Econometrics I
TA Session 2

Jukina HATAKEYAMA *

April 25, 2024

Contents

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

1 Review of Mapping

In this section, we refer to some concepts concerned with a mapping. The discussion in this section takes place in the context of Euclidean spaces. A function f from S to T, which is denoted by $f : S \to T$, is a rule that associates with each element of S, one and only one element of T. The set S is called the *domain* of the function f, and the set T its *range*. The graphical image is shown below. Then, a *continuous* function at $x \in S$ is defined as follows.

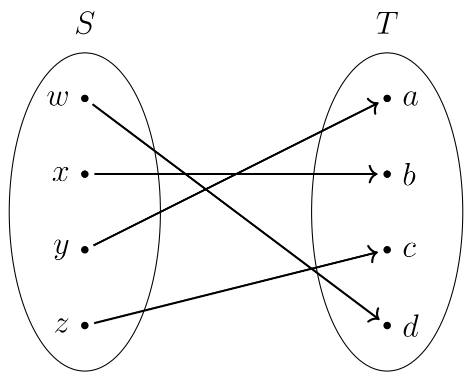

Figure 1: Mapping Diagram of Relations between S and T

Definition 1.1 (Continuous Function)**.** Let $f : S \to T$, where $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^l$. Then, a function f is said to be *continuous* at $x \in S$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $y \in S$ and $d(x, y) < \delta$ implies $d(f(x), f(y)) < \varepsilon$.

Note that $d(x, y)$ indicates the distance between x and y in $\mathbb{R}^n$, while $d(f(x), f(y)) < \varepsilon$ indicates the distance in $\mathbb{R}^l$. intuitively, f is continuous at x if the value of f at any point y that is "close" to x is a good approximation of the value of f at x.

In the following, we show an example which typifies a function continuous everywhere except some point. The function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0; \\ 1, & \text{otherwise,} \end{cases} \tag{1}$$

is continuous everywhere except at $x = 0$. At $x = 0$, every open ball $B(x, \delta)$ with center x and radius $\delta > 0$ contains at least one point $y > 0$. At all such points, $f(y) = 1 \neq 0 = f(x)$, and this approximation does not get better, no matter how close y gets to x (or no matter how small we take δ to be).

2 Review of Optimisation

In this section, we review some concepts concerned with optimisation.

2.1 Hessian

Consider a case that we aim to obtain a point $x \in \mathbb{R}$ which maximises or minimises a function $y = f(x)$ in an open interval $U \in \mathbb{R}$. In this case, if $x^0 \in \mathbb{R}$ attains the maximum or minimum, we

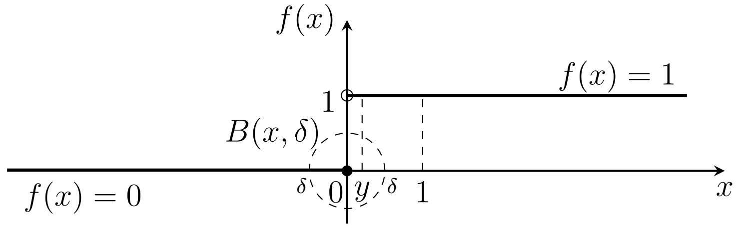

Figure 2: Graph of $f(x)$

have the following first order condition at the beginning:

$$\left.\frac{df(x)}{dx}\right|_{x=x^0} = 0. \tag{2}$$

In addition, when we consider whether the optimum is a maximum or a minimum, the sufficient condition for the optimum becomes as follows:

$$\left.\frac{d^2f(x)}{dx^2}\right|_{x=x^0} < 0 \quad \text{for a maximum;} \tag{3}$$

$$\left.\frac{d^2f(x)}{dx^2}\right|_{x=x^0} > 0 \quad \text{for a minimum.} \tag{4}$$

Here consider a function $y = g(\mathbf{x})(\in \mathbb{R})$ where $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$, denoted as $g\colon \mathbb{R}^n \to \mathbb{R}$. If an $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)' \in \mathbb{R}^n$ maximises or minimises $g(\mathbf{x})$, we apply the following theorem.

Theorem 2.1. If a function $g\colon \mathbb{R}^n \to \mathbb{R}$ is maximised (or minimised) at the point $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)$, then the following equation holds:

$$\left.\nabla_{\mathbf{x}} g(\mathbf{x})\right|_{\mathbf{x}=\mathbf{x}^0} = \begin{pmatrix} \frac{\partial g(\mathbf{x}^0)}{\partial x_1} \\ \vdots \\ \frac{\partial g(\mathbf{x}^0)}{\partial x_n} \end{pmatrix} = \mathbf{0}. \tag{5}$$

Moreover, we use the following *Hessian matrix* to discern a maximum and a minimum.

Definition 2.1 (Hessian matrix)**. A Hessian matrix** of a function $g\colon \mathbb{R}^n \to \mathbb{R}$ is defined as follows:

$$H = \nabla^2_{\mathbf{x}\mathbf{x}'} g(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix}$$

Assume that $g_{x_1}(\mathbf{x}^0) = g_{x_2}(\mathbf{x}^0) = \cdots = g_{x_n}(\mathbf{x}^0) = 0$ holds, where $g_{x_i}(\mathbf{x})$ for $i \in \{1, \dots, n\}$ denotes the partial derivative of $g(\mathbf{x})$ with respect to x_i. The following theorem is a way to distinguish whether $\mathbf{x}$ attains a maximum (or a minimum) or not.

Theorem 2.2 (Definiteness and Hessian Matrix)**.** Suppose that a smooth function $g\colon \mathbb{R}^n \to \mathbb{R}$ satisfies $g_{x_1}(\mathbf{x}^0) = \cdots = g_{x_n}(\mathbf{x}^0) = 0$. Then, we can confirm that if:

1. H is a negative definite matrix, then $\mathbf{x}^0$ is a maximum point.
2. H is a positive definite matrix, then $\mathbf{x}^0$ is a minimum point.

As for the positiveness or negativeness of a matrix, we have the following theorem.

Theorem 2.3 (Definiteness of a Matrix)**.** A necessary and sufficient condition for a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ to be positive definite is that $g_i > 0$ for all $i \in \{1, \dots, n\}$ where

$$g_i = \begin{vmatrix} a_{11} & \cdots & a_{1i} \\ \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{ii} \end{vmatrix}. \tag{6}$$

Moreover, a necessary and sufficient condition for a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ to be negative definite is that $g_1 < 0$, $g_2 > 0$, $g_3 < 0$, and so on.

For example, in the case of $f(x, y) = x^2 + 4xy + 5y^2 - 2x - 8y + 5$, we have $\partial_x f = 2x + 4y - 2$ and $\partial_y f = 4x + 10y - 8$. By solving $\partial_x f = \partial_y f = 0$, we obtain an optimum point $(x, y) = (-3, 2)$. Also, the Hessian matrix is given as follows:

$$\begin{aligned} H &= \nabla^2_{\theta,\theta'} f(x, y) \\ &= \begin{pmatrix} 2 & 4 \\ 4 & 10 \end{pmatrix}, \end{aligned} \tag{7}$$

where $\theta = (x, y)' \in \mathbb{R}^2$.
Since $|H| = 20 - 16 > 0$ and $\partial^2_{xx'} f = 2 > 0$, $(x, y) = (-3, 2)$ is a minimum point. We can analyse an optimum of a multivariable function for more variables in the same manner.

2.2 Optimisation: Jacobian

If a function $\mathbf{g}\colon \mathbb{R}^{m+n} \to \mathbb{R}^n$ satisfies

$$\mathbf{g}(\mathbf{y}, \mathbf{x}) = \begin{cases} g_1(\mathbf{y}, \mathbf{x}) = k; \\ \qquad \vdots \\ g_n(\mathbf{y}, \mathbf{x}) = k, \end{cases} \tag{8}$$

where k is a constant and $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, we call the function $\mathbf{g} \in \mathbb{R}^m$ an *implicit function*. Then, a matrix called *Jacobian matrix* is defined as follows:

Definition 2.2 (Jacobian Matrix)**.** Suppose a function $\mathbf{g}\colon \mathbb{R}^{m+n} \to \mathbb{R}^n$ is an implicit function mentioned above. Then, **the Jacobian matrix** of the function $\mathbf{g}(\mathbf{y}, \mathbf{x})$ with respect to $\mathbf{y} \in \mathbb{R}^m$ becomes as follows:

$$J = (\nabla_{\mathbf{y}} \mathbf{g}(\cdot))' = \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_m} \end{pmatrix}. \tag{9}$$

To conduct some optimisation methods such as Lagrange multiplier method, we need to know whether we can express each element of y_i in $\mathbf{y}$ for $i \in \{1, \dots, m\}$ as a unique function of $\mathbf{x} \in \mathbb{R}^n$ explicitly as follows:

$$\mathbf{y} = \phi(\mathbf{x}) = \begin{cases} y_1 = \phi_1(\mathbf{x}); \\ \quad \vdots \\ y_m = \phi_m(\mathbf{x}). \end{cases} \tag{10}$$

Roughly speaking, under some conditions, there exists a unique solution satisfying (10) if the inverse matrix of the Jacobian matrix exists.

3 Large Order and Small Order

In this section, we review some definitions of small and large order.

3.1 Convergence of a Non-stochastic Sequence of Numbers

First of all, we begin with some definitions regarding non-stochastic sequences of numbers.

Definition 3.1 (Convergence of a Sequence in the Case of Non-stochastic Numbers)**.** (1) A sequence of nonrandom numbers $\{a_n\colon n \in \mathbb{Z}_{++} = \mathbb{N}\}$(, where $\mathbb{Z}_{++} = \mathbb{N}$ represents the set of all (strictly) positive integer or natural number,) converges to a number $a \in \mathbb{R}$ if for all $\varepsilon > 0$ there exists a number n_ε such that $|a_n - a| < \varepsilon$ if $n > n_\varepsilon$. In this case, we write $a_n \to a$ as $n \to \infty$.

(2) A sequence $\{a_n\colon n \in \mathbb{Z}_{++}\}$ is *bounded* if and only if there exists some $b < \infty$ such that $|a_n| \leq b$ for all $n \in \mathbb{Z}_{++}$. Otherwise, we say that the sequence $\{a_n\colon n \in \mathbb{Z}_{++}\}$ is *unbounded.*

These definitions apply to vectors and matrices element by element. A well–known example of a sequence satisfying the first statement is the case $a_n = 1/n$ for $n \in \mathbb{Z}_{++}$. We can write $a_n \to 0$. The sequence is also bounded since $|a_n| \leq 1$ for all $n \in \mathbb{Z}_{++}$.

3.2 Large Order and Small Order

Next, we introduce the definition of "Large Order and Small Order" in the following.

Definition 3.2 (Large Order and Small Order)**.** (1) A sequence $\{a_n : n \in \mathbb{Z}_{++}\}$ is $O(n^\lambda)$ (at most of order n^λ) if $n^{-\lambda}a_n$ is bounded, that is, if there exists a number $b < \infty$ such that $|n^{-\lambda}a_n| \leq b$ for all $n \in \mathbb{Z}_{++}$. When $\lambda = 0$, the sequence is bounded (as you can confirm the fact from the definition of boundedness), and in this situation we also write $a_n = O(1)$.

(2) A sequence $\{a_n : n \in \mathbb{Z}_{++}\}$ is $o(n^\lambda)$ if $n^{-\lambda}a_n \to 0$. When $\lambda = 0$, a_n converges to zero, and we also write $a_n = o(1)$.

From these definition, it is obvious that if $a_n = o(n^\lambda)$, then $a_n = O(n^\lambda)$; in particular, if $a_n = o(1)$, then $a_n = O(1)$. In addition, if each element of a sequence of vectors or matrices is $O(n^\lambda)$, we say that the sequence of vectors or matrices is $O(n^\lambda)$, and similarly for $o(n^\lambda)$.

As for the sequence $\{a_n = 1/n : n \in \mathbb{Z}_{++}\}$, $a_n = o(1)$ since a_n converges to zero as n goes to ∞, which also indicates $a_n = O(1)$. We show, in the following, another interpretation of this sequence. $a_n = 1/n$ satisfies the following relation:

$$\frac{1}{n} = O\left(\frac{1}{n}\right) = o\left(\frac{1}{n^{1-\alpha}}\right), \tag{11}$$

for all $\alpha > 0$. When $\alpha = 0.5$, from the above equation,

$$\frac{1}{n} = o\left(\frac{1}{n^{0.5}}\right) = o\left(\frac{1}{\sqrt{n}}\right). \tag{12}$$

holds. The intuitive understanding of what $o\left(\frac{1}{\sqrt{n}}\right)$ means becomes as follows: if $a_n = o\left(\frac{1}{\sqrt{n}}\right)$, then a_n is small in the sense that $\sqrt{n}a_n$ converges to zero as n goes to ∞.

4 Basic Convergence Theory

In this section, we introduce some concepts of "convergence." Before confirming the definitions, we review the definition of probability space.

Remark 4.1 (A Probability Space)**.** When we consider a trial or a random experiment, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be constructed, where

- Ω stands for a sample space, i.e., the collection of all possible outcomes of the trial, or experiment.
- $\mathcal{F}$ stands for a subset of Ω containing $\Omega, \emptyset$ (an empty set which have no elements), in other words, a sigma-algebra of events.
- $\mathbb{P}$, denoted a probability measure on $\mathcal{F}$ and satisfies $\mathbb{P}(\Omega) = 1$.

Then, the probability of a set can be seen as a function from Ω to $\mathbb{R}$, i.e., $\mathbb{P} \colon \Omega \to \mathbb{R}$.

Let us explains the above remark by considering a case of tossing a single dice. We can define

$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, where

ω_1 : after tossing a dice, we observe that the face showing on the dice equals **1**;
ω_2 : after tossing a dice, we observe that the face showing on the dice equals **2**;
ω_3 : after tossing a dice, we observe that the face showing on the dice equals **3**;
ω_4 : after tossing a dice, we observe that the face showing on the dice equals **4**;
ω_5 : after tossing a dice, we observe that the face showing on the dice equals **5**;
ω_6 : after tossing a dice, we observe that the face showing on the dice equals **6**,

and a subset $\mathcal{F}$ which consists of all possible subsets of Ω. An example of a set $\mathcal{F}$ can be considered as $\{\omega_1, \omega_3, \omega_5\}$, implying that after tossing a dice, we observed that the faces showing on the each trial are odd numbers, i.e. 1, 3, and 5." Then, if

$$\mathrm{P}(\omega_1 \in \mathcal{F}) = \cdots = \mathrm{P}(\omega_6 \in \mathcal{F}) = \frac{1}{6}, \tag{13}$$

holds, this equation implies that "the probability that after throwing a dice we observe that the face showing on the dice equals 1 (, 2, and so on) is equal for all $\omega \in \Omega$." Similarly,

$$\mathrm{Prob}(\Omega) = \mathrm{Prob}(\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}) = 1 \tag{14}$$

indicates that "after throwing a dice the face on the dice surely shows a natural number from one to six." Eq. (13) or (14) represent the assotiation of an outcome with a real number, i.e., $\mathbb{P}\colon \Omega \to \mathbb{R}$. We can construct a random variable in a same manner. A random variable X is a function from Ω to $\mathbb{R}$, that is, $X\colon \Omega \to \mathbb{R}$. In the above example, for an outcome $\omega \in \Omega$, we can consider a random variable $X\colon \Omega \to \mathbb{R}$ taking a value as follows:

$$X(\omega) = \begin{cases} 1, & \text{if } \omega = \omega_1; \\ 2, & \text{if } \omega = \omega_2; \\ 3, & \text{if } \omega = \omega_3; \\ 4, & \text{if } \omega = \omega_4; \\ 5, & \text{if } \omega = \omega_5; \\ 6, & \text{if } \omega = \omega_6. \end{cases}$$

4.1 Convergence in Probability

First of all, we show the definition of **convergence in probability** as below.

Definition 4.1 (Convergence in Probability)**.** (1) A sequence of random variables $\{X_n \colon n \in \mathbb{Z}_{++}\}$ **convergences in probability** to a constant a if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \text{Prob}\left(|X_n - a| > \varepsilon\right) = 0 \quad \text{or} \quad \operatorname*{plim}_{n\to\infty} X_n = a. \tag{15}$$

We write $X_n \xrightarrow{p} a$ and say that a is the **probability limit (plim)** of X_n.

(2) In the special case where $a = 0$, we also say that $\{X_n\}$ is $o_p(1)$. We also write $X_n = o_p(1)$ or $X_n \xrightarrow{p} 0$.

(3) A sequence of random variables $\{X_n \colon n \in \mathbb{Z}_{++}\}$ is **convergences in probability** if and only if for every $\varepsilon > 0$, there exists a $b_\varepsilon < \infty$ and an integer n_ε such that

$$\text{Prob}\left(|X_n| \geq b_\varepsilon\right) < \varepsilon \quad \text{for all} \quad n \geq n_\varepsilon. \tag{16}$$

We write $X_n = O_p(1)$.

This concept is an important one in understanding the (Weak) Law of Large Numbers, which we will learn in later TA session.

Consider an example of tossing a (fair) coin, where the probability that the face showing on the coin becomes head equals the one that of tail. We can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as follows:

$$\Omega = \{\text{head}, \text{tail}\} =: \{\omega_1, \omega_2\}, \quad \mathcal{F} := \{\Omega, \{\omega_1\}, \{\omega_2\}, \emptyset\}, \tag{17}$$

and

$$\mathbb{P}(\omega) := \text{Prob}(\omega) = \begin{cases} \frac{1}{2} & \text{if } \omega = \omega_1; \\ \frac{1}{2} & \text{if } \omega = \omega_2. \end{cases} \tag{18}$$

Then, for $n \in Z_{++}$, we define a random variable $X_n(\omega)$ depending on whether the coin shows head/tail as follows:

$$X_n(\omega) = \begin{cases} 1, & \text{if the face of the coin shows tail at the } n\text{th trial;} \\ 0, & \text{if the face of the coin shows head at the } n\text{th trial.} \end{cases} \tag{19}$$

Also, set a random variable $X(\omega)$ as follows:

$$X(\omega) = \begin{cases} 1, & \text{if the face of the coin shows head;} \\ 0, & \text{if the face of the coin shows tail.} \end{cases} \tag{20}$$

Then, it is clear that for all $\omega \in \Omega$, $|X_n(\omega) - X(\omega)|$ holds at any trial. Therefore, for all $\eta \in (0, 1)$ and for all $n \in \{1, 2, \ldots\}$, we have

$$\text{Prob}(|X_n(\omega) - X(\omega)| > \eta) = 1. \tag{21}$$

Thus, $X_n(\omega)$ does not converge to $X(\omega)$.

4.2 Almost Surely Convergence

Next, we see the definition of **almost surely convergence**.

Definition 4.2 (Almost Surely Convergence)**.** A sequence of random variabels $\{X_n \in \mathbb{R} \colon n \in \mathbb{Z}_{++}\}$ **converges almost surely** to X if and only if for all $\omega \in \Omega$ which do not belong to events of probability 0,

$$\mathrm{Prob}\left(\{\omega;\ X_n \to X \quad \text{as} \quad n \to \infty\}\right) = 1, \tag{22}$$

or

$$\mathrm{Prob}\left(\lim_{n\to\infty} X_n \to X\right) = 1. \tag{23}$$

We write $X_n \xrightarrow{a.s.} X$.

An example is described below. Given a probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a sequence of random variable $\{X_n, n \in Z_{++}\}$ as follows:

$$X_n : \Omega \to \mathbb{R}, \quad X_n(\omega) = 1 + \frac{1}{n}. \tag{24}$$

Note that this random variable does not change for the outcome occurring from a sample space. Then, for all $\omega \in \Omega$,

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right) = 1 \tag{25}$$

holds. Therefore,

$$\mathrm{Prob}\left(\{\omega;\ X_n \to 1 \quad \text{as} \quad n \to \infty\}\right) = \mathrm{Prob}\left(\lim_{n\to\infty} X_n \to 1\right) = 1 \tag{26}$$

also holds for all $\omega \in \Omega$, which implies that X_n converges almost surely to X.

In the Case of Tossing a Dice

4.3 Convergence in Distribution

Here we check the definition of **convergence in probability**.

Definition 4.3 (Convergences in Distribution)**.** A sequence of random variabels $\{X_n \in \mathbb{R} : n \in \mathbb{Z}_{++}\}$ **converges in distribution** to a continuous random variable X if and only if

$$F_n(\xi) \to F(\xi) \quad \text{as} \quad n \to \infty, \tag{27}$$

or

$$\lim_{n\to\infty} F_n(X) = F(X), \tag{28}$$

for all $\xi \in \mathbb{R}$, where F_n is the cumulative distribution function (c.d.f.) of X_n and F is the (continuous) c.d.f. of X, and both c.d.f.s are continuous at $\xi \in \mathbb{R}$. We write $X_n \xrightarrow{d} X$.

We will learn the application of this convergence, Central Limit Theorem, in the later class of Econometrics I and TA session.

The following example displays an example of the convergence in distribution. Given a probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{X_n\}$ and X be sequences given as follows:

$$X_n : \Omega \to \mathbb{R}, \quad X_n(\omega) = \frac{1}{n}; \quad X = 0. \tag{29}$$

Then, the cumulative distribution function of X_n and that of X are assumed to be take the following form respectively:

$$F_n(x) = \begin{cases} 1 & \text{if} \quad x \geq \frac{1}{n}; \\ 0 & \text{otherwise.} \end{cases} \qquad F(x) = \begin{cases} 1 & \text{if} \quad x \geq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

In this case, we can confirm that for all $X \in \mathbb{R}/\{0\}$, $F_n(x)$ satisfies the following equation

$$\lim_{n\to\infty} F_n(X) = F(X). \tag{31}$$

Hence, X_n converges in distribution to X.

In the Case of Tossing a Dice

4.4 Convergence in L^p

Definition 4.4 (Convergences in L^p)**.** A sequence of random variabels $\{X_n \in \mathbb{R} \colon n \in \mathbb{Z}_{++}\}$ **converges in** L^p to X if and only if for all $n \in \mathbb{Z}_{++}$, $\mathbb{E}[|X_n|^p] < \infty$ and

$$\mathbb{E}[|X_n - X|^p] \to 0 \quad \text{as} \quad n \to \infty \tag{32}$$

or (in another notation,)

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0 \tag{33}$$

holds. We write $X_n \xrightarrow{\mathrm{L}^p} X$. Particularly, when $n = 2$, we say that X_n **converges in mean square** to X.

Let us consider an example of tossing a (fair) coin. Under the same assumption of a probability (or sample) space mentioned above, define a sequence of random variable X_n for $n \in Z_{++}$ as follows:

$$X_n(\omega) =: X_n = \begin{cases} 1, & \text{if } \omega = \omega_1 \text{ with probability } p; \\ 0, & \text{if } \omega = \omega_2 \text{ with probability } 1 - p. \end{cases} \tag{34}$$

In this case, this sequence of random variables becomes a sequence of a Bernoulli random variable. Then, we denote the sum of the sequence X_n from 1 to n as X:

$$X := \frac{1}{n}\left(X_1 + \cdots + X_n\right). \tag{35}$$

The expectation and variance of each X_n for $n \in Z_{++}$ becomes:

$$\begin{aligned} \mathbb{E}\left[X_n\right] &= 1 \cdot p + 0 \cdot (1-p) = p; \\ \mathbb{V}\left[X_n\right] &= \mathbb{E}\left[|X_n - \mathbb{E}\left[X_n\right]|^2\right] = p(1-p). \end{aligned}$$

Assuming that each random variable is independent with each other, a similar calculation results in:

$$\begin{aligned} \mathbb{E}\left[X\right] &= \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\mathbb{E}\left[X_1\right] + \cdots + \mathbb{E}\left[X_n\right]}{n} = p; \\ \mathbb{V}\left[X\right] &= \mathbb{V}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n^2}\mathbb{V}\left[X_1 + \cdots + X_n\right] = \frac{1}{n^2}\Big\{\mathbb{V}\left[X_1\right] + \cdots + \mathbb{V}\left[X_n\right]\Big\} = \frac{1}{n}p(1-p). \end{aligned}$$

On the other hand, we have

$$\mathbb{V}\left[X\right] = \mathbb{E}\left[|X - \mathbb{E}\left[X\right]|^2\right] = \mathbb{E}\left[|X - p|^2\right].$$

Therefore, $\mathbb{E}[|X - p|^2] = \frac{1}{n}p(1-p)$ holds, and

$$\lim_{n\to\infty} \mathbb{E}\left[|X - p|^2\right] = \lim_{n\to\infty} \frac{1}{n}p(1-p) = 0, \tag{36}$$

also holds, which indicates that X converges in L^2 to p.

4.5 Relation of Convergences

The following theorem shows significant relations of these four convergences in understanding some concepts of econometric theory such as LLN or CLT.

Theorem 4.1 (Relations of the Convergences)**.** *a.s.* convergence or L^p convergence implies convergence in probability, and the latter implies convergence in distribution.

We can summerize the above relations in the following graphical image:

$$\begin{array}{ccc} & \text{a.s.} & \\ & \big\downarrow & \\ \text{in } \mathrm{L}^p & \longrightarrow & \text{in } p \\ & \big\downarrow & \\ & \text{in } d & \end{array}$$

Figure 3: Image of the Relation between the Convergences

Then, when considering some "asymptotic convergence (in distribution)," which will be explained in a later class of the Econometrics I or TA session, the following theorems become key tools to prove some asymptotic property.

Theorem 4.2 (Continuous Mapping Theorem)**.** Let $\{\mathbf{X}_n \in \mathbb{R}^k : n \in \mathbb{Z}_{++}\}$ be sequence of $k \times 1$ random vectors such that $\mathbf{X}_n \xrightarrow[n \to \infty]{d} \mathbf{X}$. If $\mathbf{g} \colon \mathbb{R}^k \to \mathbb{R}^j$ is a continuous function, then $\mathbf{g}(\mathbf{X}_n) \xrightarrow[n \to \infty]{d} \mathbf{g}(\mathbf{X})$.

Theorem 4.3 (Slutsky's Theorem)**.** Let $\mathbf{g} \colon \mathbb{R}^k \to \mathbb{R}^j$ be a function continuous at some point $\mathbf{c} \in \mathbb{R}^k$ that does not depend on n. Let $\{\mathbf{X}_n \in \mathbb{R}^k : n \in \mathbb{Z}_{++}\}$ be sequence of $k \times 1$ random vectors such that $\mathbf{X_n} \xrightarrow{p} \mathbf{c}$. Then, $\mathbf{g}(\mathbf{X}_n) \xrightarrow{p} \mathbf{g}(\mathbf{X})$ as $n \to \infty$. In other words,

$$\operatorname*{plim}_{n \to \infty} \mathbf{g}(\mathbf{X}_n) = \mathbf{g}\left(\operatorname*{plim}_{n \to \infty} \mathbf{X}_n\right), \tag{37}$$

if $\mathbf{g}(\cdot)$ is continuous at plim $\mathbf{X}_n$.