Definition 5.5. *The sequence of probability measures* $\{P_n\}_{n\in\mathbb{N}}$ *converges generally to the measure* P*, denoted by* $P_n \Longrightarrow P$ *if*

$$P_n(E) \to P(E) \text{ as } n \to \infty \text{ for all } E \in \mathcal{B} \text{ such that } P(\partial E) = 0,$$

where $\partial E = \bar{E} \cap \overline{E^c}$ *is the boundary of* E *and* $\bar{E}$ *is the closure of* E*.*

Theorem 5.8. *The following convergence statements are equivalent:*

1. $P_n \xrightarrow{w} P$,
2. $\limsup_{n\to\infty} P_n(E) \leq P(E)$ *if* $E \in \mathcal{B}$ *is closed,*
3. $\liminf_{n\to\infty} P_n(E) \geq P(E)$ *if* $E \in \mathcal{B}$ *is open,*
4. $P_n \Longrightarrow P$.

Proof. (1. $\implies$ 2.) Let $x \in \mathbb{R}$ and define $|x - E| = \inf\{|x-y| : y \in E\}$, $E(\varepsilon) = \{x : |x - E| < \varepsilon\}$ for $\varepsilon > 0$, $f(x) = I_E(x)$,

$$g(x) = \begin{cases} 1, & \text{if } x \leq 0 \\ 1 - x, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{if } x \geq 1 \end{cases}$$

and $f_\varepsilon(x) = g\left(\frac{1}{\varepsilon}|x - E|\right)$. Note that if $x \in E(\varepsilon)$ then $\frac{1}{\varepsilon}|x-E| < 1$ and $f_\varepsilon(x) > 0$. Also, if $\varepsilon \downarrow 0$ then $E(\varepsilon) \downarrow E$. Since g is bounded and continuous, so is f_ε. Now,

$$\int_{\mathbb{R}} f dP_n = \int_{\mathbb{R}} I_E P_n = P_n(E) \leq \int_{\mathbb{R}} f_\varepsilon dP_n. \tag{5.4}$$

The inequality follows because if $x \in E$, $\varepsilon^{-1}|x - E| = 0$ and $f_\varepsilon(x) = g(0) = 1 = I_E(x)$, but if $x \notin E$ then $\varepsilon^{-1}|x-E| > 0$ and $f_\varepsilon(x) = g(\varepsilon^{-1}|x - E|) \geq 0 = I_E(x)$. Then, taking limits on both sides of equation (5.4) gives

$$\limsup_{n\to\infty} P_n(E) \leq \limsup_{n\to\infty} \int_{\mathbb{R}} f_\varepsilon dP_n = \int_{\mathbb{R}} f_\varepsilon dP$$

where the last equality follows from the fact that f_ε is continuous and bounded on $\mathbb{R}$ and the assumption that 1) holds. But

$$\int_{\mathbb{R}} f_\varepsilon dP \leq \int_{\mathbb{R}} I_{E(\varepsilon)} dP = P(E(\varepsilon)) \tag{5.5}$$

where the inequality follows from the fact that if $x \in E(\varepsilon)$ then $\varepsilon^{-1}|x - E| < 1$ and consequently $0 < f_\varepsilon(x) \leq 1 = I_{E(\varepsilon)}$. If $x \notin E(\varepsilon)$ then $f_\varepsilon(x) = 0 = I_{E(\varepsilon)}$. Consequently, combining equations (5.4) and (5.5) we obtain $\limsup_{n\to\infty} P_n(E) \leq P(E(\varepsilon))$. Given that if $\varepsilon \downarrow 0$, $E(\varepsilon) \downarrow E$, by continuity of probability measure we have $\limsup_{n\to\infty} P_n(E) \leq P(E)$.

(2. $\implies$ 3.) If E is open, then E^c is closed. Thus, from 2) $\limsup_{n\to\infty} P_n(E^c) \leq P(E^c)$. But since $P_n(E^c) = 1 - P_n(E)$ and $P(E^c) = 1 - P(E)$ we have

$$1 + \limsup_{n\to\infty} (-P_n(E)) \leq 1 - P(E) \iff 1 - \liminf_{n\to\infty} P_n(E) \leq 1 - P(E) \iff \liminf_{n\to\infty} P_n(E) \geq P(E).$$

It is evident from this argument that (3. $\implies$ 2.).

(3. $\implies$ 4.) The interior of E, denoted by $int(E)$, is open and $int(E) = E - \partial E$. Since, 2. and 3. are equivalent and $int(E)$ is open and $\bar{E}$ is closed we have

$$\limsup_{n\to\infty} P_n(E) \leq \limsup_{n\to\infty} P_n(\bar{E}) \leq P(\bar{E}), \tag{5.6}$$

$$\liminf_{n\to\infty} P_n(E) \geq \liminf_{n\to\infty} P_n(int(E)) \geq P(int(E)). \tag{5.7}$$

But if $P(\partial E) = 0$ then $P(\bar{E}) = P(int(E)) = P(E)$ and $P_n(E) \to P(E)$ whenever $P(\partial E) = 0$, i.e., $P_n \implies P$.

(4. $\implies$ 1.) Let f be bounded and continuous with $|f| < C$ and define

$$D = \{d \in \mathbb{R} : P(\{x : f(x) = d\}) > 0\}.$$

Now, choose $\{y_i\}_{i=0}^k$ such that $y_0 = -C < y_1 < \cdots < y_k = C$. $d \in D$ implies $P(f^{-1}(\{d\})) > 0$. Since f is a function, for any two $d \neq d'$ such that $d, d' \in D$ we have $f^{-1}(\{d\}) \cap f^{-1}(\{d'\}) = \emptyset$, and since $P \leq 1$, there can be at most countably many elements in D. Suppose $\{y_i\}_{i=0}^k \nsubseteq D$

and $B_i = \{x \in \mathbb{R} : y_i \leq f(x) < y_{i+1}\}$ for $i = 0, 1, \cdots, k-1$. Then,

$$\partial B_i = \{x \in \mathbb{R} : y_i = f(x)\} \cup \{x \in \mathbb{R} : y_{i+1} = f(x)\} = f^{-1}(y_i) \cup f^{-1}(y_{i+1})$$

and $P(\partial B_i) = 0$ since $\{y_i\}_{i=0}^k \nsubseteq D$. Since, $int(B_i) = B_i - \partial B_i$ we have that $P(B_i) = P(int(B_i))$ and by 4) $P_n(B_i) - P(B_i) \to 0$. Consequently,

$$\sum_{i=0}^{k-1} y_i P_n(B_i) \to \sum_{i=0}^{k-1} y_i P(B_i). \tag{5.8}$$

Now,

$$\begin{aligned}
\left| \int_{\mathbb{R}} f dP_n - \int_{\mathbb{R}} f dP \right| &\leq \left| \int_{\mathbb{R}} f dP_n - \sum_{i=0}^{k-1} y_i P_n(B_i) \right| + \left| \sum_{i=0}^{k-1} y_i P_n(B_i) - \sum_{i=0}^{k-1} y_i P(B_i) \right| \\
&+ \left| \sum_{i=0}^{k-1} y_i P(B_i) - \int_{\mathbb{R}} f dP \right| \\
&\leq 2 \max_{0 \leq i \leq k-1} (y_{i+1} - y_i) + \left| \sum_{i=0}^{k-1} y_i P_n(B_i) - \sum_{i=0}^{k-1} y_i P(B_i) \right|.
\end{aligned}$$

By equation (5.8) and the fact that $\{y_i\}_{i=0}^k$ are arbitrary we have the result. ■

Recall that with a random variable $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ we can associate a distribution function $F_X(x) : \mathbb{R} \to [0, 1]$ with the following properties:

(i) F_X is non-decreasing,

(ii) F_X is right-continuous,

(iii) $\lim_{x \to \infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$.

Let $C(F_X) = \{x \in \mathbb{R} : F_X \text{ is continuous at } x\}$ and note that $C(F_X)^c$ is a countable set.

Definition 5.6. *Let F_n, F_X be distribution functions associated with random variables X_n, X with $n = 1, 2, \ldots$. We say that X_n converges in distribution to X and write $X_n \xrightarrow{d} X$ if*

$$F_n(x) \to F_X(x), \text{ for all } x \in C(F_X).$$

In this case, we write $F_n \Longrightarrow F_X$ and say that F_n converges generally to F_X.

Theorem 5.9. *The following statements are equivalent:*

1. $P_n \xrightarrow{w} P$,

2. $P_n \Longrightarrow P$,

3. $F_n \xrightarrow{w} F$,

4. $F_n \Longrightarrow F$.

Proof. We have proved that 1. and 2. are equivalent. In addition, by construction 1. and 3. are equivalent, so we need only show that 2 and 4 are equivalent.

(2. $\Longrightarrow$ 4.) Since $P_n \Longrightarrow P$ we have, in particular, that

$$P_n((-\infty, x]) \to P((-\infty, x])$$

for all $x \in \mathbb{R}$ such that $P(\{x\}) = 0$. But this means that $F_n \Longrightarrow F$.

(4. $\Longrightarrow$ 2.) We need to prove that $P_n \Longrightarrow P$, but since by Theorem 5.8 we have that $P_n \Longrightarrow P$ is equivalent to $\liminf_{n\to\infty} P_n(E) \geq P(E)$ if $E \in \mathcal{B}$ is open, this is what we will establish. Since E is an open set in $\mathbb{R}$ it can be written as $E = \cup_{k=1}^{\infty} \mathcal{I}_k$ where $\mathcal{I}_k = (a_k, b_k)$ are component intervals (disjoint). Let $\epsilon > 0$ and for each $\mathcal{I}_k$ choose $\mathcal{I}'_k = (a'_k, b'_k]$ a subinterval such that a'_k, b'_k are points of continuity of F and $P(\mathcal{I}_k) \leq P(\mathcal{I}'_k) + 2^{-k}\epsilon$. The existence of these intervals is assured by the fact that F has at most countable many discontinuities. Now,

$$\begin{aligned}
\liminf_{n\to\infty} P_n(E) &= \liminf_{n\to\infty} \sum_{k=1}^{\infty} P_n(\mathcal{I}_k) \\
&\geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} P_n(\mathcal{I}_k) \text{ by Fatou's Lemma} \\
&\geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} P_n(\mathcal{I}'_k).
\end{aligned}$$

But by 4. we have that $P_n(\mathcal{I}'_k) = F_n(b'_k) - F_n(a'_k) \to F(b'_k) - F(a'_k) = P(\mathcal{I}'_k)$. Hence,

$$\liminf_{n\to\infty} P_n(E) \geq \sum_{k=1}^{\infty} P(\mathcal{I}'_k) \geq \sum_{k=1}^{\infty} \left(P(\mathcal{I}_k) - 2^{-k}\epsilon\right) = P(E) - \epsilon.$$

Since ϵ is arbitrary the proof is complete. ■

Remark 5.2. *1. Convergence in distribution says nothing about $X_n(\omega)$, rather it focuses on F_n, as $n \to \infty$. For example, let $X_n = (-1)^n \mathcal{Z}$ where $\mathcal{Z} \sim N(0,1)$. Then, let $f_{\mathcal{Z}}(x) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}x^2\}$ for all $x \in \mathbb{R}$. For n odd,*

$$\begin{aligned} F_n(x) &= P(\{\omega : X_n(\omega) \leq x\}) = P(\{\omega : -\mathcal{Z} \leq x\}) = P(\{\omega : \mathcal{Z} \geq -x\}) \\ &= 1 - P(\{\omega : \mathcal{Z} < -x\}) = 1 - \int_{(-\infty,-x)} f_{\mathcal{Z}}(y)dy \\ &= \int_{[-x,\infty)} f_{\mathcal{Z}}(y)dy = \int_{(-\infty,x]} f_{\mathcal{Z}}(y)dy = F_{\mathcal{Z}}(x). \end{aligned}$$

The next to last equality follows from $f_{\mathcal{Z}}(z) = f_{\mathcal{Z}}(-z)$. For n even it is obvious that $F_n(x) = F_{\mathcal{Z}}(x)$. Hence, $F_n(x) = F_{\mathcal{Z}}(x)$, for all n and trivially $F_n(x) \to F_{\mathcal{Z}}(x)$ for all $x \in \mathbb{R}$.

However, if $E_n = \{\omega : |X_n(\omega) - \mathcal{Z}(\omega)| < \epsilon\}$, then $E_1 = \{\omega : |-\mathcal{Z}(\omega) - \mathcal{Z}(\omega)| < \epsilon\} = \{\omega : |\mathcal{Z}| < \epsilon/2\}$, $E_2 = \Omega, \ldots$. Hence, there is no limit for $\{P(E_n)\}_{n=1,2,\ldots}$ and $X_n \overset{p}{\nrightarrow} \mathcal{Z}$ (neither does $X_n \overset{as}{\to} \mathcal{Z}$). This shows that convergence in distribution is a very weak mode of convergence relative to the ones we have seen so far.

2. Contrary to other modes of convergence, here there is no need to have the random variables defined in the same probability space.

Theorem 5.10. *(Continuous Mapping Theorem) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables and X be a random variable such that $X_n \overset{d}{\to} X$ as $n \to \infty$. Let $h : \mathbb{R} \to \mathbb{R}$ be continuous at every point of a set C such that $P(\{\omega : X(\omega) \in C\}) = 1$. Then,*

$$h(X_n) \overset{d}{\to} h(X).$$

Proof. For any closed set G let $E_n = \{\omega : h(X_n(\omega)) \in G\} = \{\omega : X_n(\omega) \in h^{-1}(G)\} = X_n^{-1}(h^{-1}(G))$. Note that $P(E_n) = P(X_n^{-1}(h^{-1}(G))) = P_n(h^{-1}(G))$ and

$$h^{-1}(G) \subseteq \overline{h^{-1}(G)} \subseteq h^{-1}(G) \cup C^c. \tag{5.9}$$

The first set containment follows from the fact that every set is a subset of its closure. For the second set containment, note that

$$\overline{h^{-1}(G)} \;=\; (\overline{h^{-1}(G)} \cap C) \cup (\overline{h^{-1}(G)} \cap C^c) \subseteq (\overline{h^{-1}(G)} \cap C) \cup C^c$$

Now, $(\overline{h^{-1}(G)} \cap C) = (h^{-1}(G) \cup [h^{-1}(G)]^D) \cap C = (h^{-1}(G) \cap C) \cup ([h^{-1}(G)]^D \cap C)$, where $[h^{-1}(G)]^D$ is the derived set of $h^{-1}(G)$.[1] If $x \in [h^{-1}(G)]^D$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \in h^{-1}(G) \iff \{h(x_n)\}_{n\mathbb{N}} \in G$ such that $x_n \to x$. Furthermore, if $x \in C$, then if $x_n \to x$ we have that $h(x_n) \to h(x)$ and $h(x) \in G$ since G is closed. But $x \in [h^{-1}(G)]^D$ implies $x \notin h^{-1}(G) \iff h(x) \notin G$. Hence, $[h^{-1}(G)]^D \cap C = \emptyset$ and $\overline{h^{-1}(G)} \subseteq h^{-1}(G) \cup C^c$.

Consequently,

$$\begin{aligned}\limsup_{n\to\infty} P(E_n) = \limsup_{n\to\infty} P_n(h^{-1}(G)) &\le \limsup_{n\to\infty} P_n(\overline{h^{-1}(G)}) \\ &\le P_X\left(\overline{h^{-1}(G)}\right),\end{aligned}$$

where the last inequality follows from part 2 of Theorem 5.8. Since $P_X(C^c) = 0$, we have from (5.9) that $P_X\left(\overline{h^{-1}(G)}\right) \le P_X(h^{-1}(G))$ and we have

$$\limsup_{n\to\infty} P_n(h^{-1}(G)) \le P_X(h^{-1}(G)).$$

Repeating the argument in the opposite direction completes the proof.

■

Theorem 5.11. *Let D be dense*[2] *in $\mathbb{R}$. Suppose $F_D : D \to [0,1]$ satisfies:*

[1] The collection of its limit points.

[2] A set S is dense in $\mathbb{R}$ if $\bar{S} = \mathbb{R}$ where $\bar{S} = \{x \in \mathbb{R} : S \cap B(x,\epsilon) \neq \emptyset \text{ for all } \epsilon > 0\}$ is the closure of the set S and $B(x,\epsilon) = \{y \in \mathbb{R} : |y - x| < \epsilon\}$.

1. F_D is non-decreasing on D.

2. $\lim_{x\to-\infty} F_D(x) = 0$, $\lim_{x\to\infty} F_D(x) = 1$ for $x \in D$.

Now, for all $x \in \mathbb{R}$ define

$$F(x) := \inf_{y>x, y\in D} F_D(y) = \lim_{y\downarrow x, y\in D} F_D(y).$$

Then, F is a right continuous distribution function. Thus, any two right continuous functions that coincide on a dense set D, coincide on $\mathbb{R}$.

Proof. Let $x \in \mathbb{R}$. Since D is dense in $\mathbb{R}$, for all $\delta > 0$ there exists $x' \in D$ such that $x' \in B(x, \delta)$. Take $x' > x$ and note that by right-continuity and monotonicity of F_D, there exists $\epsilon > 0$ such that

$$F_D(x') - \lim_{y\downarrow x, y\in D} F_D(y) = F_D(x') - F(x) \le \epsilon \implies F_D(x') \le F(x) + \epsilon \tag{5.10}$$

For $y \in (x, x')$, and since by definition $F(y) = \inf_{z>y, z\in D} F_D(z)$

$$F(y) \le F_D(x'). \tag{5.11}$$

Thus, equations (5.10) and (5.11) give $F(y) \le F(x) + \epsilon$ for all $y \in (x, x')$. Consequently, as $y \downarrow x$, $\lim_{y\downarrow x} F(y) \le F(x)$. But monotonicity of F gives

$$\lim_{y\downarrow x} F(y) \ge F(x).$$

Thus, the last two inequalities give $F(x) = \lim_{y\downarrow x} F(y)$, establishing right-continuity of F.

■

The next theorem establishes uniqueness of weak limits of distribution functions.

Theorem 5.12. *If $F_n \implies F$ and $F_n \implies G$, then $F = G$.*

Proof. By De Morgan's Laws $C(F)^c \cup C(G)^c = (C(F) \cap C(G))^c = \mathbb{R} - (C(F) \cap C(G))$, which implies that $C(F) \cap C(G) = \mathbb{R} - (C(F)^c \cup C(G)^c)$, where $C(F)^c \cup C(G)^c$ is a countable set. Now, if $x \in C(F) \cap C(G)$, $F_n(x) \to F(x)$ and $F_n(x) \to G(x)$, hence $F = G$ in $C(F) \cap C(G)$, since limits are unique. But note that $C(F) \cap C(G)$ is dense in $\mathbb{R}$. To see this, let $C \subset \mathbb{R}$, C countable. For each $x \in \mathbb{R}$ ($x \in C$ or not), $B(x, \epsilon)$ contains uncountable many points. Hence, for all $x \in \mathbb{R}$, the set $(\mathbb{R} - C) \cap B(x; \epsilon)$ is nonempty for all $\epsilon > 0$, so $x \in \overline{\mathbb{R} - C}$. Thus $\mathbb{R} - C \subseteq (\mathbb{R} - C) \cup C = \mathbb{R} \subseteq \overline{\mathbb{R} - C}$. Thus, F and G coincide on a dense set of $\mathbb{R}$. But since any two distribution functions coinciding on a dense set of $\mathbb{R}$ coincide everywhere, $F = G$ $\forall x \in \mathbb{R}$. ■

Theorem 5.13. *Let X_n, Y_n, W_n, X, Y be random variables defined on $(\Omega, \mathcal{F}, P)$.*

1. $X_n - Y_n \xrightarrow{p} 0$, $Y_n \xrightarrow{d} Y \implies X_n \xrightarrow{d} Y$

2. $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$

3. $X_n \xrightarrow{d} c \implies X_n \xrightarrow{p} c$ *where c is a constant*

4. $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} a$, $W_n \xrightarrow{p} b$ *where a, b are constant, then* $Y_n X_n + W_n \xrightarrow{d} aX + b$, *if* $a \neq 0$.

Proof. 1. $A_n = \{\omega : |X_n - Y_n| < \epsilon\}$, $B_n = \{\omega : X_n \leq x\}$, $C_n = \{\omega : Y_n \leq x + \epsilon\}$, $D_n = \{\omega : Y_n > x - \epsilon\}$ for any $\epsilon > 0$ and $x \in C(F_Y)$. Then,

$$\begin{aligned} F_{X_n}(x) &= P(\{\omega : X_n(\omega) \leq x\}) = P(B_n) = P(B_n \cap A_n) + P(B_n \cap A_n^c) \\ 1 - F_{X_n}(x) &= P(B_n^c) = P(B_n^c \cap A_n) + P(B_n^c \cap A_n^c). \end{aligned}$$

Now, $B_n \cap A_n = \{\omega : X_n \leq x \text{ and } |X_n - Y_n| < \epsilon\} = \{\omega : X_n \leq x \text{ and } X_n - \epsilon < Y_n < X_n + \epsilon\} \subset \{\omega : Y_n \leq x + \epsilon\} = C_n$. $B_n^c \cap A_n = \{\omega : X_n > x \text{ and } X_n - \epsilon < Y_n < X_n + \epsilon\} \subset \{\omega : x - \epsilon < Y_n\} = D_n$. Thus,

1. $F_{X_n}(x) = P(B_n) \leq P(C_n) + P(A_n^c) = F_{Y_n}(x+\epsilon) + P(A_n^c)$

2. $1 - F_{X_n}(x) = P(B_n^c) \leq P(D_n) + P(A_n^c) = 1 - F_{Y_n}(x-\epsilon) + P(A_n^c)$, or $F_{X_n}(x) \geq F_{Y_n}(x-\epsilon) - P(A_n^c)$.

That is,

$$F_{Y_n}(x-\epsilon) - P(A_n^c) \leq F_{X_n}(x) \leq F_{Y_n}(x+\epsilon) + P(A_n^c).$$

Since $x \in C(F_Y)$ and $P(A_n^c) \to 0$ as $n \to \infty$ we have that as $\epsilon \to 0$,

$$F_Y(x) \leq \liminf F_{X_n}(x) \leq \limsup F_{X_n}(x) \leq F_Y(x).$$

Hence, $\lim F_{X_n}(x)$ exists and $\lim F_{X_n}(x) = F_Y(x)$.

2. In 1. let $Y_n = X$.

3. $\{\omega : |X_n - c| > \epsilon\} = \{\omega : X_n > c+\epsilon \text{ or } X_n < c-\epsilon\} = \{\omega : X_n > c+\epsilon\} \cup \{\omega : X_n < c-\epsilon\}$ and

$$\begin{aligned} P(\{\omega : |X_n - c| > \epsilon\}) &= P(\{\omega : X_n > c+\epsilon\}) + P(\{\omega : X_n < c-\epsilon\}) \\ &= 1 - F_{X_n}(c+\epsilon) + F_{X_n}(c-\epsilon). \end{aligned}$$

Since $X_n \xrightarrow{d} c$, $F_c(x) = 0$ for all $x < c$ and $F_c(x) = 1$, for all $x \geq c$. Hence, $\lim_{n\to\infty} P(\{\omega : |X_n - c| > \epsilon\}) = 0$.

4. $W_n - b = Y_nX_n + W_n - Y_nX_n - b = Y_nX_n + W_n - (Y_nX_n + b) \xrightarrow{p} 0$ by assumption. By 1. it suffices to show that $Y_nX_n + b \xrightarrow{d} aX + b$. $Y_nX_n + b - (aX_n + b) = (Y_n - a)X_n$. If $(Y_n - a)X_n \xrightarrow{p} 0$, then it suffices to show that $aX_n + b \xrightarrow{d} aX + b$. Now, let $G_n = F_{aX_n+b}$, that is

$$\begin{aligned} G_n(x) = P(\{\omega : aX_n + b \leq x\}) &= P(\{\omega : aX_n \leq x - b\}) \\ &= P\left(\{\omega : X_n \leq \frac{x-b}{a}\}\right) \\ &= F_{X_n}\left(\frac{x-b}{a}\right). \end{aligned}$$

Then, $F_{X_n}(\frac{x-b}{a}) \to F_X(\frac{x-b}{a})$ for $\frac{x-b}{a} \in C(F_X)$. $F_X(\frac{x-b}{a}) = P(\{\omega : X \leq \frac{x-b}{a}\}) = P(aX + b \leq x) = F_{aX+b}(x)$. So, $aX_n + b \xrightarrow{d} aX + b$. We now show that $(Y_n - a)X_n = C_nX_n \xrightarrow{p} 0$. Let $c > 0$. If $-c, c \in C(F_X)$, $P(|X_n| > c) \to P(|X| > c)$. That is, $\forall \epsilon > 0$, $\exists N_\epsilon$ such that $n \geq N_\epsilon$, $-\epsilon \leq P(|X_n| > c) - P(|X| > c) \leq \epsilon$ or $P(|X| > c) - \epsilon \leq P(|X_n| > c) \leq P(|X| > c) + \epsilon$. Choose c such that $P(|X_n| > c) < \delta$, then $P(|X_n| > c) < \delta + \epsilon$. Since $Y_n - a \xrightarrow{p} 0$ and $P(|X_n| > c) < \delta + \epsilon$, $C_nX_n \xrightarrow{p} 0$. ■

Chapter 6

Laws of large numbers

We first discuss the notion of "tail equivalence" of a sequence of random variables. Here, the Borel-Cantelli Lemma is very useful. Recall that it says that if $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of events with $\sum_{n=1}^{\infty} P(E_n) < \infty$, then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

Definition 6.1. *Two sequences of random variables $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are tail equivalent if*

$$\sum_{n=1}^{\infty} P\left(\{\omega : X_n(\omega) \neq Y_n(\omega)\}\right) = \sum_{n=1}^{\infty} P(\{\omega : X_n(\omega) - Y_n(\omega) \neq 0\}) = \sum_{n=1}^{\infty} P(A_n) < \infty,$$

where $A_n = \{\omega : X_n(\omega) - Y_n(\omega) \neq 0\}$.

Theorem 6.1. *Suppose $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are tail equivalent. Then,*

1. *$\sum_{n=1}^{\infty}(X_n - Y_n)$ converges almost surely,*
2. *$\sum_{n=1}^{\infty} X_n$ converges as $\iff$ $\sum_{n=1}^{\infty} Y_n$ converges as,*
3. *If there exists $a_n \to \infty$ and if there exists a random variable X such that $a_n^{-1}\sum_{j=1}^{n} X_j \overset{as}{\to} X$, then $a_n^{-1}\sum_{j=1}^{n} Y_j \overset{as}{\to} X$.*

Proof. 1. By tail equivalence and the Borel-Cantelli Lemma $P\left(\limsup_{n\to\infty} A_n\right) = 0$. Now, recall that

$$\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m = \cap_{n=1}^{\infty} C_n.$$

Consequently,

$$(\limsup_{n\to\infty} A_n)^c = (\cap_{n=1}^{\infty} C_n)^c = \cup_{n=1}^{\infty} C_n^c = \cup_{n=1}^{\infty}(\cup_{m=n}^{\infty} A_m)^c$$
$$= \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c = \liminf_{n\to\infty} A_n^c.$$

Thus,

$$P\left(\liminf_{n\to\infty}\{\omega : X_n(\omega) = Y_n(\omega)\}\right) = P\left(\liminf_{n\to\infty} A_n^c\right) = 1 - P\left(\limsup_{n\to\infty}\ A_n\right)$$
$$= P\left(\liminf_{n\to\infty}\{\omega : X_n(\omega) - Y_n(\omega) = 0\}\right) = 1.$$

Note that $\liminf\limits_{n\to\infty} A_n^c = \{\omega : \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}$. Thus, $P\left(\{\omega : \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}\right) = 1$. Hence, there exists a set of ω's which occurs with probability 1, and in this set $X_n(\omega) = Y_n(\omega)$ for all but finitely many n. That is, for $\omega \in \{\omega : \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}$ there are only finitely many n for which $I_{\{X_n(\omega)\neq Y_n(\omega\}}(\omega) = 1$. That is, there exists $N(\omega)$ such that for all $n > N(\omega), I_{\{X_n(\omega)\neq Y_n(\omega)\}}(\omega) = 0$. Hence, in this same set,

$$\sum_{n=1}^{\infty} X_n(\omega) - \sum_{n=1}^{\infty} Y_n(\omega) = \sum_{n=1}^{N(\omega)}(X_n(\omega) - Y_n(\omega)) < \infty \text{ with probability 1 (w.p.1).}$$

Hence, $\sum_{n=1}^{\infty}(X_n(\omega) - Y_n(\omega))$ converges almost surely.

2. $\sum_{n=1}^{\infty} Y_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega) + \sum_{n=1}^{\infty} Y_n(\omega) - \sum_{n=1}^{\infty} X_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega) - \sum_{n=1}^{\infty}(X_n(\omega) - Y_n(\omega))$. But both terms on the right-hand side of the last equality converge almost surely. Hence, $\sum_{n=1}^{\infty} Y_n(\omega) < \infty\ as.$

3.

$$\frac{1}{a_n}\sum_{j=1}^{n} Y_j(\omega) = \frac{1}{a_n}\sum_{j=1}^{n}[Y_j(\omega) - X_j(\omega) + X_j(\omega)]$$
$$= \frac{1}{a_n}\sum_{j=1}^{n}(Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n}\sum_{j=1}^{n} X_j(\omega)$$
$$= \frac{1}{a_n}\sum_{j=1}^{N-1}(Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n}\sum_{j=N}^{n}(Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n}\sum_{j=1}^{n} X_j(\omega).$$

As $n \to \infty$ the last term converges as to $X(\omega)$ by assumption. The second term converges to zero since $Y_j(\omega)$ and $X_j(\omega)$ are tail equivalent (and by 1), and the first term goes to 0 as $a_n \to \infty$. Hence, $\frac{1}{a_n}\sum_{j=1}^n Y_j(\omega) \stackrel{as}{\to} X(\omega)$.

■

The following concepts and notation will be useful.

Definition 6.2. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$ and $\{s_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$. We write,*

1. *$X_n = O_p(s_n)$ if for all $\epsilon > 0$ and $n \in \mathbb{N}$, there exists $B_\epsilon > 0$ such that*

$$P\left(\left\{\omega : \frac{|X_n(\omega)|}{s_n} > B_\epsilon\right\}\right) < \epsilon$$

2. *$X_n = o_p(s_n)$ if $\frac{X_n}{s_n} \stackrel{p}{\to} 0$.*

Theorem 6.2. *(General Law of Large Numbers) Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of independent random variables defined on $(\Omega, \mathcal{F}, P)$ and $S_n = \sum_{j=1}^n X_j$. If*

1. *$\sum_{j=1}^n P(\{\omega : |X_j(\omega)| > n\}) \to 0$ as $n \to \infty$,*

2. *$\frac{1}{n^2}\sum_{j=1}^n E(X_j^2 I_{\{\omega:|X_j|\leq n\}}) \to 0$ as $n \to \infty$,*

then $\frac{S_n}{n} - \frac{1}{n}\sum_{j=1}^n E(X_j I_{\{\omega:|X_j|\leq n\}}) \stackrel{p}{\to} 0$.

Proof. Let $T_{n,j}(\omega) = X_j(\omega) I_{\{\omega:|X_j|\leq n\}}$ and $S_n' = \sum_{j=1}^n T_{n,j}$. Note that $\{\omega : X_j(\omega) \neq T_{n,j}(\omega)\} = \{\omega : |X_j(\omega)| > n\}$ and by assumption $\sum_{j=1}^n P(\{\omega : T_{n,j}(\omega) \neq X_j(\omega)\}) \to 0$ as $n \to \infty$. Note also that

$$|S_n - S_n'| = \left|\sum_{j=1}^n X_j - \sum_{j=1}^n T_{n,j}\right| \leq \sum_{j=1}^n |X_j - T_{n,j}|.$$

Thus, for all $\epsilon > 0$,

$$\{\omega : |S_n - S_n'| > \epsilon\} \subseteq \left\{\omega : \sum_{j=1}^n |X_j - T_{n,j}| > \epsilon\right\} \subseteq \cup_{j=1}^n \{\omega : |X_j(\omega) - T_{n,j}(\omega)| > \epsilon/n\}.$$

Consequently,

$$P(\{\omega : |S_n - S_n'| > \epsilon\}) \leq \sum_{j=1}^{n} P\left(\{\omega : |X_j(\omega) - T_{n,j}(\omega)| > \epsilon/n\}\right) = \sum_{j=1}^{n} P\left(\{\omega : |X_j| > n\}\right).$$

Taking limits on both sides as $n \to \infty$, we have that $S_n - S_n' \xrightarrow{p} 0$ since by assumption 1 $\sum_{j=1}^{n} P\left(\{\omega : |X_j| > n\}\right) \to 0$.

Now, since $\{X_n\}_{n=1,2,\cdots}$ is an independent sequence $E\left((T_{n,k} - E(T_{n,k}))(T_{n,l} - E(T_{n,l}))\right) = 0$ and consequently $V(S_n') = \sum_{j=1}^{n} V(T_{n,j}) \leq \sum_{j=1}^{n} E(T_{n,j}^2)$. Note also that for given n

$$E(T_{n,j}^2) = \int_{\Omega} X_j^2 I_{\{\omega:|X_j| \leq n\}} dP \leq n^2 \int_{\Omega} dP = n^2.$$

Consequently, since $V(S_n')$ exists for every n, and by Chebyshev's Inequality,

$$P\left(\left\{\omega : \left|\frac{S_n' - E(S_n')}{n}\right| > \epsilon\right\}\right) \leq \frac{V(S_n')}{n^2\epsilon^2} \leq \frac{1}{n^2\epsilon^2} \sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|<n\}}\right).$$

Taking limits on both sides as $n \to \infty$ and by the assumption that $\frac{1}{n^2}\sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|<n\}}\right) \to 0$, we have $\frac{S_n'}{n} - \frac{E(S_n')}{n} \xrightarrow{p} 0$. Now, since

$$\frac{S_n}{n} - E\left(\frac{S_n'}{n}\right) = \frac{S_n}{n} - \frac{S_n'}{n} + \frac{S_n'}{n} - E\left(\frac{S_n'}{n}\right)$$

we can immediately conclude that $\frac{S_n}{n} - E\left(\frac{S_n'}{n}\right) = o_p(1)$. Finally, from the definition of S_n' we have that $\frac{S_n}{n} - \frac{1}{n}\sum_{j=1}^{n} E(X_j I_{\{\omega:|X_j|\leq n\}}) = o_p(1)$. ■

We now provide examples where conditions 1 and 2 in the statement of Theorem 6.2 hold.

Example 6.1. *Let $\{X_n\}_{n=1,2,\cdots}$ be an independent and identically distributed sequence of random variables with $E(X_n) = \mu$, $E(X_n^2) \leq C < \infty$. Then, we verify condition 1 by noting that the identical distribution assumption and Markov's Inequality*

$$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_1| > n) \leq n\frac{E(X_1^2)}{n^2} = \frac{1}{n}E(X_1^2) \leq \frac{C}{n}.$$

Taking limits on both sides as $n \to \infty$ gives $\lim_{n\to\infty} \sum_{j=1}^{n} P(|X_j| > n) = 0$. For condition 2, note that by the identical distribution assumption

$$\frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{|X_j|\le n\}}) = \frac{1}{n}E(X_1^2 I_{\{|X_1|\le n\}}) \le \frac{1}{n}E(X_1^2) \le \frac{C}{n}.$$

Again, taking limits on both sides as $n \to \infty$ gives $\lim_{n\to\infty} \frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{|X_j|\le n\}}) = 0$. Finally, observe that

$$\frac{\sum_{j=1}^{n} E(X_j I_{\{|X_j|\le n\}})}{n} = E(X_1 I_{\{|X_1|\le n\}}) \to E(X_1) = \mu$$

as $n \to \infty$ by Lebesgue's dominated convergence theorem. Thus, $\frac{1}{n}S_n \xrightarrow{p} \mu$.

Example 6.2. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an independent and identically distributed sequence with $E(|X_1|) \le C < \infty$ and let $E(X_1) = \mu$. For condition 1, note that*

$$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_1| > n) = E(nI_{\{\omega:|X_1|>n\}}).$$

But since $nI_{\{\omega:|X_1|>n\}} \le |X_1| I_{\{\omega:|X_1|>n\}}$, we have that

$$\sum_{j=1}^{n} P(|X_j| > n) \le E(|X_1| I_{\{\omega:|X_1|>n\}})$$

Consequently, $\lim_{n\to\infty} \sum_{j=1}^{n} P(|X_j| > n) \le \lim_{n\to\infty} E(|X_1| I_{\{\omega:|X_1|>n\}})$. And since $E(|X_1|) < C$, $\lim_{n\to\infty} E(|X_1| I_{\{\omega:|X_1|>n\}}) = 0$.

For condition 2, note that by the identical distribution assumption

$$\begin{aligned}\frac{1}{n^2}\sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|\le n\}}\right) &= \frac{1}{n}E\left(X_j^2 I_{\{\omega:|X_j|\le n\}}\right)\\ &= \frac{1}{n}\left(E\left(X_j^2 I_{\{\omega:|X_j|\le \epsilon\sqrt{n}\}}\right) + E\left(X_j^2 I_{\{\omega:\epsilon\sqrt{n}\le|X_j|\le n\}}\right)\right) \text{ for any } \epsilon \in (0,1)\end{aligned}$$

Since $E(X_j^2 I_{\{\omega:|X_j|\le\epsilon\sqrt{n}\}}) = \int_\Omega X_j^2 I_{\{\omega:|X_j|\le\epsilon\sqrt{n}\}} dP \le n\epsilon^2 \int_\Omega dP = n\epsilon^2$, we have

$$\begin{aligned}\frac{1}{n^2}\sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|\le n\}}\right) &\le \epsilon^2 + \frac{1}{n}E\left(|X_j||X_j| I_{\{\omega:\epsilon\sqrt{n}\le|X_j|\le n\}}\right)\\ &\le \epsilon^2 + \frac{1}{n}E(n|X_j| I_{\{\omega:\epsilon\sqrt{n}\le|X_j|\le n\}}))\\ &\le \epsilon^2 + E(|X_j| I_{\{\omega:\epsilon\sqrt{n}\le|X_j|\}})\end{aligned}$$

Taking limits on both sides as $n \to \infty$, and noting that $E(|X_1|) < C$, we have that

$$\lim_{n\to\infty} E(|X_j| I_{\{\omega:\epsilon\sqrt{n}\leq |X_j|\}}) = 0.$$

And, since ϵ can be made arbitrarily small, $\lim_{n\to\infty} \frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{\omega:|X_j|\leq n\}}) = 0$. Consequently, $\frac{S_n}{n} - E(X_1 I_{\{\omega:|X_1|\leq n\}}) \xrightarrow{p} 0$. Lastly, note that

$$\lim_{n\to\infty}\left(\int_\Omega X_1 dP - \int_\Omega X_1 I_{\{|X_1|\leq n\}} dP\right) = \int_\Omega X_1 dP - \lim_{n\to\infty}\int_\Omega X_1 I_{\{|X_1|\leq n\}} dP = E(X_1) - E(X_1) = 0$$

by the previous example. Hence,

$$\frac{S_n}{n} - E(X_1) = \frac{S_n}{n} + E(X_1 I_{\{|X_1|\leq n\}}) - E(X_1 I_{\{|X_1|\leq n\}}) - E(X_1) = o_p(1) + o(1) = o_p(1).$$

.

Example 6.3. *Suppose $\{X_n\}_{n=1,2,\ldots}$ is an independent and identically distributed sequence with $\lim_{x\to\infty} xP(|X_1| > x) = 0$. For condition 1, given the identically distributed assumption, we have*

$$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_j| > n) \to 0$$

by assumption. For condition 2, note that

$$\begin{aligned}
\frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{\omega:|X_j|\leq n\}}) &= \frac{1}{n} E(X_1^2 I_{\{\omega:|X_j|\leq n\}}) = \frac{1}{n}\int_{|x|\leq n} x^2 dF_{X_1}(x) \\
&= \frac{2}{n}\int_{|x|\leq n}\left(\int_0^{|x|} s ds\right) dF_{X_1}(x) = \frac{2}{n}\int_0^n s\left(\int_{s<|x|\leq n} dF_{X_1}(x)\right) ds \\
&= \frac{2}{n}\int_0^n s(P(|X_1|\leq n) - P(|X_1| < s)) ds \\
&= \frac{2}{n}\int_0^n s(1 - P(|X_1| > n) - 1 + P(|X_1| \geq s)) ds \\
&= \frac{2}{n}\int_0^n s(P(|X_1|\geq s) - P(|X_1| > n)) ds \\
&= \frac{2}{n}\int_0^n \tau(s) ds - 2P(|X_1| > n)\frac{1}{n}\int_0^n s ds, \text{ where } \tau(s) = sP(|X_1| > s) \\
&= \frac{1}{n}\int_0^n \tau(s) ds - 2P(|X_1| > n)\frac{1}{n}\frac{n^2}{2} \\
&= \frac{1}{n}\int_0^n \tau(s) ds - nP(|X_1| > n) = \frac{1}{n}\int_0^n \tau(s) ds - \tau(n).
\end{aligned}$$

Since, $\tau(n) \to 0$ as $n \to \infty$, we have that for all $\epsilon > 0$ there exists N_ϵ such that if $n > N_\epsilon$, $\tau(n) \leq \epsilon$. Consequently,

$$\frac{1}{n}\int_0^n \tau(s)ds = \frac{1}{n}\int_0^{N_\epsilon} \tau(s)ds + \frac{1}{n}\int_{N_\epsilon}^n \tau(s)ds \leq \frac{1}{n}\int_0^{N_\epsilon} \tau(s)ds + \epsilon.$$

Taking limits on both sides as $n \to \infty$ gives $\frac{1}{n}\int_0^n \tau(s)ds \to 0$. Then, $\frac{S_n}{n} - E(X_1 I_{|X_1| \leq n}) \xrightarrow{p} 0$.

The following is called Markov's Law of Large Numbers.

Theorem 6.3. *(Markov's LLN) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables with $E(X_n) = \mu_n$. If for some $\delta > 0$ we have $\sum_{n=1}^\infty \frac{E|X_n - \mu_n|^{1+\delta}}{n^{1+\delta}} < \infty$,*

$$S_n - \frac{1}{n}\sum_{i=1}^n \mu_i \xrightarrow{p} 0.$$

Proof. Exercise. ■

We call Theorem 6.2 a Weak Law of Large Numbers (WLLN). A stronger result for independent and identically distributed random sequences is Kolmogorov's Strong LLN. We state it without proof.

Theorem 6.4. *(Kolmogorov's SLLN) Let $\{X_n\}_{n\geq 1}$ be a sequence of IID random variables and set $S_n = \sum_{i=1}^n X_i$. There exists a real number C such that*

$$\frac{S_n}{n} \xrightarrow{as} C \iff E(|X_1|) < \infty.$$

In this case, $C = E(X_1)$.