Chapter 7

Convergence of random variables

7.1 Convergence almost surely and in probability

Since random variables are measurable functions from a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B})$, i.e., $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$, the most natural way to define convergence of a sequence $\{X_n\}_{n\in\mathbb{N}}$ is pointwise. In this case, we say that the sequence X_n converges to X for some $\omega \in \Omega$ if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

That $X(\omega)$ is a random variable follows from Theorem 3.6. If the limit holds for all $\omega \in \Omega$ we say that X_n converges to X on Ω and write $X_n \to X$ on Ω. A weaker convergence concept requires

$$P\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

Note that $\{\omega : \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\}$ must be an event $(\neq \Omega)$ for the statement to make sense. In this case we say that X_n converges to X almost surely (or almost everywhere) and we write $X_n \overset{as}{\to} X$ (or $X_n \overset{ae}{\to} X$). Alternatively, we can require the the existence of a set $N \in \mathcal{F}$ with $P(N) = 0$ where if $\omega \in N^c$

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Note that since N is an event, N^c is an event and $P(N^c) = 1$ since $P(N) = 0$ and $P(\Omega) = 1$. Hence, we give the following definition.

Definition 7.1. *(Convergence as)* Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. Then, if there exists $N \in \mathcal{F}$ with $P(N) = 0$ such that $\lim_{n\to\infty} X_n(\omega)$ exists for all $\omega \in N^c$, we denote this limit by $X(\omega)$ and say that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ almost surely (as) and write $X_n \overset{as}{\to} X$.

The limit statement in the definition is equivalent to stating that for all $\epsilon > 0$ there exists $N(\epsilon) \in \mathbb{N}$ such that for all $n \geq N(\epsilon)$,

$$P\left(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0.$$

Letting $E_n(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$, we see that

$$\begin{aligned} P\left(\underset{j\geq n}{\cup} E_j(\epsilon)\right) &\leq \sum_{j\geq n} P\left(E_j(\epsilon)\right) \text{ by sub-additivity of } P \\ &= 0 \text{ since } P(E_j(\epsilon)) = 0 \text{ for } j \geq n. \end{aligned}$$

Recall that $\cap_{n=1}^{\infty} \underset{j\geq n}{\cup} E_j(\epsilon) = \limsup_{n\to\infty} E_n(\epsilon)$, and

$$\begin{aligned} P\left(\limsup_{n\to\infty} E_n(\epsilon)\right) &= \lim_{n\to\infty} P\left(\underset{j\geq n}{\cup} E_j(\epsilon)\right) \text{ by continuity of } P \\ &= 0. \end{aligned}$$

Hence, $X_n \overset{as}{\to} X$ is often stated as $P\left(\limsup_{n\to\infty} \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$ for all $\epsilon > 0$.

What follows is an example of a sequence of random variables that converges to 0 *as*.

Example 7.1. *Let* $(\Omega = [0,1], \mathcal{B}_{[0,1]}, \lambda)$ *where* λ *is Lebesgue measure.*

$$X_n(\omega) = \begin{cases} n & \text{if } 0 \leq \omega \leq 1/n \\ 0 & \text{if } 1/n < \omega \leq 1 \end{cases}$$

Let $N = \{0\}$ *and note that* $\lambda(N) = 0$. *If* $\omega \in N^c$ *then* $X_n(\omega) \to 0$ *as* $n \to \infty$, *but* $X_n(\omega) \not\to 0$ *everywhere on* Ω *since at* $\omega = 0$, $X_n(\omega) \to \infty$.

An even less demanding convergence concept is that of convergence in probability (convergence *ip* or convergence in measure *im*), which is given in the following definition.

Definition 7.2. *Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of random variables and* X *be a random variable defined in the same probability space* $(\Omega, \mathcal{F}, P)$*. We say that* $X_n \overset{p}{\to} X$ *if for all* $\epsilon > 0$

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0.$$

Alternatively, we can state that for all $\epsilon > 0$ and $\delta > 0$ there exists $N(\epsilon, \delta) \in \mathbb{N}$ such that for all $n \geq N(\epsilon, \delta)$, $P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) < \delta$.

Theorem 7.1. *Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of random variables and* X *be a random variable defined in the same probability space* $(\Omega, \mathcal{F}, P)$*. Then,* $X_n \overset{as}{\to} X \implies X_n \overset{p}{\to} X$*.*

Proof. Let $E_n(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$ for any $\epsilon > 0$. $X_n \overset{as}{\to} X$ implies that there exists a natural number $N(\epsilon)$ such that for all $n \geq N(\epsilon)$ we have $P(E_n(\epsilon)) = 0$. Then, from the comments following Definition 7.1

$$\begin{aligned}
P\left(\limsup_{n\to\infty} E_n(\epsilon)\right) = 0 = P\left(\lim_{n\to\infty} \cup_{m=n}^{\infty} E_m(\epsilon)\right) \\
&= \lim_{n\to\infty} P(\cup_{m=n}^{\infty} E_m(\epsilon)) \text{ by continuity of } P \\
&\geq \lim_{n\to\infty} P(E_n(\epsilon)).
\end{aligned}$$

Consequently, $\lim_{n\to\infty} P(E_n(\epsilon)) = 0$. ■

The following theorem, known as the Borel-Cantelli Lemma is the main device used to establish almost sure convergence.

Theorem 7.2. *(Borel-Cantelli Lemma) Let* $\{E_n\}_{n\in\mathbb{N}}$ *be a sequence of events. If*

$$\sum_{n=1}^{\infty} P(E_n) < \infty$$

then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

Proof.

$$\begin{aligned}
P\left(\limsup_{n\to\infty} E_n\right) &= P\left(\lim_{n\to\infty} \cup_{m\geq n} E_m\right) \\
&= \lim_{n\to\infty} P\left(\cup_{m\geq n} E_m\right) \text{ by continuity of } P \\
&\leq \limsup_{n\to\infty} \sum_{m=n}^{\infty} P\left(E_m\right) \text{ by sub-additivity of } P \\
&= 0 \text{ since } \sum_{n=1}^{\infty} P(E_n) < \infty \text{ implies } \sum_{m=n}^{\infty} P\left(E_m\right) \to 0 \text{ as } n \to \infty.
\end{aligned}$$

■

Theorem 7.3. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables and X be a random variable defined in the same probability space $(\Omega, \mathcal{F}, P)$.*

1. $X_n \stackrel{p}{\to} X \iff X_r - X_s \stackrel{p}{\to} 0$ *as n, r, $s \to \infty$ (Cauchy in probability)*

2. $X_n \stackrel{p}{\to} X \iff$ *each subsequence X_{n_k} contains a further subsequence $\{X_{n_{k(i)}}\} \stackrel{as}{\to} X$.*

Proof. 1. ($\implies$) $|X_r - X_s| = |X_r - X + X - X_s| \leq |X_r - X| + |X - X_s|$. For all $\epsilon > 0$, $\{\omega : |X_r - X_s| > \epsilon\} \subset \{\omega : |X_r - X| + |X - X_s| > \epsilon\} \subset \{\omega : |X_r - X| > \epsilon/2\} \cup \{\omega : |X_s - X| > \epsilon/2\}$. Consequently,

$$P(\{\omega : |X_r - X_s| > \epsilon\}) \leq P(\{\omega : |X_r - X| > \epsilon/2\}) + P(\{\omega : |X_s - X| > \epsilon/2\}). \tag{7.1}$$

Taking limits on both sides of the inequality as $r, s \to \infty$ and given that $X_n \stackrel{p}{\to} X$ we have that $P(\{\omega : |X_r - X_s| > \epsilon\}) \to 0$.

($\impliedby$) Let $\{X_{n(j)}\}_{j\in\mathbb{N}}$ be a subsequence of $\{X_n\}_{n\in\mathbb{N}}$. If $X_{n(j)} \stackrel{as}{\to} X$, then by equation (7.1)

$$P(\{\omega : |X_n - X| > \epsilon\}) \leq P(\{\omega : |X_n - X_{n(j)}| > \epsilon/2\}) + P(\{\omega : |X_{n(j)} - X| > \epsilon/2\}).$$

Using the fact that $\{X_n\}_{n\in\mathbb{N}}$ is Cauchy in probability $P(\{\omega : |X_n - X_{n(j)}| > \epsilon/2\}) \to 0$ as $n, n(j) \to \infty$. Also, since $X_{n(j)} \stackrel{as}{\to} X$ implies $X_{n(j)} \stackrel{p}{\to} X$ and we have that $P(\{\omega :$

$|X_{n(j)} - X| > \epsilon/2\}) \to 0$ as $n(j) \to \infty$. Thus, it suffice to show that there exists a subsequence $\{X_{n(j)}\}_{j \in \mathbb{N}}$ such that $X_{n(j)} \stackrel{as}{\to} X$. We will construct such sequence.

Let $n(1) = 1$ and define

$$n(j) = \inf\{N : N > n(j-1), P\left(\{\omega : |X_r - X_s| > 2^{-j}\}\right) < 2^{-j}, \text{ for all } r, s \geq N\}.$$

It is possible to define $\{n(j)\}$ because of the assumption that $\{X_n\}_{n \in \mathbb{N}}$ is Cauchy in probability. Also, by construction, $n(1) < n(2) < \cdots$ so that $n(j) \to \infty$. Consequently,

$$P(\{\omega : |X_{n(j)+1} - X_{n(j)}| > 2^{-j}\}) < 2^{-j}$$

and $\sum_{j=1}^{\infty} P(\{\omega : |X_{n(j)+1} - X_{n(j)}| > 2^{-j}\}) < \sum_{j=1}^{\infty} 2^{-j} < \infty$. By the Borel-Cantelli Lemma

$$P\left(\limsup_{j \to \infty} \{\omega : |X_{n(j)+1} - X_{n(j)}| > 2^{-j}\}\right) = 0$$

or

$$P\left(\liminf_{j \to \infty} \{\omega : |X_{n(j)+1} - X_{n(j)}| \leq 2^{-j}\}\right) = 1.$$

Now, $\omega \in \liminf_{j \to \infty} \{\omega : |X_{n(j)+1} - X_{n(j)}| \leq 2^{-j}\}$ means that $\omega \in \{\omega : |X_{n(j)+1} - X_{n(j)}| \leq 2^{-j}\}$ for all j sufficiently large $(j \geq J)$. Hence,

$$\sum_{j \geq J} |X_{n(j)+1}(\omega) - X_{n(j)}(\omega)| \leq \sum_{j \geq J} 2^{-j} = 2 \cdot 2^{-J}$$

Hence, for all $K > J$, $|X_{n(K)} - X_{n(J)}| \leq \sum_{j \geq J} |X_{n(j)+1} - X_{n(j)}| \leq 2 \cdot 2^{-J}$. Thus, as $J \to \infty$, $|X_{n(K)} - X_{n(J)}| \to 0$ establishing that $\{X_{n(j)}\}$ is a Cauchy sequence of real numbers with probability 1. Since $\mathbb{R}$ is complete, i.e., every Cauchy sequence in $\mathbb{R}$ has a limit in $\mathbb{R}$, $\lim_{j \to \infty} X_{n_j}(\omega)$ exists with probability 1. Hence, $X_{n_j}(\omega) \to X(\omega) = \lim\limits_{j \to \infty} X_{n_j}(\omega)$ *as*.

2. ($\Longrightarrow$) Choose a subsequence $\{X_{n(j)}\}$. Then, since $X_n \stackrel{p}{\to} X$, $X_{n(j)} \stackrel{p}{\to} X$ and $X_{n(j)}$ is Cauchy in probability by part 1. Hence, there exists $X_{n(k(i))} \stackrel{as}{\to} X$.

($\Longleftarrow$) Suppose not. If $X_n \stackrel{p}{\not\to} X$ then there exists $X_{n(j)}$ and $\epsilon, \delta > 0$ such that

$$P(\{\omega : |X_{n(j)} - X| > \epsilon\}) \geq \delta. \tag{7.2}$$

But every $X_{n(j)}$ has a subsequence $X_{n(j(i))} \stackrel{as}{\to} X$ and hence $X_{n(j(i))} \stackrel{p}{\to} X$, which contradicts equation (7.2). ■

The following theorem is often called Slutsky's Theorem. It shows that limits in probability and continuous functions can be interchanged.

Theorem 7.4. *(Slutsky's Theorem) If X_n, X are random elements defined on the same probability space and $X_n \stackrel{p}{\to} X$, $g : \mathbb{R}^K \to \mathbb{R}^L$ continuous, then $g(X_n) \stackrel{p}{\to} g(X)$.*

Proof. Recall that g is continuous at X if and only if for all $\epsilon > 0$ there exists $\delta_{\epsilon,X} > 0$ such that whenever $|X_{n,k} - X_k| < \delta_{\epsilon,X}$ for $k = 1, ..., K$, $|g_l(X_n) - g_l(X)| < \epsilon$ for $l = 1, ..., L$. Let $A_{n,k} = \{\omega : |X_{n,k} - X_k| < \delta_{\epsilon,X}\}$ and $A_n = \{\omega : |g_l(X_n) - g_l(X)| < \epsilon\}$ for all l. Note that by continuity $\cap_{k=1}^K A_{n,k} \subset A_n$, which implies that $P(\cap_{k=1}^K A_{n,k}) \leq P(A_n)$. Thus, $1 - P(A_n) \leq 1 - P(\cap_{k=1}^K A_{n,k})$ which implies that $P(A_n^c) \leq P((\cap_{k=1}^K A_{n,k})^c) = P(\cup_{k=1}^K A_{n,k}^c) \leq \sum_{k=1}^K P(A_{n,k}^c)$. Since $X_n \stackrel{p}{\to} X$, $P(A_{n,k}^c) \to 0$ and therefore $P(A_n^c) \to 0$ or $P(A_n) \to 1$. ■

Theorem 7.5. *Let X_n, X be defined in the same probability space such that $X_n \stackrel{p}{\to} X$. If there exist a random variable $0 \leq Y \in \mathcal{L}$ such that $|X_n(\omega)| \leq Y(\omega)$ for all n almost everywhere, then $E(X_n)$, $E(X) < \infty$, and $E(X_n) \to E(X)$.*

Proof. First, note that if $Y \in \mathcal{L}$, $E|X_n| < \infty$ and $X_n \in \mathcal{L}$. Also, $|X| = |X - X_n + X_n| \leq |X_n| + |X_n - X| \leq Y + |X_n - X|$ and $|X| - Y \leq |X_n - X|$. Consequently, for any $\epsilon > 0$ $\{\omega : |X(\omega)| - Y(\omega) > \epsilon\} \subset \{\omega : |X_n - X| > \epsilon\}$ and

$$P(\{\omega : |X(\omega)| - Y(\omega) > \epsilon\}) \leq P(\{\omega : |X_n - X| > \epsilon\}).$$

Taking limits as $n \to \infty$, and using the fact that $X_n \stackrel{p}{\to} X$ we obtain $P(\{\omega : |X(\omega)| - Y(\omega) > \epsilon\}) = 0$. Since this is true for any $\epsilon > 0$ we have $P(\{\omega : |X(\omega)| > Y(\omega)\}) = 0$. Consequently $|X| < Y$ almost everywhere and $E(X) < \infty$.

Since $|E(X_n) - E(X)| \leq \int_\Omega |X_n - X| dP$, we need only show that $\int_\Omega |X_n - X| dP \to 0$ as $n \to \infty$. If $Z_n = X_n - X$, then $|Z_n| \leq |X_n| + |X| < 2Y$ almost everywhere. In addition, for $\epsilon > 0$,

$$E|Z_n| = E\left(|Z_n| I_{\{\omega:|Z_n|\leq\epsilon\}}\right) + E\left(|Z_n| I_{\{\omega:|Z_n|>\epsilon\}}\right) \leq \epsilon + E\left(|Z_n| I_{\{\omega:|Z_n|>\epsilon\}}\right) \leq \epsilon + 2E\left(Y I_{\{\omega:|Z_n|>\epsilon\}}\right).$$

Since $X_n \xrightarrow{p} X$, $P(\{\omega : |Z_n| > \epsilon\}) \to 0$ as $n \to \infty$. Furthermore, since $E(Y) < \infty$, by Theorem 4.9, $E\left(Y I_{\{\omega:|Z_n|>\epsilon\}}\right) < \epsilon$ and $E|Z_n| < 3\epsilon$, completing the proof.■

The requirement that $|X_n(\omega)| \leq Y(\omega)$ for all n almost everywhere may be relaxed. A weaker requirement is given by the following definition.

Definition 7.3. *A sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables defined on $(\Omega, \mathcal{F}, P)$ is said to be uniformly integrable (u.i.) if for every $\epsilon > 0$ there exists $B_\epsilon \in [0, \infty)$ such that*

$$\sup_{n\in\mathbb{N}} E\left(|X_n| I_{\{\omega:|X_n(\omega)|\geq B_\epsilon\}}\right) < \epsilon.$$

Uniform integrability of the sequence $\{X_n\}_{n\in\mathbb{N}}$ is a weaker condition compared to the dominating condition in Theorem 7.5 . Note that if $|X_n(\omega)| \leq Y(\omega)$ for all n almost everywhere on Ω and $E(Y) < \infty$, then

$$|X_n(\omega)| I_{\{\omega:|X_n(\omega)|\geq B_\epsilon\}} \leq Y I_{\{\omega:|X_n(\omega)|\geq B_\epsilon\}} \leq Y I_{\{\omega:|Y(\omega)|\geq B_\epsilon\}}.$$

Hence, $\sup_{n\in\mathbb{N}} E\left(|X_n(\omega)| I_{\{\omega:|X_n(\omega)|\geq B_\epsilon\}}\right) \leq E\left(Y I_{\{\omega:|Y(\omega)|\geq B_\epsilon\}}\right)$. But since $E(|Y|) < \infty$, for any $\epsilon > 0$ there exists $B_\epsilon < \infty$, $E\left(Y I_{\{\omega:|Y(\omega)|\geq B_\epsilon\}}\right) < \epsilon$ and $\{X_n\}_{n\in\mathbb{N}}$ is u.i.

Theorem 7.6. *A sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables defined on $(\Omega, \mathcal{F}, P)$ is uniformly integrable if, and only if,*

1. $\sup_{n\in\mathbb{N}} E|X_n| < \infty$,
2. *for all $\epsilon > 0$, there exists $\delta > 0$ such that for all n, $E(|X_n| I_A) < \epsilon$ for any event A such that $P(A) < \delta$.*

Proof. ($\Longrightarrow$) First, let $A_n(B) = \{\omega : |X_n(\omega)| \geq B\}$ and consider

$$\begin{aligned} E(|X_n|) &= \int_\Omega |X_n| I_{A_n(B)} dP + \int_\Omega |X_n| I_{A_n^c(B)} dP \\ &< \int_\Omega |X_n| I_{A_n(B)} dP + B\,P(A_n^c(B)) \leq \int_\Omega |X_n| I_{A_n(B)} dP + B. \end{aligned}$$

Hence, $\sup\limits_{n\in\mathbb{N}} E(|X_n|) < \sup\limits_{n\in\mathbb{N}} \int_\Omega |X_n| I_{A_n(B)} dP + B$, and by uniform integrability of $\{X_n\}_{n\in\mathbb{N}}$, for any $\epsilon > 0$ there exists $B < \infty$ such that $\sup\limits_{n\in\mathbb{N}} E(|X_n|) < \epsilon + B < \infty$.

Second, let $E \in \mathcal{F}$. Then,

$$E(|X_n| I_E) = E(|X_n| I_{E\cap A_n(B)}) + E(|X_n| I_{E\cap A_n^c(B)}).$$

But $E(|X_n| I_{E\cap A_n(B)}) \leq E(|X_n| I_{A_n(B)})$ and $E(|X_n| I_{E\cap A_n^c(B)}) \leq b \int_\Omega I_{E\cup A_n^c(b)} dP \leq bP(E)$. By uniform integrability of $\{X_n\}_{n\in\mathbb{N}}$, there exists $b > 0$ such that $\sup\limits_{n\in\mathbb{N}} E(|X_n| I_{A_n(b)}) < \epsilon/2$. Furthermore, for any E such that $P(E) < \epsilon/2b$, we have $E(|X_n| I_E) < \epsilon$.

($\Longleftarrow$) By Markov's Inequality

$$P(A_n(b)) < \frac{1}{b} E(|X_n| I_{A_n(b)}) \leq \frac{1}{b} E(|X_n|).$$

Then, $\sup\limits_{n\in\mathbb{N}} P(A_n(b)) \leq \frac{1}{b} \sup\limits_{n\in\mathbb{N}} E(|X_n|) < \infty$ by condition 1. Choose, b such that $\frac{1}{b} \sup\limits_{n\in\mathbb{N}} E(|X_n|) < \delta$, implying that $P(A_n(b)) < \delta$ for all n. Then, by condition 2 it follows that $E(|X_n| I_{A_n(b)}) < \epsilon$. ■

Remark 7.1. *1. The following results follow directly from Theorem 7.3.*

$$X_n \xrightarrow{p} X,\, Y_n \xrightarrow{p} Y \implies X_n + Y_n \xrightarrow{p} X + Y$$

$$X_n \xrightarrow{p} X,\, Y_n \xrightarrow{p} Y \implies X_n Y_n \xrightarrow{p} XY.$$

2. If $E(X_n) = \mu_n < \infty$, $V(X_n) = \sigma_n^2 < \infty$. By Markov's Inequality

$$P(\{\omega : |X_n - \mu_n| \geq \epsilon\}) \leq \sigma_n^2/\epsilon^2.$$

In particular, if $E(X_t) = \mu$ and $V(X_t) = \sigma^2$, letting

$$X_n = \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu),$$

we have $E(X_n) = 0$,

$$V(X_n) = E(X_n^2) = \frac{1}{n^2}\sum_{t=1}^{n} E(X_t - \mu)^2 + \frac{1}{n^2}\sum_{t\neq\tau} E(X_t - \mu)(X_t - \mu).$$

If X_t, X_τ are independent (uncorrelated), $E(X_n^2) = \sigma^2/n$. Then,

$$P(\{\omega : |X_n| \geq \epsilon\}) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Taking limits on both sides,

$$\lim_{n\to\infty} P(\{\omega : |X_n| \geq \epsilon\}) = 0.$$

7.2 Convergence in $\mathcal{L}^p$

Definition 7.4. *Let $X, Y \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ and define $d_p(X,Y) := \|X-Y\|_p = (E(|X-Y|^p))^{1/p}$ for $p \in [1,\infty)$. We say that a sequence $\{X_n\}_{n\in\mathbb{N}} \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ converges to $X \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ in $\mathcal{L}^p$, denoted by $X_n \stackrel{\mathcal{L}^p}{\to} X$, if $d_p(X_n, X) \to 0$ as $n \to \infty$.*

The limit X in Definition 7.4 is not unique, only almost everywhere unique. If X and Y are such that $X_n \stackrel{\mathcal{L}^p}{\to} X$ and $X_n \stackrel{\mathcal{L}^p}{\to} Y$, then by the Minkowski-Riez Inequality

$$\|X - Y\|_p = \|X - X_n + X_n - Y\|_p \leq \|X - X_n\|_p + \|X_n - Y\|_p.$$

Taking limits as $n \to \infty$ we have $\|X - Y\|_p = 0$, which implies that X and Y are equal almost everywhere. We note that d_p is a (semi) metric on $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, induced by the (semi) norm $\|X\|_p = (E(|X|^p))^{1/p}$.

A sequence $\{X_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ is said to be $\mathcal{L}^p$-Cauchy if for all $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n, m \geq N(\epsilon)$ we have $d_p(X_n, X_m) < \epsilon$. Note that if $X_n \stackrel{\mathcal{L}^p}{\to} X$ we have

$$\|X_n - X_m\|_p = \|X_n - X + X - X_m\|_p \leq \|X_n - X\|_p + \|X - X_m\|_p.$$

Hence, as $n, m \to \infty$ we obtain $d_p(X_n, X_m) \to 0$, showing that convergent sequences in $\mathcal{L}^p$ are $\mathcal{L}^p$-Cauchy. The next theorem shows that every $\mathcal{L}^p$-Cauchy sequence converges to an element in $\mathcal{L}^p$, i.e., $\mathcal{L}^p$ is a complete (Banach) space.

Theorem 7.7. *(Riez-Fisher Theorem) The spaces $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ for $p \in [1, \infty)$ are complete.*

Proof. Consider a $\mathcal{L}^p$-Cauchy sequence $\{X_n\}_{n\in\mathbb{N}} \subset \mathcal{L}^p(\Omega, \mathcal{F}, P)$. We need to show that this sequence converges to a limit X in $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. That is, there exists $X \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ such that

$$\|X_n - X\|_p := \left(\int |X_n - X|^p dP\right)^{1/p} \to 0 \text{ as } n \to \infty.$$

Since $\{X_n\}_{n\in\mathbb{N}}$ is $\mathcal{L}^p$-Cauchy , we can find $1 < n(1) < n(2) < \cdots$ such that

$$\|X_{n(k+1)} - X_{n(k)}\|_p \leq \frac{1}{2^k} \text{ for } k = 1, 2, \cdots \tag{7.3}$$

Now, note that if we set $X_{n(0)} := 0$ we have that $X_{n(k+1)} = \sum_{j=0}^{k}(X_{n(j+1)} - X_{n(j)})$ are the partial sums of the series $\sum_{j=0}^{\infty}(X_{n(j+1)} - X_{n(j)})$. Recall that this series converges absolutely if the monotone sequence $\sum_{j=0}^{k} |X_{n(j+1)} - X_{n(j)}|$ converges, and in this case the series converges, that is, $\sum_{j=0}^{k}(X_{n(j+1)} - X_{n(j)})$ converges.

By Minkowski's Inequality and Beppo-Levi's Theorem

$$\begin{aligned}
\|\sum_{j=0}^{\infty} |X_{n(j+1)} - X_{n(j)}|\|_p &\leq \sum_{j=0}^{\infty} \|X_{n(j+1)} - X_{n(j)}\|_p \\
&\leq \|X_{n(1)}\|_p + \sum_{j=1}^{\infty} \frac{1}{2^j} = \|X_{n(1)}\|_p + 1 < \infty \text{ since } X_{n(1)} \text{ is in } \mathcal{L}^p .
\end{aligned}$$

Consequently, $\| \sum_{j=0}^{\infty} |X_{n(j+1)} - X_{n(j)}|\|_p^p < \infty$ and we have that $(\sum_{j=0}^{\infty} |X_{n(j+1)} - X_{n(j)}|)^p < \infty$ almost surely (almost surely real valued) and $\sum_{j=0}^{\infty}(X_{n(j+1)} - X_{n(j)})$ is almost surely (absolutely) convergent.

Letting $X = \sum_{j=0}^{\infty}(X_{n(j+1)} - X_{n(j)})$ we have that

$$\begin{aligned} \|X - X_{n(k)}\|_p &= \|\sum_{j=k}^{\infty} |X_{n(j+1)} - X_{n(j)}|\|_p \\ &\le \sum_{j=k}^{\infty} \|X_{n(j+1)} - X_{n(j)}\|_p \to 0 \text{ as } k \to \infty. \end{aligned}$$

Finally, since

$$\|X_n - X\|_p \le \|X_n - X_{n(k)}\|_p + \|X_{n(k)} - X\|_p.$$

and $\{X_n\}_{n=1,2,\cdots}$ is Cauchy we have the desired result. ■

A complete inner product space is called a Hilbert space. $\mathcal{L}^2$ is a Hilbert space but $\mathcal{L}^p$ for $p \neq 2$ is not, because the Parallelogram Law is not satisfied.

Point-wise convergence of a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables in $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ does not imply convergence in $\mathcal{L}^p$. That is,

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ for all } \omega \in \Omega \not\Rightarrow X_n \xrightarrow{\mathcal{L}^p} X.$$

However, by Lebesgue's Dominated Convergence Theorem, if there exist $0 \le Y \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ such that $|X_n| \le Y$ for all n and $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ exists almost everywhere, then

$$|X_n - X|^p \le (|X_n| + |X|)^p \le 2^p Y^p$$

and $X \in \mathcal{L}_P^p$ and $X_n \xrightarrow{\mathcal{L}^p} X$.

The next theorem shows that convergence in $\mathcal{L}_P^p$ implies convergence in probability.

Theorem 7.8. *Let X, X_n, $n = 1, 2, \cdots$ be random variables defined in the same probability space. If $X_n \xrightarrow{\mathcal{L}^p} X$, then $X_n \xrightarrow{p} X$.*

Proof. First, note that if $h : \mathbb{R} \to [0, \infty)$ and $a \ge 0$, we have $h(X) \ge aI_{h(X)\ge a}$. Then, $E(h(X)) \ge aP(h(X) \ge a)$ which implies that $P(h(X) \ge a) \le \frac{E(h(X))}{a}$. Now, choose $h(x) = |x|^p$ and set $x = X_n - X$. Then, $\{\omega : |X_n - X| \ge a\} = \{\omega : |X_n - X|^p \ge a^p\}$ and

$$P(\{\omega : |X_n - X| \ge a\}) = P(\{\omega : |X_n - X|^p \ge a^p\}) \le \frac{E(|X_n - X|^p)}{a^p}.$$

Taking limits on both sides completes the proof.■

Theorem 7.9. *Suppose that* $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of random variables defined on* $(\Omega, \mathcal{F}, P)$ *is such that* $X_n \overset{p}{\to} X$*, where* X *is defined on the same probability space. Then, the following statements are equivalent,*

1. $\{X_n\}_{n\in\mathbb{N}}$ *is uniformly integrable*

2. $E(|X_n|) < \infty$ *for all* n*,* $E(|X|) < \infty$ *and* $X_n \overset{\mathcal{L}^1}{\to} X$

3. $E(|X_n|) < \infty$ *for all* n*, and* $E(|X_n|) \to E(|X|) < \infty$*.*

Proof. ■

7.3 Convergence in distribution

Let $(\mathbb{R}, \mathcal{B}, d)$ be a metric space with $d(x, y) = |x - y|$ for all $x, y \in \mathbb{R}$ and P, P_n for $n \in \mathbb{N}$ be probability measures defined on $\mathcal{B}$.

Definition 7.5. *The sequence of probability measures* $\{P_n\}_{n\in\mathbb{N}}$ *converges weakly to the measure* P*, denoted by* $P_n \overset{w}{\to} P$ *if*

$$\int_{\mathbb{R}} f dP_n \to \int_{\mathbb{R}} f dP \text{ as } n \to \infty$$

for all $f : \mathbb{R} \to \mathbb{R}$ *that are continuous with* $|f| \le C < \infty$*.*

We note that if F_n and F are the distribution functions associated with P_n and P, we can say that

$$\int_{\mathbb{R}} f dP_n \to \int_{\mathbb{R}} f dP \Longleftrightarrow \int_{\mathbb{R}} f(x) dF_n(x) \to \int_{\mathbb{R}} f(x) dF(x)$$

and we say that $F_n \overset{w}{\to} F$.

Definition 7.6. *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 7.10. *The following convergence statements are equivalent:*

1. $P_n \xrightarrow{w} P$,
2. $\limsup_{n\to\infty} P_n(E) \le P(E)$ *if* $E \in \mathcal{B}$ *is closed,*
3. $\liminf_{n\to\infty} P_n(E) \ge P(E)$ *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 $\varepsilon > 0$, $f(x) = I_E(x)$,

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

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,

$$\int_{\mathbb{R}} f dP_n = \int_{\mathbb{R}} I_E P_n = P_n(E) \le \int_{\mathbb{R}} f_\varepsilon dP_n. \tag{7.4}$$

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 $f_\varepsilon(x) = g(\varepsilon^{-1}|x - E|) \ge 0 = I_E(x)$. Then, taking limits on both sides of equation (7.4) gives

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

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

$$\int_{\mathbb{R}} f_\varepsilon dP \leq \int_{\mathbb{R}} I_{E(\varepsilon)} dP = P(E(\varepsilon)) \tag{7.5}$$

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 (7.4) and (7.5) 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

$$\limsup_{n\to\infty} P_n(E) \leq \limsup_{n\to\infty} P_n(\bar{E}) \leq P(\bar{E}), \tag{7.6}$$

$$\liminf_{n\to\infty} P_n(E) \geq \liminf_{n\to\infty} P_n(int(E)) \geq P(int(E)). \tag{7.7}$$

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., $P_n \Longrightarrow P$.

(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 $\{y_i\}_{i=0}^k \notin D$. 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,

$$\sum_{i=0}^{k-1} y_i P_n(B_i) \to \sum_{i=0}^{k-1} y_i P(B_i). \tag{7.8}$$

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 (7.8) 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 7.7. *Let F_n, F_X be distribution functions associated with random variables X_n, X with $n \in \mathbb{N}$. 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 7.11. *The following statements are equivalent:*

1. $P_n \overset{w}{\to} P$,

2. $P_n \Longrightarrow P$,

3. $F_n \overset{w}{\to} F$,

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 7.10 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 sub-interval 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 7.2. *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 $E_1 = \{\omega : |-\mathcal{Z}(\omega) - \mathcal{Z}(\omega)| < \epsilon\} = \{\omega : |\mathcal{Z}| < \epsilon/2\}$, $E_2 = \Omega, \cdots$. Hence, there is no limit for $\{P(E_n)\}_{n\in\mathbb{N}}$ 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 7.12. *(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

$$h^{-1}(G) \subset \overline{h^{-1}(G)} \subset h^{-1}(G) \cup C^c. \tag{7.9}$$

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

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

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

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 7.10. Since $P_X(C^c) = 0$, we have from (7.9) that $P_X\left(\overline{h^{-1}(G)}\right) \le P_X(h^{-1}(G))$ and

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

which completes the proof by Theorem 7.10. ■

Theorem 7.13. *Let D be dense*[2] *in $\mathbb{R}$. Suppose $F_D : D \to [0, 1]$ satisfies:*

1. *F_D is non-decreasing on D.*

---

[1] The collection of its limit points.

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

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 definition of F, there exists $\epsilon > 0$ such that

$$F_D(x') - \lim_{y\downarrow x, y\in D} F_D(y) = F_D(x') - F(x) \leq \epsilon \implies F_D(x') \leq F(x) + \epsilon \tag{7.10}$$

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

$$F(y) \leq F_D(x'). \tag{7.11}$$

Thus, equations (7.10) and (7.11) give $F(y) \leq F(x) + \epsilon$ for all $y \in (x, x')$. Consequently, as $y \downarrow x$, $\lim_{y\downarrow x} F(y) \leq F(x)$. But monotonicity of F gives

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

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 7.14. *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 $\mathbb{R} - C \subset (\mathbb{R} - C) \cup C = \mathbb{R} \subset \overline{\mathbb{R} - C}$. 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 7.15. *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,

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

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

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

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 $x \ge c$. 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

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

Then, $F_{X_n}(\frac{x-b}{a}) \to F_X(\frac{x-b}{a})$ for $\frac{x-b}{a} \in C(F_X)$. $F_X(\frac{x-b}{a}) = P(\{\omega : X \le \frac{x-b}{a}\}) = P(aX + b \le x) = F_{aX+b}(x)$. 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 $n \ge N_\epsilon$,

$-\epsilon \leq P(|X_n| > c) - P(|X| > c) \leq \epsilon$ 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 $P(|X_n| > c) < \delta + \epsilon$, $C_n X_n \xrightarrow{p} 0$. ■

7.4 Exercises

1. Let $\{X_n\}_{n \in \mathbb{N}} \subset \mathcal{L}^p$ for $p \in [1, \infty)$ be a sequence of non-negative functions. Show that

$$\| \sum_{n=1}^{\infty} X_n \|_p \leq \sum_{n=1}^{\infty} \|X_n\|_p.$$

2. Show that if $\sum_{n \in \mathbb{N}} x_n$ converges absolutely, then it converges.

3. Prove Theorem 7.9.

4. Let $\{g_n\}_{n=1,2,\cdots}$ be a sequence of real valued functions that converge uniformly to g on an open set S, containing x, and g is continuous at x. Show that if $\{X_n\}_{n=1,2,\cdots}$ is a sequence of random variables taking values in S such that $X_n \xrightarrow{p} X$, then

$$g_n(X_n) \xrightarrow{p} g(X).$$

   Note: Recall that a sequence of real valued functions $\{g_n\}_{n=1,2,\cdots}$ converges uniformly to g on a set S if, for every $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ (depending only on ϵ) such that for all $n > N_\epsilon$, $|g_n(x) - g(x)| < \epsilon$ for every $x \in S$.

5. Show that $X_n \xrightarrow{as} X$ is equivalent to $P\left(\{\omega : \sup_{j \geq n} |X_j - X| \geq \epsilon\}\right) \to 0$ for all $\epsilon > 0$ as $n \to \infty$.

6. Prove item 1 of Remark 7.1.

7. Let $n \in \mathbb{N}$ and $h_n > 0$ such that $h_n \to 0$ as $n \to \infty$. Show that if $\sum_{n=1}^{\infty} P(\{\omega : |X_n - X| \geq h_n\}) < \infty$ then $X_n \xrightarrow{p} X$.

8. Show that if $Y_n \xrightarrow{d} Y$ then $Y_n = O_p(1)$.

9. Let $g : S \subseteq \mathbb{R}$ be continuous on S, and X_t and X_s be random variables defined on $(\Omega, \mathcal{F}, P)$ taking values in S. Show that: a) if X_t is independent of X_s, then $g(X_t)$ is independent of $g(X_s)$; b) if X_t and X_s are identically distributed, then $g(X_t)$ and $g(X_s)$ are identically distributed.

10. Let $\{X_n\}$ be a sequence of independent random variables that converges in probability to a limit X. Show that X is almost surely a constant.

11. Suppose $\frac{X_n - \mu}{\sigma_n} \xrightarrow{d} Z$ where the non-random sequence $\sigma_n \to 0$ as $n \to \infty$, and g is a function which is differentiable at μ. Then, show that $\frac{g(X_n) - g(\mu)}{g^{(1)}(\mu)\sigma_n} \xrightarrow{d} Z$.

12. Show that if $\{X_n\}_{n \in \mathbb{N}}$ and X are random variables defined on the same probability space and $r > s \geq 1$ and $X_n \xrightarrow{\mathcal{L}_r} X$, then $X_n \xrightarrow{\mathcal{L}_s} X$.