# Chapter 7 Convergence and Limit Theorems

## 7.1 Limits and Orders of Magnitude: A Review

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

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

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

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

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

## 7.1 Limits and Orders of Magnitude: A Review

- **Definition: Continuity.** The function $g : \mathbb{R} \to \mathbb{R}$ is *continuous* at point $b$ if for any sequence $\{b_n\}$ such that $\lim_{n\to\infty} b_n = b$, we have $\lim_{n\to\infty} g(b_n) = g(b)$.

- **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

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

- **Remarks:**

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

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

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

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

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

## 7.1 Limits and Orders of Magnitude: A Review

- **Definition: Order of Magnitude.**
  - A sequence $\{b_n\}$ is at most order $n^\lambda$, denoted by $b_n = O(n^\lambda)$ or $b_n/n^\lambda = O(1)$, 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 $|b_n/n^\lambda| < M$.
  - A sequence $\{b_n\}$ is of order smaller than $n^\lambda$, denoted by $b_n = o(n^\lambda)$ or $b_n/n^\lambda = o(1)$, if every real number $\varepsilon > 0$ there exists a finite integer $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$, we have $|b_n/n^\lambda| < \varepsilon$.

- **Remarks:**
  - $b_n = o(f_n)$ implies $b_n = O(f_n)$.
  - Example: Let $b_n = 4 + 2n + 6n^2$, then $b_n = O(n^2)$ and $b_n = o(n^{2+\epsilon})$ for any $\epsilon > 0$.

## 7.1 Limits and Orders of Magnitude: A Review

- **Proposition:** Let $\{a_n\}$ and $\{b_n\}$ be two sequences of real numbers.
  - If $a_n = O(n^\lambda)$ and $b_n = O(n^\tau)$, then $a_n b_n = O(n^{\lambda+\tau})$, and $a_n + b_n = O(n^\kappa)$, where $\kappa = \max\{\lambda, \tau\}$.
  - If $a_n = o(n^\lambda)$ and $b_n = o(n^\tau)$, then $a_n b_n = o(n^{\lambda+\tau})$, and $a_n + b_n = o(n^\kappa)$, where $\kappa = \max\{\lambda, \tau\}$.
  - If $a_n = O(n^\lambda)$ and $b_n = o(n^\tau)$, then $a_n b_n = o(n^{\lambda+\tau})$, and $a_n + b_n = O(n^\kappa)$, where $\kappa = \max(\lambda, \tau)$

- **Question.** For sample mean and variance, how can one measure the closeness of $\bar{X}_n$ to $\mu$ and the closeness of $S_n^2$ to $\sigma_n^2$?

## 7.2 Convergence in Quadratic Mean and $L_p$-convergence

- **Definition: Convergence in Quadratic Mean.** Let $\{Z_n, n = 1, 2, \cdots\}$ be a sequence of random variables and $Z$ be a random variable. Then $\{Z_n\}$ 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 $Z_n - Z = o_{q.m.}(1)$.

- **Definition: $L_p$-convergence** Let $0 < p < \infty$, let $Z_1, Z_2, \cdots$ be a sequence of random variables with $E|Z_n|^p < \infty$. Then we say that $Z_n$ *converges in $L_p$* to $Z$ if

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

## 7.2 Convergence in Quadratic Mean and $L_p$-convergence

- **Example:** Suppose $X_1, X_2, \cdots$ are i.i.d. with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{q.m} \mu.$$

- **Useful inequalities.**

  – 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

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

  for $p \geq 1$.

## 7.4 Convergence in Probability

- **Definition: Convergence in Probability.** A sequence of random variables $\{Z_n\}$, converges in probability to a random variable $Z$ if for every small constant $\varepsilon > 0$,

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

When $Z_n$ converges in probability to $Z$, we write $\lim_{n\to\infty} P(|Z_n - Z| > \varepsilon) = 0$ for every $\varepsilon > 0$, or $Z_n \xrightarrow{p} Z$, or $Z_n - Z = o_p(1)$, or $Z_n - Z \xrightarrow{p} 0$.

- **Remarks:**
  - When $Z_n \xrightarrow{p} b$, where $b$ is a constant, we say that $Z_n$ is *consistent* for $b$.

## 7.4 Convergence in Probability

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

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

    as $n \to \infty$.
  - $Z_n$ is consistent for $\theta$.

## 7.4 Convergence in Probability

- **Definition: Order of Convergence in Probability.** Suppose $\{f_n > 0\}$ is a sequence of real numbers.
  - We say a sequence of random variables $\{Z_n\}$ is of probability order $o_p(f_n)$, denoted by $Z_n = o_p(f_n)$, if

$$Z_n/f_n \xrightarrow{p} 0.$$

  - We say $\{Z_n\}$ is of probability order $O_p(f_n)$, denoted by $Z_n = O_p(f_n)$, if for every $\varepsilon > 0$, we can find $\Delta(\varepsilon) > 0$ such that

$$P\left[|Z_n/f_n| > \Delta(\varepsilon)\right] < \varepsilon$$

for all $n = 1, 2, \cdots$.

## 7.4 Convergence in Probability

- **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.

- **Example.** If $Z_n \sim N(0,1)$ for all $n \geq 1$. Then $Z_n = O_p(1)$ because for any give $\delta > 0$, 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 $n \geq 1$.

## 7.4 Convergence in Probability

- **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.

## 7.4 Convergence in Probability

- **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 $V_n \geq \sigma_1^2 + \cdots + \sigma_n^2$. Then for each constant $\varepsilon > 0$,

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

- **Theorem: Weak Law of Large Numbers (WLLN).** Suppose random variables $X_1, X_2, \cdots$ 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

$$P[|\bar{X}_n - \mu| \leq \varepsilon] \to 1 \text{ as } n \to \infty,$$

  or $\bar{X}_n - \mu \xrightarrow{p} 0$ or $\bar{X}_n - \mu = o_p(1)$.

- **Remark:** In WLLN, the second moment condition $\text{var}(X_i) < \infty$ (or equivalently, $E(X_i^2) < \infty$) is not necessary.

## 7.4 Convergence in Probability

- **Lemma.** Suppose $Z_n \to Z$ in $L_p$ for some $p > 0$. Then $Z_n \xrightarrow{p} Z$.

- **Proof:** By Markov's inequality, for all $\varepsilon > 0$,

$$P[|Z_n - Z| > \varepsilon] \leq \frac{E|Z_n - Z|^p}{\varepsilon^p} \to 0$$

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

- **Remarks:**
  - $Z_n \xrightarrow{p} Z$ does not imply $Z_n \to Z$ in $L_p$.
  - Example: let the pmf of d.r.v. $Z_n$ be

$$f_n(z) = \begin{cases} 1 - 1/n & \text{for } z = 0, \\ 1/n & \text{for } z = n. \end{cases}$$

then $Z_n \xrightarrow{p} 0$, but $EZ_n^2 = n \to \infty \neq 0$.

## 7.4 Convergence in Probability

- **Example.** Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID $N(\mu, \sigma^2)$ random sample. Then $S_n^2 \xrightarrow{p} \sigma^2$.

- **Lemma: Continuity.** Suppose $g : \mathbb{R} \to \mathbb{R}$ is a continuous function, and $Z_n \xrightarrow{p} Z$, then $g(Z_n) \xrightarrow{p} g(Z)$.

- **Theorem.** If function $g$ is continuous at point $b$ and $Z_n \xrightarrow{p} b$, then $g(Z_n) \xrightarrow{p} g(b)$.

- **Proof.**

  – Because $g$ is continuous at $b$, for every fixed $\varepsilon > 0$, we can find $\delta > 0$, s.t. $|z - b| < \delta$ implies $|g(z) - g(b)| < \varepsilon$.

  – For every fixed $\varepsilon > 0$, $\exists \delta$ such that

$$1 = \lim_{n \to \infty} P(|Z_n - b| < \delta) \le \lim_{n \to \infty} P(|g(Z_n) - g(b)| < \varepsilon).$$

## 7.5 Almost Sure Convergence

- **Definition: Almost Sure Convergence.** A sequence of random variables $Z_1, Z_2, \cdots$ converges almost surely to a random variable $Z$ if for every $\varepsilon > 0$,

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

or equivalently,

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

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

- **Remark:** Almost Sure Convergence is called *strong convergence*.

# 7.5 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.*,

$$P(\{s \in [a,b]\}) \stackrel{\triangle}{=} b - a$$

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

  – For $0 \le s < 1$, $Z_n(s) \to Z(s)$ as $n \to \infty$.

  – For $s = 1$, $\lim_{n\to\infty} Z_n(s) \ne Z(s)$.

  – Because $P[Z_n(s) \to Z(s)] = 1$, so $Z_n \xrightarrow{a.s.} Z$.

## 7.5 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

$$Z_n(s) = \begin{cases} 1 & \text{if } s \in [i/2^k, (i+1)/2^k] \text{ for } n = 2^k + i, \\ 0 & \text{otherwise.} \end{cases}$$

  Here $k = \lfloor \log_2(n) \rfloor$ and $i = 0, \cdots, 2^k - 1$.

  – For every $\varepsilon > 0$, $P(|Z_n - Z| > \varepsilon) \leq 1/2^k \to 0$ as $n \to \infty$, so $Z_n$ converges to $Z$ in probability.

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

  – For any $s$, $\lim_{n\to\infty} Z_n(s)$ does not exist, $Z_n$ does not converge to $Z$ almost surely.

## 7.5 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 $\varepsilon > 0$, $P(|Z_n| > \varepsilon) \leq 1/n \to 0$ as $n \to \infty$, so $Z_n$ converges to 0 in probability.
  - For any $0 < s \leq 1$, $\lim_{n\to\infty} Z_n(s) = 0$, so $Z_n$ converges to 0 almost surely.
  - $E|Z_n|^p = \frac{1}{n}e^{np}$, so $Z_n$ does not converge to 0 in $L_p$.

## 7.5 Almost Sure Convergence

- **Lemma.** If $Z_n$ converges to $Z$ almost surely, then $Z_n$ converges to $Z$ in probability.

- **Theorem: Continuity.** Suppose $g(\cdot)$ is a continuous function, and $Z_n$ converges almost surely to $Z$, then $g(Z_n)$ also converges almost surely to $g(Z)$.

- **Proof.**

  – Because $g(\cdot)$ is continuous, $\lim_{n\to\infty} Z_n(s) = Z(s)$ implies $\lim_{n\to\infty} g[Z_n(s)] = g[Z(s)]$.

  –
$$
\begin{aligned}
P\left(\{s \in S : \lim_{n\to\infty} g[Z_n(s)] = g[Z(s)]\}\right) &\geq P\left(\{s \in S : \lim_{n\to\infty} Z_n(s) = Z(s)\}\right) \\
&= 1,
\end{aligned}
$$

  therefore, $g(Z_n) \xrightarrow{a.s} g(Z)$.

- **Theorem: Strong Law of Large Numbers (SLLN).** Suppose random variables $X_1, X_2, \cdots$ are i.i.d. with finite $E\left(X_1^4\right)$. Let $\mu = E(X_1)$ and $Z_n = \frac{1}{n}\sum_{i=1}^n X_i$, then

$$Z_n \xrightarrow{a.s.} \mu.$$

- **Proof.**
  - For simplicity, assume $\mu = 0$.
  - Because $E(X_i) = 0$, $X_1, X_2, \cdots$ are independent, then

$$\begin{aligned} E\left(Z_n\right)^4 &= \frac{1}{n^4}\left[nE(X_1^4) + 3n(n-1)E(X_1^2)E(X_2^2)\right] \\ &\le \frac{C}{n^2}E\left(X_1^4\right), \end{aligned}$$

  where $C$ is a constant.

- – For every $\varepsilon > 0$,

$$
\begin{aligned}
P\left(\limsup_{n\to\infty} \{|Z_n - Z| > \varepsilon\}\right) &= \lim_{n\to\infty} P\left(\cup_{k=n}^{\infty} \{|Z_k| > \varepsilon\}\right) \\
&\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(|Z_k| > \varepsilon) \\
&\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \frac{E\left(Z_k^4\right)}{\varepsilon^4} \\
&\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \frac{C}{k^2\varepsilon^4} E\left(X_1^4\right) \\
&= 0.
\end{aligned}
$$

  – Therefore, $Z_n \xrightarrow{a.s.} 0$.

- **Remark:** In SLLN, then forth moment condition $E\left(X_1^4\right) < \infty$ is not necessary.

## 7.5 Almost Sure Convergence

- **Theorem: Kolmogorov's Strong Law of Large Numbers.** Suppose random variables $X_1, X_2, \cdots$ are i.i.d. with finite $E|X_i|$. Let $\mu = E(X_i)$ and $Z_n = \frac{1}{n}\sum_{i=1}^n X_i$, then

$$Z_n \xrightarrow{a.s.} \mu.$$

- **Theorem: Uniform Strong Law of Large Numbers (USLLN).** Suppose (a) $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample; (b) function $g(x;\theta)$ is continuous over $\Omega \times \Theta$ where $\Omega$ is the support of $X_i$ and $\Theta$ is a compact set in $\mathcal{R}^d$ with $d$ finite and fixed; (c) $E[\sup_{\theta\in\Theta} |g(X_1, \theta)|] < \infty$. Then as $n \to \infty$,

$$\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^n g(X_i, \theta) - E[g(X_1, \theta)] \right| \to 0 \text{ almost surely.}$$

Moreover, $Eg(X_1, \theta)$ is a continuous function of $\theta$ over $\Theta$.

## 7.6 Convergence in Distribution

- **Definition: Convergence in Distribution (Weak Convergence).** A sequence of random variables $Z_1, Z_2, \cdots$ *converges in distribution* to a random variable $Z$, denoted by $Z_n \xrightarrow{d} Z$, if

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

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 $\{Z_n\}$.

- **Remarks:**
  - **Weak Convergence:** $E[h(Z_n)] \to E[h(Z)]$ for all continuous bounded functions $h(\cdot)$.
  - Convergence in distribution means that the cdf's converge, not the random variables. Usually, $Z_n \xrightarrow{d} Z$ does not contain the information about $|Z_n - Z|$.
  - The cdf $F_Z(z)$ at most has countable number of discontinuous points.
  - If $Z_n \xrightarrow{d} Z$, $F_n(z)$ may not tend to $F_Z(z)$ at the discontinuous points of $F_Z$.

## 7.6 Convergence in Distribution

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

- **Lemma.** Let $Z_n$ be a random variable with distribution function $F_n(z)$, and let $Z$ be a random variable with distribution function $F_Z(z)$. If $Z_n \xrightarrow{d} Z$, then $Z_n = O_p(1)$.

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

- **Solution:**

  – For any given $u \geq 0$, we have

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

  – Therefore,

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

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

  – In this example, $\theta - Z_n = O_p(n^{-1})$.

## 7.6 Convergence in Distribution

- **Lemma.** $Z_n \xrightarrow{p} Z$ implies $Z_n \xrightarrow{d} Z$.
- **Proof.**
  - We want to show for every continuous point $z$ of $F_Z$ and every $\varepsilon > 0$, we can find $N$, such that $|F_n(z) - F_Z(z)| < \varepsilon$ for all $n > N$.
  - Because $F_Z(z)$ is continuous at $Z$, we can find $\delta > 0$ such that $|F_Z(z \pm \delta) - F_Z(z)| < \varepsilon/2$.
  - Because $Z \leq z - \delta$ implies $Z_n \leq z$ or $|Z_n - Z| > \delta$, so

$$P(Z \leq z - \delta) \leq P\left(\{Z_n \leq z\} \cup \{|Z_n - Z| > \delta\}\right).$$

    Then we have

$$F_Z(z - \delta) - P(|Z_n - Z| > \delta) \leq F_n(z).$$

  - Similarly,

$$F_n(z) \leq F_Z(z + \delta) + P(|Z_n - Z| > \delta).$$

- – Therefore,

$$F_Z(z-\delta) - P(|Z_n - Z| > \delta) \le F_n(z) \le F_Z(z+\delta) + P(|Z_n - Z| > \delta).$$

  –

$$|F_n(z) - F_Z(z)| < \varepsilon/2 + P(|Z_n - Z| > \delta).$$

  – Because $Z_n \xrightarrow{p} Z$, we can find $N$ such that $P(|Z_n - Z| > \delta) < \varepsilon/2$ for all $n > N$.

  – Then $|F_n(z) - F_Z(z)| < \varepsilon$ for all $n > N$. Hence $Z_n \xrightarrow{d} Z$.

- **Remark:** $Z_n \xrightarrow{d} Z$ does not imply $\lim_{n\to\infty} E(Z_n) = E(Z)$.

## 7.6 Convergence in Distribution

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

- **Definition: Degenerate Distribution.** A random variable $Z$ is said to have a *degenerate distribution* if $P(Z = c) = 1$ for some constant $c$.

- **Theorem.** Let $F_n(z)$ be the CDF of a random variable $Z_n$ whose distribution depends on the positive integer $n$. Let $c$ denotes a constant which does not depend upon $n$. The sequence $\{Z_n, n = 1, 2, \cdots\}$ converges in probability to constant $c$ if and only if the limiting distribution of $Z_n$ is degenerate at $z = c$.

- **Theorem: Continuous Mapping Theorem.** Suppose a sequence of $k \times 1$ random vectors $Z_n \xrightarrow{d} Z$ as $n \to \infty$ and $g : \mathcal{R}^k \to \mathcal{R}^J$ is a continuous function. Then $g(Z_n) \xrightarrow{d} g(Z)$.

## 7.7 Central Limit Theorems

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

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

- **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

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

- **Proof.**

  – Let $Y_i = (X_i - \mu)/\sigma$, then $Z_n = \sum_{i=1}^n Y_i/\sqrt{n}$.

  – Define $\phi(t) = E(e^{iY_1 t})$, then the characteristic function of $Z_n$ is

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

  .

  – Because $\phi'(0) = 0$, $\phi''(0) = -1$, $E\left(e^{itZ_n}\right) \to e^{-t^2/2}$ for all $t$ as $n \to \infty$.

  – Hence $Z_n \xrightarrow{d} N(0,1)$.

- **Remark:** We use the characteristic function rather than the moment generating function, because the mgf may not exist for some distribution with finite variance.

- **Example: Normal Approximation for the Binomial Distribution.** For a Binomial$(n,p)$ random variable $S$, we can write $Z_n = \sum_{i=1}^n X_i$, where $\{X_i\}$ are independent Bernoulli random variables with $P(X_i = 1) = p$. By the CLT, $\frac{Z_n - np}{\sqrt{np(1-p)}}$ approximately follows a $N(0,1)$ distribution.

- **Examples: Normal approximation of $\chi_n^2$.** Suppose $X_1, \cdots, X_n$ is a random sample following $N(0,1)$ distribution. Then $Z_n = \sum_{i=1}^n X_n^2$ follows a $\chi_n^2$ distribution. We have

$$E(X_1^2) = 1, \qquad \mathrm{Var}(X_1^2) = 2.$$

By the CLT, $\frac{Z_n - n}{\sqrt{2n}}$ approximately follows a $N(0,1)$ distribution when $n$ is large.

## 7.7 Central Limit Theorems

- **Theorem: Liapounov's CLT.** Suppose the random variables $X_1, X_2, \cdots$ are independent and $E|X_i - E(X_i)|^3 < \infty$ for $i = 1, 2, \cdots$. Let $\mu_i = E(X_i)$, $\sigma_i^2 = \mathrm{Var}(X_i)$, and suppose

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

  Then as $n \to \infty$, we have

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

- **Remark:** This theorem relaxes the identical distribution assumption.

## 7.7 Central Limit Theorems

- **Theorem: Slutsky's Theorem.** Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, where $c$ is a constant. Then
  - $X_n + Y_n \xrightarrow{d} X + c$.
  - $X_n - Y_n \xrightarrow{d} X - c$.
  - $X_n Y_n \xrightarrow{d} Xc$.
  - $X_n / Y_n \xrightarrow{d} X/c$. if $c \neq 0$.

- **Remarks:**
  - If $Y_n \xrightarrow{p} Y$, where $Y$ does not follow a degenerate distribution, then $X_n + Y_n$ may not converge to $X + Y$ in distribution.
  - Example: Suppose $X, X_1, X_2, \cdots$ are i.i.d. following $N(0,1)$ distribution. Let $Y_n = X$ for $n = 1, 2, \cdots$ and let $Y = X$. Then $X_n + Y_n \sim N(0,2)$, but $X + Y = 2X \sim N(0,4)$.

## 7.7 Central Limit Theorems

- **Example:** Suppose $X_1, X_2, \cdots$ are i.i.d. with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. We have
$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$
If $S_n^2 \xrightarrow{p} \sigma^2$, then
$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$$
by Slutsky's theorem.

- **Example.** Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ as $n \to \infty$. Do we have the following results?
  - $X_n \pm Y_n \xrightarrow{d} X \pm Y$ as $n \to \infty$;
  - $X_n Y_n \xrightarrow{d} XY$ as $n \to \infty$.

- **Lemma: Delta Method.** Suppose $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$, $g(\cdot)$ is a continuously differentiable ($g'(\cdot)$ is continuous) with $g'(\mu) \neq 0$. Then as $n \to \infty$,
$$\sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{d} N\left(0, \sigma^2[g'(\mu)]^2\right).$$

- **Remark:** By the Slutsky's theorem again, we have
$$\frac{\sqrt{n}[g(\bar{X}_n) - g(\mu)]}{\sigma g'(\bar{X}_n)} \xrightarrow{d} N(0,1).$$

- **Example:** Suppose $\sqrt{n}(Z_n - \mu) \xrightarrow{d} N(0, \sigma^2)$, find the limiting distribution of $\sqrt{n}(1/Z_n - 1/\mu)$.

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

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

- **Proof.**

  – Using the first order Taylor expansion,

$$n[g(\bar{X}_n) - g(\mu)] = \frac{g''(\lambda)}{2}\left[\sqrt{n}(\bar{X}_n - \mu)\right]^2,$$

  where $\lambda(s)$ is a point between $\mu$ and $Z_n(s)$. Hence $\lambda(s) \xrightarrow{p} \mu$.

  – By the Slutsky's theorem, $n[g(\bar{X}_n) - g(\mu)] \xrightarrow{d} \sigma^2 \frac{g''(\mu)}{2}\chi_1^2$

## 7.7 Central Limit Theorems

- **Lemma: 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 onlu if $a'Z_n \xrightarrow{d} a'Z$ for every constant vector $a \neq 0$.

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

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

- **THE END**

- **Theorem: Multivariate Delta Method.** Suppose $X_1, X_2, \cdots$ is a sequence of i.i.d. random vectors, where $X_i = (X_{1,i}, \cdots, X_{p,i})'$. Let $\mu_j = E(X_{j,1})$, $\sigma_{j,k} = \mathrm{Cov}(X_{j,1}, X_{k,1}) < \infty$. $g(x_1, \cdots, x_p)$ is a continuous differentiable function. Let

$$\begin{aligned}\tau^2 &= \sum_j \sum_k \sigma_{j,k} \frac{\partial g(\mu_1, \cdots, \mu_p)}{\partial x_j} \frac{\partial g(\mu_1, \cdots, \mu_p)}{\partial x_k} \\ &= \left(\frac{\partial g}{\partial \boldsymbol{x}}\right)' \Sigma \left(\frac{\partial g}{\partial \boldsymbol{x}}\right) \Big|_{\boldsymbol{x}=(\mu_1,\cdots,\mu_p)'},\end{aligned}$$

where $\Sigma = (\sigma_{j,k})_{p\times p}$ is the variance matrix of random vector $X_i$. Then

$$\sqrt{n}\left[g(\overline{X}_{1,n}, \cdots, \overline{X}_{p,n}) - g(\mu_1, \cdots, \mu_p)\right] \xrightarrow{d} N(0, \tau^2),$$

where $\overline{X}_{j,n} = \frac{1}{n}\sum_{i=1}^n X_{j,i}$.

- **Example: Ratio Estimator.** Suppose $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ are two sequence of i.i.d. random variables. The variances of $X_i$ and $Y_i$ are finite. Let $\mu_X = E(X_1)$ and $\mu_Y = E(Y_1)$.
  - We can use $\overline{X}_n/\overline{Y}_n$ as an estimator of the ratio $\mu_X/\mu_Y$.
  - Using the multivariate delta method, we have

$$\sqrt{n}\left[\overline{X}_n/\overline{Y}_n - \mu_X/\mu_Y\right] \xrightarrow{d} N(0, \tau^2),$$

where

$$\begin{aligned} \tau^2 &= \frac{1}{\mu_Y^2}\mathrm{Var}(X) - \frac{2\mu_X}{\mu_Y^3}\mathrm{Cov}(X,Y) + \frac{\mu_X^2}{\mu_Y^4}\mathrm{Var}(Y) \\ &= \frac{\mu_X^2}{\mu_Y^2}\left(\frac{\mathrm{Var}(X)}{\mu_X^2} - \frac{2\mathrm{Cov}(X,Y)}{\mu_X\mu_Y} + \frac{\mathrm{Var}(Y)}{\mu_Y^2}\right). \end{aligned}$$

## 7.7 Central Limit Theorems

- **Example: Importance Sampling.** Suppose $X$ is a random variable with pdf $f(x)$. We want to calculate $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)\,dx$ $(< \infty)$ that does not have an analytic form.

  – If we can draw i.i.d samples $X_1, \cdots, X_n$ from distribution $f(x)$, then

  $$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{a.s.} E[h(X)].$$

  – If we can't directly draw samples from $f(x)$, we can draw i.i.d samples $X_1, \cdots, X_n$ from a different distribution $g(x)$, whose support contains the support of $f(x)$. Then let $w(X_i) = f(X_i)/g(X_i)$, we have

  $$\frac{1}{n}\sum_{i=1}^{n} w(X_i)h(X_i) \xrightarrow{a.s.} E[h(X)],$$

  and (why?)

  $$\frac{\sum_{i=1}^{n} w(X_i)h(X_i)}{\sum_{i=1}^{n} w(X_i)} \xrightarrow{a.s.} E[h(X)].$$

## 7.7 Central Limit Theorems

- – In many cases, $\frac{\sum_{i=1}^n w(X_i)h(X_i)}{\sum_{i=1}^n w(X_i)}$ is a more convenient estimate of $E[h(X)]$ because we don't need to calculate some multiplicative constants in $w(X_i)$.

  – If $\mathrm{Var}[w(X_1)] < \infty$, $\mathrm{Var}[w(X_1)h(X_1)] < \infty$, and $\mathrm{Cov}[w(X_1)h(X_1), w(X_1)] < \infty$, then by the delta method,

$$\sqrt{n}\left[\frac{\sum_{i=1}^n w(X_i)h(X_i)}{\sum_{i=1}^n w(X_i)} - E[h(X)]\right] \xrightarrow{d} N(0, \tau^2),$$

  where

$$\begin{aligned}\tau^2 &= \mathrm{Var}_g[w(X_1)h(X_1)] - 2E_f[h(X)]\mathrm{Cov}_g[w(X_1)h(X_1), w(X_1)] \\ &+ E_f^2[h(X)]\mathrm{Var}_g[w(X_1)].\end{aligned}$$

  – We want to choose $g(x)$ so that $\tau^2$ is small.