Remark 4.2. *1. It follows directly from Theorem 4.3 that a finite collection of random variables $\{X_i\}_{i=1}^m$ is independent if, and only if,*

$$P(\cap_{i\in J}\{\omega : X_i(\omega) \leq x_i\}) = \prod_{i\in J} P(\{\omega : X_i(\omega) \leq x_i\}), \text{ for all } J \subset \{1, \cdots, m\}.$$

2. If X_i has a density $\{X_i\}_{i=1}^m$ are independent if, and only if,

$$P(\cap_{i\in J}\{\omega : X_i(\omega) \leq x_i\}) = \prod_{i\in J} \int_{(-\infty,x_i]} f_{X_i} d\lambda.$$

4.1 Random elements

The most common cases where we deal with random elements occur when the co-domain of the element is endowed with a metric, so that the co-domain is a metric space.

Definition 4.5. *Let $X : (\Omega, \mathcal{F}, P) \to (T, \mathcal{T} = \sigma(\mathcal{O}))$, where $\mathcal{O}$ are the open sets in T. Then, X is a random element if*

$$X^{-1}(B) \in \mathcal{F} \text{ for all } B \in \mathcal{T}.$$

In this definition, we can call $\mathcal{T}$ the collection of Borel sets of T. The following examples include definitions.

Example 4.1. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ where $k \in \mathbb{N}$. Then X is a random vector if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^k)$ and $d : \mathbb{R}^k \times \mathbb{R}^k \to [0, \infty)$ is $d(x, y) = \left(\sum_{i=1}^k (x_i - y_i)^2\right)^{1/2}$ is the metric on $\mathbb{R}^k$.*

Example 4.2. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ where $\mathbb{R}^\infty = \times_{n=1}^\infty \mathbb{R}$ and $\mathcal{B}(\mathbb{R}^\infty) = \sigma(\mathcal{C})$ with $\mathcal{C} = \{C : C = \theta_i^{-1}(B),\ B \in \mathcal{B}^i,\ \theta_i(x) = (X_1, \cdots, X_i) : \mathbb{R}^\infty \to \mathbb{R}^i,\ i \in \mathbb{N}\}$. Then X is a random sequence if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^\infty)$ and $d : \mathbb{R}^\infty \times \mathbb{R}^\infty \to [0, \infty)$ is $d(x, y) = \sum_{i=1}^\infty \frac{1}{2^i} \left(\frac{\sum_{j=1}^i |x_j - y_j|}{1+\sum_{j=1}^i |x_j - y_j|}\right)^{1/2}$ is the metric on $\mathbb{R}^\infty$.*

In Example 1.3-3 we argued that if $X(\omega) : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ is continuous, it is measurable. Below is a more general result for arbitrary metric spaces.

Theorem 4.4. *Let $f : (\mathbb{X}_1, \sigma(\mathcal{O}_1)) \to (\mathbb{X}_2, \sigma(\mathcal{O}_2))$ where $(\mathbb{X}_j, \sigma(\mathcal{O}_j))$ are metric spaces. If f is continuous, f is measurable.*

Proof. $f^{-1}(\mathcal{O}_2) \subset \mathcal{O}_1$ by continuity. But $\mathcal{O}_1 \subset \sigma(\mathcal{O}_1)$. Thus, by Theorem 1.17 f is measurable ■

Theorem 4.5. *Let $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_1, \mathcal{F}_1)$ and $g : (\mathbb{X}_1, \mathcal{F}_1) \to (\mathbb{X}_2, \mathcal{F}_2)$ be measurable functions. Let $(g \circ f) : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_2, \mathcal{F}_2)$. Then, $(g \circ f)$ is $\mathcal{F} - \mathcal{F}_2$ measurable.*

Proof. Homework. ■

Remark 4.3.

1. *Let $X \in \mathbb{R}^k$ be a random vector and $f : \mathbb{R}^k \to \mathbb{R}$ be measurable. Then, $h : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ with $h(\omega) = f(X(\omega)) = (f \circ X)(\omega)$ is a random variable since compositions of measurable functions are measurable by Theorem 4.5. In particular the result follows if f is continuous. That is, real valued continuous functions of random vectors are random variables.*

2. *In 1, if $f(X) = \pi_i(X) = X_i$ and X is random vector then X_i is a random variable for $i = 1, \ldots, k$.*

Theorem 4.6. *$X \in \mathbb{R}^k$ is a random vector $\iff$ X_i is a random variable, where X_i is the ith component of X.*

Proof. ($\Longleftarrow$) Suppose X_i is a random variable for $i = 1, \ldots, k$. Let $R_k = I_1 \times \cdots \times I_k$, where $I_i = [a_i, b_i)$ are intervals in $\mathbb{R}$. Then,

$$\begin{aligned} X^{-1}(R_k) &= \{\omega : X_i(\omega) \in [a_i, b_i) \,\forall\, i\} \\ &= \{\omega : X_i^{-1}([a_i, b_i)) \,\forall\, i\} = \cap_{i=1}^{k} X_i^{-1}(I_i). \end{aligned}$$

Since X_i is a random variable, $X_i^{-1}(I_i) \in \mathcal{F}$. Furthermore, since $\mathcal{F}$ is a σ-algebra, it is closed under intersections, and $X^{-1}(R_k) \in \mathcal{F}$. The other direction of the equivalence follows from the previous remark. ■

Remark 4.4. *1. Theorem 4.6 extends to $X = \{X_1, X_2, \dots\}$. That is, X is a random sequence if, and only if, each X_i is a random variable. Furthermore, X is a random sequence if, and only if, $(X_1 \dots X_k)$ is random vector for any k.*

2. $X^{-1}((-\infty, a_1] \times \cdots \times (-\infty, a_k]) \in \mathcal{F}$ and we write $P(X^{-1}((-\infty, a_1] \times \cdots \times (-\infty, a_k])) = P \circ X^{-1}(\times_{i=1}^k(-\infty, a_i]) = P_X(\times_{i=1}^k(-\infty, a_i])$.

Also, if there exists a non-negative Borel measurable function $f_X : \mathbb{R}^k \to \mathbb{R}$ that satisfies

$$P_X(\times_{i=1}^k(-\infty, a_i]) = \int_{C(a)} f_X d\lambda^k,$$

where $C(a) = \times_{i=1}^k(-\infty, a_i]$ and $a = (a_1 \dots a_k)^T$, we call f_X the "joint density" of X. Naturally, the joint distribution function associated with X is

$$F_X(a) : \mathbb{R}^k \to [0, 1],$$

where $F_X(a) = P(C(a))$ for $a \in \mathbb{R}^k$. We can write $C(a) = \cap_{i=1}^k\{\omega : X_i(\omega) \leq a_i\}$. That $\{\omega : X_i(\omega) \leq a_i\}$ is an element of $\mathcal{F}$ follows from Theorem 4.6.

Theorem 4.7. *Consider two random variables $X_1, X_2 : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$. X_1 and X_2 are independent if, and only if, one of the following holds:*

a) $P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) := P(X \in A_1, X \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)$, for all $A_1,\ A_2 \in \mathcal{B}$,

b) $P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)$, for all $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, where $\mathcal{A}_1, \mathcal{A}_2$ are π systems which generate $\mathcal{B}$,

c) $f(X_1)$ and $g(X_2)$ are independent for each pair (f, g) of measurable functions,

d) $E(f(X_1), g(X_2)) = E(f(X_1))E(g(X_2))$ for each pair of (f, g) of bounded measurable (or non-negative measurable) functions.

Proof. First, note that X_1 and X_2 independent means that $\sigma(X_1) = X_1^{-1}(\mathcal{B})$ and $\sigma(X_2) = X_2^{-1}(\mathcal{B})$ are independent. That is, for all $A_1, A_2 \in \mathcal{B}$,

$$\begin{aligned} P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) \quad &= \quad P(X_1^{-1}(A_1))P(X_2^{-1}(A_2)) \\ &\iff \quad P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2). \end{aligned}$$

[$a) \implies b)$] Since $\mathcal{A}_1$ generates $\mathcal{B}$ and $\mathcal{A}_2$ generates $\mathcal{B}$, $\mathcal{A}_1 \subset \mathcal{B}$ and $\mathcal{A}_2 \subset \mathcal{B}$, and if a) is true for all $A_1 \in \mathcal{B}$, $A_2 \in \mathcal{B}$, then b) is true.

[$b) \implies a)$] Let $C_1 = \{A \in \mathcal{B} : P(X_1 \in A, X_2 \in A_2) = P(X_1 \in A)P(X_2 \in A_2)$ for a given $A_2 \in \mathcal{A}_2\}$. From the proof of Theorem 4.2, C_1 is a Dynkin system. $\mathcal{A}_1 \subseteq C_1$ and $\delta(\mathcal{A}_1) = \sigma(\mathcal{A}_1) = \mathcal{B} \subseteq C_1$. Analogously, $C_2 = \{A \in \mathcal{B}_2 : P(X_1 \in A_1, X_2 \in A) = P(X_1 \in A_1)P(X_2 \in A)$ for a given $A_1 \in \mathcal{A}_1\}$ is such that $\delta(\mathcal{A}_2) = \sigma(\mathcal{A}_2) = \mathcal{B} \subseteq C_2$. Consequently, $b) \implies a)$.

[$c) \implies a)$] The identity function is measurable, therefore take $f(x) = g(x) = x$

[$a) \implies c)$] For concreteness, let $f : (\mathbb{R}, \mathcal{B}) \to (M_f, \mathcal{M}_f)$ and $g : (\mathbb{R}, \mathcal{B}) \to (M_g, \mathcal{M}_g)$. f measurable implies that for all $M \in \mathcal{M}_f$, $f^{-1}(M) \in \mathcal{B}$. But X_1 a random variable implies that $X_1^{-1}(f^{-1}(M)) \in \mathcal{F}$ which we can write as $(X_1^{-1} \circ f^{-1})(M) \in \mathcal{F}$. In addition, $X_1^{-1}(f^{-1}(M)) := (X_1^{-1} \circ f^{-1})(M) \in X_1^{-1}(\mathcal{B})$. Analogously, $X_2^{-1}(g^{-1}(M')) = X_2^{-1} \circ g^{-1}(M') \in X_2^{-1}(\mathcal{B})$, for all $M' \in \mathcal{M}_g$. But by a) $X_1^{-1}(\mathcal{B})$ and $X_2^{-1}(\mathcal{B})$ are independent. Therefore $f(X_1)$ and $g(X_2)$ are independent.

[$d) \implies a)$] Let $f = I_{A_1}$ and $g = I_{A_2}$. Then,

$$f(X_1) = \begin{cases} 1 & \text{if } X_1 \in A_1 \\ 0 & \text{if } X_1 \notin A_1 \end{cases} \text{ and } g(X_2) = \begin{cases} 1 & \text{if } X_2 \in A_2 \\ 0 & \text{if } X_2 \notin A_2. \end{cases}$$

with $E(f(X_1)) = P(X_1 \in A_1)$ and $E(g(X_2)) = P(X_2 \in A_2)$. By d)

$$E(f(X_1)g(X_2)) = P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = P(X_1 \in A_1)P(X_2 \in A_2).$$

Hence, $d) \implies a)$.

$[a) \implies d)]$ From the implication $[d) \implies a)]$ we see that if f, g are indicator functions in d) $E(f(X_1)g(X_2)) = P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\})$, which by independence a) is $P(X_1 \in A_1)P(X_2 \in A_2) = E(f(X_1))E(g(X_2))$.

Now, suppose f and g are simple functions of X_1 and X_2. Then,

$$\begin{aligned} f(X_1) &= \sum_{i=0}^{k_f} a_i^f I_{\{X_1 \in A_i^f)\}} \text{ and } E(f(X_1)) = \sum_{i=0}^{k_f} a_i^f P(X_1 \in A_i^f), \\ g(X_2) &= \sum_{i=0}^{k_g} a_i^g I_{\{X_2 \in A_i^g)\}} \text{ and } E(g(X_2)) = \sum_{i=0}^{k_g} a_i^g P(X_2 \in A_i^g) \end{aligned}$$

Consequently,

$$\begin{aligned} E(f(X_1)g(X_2)) &= E\left(\sum_{i=0}^{k_f}\sum_{j=0}^{k_g} a_i^f a_j^g I_{\{X_1 \in A_i^f\}\cap\{X_2 \in A_j^g\}}\right) \\ &= \sum_{i=0}^{k_f}\sum_{j=0}^{k_g} a_i^f a_j^g P(X_1 \in A_i^f)P(X_2 \in A_j^g) \text{ by independence} \\ &= E(f(X_1))E(g(X_2)) \end{aligned} \tag{4.4}$$

Now, let f be a measurable non-negative function such that $\{f_n\}_{n\in\mathbb{N}}$ are simple functions increasing to f and g is non-negative and simple. Then,

$$\begin{aligned} E(f(X_1)g(X_2)) &= E\left(\lim_{n\to\infty} f_n(X_1)g(X_2)\right) \\ &= \lim_{n\to\infty} E(f_n(X_1)g(X_2)) \text{ by Lebesgue's Monotone Convergence Theorem} \\ &= \lim_{n\to\infty} E(f_n(X_1))E(g(X_2)) \text{ by equation (4.4)} \\ &= E(f(X_1))E(g(X_2)) \text{ by Lebesgue's Monotone Convergence Theorem} \end{aligned} \tag{4.5}$$

Now, let f be non-negative and let $\{g_n\}_{n\in\mathbb{N}}$ be non-negative simple functions increasing to

g measurable and non-negative. Then,

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E\left(f(X_1)\lim_{n\to\infty} g_n(X_2)\right) \\
&= \lim_{n\to\infty} E(f(X_1)g_n(X_2)) \\
&= \lim_{n\to\infty} E(f(X_1))E(g_n(X_2)) \text{ by equation (4.5)} \\
&= E(f(X_1))E(g(X_2))
\end{aligned}$$

Finally, let $f = f^+ - f^-$ be bounded and measurable and g bounded and non-negative.

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E([f^+(X_1) - f^-(X_1)]g(X_2)) \\
&= E(f^+(X_1)g(X_2)) - E(f^-(X_1)g(X_2)) \\
&= E(f^+(X_1))E(g(X_2)) - E(f^-(X_1))E(g(X_2)) \\
&= E(f(X_1))E(g(X_2)).
\end{aligned}$$

To complete the proof, repeat the last argument for $g = g^+ - g^-$. ■

Chapter 5

Convergence of random variables

5.1 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 1.20. 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 $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ 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 $X_n \stackrel{as}{\to} X$ (or $X_n \stackrel{ae}{\to} X$). 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 5.1. *(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 5.1. *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 5.2. *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 $X_n \stackrel{p}{\to} X$ 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 5.1. *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, $X_n \stackrel{as}{\to} X \implies X_n \stackrel{p}{\to} X$.*

Proof. Let $E_n(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$ for any $\epsilon > 0$. $X_n \stackrel{as}{\to} X$ implies that there exists a natural number $N(\epsilon)$ such that for all $n \geq N(\epsilon)$ we have $P(E_n(\epsilon)) = 0$. Hence, if we define $E(\epsilon) = \{\omega : \sum_{n=1}^{\infty} I_{E_n(\epsilon)} < \infty\}$, then

$$P(E(\epsilon)) = P\left(\liminf_{n\to\infty} E_n^c(\epsilon)\right) = P\left(\left(\limsup_{n\to\infty} E_n(\epsilon)\right)^c\right) = 1.$$

This implies that

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

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 5.2. *(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 5.3. *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. $X_n \xrightarrow{p} X \iff X_r - X_s \xrightarrow{p} 0$ *as n, r, $s \to \infty$ (Cauchy in probability)*

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

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,

$$P(\{\omega : |X_r - X_s| > \epsilon\}) \leq P(\{\omega : |X_r - X| > \epsilon/2\}) + P(\{\omega : |X_s - X| > \epsilon/2\}). \tag{5.1}$$

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

$(\Longleftarrow)$ 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 (5.1)

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

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, $n(1) < n(2) < \ldots$ 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, $X_{n_j}(\omega) \to X(\omega) = \lim_{j\to\infty} X_{n_j}(\omega)$ *as*.

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

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

$$P(\{\omega : |X_{n(j)} - X| > \epsilon\}) \geq \delta. \tag{5.2}$$

But every $X_{n(j)}$ has a subsequence $X_{n(j(i))} \xrightarrow{as} X$ and hence $X_{n(j(i))} \xrightarrow{p} X$, which contradicts equation (5.2). ■

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

Theorem 5.4. *(Slutsky's Theorem) If X_n, X are random elements defined on the same probability space and $X_n \xrightarrow{p} X$, $g : \mathbb{R}^K \to \mathbb{R}^L$ continuous, then $g(X_n) \xrightarrow{p} g(X)$.*

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 $X_n \xrightarrow{p} X$, $P(A_{n,k}^c) \to 0$ and therefore $P(A_n^c) \to 0$ or $P(A_n) \to 1$. ■

Theorem 5.5. *Let X_n, X be defined in the same probability space such that $X_n \xrightarrow{p} X$ and $E(X_n)$, $E(X) < \infty$. If there exist a random variable $0 \leq Y \in \mathcal{L}$ such that $|X_n(\omega)| \leq Y(\omega)$ for all n, then $E(X_n) \to E(X)$.*

Proof. Since $X_n \xrightarrow{p} X$, then Theorem 5.3 says that every subsequence X_{n_k} has a further subsequence $X_{n_{k(i)}} \xrightarrow{as} X$. By Lebesgue's Dominated Convergence Theorem

$$E(X_{n_{k(i)}}) \to E(X).$$

Consequently, $E(X_{n_k}) \to E(X)$. Hence, $E(X_n) \to E(X)$. (This is so because to show $E(X_n) \to E(X)$, it suffices to show that every convergent subsequence $E(X_{n_k})$ is such that $E(X_{n_k}) \to E(X)$). ■

Remark 5.1. *1. The following results follow directly from Theorem 5.3.*

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

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

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,

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

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

Definition 5.3. *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 $X_n \xrightarrow{\mathcal{L}^p} X$, if $d_p(X_n, X) \to 0$ as $n \to \infty$.*

The limit X in Definition 5.3 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 5.6. *(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 $\{X_n\}_{n\in\mathbb{N}} \subseteq \mathcal{L}^p(\Omega, \mathcal{F}, P)$. 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

$$\|X_{n(k+1)} - X_{n(k)}\|_p \leq \frac{1}{2^k} \text{ for } k = 1, 2, \cdots \tag{5.3}$$

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

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

Finally, since

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

and $\{X_n\}_{n=1,2,\dots}$ 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,

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

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

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

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

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

Theorem 5.7. *Let X, X_n, $n = 1, 2, \ldots$ be random variables defined in the same probability space. If $X_n \overset{\mathcal{L}^p}{\to} X$, then $X_n \overset{p}{\to} X$.*

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

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

Taking limits on both sides completes the proof.■

5.3 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 5.4. *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 $|f| \leq C < \infty$.

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