MATH 5010 –Foundations of Statistical Theory and Probability

❖ Convergence Theory

- Limit Theorems

Instructor: He Wang
Department of Mathematics
Northeastern University

❖ Outline:

- **Convergences**
  - Converge in distribution
  - Converge in probability
  - Converge in mean
  - Converge almost surely
- **Weak Law of Large Numbers**
- **Strong Law of Large Numbers**
- **Central Limit Theorem**

Convergence of a Sequence of Numbers

Definition. A sequence of numbers $a_1, a_2, a_3, \cdots$ converges to a limit a if

$$\lim_{n\to\infty} a_n = a$$

More precisely, for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$|a_n - a| < \epsilon \text{ for all } n > N$$

Example:

$$a_n = \frac{n}{n+1}$$

$$\lim_{n\to\infty} a_n = 1$$

Epsilon-Delta definition of limit

Precise meaning of the limit of a function $f(x)$ as x approach to a:

If **for every** $\varepsilon > 0$, **there exists** a $\delta > 0$ such that whenever

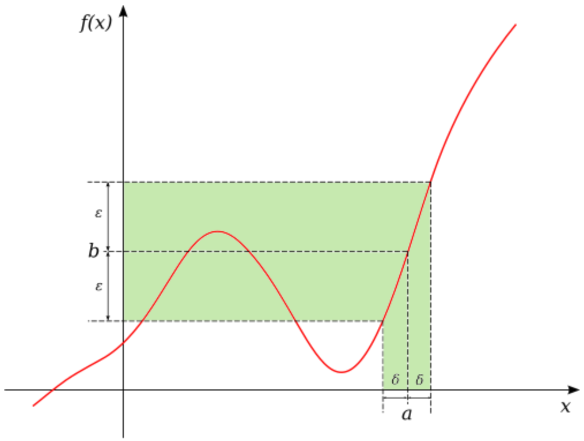
$$0 < | x - a | < \delta,$$

it follows that

$$| f(x) - L | < \varepsilon.$$

then, we say that

$$\lim_{x \to a} f(x) = L$$



Intuitive Meaning

- ε (epsilon) represents how close $f(x)$ should be to the limit L
- δ (delta) represents how close x must be to a a to make that happen.

Sequence of Random Variables

Recall that a random variable $X : S \to \mathbb{R}$ is a *function* from sample space S.

For example, if $S = \{s_1, \dots, s_k\}$ finite, a random variable is a map that

$$X(s_i) = x_i \text{ for } i = 1, \dots, k$$

A sequence of random variables $X_1, X_2, X_3, \cdots$, is in fact a sequence of functions $X_n : S \to \mathbb{R}$

$$X_n(s_i) = x_{ni} \text{ for } i = 1, \dots, k$$

It is also useful to remember that we have an underlying sample space S.

❑ Convergence in Distribution

A sequence of random variables $X_1, X_2, \dots, X_n, \dots$

The CDF of X_n is $F_n(x)$ for $i = 1, \dots, n, \dots$

Definition: The sequence X_n **converges in distribution** (or **converge weakly** or **converge in law**) to X, (Denoted by $X_n \xrightarrow{D} X$) if for every x

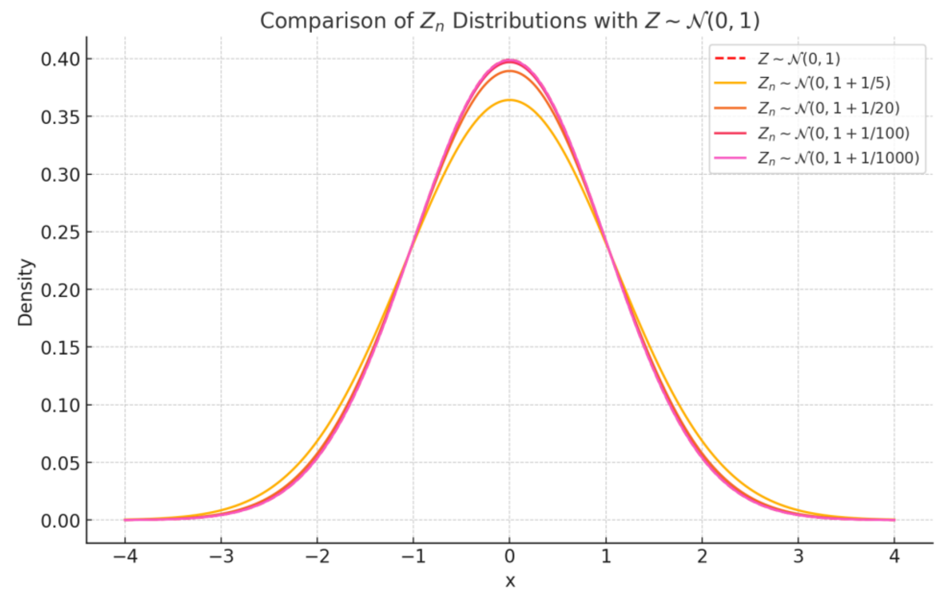
$$\lim_{n\to\infty} F_n(x) = F(x)$$

Converges in distribution: “The overall **shape** of the distribution converges.”

Example:

$$X_n \sim N\left(0, 1+\frac{1}{n}\right)$$

$$X_n \xrightarrow{D} N(0,1)$$



The CDF:

$$\Phi_n(x) = \Phi\left(\frac{x}{\sqrt{1+\frac{1}{n}}}\right) \to \Phi(x)$$

❑ Converges in probability

Definition: The sequence X_n **converges in probability** to X, if for any $\epsilon > 0$

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

(Denoted by $X_n \xrightarrow{P} X$)

Converges in probability: "Large deviations from the limit become rare."

Example: Normal

$$X_n \sim N\left(0, \frac{1}{n}\right) \qquad X_n \xrightarrow{P} 0 \qquad \sqrt{n} X_n \xrightarrow{D} N(0,1)$$

Example: Binary

$$P(X_n = k) = \begin{cases} 1 - \frac{1}{n} & if\ k = 0 \\ \frac{1}{n} & if\ k = 1 \end{cases} \qquad X_n \xrightarrow{P} 0$$

Convergence **in probability** ⇒ convergence **in distribution**

The converse is not true in general.

Example: $X \sim Normal(0,1)$ and $X_n = -X$ for $all\ n$.

$X_n \xrightarrow{D} X$ since they are the same distribution.

$$P(|X_n - X| > 1) = P(2|X| > 1) = P\left(|X| > \frac{1}{2}\right) \approx 0.62$$

Theorem: If $X_n \xrightarrow{D} c$ where c is a constant, then $X_n \xrightarrow{P.} c$

❑ Converges in Mean

Definition. Let $\boldsymbol{r} \geq 1$ be a fixed number. The sequence X_n **converges in the r-th mean** to X, if for any $\epsilon > 0$

$$\lim_{n\to\infty} E(|X_n - X|^r) = 0$$

(Denoted by $X_n \xrightarrow{L^r} X$)

Example: $X_n \sim Uniform\left(0, \frac{1}{n}\right)$ $PDF\ of\ X_n$:

$$E(|X_n - X|^r) = \int_0^{\frac{1}{n}} x^r\ n\, dx = \frac{1}{(r+1)n^r} \to 0\ as\ n \to \infty$$

Convergence in expectation $\Rightarrow$ Convergence in probability

Recall Markov's inequality

Let Y be a non-negative random variable. For any $\epsilon > 0$,

$$P(Y \geq \epsilon) \leq \frac{E[Y]}{\epsilon}$$

Let $Y = |X_n - X|$

$$P(|X_n - X| \geq \epsilon) \leq \frac{E[|X_n - X|]}{\epsilon}$$

If $E[|X_n - X|] \rightarrow 0$, then $P(|X_n - X| \geq \epsilon) \rightarrow 0$

Example (diverging expectation but convergence in probability)

$$P(X_n = k) = \begin{cases} 1 - \dfrac{1}{n} & if\ k = 0 \\ \dfrac{1}{n} & if\ k = n^2 \end{cases}$$

Then, $X_n \xrightarrow{P} 0$

The expectation, $E[X_n] = n$

❑ Converges almost surely

Definition: The sequence X_n **converges almost surely** to X, $\left(\text{Denoted by } X_n \xrightarrow{a.s.} X\right)$, if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Equivalently, *almost every* outcome ω, , the sequence $\lim_{n\to\infty} X_n(\omega) = X(\omega),\ i.e.$,

$$P\left(\left\{\omega\colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Equivalently, for each $\epsilon > 0$,

$$P\left(\lim_{n\to\infty} |X_n - X| < \epsilon\right) = 1$$

"With probability 1, the actual **sample path** converges **pointwise**."

Reason: Almost surely convergence ⇒ convergence **in probability**

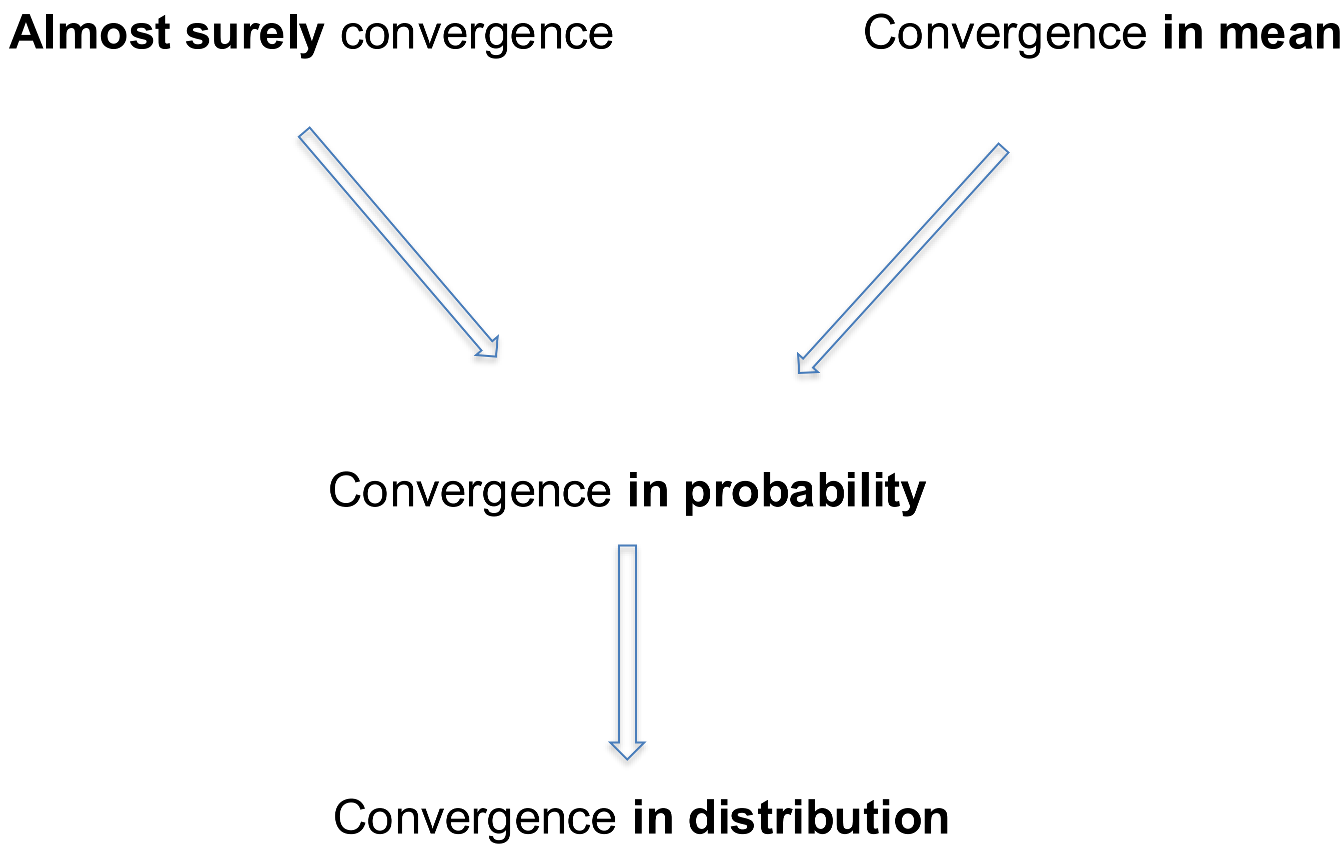
Almost surely convergence means that **almost every** $\boldsymbol{\omega}$, the sequence $X_n(\omega) \to X(\omega)$

$$P(\{\omega | X_n(\omega) \to X(\omega)\}) = 1$$

So, the set ω where $|X_n - X| > \epsilon$ must eventually become small in probability.

Define event $A_{n,\epsilon} \coloneqq \{\omega | |X_n(\omega) - X(\omega)| > \epsilon\}$

$$P\left(A_{n,\epsilon}\right) \to 0$$


Almost surely convergence
Convergence **in mean**
Convergence **in probability**
Convergence **in distribution**

- **Weak Law of Large Numbers (WLLN)**

$X_1, \cdots, X_n$ are IID (Independently, Identically Distributed) from a CDF F.

$X_1, \cdots, X_n$ is called a random sample.

Theorem (**Weak Law of Large Numbers**)

Suppose random sample $X_i \sim F$ for $i = 1, \dots, n$ has mean $\mu = E[X_i] < \infty$ and variance $Var(X_i) = \sigma^2 < \infty$, the sample average

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

converges in probability to μ.

That is $\bar{X}_n \xrightarrow{P} \mu$ $\qquad$ $\lim_{n \to \infty} (|\bar{X}_n - X| > \epsilon) = 0$

Recall Chebyshev's inequality

Let X be a random variable with finite variance. Then for any $\epsilon > 0$,

$$P(|X - E[X]| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

Proof of WLLN:

- **Strong Law of Large Numbers**

Theorem (**Strong Law of Large Numbers**)

Suppose X_i are iid random variables with a finite expected value $\mu = E[X_i] < \infty$ and variance $Var(X_i) = \sigma^2 < \infty$, the sample average

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

converges in almost surely to μ.

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

That is

$$P\left(\lim_{n\to\infty} \bar{X}_n = X\right) = 1$$

➢ Central Limit Theorem

Theorem (Central Limit Theorem)

Suppose random sample X_i for $i = 1, \ldots, n$ has mean $\mu = E[X_i] < \infty$ and variance $Var(X_i) = \sigma^2 < \infty$.

The sample average $\bar{X}_n$ *converges in distribution* to standard normal random variable $N(0,1)$.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$$

Equivalently,

$$\bar{X}_n \xrightarrow{D} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Equivalently,

$$X_1 + \cdots + X_n \xrightarrow{D} N(n\mu, n\sigma^2)$$

Continuous mapping theorem

Let g be a continuous function.

If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$

If $X_n \xrightarrow{D} X$, then $g(X_n) \xrightarrow{D} g(X)$

Slutsky's theorem.

If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$, then

$$X_n + Y_n \xrightarrow{D} X + c$$

$$X_n Y_n \xrightarrow{D} cX$$

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$$

- The convergence in probability is related to the concept of statistical consistency.
- An estimator is statistically consistent if it converges in probability toward its target population quantity.
- The convergence in distribution is often used to construct a maximum likelihood estimator, confidence interval, or perform a hypothesis test.

Concentration inequality

Concentration inequality aims at finding the function $\phi_n(\epsilon) \to 0$ such that

$$P(|X_n - E[X_n]| > \epsilon) \le \phi_n(\epsilon)$$

This automatically gives us convergence in probability.

Moreover, the **convergence rate** of $\phi_n(\epsilon)$ with respect to n is a central quantity that describes how fast X_n converges toward $E[X_n]$.

Example: Concentration of a Gaussian mean.

$$X_1, \cdots, X_n \sim N(0, \sigma^2)$$

By CLT, the sample mean $\bar{X}_n \xrightarrow{D} N\left(0, \frac{\sigma^2}{n}\right)$

Then,

$$P(|\bar{X}_n| > \epsilon) \leq 2e^{-\frac{n\epsilon^2}{2\sigma^2}}$$

Example (concentration of a maximum).

$$X_1, \cdots, X_n \sim N(0, \sigma^2)$$

Let $Z_n := \max\{|X_1|, \ldots, |X_n|\}$

Then, $$P(Z_n > \epsilon) \leq 2ne^{-\frac{\epsilon^2}{2\sigma^2}}$$

Concentration of mean

Recall that Chebyshev's inequality

$$P(|X - E[X]| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

Suppese $X_1, \ldots, X_n$ is an IID sequence with $Var(X_n) = \sigma^2$.
Apply Chebyshev's inequality to sample mean $\bar{X}_n$:

$$P(|\bar{X}_n - E[\bar{X}_n]| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Theorem (Hoeffding's inequality)

Suppose $0 \leq X_n \leq 1$, then for any $\epsilon > 0$,

$$P(|\bar{X}_n - E[\bar{X}_n]| \geq \epsilon) \leq 2e^{-2n\epsilon^2}$$

Applications:

Hoeffding's inequality gives a concentration of the order of exponential (actually it is often called a Gaussian rate) so the convergence rate is much faster than the one given by the Chebyshev's inequality.

Obtaining such an exponential rate is useful for analyzing the property of an estimator.

Many modern statistical topics, such as high-dimensional problem, nonparametric inference, semi-parametric inference, and empirical risk minimization all rely on a convergence rate of this form

Example: Consistency of estimating a high-dimensional proportion.

Consider IID binary observations: $X_1, \dots, X_n \in \{0,1\}^d$

References:

- **Book 1. [CB] Statistical Inference (Chapter 3)**, by Casella, George, Berger, Roger L, 2nd edition
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- **Book 3. Introduction to Probability**. C.M. Grinstead and J.L. Snell. American Mathematical Society, 2012
- **Book 4. Introduction to Probability Models**, S. Ross, 12th edition (published by Academic Press).

- https://www.probabilitycourse.com/ Chapter 7