Introduction
• Why large sample theory

  – studying small sample property is usually difficult and complicated

  – large sample theory studies the limit behavior of a sequence of random variables, say $X_n$.

  – example: $\bar{X}_n \to \mu$, $\sqrt{n}(\bar{X}_n - \mu)$
Modes of Convergence
• Convergence almost surely

**Definition 3.1** $X_n$ is said to converge almost surely to $X$, denoted by $X_n \to_{a.s.} X$, if there exists a set $A \subset \Omega$ such that $P(A^c) = 0$ and for each $\omega \in A$, $X_n(\omega) \to X(\omega)$ in real space.
• Equivalent condition

\[ \{ \omega : X_n(\omega) \to X(\omega) \}^c \]

\[ = \bigcup_{\epsilon > 0} \bigcap_n \{ \omega : \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \epsilon \} \]

\[ \Rightarrow X_n \to_{a.s.} X \text{ iff } \]

\[ P(\sup_{m \geq n} |X_m - X| > \epsilon) \to 0 \]
• Convergence in probability

**Definition 3.2** $X_n$ is said to converge in probability to $X$, denoted by $X_n \xrightarrow{p} X$, if for every $\epsilon > 0$,

\[ P(|X_n - X| > \epsilon) \to 0. \]
• Convergence in moments/means

**Definition 3.3** $X_n$ is said to *converge in rth mean* to $X$, denote by $X_n \to_r X$, if

$$E[|X_n - X|^r] \to 0 \text{ as } n \to \infty$$

for functions $X_n, X \in L_r(P)$, where $X \in L_r(P)$ means $\int |X|^r dP < \infty$. 
• **Convergence in distribution**

**Definition 3.4**  $X_n$ is said to *converge in distribution* of $X$, denoted by $X_n \xrightarrow{d} X$ or $F_n \xrightarrow{d} F$ (or $L(X_n) \rightarrow L(X)$ with $L$ referring to the “law” or “distribution”), if the distribution functions $F_n$ and $F$ of $X_n$ and $X$ satisfy

$$F_n(x) \rightarrow F(x) \text{ as } n \rightarrow \infty$$

for each continuity point $x$ of $F$. 


• Uniform integrability

**Definition 3.5** A sequence of random variables \( \{X_n\} \) is **uniformly integrable** if

\[
\lim_{\lambda \to \infty} \lim_{n \to \infty} \sup E \{|X_n| I(|X_n| \geq \lambda)\} = 0.
\]

• A note

  – Convergence almost surely and convergence in probability are the same as we defined in measure theory.

  – Two new definitions are
    * convergence in $r$th mean
    * convergence in distribution
• “convergence in distribution”
  – is very different from others
  – example: a sequence $X, Y, X, Y, X, Y, \ldots$ where $X$ and $Y$ are $N(0, 1)$; the sequence converges in distribution to $N(0, 1)$ but the other modes do not hold.
  – “convergence in distribution” is important for asymptotic statistical inference.
• Relationship among different modes

**Theorem 3.1**  
A. If $X_n \rightarrow_{a.s.} X$, then $X_n \rightarrow_p X$.  
B. If $X_n \rightarrow_p X$, then $X_{n_k} \rightarrow_{a.s.} X$ for some subsequence $X_{n_k}$.  
C. If $X_n \rightarrow_r X$, then $X_n \rightarrow_p X$.  
D. If $X_n \rightarrow_p X$ and $|X_n|^r$ is uniformly integrable, then $X_n \rightarrow_r X$.  
E. If $X_n \rightarrow_p X$ and $\limsup_n E|X_n|^r \leq E|X|^r$, then $X_n \rightarrow_r X$. 
F. If $X_n \rightarrow_r X$, then $X_n \rightarrow_{r'} X$ for any $0 < r' \leq r$.

G. If $X_n \rightarrow_p X$, then $X_n \rightarrow_d X$.

H. $X_n \rightarrow_p X$ if and only if for every subsequence $\{X_{n_k}\}$ there exists a further subsequence $\{X_{n_{k,l}}\}$ such that $X_{n_{k,l}} \rightarrow_{a.s.} X$.

I. If $X_n \rightarrow_d c$ for a constant $c$, then $X_n \rightarrow_p c$. 
\[ \xi_n \overset{\text{a.s.}}{\to} \xi \]

for a subsequence

\[ \xi_n \overset{p}{\to} \xi \]

\[ \xi_n \overset{d}{\to} \xi \]

\[ |\xi_n|^r \text{ uniformly integrable} \]

\[ \lim_{n \to \infty} E[|\xi_n|^r] \leq E[|\xi|^r] \]

\[ \xi \text{ is a constant} \]
Proof

A and B follow from the results in the measure theory.

Prove C. Markov inequality: for any increasing function $g(\cdot)$ and random variable $Y$, $P(|Y| > \epsilon) \leq E\left[\frac{g(|Y|)}{g(\epsilon)}\right]$.

$\Rightarrow P(|X_n - X| > \epsilon) \leq E\left[\frac{|X_n - X|^r}{\epsilon^r}\right] \to 0.$
Prove D. It is sufficient to show that for any subsequence of \( \{X_n\} \), there exists a further subsequence \( \{X_{n_k}\} \) such that
\[
E|X_{n_k} - X|^r \to 0.
\]

For any subsequence of \( \{X_n\} \), from B, there exists a further subsequence \( \{X_{n_k}\} \) such that \( X_{n_k} \to_{a.s.} X \). For any \( \epsilon \), there exists \( \lambda \) such that \( \limsup_{n_k} E[|X_{n_k}|^r I(|X_{n_k}|^r \geq \lambda)] < \epsilon \).

Particularly, choose \( \lambda \) such that \( P(|X|^r = \lambda) = 0 \)
\[
\Rightarrow |X_{n_k}|^r I(|X_{n_k}|^r \geq \lambda) \to_{a.s.} |X|^r I(|X|^r \geq \lambda).
\]

\[
\Rightarrow \text{By the Fatou’s Lemma,}
\]
\[
E[|X|^r I(|X|^r \geq \lambda)] \leq \limsup_{n_k} E[|X_{n_k}|^r I(|X_{n_k}|^r \geq \lambda)] < \epsilon.
\]
\[
\begin{align*}
\Rightarrow \\
E[|X_{n_k} - X|^r] \\
\leq E[|X_{n_k} - X|^r I(|X_{n_k}|^r < 2\lambda, |X|^r < 2\lambda)] \\
+ E[|X_{n_k} - X|^r I(|X_{n_k}|^r \geq 2\lambda, \text{ or } |X|^r \geq 2\lambda)] \\
\leq E[|X_{n_k} - X|^r I(|X_{n_k}|^r < 2\lambda, |X|^r < 2\lambda)] \\
+ 2^r E[(|X_{n_k}|^r + |X|^r)I(|X_{n_k}|^r \geq 2\lambda, \text{ or } |X|^r \geq 2\lambda)],
\end{align*}
\]

where the last inequality follows from the inequality

\[(x + y)^r \leq 2^r (\max(x, y))^r \leq 2^r (x^r + y^r), x \geq 0, y \geq 0.\]

When \(n_k\) is large, the second term is bounded by

\[2 \times 2^r \{E[|X_{n_k}|^r I(|X_{n_k}| \geq \lambda)] + E[|X|^r I(|X| \geq \lambda)]\} \leq 2^{r+1}\epsilon.\]

\[\Rightarrow \limsup_n E[|X_{n_k} - X|^r] \leq 2^{r+1}\epsilon.\]
Prove E. It is sufficient to show that for any subsequence of \( \{X_n\} \), there exists a further subsequence \( \{X_{n_k}\} \) such that
\[
E[|X_{n_k} - X|^r] \to 0.
\]

For any subsequence of \( \{X_n\} \), there exists a further subsequence \( \{X_{n_k}\} \) such that \( X_{n_k} \to_{a.s.} X \). Define
\[
Y_{n_k} = 2^r (|X_{n_k}|^r + |X|^r) - |X_{n_k} - X|^r \geq 0.
\]

\[\Rightarrow\] By the Fatou’s Lemma,
\[
\int \lim \inf_{n_k} Y_{n_k} \, dP \leq \lim \inf_{n_k} \int Y_{n_k} \, dP.
\]

It is equivalent to
\[
2^{r+1} E[|X|^r] \leq \lim \inf_{n_k} \left\{ 2^r E[|X_{n_k}|^r] + 2^r E[|X|^r] - E[|X_{n_k} - X|^r] \right\}.
\]
Prove F. The Hölder inequality:

\[
\int |f(x)g(x)|d\mu \leq \left\{ \int |f(x)|^p d\mu(x) \right\}^{1/p} \left\{ \int |g(x)|^q d\mu(x) \right\}^{1/q},
\]

\[
\frac{1}{p} + \frac{1}{q} = 1.
\]

Choose \( \mu = P \), \( f = |X_n - X|^{r'} \), \( g \equiv 1 \) and \( p = r/r' \), \( q = r/(r - r') \) in the Hölder inequality

\[
\Rightarrow \quad E[|X_n - X|^{r'}] \leq E[|X_n - X|^{r}]^{r'/r} \to 0.
\]
Prove G. \( X_n \xrightarrow{p} X \). If \( P(X = x) = 0 \), then for any \( \epsilon > 0 \),

\[
P(|I(X_n \leq x) - I(X \leq x)| > \epsilon) \\
= P(|I(X_n \leq x) - I(X \leq x)| > \epsilon, |X - x| > \delta) \\
+ P(|I(X_n \leq x) - I(X \leq x)| > \epsilon, |X - x| \leq \delta) \\
\leq P(X_n \leq x, X > x + \delta) + P(X_n > x, X < x - \delta) \\
+ P(|X - x| \leq \delta) \\
\leq P(|X_n - X| > \delta) + P(|X - x| \leq \delta).
\]

The first term converges to zero since \( X_n \xrightarrow{p} X \).
The second term can be arbitrarily small if \( \delta \) is small, since

\[
\lim_{\delta \to 0} P(|X - x| \leq \delta) = P(X = x) = 0.
\]

\( \Rightarrow I(X_n \leq x) \xrightarrow{p} I(X \leq x) \)

\( \Rightarrow F_n(x) = E[I(X_n \leq x)] \to E[I(X \leq x)] = F(x). \)
Prove H. One direction follows from B.

To prove the other direction, use the contradiction. Suppose there exists $\epsilon > 0$ such that $P(|X_n - X| > \epsilon)$ does not converge to zero. 
$\Rightarrow$ find a subsequence $\{X_{n'}\}$ such that $P(|X_{n'} - X| > \epsilon) > \delta$ for some $\delta > 0$.

However, by the condition, there exists a further subsequence $X_{n''}$ such that $X_{n''} \rightarrow_{a.s.} X$ then $X_{n''} \rightarrow_p X$ from A. Contradiction!
Prove I. Let $X \equiv c$.

\[
P(|X_n - c| > \epsilon) \leq 1 - F_n(c + \epsilon) + F_n(c - \epsilon)
\]
\[
\rightarrow 1 - F_X(c + \epsilon) + F(c - \epsilon) = 0.
\]
Some counter-examples

(Example 1) Suppose that $X_n$ is degenerate at a point $1/n$; i.e., $P(X_n = 1/n) = 1$. Then $X_n$ converges in distribution to zero. Indeed, $X_n$ converges almost surely.
(Example 2) $X_1, X_2, \ldots$ are i.i.d with standard normal distribution. Then $X_n \rightarrow_d X_1$ but $X_n$ does not converge in probability to $X_1$. 
(Example 3) Let \( Z \) be a random variable with a uniform distribution in \([0, 1]\). Let

\[
X_n = I(m2^{-k} \leq Z < (m + 1)2^{-k}) \text{ when } n = 2^k + m
\]

where \( 0 \leq m < 2^k \). Then it is shown that \( X_n \) converges in probability to zero but not almost surely. This example is already given in the second chapter.
(Example 4) Let $Z$ be $Uniform(0, 1)$ and let $X_n = 2^n I(0 \leq Z < 1/n)$. Then $E[|X_n|^r] \to \infty$ but $X_n$ converges to zero almost surely.
• Result for convergence in $r$th mean

**Theorem 3.2 (Vitali’s theorem)** Suppose that $X_n \in L_r(P)$, i.e., $\|X_n\|_r < \infty$, where $0 < r < \infty$ and $X_n \rightarrow_p X$. Then the following are equivalent:

A. $\{|X_n|^r\}$ are uniformly integrable.

B. $X_n \rightarrow_r X$.

C. $E[|X_n|^r] \rightarrow E[|X|^r]$. 
One sufficient condition for uniform integrability

Liapunov condition: there exists a positive constant $\epsilon_0$ such that $\limsup_n E[|X_n|^{r+\epsilon_0}] < \infty$

\[
E[|X_n|^r I(|X_n|^r \geq \lambda)] \leq \frac{E[|X_n|^{r+\epsilon_0}]}{\lambda^{\epsilon_0}}
\]
Integral inequalities
• Young’s inequality

\[ |ab| \leq \frac{|a|^p}{p} + \frac{|b|^q}{q}, \quad a, b > 0, \]

where the equality holds if and only if \( a = b \).

\( \log x \) is concave:

\[ \log\left(\frac{1}{p}|a|^p + \frac{1}{q}|b|^q\right) \geq \frac{1}{p} \log |a|^p + \frac{1}{q} \log |b|. \]

Geometric interpretation (insert figure here):
• Hölder inequality

\[ \int |f(x)g(x)| \, d\mu(x) \leq \left\{ \int |f(x)|^p \, d\mu(x) \right\}^{\frac{1}{p}} \left\{ \int |g(x)|^q \, d\mu(x) \right\}^{\frac{1}{q}}. \]

– in the Young’s inequality, let

\[ a = f(x)/ \left\{ \int |f(x)|^p \, d\mu(x) \right\}^{1/p} \]
\[ b = g(x)/ \left\{ \int |g(x)|^q \, d\mu(x) \right\}^{1/q}. \]

– when \( \mu = P \) and \( f = X(\omega), \, g = 1 \), \( \mu_r^{s-t} \mu_t^{r-s} \geq \mu_s^{r-t} \)

where \( \mu_r = E[|X|^r] \) and \( r \geq s \geq t \geq 0 \).

– when \( p = q = 2 \), obtain Cauchy-Schwartz inequality:

\[ \int |f(x)g(x)| \, d\mu(x) \leq \left\{ \int f(x)^2 \, d\mu(x) \right\}^{\frac{1}{2}} \left\{ \int g(x)^2 \, d\mu(x) \right\}^{\frac{1}{2}}. \]
• *Minkowski’s inequality* $r > 1$,

$$\|X + Y\|_r \leq \|X\|_r + \|Y\|_r.$$  

– derivation:

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

$$\leq E[|X|^r]^{1/r} E[|X+Y|^r]^{1-1/r} + E[|Y|^r]^{1/r} E[|X+Y|^r]^{1-1/r}.$$  

– $\| \cdot \|_r$ in fact is a norm in the linear space \{ $X : \|X\|_r < \infty$ \}. Such a normed space is denoted as $L_r(P)$.  

\begin{itemize}
  \item \textit{Markov’s inequality}
  \[ P(|X| \geq \epsilon) \leq \frac{E[g(|X|)]}{g(\epsilon)}, \]
  where \( g \geq 0 \) is a increasing function in \([0, \infty)\).
  
  \text{-} Derivation:
  \[
P(|X| \geq \epsilon) \leq P(g(|X|) \geq g(\epsilon))
  = E[I(g(|X|) \geq g(\epsilon))] \leq E\left[\frac{g(|X|)}{g(\epsilon)}\right].
  \]
  
  \text{-} When \( g(x) = x^2 \) and \( X \) replaced by \( X - \mu \), obtain \textit{Chebyshev’s inequality}:
  \[
P(|X - \mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}.
  \]
\end{itemize}
• Application of Vitali’s theorem

- \( Y_1, Y_2, ... \) are i.i.d with mean \( \mu \) and variance \( \sigma^2 \). Let \( X_n = \bar{Y}_n \).

- By the Chebyshev’s inequality,

\[
P(|X_n - \mu| > \epsilon) \leq \frac{Var(X_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0.
\]

\( \Rightarrow \) \( X_n \to_p \mu \).

- From the Liapunov condition with \( r = 1 \) and \( \epsilon_0 = 1 \), \( |X_n - \mu| \) satisfies the uniform integrability condition

\[
E[|X_n - \mu|] \to 0.
\]
Convergence in Distribution
“Convergence in distribution is the most important mode of convergence in statistical inference.”
Equivalent conditions

**Theorem 3.3 (Portmanteau Theorem)** The following conditions are equivalent.

(a). $X_n$ converges in distribution to $X$.

(b). For any bounded continuous function $g(\cdot)$, $E[g(X_n)] \to E[g(X)]$.

(c). For any open set $G$ in $R$, \[ \liminf_n P(X_n \in G) \geq P(X \in G). \]

(d). For any closed set $F$ in $R$, \[ \limsup_n P(X_n \in F) \leq P(X \in F). \]

(e). For any Borel set $O$ in $R$ with $P(X \in \partial O) = 0$ where $\partial O$ is the boundary of $O$, $P(X_n \in O) \to P(X \in O)$. 
Proof

(a) ⇒ (b). Without loss of generality, assume \( |g(x)| \leq 1 \). We choose \([-M, M]\) such that \( P(|X| = M) = 0 \).

Since \( g \) is continuous in \([-M, M]\), \( g \) is uniformly continuous in \([-M, M]\).

⇒ Partition \([-M, M]\) into finite intervals \( I_1 \cup \ldots \cup I_m \) such that within each interval \( I_k \), \( \max_{I_k} g(x) - \min_{I_k} g(x) \leq \epsilon \) and \( X \) has no mass at all the endpoints of \( I_k \) (why?).
Therefore, if choose any point $x_k \in I_k, k = 1, \ldots, m,$

\[
\begin{align*}
|E[g(X_n)] - E[g(X)]| &\leq E[|g(X_n)|I(|X_n| > M)] + E[|g(X)|I(|X| > M)] \\
+ |E[g(X_n)I(|X_n| \leq M)] - \sum_{k=1}^{m} g(x_k)P(X_n \in I_k)| \\
+ \left| \sum_{k=1}^{m} g(x_k)P(X_n \in I_k) - \sum_{k=1}^{m} g(x_k)P(X \in I_k) \right| \\
+ |E[g(X)I(|X| \leq M)] - \sum_{k=1}^{m} g(x_k)P(X \in I_k)| \\
\leq P(|X_n| > M) + P(|X| > M) \\
+ 2\epsilon + \sum_{k=1}^{m} |P(X_n \in I_k) - P(X \in I_k)|.
\end{align*}
\]

$\Rightarrow \limsup_n |E[g(X_n)] - E[g(X)]| \leq 2P(|X| > M) + 2\epsilon$. Let $M \to \infty$ and $\epsilon \to 0.$
(b) ⇒ (c). For any open set $G$, define $g(x) = 1 - \frac{\epsilon}{\epsilon + d(x, G^c)}$, where $d(x, G^c)$ is the minimal distance between $x$ and $G^c$, $\inf_{y \in G^c} |x - y|$.

For any $y \in G^c$, $d(x_1, G^c) - |x_2 - y| \leq |x_1 - y| - |x_2 - y| \leq |x_1 - x_2|$, 
⇒ $d(x_1, G^c) - d(x_2, G^c) \leq |x_1 - x_2|$.
⇒ $g(x_1) - g(x_2) \leq \epsilon^{-1}d(x_1, G^c) - d(x_2, G^c) \leq \epsilon^{-1}|x_1 - x_2|$.
⇒ $g(x)$ is continuous and bounded.
⇒ $E[g(X_n)] \rightarrow E[g(X)]$.

Note $0 \leq g(x) \leq I_G(x)$
⇒ 
$$\liminf_n P(X_n \in G) \geq \liminf_n E[g(X_n)] \rightarrow E[g(X)].$$

Let $\epsilon \rightarrow 0 \Rightarrow E[g(X)]$ converges to $E[I(X \in G)] = P(X \in G)$.

(c) ⇒ (d). This is clear by taking complement of $F$. 
(d)⇒(e). For any $O$ with $P(X \in \partial O) = 0$,

$$\limsup_{n} P(X_n \in O) \leq \limsup_{n} P(X_n \in \bar{O}) \leq P(X \in \bar{O}) = P(X \in \bar{O}),$$

$$\liminf_{n} P(X_n \in O) \geq \liminf_{n} P(X_n \in O^o) \geq P(X \in O^o) = P(X \in O).$$

(e)⇒(a). Choose $O = (-\infty, x]$ with $P(X \in \partial O) = P(X = x) = 0$. 
• **Counter-examples**
  
  – Let $g(x) = x$, a continuous but unbounded function. Let $X_n$ be a random variable taking value $n$ with probability $1/n$ and value 0 with probability $(1 - 1/n)$. Then $X_n \to_d 0$. However, $E[g(X)] = 1$ does not converge to 0.
  
  – The continuity at boundary in (e) is also necessary: let $X_n$ be degenerate at $1/n$ and consider $O = \{x : x > 0\}$. Then $P(X_n \in O) = 1$ but $X_n \to_d 0$. 
Weak Convergence and Characteristic Functions
Theorem 3.4 (Continuity Theorem) Let $\phi_n$ and $\phi$ denote the characteristic functions of $X_n$ and $X$ respectively. Then $X_n \xrightarrow{d} X$ is equivalent to $\phi_n(t) \to \phi(t)$ for each $t$. 
Proof

To prove $\Rightarrow$ direction, from (b) in Theorem 3.1,

$$
\phi_n(t) = E[e^{itX_n}] \rightarrow E[e^{itX}] = \phi(t).
$$

The proof of $\Leftarrow$ direction consists of a few tricky constructions (skipped).
• One simple example $X_1, \ldots, X_n \sim Bernoulli(p)$

$$
\phi_{\bar{X}_n}(t) = E[e^{it(X_1+\ldots+X_n)/n}] = (1 = p + pe^{it/n})^n
$$

$$
= (1 - p + p + itp/n + o(1/n))^n \to e^{itp}.
$$

Note the limit is the c.f. of $X = p$. Thus, $\bar{X}_n \to_d p$ so $\bar{X}_n$ converges in probability to $p$.  

Generalization to multivariate random vectors

- \( X_n \rightarrow_d X \) if and only if
  \[ E[\exp\{it'X_n\}] \rightarrow E[\exp\{it'X\}] \]
  where \( t \) is any \( k \)-dimensional constant

- Equivalently, \( t'X_n \rightarrow_d t'X \) for any \( t \)

- to study the weak convergence of random vectors, we can reduce to study the weak convergence of one-dimensional linear combination of the random vectors

- This is the well-known Cramér-Wold’s device
Theorem 3.5 (The Cramér-Wold device) Random vector $X_n$ in $\mathbb{R}^k$ satisfy $X_n \rightarrow_d X$ if and only $t'X_n \rightarrow_d t'X$ in $\mathbb{R}$ for all $t \in \mathbb{R}^k$. 
Properties of Weak Convergence
Theorem 3.6 (Continuous mapping theorem)
Suppose $X_n \rightarrow_{a.s.} X$, or $X_n \rightarrow_{p} X$, or $X_n \rightarrow_{d} X$. Then for any continuous function $g(\cdot)$, $g(X_n)$ converges to $g(X)$ almost surely, or in probability, or in distribution.
Proof

If $X_n \rightarrow_{a.s.} X$, then $g(X_n) \rightarrow_{a.s.} g(X)$.

If $X_n \rightarrow_{p} X$, then for any subsequence, there exists a further subsequence $X_{n_k} \rightarrow_{a.s.} X$. Thus, $g(X_{n_k}) \rightarrow_{a.s.} g(X)$. Then $g(X_n) \rightarrow_{p} g(X)$ from (H) in Theorem 3.1.

To prove that $g(X_n) \rightarrow_{d} g(X)$ when $X_n \rightarrow_{d} X$, use (b) of Theorem 3.1.
• **One remark**

Theorem 3.6 concludes that \( g(X_n) \to_d g(X) \) if \( X_n \to_d X \) and \( g \) is continuous. In fact, this result still holds if \( P(X \in C(g)) = 1 \) where \( C(g) \) contains all the continuity points of \( g \). That is, if \( g \)'s discontinuity points take zero probability of \( X \), the continuous mapping theorem holds.
Theorem 3.7 (Slutsky theorem) Suppose $X_n \rightarrow_d X$, $Y_n \rightarrow_p y$ and $Z_n \rightarrow_p z$ for some constant $y$ and $z$. Then $Z_n X_n + T_n \rightarrow_d z X + y$. 
Proof

First show that $X_n + Y_n \rightarrow_d X + y$.

For any $\epsilon > 0$,

$$P(X_n + Y_n \leq x) \leq P(X_n + Y_n \leq x, |Y_n - y| \leq \epsilon) + P(|Y_n - y| > \epsilon)$$

$$\leq P(X_n \leq x - y + \epsilon) + P(|Y_n - y| > \epsilon).$$

$$\Rightarrow \lim \sup_n F_{X_n+Y_n}(x) \leq \lim \sup_n F_{X_n}(x - y + \epsilon) \leq F_X(x - y + \epsilon).$$
On the other hand,

\[
P(X_n + Y_n > x) = P(X_n + Y_n > x, |Y_n - y| \leq \epsilon) + P(|Y_n - y| > \epsilon) \\
\leq P(X_n > x - y - \epsilon) + P(|Y_n - y| > \epsilon).
\]

\[\Rightarrow\]

\[
\limsup_n (1 - F_{X_n+Y_n}(x)) \leq \limsup_n P(X_n > x - y - \epsilon) \\
\leq \limsup_n P(X_n \geq x - y - 2\epsilon) \leq (1 - F_X(x - y - 2\epsilon)).
\]

\[\Rightarrow\ F_X(x - y - 2\epsilon) \leq \liminf_n F_{X_n+Y_n}(x) \leq \limsup_n F_{X_n+Y_n}(x) \leq F_X(x + y + \epsilon).
\]

\[\Rightarrow\]

\[
F_{X+Y}(x-) \leq \liminf_n F_{X_n+Y_n}(x) \leq \limsup_n F_{X_n+Y_n}(x) \leq F_{X+Y}(x).
\]
To complete the proof,

\[ P(\left|(Z_n - z)X_n\right| > \epsilon) \leq P(\left|Z_n - z\right| \geq \epsilon^2) + P(\left|Z_n - z\right| \leq \epsilon^2, |X_n| > \frac{1}{\epsilon}). \]

\[ \Rightarrow \]

\[ \limsup_n P(\left|(Z_n - z)X_n\right| > \epsilon) \leq \limsup_n P(\left|Z_n - z\right| > \epsilon^2) \]

\[ + \limsup_n P(\left|X_n\right| \geq \frac{1}{2\epsilon}) \rightarrow P(\left|X\right| \geq \frac{1}{2\epsilon}). \]

\[ \Rightarrow \text{that } (Z_n - z)X_n \rightarrow_p 0. \]

Clearly \( zX_n \rightarrow_d zX \Rightarrow Z_nX_n \rightarrow_d zX \) from the proof in the first half.

Again, using the first half’s proof, \( Z_nX_n + Y_n \rightarrow_d zX + y. \)
• **Examples**

  - Suppose \(X_n \to_d N(0, 1)\). Then by continuous mapping theorem, \(X_n^2 \to_d \chi_1^2\).

  - This example shows that \(g\) can be discontinuous in Theorem 3.6. Let \(X_n \to_d X\) with \(X \sim N(0, 1)\) and \(g(x) = 1/x\). Although \(g(x)\) is discontinuous at origin, we can still show that \(1/X_n \to_d 1/X\), the reciprocal of the normal distribution. This is because \(P(X = 0) = 0\). However, in Example 3.6 where \(g(x) = I(x > 0)\), it shows that Theorem 3.6 may not be true if \(P(X \in C(g)) < 1\).
The condition $Y_n \rightarrow_p y$, where $y$ is a constant, is necessary. For example, let $X_n = X \sim Uniform(0, 1)$. Let $Y_n = -X$ so $Y_n \rightarrow_d -\tilde{X}$, where $\tilde{X}$ is an independent random variable with the same distribution as $X$. However $X_n + Y_n = 0$ does not converge in distribution to $X - \tilde{X}$. 
Let $X_1, X_2, \ldots$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2 > 0$,

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2),$$

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \xrightarrow{a.s} \sigma^2.$$

$$\Rightarrow \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \xrightarrow{d} \frac{1}{\sigma} N(0, \sigma^2) \approx N(0, 1).$$

$$\Rightarrow \text{in large sample, } t_{n-1} \text{ can be approximated by a standard normal distribution.}$$
Representation of Weak Convergence
Theorem 3.8 (Skorohod’s Representation Theorem) Let \( \{X_n\} \) and \( X \) be random variables in a probability space \((\Omega, \mathcal{A}, P)\) and \( X_n \xrightarrow{d} X \). Then there exists another probability space \((\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})\) and a sequence of random variables \(\tilde{X}_n\) and \(\tilde{X}\) defined on this space such that \(\tilde{X}_n\) and \(X_n\) have the same distributions, \(\tilde{X}\) and \(X\) have the same distributions, and moreover, \(\tilde{X}_n \xrightarrow{a.s.} \tilde{X}\).
• Quantile function

\[ F^{-1}(p) = \inf\{x : F(x) \geq p\}. \]

**Proposition 3.1**

(a) \( F^{-1} \) is left-continuous.

(b) If \( X \) has continuous distribution function \( F \), then \( F(X) \sim Uniform(0, 1) \).

(c) Let \( \xi \sim Uniform(0, 1) \) and let \( X = F^{-1}(\xi) \). Then for all \( x \), \( \{X \leq x\} = \{\xi \leq F(x)\} \). Thus, \( X \) has distribution function \( F \).
Proof

(a) Clearly, $F^{-1}$ is nondecreasing. Suppose $p_n$ increases to $p$ then $F^{-1}(p_n)$ increases to some $y \leq F^{-1}(p)$. Then $F(y) \geq p_n$ so $F(y) \geq p$. $\Rightarrow F^{-1}(p) \leq y \Rightarrow y = F^{-1}(p)$.

(b) $\{X \leq x\} \subset \{F(X) \leq F(x)\} \Rightarrow F(x) \leq P(F(X) \leq F(x))$.
$\{F(X) \leq F(x) - \epsilon\} \subset \{X \leq x\} \Rightarrow P(F(X) \leq F(x) - \epsilon) \leq F(x) \Rightarrow P(F(X) \leq F(x) - \epsilon) \leq F(x)$.

Then if $X$ is continuous, $P(F(X) \leq F(x)) = F(x)$.

(c) $P(X \leq x) = P(F^{-1}(\xi) \leq x) = P(\xi \leq F(x)) = F(x)$. 
Proof

Let $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ be $([0, 1], \mathcal{B} \cap [0, 1], \lambda)$. Define $\tilde{X}_n = F^{-1}_n(\xi)$, $\tilde{X} = F^{-1}(\xi)$, where $\xi \sim Uniform(0, 1)$. $\tilde{X}_n$ has a distribution $F_n$ which is the same as $X_n$.

For any $t \in (0, 1)$ such that there is at most one value $x$ such that $F(x) = t$ (it is easy to see $t$ is the continuous point of $F^{-1}$),

$\Rightarrow$ for any $z < x$, $F(z) < t$
$\Rightarrow$ when $n$ is large, $F_n(z) < t$ so $F^{-1}_n(t) \geq z$.
$\Rightarrow$ $\lim \inf F^{-1}_n(t) \geq z \Rightarrow \lim \inf F^{-1}_n(t) \geq x = F^{-1}(t)$.

From $F(x + \epsilon) > t$, $F_n(x + \epsilon) > t$ so $F^{-1}_n(t) \leq x + \epsilon$.
$\Rightarrow$ $\lim \sup F^{-1}_n(t) \leq x + \epsilon \Rightarrow \lim \sup F^{-1}_n(t) \leq x$.

Thus $F^{-1}_n(t) \to F^{-1}(t)$ for almost every $t \in (0, 1) \Rightarrow \tilde{X}_n \to a.s. \tilde{X}$. 
• Usefulness of representation theorem
  
  – For example, if $X_n \to_d X$ and one wishes to show some function of $X_n$, denote by $g(X_n)$, converges in distribution to $g(X)$:

  – see the diagram in Figure 2.
\[ \bar{x}_n \xrightarrow{d} \bar{x} \quad \Rightarrow \quad g(\bar{x}_n) \xrightarrow{d} g(\bar{x}) \]

\[ \bar{x}_* \xrightarrow{a.s.} \bar{X}_* \quad \Rightarrow \quad g(\bar{X}_*) \xrightarrow{a.s.} g(\bar{X}) \]
Alternative Proof for Slutsky Theorem

First, show \((X_n, Y_n) \to_d (X, y)\).

\[
|\phi_{(X_n,Y_n)}(t_1,t_2) - \phi_{(X,y)}(t_1,t_2)| = |E[e^{it_1X_n}e^{it_2Y_n}] - E[e^{it_1X}e^{it_2y}]| \\
\leq |E[e^{it_1X_n}(e^{it_2Y_n} - e^{it_2y})]| + |e^{it_2y}| |E[e^{it_1X_n}] - E[e^{it_1X}]| \\
\leq E[|e^{it_2Y_n} - e^{it_2y}|] + |E[e^{it_1X_n}] - E[e^{it_1X}]| \to 0.
\]

Thus, \((Z_n, X_n) \to_d (z, X)\). Since \(g(z, x) = zx\) is continuous,

\[
\Rightarrow Z_nX_n \to_d zX.
\]

Since \((Z_nX_n, Y_n) \to_d (zX, y)\) and \(g(x, y) = x + y\) is continuous,

\[
\Rightarrow Z_nX_n + Y_n \to_d zX + y.
\]
Summation of Independent R.V.s
• Some preliminary lemmas

**Proposition 3.2 (Borel-Cantelli Lemma)** For any events $A_n$,

$$\sum_{i=1}^{\infty} P(A_n) < \infty$$

implies $P(A_n, i.o.) = P(\{A_n\} \text{ occurs infinitely often}) = 0$; or equivalently, $P(\cap_{n=1}^{\infty} \cup_{m\geq n} A_m) = 0$.

**Proof**

$$P(A_n, i.o) \leq P(\cup_{m\geq n} A_m) \leq \sum_{m\geq n} P(A_m) \rightarrow 0, \text{ as } n \rightarrow \infty.$$
• One result of the first Borel-cantelli lemma

If for a sequence of random variables, \( \{Z_n\} \), and for any \( \epsilon > 0 \), \( \sum_n P(|Z_n| > \epsilon) < \infty \), then \( |Z_n| > \epsilon \) only occurs finite times.

\[ \Rightarrow Z_n \to_{a.s.} 0. \]
Proposition 3.3 (Second Borel-Cantelli Lemma)

For a sequence of independent events $A_1, A_2, \ldots$, $\sum_{n=1}^{\infty} P(A_n) = \infty$ implies $P(A_n, i.o.) = 1$.

**Proof** Consider the complement of $\{A_n, i.o\}$. 

\[
P(\bigcup_{n=1}^{\infty} \cap_{m\geq n} A_m^c) = \lim_{n} P(\cap_{m\geq n} A_m^c) = \lim_{n} \prod_{m\geq n} (1 - P(A_m))
\]

\[
\leq \limsup_{n} \exp\{- \sum_{m \geq n} P(A_m)\} = 0.
\]
• **Equivalence lemma**

**Proposition 3.4** $X_1, ..., X_n$ are i.i.d with finite mean.
Define $Y_n = X_n I(|X_n| \leq n)$. Then

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty.$$
Proof Since $E[|X_1|] < \infty$,

$$
\sum_{n=1}^{\infty} P(|X| \geq n) = \sum_{n=1}^{\infty} nP(n \leq |X| < (n + 1)) \leq \sum_{n=1}^{\infty} E[|X|] < \infty.
$$

From the Borel-Cantelli Lemma, $P(X_n \neq Y_n, i.o) = 0$.

For almost every $\omega \in \Omega$, when $n$ is large enough, $X_n(\omega) = Y_n(\omega)$. 
Weak Law of Large Numbers
Theorem 3.9 (Weak Law of Large Number) If $X, X_1, ..., X_n$ are i.i.d with mean $\mu$ (so $E[|X|] < \infty$ and $\mu = E[X]$), then $\bar{X}_n \to_p \mu$. 
Proof

Define \( Y_n = X_n I(-n \leq X_n \leq n) \). Let \( \bar{\mu}_n = \frac{\sum_{k=1}^{n} E[Y_k]}{n} \).

\[
P(|\bar{Y}_n - \bar{\mu}_n| \geq \epsilon) \leq \frac{\text{Var}(\bar{Y}_n)}{\epsilon^2} \leq \frac{\sum_{k=1}^{n} \text{Var}(X_k I(|X| \leq k))}{n^2 \epsilon^2}.
\]

\[
\text{Var}(X_k I(|X| \leq k)) \leq E[X_k^2 I(|X| \leq k)]
\]

\[
= E[X_k^2 I(|X| \leq k, |X| \geq \sqrt{k \epsilon^2})] + E[X_k^2 I(|X| \leq k, |X| \leq \sqrt{k \epsilon^2})]
\]

\[
\leq k E[|X_k| I(|X_k| \geq \sqrt{k \epsilon^2})] + k \epsilon^4,
\]

\[
\Rightarrow \text{Var}(X_k I(|X| \leq k)) \leq \frac{\sum_{k=1}^{n} E[|X| I(|X| \geq \sqrt{k \epsilon^2})]}{n^2 \epsilon^2} + \epsilon^2 \frac{n(n+1)}{2n^2}.
\]

\[
\limsup_n P(|\bar{Y}_n - \mu| \geq \epsilon) \leq \frac{\epsilon^2}{2n^2} \Rightarrow \bar{Y}_n - \bar{\mu}_n \rightarrow_p 0.
\]

\[
\bar{\mu}_n \rightarrow \mu \Rightarrow \bar{Y}_n \rightarrow_p \mu. \text{ From Proposition 3.4 and subsequence arguments,}
\]

\[
\bar{X}_{nk} \rightarrow a.s. \mu \Rightarrow X_n \rightarrow_p \mu.
\]
Strong Law of Large Numbers
Theorem 3.10 (Strong Law of Large Number) If $X_1, \ldots, X_n$ are i.i.d with mean $\mu$ then $\bar{X}_n \to_{a.s.} \mu$. 
Proof

Without loss of generality, we assume $X_n \geq 0$ since if this is true, the result also holds for any $X_n$ by $X_n = X_n^+ - X_n^-$. Similar to Theorem 3.9, it is sufficient to show $\bar{Y}_n \to_{a.s.} \mu$, where $Y_n = X_n I(X_n \leq n)$.

Note $E[Y_n] = E[X_1 I(X_1 \leq n)] \to \mu$ so

$$\sum_{k=1}^{n} E[Y_k]/n \to \mu.$$ 

$\Rightarrow$ if we denote $\tilde{S}_n = \sum_{k=1}^{n} (Y_k - E[Y_k])$ and we can show $\tilde{S}_n/n \to_{a.s.} 0$, then the result holds.
\[
\text{Var}(\tilde{S}_n) = \sum_{k=1}^{n} \text{Var}(Y_k) \leq \sum_{k=1}^{n} E[Y_k^2] \leq nE[X_1^2 I(X_1 \leq n)].
\]

By the Chebyshev’s inequality,
\[
P\left( \left| \frac{\tilde{S}_n}{n} \right| > \epsilon \right) \leq \frac{1}{n^2 \epsilon^2} \text{Var}(\tilde{S}_n) \leq \frac{E[X_1^2 I(X_1 \leq n)]}{n \epsilon^2}.
\]

For any \( \alpha > 1 \), let \( u_n = [\alpha^n] \).
\[
\sum_{n=1}^{\infty} P\left( \left| \frac{\tilde{S}_{u_n}}{u_n} \right| > \epsilon \right) \leq \sum_{n=1}^{\infty} \frac{1}{u_n \epsilon^2} E[X_1^2 I(X_1 \leq u_n)] \leq \frac{1}{\epsilon^2} E[X_1^2 \sum_{u_n \geq X_1} \frac{1}{u_n}].
\]

Since for any \( x > 0 \), \( \sum_{u_n \geq x} \{\mu_n\}^{-1} < 2 \sum_{n \geq \log x / \log \alpha} \alpha^{-n} \leq K x^{-1} \)
for some constant \( K \), \( \Rightarrow \)
\[
\sum_{n=1}^{\infty} P\left( \left| \frac{\tilde{S}_{u_n}}{u_n} \right| > \epsilon \right) \leq \frac{K}{\epsilon^2} E[X_1] < \infty,
\]
\[
\Rightarrow \tilde{S}_{u_n} / u_n \rightarrow_{a.s.} 0.
\]
For any $k$, we can find $u_n < k \leq u_{n+1}$. Thus, since $X_1, X_2, \ldots \geq 0$,

$$\frac{\tilde{S}_{u_n}}{u_n} \frac{u_n}{u_{n+1}} \leq \frac{\tilde{S}_k}{k} \leq \frac{\tilde{S}_{u_{n+1}}}{u_{n+1}} \frac{u_{n+1}}{u_n}.$$ 

\[ \Rightarrow \]

$$\frac{\mu}{\alpha} \leq \liminf_k \frac{\tilde{S}_k}{k} \leq \limsup_k \frac{\tilde{S}_k}{k} \leq \mu \alpha.$$ 

Since $\alpha$ is arbitrary number larger than 1, let $\alpha \to 1$ and we obtain $\lim_k \frac{\tilde{S}_k}{k} = \mu$. 
Central Limit Theorems
• Preliminary result of c.f.

**Proposition 3.5** Suppose $E[|X|^m] < \infty$ for some integer $m \geq 0$. Then

$$|\phi_X(t) - \sum_{k=0}^{m} \frac{(it)^k}{k!} E[X^k]|/|t|^m \to 0, \text{ as } t \to 0.$$
Proof

\[ e^{itx} = \sum_{k=1}^{m} \frac{(itx)^k}{k!} + \frac{(itx)^m}{m!} [e^{it\theta x} - 1], \]

where \( \theta \in [0, 1] \).

\[ \Rightarrow \]

\[ |\phi_X(t) - \sum_{k=0}^{m} \frac{(it)^k}{k!} E[X^k]|/|t|^m \leq E[|X|^m |e^{it\theta X} - 1|]/m! \to 0, \]

as \( t \to 0. \)
• Simple versions of CLT

Theorem 3.11 (Central Limit Theorem) If $X_1, ..., X_n$ are i.i.d with mean $\mu$ and variance $\sigma^2$ then
$$\sqrt{n} (\bar{X}_n - \mu) \rightarrow_d N(0, \sigma^2).$$
Proof

Denote $Y_n = \sqrt{n}(\bar{X}_n - \mu)$.

$$\phi_{Y_n}(t) = \left\{ \phi_{X_1-\mu}(t/\sqrt{n}) \right\}^n.$$ 

$\Rightarrow \phi_{X_1-\mu}(t/\sqrt{n}) = 1 - \sigma^2 t^2 / 2n + o(1/n).$

$\Rightarrow$

$$\phi_{Y_n}(t) \rightarrow \exp\left\{-\frac{\sigma^2 t^2}{2}\right\}.$$
Theorem 3.12 (Multivariate Central Limit Theorem) If $X_1, \ldots, X_n$ are i.i.d random vectors in $\mathbb{R}^k$ with mean $\mu$ and covariance $\Sigma = E[(X - \mu)(X - \mu)']$, then $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \Sigma)$.

Proof

Use the Cramér-Wold’s device.
Liapunov CLT

Theorem 3.13 (Liapunov Central Limit Theorem)
Let $X_{n1}, ..., X_{nn}$ be independent random variables with $\mu_{ni} = E[X_{ni}]$ and $\sigma^2_{ni} = Var(X_{ni})$. Let $\mu_n = \sum_{i=1}^{n} \mu_{ni}$, $\sigma^2_n = \sum_{i=1}^{n} \sigma^2_{ni}$. If

$$\sum_{i=1}^{n} \frac{E[|X_{ni} - \mu_{ni}|^3]}{\sigma^3_n} \to 0,$$

then $\sum_{i=1}^{n} (X_{ni} - \mu_{ni})/\sigma_n \to_d N(0, 1)$. 
• Lindeberg-Feller CLT

**Theorem 3.14 (Lindeberg-Fell Central Limit Theorem)** Let $X_{n1}, \ldots, X_{nn}$ be independent random variables with $\mu_{ni} = E[X_{ni}]$ and $\sigma_{ni}^2 = Var(X_{ni})$. Let $\sigma_n^2 = \sum_{i=1}^{n} \sigma_{ni}^2$. Then both $\sum_{i=1}^{n} \frac{(X_{ni} - \mu_{ni})}{\sigma_n} \xrightarrow{d} N(0,1)$ and $\max \{ \frac{\sigma_{ni}^2}{\sigma_n^2} : 1 \leq i \leq n \} \to 0$ if and only if the Lindeberg condition

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{n} E[|X_{ni} - \mu_{ni}|^2 I(|X_{ni} - \mu_{ni}| \geq \epsilon \sigma_n)] \to 0, \text{ for all } \epsilon > 0$$

holds.
Proof of Liapunov CLT using Theorem 3.14

\[
\frac{1}{\sigma^2_n} \sum_{i=1}^{n} E[|X_{nk} - \mu_{nk}|^2 I(|X_{nk} - \mu_{nk}| > \epsilon \sigma_n)] \\
\leq \frac{1}{\epsilon^3 \sigma^3_n} \sum_{k=1}^{n} E[|X_{nk} - \mu_{nk}|^3].
\]
• Examples

- This is one example from a simple linear regression \( X_j = \alpha + \beta z_j + \epsilon_j \) for \( j = 1, 2, \ldots \) where \( z_j \) are known numbers not all equal and the \( \epsilon_j \) are i.i.d with mean zero and variance \( \sigma^2 \).

\[
\hat{\beta}_n = \frac{\sum_{j=1}^{n} X_j (z_j - \bar{z}_n)}{\sum_{j=1}^{n} (z_j - \bar{z}_n)^2} = \beta + \frac{\sum_{j=1}^{n} \epsilon_j (z_j - \bar{z}_n)}{\sum_{j=1}^{n} (z_j - \bar{z}_n)^2}.
\]

Assume

\[
\max_{j \leq n} (z_j - \bar{z}_n)^2/\sum_{j=1}^{n} (z_j - \bar{z}_n)^2 \rightarrow 0.
\]

\[
\Rightarrow \sqrt{n} \sqrt{\frac{\sum_{j=1}^{n} (z_j - \bar{z}_n)^2}{n}} (\hat{\beta}_n - \beta) \rightarrow_d N(0, \sigma^2).
\]
The example is taken from the randomization test for paired comparison. Let \((X_j, Y_j)\) denote the values of \(j\)th pairs with \(X_j\) being the result of the treatment and \(Z_j = X_j - Y_j\). Conditional on \(|Z_j| = z_j\), 
\[Z_j = |Z_j| \text{sgn}(Z_j)\] 
is independent taking values \(\pm |Z_j|\) with probability \(1/2\), when treatment and control have no difference. Conditional on \(z_1, z_2, \ldots\), the randomization \(t\)-test is the \(t\)-statistic \(\sqrt{n - 1} \bar{Z}_n / s_z\) where \(s_z^2\) is \(1/n \sum_{j=1}^n (Z_j - \bar{Z}_n)^2\). When

\[\max_{j \leq n} \frac{z_j^2}{\sum_{j=1}^n z_j^2} \to 0,\]

this statistic has an asymptotic normal distribution \(N(0, 1)\).
Delta Method
Theorem 3.15 (Delta method) For random vector $X$ and $X_n$ in $\mathbb{R}^k$, if there exists two constant $a_n$ and $\mu$ such that $a_n(X_n - \mu) \xrightarrow{d} X$ and $a_n \to \infty$, then for any function $g : \mathbb{R}^k \mapsto \mathbb{R}^l$ such that $g$ has a derivative at $\mu$, denoted by $\nabla g(\mu)$

$$a_n(g(X_n) - g(\mu)) \xrightarrow{d} \nabla g(\mu)X.$$
Proof

By the Skorohod representation, we can construct \( \tilde{X}_n \) and \( \tilde{X} \) such that \( \tilde{X}_n \sim_d X_n \) and \( \tilde{X} \sim_d X \) (\( \sim_d \) means the same distribution) and \( a_n(\tilde{X}_n - \mu) \to_{a.s.} \tilde{X} \).

\[
\Rightarrow \\
a_n(g(\tilde{X}_n) - g(\mu)) \to_{a.s.} \nabla g(\mu) \tilde{X}
\]

\[
\Rightarrow \\
a_n(g(X_n) - g(\mu)) \to_d \nabla g(\mu) X
\]
• Examples

- Let $X_1, X_2, \ldots$ be i.i.d with fourth moment and $s_n^2 = (1/n) \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$. Denote $m_k$ as the $k$th moment of $X_1$ for $k \leq 4$. Note that $s_n^2 = (1/n) \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i/n)^2$ and

$$
\sqrt{n} \left[ \left( \frac{\bar{X}_n}{(1/n) \sum_{i=1}^{n} X_i^2} \right) - \left( \begin{array}{c} m_1 \\ m_2 \end{array} \right) \right] 
\rightarrow_d \mathcal{N} \left( 0, \begin{pmatrix} m_2 - m_1 & m_3 - m_1 m_2 \\ m_3 - m_1 m_2 & m_4 - m_2^2 \end{pmatrix} \right),
$$

the Delta method with $g(x, y) = y - x^2$

$$
\Rightarrow \sqrt{n}(s_n^2 - \text{Var}(X_1)) \rightarrow_d \mathcal{N}(0, m_4 - (m_2 - m_1^2)^2).
$$
Let \((X_1, Y_1), (X_2, Y_2), \ldots\) be i.i.d bivariate samples with finite fourth moment. One estimate of the correlation among \(X\) and \(Y\) is

\[
\hat{\rho}_n = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}},
\]

where \(s_{xy} = (1/n) \sum_{i=1}^{n} (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)\), \(s_x^2 = (1/n) \sum_{i=1}^{n} (X_i - \bar{X}_n)^2\) and \(s_y^2 = (1/n) \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2\). To derive the large sample distribution of \(\hat{\rho}_n\), first obtain the large sample distribution of \((s_{xy}, s_x^2, s_y^2)\) using the Delta method then further apply the Delta method with

\[
g(x, y, z) = x/\sqrt{y z}.
\]
The example is taken from the Pearson’s Chi-square statistic. Suppose that one subject falls into $K$ categories with probabilities $p_1, \ldots, p_K$, where $p_1 + \ldots + p_K = 1$. The Pearson’s statistic is defined as

$$\chi^2 = n \sum_{k=1}^{K} \left( \frac{n_k}{n} - p_k \right)^2 / p_k,$$

which can be treated as

$$\sum (\text{observed count} - \text{expected count})^2 / \text{expected count}.$$

Note $\sqrt{n}(n_1/n - p_1, \ldots, n_K/n - p_K)$ has an asymptotic multivariate normal distribution. Then we can apply the Delta method to $g(x_1, \ldots, x_K) = \sum_{i=1}^{K} x_i^2.$
CHAPTER 3 LARGE SAMPLE THEORY

U-statistics
Definition 3.6 A \( U \)-statistics associated with \( \tilde{h}(x_1, \ldots, x_r) \) is defined as

\[
U_n = \frac{1}{r! \left( \begin{array}{c} n \\ r \end{array} \right)} \sum_{\beta} \tilde{h}(X_{\beta_1}, \ldots, X_{\beta_r}),
\]

where the sum is taken over the set of all unordered subsets \( \beta \) of \( r \) different integers chosen from \( \{1, \ldots, n\} \).
• Examples

- One simple example is \( \tilde{h}(x, y) = xy \). Then
  \[ U_n = (n(n - 1))^{-1} \sum_{i \neq j} X_i X_j. \]

- \( U_n = E[\tilde{h}(X_1, \ldots, X_r) | X_{(1)}, \ldots, X_{(n)}] \).

- \( U_n \) is the summation of non-independent random variables.

- If define \( h(x_1, \ldots, x_r) \) as \( (r!)^{-1} \sum (\tilde{x}_1, \ldots, \tilde{x}_r) \tilde{h}(\tilde{x}_1, \ldots, \tilde{x}_r) \),
  then \( h(x_1, \ldots, x_r) \) is permutation-symmetric
  \[ U_n = \frac{1}{\binom{n}{r}} \sum_{\beta_1 < \ldots < \beta_r} h(\beta_1, \ldots, \beta_r). \]

- \( h \) is called the kernel of the U-statistic \( U_n \).
• CLT for U-statistics

**Theorem 3.16** Let \( \mu = E[h(X_1, \ldots, X_r)] \). If \( E[h(X_1, \ldots, X_r)^2] < \infty \), then

\[
\sqrt{n}(U_n - \mu) - \sqrt{n} \sum_{i=1}^{n} E[U_n - \mu | X_i] \rightarrow_p 0.
\]

Consequently, \( \sqrt{n}(U_n - \mu) \) is asymptotically normal with mean zero and variance \( r^2 \sigma^2 \), where, with \( X_1, \ldots, X_r, \tilde{X}_1, \ldots, \tilde{X}_r \) i.i.d variables,

\[
\sigma^2 = Cov(h(X_1, X_2, \ldots, X_r), h(X_1, \tilde{X}_2, \ldots, \tilde{X}_r)).
\]
- Some preparation

- Linear space of r.v.: let $\mathcal{S}$ be a linear space of random variables with finite second moments that contain the constants; i.e., $1 \in \mathcal{S}$ and for any $X, Y \in \mathcal{S}$, $aX + bY \in \mathcal{S}_n$ where $a$ and $b$ are constants.

- Projection: for random variable $T$, a random variable $S$ is called the projection of $T$ on $\mathcal{S}$ if $E[(T - S)^2]$ minimizes $E[(T - \tilde{S})^2]$, $\tilde{S} \in \mathcal{S}$. 
Proposition 3.7  Let $\mathcal{S}$ be a linear space of random variables with finite second moments. Then $\mathcal{S}$ is the projection of $T$ on $\mathcal{S}$ if and only if $S \in \mathcal{S}$ and for any $\tilde{S} \in \mathcal{S}$, $E[(T - S)\tilde{S}] = 0$. Every two projections of $T$ onto $\mathcal{S}$ are almost surely equal. If the linear space $\mathcal{S}$ contains the constant variable, then $E[T] = E[S]$ and $Cov(T - S, \tilde{S}) = 0$ for every $\tilde{S} \in \mathcal{S}$.
Proof For any $S$ and $\tilde{S}$ in $S$,

$$E[(T - \tilde{S})^2] = E[(T - S)^2] + 2E[(T - S)\tilde{S}] + E[(S - \tilde{S})^2].$$

⇒ if $S$ satisfies that $E[(T - S)\tilde{S}] = 0$, then $E[(T - \tilde{S})^2] \geq E[(T - S)^2]$. ⇒ $S$ is the projection of $T$ on $S$.

If $S$ is the projection, for any constant $\alpha$, $E[(T - S - \alpha\tilde{S})^2]$ is minimized at $\alpha = 0$. Calculate the derivative at $\alpha = 0$ ⇒ $E[(T - S)\tilde{S}] = 0$.

If $T$ has two projections $S_1$ and $S_2$, ⇒ $E[(S_1 - S_2)^2] = 0$. Thus, $S_1 = S_2$, a.s. If the linear space $S$ contains the constant variable, choose $\tilde{S} = 1$ ⇒ $0 = E[(T - S)\tilde{S}] = E[T] - E[S]$. Clearly, $Cov(T - S, \tilde{S}) = E[(T - S)\tilde{S}] = 0$. 
• Equivalence with projection

Proposition 3.8 Let \( S_n \) be linear space of random variables with finite second moments that contain the constants. Let \( T_n \) be random variables with projections \( S_n \) on to \( S_n \). If \( \text{Var}(T_n)/\text{Var}(S_n) \to 1 \) then

\[
Z_n \equiv \frac{T_n - E[T_n]}{\sqrt{\text{Var}(T_n)}} - \frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} \to_p 0.
\]
Proof. $E[Z_n] = 0$. Note that

$$Var(Z_n) = 2 - 2 \frac{Cov(T_n, S_n)}{\sqrt{Var(T_n)Var(S_n)}}.$$ 

Since $S_n$ is the projection of $T_n$,

$Cov(T_n, S_n) = Cov(T_n - S_n, S_n) + Var(S_n) = Var(S_n)$. We have

$$Var(Z_n) = 2(1 - \sqrt{\frac{Var(S_n)}{Var(T_n)}}) \to 0.$$ 

By the Markov’s inequality, we conclude that $Z_n \to_p 0$. 
• **Conclusion**

  - if $S_n$ is the summation of i.i.d random variables such that $\frac{(S_n - E[S_n])}{\sqrt{\text{Var}(S_n)}} \to_d N(0, \sigma^2)$, so is $\frac{(T_n - E[T_n])}{\sqrt{\text{Var}(T_n)}}$. The limit distribution of U-statistics is derived using this lemma.
• Proof of CLT for U-statistics

**Proof**

Let \( \tilde{X}_1, ..., \tilde{X}_r \) be random variables with the same distribution as \( X_1 \) and they are independent of \( X_1, ..., X_n \). Denote \( \tilde{U}_n \) by

\[
\sum_{i=1}^n E[U - \mu | X_i].
\]

We show that \( \tilde{U}_n \) is the projection of \( U_n \) on the linear space

\[
S_n = \{ g_1(X_1) + ... + g_n(X_n) : E[g_k(X_k)^2] < \infty, k = 1, ..., n \},
\]

which contains the constant variables. Clearly, \( \tilde{U}_n \in S_n \). For any \( g_k(X_k) \in S_n \),

\[
E[(U_n - \tilde{U}_n)g_k(X_k)] = E[E[U_n - \tilde{U}_n | X_k]g_k(X_k)] = 0.
\]
\[
\tilde{U}_n = \sum_{i=1}^{n} \frac{\binom{n-1}{r-1}}{\binom{n}{r}} E[h(\tilde{X}_1, \ldots, \tilde{X}_{r-1}, X_i) - \mu|X_i] \\
= \frac{r}{n} \sum_{i=1}^{n} E[h(\tilde{X}_1, \ldots, \tilde{X}_{r-1}, X_i) - \mu|X_i].
\]

\[
\Rightarrow \quad Var(\tilde{U}_n) = \frac{r^2}{n^2} \sum_{i=1}^{n} E[(E[h(\tilde{X}_1, \ldots, \tilde{X}_{r-1}, X_i) - \mu|X_i)]^2]
\]

\[
= \frac{r^2}{n} Cov(E[h(\tilde{X}_1, \ldots, \tilde{X}_{r-1}, X_1)|X_1], E[h(\tilde{X}_1, \ldots, \tilde{X}_{r-1}, X_1)|X_1])
\]

\[
= \frac{r^2}{n} Cov(h(X_1, \tilde{X}_2, \ldots, \tilde{X}_r), h(X_1, X_2, \ldots, X_r)) = \frac{r^2\sigma^2}{n}.
\]
Furthermore,

\[
Var(U_n) = \binom{n}{r}^2 \sum_{\beta} \sum_{\beta'} Cov(h(X_{\beta_1}, \ldots, X_{\beta_r}), h(X_{\beta'_1}, \ldots, X_{\beta'_r}))
\]

\[
= \binom{n}{r}^2 \sum_{k=1}^{r} \sum_{\text{\beta and \beta' share } k \text{ components}} Cov(h(X_1, X_2, \ldots, X_k, X_{k+1}, \ldots, X_r), h(X_1, X_2, \ldots, X_k, \tilde{X}_{k+1}, \ldots, \tilde{X}_r)).
\]

\[
\Rightarrow Var(U_n) = \sum_{k=1}^{r} \frac{r!}{k!(r-k)!} \frac{(n-r)(n-r+1)\cdots(n-2r+k+1)}{n(n-1)\cdots(n-r+1)} c_k.
\]

\[
\Rightarrow Var(U_n) = \frac{r^2}{n} Cov(h(X_1, X_2, \ldots, X_r), h(X_1, \tilde{X}_2, \ldots, \tilde{X}_r)) + O\left(\frac{1}{n^2}\right).
\]

\[
\Rightarrow Var(U_n) / Var(\tilde{U}_n) \to 1.
\]

\[
\Rightarrow \frac{U_n - \mu}{\sqrt{Var(U_n)}} - \frac{\tilde{U}_n}{\sqrt{Var(\tilde{U}_n)}} \to_p 0.
\]
• Example

- In a bivariate i.i.d sample \((X_1, Y_1), (X_2, Y_2), \ldots\), one statistic of measuring the agreement is called *Kendall’s \(\tau\)-statistic*

  \[
  \hat{\tau} = \frac{4}{n(n - 1)} \sum \sum_{i<j} I \{(Y_j - Y_i)(X_j - X_i) > 0\} - 1.
  \]

  \[
  \Rightarrow \hat{\tau} + 1 \text{ is a U-statistic of order 2 with the kernel}
  \]

  \[
  2I \{(y_2 - y_1)(x_2 - x_1) > 0\}.
  \]

  \[
  \Rightarrow \sqrt{n}(\hat{\tau}_n + 1 - 2P((Y_2 - Y_1)(X_2 - X_1) > 0)) \text{ has an asymptotic normal distribution with mean zero.}
  \]
Rank Statistics
• Some definitions

- $X_1 \leq X_2 \leq \ldots \leq X_n$ is called order statistics

- The rank statistics, denoted by $R_1, \ldots, R_n$ are the ranks of $X_i$ among $X_1, \ldots, X_n$. Thus, if all the $X$’s are different, $X_i = X_{(R_i)}$.

- When there are ties, $R_i$ is defined as the average of all indices such that $X_i = X_{(j)}$ (sometimes called midrank).

- Only consider the case that $X$’s have continuous densities.
• More definitions

  – a rank statistic is any function of the ranks

  – a linear rank statistic is a rank statistic of the special form \( \sum_{i=1}^{n} a(i, R_i) \) for a given matrix \((a(i, j))_{n \times n}\).

  – if \( a(i, j) = c_i a_j \), then such statistic with form \( \sum_{i=1}^{n} c_i a_{R_i} \) is called simple linear rank statistic: \( c \) and \( a \)'s are called the coefficients and scores.
• **Examples**

  - In two independent sample $X_1, ..., X_n$ and $Y_1, ..., Y_m$, a Wilcoxon statistic is defined as the summation of all the ranks of the second sample in the pooled data $X_1, ..., X_n, Y_1, ..., Y_m$, i.e.,

    $$W_n = \sum_{i=n+1}^{n+m} R_i.$$ 

    Other choices for rank statistics: for instance, the van der Waerden statistic $\sum_{i=n+1}^{n+m} \Phi^{-1}(R_i)$. 

• Properties of rank statistics

Proposition 3.9 Let \( X_1, \ldots, X_n \) be a random sample from continuous distribution function \( F \) with density \( f \). Then

1. the vectors \((X_{(1)}, \ldots, X_{(n)})\) and \((R_1, \ldots, R_n)\) are independent;

2. the vector \((X_{(1)}, \ldots, X_{(n)})\) has density \( n! \prod_{i=1}^{n} f(x_i) \) on the set \( x_1 < \ldots < x_n \);

3. the variable \( X_{(i)} \) has density

\[
\binom{n-1}{i-1} F(x)^{i-1} (1 - F(x))^{n-i} f(x); \text{ for } F \text{ the uniform distribution on } [0, 1], \text{ it has mean } i/(n+1) \text{ and variance } i(n-i+1)/[(n+1)^2(n+2)];
\]
4. the vector \((R_1, \ldots, R_n)\) is uniformly distributed on the set of all \(n!\) permutations of 1, 2, \ldots, \(n\);

5. for any statistic \(T\) and permutation \(r = (r_1, \ldots, r_n)\) of 1, 2, \ldots, \(n\),

\[
E[T(X_1, \ldots, X_n)|(R_1, \ldots, R_n) = r] = E[T(X_{(r_1)}, \ldots, X_{(r_n)})];
\]

6. for any simple linear rank statistic \(T = \sum_{i=1}^{n} c_i a_{R_i}\),

\[
E[T] = n\bar{c}_n\bar{a}_n, \quad Var(T) = \frac{1}{n-1} \sum_{i=1}^{n} (c_i - \bar{c}_n)^2 \sum_{i=1}^{n} (a_i - \bar{a}_n)^2.
\]
• CLT of rank statistics

Theorem 3.17 Let \( T_n = \sum_{i=1}^{n} c_i a_{R_i} \) such that

\[
\max_{i \leq n} \frac{|a_i - \bar{a}_n|}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a}_n)^2}} \rightarrow 0, \quad \max_{i \leq n} \frac{|c_i - \bar{c}_n|}{\sqrt{\sum_{i=1}^{n} (c_i - \bar{c}_n)^2}} \rightarrow 0.
\]

Then \( (T_n - E[T_n]) / \sqrt{Var(T_n)} \rightarrow_d N(0, 1) \) if and only if for every \( \epsilon > 0 \),

\[
\sum_{(i,j)} I \left\{ \sqrt{n} \frac{|a_i - \bar{a}_n| |c_i - \bar{c}_n|}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a}_n)^2} \sum_{i=1}^{n} (c_i - \bar{c}_n)^2} > \epsilon \right\} \times \frac{|a_i - \bar{a}_n|^2 |c_i - \bar{c}_n|^2}{\sum_{i=1}^{n} (a_i - \bar{a}_n)^2 \sum_{i=1}^{n} (c_i - \bar{c}_n)^2} \rightarrow 0.
\]
• More on rank statistics

  – a simple linear *signed rank statistic*

  \[ \sum_{i=1}^{n} a_{R_i} \text{sign}(X_i), \]

  where \( R_1^+, \ldots, R_n^+ \), *absolute rank*, are the ranks of \( |X_1|, \ldots, |X_n| \).

  – In a bivariate sample \((X_1, Y_1), \ldots, (X_n, Y_n)\),

  \[ \sum_{i=1}^{n} a_{R_i} b_{S_i} \text{ where } (R_1, \ldots, R_n) \text{ and } (S_1, \ldots, S_n) \text{ are respective ranks of } (X_1, \ldots, X_n) \text{ and } (Y_1, \ldots, Y_n). \]
Martingales
**Definition 3.7** Let \( \{Y_n\} \) be a sequence of random variables and \( \mathcal{F}_n \) be sequence of \( \sigma \)-fields such that \( \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \). Suppose \( E[|Y_n|] < \infty \). Then the pairs \( \{(Y_n, \mathcal{F}_n)\} \) is called a ***martingale*** if

\[
E[Y_n|\mathcal{F}_{n-1}] = Y_{n-1}, \quad \text{a.s.}
\]

\( \{(Y_n, \mathcal{F}_n)\} \) is a ***submartingale*** if

\[
E[Y_n|\mathcal{F}_{n-1}] \geq Y_{n-1}, \quad \text{a.s.}
\]

\( \{(Y_n, \mathcal{F}_n)\} \) is a ***supmartingale*** if

\[
E[Y_n|\mathcal{F}_{n-1}] \leq Y_{n-1}, \quad \text{a.s.}
\]
• Some notes on definition

- \( Y_1, ..., Y_n \) are measurable in \( \mathcal{F}_n \). Sometimes, we say \( Y_n \) is adapted to \( \mathcal{F}_n \).

- One simple example: \( Y_n = X_1 + ... + X_n \), where \( X_1, X_2, ... \) are i.i.d with mean zero, and \( \mathcal{F}_n \) is the \( \sigma \)-filed generated by \( X_1, ..., X_n \).
Convex function of martingales

**Proposition 3.9** Let \( \{(Y_n, \mathcal{F}_n)\} \) be a martingale. For any measurable and convex function \( \phi \), \( \{\phi(Y_n), \mathcal{F}_n\} \) is a submartingale.
Proof Clearly, $\phi(Y_n)$ is adapted to $\mathcal{F}_n$. It is sufficient to show

$$E[\phi(Y_n)|\mathcal{F}_{n-1}] \geq \phi(Y_{n-1}).$$

This follows from the well-known Jensen's inequality: for any convex function $\phi$,

$$E[\phi(Y_n)|\mathcal{F}_{n-1}] \geq \phi(E[Y_n|\mathcal{F}_{n-1}]) = \phi(Y_{n-1}).$$
Jensen’s inequality

Proposition 3.10 For any random variable $X$ and any convex measurable function $\phi$,

$$E[\phi(X)] \geq \phi(E[X]).$$
Proof

Claim that for any $x_0$, there exists a constant $k_0$ such that for any $x$, $\phi(x) \geq \phi(x_0) + k_0(x - x_0)$.

By the convexity, for any $x' < y' < x_0 < y < x$,

$$\frac{\phi(x_0) - \phi(x')}{x_0 - x'} \leq \frac{\phi(y) - \phi(x_0)}{y - x_0} \leq \frac{\phi(x) - \phi(x_0)}{x - x_0}.$$

Thus, $\frac{\phi(x) - \phi(x_0)}{x - x_0}$ is bounded and decreasing as $x$ decreases to $x_0$. Let the limit be $k_0^+ \Rightarrow \frac{\phi(x) - \phi(x_0)}{x - x_0} \geq k_0^+ \Rightarrow \phi(x) \geq k_0^+(x - x_0) + \phi(x_0)$. 
Similarly,

\[
\frac{\phi(x') - \phi(x_0)}{x' - x_0} \leq \frac{\phi(y') - \phi(x_0)}{y' - x_0} \leq \frac{\phi(x) - \phi(x_0)}{x - x_0}.
\]

Then \(\frac{\phi(x') - \phi(x_0)}{x' - x_0}\) is increasing and bounded as \(x'\) increases to \(x_0\). Let the limit be \(k_0^- \Rightarrow \phi(x') \geq k_0^- (x' - x_0) + \phi(x_0)\).

Clearly, \(k_0^+ \geq k_0^-\). Combining those two inequalities,

\[
\phi(x) \geq \phi(x_0) + k_0 (x - x_0)
\]

for \(k_0 = (k_0^+ + k_0^-)/2\).

Choose \(x_0 = E[X]\) then \(\phi(X) \geq \phi(E[X]) + k_0 (X - E[X])\).
• Decomposition of submartingale

\[ Y_n = (Y_n - E[Y_n|\mathcal{F}_{n-1}]) + E[Y_n|\mathcal{F}_{n-1}] \]

– any submartingale can be written as the summation of a martingale and a random variable predictable in \( \mathcal{F}_{n-1} \).
• Convergence of martingales

**Theorem 3.18 (Martingale Convergence Theorem)**

Let \( \{(X_n, \mathcal{F}_n)\} \) be submartingale. If \( K = \sup_n E[|X_n|] < \infty \), then \( X_n \to_{a.s.} X \) where \( X \) is a random variable satisfying \( E[|X|] \leq K \).
Corollary 3.1 If $\mathcal{F}_n$ is increasing $\sigma$-field and denote $\mathcal{F}_\infty$ as the $\sigma$-field generated by $\bigcup_{n=1}^\infty \mathcal{F}_n$, then for any random variable $Z$ with $E[|Z|] < \infty$, it holds

$$E[Z|\mathcal{F}_n] \to_{a.s.} E[Z|\mathcal{F}_\infty].$$
• CLT for martingale

Theorem 3.19 (Martingale Central Limit Theorem) Let \((Y_{n1}, \mathcal{F}_{n1}), (Y_{n2}, \mathcal{F}_{n2}), \ldots\) be a martingale. Define \(X_{nk} = Y_{nk} - Y_{n,k-1}\) with \(Y_{n0} = 0\) thus \(Y_{nk} = X_{n1} + \ldots + X_{nk}\). Suppose that

\[
\sum_k E[X^2_{nk}|\mathcal{F}_{n,k-1}] \rightarrow_p \sigma^2
\]

where \(\sigma\) is a positive constant and that

\[
\sum_k E[X^2_{nk} I(|X_{nk}| \geq \epsilon)|\mathcal{F}_{n,k-1}] \rightarrow_p 0
\]

for each \(\epsilon > 0\). Then

\[
\sum_k X_{nk} \rightarrow_d N(0, \sigma^2).
\]
Some Notation
\begin{itemize}
  \item \( o_p(1) \) and \( O_p(1) \)
    \begin{itemize}
      \item \( X_n = o_p(1) \) denotes that \( X_n \) converges in probability to zero,
      \item \( X_n = O_p(1) \) denotes that \( X_n \) is bounded in probability; i.e.,
        \[
        \lim_{M \to \infty} \limsup_{n} P(|X_n| \geq M) = 0.
        \]
      \item for a sequence of random variable \( \{r_n\} \), \( X_n = o_p(r_n) \)
        means that \( |X_n|/r_n \to_p 0 \) and \( X_n = O_p(r_n) \) means that \( |X_n|/r_n \) is bounded in probability.
    \end{itemize}
\end{itemize}
• Algebra in $o_p(1)$ and $O_p(1)$

\[ o_p(1) + o_p(1) = o_p(1) \]
\[ O_p(1) + O_p(1) = O_p(1), \]
\[ O_p(1) o_p(1) = o_p(1) \ (1 + o_p(1))^{-1} = 1 + o_p(1) \]
\[ o_p(R_n) = R_n o_p(1) \quad O_p(R_n) = R_n O_p(1) \]
\[ o_p(O_p(1)) = o_p(1). \]

If a real function $R(\cdot)$ satisfies that $R(h) = o(|h|^p)$ as $h \to 0$, \(\Rightarrow\) $R(X_n) = o_p(|X_n|^p)$.

If $R(h) = O(|h|^p)$ as $h \to 0$, \(\Rightarrow\) $R(X_n) = O_p(|X_n|^p)$. 
READING MATERIALS: Lehmann and Casella, Section 1.8, Ferguson, Part 1, Part 2, Part 3 12-15