Econometrics I
TA Session 6

Jukina HATAKEYAMA*

May 21, 2024

Contents

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

1 Review of Some Concepts for a Multivariate Normal Random Variable

Theorem 1.1 (Multivariate Normal Distribution)**.** Let the vector $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathbb{R}^k$ be the set of n random variables, μ their mean vector, and Σ their variance–covariance matrix. The general form of the joint distribution is given by

$$f(\mathbf{x}) = (2\pi)^{-k/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}.$$

In the special case where $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathbb{R}^k$ and x_i for $i \in \{1, \ldots, k\}$ is an i.i.d.random variable with mean 0 and finite variance $\sigma_i^2 < \infty$, we have

$$f(\mathbf{x}) = (2\pi)^{-k/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}\right\}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_k^2).$$

The proof is shown in Appendix A.

Adding to the theorem, we can construct the *characteristic function* and *moment generating function* for this random variable as follows.

Theorem 1.2 (Characteristic Function and Moment Generating Function)**.** For a random variable $\mathbf{x}\colon \Omega \to \mathbb{R}^k$ which follows a multibvariate normal distribution with mean $\mu \in \mathbb{R}^k$ and variance–covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$, by using a parameter $\theta \in \mathbb{R}^k$, we can define a function $\varphi_{\mathbf{X}}\colon \mathbb{R}^k \to \mathbb{C}$:

$$\varphi_{\mathbf{x}}(\theta) := \mathbb{E}[e^{i\theta'\mathbf{x}}] = \exp\left(i\theta'\mu - \frac{1}{2}\theta'\Sigma\theta\right), \tag{1}$$

which is called the *characteristic function* of $\mathbf{x}$. In addition, there exists a function $\phi_{\mathbf{x}}\colon \mathbb{R}^k \to \mathbb{R}$ defined as

$$\phi(\theta) = \mathbb{E}\left[\exp\left(\theta'\mathbf{x}\right)\right] = \exp\left(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta\right), \tag{2}$$

which is called the *moment generating function* of $\mathbf{x}$.

2 Multiple Regression Model

$$y_i = b_1 x_{i,1} + \cdots + b_{i,k}x_{i,k} + u_i = x_i b + u_i, \tag{3}$$

where $x_i = (x_{i,1}, \ldots, x_{i,k})$ is a $1 \times K$ vector for $i \in \{1, \ldots, n\}$ and $b = (b_1, \ldots, b_k)'$ is a $k \times 1$

vector. Denoting by

$$
\begin{aligned}
y &:= (y_1, \ldots, y_n)' \in \mathbb{R}^n, \\
u &:= (u_1, \ldots, u_n)' \in \mathbb{R}^n, \\
x &:= \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{pmatrix} \in \mathbb{R}^{n \times k},
\end{aligned}
$$

we can write the stacked regression system as follows:

$$
y = xb + u \left(\iff \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\in \mathbb{R}^n} = \underbrace{\begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{pmatrix}}_{\in \mathcal{M}_{n \times k}(\mathbb{R})} \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}}_{\in \mathbb{R}^k} + \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}}_{\in \mathbb{R}^n} \right).
$$

2.1 Derivation of the OLS Estimator

In this subsection, we derive the OLS estimator, which is defined as follows.

Definition 2.1 (Ordinary Least Squares (OLS) Estimator for a Multivariate Regression Model)**.** The OLS estimator $\hat{b}$ for a multivariate regression model is a vector $\hat{b} \in \mathbb{R}^k$ which satisfies the minimum distance between y and the vectorial space of $\mathbb{R}^n$ generated by X for the Euclidian norm:

$$
\hat{b} = \arg\min_b \|y - Xb\|_2^2 = \arg\min_b (y - Xb)'(y - Xb) = \arg\min_b \sum_{i=1}^{n} \left(y_i - \sum_{l=1}^{k} b_l X_{i,l} \right)^2.
$$

The residual is defined by $\hat{u}_i = y_i - \hat{y}_i = \sum_{l=1}^{k} \hat{b}_l X_{i,l}$. Therefore, the above definition can be written as

$$
\hat{b} = \arg\min_b \sum_{i=1}^{n} \hat{u}_i^2.
$$

This implies that the OLS estimator is an estimator which minimizes the sum of the residual sum of squares. The OLS estimator obtained from the above definition becomes as follows.

Theorem 2.1 (Ordinary Least Squares (OLS) Estimator for a Multivariate Regression Model)**.** Suppose

H1: $X_1, \ldots, X_k$ are independent,

then the OLS estimator $\hat{b}$ exists uniquely and satisfies

$$
\hat{b} = (X'X)^{-1}(X'y). \tag{4}
$$

Proof. To obtain the OLS estimator, we have to confirm the first and second order condition for the minimization problem of the following loss function $S(b)$:

$$
\arg\min_b \|y - Xb\|_2^2 =: \arg\min_b S(b).
$$

The first order condition becomes

$$\begin{aligned}\nabla_b \|y - X\hat{b}\|_2^2 &= \nabla_b \left(y - X\hat{b}\right)' \left(y - X\hat{b}\right) \\ &= -2X'(y - X\hat{b}) = \mathbf{0}.\end{aligned}$$

The OLS estimator, denoted as $\hat{b}$, satisfies this equation, and hence

$$(X'X)\,\hat{b} = X'y.$$

From the assumption **H1**, the inverse matrix $(X'X)^{-1}$ exists, with $X = (X_1', \dots, X_k')' \in \mathcal{M}_{n\times k}(\mathbb{R})$, whose columns are independent so that $X'X$ is a full rank matrix, and therefore we can obtain the OLS estimator in the form of (4). The second order condition becomes

$$\nabla^2_{b,b'} \|y - X\hat{b}\|_2^2 = 2X'X > 0.$$

By assumption **H1**, $X'X$ is a positive definite matrix. This shows that the loss function $S(b)$ has a minimum at the OLS estimator $\hat{b}$. □

From this theorem, we can confirm that the OLS estimator expressed as (4) is a random variable since we can rewite it as follows:

$$\hat{b} = b + (X'X)^{-1} X'u. \tag{5}$$

Therefore, we can consider the mean and variance of the OLS estimator. First, we see the mean of the OLS estimator, which will be used to prove that the OLS estimator is an unbiased estimator.

Proposition 2.1 (Mean of the OLS Estimator)**.** Suppose

H2: $\mathbb{E}[u_i | X] = 0$ for all $i \in \{1, \dots, n\}$,

then the conditional expectation of the OLS estimator $\hat{b}$ becomes

$$\mathbb{E}[\hat{b}|X] = b. \tag{6}$$

Proof. Calculating the expectation of $\hat{b}$ yields

$$\begin{aligned}\mathbb{E}[\hat{b}|X] &= \mathbb{E}\left[(X'X)^{-1}(X'y)\,\middle|\,X\right] \\ &= b + \mathbb{E}\left[(X'X)^{-1} X'u \middle| X\right] \\ &= b + (X'X)^{-1} X' \underbrace{\mathbb{E}\left[u|X\right]}_{=0(\text{from } \mathbf{H2})} \\ &= b,\end{aligned}$$

which proves (6). □

Remark 2.1 (Unconditional Expectation of the OLS estimator)**.** The conditional expectation of the OLS estimator is same as the unconditional one:

$$\mathbb{E}[\hat{b}] = b.$$

from the *law of iterated expectation* mentioned below.

Lemma 2.1 (Law of Iterated Expectation)**.** For any two random variables x and y,

$$\mathbb{E}[y] = \mathbb{E}_x\left[\mathbb{E}[y|X]\right], \tag{7}$$

where $\mathbb{E}_x$ is the expectation over the values of x.

The proof is omitted (left as an exercise for students). From this, we have

$$\mathbb{E}[\hat{b}] = \mathbb{E}\big[\underbrace{\mathbb{E}[\hat{b}|X]}_{=b}\big] = \mathbb{E}[b] = b.$$

The variance of the OLS estimator, which is the minimum variance in the class of linear OLS estimator, becomes as follows.

Proposition 2.2 (Variance of the OLS Estimator)**.** Suppose [**H1**–**H2**] holds and assume

H3 $\mathbb{V}[u_i|X] = \sigma^2$ for all $i \in \{1, \ldots, n\}$;

H4 $\mathbb{E}[u_i u_j|X] = 0$ for all $i \neq j$ and $i, j \in \{1, \ldots, n\}$,

the conditional variance of the OLS estimator $\hat{b}$ becomes

$$\mathbb{V}[\hat{b}|X] = \sigma^2 \left(X'X\right)^{-1}, \tag{8}$$

and the unconditional variance becomes

$$\mathbb{V}[\hat{b}] = \sigma^2 \mathbb{E}\left[\left(X'X\right)^{-1}\right]. \tag{9}$$

Proof. From the Eq, (5) and Eq, (6),

$$\hat{b} - \mathbb{E}[\hat{b}|X] = \hat{b} - b = \left(X'X\right)^{-1} X'u.$$

Therefore,

$$\begin{aligned}
\mathbb{V}[\hat{b}|X] &= \mathbb{E}\left[\left(\hat{b} - \mathbb{E}[\hat{b}|X]\right)\left(\hat{b} - \mathbb{E}[\hat{b}|X]\right)' \Big| X\right] \\
&= \mathbb{E}\left[\left(X'X\right)^{-1} X'uu'X \left(X'X\right)^{-1} \Big| X\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{E}\left[uu'|X\right] X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1}.
\end{aligned}$$

This implies (8) holds. Thus,

$$\begin{aligned}
\mathbb{V}[\hat{b}] &= \mathbb{E}\left[\mathbb{V}[\hat{b}|X]\right] + \mathbb{V}\left[\mathbb{E}[\hat{b}|X]\right] \\
&= \mathbb{E}\left[\sigma^2 \left(X'X\right)^{-1}\right] + \underbrace{\mathbb{V}[b]}_{=0} \\
&= \sigma^2 \mathbb{E}\left[\left(X'X\right)^{-1}\right],
\end{aligned}$$

which porves (9). See the Appendix B for the proof of the first equality. □

2.2 Properties of the OLS Estimator

Here we exhibit some properties of the OLS estimator.

Theorem 2.2 (Properties of the OLS Estimator)**.** The OLS estimator obtained above has the following properties.

(i) **unbiasedness** Under the assumption **H2**, the OLS estimator $\hat{b}$ becomes an unbiased estimator:

$$\mathbb{E}[\hat{b}] = b. \tag{10}$$

(ii) **consistency** Under the following assumption:

H5 $X'X$ is positive definite;

H6 For all i, for all k, l, the moments of $\mathbb{E}\left[|X_{ik}X_{il}|\right]$ exist and $\mathbb{E}\left[X'X\right]$ is positive definite,

as well as [**H1**–**H4**], the OLS estimator $\hat{b} = \left(X'X\right)^{-1}\left(X'y\right)$ satisfies

$$\hat{b} \xrightarrow[n\to\infty]{p} b \quad \text{or} \quad \operatorname*{plim}_{n\to\infty} \hat{b} = b. \tag{11}$$

(iii) **efficiency** Under the assumption [**H1**–**H4**], the variance of the OLS estimator is the minimum one in the class of linear unbiased estimator.

Proof. We can derive these properties via a similar calculation in the case of a simple regression model.

(i) **unbiasedness** This property is shown above (in Remark 2.1).

(ii) **consistency** From (5), we have:

$$\begin{aligned}\hat{b} &= b + \left(X'X\right)^{-1} X'u \\ &= b + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right).\end{aligned}$$

By WLLN and CMT, we have:

$$\hat{b} = b + \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'u\right) \tag{12}$$

$$= b + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right) \tag{13}$$

$$\xrightarrow[n\to\infty]{\mathbb{P}} b + \mathbb{E}\left[X_i'X_i\right]^{-1}\mathbb{E}\left[X_i'u_i\right]. \tag{14}$$

Here we apply the *convergence of the product of random variables in probability*, which we will discuss in the following. From the **weak law of large numbers** (WLLN),

$$\frac{1}{n}\sum_{i=1}^{n} X_i' X_i \xrightarrow[n\to\infty]{p} \mathbb{E}\left[X_i' X_i\right] < \infty; \tag{15}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i' u_i \xrightarrow[n\to\infty]{p} \mathbb{E}\left[X_i' u_i\right] = \mathbf{0} \in \mathbb{R}^k. \tag{16}$$

$\mathbb{E}\left[X_i' u_i\right] = 0$ holds from the *orthogonality condition* with respect to X and u. In addition,

$$\left(\frac{1}{n} X_i' X_i\right)^{-1} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\left[X_i' X_i\right]^{-1} \tag{17}$$

holds from the *continuous mapping theorem* shown as below. Thus, substituting (15) and (16) into (14) results in

$$\hat{b} \xrightarrow[n\to\infty]{P} b + \mathbb{E}\left[X_i' X_i\right]^{-1} \mathbf{0} = b,$$

which indicates that $\hat{b} \xrightarrow[n\to\infty]{p} b$.

(iii) **efficiency** As for the efficiency of the OLS estimator, the following *Gauss–Markov theorem* for a multiple regression model support the efficiency.

□

The *convergence of the product of random variables in probability* and *continuous mapping theorem* are respectively given as follows.

Lemma 2.2 (Convergence of the Product of Random Variables in Probability)**.** Suppose a sequence of random vector $\mathbf{X}_n$ converges in probability to $\mathbf{X}$ and $\mathbf{y}_n$ to $\mathbf{y}$, respectively. Then, the product of the two random variable $\mathbf{X}_n\mathbf{y}_n$ also converges in probability to the product of the each probability limit:

$$\mathbf{X}_n\mathbf{y}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbf{X}\mathbf{y}.$$

In another notation,

$$\operatorname*{plim}_{n\to\infty} \mathbf{X}_n\mathbf{y}_n = \operatorname*{plim}_{n\to\infty} \mathbf{X}_n \operatorname*{plim}_{n\to\infty} \mathbf{y}_n.$$

Lemma 2.3 (Continuous Mapping Theorem)**.** Suppose a sequence of random vector $\mathbf{x}_n \in \mathcal{S}$ converges in probability to $\mathbf{x}$. Then, for any continuous mapping $\mathbf{g}\colon \mathcal{S} \to \mathbb{R}^l$, the following relation holds:

$$\operatorname*{plim}_{n\to\infty} \mathbf{g}\left(\mathbf{x}_n\right) = \mathbf{g}\left(\operatorname*{plim}_{n\to\infty} \mathbf{x}_n\right) = \mathbf{g}\left(\mathbf{x}\right).$$

3 Gauss–Markov Theorem for a Multiple Regression Model

Here we will obtain a general result for the class of linear unbiased estimators of **b**. It can be conducted via a direct method.

Theorem 3.1 (Gauss–Markov Theorem for a Multiple Regression Model)**.** Under the assumption [**H1**–**H4**], the OLS estimater $\hat{b}$ of the multiple regression model

$$y_i = X_i b + u_i, \tag{18}$$

for all $i \in \{1, \ldots, n\}$ is of minimum variance among the class of linear unbiased estimator.

Proof. Let us assume another unbiased linear estimator of b, say $\tilde{b}$. Thus, there exists a matrix $A \in \mathbb{R}^{k\times n}$ such that $\tilde{b} = Ay$. Since $\tilde{b}$ is an unbiased estimator,

$$\mathbb{E}[\tilde{b}] = b \tag{19}$$

holds, which yields

$$\mathbb{E}\left[A\left\{Xb + u\right\}\right] = b \iff AXb = b. \tag{20}$$

Therefore, $AX = I_k$ must hold. Moreover, from the equation:

$$\tilde{b} - \mathbb{E}[\tilde{b}] = A\left\{y - Xb\right\} = Au. \tag{21}$$

the variance $\mathbb{V}[\tilde{b}]$ becomes

$$\mathbb{V}[\tilde{b}] = \mathbb{V}\left[Au\right] = A\mathbb{V}\left[u\right]A' = A\left(\sigma^2 I_n\right)A' = \sigma^2 AA', \tag{22}$$

from the assumption $\mathbb{V}\left[u\right] = \sigma^2 I_n$. Using the *projection matrix*:

$$\mathcal{M}_X := I_n - X\left(X'X\right)^{-1}X' \left(\iff I_n = \mathcal{M}_X + X\left(X'X\right)^{-1}X'\right), \tag{23}$$

we can rewrite (22) as follows:

$$\begin{aligned}
\mathbb{V}[\tilde{b}] &= A\left(\sigma^2 I_n\right)A' \\
&= \sigma^2 A\left(X\left(X'X\right)^{-1}X' + \mathcal{M}_X\right)A' \\
&= \sigma^2\left(AX\left(X'X\right)^{-1}X'A' + A\mathcal{M}_X A'\right).
\end{aligned}$$

Substituting $AX(= X'A') = I_k$ and $\mathbb{V}[\hat{b}] = \sigma^2\left(X'X\right)^{-1}$ into the above equation results in

$$\mathbb{V}[\tilde{b}] = \mathbb{V}[\hat{b}] + \sigma^2 A\mathcal{M}_X A' \iff \mathbb{V}[\tilde{b}] - \mathbb{V}[\hat{b}] = \sigma^2 A\mathcal{M}_X A'.$$

Hence, the difference of ith diagonal elements of variance–covariance matrices becomes

$$\mathbb{V}[\tilde{b}]_{ii} - \mathbb{V}[\hat{b}]_{ii} = a_i'\mathcal{M}a_i > 0$$

for any column vector a_i in A for $i \in \{1, \ldots, k\}$, which proves the theorem. □

4 Asymptotic Normality for the OLS Estimator of a Multiple Regression Model

In this section, we derive the asymptotic distribution of an OLS estimator to observe how the distribution changes as $n \to \infty$.

Theorem 4.1 (Asymptotic Normality of an OLS Estimator)**.** Let $\hat{b}$ be the OLS estimator obtained under the assumption [**H1**–**H6**]. Then, the OLS estimator asymptotically follows a normal distribution as follows:

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \left(\mathbb{E}\left[X_i' X_i\right]\right)^{-1}\right).$$

Proof. From (5), we have

$$\begin{aligned}\hat{b} &= b + \left(X'X\right)^{-1} X'u \\ &= b + \left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i' u_i\right).\end{aligned}$$

Therefore,

$$\sqrt{n}(\hat{b} - b) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' u_i\right). \tag{24}$$

From the **Lindeberg–Feller central limit theorem** (Lindeberg–Feller CLT) as well as the **weak law of large numbers** (WLLN) and *continuous mapping theorem*, we have

$$\begin{aligned}\left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1} &\xrightarrow[n\to\infty]{P} \mathbb{E}\left[X_i' X_i\right]^{-1}; \\ \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' u_i\right) = &\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i' u_i - \mathbf{0}\right) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \mathbb{V}[X_i' u_i]\right),\end{aligned}$$

since from the orthogonality condition,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i' u_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i' u_i] = \mathbf{0}.$$

Then,

$$\begin{aligned}\mathbb{V}[X_i' u_i] &= \mathbb{E}\left[\mathbb{V}[X_i' u_i \middle| X_i]\right] + \mathbb{V}\big[\underbrace{\mathbb{E}[X_i' u_i \middle| X_i]}_{=0}\big] \\ &= \mathbb{E}\left[X_i' \mathbb{V}[u_i \middle| X_i] X_i\right] \\ &= \mathbb{E}\left[X_i' \sigma^2 X_i\right] \\ &= \sigma^2 \mathbb{E}\left[X_i' X_i\right] < \infty,\end{aligned}$$

Therefore, from (24) and the **Slutsky's theorem**,

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} \mathbb{E}\left[X_i' X_i\right]^{-1} b,$$

where

$$b \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]\right).$$

From the following relation:

$$b \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]\right) \Longrightarrow \mathbb{E}\left[X_i' X_i\right]^{-1} b \sim N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]^{-1}\right),$$

we obtain

$$\sqrt{n}(\hat{b} - b) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^k}\left(\mathbf{0}, \sigma^2 \mathbb{E}\left[X_i' X_i\right]^{-1}\right).$$

□

Here we review the *Slutsky's Theorem*.

Lemma 4.1 (Slutsky's Theorem)**.** Suppose a sequence of random vector $\mathbf{X}_n \xrightarrow[n\to\infty]{P} \mathbf{X}$ and $\mathbf{y}_n \xrightarrow[n\to\infty]{d} \mathbf{y}$, respectively. Then, the product of the two random variable $\mathbf{X}_n \mathbf{y}_n$ also converges in distribution as follows:

$$\mathbf{X}_n \mathbf{y}_n \xrightarrow[n\to\infty]{d} \mathbf{X}\mathbf{y}. \tag{25}$$

Appendix

A The Probability Density Function for a Multivariate Normal Distribution

A.1 Independent Univariate Normals

To derive the general case of *probability density function for a multivariate normal distribution*, we will start with a vector consisting of k independent and normally distributed random variables with mean 0: $\mathbf{x} = (x_1, \ldots, x_k)$ where

$$x_i \sim N_{\mathbb{R}}\left(0, \sigma_i^2\right).$$

Let us denote, by f_{x_i}, the probability density function for a single normal random variable x_i for $i \in \{1, \ldots, k\}$. Then, since the variables are independent, the joint probability density function, $f_{\mathbf{x}}$, of all k variables will just be the product of their densities:

$$\begin{aligned} f_{\mathbf{x}} &= \Pi_{i=1}^k f_{x_i} \\ &= \Pi_{i=1}^k \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{x_i^2}{2\sigma_i^2}\right\} \\ &= \frac{1}{\sqrt{(2\pi)^k \Pi_{i=1}^k \sigma_i^2}} \exp\left\{-\frac{1}{2}\mathbf{x}' \mathrm{diag}(\sigma_1^2, \ldots, \sigma_k^2)^{-1} \mathbf{x}\right\} \\ &= \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}' \Sigma^{-1} \mathbf{x}\right\} \end{aligned}$$

where $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_k^2)$. In this case we say that $\mathbf{x} \sim N_{\mathbb{R}^k}(\mathbf{0}, \Sigma)$. Unfortunately, this derivation is restricted to the case where these entries are independent and 0-centered. Thus, we will see that we can derive the general case using this result.

A.2 Affine Transformations of a Random Vector

Consider an affine transformation $L\colon \mathbb{R}^k \to \mathbb{R}^k$, $L(\mathbf{x}) = A\mathbf{x} + b$ for an invertible matrix $A \in \mathbb{R}^{k\times k}$ and a constant vector $b \in \mathbb{R}^k$. It is easy to verify that when we apply this transformation to a random variable $\mathbf{z} = (z_1, \ldots, z_k)$ with mean $\mu \in \mathbb{R}^k$ and variance–covariance matrix $\Sigma_{\mathbf{z}} \in \mathbb{R}^{k\times k}$ we get a new random variable $\mathbf{x} = L(\mathbf{z})$ such that

$$\begin{aligned}\mathbb{E}[L(\mathbf{x})] &= L\left(\mathbb{E}[\mathbf{z}]\right);\\ \mathbb{V}[L(\mathbf{x})] &= \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])\left(\mathbf{x} - \mathbb{E}[\mathbf{x}]\right)'\right] = A\Sigma_{\mathbf{z}}A'.\end{aligned}$$

In this case, for a symmetric, positive definite matrix Σ and constant vector μ, we will be looking at the transformation $\mathbf{x} = \Sigma^{1/2}\mathbf{z}+\mu$. It is interesting to note that, given an orthogonal decomposition $\Sigma = U\Lambda U'$, where U is orthogonal and Λ is a diagonal matrix consisting of the eigenvalues of Σ, entry x_i of the new random vector is a weighted sum of originally independent random variables in $\mathbf{z}$. Let u_i denote the ith row of a matrix U. Then,

$$x_i = \left(\Sigma_{\mathbf{z}}^{1/2}\mathbf{z} + \mu\right)_i = \sqrt{\lambda_i}u_i\mathbf{z} + \mu_i = \sum_{j=1}^{k} \lambda_i u_{ij} z_j + \mu_i.$$

We now just need one more fact about a change of variables to derive the general multivariate normal probability density function for this new random vector.

A.3 Probability Density Function of a Transformed Random Vector

Suppose that $\mathbf{z}$ is a random vector taking on values in a subset $S \in \mathbb{R}^k$, with a continuous probability density function f. Suppose $\mathbf{x} = r(\mathbf{z})$ where r is a differentiable function from S onto some other subset $T \in \mathbb{R}^k$. Then, the probability density function g of $\mathbf{x}$ is given by

$$g(\mathbf{x}) = f(\mathbf{z})\left|\det\left(\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}\mathbf{x}}\right)\right| = f\left(r(\mathbf{x})^{-1}\right)\left|\det\left(\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}\mathbf{x}}\right)\right|,$$

where $\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}\mathbf{x}}$ stands for the *Jacobian* of the inverse of r, and det() the determinant of a matrix. Returning to our previous discussion, where $\mathbf{x} = \Sigma^{1/2}\mathbf{z} + \mu$, we can see that the inverse transformation is given by $\mathbf{z} = \Sigma^{-1/2}(\mathbf{x} - \mu)$. Thus, the determinant of the Jacobian of this inverse becomes $\det(\Sigma^{-1/2}) = \frac{1}{\sqrt{\det(\Sigma)}}$. (You should check this equality using some porperties about the determinant.)

A.4 The Multivariate Normal Probability Density Function

Consider the random vector $\mathbf{z} \sim N_{\mathbb{R}^k}\left(\mathbf{0}, I\right)$ where I is is the identity matrix. As before we let $\mathbf{x} := \Sigma^{1/2}\mathbf{z} + \mu$ for positive definite Σ and a constant vector μ. We can now find the density function g of $\mathbf{x}$ from the known density function f for $\mathbf{z}$.

$$
\begin{aligned}
g(\mathbf{x}) &= f\left(r(\mathbf{x})^{-1}\right)\left|\det\left(\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}\mathbf{x}}\right)\right| \\
&= f\left(\Sigma^{-1/2}(\mathbf{x}-\mu)\right)\frac{1}{\sqrt{\det(\Sigma)}} \\
&= \frac{1}{\sqrt{(2\pi)^k}}\frac{1}{\sqrt{\det(\Sigma)}}\exp\left\{-\frac{1}{2}\left(\Sigma^{-1/2}(\mathbf{x}-\mu)\right)'\left(\Sigma^{-1/2}(\mathbf{x}-\mu)\right)\right\} \\
&= \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}.
\end{aligned}
$$

This is the probability density function for a multivariate normal distribution with mean vector μ and a covariance matrix Σ. We say that $\mathbf{x} \sim N_{\mathbb{R}^k}(\mu, \Sigma)$.

B Properties of Conditional Variances

Theorem B.1 (Properties of Conditional Variances)**.**

$$
\mathbb{V}[\mathbf{y}] = \mathbb{E}\left[\mathbb{V}[\mathbf{y}|\mathbf{X}]\right] + \mathbb{V}\left[\mathbb{E}[\mathbf{y}|\mathbf{X}]\right]. \tag{26}
$$

Proof. We can derive this relation in direct procedure shown as follows.

$$
\begin{aligned}
\mathbb{V}[\mathbf{y}] &\equiv \mathbb{E}\left[(\mathbf{y}-\mathbb{E}[\mathbf{y}])^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]+\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)^2\right] + 2\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)\right].
\end{aligned}
$$

Here we have the following calculation:

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\big|\mathbf{X}\right] &= \mathbb{E}\left[\mathbf{y}|\mathbf{X}\right] - \mathbb{E}\left[\mathbb{E}[\mathbf{y}|\mathbf{X}]\big|\mathbf{X}\right] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbf{y}|\mathbf{X}] - \mathbb{E}[\mathbf{y}|\mathbf{X}]\big|\mathbf{X}\right] \\
&= \mathbf{0}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)\Big|\mathbf{X}\right]\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\Big|\mathbf{X}\right]\right] \\
&= \mathbf{0}.
\end{aligned}
$$

Note that $\mathbb{E}[c|\mathbf{X}] = c$ for any constant c and $\mathbb{E}[\mathbf{y}|\mathbf{X}]$ is a random variable of $\mathbf{x}$. Therefore, by using the *law of iterated expectations*,

$$
\begin{aligned}
\mathbb{V}[\mathbf{y}] &= \mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}[\mathbf{y}]\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)^2\Big|\mathbf{X}\right]\right] + \mathbb{E}\left[\left(\mathbb{E}[\mathbf{y}|\mathbf{X}]-\mathbb{E}\left[\mathbb{E}[\mathbf{y}|\mathbf{X}]\right]\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{V}[\mathbf{y}|\mathbf{X}]\right] + \mathbb{V}\left[\mathbb{E}[\mathbf{y}|\mathbf{X}]\right],
\end{aligned}
$$

which proves (26). □

Lemma B.1. In general, we have the following equation:

$$\mathbb{E}\left[\mathbf{g}(\mathbf{x})\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\right]=\mathbf{0}.$$

Proof. Using the fact $\mathbb{E}\left[\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\middle|\mathbf{X}\right]=0$, we have

$$\begin{aligned}\mathbb{E}\left[\mathbf{g}(\mathbf{x})\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)\right]&=\mathbb{E}\left[\mathbb{E}\left[\mathbf{g}(\mathbf{x})\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)|\mathbf{X}\right]\right]\\&=\mathbb{E}\left[\mathbf{g}(\mathbf{x})\mathbb{E}\left[\left(\mathbf{y}-\mathbb{E}[\mathbf{y}|\mathbf{X}]\right)|\mathbf{X}\right]\right]\\&=\mathbf{0}.\end{aligned}$$

□