Section 25

Convergence of Distributions

Po-Ning Chen, Professor
Institute of Communications Engineering
National Chiao Tung University
Hsin Chu, Taiwan 30010, R.O.C.
Convergence of distributions

Definition (convergence in distribution) Distribution function $F_n(\cdot)$ is said to converge weakly to distribution function $F$, if

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

for every continuity point $x$ of $F(\cdot)$.

In notations, we write $F_n \Rightarrow F$.

**Why** the definition only require convergence at continuity point?

**Answer:** If not, there will be quite a few distributions do not converge.
Convergence of distributions

**Example 14.4** Let $X_1, X_2, \ldots$ be i.i.d.

\[
\Pr[X_n = 1] = \Pr[X_n = -1] = \frac{1}{2}.
\]

Then

\[
F_{(X_1 + \cdots + X_n)/n}(x) \Rightarrow \Delta(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1, & \text{if } x \geq 0. \end{cases}
\]

By symmetry,

\[
\Pr \left[ \frac{X_1 + \cdots + X_n}{n} > 0 \right] = \Pr \left[ \frac{X_1 + \cdots + X_n}{n} < 0 \right] = 1 - \Pr \left[ \frac{X_1 + \cdots + X_n}{n} = 0 \right].
\]

Accordingly,

\[
F_{(X_1 + \cdots + X_n)/n}(0) = \Pr \left[ \frac{X_1 + \cdots + X_n}{n} \leq 0 \right] = \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \left( \frac{2}{e} \right)^n < \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \left( 1 + \frac{1}{12(n-1)} \right).
\]

\[
\Rightarrow 2^{-2k} \left( \frac{2k}{k} \right) \leq \frac{1}{\sqrt{k\pi}} \left( 1 + \frac{1}{24k-1} \right)
\]

\[
\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \left( \frac{2}{e} \right)^n \rightarrow \frac{1}{2} \neq \Delta(0) = 1.
\]
Vague convergence

**Definition (vague convergence)** A sequence of measures \( \{\mu_n\}_{n=1}^{\infty} \) is said to converge vaguely to measure \( \mu \), if

\[
\mu_n(a, b) \to \mu(a, b),
\]

for every finite interval for which \( \mu\{a\} = \mu\{b\} = 0 \).

In notations, we write \( \mu_n \xrightarrow{v} \mu \).

**Observation** If \( \mu_n \) and \( \mu \) are both probability measure, then \( \mu_n \xrightarrow{v} \mu \) is equivalent to \( F_n \Rightarrow F \), where \( F_n(x) = \mu_n(-\infty, x] \) and \( F(x) = \mu(-\infty, x] \).

**Example 25.1 (converge vaguely \( \not\Rightarrow \) converge in distribution)**
\[
F_n(x) = I_{[n, \infty)}.
\]
Then \( F_n \xrightarrow{v} F \equiv 0 \)
but we cannot write \( F_n \Rightarrow F \), since \( \lim_{x \uparrow \infty} F(x) = 0 \).
Vague convergence

The condition “for every finite interval for which \( \mu\{a\} = \mu\{b\} = 0 \)” is essential for vague convergence.

**Example 25.3**

\( \mu_n \) places mass \( 1/n \) at each point \( k/n \) for \( k = 0, 1, \ldots, n - 1 \).

Then \( F_n(x) = \mu_n(-\infty, x] = \begin{cases} 0, & \text{if } x < 0; \\ \frac{\lfloor nx \rfloor + 1}{n}, & \text{if } 0 \leq x < 1; \\ 1, & \text{if } x \geq 1. \end{cases} \)

Accordingly,

\( F_n(x) \Rightarrow F(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \leq x < 1; \\ 1, & \text{if } x \geq 1. \end{cases} \)

So \( \mu_n \Rightarrow \mu \), where \( \mu \) is Lebesgue measure confined in \([0, 1)\).

Let \( \mathbb{Q} \) be the set of all rational numbers.

Then \( \mu_n(\mathbb{Q}) = 1 \) for every \( n \).

But \( \mu(\mathbb{Q}) = 0 \).

However, this does not violate \( \mu_n \Rightarrow \mu \).
Poisson approximation to the binomial

Theorem 23.2 $Z_{n,1}, Z_{n,2}, \ldots, Z_{n,r_n}$ are independent random variables. \Pr[Z_{n,k} = 1] = p_{n,k}$ and $\Pr[Z_{n,k} = 0] = 1 - p_{n,k}$. Then

\begin{align*}
(i) \quad & \lim_{n \to \infty} \sum_{k=1}^{r_n} p_{n,k} = \lambda > 0 \quad \Rightarrow \quad \Pr \left[ \sum_{k=1}^{r_n} Z_{n,k} = i \right] \to e^{-\lambda} \frac{\lambda^i}{i!} \quad \text{for } i = 0, 1, 2, \ldots \\
(ii) \quad & \lim_{n \to \infty} \max_{1 \leq k \leq r_n} p_{n,k} = 0
\end{align*}

and

\begin{align*}
(i) \quad & \lim_{n \to \infty} \sum_{k=1}^{r_n} p_{n,k} = 0 \quad \Rightarrow \quad \Pr \left[ \sum_{k=1}^{r_n} Z_{n,k} = i \right] \to \begin{cases} 1, & \text{if } i = 0; \\ 0, & \text{if } i = 1, 2, \ldots \end{cases} \\
(ii) \quad & \lim_{n \to \infty} \max_{1 \leq k \leq r_n} p_{n,k} = 0
\end{align*}

If $r_n = n$, then Theorem 23.2 reduces to Poisson approximation to the binomial.
Poisson approximation to the binomial

Example 25.2 (Poisson approximation to the binomial)
Take $p_{n,k} = \lambda/n$.

$$
\mu_n\{k\} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}
$$
for $0 \leq k \leq n$.

Then

$$
\mu_n \Rightarrow \text{Poisson}(\lambda).
$$

Example 25.4 $\mu_n\{x_n\} = 1$ and $\mu\{x\} = 1$.
Then

$$
\mu_n \Rightarrow \mu \text{ if, and only if } x_n \xrightarrow{n \to \infty} x.
$$

If $x_n > x$ for every $n$, then (at the discontinuity point $x$ of $F(\cdot)$)

$$
F_n(x) = 0 \text{ for every } n, \text{ but } F(x) = 1.
$$
Uniform distribution modulo 1

Fix a sequence of real numbers $x_1, x_2, \ldots$.

Define a counting probability measure as:

$$
\mu_n(A) = \frac{\text{number of } " (x_n - \lfloor x_n \rfloor ) \in A " \text{ in } x_1, \ldots, x_n }{n}.
$$

(If $x_i - \lfloor x_i \rfloor = x_j - \lfloor x_j \rfloor \in A$ for some $i \neq j$, then their probability masses add to $\mu_n(A)$.)

**Definition (Uniformly distributed modulo 1 for a deterministic sequence)** If $\mu_n$, defined above, satisfies $\mu_n \Rightarrow \mu$, where $\mu$ is a Lebesgue measure restricted to the unit interval, then $x_1, x_2, \ldots$ is said to uniformly distributed modulo 1.

**Theorem 25.1** For any irrational number $\theta$,

$$
\theta, 2\theta, 3\theta, 4\theta, \ldots,
$$

is uniformly distributed modulo 1.

**Proof:** Will be given in Section 26.

• It forms the basis for numerically generating a Lebesgue measure restricted to the unit interval.
**Definition** Let random variables $X_n$ and $X$ have distributions $F_n(\cdot)$ and $F(\cdot)$, respectively. Then $X_n$ is said to *converge in distribution* or *converge in law* to $X$, if

$$F_n \Rightarrow F,$$

or equivalently,

$$\lim_{n \to \infty} \Pr[X_n \leq x] = \Pr[X \leq x]$$

for every $x$ such that $\Pr[X = x] = 0$. 
Convergence in distribution

Example 25.5 (also, Example 14.1) Let $X_1, X_2, \ldots$ be i.i.d. with

$$\Pr[X_n \geq x] = \begin{cases} e^{-\alpha x}, & \text{if } x \geq 0; \\ 1, & \text{for } x < 0. \end{cases}$$

Then

$$\Pr \left[ \max\{X_1, X_2, \ldots, X_n\} - \frac{1}{\alpha} \log(n) \leq x \right]$$

$$= \Pr \left[ \left( X_1 \leq x + \frac{1}{\alpha} \log(n) \right) \land \cdots \land \left( X_n \leq x + \frac{1}{\alpha} \log(n) \right) \right]$$

$$= \begin{cases} (1 - e^{-(\alpha x + \log(n))})^n, & \text{if } \alpha x \geq -\log(n); \\ 0, & \text{if } \alpha x < -\log(n) \end{cases}$$

$$= \begin{cases} \left(1 - \frac{e^{-\alpha x}}{n}\right)^n, & \text{if } \alpha x \geq -\log(n); \\ 0, & \text{if } \alpha x < -\log(n) \end{cases}$$

$$\xrightarrow{n \to \infty} e^{-e^{-\alpha x}} = \Pr[X \leq x] \text{ for all } x \in \mathbb{R}.$$
\( X_n \xrightarrow{p} X \) implies \( X_n \Rightarrow X \)

**Theorem** \( X_n \xrightarrow{p} X \) implies \( X_n \Rightarrow X \).

**Proof:** \( X_n \xrightarrow{p} X \) means that

\[
\lim_{n \to \infty} \Pr[|X_n - X| > \varepsilon] = 0 \text{ for any positive } \varepsilon.
\]

Observe that

\[
\Pr[A \leq a] - \Pr[|A - B| > b] \leq \Pr[(A \leq a) \land (|A - B| > b)^c] = \Pr[(A \leq a) \land (|A - B| \leq b)] = \Pr[(A + b \leq a + b) \land (A - b \leq B \leq A + b)] \leq \Pr[B \leq a + b].
\]

\[
\Pr[X \leq x - \varepsilon] - \Pr[|X_n - X| > \varepsilon] \leq \Pr[X_n \leq (x - \varepsilon) + \varepsilon] = \Pr[X_n \leq x],
\]
and

\[
\Pr[X_n \leq x] - \Pr[|X_n - X| > \varepsilon] \leq \Pr[X \leq x + \varepsilon].
\]

Hence,

\[
\Pr[X \leq x - \varepsilon] - \Pr[|X_n - X| > \varepsilon] \leq \Pr[X_n \leq x] \leq \Pr[X \leq x + \varepsilon] + \Pr[|X_n - X| > \varepsilon],
\]
\( X_n \xrightarrow{p} X \) implies \( X_n \Rightarrow X \)

which implies that

\[
\Pr[X \leq x - \varepsilon] \leq \liminf_{n \to \infty} \Pr[X_n \leq x] \leq \limsup_{n \to \infty} \Pr[X_n \leq x] \leq \Pr[X \leq x + \varepsilon].
\]

Consequently, for every continuous point of \( \Pr[X \leq x] \) (i.e., \( \lim_{\varepsilon \downarrow 0} \Pr[X \leq x + \varepsilon] = \lim_{\varepsilon \downarrow 0} \Pr[X \leq x - \varepsilon] \)),

\[
\lim_{n \to \infty} \Pr[X_n \leq x] = \Pr[X \leq x].
\]
Counterexample for $X_n \Rightarrow X$ implying $X_n \xrightarrow{p} X$ 25-12

Counterexample $X \perp \perp Y$ and

$$\Pr[X = 0] = \Pr[X = 1] = \Pr[Y = 0] = \Pr[Y = 1] = \frac{1}{2}.$$ 

Let $X_n = Y$ for each $n$.

Then apparently, $X_n \Rightarrow X$.

However, for $0 < \varepsilon < 1$,

$$\Pr[|X_n - X| > \varepsilon] = \Pr[|Y - X| > \varepsilon]$$
$$= \Pr[X = 0 \land Y = 1] + \Pr[X = 1 \land Y = 0]$$
$$= \Pr[X = 0] \Pr[Y = 1] + \Pr[X = 1] \Pr[Y = 0]$$
$$= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}.$$ 

From the above, you may already get that it is really easy to construct a counterexample for $X_n \Rightarrow X$ implies $X_n \xrightarrow{p} X$. So a general condition for which $X_n \Rightarrow X$ implies $X_n \xrightarrow{p} X$ may be hard to create!
Counterexample for $X_n \Rightarrow X$ implying $X_n \xrightarrow{p} X$

Another note for counterexample construction for $X_n \Rightarrow X$ implying $X_n \xrightarrow{p} X$ is that:

- $X_n \xrightarrow{p} X$ requires that $X_1, X_2, X_3, \ldots$ must be random variables defined on \textbf{the same} probability space.
- But $X_n \Rightarrow X$ allows $X_1, X_2, X_3, \ldots$ to be defined over \textbf{distinct} probability space.

There is however an exception:

\textbf{Theorem} Suppose $\Pr[X = a] = 1$ and $X_1, X_2, \ldots$ are random variables defined over the same probability space. Then

$$X_n \xrightarrow{p} X \text{ if, and only if, } X_n \Rightarrow X.$$

- Notable, since $X$ is a degenerated random variable, $X_n \xrightarrow{p} X$ means that for some $a$,

$$\lim_{n \to \infty} \Pr[|X_n - a| \geq \varepsilon] = 0 \text{ for any } \varepsilon > 0.$$  

However, the validity of above inequality does not require $X_1, X_2, \ldots$ to be defined over \textbf{the same} probability space.
Counterexample for \( X_n \Rightarrow X \) implying \( X_n \overset{p}{\rightarrow} X \)

So we can rewrite the above theorem as:

**Theorem** Suppose \( \Pr[X = a] = 1 \). Then

\[
\lim_{n \rightarrow \infty} \Pr[|X_n - a| \geq \varepsilon] = 0 \text{ for any } \varepsilon > 0 \text{ if, and only if, } X_n \Rightarrow X.
\]
Properties regarding convergence in distribution

**Theorem** $X_n \Rightarrow X$ and $\delta_n \xrightarrow{n \to \infty} 0$ jointly imply that $\delta_n X_n \Rightarrow 0$.

**Proof:**

- For any $\eta > 0$ given, choose $x > 0$ such that
  \[ \Pr[|X| \geq x] < \eta \quad \text{and} \quad \Pr[X = \pm x] = 0. \]
  Imagine that $\eta$ small implies $x$ large for general $X$.
  In case $X$ is a degenerated random variable with $\Pr[X = x_0] = 1$, any $x > x_0$ will give $\Pr[|X| \geq x] = 0 < \eta$.

- For any $\varepsilon > 0$ given, choose $N_0$ such that
  \[ \delta_n < \frac{x}{\varepsilon} \text{ for } n \geq N_0. \]

- Since $\Pr[X = \pm x] = 0$ and $X_n \Rightarrow X$,
  \[ |\Pr[|X_n| \geq x] - \Pr[|X| \geq x]| \xrightarrow{n \to \infty} 0. \]
  Therefore, there exists $N_1$ such that for $n > N_1$,
  \[ |\Pr[|X_n| \geq x] - \Pr[|X| \geq x]| < \eta. \]
Properties regarding convergence in distribution

• Then for $n > \max\{N_0, N_1\}$,

$$
\Pr[|\delta_n X_n| \geq \varepsilon] = \Pr[|\delta_n| \cdot |X_n| \geq \varepsilon] \leq \Pr\left[\frac{\varepsilon}{x} |X_n| \geq \varepsilon\right] = \Pr[|X_n| \geq x] \\
\leq \Pr[|X| \geq x] + \eta < 2\eta.
$$

Hence,

$$
\limsup_{n \to \infty} \Pr[|\delta_n X_n| \geq \varepsilon] < 2\eta.
$$

• As $\eta$ can be chosen arbitrarily small, independent of $\varepsilon$,

$$
\limsup_{n \to \infty} \Pr[|\delta_n X_n| \geq \varepsilon] = 0.
$$
Properties regarding convergence in distribution

**Theorem 25.4** If \( X_n \xrightarrow{} X \) and \( X_n - Y_n \xrightarrow{} 0 \), then \( Y_n \xrightarrow{} X \).

**Proof:** For any \( x \) and arbitrarily small (but carefully chosen) \( \varepsilon > 0 \) (such that \( \Pr[X = y'] = \Pr[X = y''] = 0 \)), let \( y' = x - \varepsilon \) and \( y'' = x + \varepsilon \). Observe that

\[
\Pr[X_n \leq y'] - \Pr[|X_n - Y_n| > \varepsilon] \leq \left( \Pr[Y_n \leq y' + \varepsilon] = \right) \Pr[Y_n \leq x],
\]

and

\[
\Pr[Y_n \leq x] - \Pr[|X_n - Y_n| > \varepsilon] \leq \left( \Pr[X_n \leq x + \varepsilon] = \right) \Pr[X_n \leq y''].
\]

Hence,

\[
\Pr[X_n \leq y'] - \Pr[|X_n - Y_n| > \varepsilon] \leq \Pr[Y_n \leq x] \leq \Pr[X_n \leq y''] + \Pr[|X_n - Y_n| > \varepsilon],
\]

which implies that

\[
\Pr[X \leq x - \varepsilon] \leq \liminf_{n \to \infty} \Pr[Y_n \leq x] \leq \limsup_{n \to \infty} \Pr[Y_n \leq x] \leq \Pr[X \leq x + \varepsilon].
\]

Hence, the desired \( Y_n \xrightarrow{} X \) is obtained. \( \square \)
Properties regarding convergence in distribution

**Theorem 25.5** If

1. $X_{n,m} \xrightarrow{n \to \infty} X_m$,
2. $X_m \xrightarrow{m \to \infty} X$, and
3. $\lim_{m \to \infty} \limsup_{n \to \infty} \Pr[|X_{n,m} - Y_n| > \varepsilon] = 0$ for any positive $\varepsilon$,

then $Y_n \Rightarrow X$.

**Proof:**

- For any $x$, we can choose $\varepsilon$ arbitrarily small such that
  \[
  \Pr[X = y'] = \Pr[X_1 = y'] = \Pr[X_2 = y'] = \cdots = 0
  \]
  and
  \[
  \Pr[X = y''] = \Pr[X_1 = y''] = \Pr[X_2 = y''] = \cdots = 0,
  \]
  where $y' = x - \varepsilon$ and $y'' = x + \varepsilon$.

- We can then derive
  \[
  \Pr[X_{n,m} \leq y'] - \Pr[|X_{n,m} - Y_n| > \varepsilon] \leq \Pr[Y_n \leq x] \leq \Pr[X_{n,m} \leq y''] + \Pr[|X_{n,m} - Y_n| > \varepsilon].
  \]
Properties regarding convergence in distribution

Hence,

\[ \liminf_{n \to \infty} \left( \Pr[X_{n,m} \leq y'] - \Pr[|X_{n,m} - Y_n| > \varepsilon] \right) \]
\[ \leq \liminf_{n \to \infty} \Pr[Y_n \leq x] \]
\[ \leq \limsup_{n \to \infty} \Pr[Y_n \leq x] \]
\[ \leq \limsup_{n \to \infty} \left( \Pr[X_{n,m} \leq y'''] + \Pr[|X_{n,m} - Y_n| > \varepsilon] \right) , \]

which gives:

\[ \Pr[X_m \leq y'] - \limsup_{n \to \infty} \Pr[|X_{n,m} - Y_n| > \varepsilon] \]
\[ \leq \liminf_{n \to \infty} \Pr[Y_n \leq x] \]
\[ \leq \limsup_{n \to \infty} \Pr[Y_n \leq x] \]
\[ \leq \Pr[X_m \leq y'''] + \limsup_{n \to \infty} \Pr[|X_{n,m} - Y_n| > \varepsilon]. \]

• Taking \( m \) to infinity in the above inequality, we obtain:

\[ \Pr[X \leq y'] \leq \liminf_{n \to \infty} \Pr[Y_n \leq x] \leq \limsup_{n \to \infty} \Pr[Y_n \leq x] \leq \Pr[X \leq y'']. \]
Properties regarding convergence in distribution

- A random sequence cannot have two distinct weak limits.

**Theorem** Let $F_n$, $F$ and $G$ be cdfs of some random variables. If $F_n \Rightarrow F$ and $F_n \Rightarrow G$, then $F(x) = G(x)$ for every $x \in \mathbb{R}$.

**Proof:** By the definition of convergence in distribution, $F(x) = G(x)$ must coincides at every continuous points of $F(x)$ and $G(x)$. By definitions, cdfs must be right-continuous. So $F'(x)$ and $G'(x)$ coincides also at discontinuous points. $\square$
Theorem 25.6 (Skorohod’s theorem) Suppose $\mu_n$ and $\mu$ are probability measures on $(\mathbb{R}, \mathcal{B})$, and $\mu_n \Rightarrow \mu$. Then there exist random variables $Y_n$ and $Y$ such that:

1. they are both defined on common probability space $(\Omega, \mathcal{F}, P)$;
2. $\Pr[Y_n \leq y] = \mu_n(-\infty, y]$ for every $y$;
3. $\Pr[Y \leq y] = \mu(-\infty, y]$ for every $y$;
4. $\lim_{n \to \infty} Y_n(\omega) = Y(\omega)$ for every $\omega$.

**Implication:** Again, cdfs are sufficient; we do not need to rely on the inherited probability space.
Fundamental theorems (without proofs)

**Theorem (A simplified version of mapping theorem)** Suppose that a real-valued function $h$ is $\mathcal{B}/\mathcal{B}$-measurable, and the set $\mathcal{D}_h$ of its discontinuities is $\mathcal{B}$-measurable. Then

\[ X_n \Rightarrow X \text{ and } \Pr[X \in \mathcal{D}_h] = 0 \implies h(X_n) \Rightarrow h(X). \]

**Theorem** If $X_n \Rightarrow a$ and function $h$ is continuous at $a$, then $h(X_n) \Rightarrow h(a)$.

**Example** $X_n \Rightarrow X$ and $h(x) = ax + b$ imply $aX_n + b \Rightarrow aX + b$.

**Example** Suppose $X_n \Rightarrow X$ and $h(x) = ax + b$ and $a_n \to a$ and $b_n \to b$. Then (by Theorem 25.4)

\[
\begin{align*}
(aX_n + b) - (a_nX_n + b_n) &= (a - a_n)X_n + (b - b_n) \Rightarrow 0 \\
(aX_n + b) &\Rightarrow aX + b
\end{align*}
\]

imply $a_nX_n + b_n \Rightarrow aX + b$. 
Fundamental theorems (without proofs)

**Theorem 25.8 (A rephrased version)** The following two conditions are equivalent.

- \( F_n \Rightarrow F \);
- \( \lim_{n \to \infty} \int_{\mathbb{R}} f(x) dF_n(x) = \int_{\mathbb{R}} f(x) dF(x) \) for every bounded, continuous real function \( f \).

**Counterexample**

- \( X_n \) is uniformly distributed over \( \{0, 1/n, 2/n, \ldots, (n-1)/n\} \), and \( X \) is uniformly distributed over \([0, 1)\).
- \( F_n(x) = \text{Pr}[X_n \leq x] \) and \( F(x) = \text{Pr}[X \leq x] \).
- \( \mathcal{A} = \) set of all rational numbers in \([0, 1)\).
- \( f(x) = 1 \) if \( x \in \mathcal{A} \), and \( f(x) = 0 \), otherwise.
- Since \( f(\cdot) \) is not continuous (though bounded),

\[
1 = \int_{\mathbb{R}} f(x) dF_n(x) \not\Rightarrow \int_{\mathbb{R}} f(x) dF(x) = 0.
\]
Helly’s theorem

**Theorem 25.9 (Helly’s theorem)** For every sequence \( \{F_n\}_{n=1}^{\infty} \) of distribution functions, there exists a subsequence \( \{F_{n_k}\}_{k=1}^{\infty} \) and a non-decreasing, right-continuous function \( F \) (not necessarily a cdf) such that

\[
\lim_{k \to \infty} F_{n_k}(x) = F(x)
\]

for every continuous points of \( F \).
Helly’s theorem

**Theorem (The diagonal method)** Give a bounded sequence of real numbers:

\[
\begin{align*}
x_{1,1} & \quad x_{1,2} & \quad x_{1,3} & \quad \cdots \\
x_{2,1} & \quad x_{2,2} & \quad x_{2,3} & \quad \cdots \\
\vdots & \quad \vdots & \quad \vdots & \quad \vdots
\end{align*}
\]

There exists an increasing sequence \( n_1, n_2, \ldots \) such that the limit \( \lim_{k \to \infty} x_{m,n_k} \) exists for each \( m = 1, 2, 3, \ldots \).

**Proof:**

- For \( x_{1,1}, x_{1,2}, x_{1,3}, \ldots \), there exists \( n_{1,1}, n_{1,2}, n_{1,3}, \ldots \) such that \( \lim_{k \to \infty} x_{1,n_{1,k}} \) exists.
- For \( x_{2,n_{1,1}}, x_{2,n_{1,2}}, x_{2,n_{1,3}}, \ldots \), there exists \( n_{2,1}, n_{2,2}, n_{2,3}, \ldots \) such that \( \lim_{k \to \infty} x_{2,n_{2,k}} \) exists (and still, \( \lim_{k \to \infty} x_{1,n_{2,k}} \) exists).
- Repeat the process to obtain:

\[
\begin{align*}
n_{1,1} & \quad n_{1,2} & \quad n_{1,3} & \quad \cdots \\
n_{2,1} & \quad n_{2,2} & \quad n_{2,3} & \quad \cdots \\
\vdots & \quad \vdots & \quad \vdots & \quad \vdots
\end{align*}
\]

Since each row is a subsequence of the previous row in the above \( n \)-list, \( n_{k,k} \) is increasing in \( k \). Finally, \( n_{k,k}, n_{k+1,k+1}, n_{k+2,k+2}, \ldots \) satisfies that \( \lim_{k \to \infty} x_{m,n_{k,k}} \) exists. \( \square \).
Helly’s theorem

Proof of Helly’s theorem:

- List the two dimensional array of \( F_n(r) \) for \( r \) rational. Then by the diagonal method, there exists \( n_1, n_2, \ldots \) such that \( \lim_{k \to \infty} F_{n_k}(r) \) exists for every rational \( r \).

- Let \( G(r) = \lim_{k \to \infty} F_{n_k}(r) \) for every rational \( r \) (So for two rationals \( s < r \), \( G(s) \leq G(r) \)) and define

  \[
  F(x) = \inf \{ G(r) : r > x \text{ and } r \text{ rational} \}.
  \]

  Thus, \( F(x) \) is clearly non-decreasing, since taking infimum over a smaller set yields a larger value. (So for any \( r > x \), \( G(r) \geq F(x) \).)

- By definition of infimum, for a given \( \varepsilon > 0 \), there exists a rational \( r > x \) such that

  \[
  G(r) < F(x) + \varepsilon.
  \]

- (Base on the above \( r, x \) and \( \varepsilon \).) For any \( \mu \) satisfying \( x < x + \mu < r \), \( F(x) \leq F(x + \mu) \leq G(r) \) \( (< F(x) + \varepsilon) \).

  So

  \[
  F(x) \leq \lim_{\mu \downarrow 0} F(x + \mu) < F(x) + \varepsilon.
  \]
Helly’s theorem

(The limit of \( \lim_{\mu \downarrow 0} F(x + \mu) \) must exist. Why?)

Since the above inequality is valid for any \( \varepsilon > 0 \),

\[
\lim_{\mu \downarrow 0} F(x + \mu) = F(x),
\]

which means that \( F(\cdot) \) is right-continuous.

• Finally, suppose that \( F(\cdot) \) is continuous at \( x \).

Then again, by definition of infimum, for a given \( \varepsilon > 0 \), there exists a rational \( r > x \) such that

\[
G(r) < F(x) + \varepsilon.
\]

Also, by continuity, for this \( \varepsilon \), there exists \( y < x \) such that

\[
F(x) - \varepsilon < F(y).
\]

Choose another rational \( s \) satisfying \( y < s < x \) \((< r)\). Apparently, \( F(y) \leq G(s) \) and \( G(s) \leq G(r) \).

Therefore, we have:

\[
F(x) - \varepsilon < G(s) \leq G(r) < F(x) + \varepsilon.
\]

On the other hand,

\[
F_n(s) \leq F_n(x) \leq F_n(r)
\]
Helly’s theorem

implies that

\[ G(s) = \lim_{k \to \infty} F_{n_k}(s) \leq \liminf_{k \to \infty} F_{n_k}(x) \leq \limsup_{k \to \infty} F_{n_k}(x) \leq \lim_{k \to \infty} F_{n_k}(r) = G(r). \]

The above concludes to:

\[ F(x) - \varepsilon \leq \liminf_{k \to \infty} F_{n_k}(x) \leq \limsup_{k \to \infty} F_{n_k}(x) \leq F(x) + \varepsilon. \]

The proof is completed by noting that \( \varepsilon \) can be made arbitrarily small. \( \Box \).

- In the above theorem, the limit \( F(\cdot) \) is not necessarily a cdf!

Example \( F_n(x) = 0 \) for \( x < n \) and \( F_n(x) = 1 \) for \( x \geq n \). Then

\[ \lim_{n \to \infty} F_n(x) = 0 \]

for every \( x \in \mathbb{R} \).
Definition (tightness) A sequence of cdf's is said to be tight if for any $\varepsilon > 0$, there exist $x$ and $y$ such that

$$F_n(x) < \varepsilon \text{ and } F_n(y) > 1 - \varepsilon \text{ for all sufficiently large } n.$$ 

- It can be shown that the limit $F(\cdot)$ in Helly’s theorem satisfies $0 \leq F(x) \leq 1$.
- Also, $F(\cdot)$ is right-continuous and non-decreasing.
- So if $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. Then $F(\cdot)$ becomes a cdf.
- Tightness is a condition to prevent the probability mass from escaping to infinity.
**Theorem 25.10 (rephrased version)** Tightness of \( \{F_{n_k}\}_{k=1}^{\infty} \) is a necessary and sufficient condition for the limit \( F(\cdot) \) in Helly’s theorem to be a cdf.

**Proof:**

1. **Sufficiency:** Suppose \( \{F_{n_k}(\cdot)\}_{k=1}^{\infty} \) is tight. Then for any \( \varepsilon > 0 \), we can find \( x \) and \( y \) such that

\[
F_{n_k}(x) < \varepsilon \quad \text{and} \quad F_{n_k}(y) > 1 - \varepsilon \quad \text{for all sufficiently large} \quad k.
\]

Hence,

\[
F(x) = \lim_{k \to \infty} F_{n_k}(x) \leq \varepsilon \quad \text{and} \quad F(y) = \lim_{k \to \infty} F_{n_k}(y) \geq 1 - \varepsilon,
\]

which implies

\[
\lim_{x \downarrow -\infty} F(x) \leq \varepsilon \quad \text{and} \quad \lim_{y \uparrow \infty} F(y) \geq 1 - \varepsilon.
\]

The proof is completed by noting that \( \varepsilon \) can be made arbitrarily small.
2. **Necessity:** Suppose that $F(\cdot)$ is a cdf. Then for any $\varepsilon > 0$, there exist $x$ and $y$ such that

$$F(x) < \varepsilon \text{ and } F(y) > 1 - \varepsilon.$$ 

In other words,

$$\lim_{k \to \infty} F_{n_k}(x) < \varepsilon \text{ and } \lim_{k \to \infty} F_{n_k}(y) > 1 - \varepsilon.$$ 

Therefore, for all sufficiently large $k$,

$$F_{n_k}(x) < \varepsilon \text{ and } F_{n_k}(y) > 1 - \varepsilon.$$ 

\[ \square \]

To let you have some feeling on **tightness**, we provide the next observation.

**Observation** Suppose $F_n(\cdot)$ is a degenerated cdf at $x_n$. Then $\{F_n\}_{n=1}^\infty$ is tight if, and only if, $\{x_n\}_{n=1}^\infty$ is bounded.

- **Final remark on tightness:** *Tightness* on sequences of probability measures is similar to *boundedness* on sequences of real numbers.
Example 25.10 Let \( \{X_n\}_{n=1}^{\infty} \) be a sequence of normal distribution with mean \( m_n \) and variance \( \sigma_n^2 \).

- If \( \{m_n\}_{n=1}^{\infty} \) and \( \{\sigma_n\}_{n=1}^{\infty} \) are bounded, then \( \{X_n\}_{n=1}^{\infty} \) is tight.

**Proof:** By Markov’s inequality,

\[
\Pr[|X_n| > a] \leq \frac{E[X_n^2]}{a^2} = \frac{\sigma_n^2 + m_n^2}{a^2} \leq \frac{\sigma_{\max}^2 + m_{\max}^2}{a^2}.
\]

So for any \( \varepsilon > 0 \),

\[
x = -\sqrt{\frac{\sigma_{\max}^2 + m_{\max}^2}{\varepsilon}}
\]

and

\[
y = \sqrt{\frac{\sigma_{\max}^2 + m_{\max}^2}{\varepsilon}}
\]

satisfy the tightness condition. \( \Box \).
Example for tightness vs boundedness

Example 25.10 (cont.)

- If \( \{m_n\}_{n=1}^{\infty} \) is unbounded, then \( \{X_n\}_{n=1}^{\infty} \) is not tight!

  \textit{Proof:} This can be easily seen from \( \Pr[X_n \geq m_n] = \Pr[X_n \leq m_n] = 1/2. \) \( \square \)
Moment and in-distribution convergence

- Convergence in mean implies convergence in distribution. But the reverse is not necessarily true.
- However, we can still say “something” in the reverse direction.

**Theorem 25.11** If $X_n \Rightarrow X$, then

$$E[|X|] \leq \liminf_{n \to \infty} E[|X_n|].$$

**Lemma (Fatou’s lemma)** If $\{f_n(\cdot)\}_{n=1}^{\infty}$ is a sequence of non-negative measurable functions, and $\lim_{n \to \infty} f_n(x) = f(x)$ for every $x \in \mathcal{E}$ except on a set of Lebesgue measure zero, then

$$\int_{\mathcal{E}} f(x) dx \leq \liminf_{n \to \infty} \int_{\mathcal{E}} f_n(x) dx.$$

**Proof:** By Fatou’s lemma,

$$\int_{\mathbb{R}} |x|dF(x) \leq \liminf_{n \to \infty} \int_{\mathbb{R}} |x|dF_n(x).$$
Moment and in-distribution convergence

**Definition (Integrability)** A random variable $X$ is integrable, if

$$\lim_{\alpha \to \infty} \int_{|x| \geq \alpha} |x|dF_X(x) = 0.$$  

**Lemma** A random variable $X$ is integrable if, and only if, $E[|X|] < \infty$.

**Proof:**

\[
\lim_{\alpha \to \infty} \int_{|x| \geq \alpha} |x|dF_X(x) = 0
\]

\[
\Rightarrow (\exists \alpha') \int_{|x| \geq \alpha'} |x|dF_X(x) < \varepsilon \text{ for a given } \varepsilon > 0
\]

\[
\Rightarrow E[|X|] = \int_{|x| < \alpha'} |x|dF_X(x) + \int_{|x| \geq \alpha'} |x|dF_X(x) \leq \alpha' + \varepsilon < \infty.
\]

and

\[
\lim_{\alpha \to \infty} \int_{|x| < \alpha} |x|dF_X(x) = \int_{\mathbb{R}} |x|dF_X(x) < \infty
\]

\[
\Rightarrow \lim_{\alpha \to \infty} \int_{|x| \geq \alpha} |x|dF_X(x) = \lim_{\alpha \to \infty} \left( E[|X|] - \int_{|x| < \alpha} |x|dF_X(x) \right) = 0.
\]

• Hence, integrability can also be defined directly through $E[|X|] < \infty$.  

Moment and in-distribution convergence

**Definition (Uniform integrability)** A sequence of random variables \( \{X_n\}_{n=1}^{\infty} \) (defined over the same probability space) is *uniformly integrable* if

\[
\lim_{\alpha \to \infty} \sup_{n \geq 1} \int_{\{ |x| \geq \alpha \}} |x|dF_{X_n}(x) = 0.
\]

- The necessity of the condition of defining over the same probability space \((\Omega, \mathcal{F}, P)\) is more obvious, if we write the above equation as:

\[
\lim_{\alpha \to \infty} \sup_{n \geq 1} \int_{\{ \omega \in \Omega : |x_n(\omega)| \geq \alpha \}} |x_n(\omega)|dP(\omega) = 0.
\]

- However, I personally think that the condition of defining over the same probability space can be relaxed since in-distribution convergence does not require this condition.
Moment and in-distribution convergence

**Lemma** Uniform integrability implies that

\[
\sup_{n \geq 1} E[|X_n|] < \infty.
\]

**Proof:**

\[
\lim_{\alpha \to \infty} \sup_{n \geq 1} \int_{|x| \geq \alpha} |x| dF_{X_n}(x) = 0
\]

\[\Rightarrow (\exists \alpha') \sup_{n \geq 1} \int_{|x| \geq \alpha'} |x| dF_{X_n}(x) < \varepsilon \text{ for a given } \varepsilon > 0\]

\[\Rightarrow \sup_{n \geq 1} E[|X_n|] = \sup_{n \geq 1} \left( \int_{|x| < \alpha'} |x| dF_{X_n}(x) + \int_{|x| \geq \alpha'} |x| dF_{X_n}(x) \right) \leq \alpha' + \varepsilon < \infty.\]

\[\square.\]

- Although the converse statement for **integrability** holds, the converse statement for the **uniform integrability** is not necessarily valid.
Lemma

\[ \sup_{n \geq 1} E[|X_n|] < \infty \]

does not necessarily imply uniform integrability.

Proof: Let \( \text{Pr}[X_n = 0] = 1 - (1/n) \) and \( \text{Pr}[X_n = n] = 1/n \).

Then, \( E[|X_n|] = 1 \) for every \( n \), but

\[ \int_{|x| > \alpha} |x| dF_n(x) = \begin{cases} 
0, & n < \alpha; \\
1, & n > \alpha.
\end{cases} \]

We therefore have

\[ \sup_{n \geq 1} E[|X_n|] = 1 < \infty \quad \text{but} \quad \lim_{\alpha \to \infty} \sup_{n \geq 1} \int_{|x| > \alpha} |x| dF_n(x) = 1 \not\to 0. \]
Moment and in-distribution convergence

Remark:

• In the above example, we actually have

\[ E[|X_n|] = \int_{|x|>\alpha} |x|dF_n(x) \quad \text{for } n > \alpha. \]

Hence, the uniform “boundedness” of \( E[|X_n|] \) for \( n > \alpha \) (i.e., \( \sup_{n \geq \alpha} E[|X_n|] < \infty \)) does not imply the uniform “close-to-zero” of \( E[|X_n|] \) (i.e.,

\[ \sup_{n \geq \alpha} E[|X_n|] \to 0 \quad \text{as } \alpha \to \infty. \]

• 

\[ \sup_{n \geq 1} E[|X_n|] < \infty \]

does not necessarily imply uniform integrability.

But

\[ \sup_{n \geq 1} E[|X_n|^{1+\varepsilon}] < \infty \]

does. (This can be proved by the generalized Markov inequality introduced in the next slide with \( b = 1 \) and \( k = \varepsilon \).)
Generalization of Markov’s inequality

Markov’s inequality

\[ \int_{[|x| \geq \alpha]} dF_X(x) \leq \frac{1}{\alpha^k} E[|X|^k]. \]

Generalized Markov’s inequality

\[ \int_{[|x| \geq \alpha]} |x|^b dF_X(x) \leq \frac{1}{\alpha^k} E[|X|^{b+k}]. \]

Proof:

\[ E[|X|^{b+k}] = \int_{\mathbb{R}} |x|^{b+k} dF_X(x) \]
\[ \geq \int_{[|x| \geq \alpha]} |x|^{b+k} dF_X(x) \]
\[ \geq \alpha^k \int_{[|x| \geq \alpha]} |x|^b dF_X(x). \]
More on uniform integrability

**Lemma** If there exists an integrable random variable $Z$ with

$$\Pr[|X_n| \geq t] \leq \Pr[|Z| \geq t] \text{ for all } t \text{ and } n,$$

then $\{X_n\}_{n=1}^\infty$ is uniformly integrable.

**Proof:**

$$\int_{[x \geq \alpha]} x dF_X(x) = \alpha \Pr[X \geq \alpha] + \int_{\alpha}^{\infty} \Pr[X \geq t] dt.$$

$$\int_{|[x| \geq \alpha]} |x| dF_{X_n}(x) = \alpha \Pr[|X_n| \geq \alpha] + \int_{\alpha}^{\infty} \Pr[|X_n| \geq t] dt$$

$$\leq \alpha \Pr[|Z| \geq \alpha] + \int_{\alpha}^{\infty} \Pr[|Z| \geq t] dt$$

$$= \int_{|[z| \geq \alpha]} |z| dF_Z(z).$$
Theorem 25.12 If $X_n \Rightarrow X$ and $\{X_n\}_{n=1}^{\infty}$ uniformly integrable, then $X$ is integrable, and $E[X_n] \xrightarrow{n \to \infty} E[X]$.

Proof:

- By uniform integrability,
  
  $$E[|X|] \leq \liminf_{n \to \infty} E[|X_n|] \leq \sup_{n \geq 1} E[|X_n|] < \infty.$$ 

  Hence, $X$ is integrable.

- Define $Y_n = X_n I_{|X_n| < \alpha}$ and $Y = X I_{|X| < \alpha}$.
  Observe that

  \[
  \begin{align*}
  Y_n^+ &\Rightarrow Y^+ \\
  \alpha - Y_n^- &\Rightarrow \alpha - Y^+
  \end{align*}
  \]

  imply

  \[
  \begin{align*}
  E[Y^+] &\leq \liminf_{n \to \infty} E[Y_n^+] \\
  \alpha - E[Y^+] &\leq \liminf_{n \to \infty} (\alpha - E[Y_n^+]) = \alpha - \limsup_{n \to \infty} E[Y_n^+].
  \end{align*}
  \]

  Hence, $\lim_{n \to \infty} E[Y_n^+] = E[Y^+]$.

  Similarly, we have $\lim_{n \to \infty} E[Y_n^-] = E[Y^-]$.

  Accordingly, $\lim_{n \to \infty} E[Y_n] = E[Y]$. 

Moment and in-distribution convergence

\[ \left| \int_{\mathbb{R}} xdF_{X_n}(x) - \int_{\mathbb{R}} xdF_X(x) \right| = \left| \int_{|x|<\alpha} xdF_{X_n}(x) - \int_{|x|<\alpha} xdF_X(x) + \int_{|x|\geq\alpha} xdF_{X_n}(x) - \int_{|x|\geq\alpha} xdF_X(x) \right| \\
= \left| \int_{\mathbb{R}} ydF_{Y_n}(y) - \int_{\mathbb{R}} ydF_Y(y) + \int_{|x|\geq\alpha} xdF_{X_n}(x) - \int_{|x|\geq\alpha} xdF_X(x) \right| \\
\leq \left| \int_{\mathbb{R}} ydF_{Y_n}(y) - \int_{\mathbb{R}} ydF_Y(y) \right| + \sup_{n \geq 1} \int_{|x|\geq\alpha} |x|dF_{X_n}(x) + \int_{|x|\geq\alpha} |x|dF_X(x) \\
\]

Therefore,

\[ \limsup_{n \to \infty} \left| \int_{\mathbb{R}} xdF_{X_n}(x) - \int_{\mathbb{R}} xdF_X(x) \right| \leq \sup_{n \geq 1} \int_{|x|\geq\alpha} |x|dF_{X_n}(x) + \int_{|x|\geq\alpha} |x|dF_X(x). \]

The proof is completed by taking \( \alpha \) to the infinity. \( \Box \)
Corollary Let $r$ be a positive integer. If $X_n \Rightarrow X$ and $\sup_{n \geq 1} E[|X_n|^{r+\varepsilon}] < \infty$, where $\varepsilon > 0$, then

$$|X|^r$$ integrable, and $E[|X_n|^r] \xrightarrow{n \to \infty} E[|X|^r]$.

Proof: This is a direct consequence of Theorem 25.12 by noting that:

1. $X_n \Rightarrow X$ implies $|X_n|^r \Rightarrow |X|^r$, and

2. $\sup_{n \geq 1} E[|X_n|^{r+\varepsilon}] < \infty$ implies $\{|X_n|^r\}_{n=1}^{\infty}$ is uniformly integrable.