly equal to each other, i.e. $M_x(t) = M_y(t)$ in a neighborhood of 0.

Definition 1.3. The **moment generating function** of a continuous random variable X, with its pdf $f(x)$, is defined as:

$$m_x(t) = \mathbb{E}[e^{tX}] = \int e^{tx} f(x)dx. \tag{2}$$

Then, we can derive the m^{th} moment of X by using this function. The first derivative of $m_x(t)$ with respect to t can specify the mean of X and the second derivative determines the second moment, $\mathbb{E}[X^2]$.

For instance, the m.g.f. of X which follows a distribution $X \sim N(\mu, \sigma^2)$ is represented as follows:

$$\mathbb{E}[\exp(tX)] = \exp(\mu t + \frac{1}{2}\sigma^2 t^2). \tag{3}$$

Actually, numerous distributions do not have moment-generating function. Instead, we usually use characteristic function which exists for every distribution. More details about the mgf and the normal distribution is written in the statistics textbooks.

1.2 Definition of Multivariate Normal Distribution

Suppose a n-dimensional random vector X whose mean vector and variance-covariance matrix is given as follows:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad and \quad \Sigma = \mathbb{E}[(x-\mu)(x-\mu)'] = \begin{bmatrix} Var(x_1) & Cov(x_1,x_2) & \cdots & Cov(x_1,x_n) \\ Cov(x_2,x_1) & Var(x_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ Cov(x_n,x_1) & Cov(x_n,x_2) & \cdots & Var(x_n) \end{bmatrix}.$$

Then, we have the following definition.

Definition 1.4. Let X a n-dimensional random vector. When X follows **a multivariate normal distribution**, denoted as $X \sim N_{\mathbb{R}^n}(\mu, \Sigma^2)$, its pdf is defined as:

$$f(X) = \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}}}\exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)) \tag{4}$$

.

Next, we want to mention about the properties of multivariate normal distribution. Assume that $Y = (Y_1, Y_2, ..., Y_n)'$ is given as follows:

$$Y = aX + b. \tag{5}$$

Where $b \in \mathbb{R}^{n\times 1}$ is a non-stochastic vector and $a \in \mathbb{R}^{n\times n}$ is a non-stochastic regular matrix. In this case, using Change of Valiables, we can derive its distribution.:

$$Y \sim N_{\mathbb{R}^n}(a\mu + b, a\Sigma a')$$

2 Ordinary Least Squares

Consider a following regression model:

$$y_i = b_1 + b_2 x_i + u_i. \tag{6}$$

for $i = (1, \dots, n)$ representing individuals. We stack above equation for i, then we have

$$y = xb + u. \tag{7}$$

We define $y = (y_1, y_2, \dots, y_n)' \in \mathbb{R}^{n\times 1}$, $z = (\imath, x)$, $x = (x_1, x_2, \dots, x_n)' \in \mathbb{R}^{n\times 1}$, $\imath = (1, 1, \dots, 1)' \in \mathbb{R}^{n\times 1}$, $b = (b_1, b_2)'$ and $u = (u_1, u_2, \dots, u_n)' \in \mathbb{R}^{n\times 1}$.

2.1 Assumptions of the Classical Linear Regression Model

The OLS estimator is based on important assumptions. If these assumptions do not hold (especially exogeneity), we must rely on other methods such as GLS and/or IV.

1. Linear model
2. Full Rank: The data matrix z has full rank.
3. Exogeneity of the Independent Variables: $E[u|z] = 0$
4. Homoscedasticity and No-Autocorrelation: $E[u_i^2|z] = \sigma^2$ *and* $E[u_i u_j|z] = 0$ for all i, j.

2.2 Derivation of the OLSE

When we estimate (6) or (7), the OLS is applied under some assumptions.

Definition 2.1. The OLS estimator is derived as a minimum distance between y and the vectorial space spanned by i and $\underline{x}$ for the Euclidian norm:

$$\hat{b} = \arg\min_b ||y - xb||_2^2 = \arg\min_b (y - xb)'(y - xb) \tag{8}$$

$$= \arg\min_b \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2. \tag{9}$$

$\hat{b} = (\hat{b_1}, \hat{b_2})'$ is **the OLS estimator** of b.

Next, we explain the meaning of the above definition. Consider that we observed data and estimate (6). Then, it is reasonable to reduce the distance between observed data and our estimation of $\mathbb{E}[y_i|x_i] = \hat{b_1} + \hat{b_2} x_i$. A convenient method is the OLS estimation, which minimizes the sum of the squared of the residuals:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (\hat{b_1} + \hat{b_2} x_i)]^2. \tag{10}$$

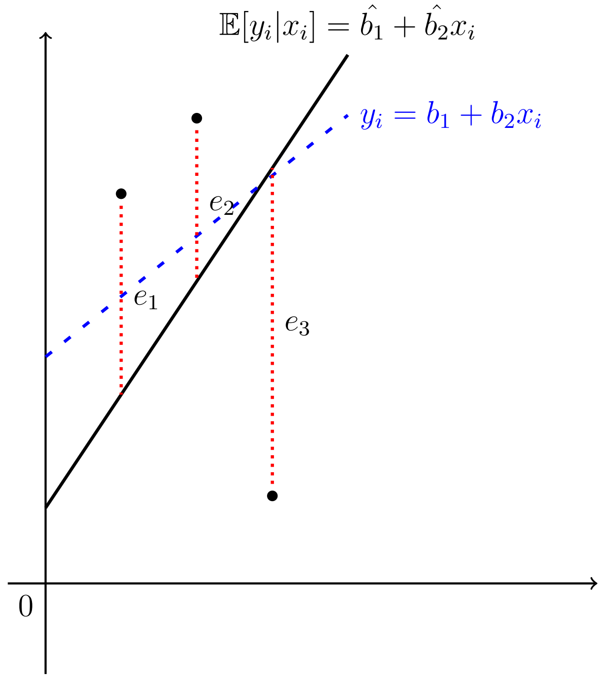

Figure 1: Meaning of OLS

Consider the stacked version. The loss function is given as follows:

$$l(b_1, b_2) = (y - xb)'(y - xb). \tag{11}$$

The first order condition and the second order condition are derived as follows:

$$\nabla_b l(b_1, b_2) = -2x'(y - xb) = 0.$$

$$\nabla^2_{bb'} l(\cdot) = 2x'x > 0 \tag{12}$$

Keep in mind that the design matrix is invertible. A matrix $(x'x)^{-1}$ is positive definite and full rank.

Theorem 2.2. Suppose that $\imath, \underline{x}$ are independent, then a unique OLSE is given as follows:

$$\hat{b} = (x'x)^{-1}x'y. \tag{13}$$

2.3 Properties of OLSE

Under suitable assumptions, the OLSE has some important properties.

1. **Unbiasedness**
2. **Consistency**
3. **Minimum Variance**

2.3.1 Unbiasedness

Suppose the regression model like (7). Then, the conditional expectation of OLSE is:

$$\begin{aligned} \mathbb{E}[\hat{b}|x] &= \mathbb{E}[(x'x)^{-1}(x'y)|x] \\ &= \mathbb{E}[(x'x)^{-1}x'(xb + u)|x] \\ &= b + \mathbb{E}[(x'x)^{-1}x'u|x] \\ &= b + (x'x)^{-1}\mathbb{E}[x'u|x] = b. \end{aligned} \tag{14}$$

In the same manner, we can derive the conditional covariance matrix of $\hat{b}$:

$$\begin{aligned} Var[\hat{b}|x] &= \mathbb{E}[(\hat{b} - \underline{b})(\hat{b} - \underline{b})'|x] \\ &= \mathbb{E}[(x'x)^{-1}x'uu'x(x'x)^{-1}|x] \\ &= (x'x)^{-1}x'\mathbb{E}[uu'|x]x(x'x)^{-1} \\ &= \sigma^2(x'x)^{-1} \end{aligned} \tag{15}$$

.

2.3.2 Consistency

A consistent parameter $\hat{b}$ means that:

$$\lim_{n\to\infty} \mathbb{P}(\|\hat{b} - b\| > \epsilon) = 0, \ \forall \epsilon > 0.$$

In practice, we can confirm whether $\hat{b}$ is consistent or not as follows:

$$\begin{aligned} \hat{b} &= (x'x)^{-1}x'(xb+u) \\ &= b + (x'x)^{-1}x'u \\ &= b + (\frac{1}{n}x'x)^{-1}(\frac{1}{n}x'u). \end{aligned} \tag{16}$$

The third line of above equation is the application of Slutsky's theorem. Additionaly, we use LLN in the third line of (16):

$$\frac{1}{n}x'u \xrightarrow[n\to\infty]{\text{p}} E[x'u] = 0. \tag{17}$$

This is called as a moment of probability convergence. Assume there exists a regular matrix Q such that:

$$(\frac{1}{n}x'x)^{-1} \xrightarrow[n\to\infty]{\text{p}} E[x'x]^{-1}. \tag{18}$$

From these equations, we can derive the probability convergence of $\hat{b}$ as follows:

$$\hat{b} \xrightarrow[n\to\infty]{\text{p}} b + E[x'x]^{-1} \times 0 = b. \tag{19}$$

2.3.3 Useful Operators

Definition 2.3. $P_z = x(x'x)^{-1}x'$ is the operator to make a projection vector which can derive $\hat{y} = x\hat{b} = x(x'x)^{-1}x'y = P_zy$. In addition, $M_z = I - P_z$ is the operator to make a residual because $e = y - \hat{y} = y - P_zy = M_zy$.

Note that these operators have particular properties.

Corollary 2.4.

- P_z and M_z are symmetric and idempotemt.
- $P_zx = x$ and $M_zx = 0$ is derived.
- $P_zM_z = M_zP_z = 0$

2.3.4 Minimum Variance

Generally, the OLSE of the classical regression model is the best linear unbiased estimator (: BLUE). In other words, the OLSE has the smallest variance among the estimators which have linearity and unbiasedness. This is a reason why we adopt OLS to the classical regression model.

Theorem 2.5. (Gauss-Markov Theorem) OLSE is a **BLUE** when we estimate a classic regression model.

This theorem is also applied to the $n \geq 3$ variable case. This sketch of proof is explained in chapter 4 of Greene(2011).

Let $b_0 = Cy$ be another linear unbiased estimator of $\underline{b}$, where C is a $2 \times n$ matrix. Because b_0 is unbiased, we can derive its expectation as follows:

$$E[Cy|x] = E[Cx\underline{b} + Cu|x] = \underline{b}. \tag{20}$$

This equation implies that $Cx = I$. There are many canditates of C. For example, we can consider the case of $C = [\underline{z_0}^{-1}|0]$ where $\underline{z_0}^{-1}$ is the inverse for first k_0 rows of x [1]. The variance-covariance matrix of b_0 equalls to $\sigma^2 CC'$ and we can get following result:

$$\begin{aligned} Var[b_0|x] - Var[\hat{b}|x] &= \sigma^2 CC' - \sigma^2(x'x)^{-1} \\ &= \sigma^2 C(P_z + M_z)C' - \sigma^2(x'x)^{-1} \\ &= \sigma^2[Cx(x'x)^{-1}x'C + CM_zC'] - \sigma^2(x'x)^{-1} \\ &= \sigma^2 I_z(x'x)^{-1}I_z + \sigma^2 CM_zC' - \sigma^2(x'x)^{-1} \\ &= \sigma^2 CM_zC' \geq 0 \end{aligned} \tag{21}$$

Because M_z is a symmetric and idempotent matrix, $\sigma^2 CM_zC'$ is positive-semidefinite. Therfore, $Var[b_0|x] \geq Var[\hat{b}|x]$ holds.

[1] We can say that this method is also applied to the general case, OLS with $k \geq 3$ variables. In this class, we should assume $k_0 = 1$.

3 R Exercise

In this section, we will explain how to use R.[2] Now, we use data of the real estate price and the location of each building in HongKong.[3] Consider the following regression model:

$$\begin{aligned}(price)_i &= a + b_1(housage)_i + b_2(distance\ to\ the\ station)_i \\ &\quad + b_3(a\ number\ of\ convinience\ stores)_i + b_4(latitude)_i + b_5(longitude)_i + u_i. \qquad (22)\end{aligned}$$

Table 1: A multiple regression model of real estate and location

	Dependent variable:
	price
houseage	-0.269^{***}
	(0.039)
distance	-0.004^{***}
	(0.001)
conviniencestore	1.163^{***}
	(0.190)
latitude	237.767^{***}
	(44.948)
longtitude	-7.805
	(49.149)
Constant	$-4{,}945.595$
	(6,211.157)
Observations	414
R^2	0.571
Adjusted R^2	0.566
Residual Std. Error	8.965 (df = 408)
F Statistic	108.682^{***} (df = 5; 408)
Note:	*p<0.1; **p<0.05; ***p<0.01

R code is given in the appendix. The "lm" function is a default function of R.

[2] The data set used in this class is uploaded in the UCI Machine Learning Repository.

[3] Yeh, I. C., & Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. Applied Soft Computing, 65, 260-271.

```
#Firstly, please decide the directory where you put a data set.
#Yeh, I. C., & Hsu, T. K. (2018).
#Building real estate valuation models with comparative approach
#through case-based reasoning.
#Applied Soft Computing, 65, 260-271.

rm(list=ls(all=TRUE)) #A kind of magic spell

install.packages("stargazer")
#If you use some packages which are not installed in your R,
#please write this command on the script.
#The stargazer package includes a function to output the result.


library(stargazer)        #A command to call packages
variableset<-read.csv("Real estate valuation data set.csv",header=T)
#A command to read a data set with a header.
#We recommend to use a csv file.
variableset<-data.frame(variableset)


houseage<-variableset[,3]
distance<-variableset[,4]
conviniencestore<-variableset[,5]
latitude<-variableset[,6]
longtitude<-variableset[,7]
price<-variableset[,8]

b_1<-lm(price~houseage + distance +
conviniencestore + latitude + longtitude, data=variableset)


#A default function to estimate a regression model.

stargazer(b_1,
title="A multiple regression model of real estate and location")

#If you want to write a table by tex, stargazer is useful.
#When you want to check the result quickly,
#you can add a command type="text" in the above function.
```