Convergences and Limit Theorems

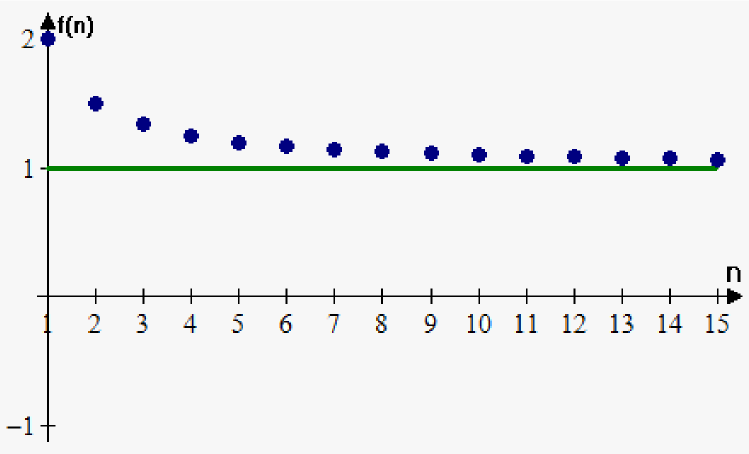
Professor Yongmiao Hong
Cornell University
April 16, 2020

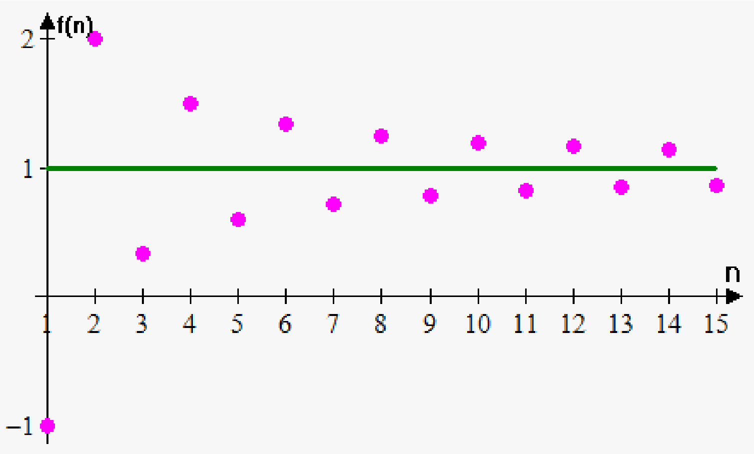
CONTENTS

Limits and Orders of Magnitude: A Review

Definition 1 (7.1). [Limit]

Let $\{b_n, n = 1, 2, \cdots\}$ be a sequence of nonstochastic real numbers. If there exists a real number b such that for every real number $\epsilon > 0$, there exists a finite integer $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, we have $|b_n - b| < \epsilon$, then b is the limit of the sequence $\{b_n, n = 1, 2, \cdots\}$.





Limits and Orders of Magnitude: A Review

Remarks:

- When $\{b_n\}$ converges to b as $n \to \infty$, we write

$$b_n \to b \text{ as } n \to \infty,$$

or

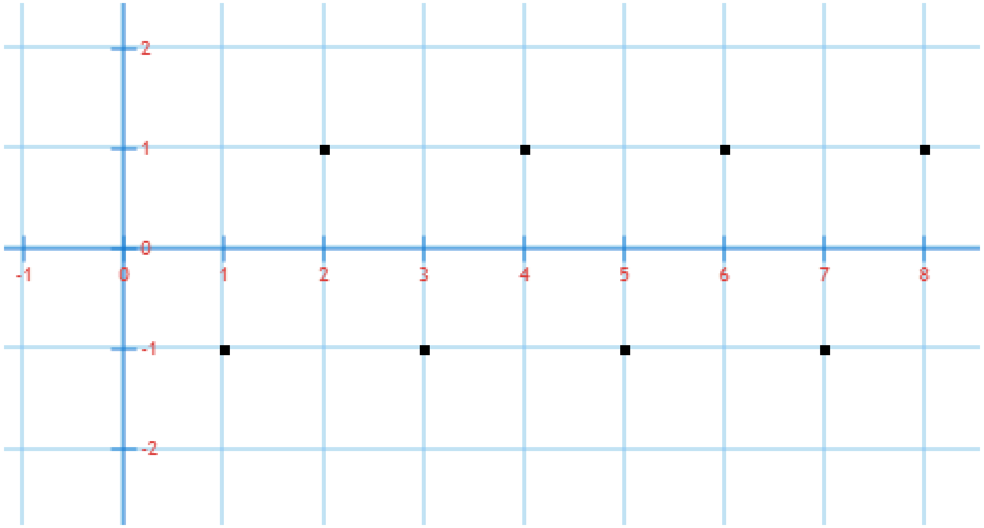
$$\lim_{n\to\infty} b_n = b.$$

- The constant $\epsilon > 0$ can be set to be arbitrarily small. The smaller ϵ is, the larger $N(\epsilon)$ will be. One can interpret ϵ as a prespecified **tolerance level** for the discrepancy between b_n and b.

Limits and Orders of Magnitude: A Review

Example 1 (7.1)

$b_n = (-1)^n$. Then $\{b_n, n = 1, 2, \cdots\}$ is bounded by a constant in the sense that $|b_n| \leq M$ for some constant $M > 1$ and for all $n \geq 1$. However, its limit does not exist.

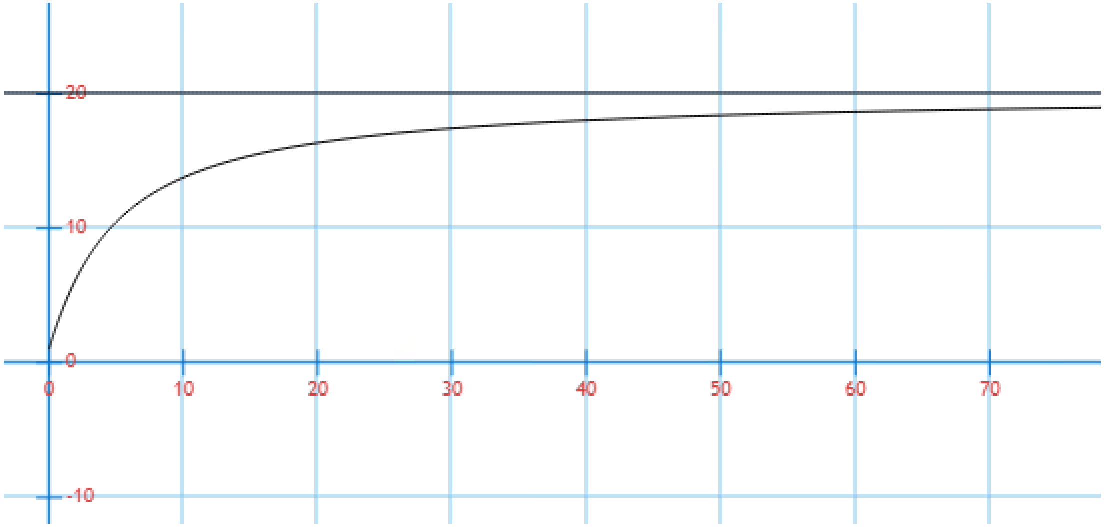


Limits and Orders of Magnitude: A Review

Example 2 (7.2)

Let $b_n = (1 + a/n)^n$, where a is a constant. Then

$$b_n \to e^a \text{ as } n \to \infty.$$



$$(1 + 3/n)^n \to e^3 \text{ as } n \to \infty.$$

Limits and Orders of Magnitude: A Review

Example 3 (7.3)

$b_n = (-1)^n$. Then $\{b_n, n = 1, 2, \cdots\}$ is bounded by a constant in the sense that $|b_n| \leq M$ for some constant $M > 1$ and for all $n \geq 1$. However, its limit does not exist.

Solution

Set $\epsilon = \frac{1}{2}$. Then there exist no b and $N(\epsilon)$ such that for all $n > N(\epsilon)$, we can have $|b_n - b| < \epsilon$.

Limits and Orders of Magnitude: A Review

Definition 2 (7.2). [Continuity]

The function $g : \mathbb{R} \to \mathbb{R}$ is continuous at point b if for any sequence $\{b_n\}$ such that $b_n \to b$ as $n \to \infty$, we have

$$g(b_n) \to g(b) \text{ as } n \to \infty.$$

Limits and Orders of Magnitude: A Review

Remarks:

- An alternative but equivalent definition of continuity: for each given $\epsilon > 0$ there exists a $\delta = \delta(\epsilon)$ such that whenever $|b_n - b| < \delta$, we have $|g(b_n) - g(b)| < \epsilon$.
- When $g(\cdot)$ is continuous at b, we can write

$$\lim_{b_n \to b} g(b_n) = g(\lim_{n \to \infty} b_n) = g(b).$$

In other words, the limit of a sequence of values for a continuous function is equal to the value of the function at the limit.

Limits and Orders of Magnitude: A Review

Example 4 (7.4)

Suppose $a_n \to a$ and $b_n \to b$ as $n \to \infty$. Then as $n \to \infty$,

(1) $a_n + b_n \to a + b$;

(2) $a_n b_n \to ab$;

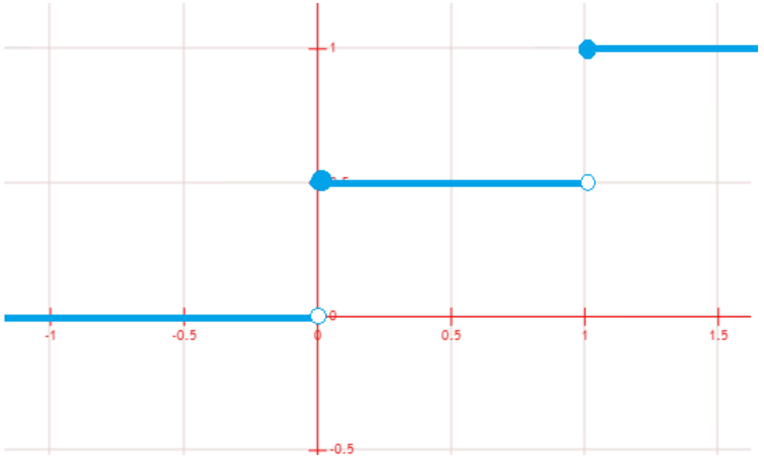
(3) $a_n / b_n \to a/b$ if $b \neq 0$.

Limits and Orders of Magnitude: A Review

Example 5 (7.5)

Define a function

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{2} & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$



Here, $F(x)$ is neither continuous at 0 nor at 1. This is because for at least one sequence $\{b_n\}$ such that $b_n = -\frac{1}{n}$, we have $b_n \to b = 0$, but $F(b_n) = F(-\frac{1}{n}) = 0$ for all $n \ge 1$, and $\lim_{n\to\infty} F(-\frac{1}{n}) = 0 \neq F(0) = \frac{1}{2}$.

Limits and Orders of Magnitude: A Review

Definition 3 (7.3). [Order of Magnitude]

(1) A sequence $\{b_n\}$ is at most of order n^λ, denoted $b_n = O(n^\lambda)$ or $n^{-\lambda}b_n = O(1)$, if for **some** (sufficiently large) real number $M < \infty$, there exists a finite integer $N(M)$ such that for all $n \geq N(M)$, we have

$$\left|n^{-\lambda}b_n\right| < M.$$

(2) A sequence $\{b_n\}$ is of order smaller than n^λ, denoted $b_n = o(n^\lambda)$ or $n^{-\lambda}b_n = o(1)$, if for **every** real number $\epsilon > 0$ there exists a finite integer $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, we have

$$|n^{-\lambda}b_n| < \epsilon.$$

Limits and Orders of Magnitude: A Review

Remarks:

- In the definition of $b_n = O(n^\lambda)$, the constant M is usually set to be a very big number. Note that it suffices to find one constant M only.

- For $\lambda > 0$, $b_n = O(n^\lambda)$ implies that b_n grows to infinity at a rate slower than or at most equal to n^λ. In particular, if $\lim_{n\to\infty} \frac{b_n}{n^\lambda} = C < \infty$,then $b_n = O(n^\lambda)$ or $n^{-\lambda}b_n = O(1)$.

Limits and Orders of Magnitude: A Review

Example 6 (7.6)

$b_n = 4 + 2n + 6n^2$. Then $b_n = O(n^2)$, because

$$\begin{aligned} \frac{b_n}{n^2} &= \frac{4}{n^2} + \frac{2n}{n^2} + \frac{6n^2}{n^2} \\ &\to 6 < 2M = 2 \cdot 6 \text{ (say)} \end{aligned}$$

for all n sufficiently large. Intuitively, the order of b_n is determined by the dominant term (i.e., n^2) that grows to infinity fastest.

Remark

- It is possible that $\lim_{n\to\infty} b_n/n^\lambda$ does not exist but $|b_n/n^\lambda|$ is bounded; in this case, we still have $b_n = O(n^\lambda)$.

Limits and Orders of Magnitude: A Review

Example 7 (7.7)

$b_n = (-1)^n$. Then $b_n = O(1)$.

$$|b_n| = 1 < M \equiv 1.01 \text{ for all } n \geq 1.$$

Remark

- In the definition of $b_n = o(n^\lambda)$, the constant ϵ can be set to be a very small value. Intuitively, $b_n = o(n^\lambda)$ implies that b_n grows at a rate strictly slower than n^λ. That is,

$$\lim_{n\to\infty} \frac{b_n}{n^\lambda} = 0.$$

Limits and Orders of Magnitude: A Review

Example 8 (7.8)

$b_n = 4 + 2n + 6n^2$. Then $b_n = o(n^{2+\delta})$ for all $\delta > 0$.

Remark

- If $b_n = o(n^\lambda)$, then $b_n = O(n^\lambda)$. Intuitively, if b_n grows at a rate slower than n^λ, it will grow at most at a rate of n^λ.

Limits and Orders of Magnitude: A Review

Lemma 1 (7.1)

Let a_n and b_n be scalars.

(1) If $a_n = O(n^\lambda)$ and $b_n = O(n^\mu)$, then $a_n b_n = O(n^{\lambda+\mu})$, $a_n + b_n = O(n^\kappa)$, where $\kappa = \max(\lambda, \mu)$.

(2) If $a_n = o(n^\lambda)$ and $b_n = o(n^\mu)$, then $a_n b_n = o(n^{\lambda+\mu})$, $a_n + b_n = o(n^\kappa)$, where $\kappa = \max(\lambda, \mu)$.

(3) If $a_n = O(n^\lambda)$ and $b_n = o(n^\mu)$, then $a_n b_n = o(n^{\lambda+\mu})$, $a_n + b_n = O(n^\kappa)$, where $\kappa = \max(\lambda, \mu)$.

Proof

Limits and Orders of Magnitude: A Review

Proof:

- (1) Because a_n grows at most at rate n^λ, b_n grows at most at rate n^μ, the product $a_n b_n$ will grow at most at rate $n^{\lambda+\mu}$. This is because for all n sufficiently large (i.e., for all $n \geq N(M)$),

$$\begin{aligned} \left|\frac{a_n b_n}{n^{\lambda+\mu}}\right| &= \left|\frac{a_n}{n^\lambda}\frac{b_n}{n^\mu}\right| \\ &\leq M \cdot M = M^2. \end{aligned}$$

On the other hand, the sum $a_n + b_n$ will be dominated by the term that grows to infinity faster: For all n sufficiently large,

$$\begin{aligned} \left|\frac{a_n + b_n}{n^\kappa}\right| &= \left|\frac{a_n}{n^\kappa} + \frac{b_n}{n^\kappa}\right| \\ &\leq \epsilon + M \leq 2M. \end{aligned}$$

Therefore, $a_n + b_n = O(n^\kappa)$.

To be Continued

Limits and Orders of Magnitude: A Review

Proof:

- (2) Similar to the proof of Result (1).
- (3) The product $a_n b_n = o(n^{\lambda+\mu})$ because

$$\begin{aligned} \left|\frac{a_n b_n}{n^{\lambda+\mu}}\right| &= \left|\frac{a_n}{n^{\lambda}} \frac{b_n}{n^{\mu}}\right| \\ &\leq M \cdot \left|\frac{b_n}{n^{\mu}}\right| \\ &\to 0 \text{ as } n \to \infty \end{aligned}$$

given $a_n = O(n^{\lambda})$ and $b_n = o(n^{\mu})$.

Limits and Orders of Magnitude: A Review

Example 9 (7.9)

Suppose $a_n = O(1)$, and $b_n = o(1)$. Then

$$\begin{aligned} a_n b_n &= o(1), \\ a_n + b_n &= O(1). \end{aligned}$$

CONTENTS

Motivation for Convergence Concepts

Question: Why do we need convergence concepts in econometrics and statistics?

- A random sample $\mathbf{X}^n = (X_1, \cdots, X_n)$ of size n is a sequence of random variables $X_1, \cdots, X_n$. It can be viewed as an n-dimensional real-valued random vector, where the dimension n may go to infinity.

- Its realization is an n-dimensional vector $\mathbf{x}^n = (x_1, ..., x_n)$. A realization $\mathbf{x}^n$ of $\mathbf{X}^n$ is usually called a sample point or a data set generated from the random sample $\mathbf{X}^n$.

Motivation for Convergence Concepts

- When $\mathbf{X}^n$ is an IID random sample from a population PMF/PDF $f_X(\cdot)$, the joint PMF/PDF of the random sample $\mathbf{X}^n$ is given by

$$f_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_{i=1}^{n} f_X(x_i).$$

  It completely describes the probability law of the random sample $\mathbf{X}^n$.

- Often, the population distribution $f_X(x)$ is assumed to be a parametric model in the sense that $f_X(x) = f(x, \theta)$ for some value of a finite-dimensional parameter θ, where the functional form of $f(\cdot, \cdot)$ is known but θ is unknown.

Motivation for Convergence Concepts

– For example, if $f_X(x)$ is assumed to follow a $N(\mu, \sigma^2)$ distribution, we have

$$\begin{aligned} f_X(x) &= f(x, \theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \end{aligned}$$

where $\theta = (\mu, \sigma^2)$.

- One important objective of statistical analysis is to estimate the unknown parameter θ when we are given a data set $\mathbf{x}^n$, which is a realization of the random sample $\mathbf{X}^n$. An estimator for θ is a function of $\mathbf{X}^n$, and so it is a statistic.

Motivation for Convergence Concepts

- To motivate the importance of various convergence concepts, we consider two simple statistics—the sample mean and sample variance. They are used to estimate population mean μ and population variance σ^2 respectively.
  - The sample mean estimator

$$\bar{X}_n = T(\mathbf{X}^n) = n^{-1} \sum_{i=1}^{n} X_i.$$

  - The sample variance estimator

$$S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

  - If the sample size n is sufficiently large, then it is expected that $\bar{X}_n$ will be "close" to μ and S_n^2 will be "close" to σ^2.

Motivation for Convergence Concepts

Question: How can one measure the closeness of $\bar{X}_n$ to μ and the closeness of S_n^2 to σ^2?

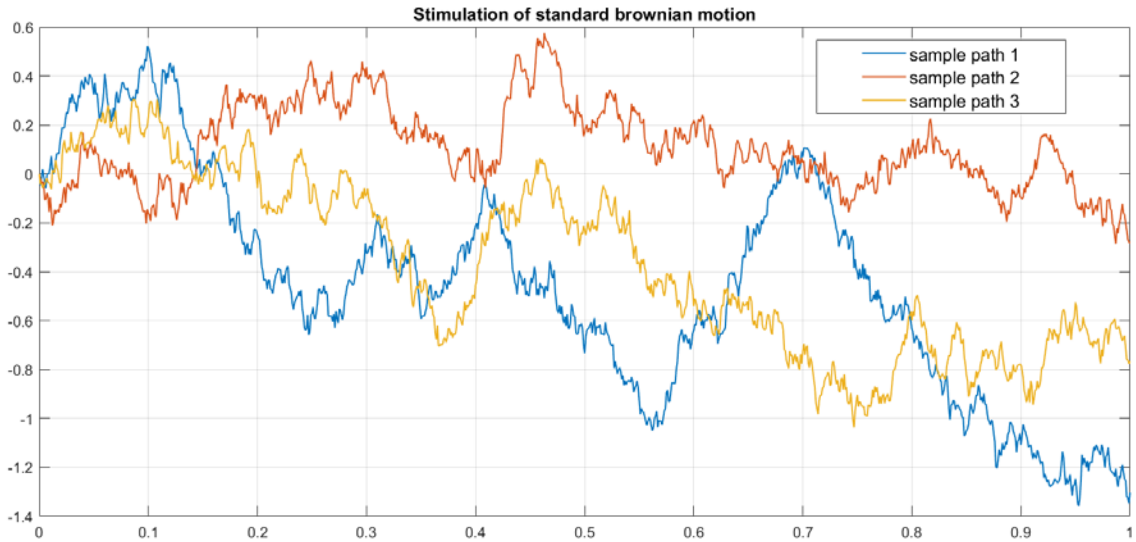
- [**Randomness of Estimators**] Since $\bar{X}_n$ and S_n^2 are random variables, the convergence concepts for nonstochastic sequences do not apply.
- Both $\bar{X}_n$ and S_n^2 are mappings from S to the real line; that is, $\bar{X}_n : S \to \mathbb{R}$ and $S_n^2 : S \to \mathbb{R}^+$, where S is the sample space of the underlying random experiment.

Motivation for Convergence Concepts

- Suppose a random experiment is conducted, a basic outcome $s \in S$ occurs. Then we observe a data set $\mathbf{x}^n = (x_1, \cdots, x_n)$, where $x_i = X_i(s)$. This is a realization of the random sample $\mathbf{X}^n$. From the data set $\mathbf{x}^n$, we can compute an estimate $\bar{x}_n = \bar{X}_n(s)$ for μ and an estimate $s_n^2 = S_n^2(s)$ for σ^2. Different outcomes s will yield different estimates for μ and σ^2 respectively for any given n.

Motivation for Convergence Concepts

- **[Sample Path]** In fact, each basic outcome $s \in S$ will generate a sequence of real numbers $\{\bar{x}_n = \bar{X}_n(s), n = 1, 2, \cdots\}$ and a sequence of real numbers $\{s_n^2 = S_n^2(s), n = 1, 2, \cdots\}$ respectively. These nonstochastic sequences are called a **sample path** for the sample mean $\bar{X}_n$ and a sample path for the sample variance S_n^2 respectively, when a basic outcome s occurs. There are many such sample paths for both $\bar{X}_n$ and S_n^2 respectively, corresponding to different basic outcomes $s \in S$.



Motivation for Convergence Concepts

- We will develop various suitable convergence concepts and distance measures between $\bar{X}_n$ and μ, or between S_n^2 and σ^2. A **common feature** of these different convergence concepts is that they define that **most nonstochastic sequences** $\{\bar{X}_n(s), n = 1, 2, \cdots\}$ and $\{S_n^2(s), n = 1, 2, \cdots\}$ **converge** to μ and σ^2 respectively.

Question: What is meant by ”most nonstochastic sequences” for $\bar{X}_n$ or $\bar{S}_n^2$?

CONTENTS

Convergence in Quadratic Mean and L_p-Convergence

Definition 4 (7.4). [Convergence in Quadratic Mean]

Let $\{Z_n, n = 1, 2, \cdots\}$ be a sequence of random variables and Z be a random variable. Then the sequence $\{Z_n\}$ converges in quadratic mean (or converges in mean square) to Z if

$$E(Z_n - Z)^2 \to 0 \text{ as } n \to \infty,$$

or equivalently,

$$\lim_{n\to\infty} E(Z_n - Z)^2 = 0.$$

It is also denoted as

$$Z_n \overset{q.m.}{\to} Z$$

or

$$Z_n - Z = o_{q.m.}(1).$$

Convergence in Quadratic Mean and L_p-Convergence

Remarks:

- The convergence in quadratic mean means that the **weighted average** of the **squared deviations** between $Z_n(s)$ and $Z(s)$ vanishes to 0 as $n \to \infty$, where the average is taken over all possible basic outcomes $\{s\}$ **weighted** by their **probabilities of occuring**.
- When Z_n converge to Z in quadratic mean, it is possible that there exist some sample paths for which $Z_n(s)$ does not converge to $Z(s)$. However, the quadratic deviations of these sample paths all together weighted by the probabilities of their occurings become negligible as n becomes large.

Convergence in Quadratic Mean and L_p-Convergence

Example 10 (7.10)

Suppose $X^n = (X_1, \cdots, X_n)$ is an IID random sample from a population with mean μ and variance σ^2. Define $Z_n = \bar{X}_n$. Show

$$\bar{X}_n \stackrel{q.m.}{\to} \mu.$$

Solution

Convergence in Quadratic Mean and L_p-Convergence

Solution

- It suffices to show $\lim_{n\to\infty} E(\bar{X}_n - \mu)^2 = 0$. Noting $E(\bar{X}_n) = \mu$, we have

$$
\begin{aligned}
E(\bar{X}_n - \mu)^2 &= \operatorname{var}(\bar{X}_n) \\
&= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(X_i) \\
&= \frac{\sigma^2}{n} \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}
$$

Convergence in Quadratic Mean and L_p-Convergence

Definition 5 (7.5). [L_p-Convergence]

Let $0 < p < \infty$, and let $\{Z_n, n = 1, 2, \cdots\}$ be a sequence of random variables with $E|Z_n|^p < \infty$, and let Z be a random variable with $E|Z|^p < \infty$. Then Z_n converges in L_p to Z if

$$\lim_{n\to\infty} E|Z_n - Z|^p = 0.$$

Convergence in Quadratic Mean and L_p-Convergence

Remarks:

- In connection with L_p-convergence, the following inequalities are useful:

- Holder's inequality

$$E|XY| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q},$$

where $p > 1$ and $1/p + 1/q = 1$.

- Minkowski's inequality

$$E|X+Y|^p \leq \left[(E|X|^p)^{1/p} + (E|Y|^p)^{1/p} \right]^p$$

for $p \geq 1$.

Convergence in Quadratic Mean and ***L_p***-Convergence

Question: What happens if Z_n and Z are $d \times 1$ random vectors, where d is fixed (i.e., d does not change as $n \to \infty$)?

- A sequence of random vectors $\{Z_n\}$ converges to Z in L_p, if each component of the vector Z_n, $Z_{in} \xrightarrow{L_p} Z_i$ for $i = 1, \cdots, d$.
- In other words, component-wise convergences ensure joint convergence of the entire vector Z_n, and vice versa.

CONTENTS

Convergence in Probability

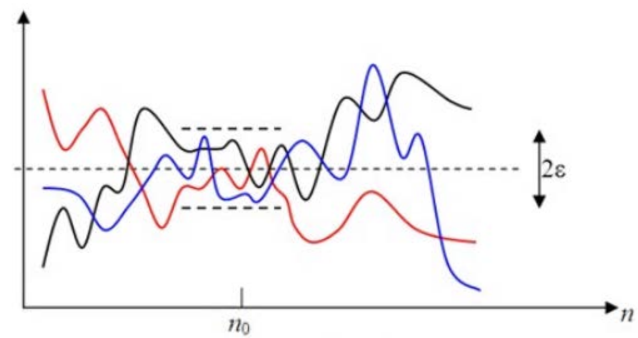
Definition 6 (7.6). [Convergence in Probability]

A sequence of random variables $\{Z_n, n = 1, 2, \cdots\}$ converges in probability to a random variable Z if for every small constant $\epsilon > 0$,

$$P[|Z_n - Z| > \epsilon] \to 0 \text{ as } n \to \infty.$$

When Z_n converges in probability to Z, we write

$$\lim_{n\to\infty} P(|Z_n - Z| > \epsilon) = 0$$



for every $\epsilon > 0$, or $p\lim_{n\to\infty} Z_n = Z$, or $Z_n \xrightarrow{p} Z$, or $Z_n - Z = o_P(1)$, or $Z_n - Z \xrightarrow{p} 0$.

Convergence in Probability

Remarks:

- Convergence in probability is also called **weak convergence**.
- The constant $\epsilon > 0$ could be viewed as a prespecified **cutoff point** such that the difference $|Z_n - Z|$ will be considered as a "**large deviation**" if $|Z_n - Z| > \epsilon$, and as a "**small deviation**" if $|Z_n - Z| \leq \epsilon$.
- An alternative definition of convergence in probability: Given any $\epsilon > 0$ and any $\delta > 0$, there exists a finite integer $N = N(\epsilon, \delta)$ such that for all $n > N$, we have

$$P(|Z_n - Z| > \epsilon) < \delta.$$

Convergence in Probability

- For n sufficiently large, the probability that the difference $|Z_n - Z|$ takes large values (i.e., larger than ϵ) is rather small. In other words, if Z_n converges in probability to Z, Z_n will be arbitrarily close to Z with a very high probability when n is sufficiently large.

- Define a set in sample space S :

$$A_n(\epsilon) = \{s \in S : |Z_n(s) - Z(s)| \leq \epsilon\},$$

i.e., $A_n(\epsilon)$ is a subset of S that consists of all basic outcomes $s \in S$ such that the difference $|Z_n(s) - Z(s)|$ is small. When Z_n converges in probability to Z, the probability of "small deviations"

$$\begin{aligned} P[|Z_n - Z| &\leq \epsilon] = P[A_n(\epsilon)] \\ &\to 1 \text{ as } n \to \infty. \end{aligned}$$

For this reason, convergence in probability is also called **convergence with probability approaching 1**.

Convergence in Probability

- When $Z_n \xrightarrow{p} b$, where b is a constant, we say that Z_n is consistent for b, and b is the probability limit of Z_n, denoted as
$$b = p \lim_{n\to\infty} Z_n.$$

Convergence in Probability

Example 11 (7.11)

Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample from a $U[0, \theta]$ distribution, where $\theta > 0$ is an unknown parameter. Define a statistic $Z_n = \max_{1 \leq i \leq n}(X_i)$. Is Z_n consistent for θ?

Solution

Convergence in Probability

Solution

- Given $\{|Z_n - Z| > \epsilon\} = \{Z_n - Z > \epsilon\} \cup \{Z_n - Z < -\epsilon\}$, we have

$$\begin{aligned}
& P(|Z_n - \theta| > \epsilon) \\
&= P(Z_n > \theta + \epsilon) + P(Z_n < \theta - \epsilon) \\
&= P(Z_n < \theta - \epsilon) \\
&= P\left[\max_{1 \le i \le n}(X_i) < \theta - \epsilon\right] \\
&= P(X_1 < \theta - \epsilon, X_2 < \theta - \epsilon, \cdots, X_n < \theta - \epsilon) \\
&= \prod_{i=1}^{n} P(X_i < \theta - \epsilon) \text{ by independence} \\
&= \left(\frac{\theta - \epsilon}{\theta}\right)^n \\
&= \left(1 - \frac{\epsilon}{\theta}\right)^n \\
&\to 0 \text{ as } n \to \infty \text{ for any given } \epsilon > 0.
\end{aligned}$$

It follows that Z_n is consistent for θ.

Convergence in Probability

- **Remarks:**
  - $Z_n = \max_{1 \leq i \leq n} |X_i|$ is called an order statistic which involves some sort of ranking for the n random variables in the random sample $\mathbf{X}^n$.
- **Convergence in probability with order** n^α, where α can be a positive or negative number:
  - The sequence of random variables $\{Z_n\}$ is said to be of order smaller than n^α in probability if

$$Z_n/n^\alpha \xrightarrow{p} 0 \text{ as } n \to \infty.$$

This is denoted as $Z_n = o_P(n^\alpha)$.

  - The sequence of random variables $\{Z_n\}$ is said to be at most of order n^α in probability if for any given $\delta > 0$, there exist a constant $M = M(\delta) < \infty$ and a finite integer $N = N(\delta)$, such that

$$P(|Z_n/n^\alpha| > M) < \delta$$

for all $n > N$. This is denoted as $Z_n = O_P(n^\alpha)$.

Convergence in Probability

- **Remarks:**
  - For $Z_n = O_P(n^\alpha)$ with $\alpha > 0$, the order n^α is the fastest growth rate at which Z_n goes to infinity with probability approaching one.
  - When $\alpha < 0$, the order n^α is the fastest convergence rate at which Z_n vanishes to 0 with probability approaching one.
  - $Z_n = O_P(n^\alpha)$ involves the concept of boundedness in probability.

Convergence in Probability

Definition 7 (7.7). [Boundedness in Probability]

For every constant $\delta > 0$, there exists a constant $M = M(\delta)$ and an integer $N = N(\delta)$ such that

$$P(|Z_n| > M) < \delta$$

for all $n \geq N$. Then $Z_n = O_P(1)$ and Z_n is called bounded in probability.

Convergence in Probability

Remarks:

- $Z_n = O_P(1) = O_P(n^0)$ implies that for n sufficiently large, the event that $|Z_n|$ takes values larger than a very large constant has a tiny probability.
- In other words, $|Z_n|$ is bounded by a constant with a very high probability for all n sufficiently large.

Convergence in Probability

Example 12 (7.12)

If $Z_n \sim N(0,1)$ for all $n \geq 1$. Then $Z_n = O_P(1)$ because for any given $\delta > 0$, there exists a finite constant $M = \Phi^{-1}(1 - \frac{\delta}{2}) < \infty$, where $\Phi(\cdot)$ is the N(0,1) CDF, such that

$$P(|Z_n| > M) = 2[1 - \Phi(M)] = \delta < 2\delta$$

for all $n \geq 1$.

Remarks

What is the probability that a $N(0,1)$ random variable takes values less than 3 in abolute value?

$$P(|X| < 3) = 0.997?$$

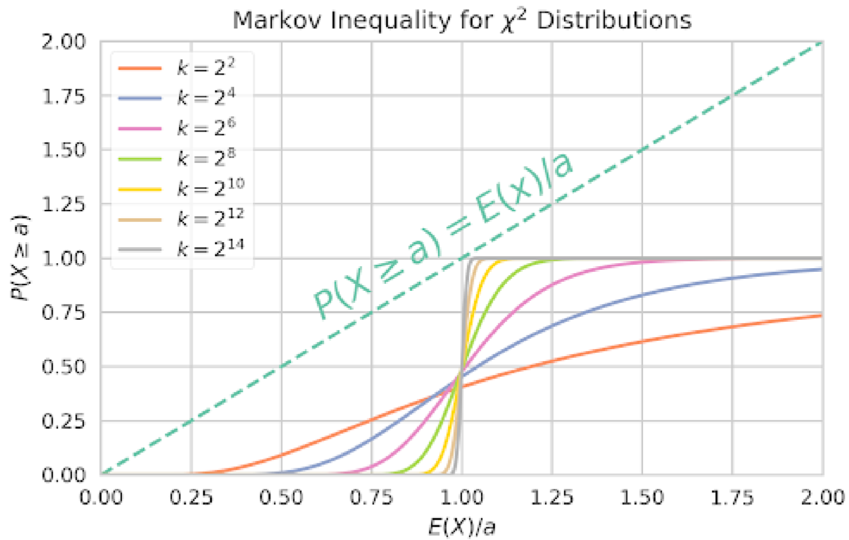
Convergence in Probability

Lemma 2 (7.2). [Markov's Inequality]

Suppose X is a random variable and $g(X)$ is a nonnegative function. Then for any $\epsilon > 0$, and any $k > 0$, we have

$$P[g(X) \geq \epsilon] \leq \frac{E[g(X)^k]}{\epsilon^k}.$$

安德雷 • 马尔可夫



Proof

Convergence in Probability

Proof:

- Let $\mathbf{1}(\cdot)$ be the indicator function that takes values 1 and 0, depending on whether the statement is true.
- Then

$$
\begin{aligned}
P[g(X) > \epsilon] \\
&= \int_{\{x:g(x)>\epsilon\}} dF_X(x) \\
&= \int_{-\infty}^{\infty} \mathbf{1}[g(x) > \epsilon] dF_X(x) \\
&\leq \int_{-\infty}^{\infty} \mathbf{1}[g(x) > \epsilon] \frac{g(x)^k}{\epsilon^k} dF_X(x) \\
&\leq \int_{-\infty}^{\infty} \frac{g(x)^k}{\epsilon^k} dF_X(x) \\
&= \frac{1}{\epsilon^k} E[g(X)^k].
\end{aligned}
$$

Convergence in Probability

Remarks:

- Markov's inequality bounds the **tail probability** by a **moment** condition. The thickness of the tail probability depends on the magnitude of moments of the distribution.
- When $k = 2$ and $g(x) = |x|$, Markov's inequality is called **Chebyshev's inequality**.

$$P[|X| \geq \epsilon] \leq \frac{E[|X|^2]}{\epsilon^2}.$$

巴夫尼提・列波维奇・切比雪夫

Convergence in Probability

Lemma 3 (7.3). [Bernstein's inequality]

Let $X_1, \cdots, X_n$ be independent random variables with mean zero and bounded support: $|X_i| < M$ for all $i = 1, \cdots, n$. Let $\sigma_i^2 = \text{var}(X_i)$. Suppose $V_n \geq \Sigma_{i=1}^n \sigma_i^2$. Then for each constant $\epsilon > 0$,

$$P\left[\left|\sum_{i=1}^{n} X_i\right| > \epsilon\right] \leq 2e^{-\frac{1}{2}\epsilon^2/(V_n + \frac{1}{3}M\epsilon)}.$$

Convergence in Probability

Remarks:

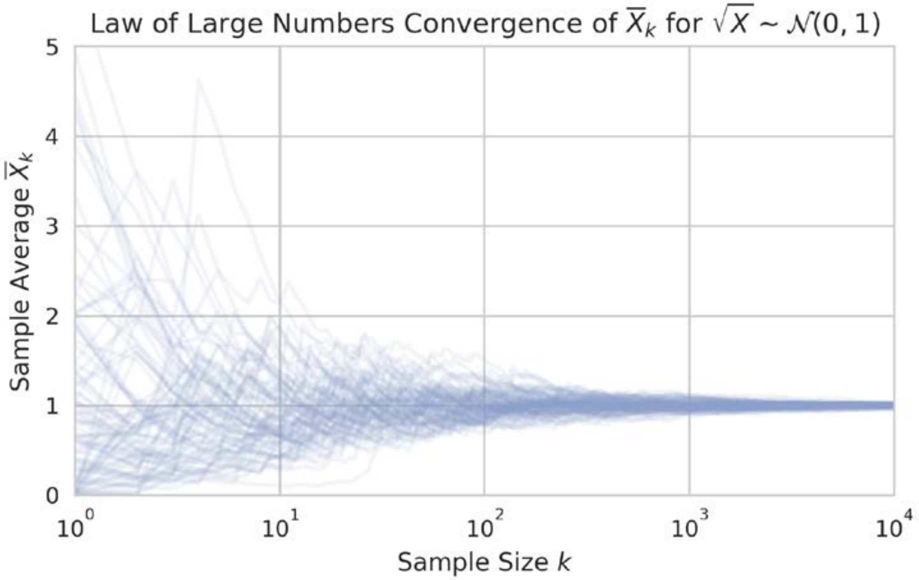
- **Bernstein's inequality** gives a **sharper exponentially decaying bound** on the tail probability when the support of X_i is bounded.
- The extensions of Bernstein's inequality to unbounded supports or time series dependent random samples can be found in White and Wooldridge (1990).

Convergence in Probability

Theorem 4 (7.4). [Weak Law of Large Numbers (WLLN)]

Let $\mathbf{X}^n = (X_1, \cdots, X_n)$ be an IID random sample with $E(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Then for any given constant $\epsilon > 0$ and as $n \to \infty$,

$$P[|\bar{X}_n - \mu| \leq \epsilon] \to 1 \quad or \quad \bar{X}_n - \mu \xrightarrow{p} 0.$$



Proof

Convergence in Probability

Proof:

- First, noting $E(\bar{X}_n) = \mu$ and $\text{var}(\bar{X}_n) = \sigma^2/n$, we have by Chebychev's inequality

$$\begin{aligned} P[|\bar{X}_n - \mu| &> \epsilon] \\ &\leq \frac{E(\bar{X}_n - \mu)^2}{\epsilon^2} \\ &= \frac{\text{var}(\bar{X}_n)}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2}. \end{aligned}$$

It follows that

$$\begin{aligned} P[|\bar{X}_n - \mu| &\leq \epsilon] \\ &= 1 - P[|\bar{X}_n - \mu| > \epsilon] \\ &\geq 1 - \frac{\sigma^2}{n\epsilon^2} \\ &\to 1 \end{aligned}$$

as $n \to \infty$. Therefore, $\bar{X}_n \xrightarrow{p} \mu$.

Convergence in Probability

Remarks:

- WLLN says that the sample mean $\bar{X}_n$ will approach the population mean μ with probability approaching one as the sample size $n \to \infty$.
- Given any constant $\epsilon > 0$, it is possible that the difference $|\bar{X}_n - \mu|$ is larger than any pre-specified constant ϵ for any finite n, but this becomes less and less likely as the sample size n increases.
- When n is sufficiently large, we can be practically certain that the error made with the sample mean $\bar{X}_n$ will be less than any pre-assigned positive constant ϵ.

Convergence in Probability

Remarks:

- In the WLLN theorem, we have assumed a finite variance. Although such an assumption is true and desirable in most applications, it is, in fact, a stronger assumption than is needed. The only moment condition needed is that $E|X_i| < \infty$ (see Resnik 1999, Chapter 7, or Billingsley 1995, Section 22).
- To appreciate why a moment condition is needed for the WLLN, one can consider the example of an IID random sample from the Cauchy distribution whose moments do not exist.

Convergence in Probability

Example 13 (7.13)

Suppose $\mathbf{X}^n$ is an IID random sample from a Cauchy distribution. Then for all integers $n > 0$,

$$\bar{X}_n \sim \text{ Cauchy}(0, 1).$$

Thus, the sample $\bar{X}_n$ does not converge to a constant even

when $n \to \infty$.

Convergence in Probability

Example 13 (7.13). [WLLN and Buy & Hold Trading Strategy]

In finance, there is a popular investment strategy called **buy-and-hold** trading strategy. An investor buys an asset at some day and then hold it for a long time period before he sells it out. This is called a buy-and-hold trading strategy. What is the average daily return of this trading strategy?

Suppose X_i is the return of the asset on period i, and the returns over different time periods are IID(μ, σ^2). Also assume the investor holds the stock for a total of n periods from $i = 1$ to $i = n$. Then the average return in each time period is the sample mean

Convergence in Probability

Example 13 (7.13). [WLLN and Buy & Hold Trading Strategy]

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

When the number n of holding periods is large, we have

$$\bar{X}_n \xrightarrow{p} \mu = E(X_i)$$

as $n \to \infty$. That is, the average return of the buy-and-hold trading strategy is approximately equal to the population μ when n is sufficiently large. In other words, the population mean μ can be viewed as the long-run average return for the buy-and-hold trading strategy.

Convergence in Probability

Question: What is the relationship between L_p-convergence and convergence in probability?

Convergence in Probability

Lemma 5 (7.5)

Suppose $Z_n \to Z$ in L_p as $n \to \infty$. Then $Z_n \xrightarrow{p} Z$ as $n \to \infty$.

Proof

Convergence in Probability

Proof:

By Markov's inequality, we have for all $\epsilon > 0$,

$$\begin{aligned} P[|Z_n - Z| &> \epsilon] \\ &\leq \frac{E|Z_n - Z|^p}{\epsilon^p} \\ &\to 0 \end{aligned}$$

if $\lim_{n\to\infty} E|Z_n - Z|^p = 0$.

Convergence in Probability

Example 14 (7.14)

Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID $N(\mu, \sigma^2)$ random sample. Show

$$S_n^2 \xrightarrow{p} \sigma^2.$$

Solution

Convergence in Probability

Solution

- We have shown that under a normal random sample $\mathbf{X}^n$,

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

for all $n > 1$.

- It follows that $E(S_n^2) = \sigma^2$ and $\mathrm{var}(S_n^2) = 2\sigma^4/(n-1)$. Hence, we have

$$\begin{aligned} E(S_n^2 - \sigma^2)^2 &= \mathrm{var}(S_n^2) \\ &= \frac{2\sigma^4}{n-1} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

- It follows from Lemma 7.5 that $S_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$.

Convergence in Probability

Question: Does convergence in probability imply L_p-convergence?

- Convergence in probability does not necessarily imply L_p-convergence.

Convergence in Probability

Example 15 (7.15)

Suppose a sequence of binary random variables $\{Z_n\}$ is defined as

Z_n	$\frac{1}{n}$	n
$f_{Z_n}(z_n)$	$1-\frac{1}{n}$	$\frac{1}{n}$

(1) Does Z_n converge in quadratic mean to 0? Give your reasoning clearly.
(2) Does Z_n converge in probability to 0? Give your reasoning clearly.

Solution

Convergence in Probability

Solution

(1) Z_n does not converge in quadratic mean to 0 because

$$\begin{aligned} E(Z_n-0)^2 &= \sum (z_n-0)^2 f_{Z_n}(z_n) \\ &= n^{-2}(1-n^{-1})+n^2(n^{-1}) \\ &> n \to \infty. \end{aligned}$$

(2) Given any $\epsilon > 0$, for all $n > N(\epsilon) = [\epsilon^{-1}] + 1$ (so $n^{-1} < \epsilon$),

$$\begin{aligned} P(|Z_n - 0| &\leq \epsilon) \\ &= P(Z_n = n^{-1}) \\ &= 1 - n^{-1} \\ &\to 1 \text{ as } n \to \infty. \end{aligned}$$

Therefore, Z_n converges to 0 in probability.

Convergence in Probability

Remarks:

- Z_n converges to 0 with probability $1-n^{-1}$ approaching one, so $Z_n \xrightarrow{p} 0$ as $n \to \infty$.
- However, the square of Z_n grows to infinity at a rate of n^2, which is faster than the rate n^{-1} of the vanishing probability $P(Z_n = n)$. As a result, the second moment does not exist.
- More generally, convergence in probability implies that there may exist a set of outcomes in sample space with "large deviations" and the probability of this set shrinks to zero as $n \to \infty$ but it is nonzero for a finite n. Some large deviations may be even explosive at a rate faster than their shrinking probabilities. As a result, the L_p-convergence may not exist.

Convergence in Probability

Lemma 6 (7.6). [Continuity]

Suppose $g(\cdot)$ is a continuous function, and Z_n converges in probability to Z. Then $g(Z_n)$ also converges in probability to $g(Z)$. That is, if $g(\cdot)$ is continuous, then $Z_n \xrightarrow{p} Z$ as $n \to \infty$ implies

$$g(Z_n) \xrightarrow{p} g(Z) \text{ as } n \to \infty$$

or equivalently

$$p\lim g(Z_n) = g(p\lim Z_n).$$

Proof

Convergence in Probability

Proof:

- By continuity of function $g(\cdot)$: given any $\epsilon > 0$, there exists a constant $\delta = \delta(\epsilon)$, such that whenever $|Z_n - Z| \leq \delta$, we have
$$|g(Z_n) - g(Z)| < \epsilon.$$

- Now, define two events

$$\begin{aligned} A_n(\delta) &\equiv \{s \in S : |Z_n(s) - Z(s)| \leq \delta\}, \\ B_n(\epsilon) &\equiv \{s \in S : |g[Z_n(s)] - g[Z(s)]| \leq \epsilon\}. \end{aligned}$$

Then continuity of $g(\cdot)$ implies $A_n(\delta) \subseteq B_n(\epsilon)$, that is, $A_n(\delta)$ is a subset of $B_n(\epsilon)$.

To be Continued

Convergence in Probability

Proof:

- It follows that $P[A_n(\delta)] \leq P[B_n(\epsilon)]$, and so

$$P[B_n(\epsilon)^c] \leq P[A_n(\delta)^c] \to 0 \text{ as } n \to \infty.$$

where $B_n(\epsilon)^c$ and $A_n(\delta)^c$ are the complements of $B_n(\epsilon)$ and $A_n(\delta)$ respectively. Because ϵ is arbitrary, so is δ. It follows that $g(Z_n) \xrightarrow{p} g(Z)$ as $n \to \infty$.

Convergence in Probability

Remarks:

- The $p\lim$ operator passes through nonlinear functions, provided they are continuous.
- This is analogous to the well-known result in calculus that the limit of a continuous function is equal to the function of the limit.
- The expectation operator $E(\cdot)$, which is used in the L_p-convergence, does not have this feature, and this makes finite sample analysis difficult for many statistics.

Convergence in Probability

Example 16 (7.16)

Let Z_n converge in probability to a constant $c \neq 0$. Then show that the random variable Z_n/c converges in probability to 1.

Solution

The result follows from Lemma 7.6 and the fact that the function $g(z) = z/c$ is a continuous function when $c \neq 0$.

Convergence in Probability

Example 17 (7.17)

Let Z_n converge in probability to a constant $c > 0$ and let $P(Z_n < 0) = 0$ for every n. Show the random variable $\sqrt{Z_n}$ converges in probability to $\sqrt{c}$.

Solution

The result follows immediately from Lemma 7.6 and the fact that the square root function $g(z) = \sqrt{z}$ is continuous.

Convergence in Probability

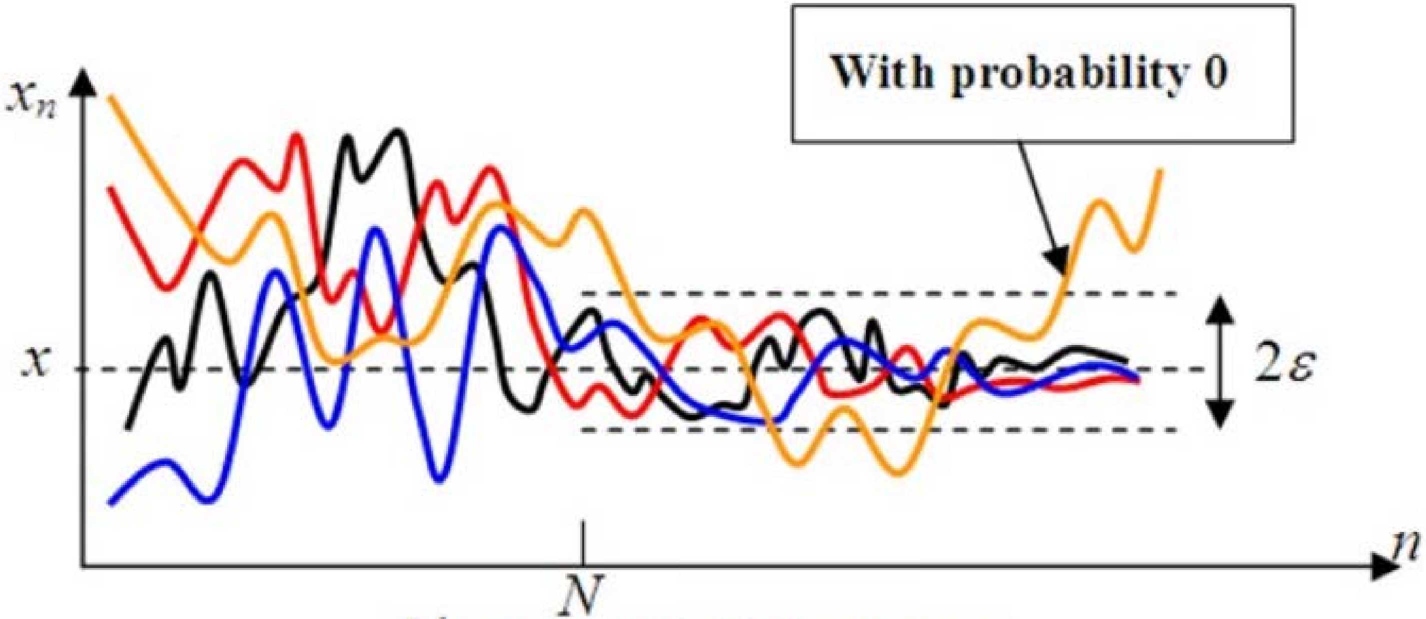
Remarks:

- One application is $Z_n = S_n^2$. Then $\sqrt{Z_n} = S_n$ is the sample standard deviation.
- It follows that $S_n \xrightarrow{p} \sigma$ as $n \to \infty$ given $S_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$.

CONTENTS

Almost Sure Convergence

- Almost sure convergence is the standard notion of **pointwise convergence** that one studies in first year calculus with one exception: convergence is allowed to fail on a subset in S that occurs with **probability zero**.



Almost-sure convergence

Almost Sure Convergence

Definition 8 (7.8). [Almost Sure Convergence]

A sequence of random variables $\{Z_n\}$ converges almost surely to a random variable Z if for every given constant $\epsilon > 0$,

$$P\left[\lim_{n\to\infty} |Z_n - Z| > \epsilon\right] = 0$$

or equivalently,

$$P\left[s \in S : \lim_{n\to\infty} |Z_n(s) - Z(s)| \leq \epsilon\right] = 1,$$

where S is the sample space. When Z_n converges almost surely to Z, we write $Z_n \stackrel{a.s.}{\to} Z$, or $Z_n - Z = o_{a.s.}(1)$, or $Z_n - Z \stackrel{a.s.}{\to} 0$.

Almost Sure Convergence

Remarks:

- Almost sure convergence can also be expressed as

$$P\left(\lim_{n\to\infty} |Z_n - Z| = 0\right) = 1.$$

  For this reason, almost sure convergence is also called **convergence with probability 1.**

- When $Z = b$, a constant, we say that Z_n is **strongly consistent** for b if Z_n converges to b with probability one.

Almost Sure Convergence

- Define a set

$$\begin{aligned} A(\epsilon) &= \left\{s \in S : \lim_{n\to\infty} |Z_n(s) - Z(s)| \le \epsilon\right\} \\ &= \{s \in S : |Z_n(s) - Z(s)| \le \epsilon \text{ for all } n > N(\epsilon, s)\}, \end{aligned}$$

i.e. $A(\epsilon)$ is a subset of S that consists of all basic outcomes $s \in S$ such that $\lim_{n\to\infty} |Z_n(s) - Z(s)| < \epsilon$. Intuitively $A(\epsilon)$ is a convergence set in S in the sense that for each $s \in A(\epsilon)$, the sample path $\{Z_n(s)\}$ converges to $Z(s)$ as $n \to \infty$. Then

$$\begin{aligned} P\left(\lim_{n\to\infty} |Z_n - Z| \le \epsilon\right) &= P[A(\epsilon)] \\ &= 1 \end{aligned}$$

when almost sure convergence holds. Almost sure convergence requires that the convergence set have probability one to occur.

Almost Sure Convergence

- There may exist a set, $A(\epsilon)^c$, for which the sample path $Z_n(s)$ does not converge to $Z(s)$ as $n \to \infty$, but such a set has probability zero.
- When the sample space S consists of only a finite number of basic outcomes with positive probabilities, almost sure convergence implies pointwise convergence.
- When S consists of continuous basic outcomes, then the convergence set $A(\epsilon) = S - \Lambda$ almost covers the entire sample space S, where Λ is a subset in S with $P(\Lambda) = 0$. This arises when Λ contains a finite or infinite but countable number of points $s \in S$. Because $A(\epsilon)$ contains "almost all points in S" except for a set with zero probability, we call such convergence "almost sure convergence".

Almost Sure Convergence

Example 18 (7.18)

Suppose $S = [0,1]$ is a sample space with basic outcome s following a uniform distribution on $[0,1]$. Define two random variables

$$Z_n(s) = s + s^n \text{ and } Z(s) = s.$$

Show that $Z_n - Z \xrightarrow{a.s.} 0.$

Almost Sure Convergence

Remarks:

- Almost sure convergence can also be expressed as

$$P\left(\lim_{n\to\infty} |Z_n - Z| = 0\right) = 1.$$

  For this reason, almost sure convergence is also called **convergence with probability 1.**

- When $Z = b$, a constant, we say that Z_n is **strongly consistent** for b if Z_n converges to b with probability one.

Almost Sure Convergence

Solution

- For every $s \in [0,1)$, $s^n \to 0$ as $n \to \infty$. It follows that for all $s \in A(\epsilon) = [0,1)$, we have

$$Z_n(s) = s + s^n \to s = Z(s) \text{ as } n \to \infty.$$

That is, for any $\epsilon > 0$ and any s in [0,1), there exists a $N(\epsilon, s) = [\ln \epsilon / \ln s] + 1$ such that for all $n > N(\epsilon, s)$

$$|Z_n(s) - Z(s)| = s^n < \epsilon.$$

Note that the dependence of $N(\epsilon, s)$ on basic outcome s indicates that the convergence rate of $Z_n(s)$ may depend on s.

To be Continued

Almost Sure Convergence

Solution

- On the other hand, when $s = 1$, we have

$$\begin{aligned} Z_n(s) &= s + s^n = 1 + 1^n = 2, \\ Z(s) &= s = 1, \\ |Z_n(s) - Z(s)| &= 1 > \epsilon = \frac{1}{2} \text{ (say)}. \end{aligned}$$

- There exists $\Lambda = \{1\}$ such that $Z_n(1) = 2$ for all n so that $Z_n(1) - Z(1) \nrightarrow 0$ as $n \to \infty$. However, $P(\Lambda) = 0$ and $P[A(\epsilon)] = 1$ because s follows a continuous distribution. Thus, $Z_n - Z = o_{a.s.}(1)$.

Almost Sure Convergence

- **Almost sure convergence with order** n^α, where α can be positive or negative:
  - The sequence of random variables $\{Z_n\}$ is said to be of order smaller than n^α with probability one if

    $$Z_n/n^\alpha \stackrel{a.s.}{\to} 0 \text{ as } n \to \infty.$$

    This is denoted as $Z_n = o_{a.s.}(n^\alpha)$.
  - The sequence of random variables $\{Z_n\}$ is said to be at most of order n^α with probability one if there exists some constant $M < \infty$ such that

    $$P\left(\lim_{n\to\infty} |Z_n/n^\alpha| > M\right) = 0.$$

    This is denoted as $Z_n = O_{a.s.}(n^\alpha)$.

Almost Sure Convergence

Remarks:

- $Z_n = O_{a.s.}(1)$ implies that with probability one, Z_n is bounded by some large constant for all n sufficiently large.

Question: What is the relationship between convergence in probability and almost sure convergence.

Almost Sure Convergence

- **[Notational Difference]** We first compare the difference in notations between almost sure convergence

$$P\left[\lim_{n\to\infty} |Z_n - Z| > \epsilon\right] = 0$$

and convergence in probability

$$\lim_{n\to\infty} P[|Z_n - Z| > \epsilon] = 0.$$

- The fact that the set $A_n(\epsilon)$ of "large" differences for $|Z_n - Z|$ may have a nonzero probability for any finite n implies that convergence in probability is weaker than almost sure convergence.
- If for some $s \in S$, $Z_n(s) \to Z(s)$ as $n \to \infty$, then the difference $|Z_n(s) - Z(s)|$ will eventually become "small" (i.e., smaller than ϵ) for all n sufficiently large. Hence, almost sure convergence implies convergence in probability.

Almost Sure Convergence

Lemma 7 (7.7)

Suppose $Z_n \overset{a.s.}{\to} Z$. Then $Z_n \overset{p}{\to} Z$.

Almost Sure Convergence

Example 19 (7.19)

Let the sample space S be the closed interval [0,1] with the occurring of basic outcomes following a uniform probability distribution on [0,1]. Define a random variable $Z(s) = s$ for all $s \in [0, 1]$. Also, for $n = 1, 2, \cdots$ define a sequence of random variables

$$Z_n(s) = \begin{cases} s + s^n & \text{if } 0 \leq s \leq 1 - \frac{1}{n}, \\ s + 1 & \text{if } 1 - \frac{1}{n} < s \leq 1. \end{cases}$$

(1) Does Z_n converge almost surely to Z? Give your reasoning;
(2) Does Z_n converge in probability to Z? Give you reasoning;
(3) Does Z_n converges in L_p to Z? Give your reasoning.

Almost Sure Convergence

Solution

- (1) Consider the set

$$A^c(\epsilon) = \{s \in S : \lim_{n\to\infty} |Z_n(s) - Z(s)| > \epsilon\}.$$

For any given $s \in [0, 1)$, it will eventually falls into the region $[0, 1 - n^{-1}]$ if n is sufficiently large, i.e, if $n > N(s) = [1/(1 - s)] + 1$. Thus, we have

$$\lim_{n\to\infty} |Z_n(s) - Z(s)| = \lim_{n\to\infty} s^n = 0$$

for any $s \in [0, 1)$. Hence, the complement set $A^c(\epsilon)$ at most contains one basic outcome, namely $s = 1$. Since the occurring of basic outcome s follows a continuous distribution, we have $P[A^c(\epsilon)] = 0$. It follows that Z_n converges to Z almost surely.

To be Continued

Almost Sure Convergence

Solution

- (2) Because almost sure convergence implies convergence in probability, we have $Z_n \xrightarrow{p} Z$.
- Alternatively, we can consider the set

$$A_n^c(\epsilon) = \{s \in S : |Z_n(s) - Z(s)| > \epsilon\}.$$

Without loss of generality, we assume $0 < \epsilon < 1$. Then for each n, the interval $[1 - \frac{1}{n}, 1]$ is contained in $A_n^c(\epsilon)$. Furthermore, if $|Z_n(s) - Z(s)| = s^n > \epsilon$, then $s \geq e^{n^{-1}\ln\varepsilon}$. It follows that

$$A_n^c(\epsilon) = [\min(1 - n^{-1}, e^{n^{-1}\ln\epsilon}), 1]$$

and

To be Continued

$$P[A_n^c(\epsilon)] = 1 - \min(1 - n^{-1}, e^{n^{-1}\ln\epsilon}) \to 0 \text{ as } n \to \infty.$$

Almost Sure Convergence

Solution

- (3) For $p > 0$,

$$\begin{aligned} E|Z_n - Z|^p &= \int_0^1 |Z_n(s) - Z(s)|^p ds \\ &= \int_0^{1-n^{-1}} s^{pn} ds + \int_{1-n^{-1}}^1 1 \cdot ds \\ &= \frac{1}{pn+1}\left(1 - \frac{1}{n}\right)^{pn+1} + \frac{1}{n} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence, $Z_n \to^{L_p} Z$ as $n \to \infty$.

Almost Sure Convergence

Remarks:

- While almost sure convergence implies convergence in probability, the converse may not be true.

Almost Sure Convergence

Example 20 (7.20)

Suppose $\{S, \mathcal{B}, P\}$ is a probability space, where the sample space $S = [0, 1]$, $\mathcal{B}$ is a σ-field, and P is a uniform probability measure on S. Define a sequence of random varianles as follow:

$$
\begin{aligned}
Z_1(s) &= 1 \text{ for } 0 \leq s \leq 1, \\
Z_2(s) &= \begin{cases} 1, & 0 \leq s \leq \frac{1}{2}, \\ 0, & \frac{1}{2} < s \leq 1, \end{cases} \\
Z_3(s) &= \begin{cases} 0, & 0 \leq s \leq \frac{1}{2}, \\ 1, & \frac{1}{2} < s \leq 1, \end{cases}
\end{aligned}
$$

Almost Sure Convergence

Example 20 (7.20)

$$Z_4(s) = \begin{cases} 1, & 0 \le s \le \frac{1}{3}, \\ 0, & \frac{1}{3} < s \le 1, \end{cases}$$

$$Z_5(s) = \begin{cases} 0, & 0 \le s \le \frac{1}{3}, \\ 1, & \frac{1}{3} < s \le \frac{2}{3}, \\ 0, & \frac{2}{3} < s \le 1, \end{cases}$$

$$Z_6(s) = \begin{cases} 0, & 0 \le s \le \frac{2}{3}, \\ 1, & \frac{2}{3} < s \le 1. \end{cases}$$

Show that $\{Z_n, n = 1, 2, \cdots\}$ converges in probability to 0 but does not converge to 0 almost surely.

Almost Sure Convergence

Example 21 (7.21)

Suppose $\{S, \mathcal{B}, P\}$ is a probability space, where the sample space $S = [0, 1]$, $\mathcal{B}$ is a σ-field, and P is a uniform probability measure on S. Define $Z(s) = s$ and

$$Z_n(s) = \begin{cases} 1, & \text{if } s \in \left[\frac{i}{2^k}, \frac{i+1}{2^k}\right] \text{ for } i = n - 2^k, 1 \le i \le 2^k, \\ s, & \text{otherwise.} \end{cases}$$

where $k = [\log_2 n]$ is the integer part of $\log_2 n$, and $i = 1, \cdots, 2^k$. Then

(1) For every $\epsilon > 0$, $P(|Z_n - Z| > \epsilon) \le 1/2^k \to 0$ as $n \to \infty$, so $Z_n \xrightarrow{p} Z$ as $n \to \infty$.

(2) $E|Z_n - Z|^p = 1/2^k \to 0$ as $n \to \infty$, so $Z_n \to Z$ in L_p.

(3) For any $s \in [0, 1]$, $\lim_{n\to\infty} Z_n(s)$ does not exist, so Z_n does not converge to Z almost surely.

Almost Sure Convergence

What is the relationship between almost sure convergence and L_p-convergence?

- Almost sure convergence does not imply L_p-convergence and L_p-convergence does not imply almost sure convergence.

Almost Sure Convergence

Example 22 (7.22)

Let the sample space S be the closed interval [0,1] with the occurring of basic outcome s following a uniform probability distribution on [0,1]. For $n = 1, 2, \cdots$, define a sequence of random variables

$$Z_n(s) = \begin{cases} 0, & \text{if } s \in [0, 1 - n^{-2}], \\ e^n, & \text{if } s \in (1 - n^{-2}, 1]. \end{cases}$$

Answer the following questions and provide reasoning:

(1) $Z_n \stackrel{q.m.}{\to} 0$?

(2) $Z_n \stackrel{p}{\to} 0$?

(3) $Z_n \stackrel{a.s.}{\to} 0$?

Solution

(1) no; (2) yes; (3) yes.

Almost Sure Convergence

Lemma 8 (7.8)

Suppose $g(Z)$ is a continuous function, and Z_n converges almost surely to Z. Then $g(Z_n)$ also converges almost surely to $g(Z)$.

Proof

Almost Sure Convergence

Proof:

- The proof is similar to the proof of Lemma 7.6 for convergence in probability of a continuous function.

- Let $s \in S$ be a basic outcome. Since $Z_n(s) \to Z(s)$ as $n \to \infty$ implies $g[Z_n(s)] \to g[Z(s)]$ as $n \to \infty$ by the continuity of $g(\cdot)$, we have

$$\{s \in S : Z_n(s) \to Z(s)\} \subseteq \{s \in S : g[Z_n(s)] \to g[Z(s)]\}.$$

Hence,

$$P\left[s \in S : g[Z_n(s)] \to g[Z(s)]\right] \geq P\left[s \in S : Z_n(s) \to Z(s)\right] \to 1.$$

It follows that $g(Z_n) \stackrel{a.s.}{\to} g(Z)$.

Almost Sure Convergence

Theorem 9 (7.9). [Kolmogorov‘s Strong Law of Large Numbers (SLLN)]

Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ be an IID random sample with $E(X_i) = \mu$ and $E|X_i| < \infty$. Define $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then

$$\bar{X}_n \overset{a.s.}{\to} \mu \text{ as } n \to \infty.$$

Proof

See (e.g.) Gallant (1997, pp.132-135).

Almost Sure Convergence

Theorem 10 (7.10). [Uniform Strong Law of Large Numbers (USLLN)]

Suppose
(1) $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an **IID** random sample;
(2) function $g(x, \theta)$ is **continuous** over $\Omega \times \Theta$ where Ω is the support of X_i and Θ is a **compact** set in $\mathbb{R}^d$ with d finite and fixed; and
(3) $E\left[\sup_{\theta\in\Theta} |g(X_i, \theta)|\right] < \infty$, where the expectation $E(\cdot)$ is taken over the population distribution of X_i.
Then as $n \to \infty$,

$$\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} g(X_i, \theta) - E[g(X_i, \theta)] \right| \to 0 \text{ almost surely.}$$

Moreover, $E[g(X_i, \theta)]$ is a **continuous** function of θ over Θ.

Almost Sure Convergence

Remarks:

- Different from the SLLN, $g(X_i, \theta)$ depends on both random variable X_i and parameter θ.
- USLLN says that, for each n, the worst deviation of the sample average $n^{-1}\sum_{i=1}^{n} g(X_i, \theta)$ from the population mean $E[g(X_i, \theta)]$ that one could find over all possible values in Θ converges to zero almost surely. The uniform convergence is with respect to the parameter space Θ.
- USLLN is rather useful when investigating the asymptotic behavior of nonlinear econometric estimators.

CONTENTS

Convergence in Distribution

Remarks:

- When the random variable X_i is not normally distributed, the sampling distribution of $\bar{X}_n$ is generally unknown or very complicated.
- In this case, one may only like to know what is the limiting distribution of $\bar{X}_n$ when the sample size $n \to \infty$.
- Convergence in distribution is the tool that is used to obtain an asymptotic approximation to the exact distribution of a statistic like $\bar{X}_n$.

Convergence in Distribution

Definition 9 (7.9). [Convergence in Distribution]

Let $\{Z_n\}$ be a sequence of random variables with a sequence of corresponding CDF's $\{F_n(z)\}$, and let Z be a random variable with CDF $F(z)$. Then Z_n converges in distribution to Z as $n \to \infty$ if the CDF $F_n(z)$ converges to $F(z)$ at every continuity point $z \in (-\infty, \infty)$, namely

$$\lim_{n\to\infty} F_n(z) = F(z)$$

at every point z where $F_Z(z)$ is continuous. Here, $F(z)$ is called a **limiting** or **asymptotic distribution** of the sequence of random variables $\{Z_n\}$. Convergence in distribution is denoted as $Z_n \xrightarrow{d} Z$ as $n \to \infty$.

Convergence in Distribution

Remarks:

- When the random variable X_i is not normally distributed, the sampling distribution of $\bar{X}_n$ is generally unknown or very complicated.
- Although we refer to a sequence of random variables, $\{Z_n\}$, converging in distribution to a random variable Z, it is actually the sequence of CDF's $\{F_n(\cdot), n = 1, 2, ...\}$ that converges to the CDF $F(\cdot)$. In other words, convergence in distribution means that their **CDF's converge**, not the random variables $\{Z_n\}$ themselves.
- This is different from the concepts of convergence in L_p, convergence in probability, and almost sure convergence. The latter all characterize the **convergence** or closeness of the **random variables** $\{Z_n\}$ to random variable Z.

Convergence in Distribution

- It should be emphasized that the **limiting distribution** $F(\cdot)$ might **not** be obtained by taking the **limit of** $F_n(\cdot)$.

- For example, suppose $Z_n \sim N(0, \frac{1}{n})$. Then it has the distribution function

$$
\begin{aligned}
F_n(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{1/n}\sqrt{2\pi}} e^{-nu^2/2} du \\
&= \int_{-\infty}^{\sqrt{n}z} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv \\
&= \Phi\left(\sqrt{n}z\right),
\end{aligned}
$$

where $\Phi(\cdot)$ is the $N(0,1)$ CDF. Obviously, we have

Convergence in Distribution

$$\lim_{n\to\infty} F_n(z) = \begin{cases} 0 & \text{if } z < 0, \\ \frac{1}{2} & \text{if } z = 0, \\ 1 & \text{if } z > 0. \end{cases}$$

Now define the function

$$F(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0. \end{cases}$$

Then $F(z)$ is a CDF and $\lim_{n\to\infty} F_n(z) = F(z)$ at every continuity point of $F(z)$. (The function $F(z)$ is not continuous at point $z = 0$.) Therefore, $F(\cdot)$ is a limiting distribution of Z_n.

Convergence in Distribution

- However, $F(\cdot)$ cannot be obtained by taking the limit of $F_n(\cdot)$, because $\lim_{n\to\infty} F_n(0) \neq F(0)$ at zero. Note that in this example $\lim_{n\to\infty} F_n(z)$ is not a CDF because it is not right-continuous.

- The concept of convergence in distribution is rather useful in practice. In general, $F_n(z)$ is usually unknown or very complicated for any finite n, but $F(z)$ is known and is often simple. The convergence in distribution thus permits us to approximate the CDF $F_n(z)$ by using the CDF $F(z)$.

- The distribution $F(z)$ is called the limiting distribution or asymptotic distribution of Z_n. Suppose $F(\cdot)$ has mean μ and variance σ^2. Then they will be called, respectively, the asymptotic mean and asymptotic variance of the distribution $F_n(\cdot)$. Since $F(\cdot)$ is not the limit of $F_n(\cdot)$, the asymptotic mean and variance may not be the limits of the mean and variances of the distribution $F_n(\cdot)$ respectively, even if the latter exist.

Convergence in Distribution

Example 23 (7.23)

Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID $U[0, \theta]$ random sample, where θ is an unknown parameter. Let $Z_n = \max_{1 \le i \le n}(X_i)$ be an estimator of θ. Derive the limiting distribution of $n(\theta - Z_n)$.

Solution

Convergence in Distribution

Solution

- For any given $u \geq 0$, we have

$$\begin{aligned} P\left[n(\theta - Z_n) > u\right] &= P\left(Z_n < \theta - \frac{u}{n}\right) \\ &= P\left(X_1 < \theta - \frac{u}{n}, \cdots, X_n < \theta - \frac{u}{n}\right) \\ &= \prod_{i=1}^{n} P\left(X_i < \theta - \frac{u}{n}\right) \\ &= \left(1 - \frac{u}{n\theta}\right)^n \\ &\rightarrow e^{-u/\theta} \end{aligned}$$

To be Continued

Convergence in Distribution

Solution

as $n \to \infty$, where we have used the formula $(1 - \frac{a}{n})^n \to e^{-a}$ as $n \to \infty$. It follows that for $u \geq 0$,

$$\begin{aligned} F_n(u) &= 1 - P[n(\theta - Z_n) > u] \\ &\to 1 - e^{-u/\theta} \end{aligned}$$

as $n \to \infty$. This implies that $n(\theta - Z_n)$ converges in distribution to the exponential distribution with parameter θ, denoted as $\text{EXP}(\theta)$.

Convergence in Distribution

Question:

What is the relationship between convergence in distribution and convergence in probability.

Convergence in Distribution

Lemma 11 (7.11)

Let Z_n be a random variable with CDF $F_n(\cdot)$, and let Z be a random variable with a continuous CDF $F(\cdot)$. If $Z_n \xrightarrow{d} Z$ as $n \to \infty$, then $Z_n = O_P(1)$.

Proof

Convergence in Distribution

Proof:

- For any given constant $\epsilon > 0$, let $M = M(\epsilon)$ be a (large) constant such that $P(|Z| > M) < \epsilon$. Let $F_n(z)$ be the CDF of Z_n. Given $Z_n \xrightarrow{d} Z$, and $F(z)$ is continuous everywhere, we have $|F_n(z) - F(z)| \leq \epsilon$ for any point $z \in (-\infty, \infty)$ and for all n sufficiently large. This implies that for all n sufficiently large, we have

$$\begin{aligned} P(Z_n &> M) - P(Z > M) \leq \epsilon, \\ P(Z_n &\leq -M) - P(Z \leq -M) \leq \epsilon. \end{aligned}$$

To be Continued

Convergence in Distribution

Proof:

It follows that

$$\begin{aligned} P(Z_n > M) + P(Z_n < -M) &\leq P(Z_n > M) + P(Z_n \leq -M) \\ &< P(Z > M) + P(Z \leq -M) + 2\epsilon \\ &= P(Z > M) + P(Z < -M) + 2\epsilon, \end{aligned}$$

where $P(Z = -M) = 0$ given that Z follows a continuous distribution. Therefore,

$$P(|Z_n| > M) < P(|Z| > M) + 2\epsilon < 3\epsilon \equiv \delta.$$

Because ϵ is arbitrary, so is δ, and therefore $Z_n = O_P(1)$.

Convergence in Distribution

Remarks:

- If the probability distribution of Z_n converges to a well-defined continuous probability distribution as $n \to \infty$, then Z_n is bounded in probability.

Convergence in Distribution

Example 24 (7.24)

Recall from Example 7.23 that for $Z_n = \max_{1\le i\le n}(X_i)$, where $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample from a $U[0,\theta]$ distribution, we have shown that $n(\theta - Z_n) \xrightarrow{d} \mathrm{EXP}(\theta)$ as $n \to \infty$. Therefore, $n(\theta - Z_n) = O_P(1)$ and $Z_n - \theta = O_P(n^{-1})$. This implies that the convergence rate of Z_n to θ in probability is at most n^{-1}, which is rather rapid.

Convergence in Distribution

Remarks:

- The observations of n random variables $\{X_i\}_{i=1}^n$ will be more or less equally spread over the interval $[0, \theta]$. Therefore, the maximal observation of $\{X_i\}_{i=1}^n$ will approach the upper bound of θ at a rate equal to n.

Convergence in Distribution

Lemma 12 (7.12)

If $Z_n \xrightarrow{p} Z$ as $n \to \infty$, then $Z_n \xrightarrow{d} Z$ as $n \to \infty$.

Proof:

This is left as an exercise.

Convergence in Distribution

Remarks:

- When Z_n converges to Z in probability as $n \to \infty$, the random variable Z_n will be arbitrarily close to random variable Z for n sufficiently large. Therefore, the probability law of Z_n will be arbitrarily close to the probability law of Z for n sufficiently large.

Convergence in Distribution

Example 25 (7.25)

Suppose $\{Z_n\}$ is an IID random sample from a population distribution $F(z)$, and Z is a random variable independent of the sequence $\{Z_n\}$ but has the same distribution $F(z)$. Assume $\text{var}(Z) = \sigma^2 < \infty$.

(1) Does Z_n converge in distribution to Z?

(2) Does Z_n converge in quadratic mean to Z?

(3) Does Z_n converge in probability to Z?

Solution

Convergence in Distribution

Solution

- (1) To show $Z_n \xrightarrow{d} Z$, it suffices to show

$$\lim_{n\to\infty} F_n(z) = F(z) \text{ for all continuity points } \{z\}.$$

(By continuity point z, it means the point z at which $F(z)$ is continuous.) Given the identical distribution assumption, we have

$$F_n(z) = F(z) \text{ for all } z \in (-\infty, \infty) \text{ and all } n > 0.$$

Hence, we trivially have

$$\lim_{n\to\infty} F_n(z) = F(z) \text{ for all continuity points } \{z\}.$$

Hence, we have $Z_n \xrightarrow{d} Z$ as $n \to \infty$.

To be Continued

Convergence in Distribution

Solution

- (2) Because Z_n and Z are independent, we have

$$
\begin{aligned}
E(Z_n - Z)^2 &= \text{var}(Z_n - Z) \\
&= \text{var}(Z_n) + \text{var}(Z) \\
&= 2\text{var}(Z) \\
&= 2\sigma^2 > 0.
\end{aligned}
$$

  Thus, Z_n does not converge to Z in quadratic mean, although it converges in Z in distribution.

- (3) No. (Why?)

Convergence in Distribution

Lemma 13 (7.13). [Asymptotic Equivalence]

If $Y_n - Z_n \xrightarrow{p} 0$ and $Z_n \xrightarrow{d} Z$ as $n \to \infty$, then $Y_n \xrightarrow{d} Z$.

Convergence in Distribution

Remarks:

- If two random variables Y_n and Z_n are very close with probability approaching one as $n \to \infty$, they will follow the same large sample distribution.
- This lemma is very useful when one is interested in deriving the asymptotic distribution of Y_n.
- For example, suppose we are interested in deriving the asymptotic distribution of the following scaled sample Sharpe ratio

$$Y_n = \frac{\sqrt{n}\bar{X}_n}{S_n}$$

Convergence in Distribution

Remarks:

under the hypothesis that $\mu = 0$, where $\bar{X}_n$ and S_n are the sample mean and sample standard deviation for an IID random sample $\mathbf{X}^n$. We can first establish the asymptotic equivalence in probability between Y_n and the random variable

$$Z_n = \frac{\sqrt{n}\bar{X}_n}{\sigma}$$

and then derive the asymptotic distribution of Z_n, which is much simpler because we do not have to deal with a random denominator.

Convergence in Distribution

Definition 10 (7.10). [Degenerate Distribution]

A random variable Z is said to have a degenerate distribution if $P(Z = c) = 1$ for some constant c.

Convergence in Distribution

Theorem 14 (7.14)

Let $F_n(z)$ be the CDF of a random variable Z_n whose distribution depends on the positive integer n. Let c denote a constant which does not depend upon n. The sequence $\{Z_n\}$ converges in probability to constant c if and only if the limiting distribution of Z_n is degenerate at $z = c$.

Proof

Convergence in Distribution

Proof:

(1) $[Necessity]$: First, Suppose $\lim_{n\to\infty} P(|Z_n - c| < \epsilon) = 1$ for any given $\epsilon > 0$. Then we shall show

$$\lim_{n\to\infty} F_n(z) = \begin{cases} 0 & \text{if } z < c, \\ 1 & \text{if } z > c, \end{cases}$$

from which we can define an asymptotic (i.e., limiting) distribution

$$F(z) = \begin{cases} 0 & \text{if } z < c, \\ 1 & \text{if } z \geq c. \end{cases}$$

We first observe that

$$P(|Z_n - c| \leq \epsilon) = F_n(c+\epsilon) - F_n(c-\epsilon) + P(Z_n = c - \epsilon).$$

To be Continued

Convergence in Distribution

Proof:

Because $0 \le F_n(z) \le 1$ and $\lim_{n\to\infty} \Pr(|Z_n - c| < \epsilon) = 1$, we must have that for all $\epsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} F_n(c+\epsilon) &= 1, \\ \lim_{n\to\infty} F_n(c-\epsilon) &= 0, \\ \lim_{n\to\infty} P(Z_n = c-\epsilon) &= 0. \end{aligned}$$

It follows that

$$\lim_{n\to\infty} F_n(z) = \begin{cases} 0, & \text{if } z < c, \\ 1, & \text{if } z > c. \end{cases}$$

To be Continued

Convergence in Distribution

Proof:

Thus, we can define an asymptotic (i.e., limiting) distribution as follows:

$$F(z) = \begin{cases} 0, & \text{if } z < c, \\ 1, & \text{if } z \geq c. \end{cases}$$

which is the CDF for Z such that $P(Z = c) = 1$. Since $\lim_{n\to\infty} F_n(z) = F(z)$ at all continuity points $\{z\}$ on the real line (only $z = 0$ is not a continuity point), we have $Z_n \xrightarrow{d} c$ as $n \to \infty$.

To be Continued

Convergence in Distribution

Proof:

(2) [*Sufficiency*]: To complete the proof of the theorem, now suppose

$$\lim_{n\to\infty} F_n(z) = \begin{cases} 0 & \text{if } z < c, \\ 1 & \text{if } z > c. \end{cases}$$

We shall prove that $\lim_{n\to\infty} \Pr(|Z_n - c| \le \epsilon) = 1$ for all $\epsilon > 0$. Because

$$\begin{aligned} 1 &\ge \Pr(|Z_n - c| \le \epsilon) = F_n(c+\epsilon) - F_n(c-\epsilon) + P(Z_n = c - \epsilon) \\ &\to 1 - 0 + 0 = 1 \text{ as } n \to \infty \end{aligned}$$

for all $\epsilon > 0$, we then obtain the desired result immediately.

To be Continued

Convergence in Distribution

Theorem 15 (7.15). [Continuous Mapping Theorem]

Suppose a sequence of $k \times 1$ random vectors $Z_n \xrightarrow{d} Z$ as $n \to \infty$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is a continuous vector-valued function. Then $g(Z_n) \xrightarrow{d} g(Z)$.

Proof

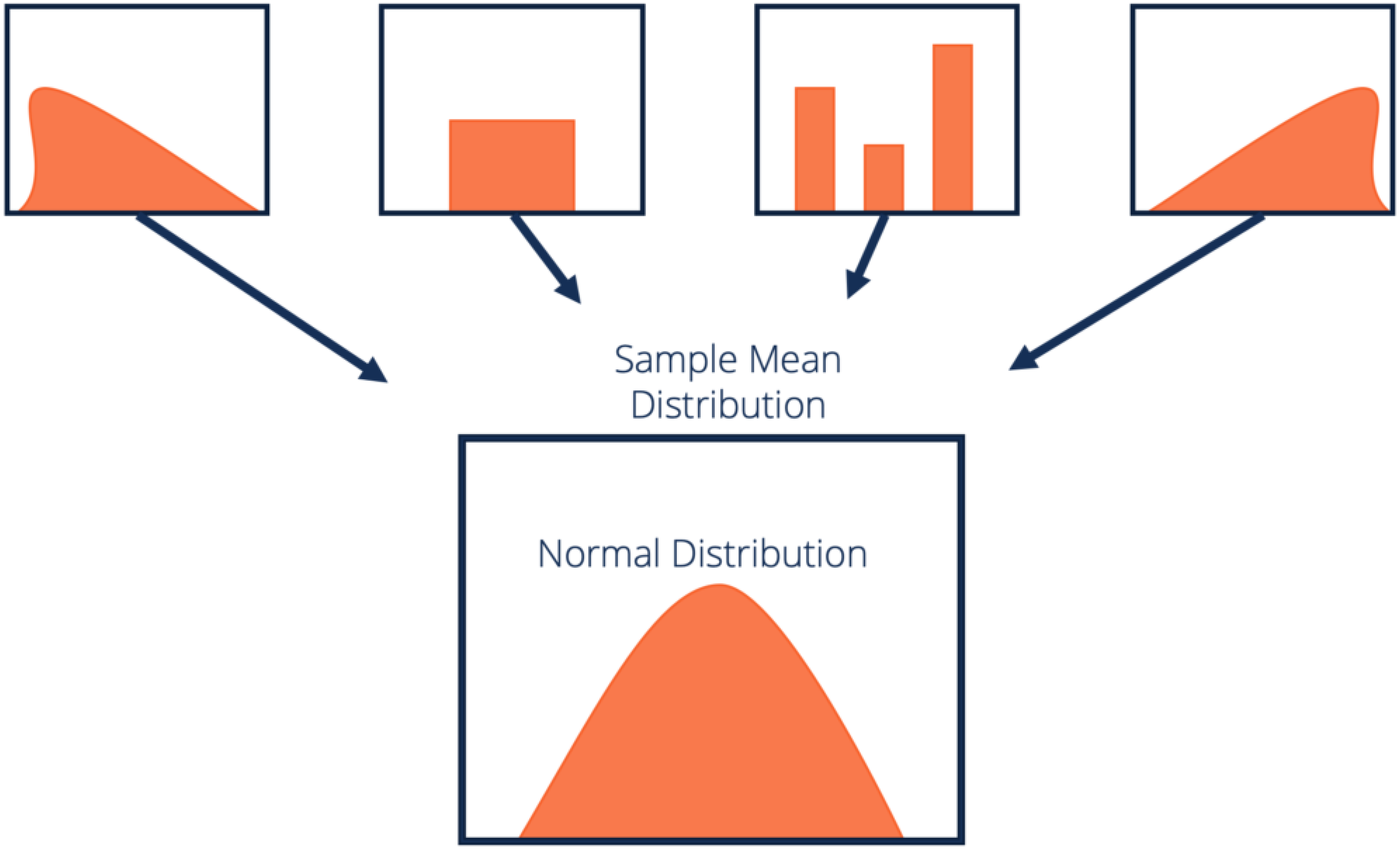
Convergence in Distribution

Remarks:

- Once we know the limiting distribution of Z_n, we can find the limiting distribution of many interesting functions of Z_n. This is particularly useful for deriving the limiting distributions of statistic $T(Z_n)$ once the limiting distribution of Z_n is known.

CONTENTS

Central Limit Theorems



Central Limit Theorems

Theorem 16 (7.16). [Lindeberg-Levy's Central Limit Theorem (CLT)]

Let $\mathbf{X}^n = (X_1, \cdots, X_n)'$ be an IID random sample from a population with mean μ and variance $0 < \sigma^2 < \infty$. Define $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Then the standardized sample mean

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{var}(\bar{X}_n)}} \\ &= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \\ &= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \\ &\xrightarrow{d} N(0,1) \text{ as } n \to \infty. \end{aligned}$$

Central Limit Theorems

Remarks:

- Let $Z \sim N(0,1)$. The $N(0,1)$ CDF is denoted as

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

CLT says that when $n \to \infty$, $Z_n \xrightarrow{d} Z$, i.e.,

$$F_n(z) \equiv P(Z_n \leq z) \to \Phi(z)$$

for all $z \in (-\infty, \infty)$.

Central Limit Theorems

Proof:

- Define a standardized random variable

$$Y_i = \frac{X_i - \mu}{\sigma}, \qquad i = 1, \cdots, n,$$

with characteristic function $\varphi_Y(t) = E(e^{\mathbf{i}tY_i})$, where $\mathbf{i} = \sqrt{-1}$. Then Y_i has zero mean and unit variance.

- It follows that

$$\begin{aligned} \varphi'_Y(0) &= \mathbf{i} \cdot 0 = 0, \\ \varphi''_Y(0) &= \mathbf{i}^2 \cdot \sigma_Y^2 = -1. \end{aligned}$$

To be Continued

Central Limit Theorems

Proof:

- We now write the standardized sample mean

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \\ &= \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \\ &= \sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma}\right) \\ &= \sqrt{n}\bar{Y}_n \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i. \end{aligned}$$

To be Continued

Central Limit Theorems

Proof:

- Since X_i may not have a well-defined MGF, we take a characteristic function approach, that is, we shall show that $\varphi_n(t) \to e^{-\frac{1}{2}t^2}$ as $n \to \infty$, where $\varphi_n(t) = E(e^{\mathbf{i}tZ_n})$ is the characteristic function of Z_n and $e^{-\frac{1}{2}t^2}$ is the characteristic function of $N(0,1)$.

To be Continued

Central Limit Theorems

Proof:

- Given the IID assumption, we have

$$\begin{aligned}
\varphi_n(t) &= E(e^{\mathrm{i}tZ_n}) \\
&= E\left(e^{\mathrm{i}t\sqrt{n}\bar{Y}_n}\right) \\
&= E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}\Sigma_{i=1}^{n}Y_i}\right) \\
&= E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_1}e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_2}\cdots e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_n}\right) \\
&= E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_1}\right)E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_2}\right)\cdots E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_n}\right) \\
&= \left[E\left(e^{\frac{\mathrm{i}t}{\sqrt{n}}Y_1}\right)\right]^n \\
&= \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n,
\end{aligned}$$

where $\varphi_Y(t) = E(e^{\mathrm{i}tY_i})$.

To be Continued

Central Limit Theorems

Proof:

- Now, write

$$\ln\left\{[\varphi_Y(t/\sqrt{n})]^n\right\} = n\ln[\varphi_Y(t/\sqrt{n})] = \frac{\ln[\varphi_Y(t/\sqrt{n})]}{1/n}.$$

Noting $\ln[\varphi_Y(t/\sqrt{n})] \to 0$ given $\varphi_Y(0) = 1$, and $1/n \to 0$, we have for any given $t \in (-\infty, \infty)$,

$$\lim_{n\to\infty} \frac{\ln[\varphi_Y(t/\sqrt{n})]}{1/n} = \lim_{n\to\infty} \frac{\frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})}(-\frac{t}{2n\sqrt{n}})}{-1/n^2} = \frac{t}{2}\lim_{n\to\infty}\frac{\frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})}}{1/\sqrt{n}}$$

by L'Hospital's rule. Since

To be Continued

Central Limit Theorems

Proof:

$$\frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})} \to 0$$

given $\varphi'_Y(0) = 0$ and $1/\sqrt{n} \to 0$, we use L'Hospital's rule again and obtain

$$\begin{aligned}
& \lim_{n\to\infty} \frac{\frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})}}{1/\sqrt{n}} \\
= & \lim_{n\to\infty} \frac{\frac{\varphi''_Y(t/\sqrt{n})\varphi_Y(t/\sqrt{n}) - [\varphi'_Y(t/\sqrt{n})]^2}{[\varphi_Y(t/\sqrt{n})]^2} \cdot \left(-\frac{t}{2n\sqrt{n}}\right)}{-\frac{1}{2n\sqrt{n}}} \\
= & \ t \lim_{n\to\infty} \frac{\varphi''_Y(t/\sqrt{n})\varphi_Y(t/\sqrt{n}) - [\varphi'_Y(t/\sqrt{n})]^2}{[\varphi_Y(t/\sqrt{n})]^2} \\
= & -t,
\end{aligned}$$

where we have used the fact that $\varphi_Y(0) = 1, \varphi'_Y(0) = 0$ and $\varphi''_Y(0) = -1$.

To be Continued

Central Limit Theorems

Proof:

- It follows that

$$\lim_{n\to\infty} \ln \varphi_n(t) = \lim_{n\to\infty} \frac{\ln[\varphi_Y(t/\sqrt{n})]}{1/n} = -\frac{1}{2}t^2.$$

Because the limit of a continuous function (here, the exponential function) is equal to the function of the limit, we then have

$$\lim_{n\to\infty} \varphi_n(t) = e^{-\frac{t^2}{2}}.$$

Therefore, $Z_n \xrightarrow{d} N(0,1)$ as $n \to \infty$.

Central Limit Theorems

Remark: An alternative heuristic proof

- The characteristic function of Z_n

$$
\begin{aligned}
\varphi_n(t) &= E\left(e^{\mathbf{i}t\sqrt{n}\bar{Y}_n}\right) \\
&= [E(e^{\mathbf{i}tY_1/\sqrt{n}})]^n \\
&= [\varphi_Y(t/\sqrt{n})]^n \\
&= \left[\varphi_Y(0) + \varphi_Y'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}\varphi_Y''(0)\left(\frac{t}{\sqrt{n}}\right)^2 + r\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n \\
&\to e^{-t^2/2},
\end{aligned}
$$

where $r(t/\sqrt{n})$ denotes a reminder term, and we have used the formula $(1+\frac{a}{n})^n \to e^a$ as $n \to \infty$.

Central Limit Theorems

Remarks:

- In the proof of CLT, we have used the characteristic function rather than the more familiar moment generating function, which covers the population distributions (e.g., the lognormal and Student's t-distributions) whose moment generating functions do not exist.
- Historically, CLT was first established for a random sample from a Bernoulli distribution by A. de Moivre in the early eighteenth century. The proof for a random sample from an arbitrary distribution was given independently by J.W. Lindeberg and P. Levy in the early 1920s.

Central Limit Theorems

- CLT says that if a large random sample is taken from any population distribution with finite variance, regardless of whether this population distribution is discrete or continuous, then the distribution of the standardized sample mean

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

will approximately follow a $N(0,1)$ distribution when n is large. Intuitively, the sample mean $\bar{X}_n$ may be viewed as the superposition of many "small" independent noises or perturbations, all of the same kind. Although each perturbation follows a distribution, the independent superposition of the perturbations leads to the fact that the final result is approximately normally distributed. Therefore, for each sufficiently large integer n, the distribution of $\bar{X}_n$ will be approximately a $N(\mu, \sigma^2/n)$ or equivalently, the distribution of the sum $\sum_{i=1}^{n} X_i$ will be approximately a $N(n\mu, n\sigma^2)$ distribution.

Central Limit Theorems

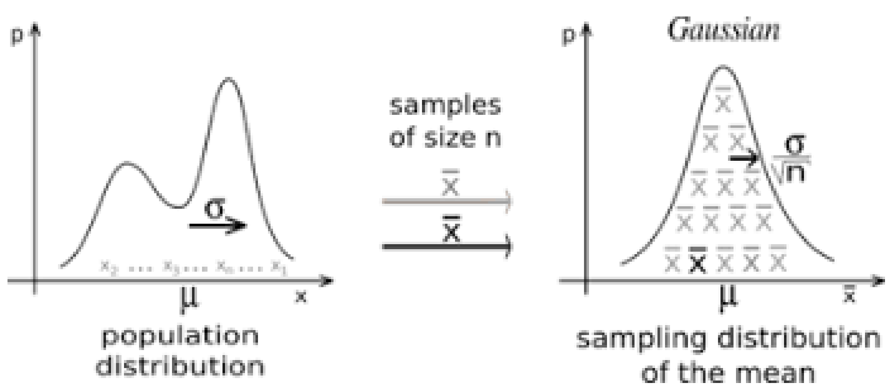
- Sometimes CLT is interpreted incorrectly as implying that the distribution of $\bar{X}_n$ approaches a normal distribution as $n \to \infty$. This is incorrect because $\text{var}(\bar{X}_n) \to 0$ and $\bar{X}_n$ converges to a degenerate distribution $F(\cdot)$ such that $F(x) = 0$ if $x < \mu$ and $F(x) = 1$ if $x \geq \mu$.

- CLT provides a plausible explanation for the fact that the distribution of some random variables in economic systems and physical experiments are approximately normal. In economics, many aggregate economic variables are the sums of individual counterparts. In physics, the observed values for important physical variables be the averages of the measurements in a large number of repeated experiments. In these cases, the variables of interest may be approximately normally distributed if the independence assumption is approximately true.

Central Limit Theorems

- CLT occupies a central position in statistical inference. Although CLT provides a simple and useful general approximation, there is no automatic way of knowing how good the approximation is in general. In fact, the goodness of the approximation is a function of the sample size n and the original population distribution, and differs case by case.

- With rapid advance in computing technology, the importance of CLT is somewhat lessened. One can for example use the bootstrap, a computer-based resampling method, to accurately approximate the finite sample distribution of $\bar{X}_n$ for any finite n.



Central Limit Theorems

Example 26 (7.26). [Life Insurance]

Suppose 10,000 people buy insurance from a particular insurance company. The insurance premium is 12 dollars every person per year. The probability of a man dying in one year is 0.006. If the insurance applicant dead in this year, his families will get 1000 dollar. What is the probability of the insurance company losses?

Solution

Central Limit Theorems

Solution

Denote:

$$X_i = \begin{cases} 1, & \text{if person } i \text{ dies within a year,} \\ 0, & \text{otherwise.} \end{cases}$$

That is, X_i ~Bernoulli(0.006) distribution, where $i = 1, ..., n = 10000$. By CLT,

$$\sqrt{n}\frac{\bar{X}_n - 0.006}{\sqrt{0.006 \times (1 - 0.006)}} \xrightarrow{d} N(0,1)$$

That is,

$$\frac{\sum_{i=1}^{n} X_i - n \times 0.006}{\sqrt{n \times 0.006 \times (1 - 0.006)}} \xrightarrow{d} N(0,1)$$

To be Continued

Central Limit Theorems

Solution

Define the number of the deaths among the insured persons within one year as $Z = \sum_{i=1}^{n} X_i$. If

$$12 \times n - 1000Z < 0 \text{ or } Z > 120,$$

the company will lose money. So, the probability of the insurance company losses:

$$\begin{aligned}
P(Z > 120) &= P\left(\frac{Z - n \times 0.006}{\sqrt{n \times 0.006 \times (1 - 0.006)}} > \frac{120 - 10000 \times 0.006}{\sqrt{10000 \times 0.006 \times (1 - 0.006)}}\right) \\
&\approx 1 - \Phi\left(\frac{120 - n \times 0.006}{\sqrt{n \times 0.006 \times (1 - 0.006)}}\right) \\
&= 1 - \Phi(7.769) \\
&\approx 0.
\end{aligned}$$

Central Limit Theorems

Example 27 (7.27). [Normal Approximation for the Binomial Distribution *B(n,p)* When *n* is Large]

For a binomial $B(n,p)$ random variable Z_n, we can write $Z_n = \sum_{i=1}^{n} X_i$, where $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample from a Bernoulli(p) distribution with $P(X_i = 1) = p \in (0,1)$. By CLT, we have that as $n \to \infty$, the standardized random variable

$$\frac{Z_n - E(Z_n)}{\sqrt{\mathrm{var}(Z_n)}} = \frac{Z_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

Although this result applies only when $n \to \infty$, the normal distribution is often used to approximate binomial probabilities in practice even when n is fairly small. Figure 7.3 plots the case of $p = 0.4$ for which the normal approximation works well for $n \geq 50$.

Central Limit Theorems

Remarks:

- In Chapter 4, we approximate the Binomial(n, p) distribution by a Poisson distribution, which is called the **law of small numbers**. Here we use a normal approximation due to the **law of large numbers**.
- A Poisson approximation is better when p is small, while a normal approximation works better when both np and $n(1-p)$ are both larger than 5.

Central Limit Theorems

Example 28 (7.28). [Normal Approximation of χ_n^2 When n is Large]

Suppose $\mathbf{X}^n$ is an IID $N(0,1)$ random sample. Put $Y = X_i^2$, $i = 1, \cdots, n$. Then as $n \to \infty$, the standardized random variable

$$\frac{\sum_{i=1}^n Y_i - n\mu_Y}{\sqrt{n\sigma_Y^2}} = \frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}} \xrightarrow{d} N(0,1).$$

Solution

Central Limit Theorems

Solution

- Put $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$. Since $E(Y_i) = 1$ and $\text{var}(Y_i) = 2$, we have from CLT that as $n \to \infty$, the standardized sample mean

$$\frac{\bar{Y}_n - \mu_Y}{\sigma_Y/\sqrt{n}} = \frac{\bar{Y}_n - 1}{\sqrt{2}/\sqrt{n}} \xrightarrow{d} N(0,1),$$

where

$$\begin{aligned}\frac{\bar{Y}_n - 1}{\sqrt{2}/\sqrt{n}} &= \frac{\sqrt{n}(\bar{Y}_n - 1)}{\sqrt{2}} \\ &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - 1)}{\sqrt{2}} \\ &= \frac{\sum_{i=1}^n Y_i - n}{\sqrt{2n}} \\ &= \frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}}.\end{aligned}$$

To be Continued

Central Limit Theorems

Solution

- Noting $\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$, we can approximate a χ_n^2 distribution by a $N(n, 2n)$ distribution when the degree n of freedom is large. As an example, when $\mathbf{X}^n$ is an IID $N(\mu, \sigma^2)$ random sample, we have $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$ for all $n > 1$. It follows that as $n \to \infty$,

$$\left[\frac{(n-1)S_n^2}{\sigma^2} - (n-1)\right] / \sqrt{2(n-1)}$$
$$\xrightarrow{d} N(0,1).$$

Central Limit Theorems

Question:

How important is the assumption of $\text{var}(X_i) = \sigma^2 < \infty$ in the CLT?

Central Limit Theorems

Example 29 (7.29). [Sum of Independent Cauchy Random Variables]

Suppose $\mathbf{X}^n$ is an IID random sample from the Cauchy$(0,1)$ distribution. Then it can be shown that $\bar{X}_n \sim$ Cauchy$(0,1)$ for all $n \geq 1$.

Solution

Central Limit Theorems

Solution

- We use the characteristic function to prove this result. By the IID property and the fact that the characteristic function of a Cauchy(0,1) random variables is $\varphi(t) = \exp(-|t|)$, as given in Chapter 4, we have

$$\begin{aligned}\varphi_n(t) &= E\left(e^{\mathbf{i}t\bar{X}_n}\right)\\ &= \left[\varphi\left(\frac{t}{n}\right)\right]^n\\ &= e^{-|t|}\\ &= \varphi(t).\end{aligned}$$

Therefore, $\bar{X}_n \sim$ Cauchy(0,1) for all $n \geq 1$. Thus, the sample mean of an IID Cauchy random sample will not converge to $N(0,1)$ when $n \to \infty$.

Central Limit Theorems

Remarks:

- The assumption of a finite variance is essentially necessary for CLT. It implies that we always obtain an approximate normality from the sum of "small" (finite variance) independent disturbances. Although the finite variance assumption can be relaxed somewhat, it cannot be eliminated.
- On the other hand, the identical distribution assumption can be relaxed. In other words, CLT continues to hold when there exist certain degrees of heterogeneity in observations.

Central Limit Theorems

Theorem 17 (7.17). [Liapounov (1901) CLT for Independent Random Variables]

Suppose the random variables $X_1, \cdots, X_n$ are jointly independent and $E|X_i - \mu_i|^3 < \infty$ for $i = 1, \cdots, n$, where $E(X_i) = \mu_i$. Also, suppose

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E|X_i - \mu_i|^3}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0.$$

Then as $n \to \infty$, we have the standardized random variable

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \xrightarrow{d} N(0,1).$$

Central Limit Theorems

Remarks:

- CLT also holds when there exists certain degree of dependence among the random variables $X_1, \cdots, X_n$.
- Although the dependence allowed cannot be too strong (see what happens if $X_1 = X_2 = \cdots = X_n$ in an extreme case), one can relax the independence assumption to some extent. See, for example, Billingsley (1995, Section 27). Also see White (1999) for CLT for dependent random samples. This allows application of CLT to some time series data.

Central Limit Theorems

Example 31 (7.31)

Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ as $n \to \infty$. Do we have the following results? Give your reasoning:

(1) $X_n \pm Y_n \xrightarrow{d} X \pm Y$ as $n \to \infty$;

(2) $X_n Y_n \xrightarrow{d} XY$ as $n \to \infty$.

Solution

Central Limit Theorems

Solution

- The answer is generally no, because the dependence between X_n and Y_n is not taken into account.
- In other words, convergence in marginal distribution does not imply convergence in joint distribution. This is different from other convergence concepts, such as convergence in quadratic mean, convergence in probability and almost sure convergence. For these convergences, element-by-element convergences are equivalent to joint convergence.

Central Limit Theorems

Lemma 19 (7.19). [Delta Method]

Suppose $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$ as $n \to \infty$, and function $g(\cdot)$ is continuously differentiable with $g'(\mu) \neq 0$. Then as $n \to \infty$,

$$\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right] \xrightarrow{d} N(0, \sigma^2[g'(\mu)]^2)$$

and

$$\frac{\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right]}{\sigma g'(\mu)} \xrightarrow{d} N(0,1).$$

Proof

Central Limit Theorems

Proof:

- First, by Lemma 7.11, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$ implies $\sqrt{n}(\bar{X}_n - \mu)/\sigma = O_P(1)$. Therefore, we have $\bar{X}_n - \mu = O_P(n^{-1/2}) = o_P(1)$.

- Next, by the mean value theorem, we have

$$Y_n = g(\bar{X}_n) = g(\mu) + g'(\bar{\mu}_n)(\bar{X}_n - \mu),$$

where $\bar{\mu}_n = \lambda\mu + (1 - \lambda)\bar{X}_n$ for some $\lambda \in [0, 1]$. Note that $|\bar{\mu}_n - \mu| = |(1 - \lambda)(\bar{X}_n - \mu)| \leq |\bar{X}_n - \mu| = o_P(1)$.

To be Continued

Central Limit Theorems

Proof:

- It follows by the Slutsky theorem that

$$\begin{aligned}\sqrt{n}\left[\frac{g(\bar{X}_n)-g(\mu)}{\sigma}\right] &= g'(\bar{\mu}_n)\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \\ &\xrightarrow{d} N\left[0, g'(\mu)^2\right],\end{aligned}$$

  where $g'(\bar{\mu}_n) \xrightarrow{p} g'(\mu)$ by Lemma 7.6 given $\bar{\mu}_n \xrightarrow{p} \mu$ and the continuity of the first derivative $g'(\cdot)$.

- By the Slutsky theorem again, we have

$$\frac{\sqrt{n}[g(\bar{X}_n)-g(\mu)]}{\sigma g'(\bar{X}_n)} \xrightarrow{d} N(0,1).$$

Central Limit Theorems

Remarks:

- The Delta method can be viewed as a Taylor series approximation in a statistical context. It linearizes a smooth (i.e., continuously differentiable) nonlinear statistic so that CLT can be applied to the linearized statistic. Therefore, it can be viewed as a generalization of CLT.

Central Limit Theorems

Example 32 (7.32)

Suppose $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$ as $n \to \infty$ and $\mu \neq 0$ and $0 < \sigma < \infty$. Find the limiting distribution of $\sqrt{n}(\bar{X}_n^{-1} - \mu^{-1})$.

Solution

Central Limit Theorems

Solution

- We apply the Delta method with $g(z) = z^{-1}$. Because $\mu \neq 0$, $g(z) = z^{-1}$ is continuously differentiable at $z = \mu$, and its derivative

$$g'(\mu) = \frac{1}{\mu^2}.$$

- It follows from the delta method that

$$\frac{\sqrt{n}(\bar{X}_n^{-1} - \mu^{-1})}{\sigma} \xrightarrow{d} N(0, \mu^{-4}) \text{ as } n \to \infty.$$

- By the Slutsky theorem, we have

$$\frac{\bar{X}_n^2\sqrt{n}(\bar{X}_n^{-1} - \mu^{-1})}{\sigma} \xrightarrow{d} N(0, 1) \text{ as } n \to \infty.$$

Central Limit Theorems

Question:

To apply the delta method, it is required that $g'(\mu) \neq 0$. What happens to the delta method if $g'(\mu) = 0$?

Central Limit Theorems

Lemma 20 (7.20). [Second Order Delta Method]

Suppose random variables $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$, and function $g(\cdot)$ is twice continuously differentiable such that $g'(\mu) = 0$ and $g''(\mu) \neq 0$. Then as $n \to \infty$,

$$\frac{n\left[g(\bar{X}_n) - g(\mu)\right]}{\sigma^2} \xrightarrow{d} \frac{g''(\mu)}{2}\chi_1^2.$$

Central Limit Theorems

Question:

We have considered the sequence of scalar random variables $\{\bar{X}_n, n = 1, 2, \cdots\}$. How can one derive the asymptotic distribution for a random vector Z_n?

- The following Cramer-Wold device will allow us to derive the asymptotic distribution of a sequence of random vectors.

Central Limit Theorems

Lemma 21 (7.21). [Cramer-Wold Device]

Let d be a fixed positive integer. A sequence of random vectors $Z_n = (Z_{1n}, \cdots, Z_{dn})'$ converges in distribution to a random vector Z if $\lim_{n\to\infty} F_n(z) = F(z)$ at every point z where $F(z)$ is continuous, where $F_n(z)$ is the CDF of Z_n and $F(z)$ is the CDF of Z. Then a sequence of random vectors Z_n converges in distribution to a random vector Z if and only if $a'Z_n \xrightarrow{d} a'Z$ for every constant vector $a \neq 0$.

Central Limit Theorems

Example 33 (7.33)

Suppose $Z_n \xrightarrow{d} Z \sim N(0, \Sigma)$, where Z is an $m \times 1$ random vector and Σ is an $m \times m$ nonsingular matrix, where the dimension m is fixed. If $\hat{\Sigma}_n \xrightarrow{p} \Sigma$ as $n \to \infty$, then the quadratic form

$$Z_n{}'\hat{\Sigma}_n^{-1} Z_n \xrightarrow{d} Z'\Sigma^{-1} Z \sim \chi^2_m.$$

Proof

Central Limit Theorems

Proof:

- First, by the Cramer-Wold device and the Slutsky theorem, we can show that

$$\hat{\Sigma}^{-\frac{1}{2}} Z_n \xrightarrow{d} \Sigma^{-1/2} Z \sim N(0, I_m) \text{ as } n \to \infty,$$

where I_m is an $m \times m$ identity matrix.

- It follows from the continuous mapping theorem that

$$\begin{aligned} \left(\hat{\Sigma}^{-\frac{1}{2}} Z_n\right)' \left(\hat{\Sigma}^{-\frac{1}{2}} Z_n\right) &= Z_n' \hat{\Sigma}^{-1} Z_n \\ &\xrightarrow{d} Z' \Sigma^{-1} Z \sim \chi^2_m. \end{aligned}$$

CONTENTS

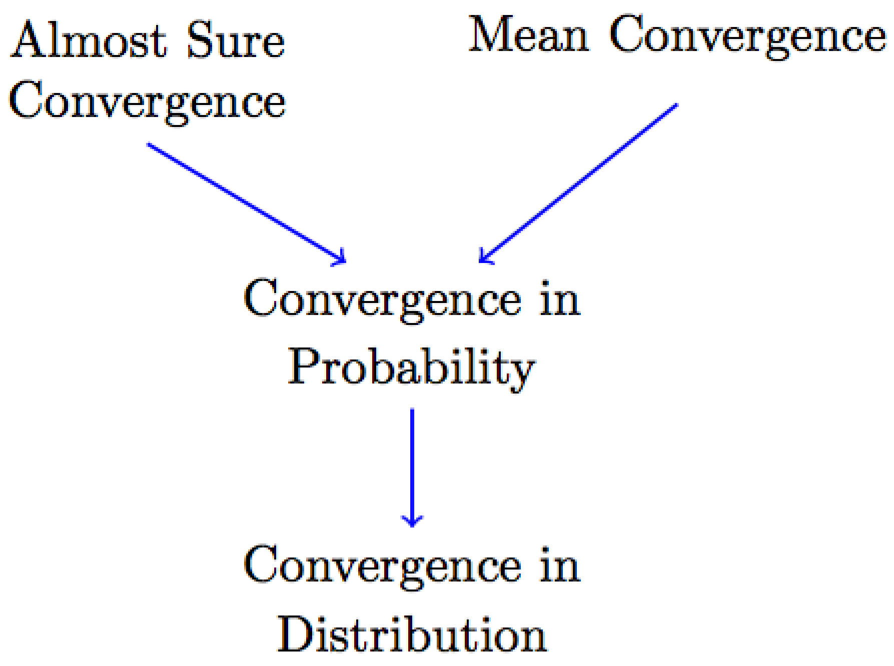
Conclusion

- An important empirical stylized fact in economics is that most economic variables follow a probability distribution with heavy tails.
- The finite sample distributions of most statistics in econometrics are generally unknown or rather complicated, when the random sample are not generated from a normal population.
- One may like to know the limiting behaviors of econometric statistics when the sample size grows to infinity. This is usually called the asymptotic analysis or asymptotic theory.
- In this chapter we have introduced basic concepts and analytic tools for asymptotic theory.

Conclusion

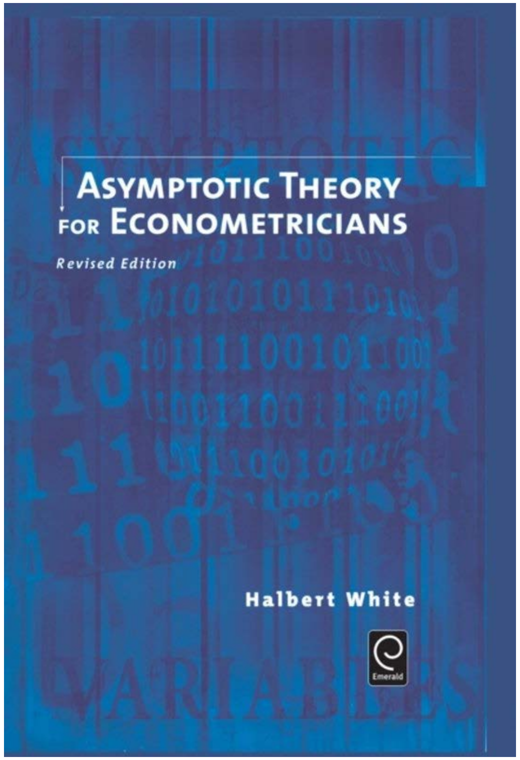
- First, we have introduced four convergence concepts—convergence in quadratic mean, convergence in probability, almost sure convergence, and convergence in distribution. The first three convergences characterize, in a different manner, closeness between a sequence of random variables, and a random variable, while the last convergence concept characterizes the closeness between the CDFs of a sequence of random variables and the CDF of a random variable, rather than the closeness of random variables themselves. Relationships among these convergence concepts are discussed.
- We also introduce and show two limit theorems—the law of large numbers and the central limit theorem.

Conclusion


Almost Sure Convergence
Mean Convergence
Convergence in Probability
Convergence in Distribution

Conclusion

- These asymptotic tools and methods are rather useful in econometrics and statistics. For more discussion on asymptotic theory, see White (1999).



International Symposium on Econometrics of Specification Tests in 30 Years, WISE, Xiamen University,2010

Thank You !