

Probability Theory

Chapter 8—Limit Theorems

Nan Liu¹

¹These slides are based on Prof. Mui Li's. Any mistakes are my own. If you find an error, please send an email at nanliu@xmu.edu.cn

8.0 Limits and Continuity: A Review

- **Definition: Limit.** Let b_n , $n = 1, 2, \dots$, be a sequence of real numbers. If there exists a real number b such that for every $\varepsilon > 0$ ($\forall \varepsilon$), we can find a finite integer $N(\varepsilon)$ ($\exists N(\varepsilon)$) such that

$$|b_n - b| < \varepsilon$$

for all $n > N(\varepsilon)$, then b is called the *limit* of the sequence $\{b_n\}$. We write $b_n \rightarrow b$ as $n \rightarrow \infty$, or

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

- **Remarks:** If $a_n \rightarrow a$ and $b_n \rightarrow b$ as $n \rightarrow \infty$, then when $n \rightarrow \infty$,
 - $a_n + b_n \rightarrow a + b$.
 - $a_n b_n \rightarrow ab$.
 - $a_n / b_n \rightarrow a / b$ if $b \neq 0$.

8.0 Limits and Continuity: A Review

- **Definition: Continuity.** The function $g : \mathbb{R} \rightarrow \mathbb{R}$ is *continuous* at point b if for any sequence $\{b_n\}$ such that $\lim_{n \rightarrow \infty} b_n = b$, we have $\lim_{n \rightarrow \infty} g(b_n) = g(b)$.
- **Remarks:**
 - An alternative definition of continuity: If $\forall \varepsilon, \exists \delta(\varepsilon) > 0$ such that $|g(x) - g(b)| < \varepsilon$ for all $|x - b| < \delta(\varepsilon)$, then g is continuous at b , denoted by

$$\lim_{x \rightarrow b} g(x) = g(b).$$

8.0 Limits and Continuity: A Review

- **Remarks:**

- If $\forall \varepsilon, \exists \delta(\varepsilon) > 0$ such that $|g(x) - g(b)| < \varepsilon$ for all $b < x < b + \delta(\varepsilon)$, then we say g is *right continuous* at b , denoted by

$$\lim_{x \rightarrow b+} g(x) = g(b).$$

- If $\forall \varepsilon, \exists \delta(\varepsilon) > 0$ such that $|g(x) - g(b)| < \varepsilon$ for all $b - \delta(\varepsilon) < x < b$, then g is *left continuous* at b , denoted by

$$\lim_{x \rightarrow b-} g(x) = g(b).$$

- Function g is continuous at b if and only if g is left continuous and right continuous at b .

8.1 Introduction

- Question: How can we measure the closeness of sample mean $\bar{X}_n$ to μ ?
- Laws of large numbers : the average of a sequence of random variables converges (in some sense) to the expected average.
- Central limit theorems: the sum of a large number of random variables has a probability distribution that is approximately normal.

8.2 Convergence in Probability and Weak Law of Large Numbers

- **Convergence in Probability.** A sequence of random variables $\{Z_n\}$, converges in probability to a random variable Z , denoted by $Z_n \xrightarrow{P} Z$, if for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|Z_n - Z| > \varepsilon) = 0.$$

- **Remarks:**

- When $Z_n \xrightarrow{P} b$, where b is a constant, we say that Z_n is *consistent* for b .
- **Continuous Mapping Theorem:** Suppose $g : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function, and $Z_n \xrightarrow{P} Z$, then $g(Z_n) \xrightarrow{P} g(Z)$.

8.2 Example: Convergence in Probability

- **Example:** Suppose $X_1, X_2, \dots$ are i.i.d. random variable from a $U[0, \theta]$ distribution. Let $Z_n = \max\{X_1, \dots, X_n\}$. Is Z_n consistent for θ ?
- **ANS:**
 - For any $\varepsilon > 0$,

$$\begin{aligned}P(|Z_n - \theta| > \varepsilon) &= P(Z_n < \theta - \varepsilon) \\&= P(X_1 < \theta - \varepsilon, \dots, X_n < \theta - \varepsilon) \\&= \prod_{i=1}^n P(X_i < \theta - \varepsilon) \\&= \left(\frac{\theta - \varepsilon}{\theta}\right)^n \rightarrow 0\end{aligned}$$

as $n \rightarrow \infty$.

- Z_n is consistent for θ .

8.2 Markov's Inequality

- **Markov's Inequality.** Suppose X is a random variable and $g(X)$ is a nonnegative function. Then for any $\varepsilon > 0$, and any $p > 0$, we have

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

- **Proof.** Because

$$\begin{aligned} E[g(X)]^p &= \int_{-\infty}^{\infty} g^p(x) dF(x) \\ &\geq \int_{\{x: g(x) \geq \varepsilon\}} g^p(x) dF(x) \\ &\geq \int_{\{x: g(x) \geq \varepsilon\}} \varepsilon^p dF(x) \\ &= \varepsilon^p P[g(X) \geq \varepsilon], \end{aligned}$$

the conclusion holds.

8.2 Chebyshev's inequality

- **Chebyshev's inequality.** If X is a random variable with finite mean μ and variance σ^2 , then, for any value $\varepsilon > 0$,

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- **Proposition.** If $\text{Var}(X) = 0$, then

$$P(X = E[X]) = 1$$

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

8.2 Example: Markov's and Chebyshev's inequality

- Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50 . (a) What can be said about the probability that this week's production will exceed 75 ? (b) If the variance of a week's production is known to equal 25 , then what can be said about the probability that this week's production will be between 40 and 60 ?
- Let X be the number of items that will be produced in a week.
(a) By Markov's inequality,

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) By Chebyshev's inequality,

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4} \Rightarrow P\{|X - 50| < 10\} \geq \frac{3}{4}$$

so the probability that this week's production will be between 40 and 60 is at least .75.

8.2 Weak Law of Large Numbers

- **Theorem: Weak Law of Large Numbers (WLLN).** Suppose random variables $X_1, X_2, \dots$ are i.i.d. with mean $E(X_i) = \mu$ and finite variance $\text{Var}(X_i) = \sigma^2$. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, then

$$\bar{X}_n \xrightarrow{P} \mu.$$

- **Remark:** In WLLN, the second moment condition $\text{Var}(X_i) < \infty$ (or equivalently, $E(X_i^2) < \infty$) is not necessary.
Khintchine LLN Suppose random variables $X_1, X_2, \dots$ are i.i.d. with finite mean $E(X_i) = \mu < \infty$. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, then

$$\bar{X}_n \xrightarrow{P} \mu.$$

- By CMT,

$$1/\bar{X}_n \xrightarrow{P} 1/\mu.$$

8.3 Convergence in Distribution and the Central Limit Theory

- **Definition: Convergence in Distribution (Weak Convergence).**

A sequence of random variables $Z_1, Z_2, \dots$ *converges in distribution* to a random variable Z , denoted by $Z_n \xrightarrow{d} Z$, if

$$\lim_{n \rightarrow \infty} F_n(z) = F_Z(z)$$

for all z where $F_Z(z)$ is **continuous**. Here $F_n(z)$ and $F_Z(z)$ are the cdf's of Z_n and Z , respectively. $F_Z(z)$ is called the *limiting (or asymptotic) distribution* of $\{Z_n\}$.

8.3 Convergence in Distribution

- **Remarks:**

- The cdf $F_Z(z)$ at most has countable number of discontinuous points.
- If $Z_n \xrightarrow{d} Z$, $F_n(z)$ may not converge to $F_Z(z)$ at the discontinuous points of F_Z .
- Example: Let $Z_n \sim N\left(0, \frac{1}{n}\right)$ and $Z = 0$. Then

$$F_Z(z) = \begin{cases} 0 & \text{for } z < 0, \\ 1 & \text{for } z \geq 0. \end{cases}$$

Because for any $z < 0$, $\lim_n F_n(z) \rightarrow 0$ and for any $z > 0$, $\lim_n F_n(z) \rightarrow 1$, so $Z_n \xrightarrow{d} Z$. However, $F_n(0) = 0.5 \neq F_Z(0)$.

8.3 Example: Convergence in Distribution

- **Example:** Suppose $\{X_i\}$ is a random sample following $U[0, \theta]$ distribution. Let $Z_n = \max\{X_1, \dots, X_n\}$. Derive the limiting distribution of $n(\theta - Z_n)$.

- **ANS:**

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

$$\begin{aligned}P[n(\theta - Z_n) > u] &= P\left(Z_n < \theta - \frac{u}{n}\right) \\&= \left(1 - \frac{u}{n\theta}\right)^n.\end{aligned}$$

- Therefore,

$$F_n(u) = P[n(\theta - Z_n) \leq u] \rightarrow 1 - e^{-u/\theta}.$$

The limiting distribution of $n(\theta - Z_n)$ is an exponential(θ) distribution.

8.3 Central Limit Theory

- **Theorem: Lindeberg-Levys Central Limit Theorem (CLT).** Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with $E(X_1) = \mu$ and finite variance $\sigma^2 = \text{Var}(X_1)$. Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, then

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

- **Remarks:**

- $E(Z_n) = 0$ and $\text{Var}(Z_n) = 1$ for all n .
- X_i can be discrete or continuous.
- The CLT says that

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad \text{for all } z \in \mathbb{R}.$$

8.3 Central Limit Theory

• Proof.

- Let $Y_i = (X_i - \mu)/\sigma$, then $Z_n = \sum_{i=1}^n Y_i / \sqrt{n}$.
- Define $M(t) = E(e^{Y_1 t})$, then the MGF of Z_n is

$$\begin{aligned} E\left(e^{tZ_n}\right) &= \prod_{i=1}^n E\left(e^{tY_i/\sqrt{n}}\right) = [M(t/\sqrt{n})]^n \\ &= \left[1 + M'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}M''(0)\frac{t^2}{n} + o(1/n)\right]^n \end{aligned}$$

- Because $M'(0) = 0$, $M''(0) = 1$, $E\left(e^{tZ_n}\right) \rightarrow e^{t^2/2}$ for all t as $n \rightarrow \infty$.
- Hence $Z_n \xrightarrow{d} N(0, 1)$.

8.3 Central Limit Theory

- **Example: Normal Approximation for the Binomial Distribution.**

For a Binomial(n, p) random variable Y_n , we can write as

$Y_n = \sum_{i=1}^n X_i$, where $\{X_i\}$ are independent Bernoulli random variables with $P(X_i = 1) = p$. By the CLT, $\frac{Y_n - np}{\sqrt{np(1-p)}}$ (or $\frac{\bar{X}_n - p}{\sqrt{p(1-p)/n}}$) approximately follows a $N(0, 1)$ distribution.

- **Examples: Normal approximation of χ_n^2 .** Suppose $Z_1, \dots, Z_n$ is a random sample following $N(0, 1)$ distribution. Let $X_i = Z_i^2$. Then $Y_n = \sum_{i=1}^n X_i^2$ follows a χ_n^2 distribution. We have

$$E(X_1^2) = 1, \quad \text{Var}(X_1^2) = 2.$$

By the CLT, $\frac{Y_n - n}{\sqrt{2n}}$ (or $\frac{\bar{X}_n - 1}{\sqrt{2/n}}$) approximately follows a $N(0, 1)$ distribution when n is large.

8.3 Application of the Central Limit Theory

- If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.
- Solution: Let X_i denote the value of the i th die, $i = 1, 2, \dots, 10$.
Since

$$E(X_i) = \frac{7}{2}, \quad \text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{35}{12},$$

the central limit theorem yields

$$\begin{aligned} P\{29.5 \leq X \leq 40.5\} &= P\left\{ \frac{29.5 - 35}{\sqrt{\frac{350}{12}}} \leq \frac{X - 35}{\sqrt{\frac{350}{12}}} \leq \frac{40.5 - 35}{\sqrt{\frac{350}{12}}} \right\} \\ &\approx 2\Phi(1.0184) - 1 \\ &\approx .692 \end{aligned}$$

8.3 Application of the Central Limit Theory

- Let $X_i, i = 1, \dots, 10$, be independent random variables, each uniformly distributed over $(0, 1)$. Calculate an approximation to $P \left\{ \sum_{i=1}^{10} X_i > 6 \right\}$.
- Solution: Since $E[X_i] = \frac{1}{2}$ and $\text{Var}(X_i) = \frac{1}{12}$, we have, by the central limit theorem,

$$\begin{aligned} P \left\{ \sum_1^{10} X_i > 6 \right\} &= P \left\{ \frac{\sum_1^{10} X_i - 5}{\sqrt{10 \left(\frac{1}{12} \right)}} > \frac{6 - 5}{\sqrt{10 \left(\frac{1}{12} \right)}} \right\} \\ &\approx 1 - \Phi(\sqrt{1.2}) \\ &\approx .1367 \end{aligned}$$

Hence, $\sum_{i=1}^{10} X_i$ will be greater than 6 only 14 percent of the time.

8.3 Central limit theorem for independent random variables

- Let $X_1, X_2, \dots$ be a sequence of independent random variables having respective means and variances $\mu_i = E[X_i]$, $\sigma_i^2 = \text{Var}(X_i)$. If (a) the X_i are uniformly bounded-that is, if for some M , $P\{|X_i| < M\} = 1$ for all i , and (b) $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$ -then

$$P \left\{ \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq a \right\} \rightarrow \Phi(a) \quad \text{as } n \rightarrow \infty$$

8.4 Almost Sure Convergence and the Strong Law of Large Numbers

- **Definition: Almost Sure Convergence.** In probability space $\{S, \mathcal{B}, P\}$, $Z_1, Z_2, \dots$ is a sequence of random variables. If

$$P\left(\left\{s \in S : \lim_{n \rightarrow \infty} Z_n(s) = Z(s)\right\}\right) = 1,$$

then we say that Z_n converges to Z *almost surely* (a.s.), *almost everywhere* (a.e.), or *with probability 1*, denoted by

$$Z_n \xrightarrow{\text{a.s.}} Z.$$

- **Remark:** Almost Sure Convergence is also called *strong convergence*.
- **Theorem:** If $Z_n \xrightarrow{\text{a.s.}} Z$, then $Z_n \xrightarrow{P} Z$.
- **Theorem:** Suppose $g(\cdot)$ is a continuous function, and $Z_n \xrightarrow{\text{a.s.}} Z$, then $g(Z_n) \xrightarrow{\text{a.s.}} g(Z)$.

8.4 Example: Almost Sure Convergence

- **Example:** Suppose the probability space is $\{S, \mathcal{B}, P\}$, where the sample space $S = [0, 1]$, $\mathcal{B}$ contains all the Borel sets B such that $B \subset S$, P is the Lebesgue measure on S , i.e.,

$$P(\{s \in [a, b]\}) \triangleq b - a$$

for any $0 \leq a \leq b \leq 1$. Then $Z_n(s) = s + s^n$, $n = 1, 2, \dots$, and $Z(s) = s$ are random variables.

- For $0 \leq s < 1$, $Z_n(s) \rightarrow Z(s)$ as $n \rightarrow \infty$.
- For $s = 1$, $\lim_{n \rightarrow \infty} Z_n(s) \neq Z(s)$.
- Because $P[Z_n(s) \rightarrow Z(s)] = 1$, so $Z_n \xrightarrow{a.s.} Z$.

8.4 Example: Almost Sure Convergence

- **Example:** Suppose the probability space is $\{S, \mathcal{B}, P\}$, where the sample space $S = [0, 1]$, $\mathcal{B}$ contains all the Borel sets B such that $B \subset S$, P is the Lebesgue measure on S . Let $Z(s) = 0$ and

$$\begin{cases} Z_1(s) = I(0 < s \leq 1), \text{ where } I(\cdot) \text{ is the indicator function;} \\ Z_2(s) = I(0 < s \leq \frac{1}{2}), Z_3(s) = I(\frac{1}{2} < s \leq 1); \\ Z_n(s) = I(\frac{n-4}{4} < s \leq \frac{n-3}{4}), n = 4, \dots, 7; \\ Z_n(s) = I(\frac{n-8}{8} < s \leq \frac{n-7}{8}), n = 8, \dots, 15; \\ \vdots \end{cases}$$

- For every $\varepsilon > 0$, $P(|Z_n - Z| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$, so $Z_n \xrightarrow{P} Z$.
- for any $s > 0$, $\lim_{n \rightarrow \infty} Z_n(s)$ does not exist, Z_n does not converge to Z almost surely.

8.4 Example: Almost Sure Convergence

- **Example:** Suppose the probability space is $\{S, \mathcal{B}, P\}$, where the sample space $S = [0, 1]$, $\mathcal{B}$ contains all the Borel sets B such that $B \subset S$, P is the Lebesgue measure on S . Define $Z_n(s)$ as

$$Z_n(s) = \begin{cases} 0 & \text{for } 1/n < s \leq 1, \\ e^n & \text{for } 0 \leq s \leq 1/n. \end{cases}$$

- For every $\varepsilon > 0$, $P(|Z_n| > \varepsilon) \leq 1/n \rightarrow 0$ as $n \rightarrow \infty$, so Z_n converges to 0 in probability.
- For any $0 < s \leq 1$, $\lim_{n \rightarrow \infty} Z_n(s) = 0$, so Z_n converges to 0 almost surely.

8.4 Strong Law of Large Numbers

- **Theorem: Kolmogorov's Strong Law of Large Numbers.**

Suppose random variables $X_1, X_2, \dots$ are i.i.d. with finite $E|X_i|$. Let $\mu = E(X_i)$ and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, then

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

8.5 Other Inequalities

- **One-sided Chebyshev inequality:**

If $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$, then, for $a > 0$,

$$P\{X \geq \mu + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

8.5 Example

- If the number of items produced in a factory during a week is a random variable with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be at least 120.
- Solution: It follows from the one-sided Chebyshev inequality that

$$P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400 + (20)^2} = \frac{1}{2}$$

Hence, the probability that this week's production will be 120 or more is at most $\frac{1}{2}$. If we attempted to obtain a bound by applying Markov's inequality, then we would have obtained

$$P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}$$

8.5 Other Inequalities

- When the MGF of the random variable X exist, we can obtain even more effective bounds on $P\{X \geq a\}$. Let

$$M(t) = E \left[e^{tX} \right]$$

be the moment generating function of the random variable X . Then, for $t > 0$,

$$\begin{aligned} P\{X \geq a\} &= P\left\{e^{tX} \geq e^{ta}\right\} \\ &\leq E \left[e^{tX} \right] e^{-ta} \quad \text{by Markov's inequality} \end{aligned}$$

Similarly, for $t < 0$

$$\begin{aligned} P\{X \leq a\} &= P\left\{e^{tX} \geq e^{ta}\right\} \\ &\leq E \left[e^{tX} \right] e^{-ta} \end{aligned}$$

8.5 Other Inequalities

- **Chernoff bounds**

$$P\{X \geq a\} \leq e^{-ta}M(t) \quad \text{for all } t > 0$$

$$P\{X \leq a\} \leq e^{-ta}M(t) \quad \text{for all } t < 0$$

Since the Chernoff bounds hold for all t in either the positive or negative quadrant, we obtain the best bound on $P\{X \geq a\}$ by using the t that minimizes $e^{-ta}M(t)$.

8.5 Chernoff bounds for the standard normal

- If Z is a standard normal random variable, then its moment generating function is $M(t) = e^{t^2/2}$, so the Chernoff bound on $P\{Z \geq a\}$ is given by

$$P\{Z \geq a\} \leq e^{-ta} e^{t^2/2} \text{ for all } t > 0$$

Now the value of t , $t > 0$, that minimizes $e^{t^2/2 - ta}$ is the value that minimizes $t^2/2 - ta$, which is $t = a$. Thus, for $a > 0$, we have

$$P\{Z \geq a\} \leq e^{-a^2/2}$$

Similarly, we can show that, for $a < 0$,

$$P\{Z \leq a\} \leq e^{-a^2/2}$$

8.5 Chernoff bounds for the Poisson random variable

- If X is a Poisson random variable with parameter λ , then its moment generating function is $M(t) = e^{\lambda(e^t-1)}$. Hence, the Chernoff bound on $P\{X \geq i\}$ is

$$P\{X \geq i\} \leq e^{\lambda(e^t-1)} e^{-it} \quad t > 0$$

Minimizing the right side of the preceding inequality is equivalent to minimizing $\lambda(e^t - 1) - it$, and calculus shows that the minimal value occurs when $e^t = i/\lambda$. Provided that $i/\lambda > 1$, this minimizing value of t will be positive. Therefore, assuming that $i > \lambda$ and letting $e^t = i/\lambda$ in the Chernoff bound yields

$$P\{X \geq i\} \leq e^{\lambda(i/\lambda-1)} \left(\frac{\lambda}{i}\right)^i$$

or, equivalently,

$$P\{X \geq i\} \leq \frac{e^{-\lambda} (e\lambda)^i}{i^i}$$

8.5 Other Inequalities

- **Jensen's Inequality:** If $f(x)$ is a convex function, then

$$E[f(X)] \geq f(E[X])$$

if the expectations exist and are finite.

- Remark: A twice-differentiable real-valued function $f(x)$ is said to be convex if $f''(x) \geq 0$ for all x ; similarly, it is said to be concave if $f''(x) \leq 0$.
- Some examples of convex functions are $f(x) = x^2$, $f(x) = e^{ax}$, and $f(x) = -x^{1/n}$ for $x \geq 0$. If $f(x)$ is convex, then $g(x) = -f(x)$ is concave, and vice versa.

8.5 Other Inequalities

- **Hölder's inequality**

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

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

Special case: $E|X| \leq (E|X|^2)^{1/2} \leq (E|X|^3)^{1/3} \leq \dots$

- **Minkowski's inequality**

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

for $p \geq 1$.