Econometrics I
TA Session 4

Jukina HATAKEYAMA *

May 7, 2024

Contents

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

1 Lebesgue-Stieltjes Expression

From the definition of the expectation of a random variable, we can symbolically write the expectation, by means of Lebesgue-Stieltjes integral, as follows:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \mathrm{d}F(x) := \begin{cases} \sum_i x_i P(X = x_i) & : \text{discrete random variable version;} \\ \int_{-\infty}^{\infty} x f(x) \mathrm{d}x & : \text{continuous random variable version.} \end{cases}$$

Here $P\colon \Omega \to \mathbb{R}$ stands for the probability (: probability mass function) that the realized value of X becomes x_i on a probability space $(\Omega, \mathcal{F}, P)$. Also, $f(x)$ stands for the probability density function defined as the derivative of the cummurative density function $F(x)\colon \mathbb{R} \to \mathbb{R}$:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x), \tag{1}$$

if the derivative exists.

2 Markov's inequality and Chebyshev's Inequality

In this section, we introduce two useful theorems which associate the distribution function of a random variable (or a function of a random variable) with the variables' expectation. First, the *Markov's inequality* is shown as follows.

Theorem 2.1 (Markov's Inequality)**.** If X is a non-negative random variable and δ is a positive constant, then

$$\mathbb{P}[X \geq \delta] \leq \frac{\mathbb{E}[X]}{\delta}. \tag{2}$$

Moreover, If $\phi\colon \mathbb{R} \to \mathbb{R}$ is a monotonically increasing nonnegative function for the non-negative reals, $X\colon \Omega \to \mathbb{R}$ is a random variable, $\delta \geq 0$, and $\phi(\delta) > 0$, then

$$\mathbb{P}[|X| \geq \delta] = \mathbb{P}[\phi(|X|) \geq \phi(\delta)] =\leq \frac{\mathbb{E}[\phi(|X|)]}{\phi(\delta)}. \tag{3}$$

Proof. We prove (2). Since the random variable X is a nonnegative random variable,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \mathrm{d}F(x) = \int_{0}^{\infty} x \mathrm{d}F(x). \tag{4}$$

From this we can derive

$$\begin{aligned} \mathbb{E}[X] &= \int_{0}^{\infty} x \mathrm{d}F(x) = \int_{0}^{\delta} x \mathrm{d}F(x) + \int_{\delta}^{\infty} x \mathrm{d}F(x) \\ &\geq \int_{\delta}^{\infty} x \mathrm{d}F(x) \geq \int_{\delta}^{\infty} \delta \mathrm{d}F(x) = \delta \int_{\delta}^{\infty} \mathrm{d}F(x) = \delta \mathbb{P}[X \geq \delta]. \quad (\because \delta \leq x \leq \infty) \end{aligned} \tag{5}$$

From this it is easy to see that (2) holds. A similar calculation yields the extended (or general) form of the Markov's inequality or (3). □

The Markov's inequality gives an upper bound for the probability that a non–negative function of a random variable is greater than or equal to some positive constant. Next, we present the *Chebyshev's inequality.*

Theorem 2.2 (Chebyshev's Inequality)**.** Let X be a (integrable) random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $\varepsilon > 0$,

$$\mathbb{P}[|X-\mu| \geq \varepsilon] \leq \frac{\mathbb{E}[|X-\mu|^2]}{\varepsilon^2} = \frac{\mathbb{V}[X]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}. \tag{6}$$

Proof. It can also be proved directly. Using the indicator function:

$$I(A) = \begin{cases} 1 & \text{if the event } A \text{ occurs;} \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

we have

$$\begin{aligned} \mathbb{P}[|X-\mu| \geq \varepsilon] &= \mathbb{E}[I(|X-\mu| \geq \varepsilon)] \\ &= \int_{-\infty}^{\infty} (|X-\mu| \geq \varepsilon) \, \mathrm{d}F(x) \\ &= \int_{-\infty}^{\infty} I\left(\left|\frac{X-\mu}{\varepsilon}\right| \geq 1\right) \mathrm{d}F(x) \\ &\leq \int_{-\infty}^{\infty} \left|\frac{X-\mu}{\varepsilon}\right| I\left(|\frac{X-\mu}{\varepsilon}| \geq 1\right) \mathrm{d}F(x) \\ &\leq \int_{-\infty}^{\infty} \left|\frac{X-\mu}{\varepsilon}\right|^2 I\left(|\frac{X-\mu}{\varepsilon}| \geq 1\right) \mathrm{d}F(x) \\ &\leq \int_{-\infty}^{\infty} \left|\frac{X-\mu}{\varepsilon}\right|^2 \mathrm{d}F(x) \\ &= \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} |X-\mu|^2 \, \mathrm{d}F(x) \\ &= \frac{\mathbb{E}[|X-\mu|^2]}{\varepsilon^2}. \end{aligned} \tag{8}$$

Rewriting this yields the Chebyshev's inequality. □

Chebyshev's inequality guarantees that, for a wide class of probability distributions, no more than a certain fraction of values can be more than a certain distance from the mean.

3 Law of Large Numbers

In this section, we will discuss important theorems, so called the law of large numbers (: LLN), which has a very important role in probability and statistics. LLN states that the average of a large number of i.i.d. random variables converges to the expected value. There are two main versions of the law, which are called the Weak and Strong LLN. The difference between them is mostly theoretical.

3.1 Strong Law of Large Numbers

The Strong Law of Large Numbers (: SLLN) states that the average of a large number of i.i.d. random variables converges almost surely to its expected value.

Theorem 3.1 (Strong Law of Large Numbers)**.** Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}\left[|X_1|^4\right] < \infty$ and $\mathbb{E}[X_1] = \mu$. Then,

$$\overline{X}_n \xrightarrow[n\to\infty]{a.s.} \mu,$$

where $\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.

The proof of this theorem is a little difficult. See Appendix A if you want to see the proof.

3.2 Weak Law of Large Numbers

The Weak Law of Large Numbers (: WLLN) states that the average of a large number of i.i.d. random variables converges in probability to the expected value.

Theorem 3.2 (Weak Law of Large Numbers)**.** Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}\left[|X_1|^2\right] < \infty$. Then,

$$\overline{X}_n \xrightarrow[n\to\infty]{p} \mu,$$

where $\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean..

Proof. We show that for all $\varepsilon > 0$, the following equality holds:

$$\lim_{n\to\infty} \mathbb{P}\left(\{|X_n - \mu| > \varepsilon\}\right) = 0, \tag{9}$$

From the assumption, we have

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\mu = \mu;$$
$$\mathbb{V}\left[\overline{X}_n\right] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2.$$

Substituting these into the Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}[X]}{\varepsilon^2} \Longleftrightarrow \mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}. \tag{10}$$

Therefore, taking a limit with respect to n results in

$$\lim_{n\to\infty} \mathbb{P}\left(\{|X_n - \mu| > \varepsilon\}\right) = 0, \tag{11}$$

which implies that $\overline{X}_n \xrightarrow{p} \mu$. □

SLLN is the probability of the limit, and WLLN is the limit of probability. This is why we call them weak and strong respectively.

3.3 Intuitive Understanding

To understand the LLN intuitively, we use the following example.

Example 3.1 (Tossing a Dice)**.** We consider tossing a (fair) dice (infinitely). Let $X_i \colon \Omega \to \mathbb{R}$ for $i \in \{1, 2, \ldots\}$ be a independently and identically distributed random variable with mean

$$\mathbb{E}\left[X_n\right] = \mu = \frac{1+2+3+4+5+6}{6} = 3.5. \tag{12}$$

This random variable represents the face showing after tossing a dice at the n th trial. Clearly, X_i and X_j for $i \neq j$ are independent. Our focus is placed on how the sample mean:

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

move as the number of trial n increases (infinitely). The graphical image is shown below.

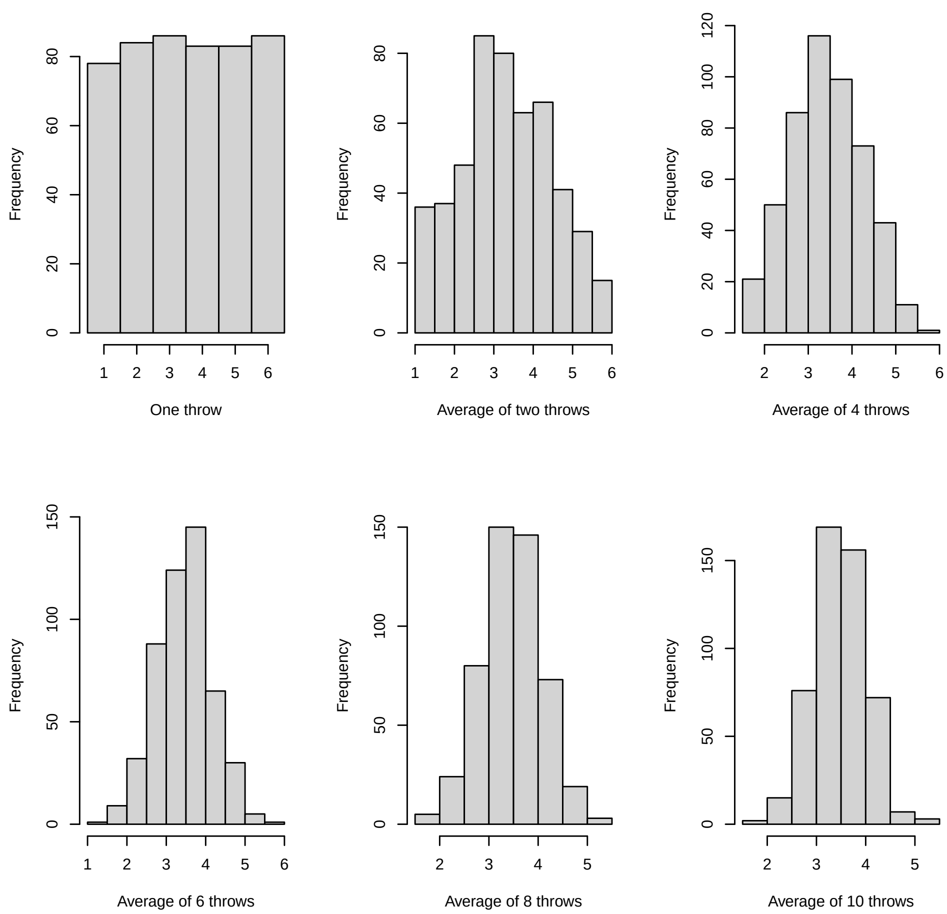

Figure 1: Example: dicetest

From this, we can confirm that the sample mean converges to the mean, i.e.,

$$\frac{1}{n}\sum_{i=1}^{\infty} X_i \xrightarrow[n\to\infty]{p} \mu. \tag{13}$$

From this, we can conclude that if you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value.

4 Characteristic Function and Moment Generating Function of a Random Variable

In this section, we introduce two important functions: characteristic function and moment generating function. The definition of these functions are given as follows:

Definition 4.1 (Characteristic Function and Moment Generating Function)**.** For a random variable $X\colon \Omega \to \mathbb{R}$, a function $\varphi_X\colon \mathbb{R} \to \mathbb{C}$ defined as

$$\varphi_X(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)] \tag{14}$$

is called the *characteristic function* of X. In addition, if there exists a function $M_X\colon \mathbb{R} \to \mathbb{R}$ defined as

$$M_X(t) := \mathbb{E}[e^{tX}] < \infty, \tag{15}$$

then the function M_X is called the *moment generating function* of X.

The characteristic function of a random variable always exists, since it is an integral of a bounded continuous function over a space whose measure is finite. However, the moment generating function does not always exists, that is, for some random variables Y, $M_Y(t) := \mathbb{E}[e^{tY}] = \infty$.

Proposition 4.1. For a random variable X which follows a normal distribution with mean 0 and variance $\sigma^2 < \infty$, we have

$$\varphi_X(t) = \exp\left\{-\frac{t^2\sigma^2}{2}\right\} \tag{16}$$

Proof. Directly calculating the characteristic function φ_X yields

$$\begin{aligned}
\varphi_X(t) = E[e^{itX}] &= \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{x^2 - 2it\sigma^2 x}{2\sigma^2}\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x - it\sigma^2)^2 - (it\sigma^2)^2}{2\sigma^2}\right\}\mathrm{d}x \\
&= \exp\left\{\frac{-(it\sigma^2)^2}{2\sigma^2}\right\}\underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x - it\sigma^2)^2}{2\sigma^2}\right\}\mathrm{d}x}_{=1} \\
&= \exp\left\{-\frac{t^2\sigma^2}{2}\right\}.
\end{aligned}$$

□

A similar calculation enables us to derive the moment generating function $M_X = \exp\left\{\frac{t^2\sigma^2}{2}\right\}$.

Proposition 4.2. For any two random variables, $X\colon \Omega_X \to \mathbb{R}$, and $Y\colon \Omega_Y \to \mathbb{R}$, the following statements are equavalent:

1. Both random variables follow the same distribution: $\mathbb{P}_X = \mathbb{P}_Y$;
2. The cummurative distribution function of X, F_X is the same as that of Y, F_Y: $F_X = F_Y$;
3. The characteristic function of X, φ_X is equal to that of Y, φ_X: $\varphi_X = \varphi_Y$

5 Central Limit Theorem

The Central Limit Theorem, one of the most striking and useful results in both probability and statistics, explains why the normal distribution appears in areas as diverse as gambling, measurement error, sampling, and statistical mechanics. Essentially, the Central Limit Theorem states that the normal distribution applies whenever one is approximating probabilities for a sum of many independent contributions all of which are roughly the same size. It is the Lindeberg–Feller Central Limit Theorem which makes this statement more precise in providing the sufficient, and in some sense necessary, Lindeberg condition whose satisfaction accounts for the ubiquitous appearance of the bell–shaped normal(, although here we omit the argument of the condition since it needs some intricate mathematics to understand).

In a Central Limit Theorem, we first standardise the sample mean $\overline{X}$ of a sequence of random variable X_i for $i \in \{1, \dots, n\}$, that is, we subtract from it its expected value $\mathbb{E}\left[\overline{X}\right]$ and we divide it by its standard deviation $\sqrt{\mathbb{V}\left[\overline{X}\right]}$. Then, we analyse the behaviour of its distribution as the sample size gets large. What happens is that the standardised sample mean converges in distribution to a normal distribution:

$$\frac{\overline{X} - \mathbb{E}\left[\overline{X}\right]}{\sqrt{\mathbb{V}\left[\overline{X}\right]}} \xrightarrow[n \to \infty]{d} Z, \tag{17}$$

where Z is a standard normal random variable. Intuitively, in the Law of Large Numbers, the variance of the sample mean converges to zero, while in the Central Limit Theorem the sample mean is multiplied by $\sqrt{n}$ so that its variance stays constant.

5.1 Lindeberg-Levy Central Limit Theorem

In the important case in which the variables X_i for $i \in \{1, \dots, n\}$ are independently and identically distributed (IID), the above formula becomes the one shown as follows.

Theorem 5.1 (Lindeberg–Levy Central limit theorem for a univariate random variable)**.** Assume that X_i for $ii \in \{1, \dots, n\}$ are a random sample from a probability distrubution with finite mean μ and finite positive variance σ^2, i.e.,

$$X_i \overset{i.i.d.}{\sim} (\mu, \sigma^2). \tag{18}$$

Define $\overline{X} = \frac{1}{n} \sum_1^n X_i$. Then,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow[n \to \infty]{d} N_{\mathbb{R}}(0, \sigma^2) \tag{19}$$

Proof. In this proof, we use the fact that a characteristic function of $\overline{X}$ converges to that of a random variable which follows a normal distribution. To prove the Lindberg–Levy CLT, we need to check whether

$$\mathbb{E}\left[\mathrm{e}^{it\sqrt{n}\overline{X}_n}\right] \to \exp\left\{-\frac{t^2\sigma^2}{2}\right\} \quad \text{as} \quad n \to \infty, \tag{20}$$

holds where

$$\overline{X}_n = \frac{1}{n}\sum_{s=1}^{n} X_i \quad \text{and} \quad X_i \overset{i.i.d.}{\sim} (0, \sigma^2). \tag{21}$$

First of all, from the assumption that $X_i \overset{i.i.d.}{\sim} (0, \sigma^2)$, we have

$$\mathbb{E}\left[\mathrm{e}^{it\sqrt{n}\overline{X}}\right] = \mathbb{E}\left[\mathrm{e}^{i\frac{t}{\sqrt{n}}\sum_{s=1}^{n} X_s}\right] \tag{22}$$

$$= \prod_{s=1}^{n} \underbrace{\mathbb{E}\left[\mathrm{e}^{i\frac{t}{\sqrt{n}}X_s}\right]}_{(a)}. \tag{23}$$

Here we apply the following theorem.

Lemma 5.1. The moment generating function of a sum of independent random variables is just the product of their moment generating functions. That is, for independent random variables X_i for $i \in \{1, \ldots, n\}$,

$$\mathbb{E}\left[\mathrm{e}^{i\sum_{i=1}^{n} X_n}\right] = \prod_{s=1}^{n} \mathbb{E}\left[\mathrm{e}^{iX_s}\right]. \tag{24}$$

(a) can be rewritten as below.

$$\mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}X_s}\right] = \mathbb{E}\left[1 + \underbrace{\frac{it}{\sqrt{n}}X_s}_{canceled} - \frac{t^2}{2n}X_s^2\right] + o\left(\frac{t^2}{n}\right) \tag{25}$$

$$\xrightarrow[n \to \infty]{p} \mathbb{E}\left[1 - \frac{t^2}{2n}X_s^2\right] \tag{26}$$

$$= 1 - \frac{t^2\sigma^2}{2n} \tag{27}$$

Thus, we can find that

$$E\left[e^{it\sqrt{n}\overline{X}}\right] = \prod_{s=1}^{n} \mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}X_s}\right] \tag{28}$$

$$= \left(1 - \frac{t^2\sigma^2}{2n}\right)^n \tag{29}$$

$$\xrightarrow[n \to \infty]{p} \mathrm{e}^{-\frac{t^2\sigma^2}{2}} = \mathbb{E}\left[\mathrm{e}^{itZ}\right], \tag{30}$$

where $Z \sim N_{\mathbb{R}}(0, \sigma^2)$. Recall that

$$\lim_{n\to\infty} \left(1 + \frac{m}{n}\right)^n = \mathrm{e}^m. \tag{31}$$

Thus, we can find (20) holds. □

The result is quite remarkable as it holds regardless of the parent distribution. An improtant extension allows us to relax the assumption of equal variances. The Lindeberg–Feller form of the central limit theorem, which appears in the following, is the centerpiece of most of the analysis in econometrics.

5.2 Lindeberg-Feller Central Limit Theorem

The Lindeberg-Feller Central Limit Theorem states in part that sums of independent random variables, properly standardised, converge in distribution to standard normal as long as a certain condition, called the Lindeberg condition, is satisfied.

Theorem 5.2 (Lindeberg–Feller Central Limit Theorem for a Univariate Random Variable)**.** Assume that X_i for $i \in \{1, \ldots, n\}$ are independent random variables from a probability distrubution with finite mean μ_i and finite positive variance σ_i^2, i.e.,

$$X_i \sim (\mu_i, \sigma_i^2). \tag{32}$$

Define

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i, \quad \text{and} \quad \overline{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2.$$

If no single term dominates this average variance, that is,

$$\lim_{n\to\infty} \frac{\max(\sigma_i)}{n\overline{\sigma}_n} = 0, \tag{33}$$

and if the average variance converges to a finite constant,

$$\overline{\sigma}^2 := \lim_{n\to\infty} \overline{\sigma}_n^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \sigma_i^2 < \infty \tag{34}$$

holds, then

$$\sqrt{n}(\overline{X}_n - \overline{\mu}) \xrightarrow[n \to \infty]{d} N_{\mathbb{R}}(0, \overline{\sigma}^2). \tag{35}$$

This result generalizes the Central Limit Theorem for independent and identically distributed sequences. In practical terms, the theorem states that sums of random variables, regardless of their form, will tend to be normally distributed. The result is yet more remarkable in that it does not require the variables in the sum to come from the same underlying distribution. It requires, essentially, only that the mean be a mixture of many random variables, none of which is large compared with their sum. Because nearly all the estimators we construct in econometrics fall under the purview of the central limit theorem, it is obviously an important result.

Remark 5.1 (Intuitive Understanding of the Lindeberg Condition)**.** Proof of the Lindeberg–Feller theorem requires some quite intricate mathematics that are well beyond the scope of our work here. We do note an important consideration in this theorem. The result rests on a condition known as the Lindeberg condition. The sample mean computed in the theorem is a mixture of random variables from possibly different distributions. The Lindeberg condition, in words, states that the contribution of the tail areas of these underlying distributions to the variance of the sum must be negligible in the limit. The condition formalizes the assumption in the above theorem that the average variance be positive and not be dominated by any single term. The condition is essentially impossible to verify in practice, so it is useful to have a simpler version of the theorem that encompasses it.

Appendix

A Proof of Theorem 1.3

The event

$$A := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n (= A_n \ i.o.) \tag{36}$$

represents "the event that infinitely many of the events A_n occurs." Then, we have the following theorem.

Theorem A.1 (Borel-Cantelli Lemma)**.** If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A) = 0$.

Proof. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then using the fact $A \subset \bigcup_{n=k}^{\infty} A_n$,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) = \lim_{k\to\infty} \mathbb{P}\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} \mathbb{P}(A_n) = 0. \tag{37}$$

□

Adding to the Borel-Cantelli Lemma, we use the following lemma in the following proof.

Lemma A.1. For a sequence of independent and identically distributed (i.i.d) random variables X_n with $\mathbb{E}[|X_1|^4] < \infty$ and $\mathbb{E}[X_1] = \mu$, there exists a constant $K < \infty$ such that for all $N \in Z_{++}$

$$\mathbb{E}\left[|S_n - n\mu|^4\right] \leq Kn^2. \tag{38}$$

where

$$S_n := X_1 + \cdots + X_n = \sum_{k=1}^{n} X_k. \tag{39}$$

Now we prove Theorem 1.3. For a sequence of independent and identically distributed (i.i.d) random variables X_n with $\mathbb{E}[|X_n|^4] < \infty$ and $\mathbb{E}[X_n] = \mu$ for $n \in Z_{++}$, let S_n be the sum from X_1 to X_n in the sequence:

$$S_n := X_1 + \cdots + X_n = \sum_{k=1}^{n} X_k. \tag{40}$$

By Markov's inequality,

$$\mathbb{P}\left(\left\{\frac{1}{n}|S_n - n\mu| \geq n^{-\gamma}\right\}\right) \leq \mathbb{E}\left[\frac{|S_n/n - \mu|^4}{n^{-4\gamma}}\right] \leq Kn^{-2+4\gamma} \tag{41}$$

holds. Also define $\gamma \in (0, 1/4)$ and

$$A_n := \left\{ \frac{1}{n} |S_n - n\mu| \geq n^{-\gamma} \right\}. \tag{42}$$

Then, by Borel-Cantelli Lemma, since

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) \leq \sum_{n=1}^{\infty} K n^{-2+4\gamma} < \infty \tag{43}$$

holds, we have

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \right) = 0. \tag{44}$$

(The regorous argument for (43) are shown in the following Remark.) On the other hand, the event A^c happens if and only if there exists a real number N such that for all $n \geq N$,

$$\left| \frac{S_n}{n} - \mu \right| < n^{-\gamma}, \tag{45}$$

which implies that $S_n/n = \overline{X}_n \xrightarrow{a.s.} \mu$. □

Remark A.1 (Almost sure convergence)**.** If a random variable X_n cnoverges to X almost surely, then

$$\text{Prob}\left(\{\omega;\ X_n \to X \quad \text{as} \quad n \to \infty\} \right) = 1, \tag{46}$$

or in another notation,

$$\mathbb{P}\left(\limsup_{n \to \infty} \{\omega; |X_n(\omega) - X(\omega)| > \varepsilon\} \right) = 0, \quad \text{for all } \varepsilon > 0. \tag{47}$$

Remark A.2 (Infinitely Sum of Series)**.** Consider the following summation:

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots. \tag{48}$$

Then, we can rewrite the above equation as follows:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^p} &= 1 + \frac{1}{2^p} + \frac{1}{3^p} + \frac{1}{4^p} + \frac{1}{5^p} + \cdots \\
&= 1 + \left\{ \frac{1}{2^p} + \frac{1}{3^p} \right\} + \left\{ \frac{1}{4^p} + \cdots + \frac{1}{7^p} \right\} + \cdots \\
&< 1 + \left\{ \frac{1}{2^p} + \frac{1}{2^p} \right\} + \left\{ \frac{1}{4^p} + \cdots + \frac{1}{4^p} \right\} + \cdots \\
&= 1 + \frac{1}{2^{p-1}} + \frac{1}{4^{p-1}} + \frac{1}{8^{p-1}} + \cdots \\
&= \sum_{n=1}^{\infty} 2^{(1-p)(n-1)} \\
&= \begin{cases} \infty & \text{if} \quad 0 \leq p \leq 1; \\ \frac{1}{1-2^{(1-p)}} (< \infty) & \text{if} \quad p \geq 1. \end{cases}
\end{aligned}$$

In fact, when $0 \leq p \leq 1$, (48) tends to infinity(, although the proof is omitted).

B Program for Coin Tossing Figure.

```
N <- 500

# Layout the plots in 2 rows and 3 columns
m <- t(matrix(seq(1, 6), 3, 2))

# Open a PDF file for output
pdf("CLT_dice_tossing.pdf")

layout(m)

# Throw the dice several times
s1 <- as.integer(runif(N, 1, 7))
s2 <- as.integer(runif(N, 1, 7))
s3 <- as.integer(runif(N, 1, 7))
s4 <- as.integer(runif(N, 1, 7))
s5 <- as.integer(runif(N, 1, 7))
s6 <- as.integer(runif(N, 1, 7))
s7 <- as.integer(runif(N, 1, 7))
s8 <- as.integer(runif(N, 1, 7))
s9 <- as.integer(runif(N, 1, 7))
s10 <- as.integer(runif(N, 1, 7))

bins <- 8

# Plot each histogram
hist(s1, main = "",
        xlab = "One throw",
        breaks = seq(0, 6) + 0.5)
hist((s1 + s2) / 2,
        breaks = bins, main = "",
        xlab = "Average of two throws")
hist((s1 + s2 + s3 + s4) / 4,
        breaks = bins,
        main = "",
        xlab = "Average of 4 throws")
hist((s1 + s2 + s3 + s4 + s5 + s6) / 6,
        breaks = bins,
        main = "",
        xlab = "Average of 6 throws")

bins <- 12

hist((s1 + s2 + s3 + s4 + s5 + s6 + s7 + s8) / 8,
         breaks = bins, main = "",
         xlab = "Average of 8 throws")
hist((s1 + s2 + s3 + s4 + s5 + s6 + s7 + s8 + s9 + s10) / 10,
        breaks = bins, main = "",
        xlab = "Average of 10 throws")

# Close the PDF file
dev.off()
```