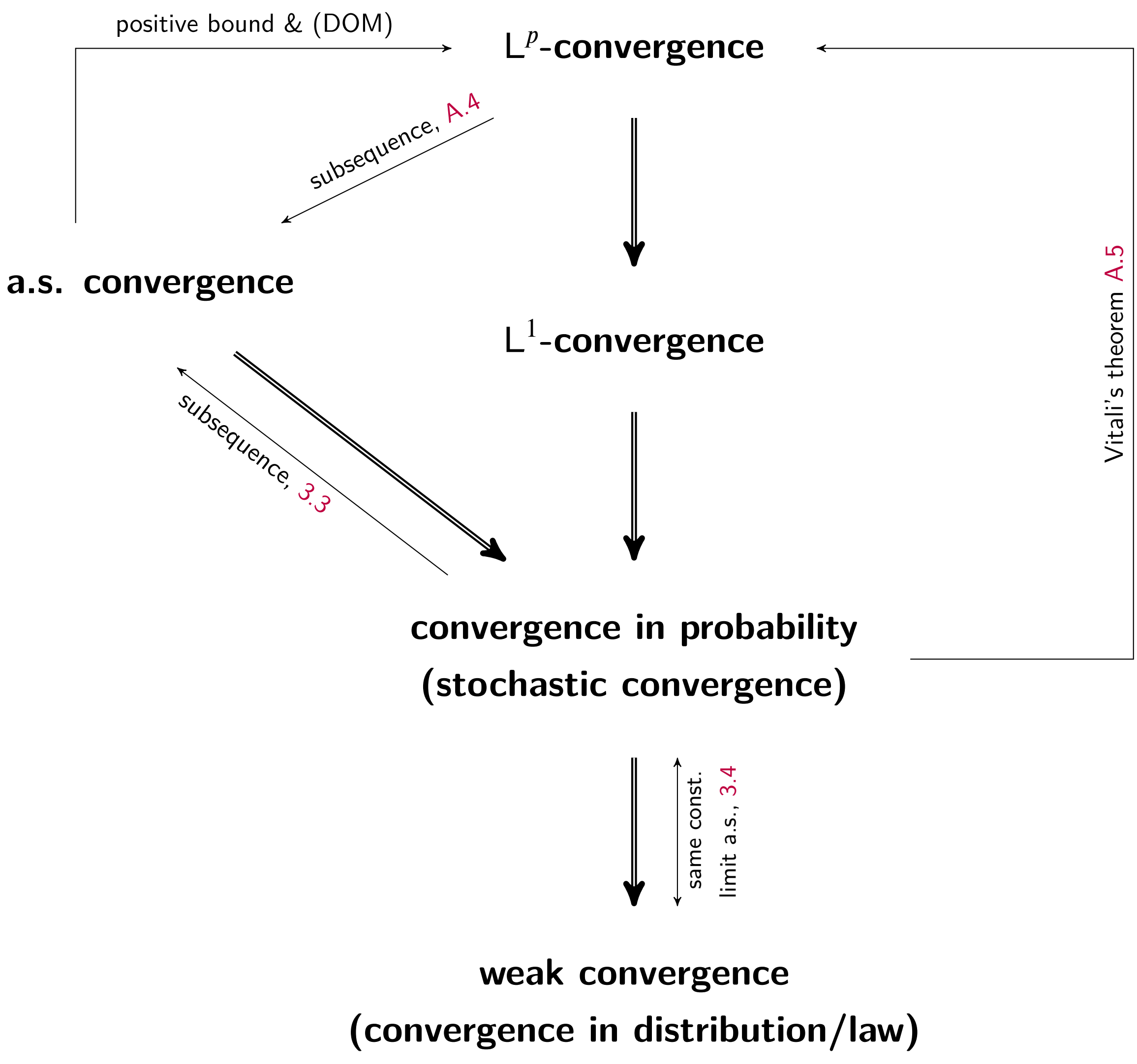
Convergence in Probability

Robert Baumgarth[1]

[1]Mathematics Research Unit, FSTM, University of Luxembourg, Maison du Nombre, 6, Avenue de la Fonte, 4364 Esch-sur-Alzette, Grand-Duché de Luxembourg



1 Definitions

Definition 1.1. Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$) defined on the same probability space $(\Omega, \mathscr{A}, \mathbb{P})$. We say that

(i) $X_n \xrightarrow{\text{a.s.}} X$, X_n **converges to** X **almost surely (a.s.)** or $X_n \xrightarrow{n\uparrow\infty} X$ **with probability one**

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

▷ «presque sûrement» (p.s.), «fast sicher» (f.s.) or «avec probabilité 1», «mit Wahrscheinlichkeit 1».

(ii) $X_n \xrightarrow{\mathbb{P}} X$, X_n **converges to** X **in probability** or X_n **converges stochastically to** X

$$:\iff \forall \varepsilon > 0 : \quad \lim_{n\to\infty} \mathbb{P}\big(|X_n - X| > \varepsilon\big) = 0$$

▷ en probabilité, in Wahrscheinlichkeit or –, stochastisch.

(iii) $X_n \xrightarrow{\mathsf{L}^p} X$, X_n **converges to** X **in** L^p or X_n **converges to** X **in the** p^{th} **mean**

$$:\iff \forall X, X_n \in \mathsf{L}^p(\mathbb{P}) : \quad \lim_{n\to\infty} \|X_n - X\|_{\mathsf{L}^p} = 0,$$

where, $\|X\|_{\mathsf{L}^p}(\mathbb{P}) := \left(\mathbb{E}\,|X|^p\right)^{1/p}$, i.e.

$$:\iff \forall X, X_n \in \mathsf{L}^p : \quad \lim_{n\to\infty} \mathbb{E}\,|X_n - X|^p = 0.$$

▷ dans L^p, in L^p or en moyenne d'ordre p, im p-ten Mittel.

Definition 1.2. Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$) *not necessarily* defined on the same probability space. We say that

(i) $X_n \xrightarrow{\text{d}/\mathscr{D}/\mathscr{L}} X$, X_n **converges to** X **in distribution/law**

$$:\iff \lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mathbb{P}\big(X_n \leqslant x\big) = \mathbb{P}(X \leqslant x) = F(x)$$

for all points $x \in \mathbb{R}$, where F is continuous.

▷ en loi, in Verteilung.

(ii) $X_n \xrightarrow{\text{d/w}} X$, X_n **converges to** X **weakly**

$$:\iff \forall f \in \mathsf{C}_b(\mathbb{R}) : \quad \lim_{n\to\infty} \mathbb{E} f(X_n) = \mathbb{E} f(X). \tag{1.1}$$

▷ faible, schwach.

Remark 1.3. (1) Since the real random variables X_n, X do not have to be defined on the same probability space in the definition of weak convergence, we actually should emphasise the dependence by writing $\mathbb{P}$, $\mathbb{P}^n$ and $\mathbb{E}$, $\mathbb{E}^n$ respectively. It is a common convention to omit this dependence.

(2) As already indicated in the notation in Definition 1.2, (i) and (ii) are equivalent (⇝ Portmanteau theorem). We can equivalently write (1.1) as

$$\forall f \in \mathsf{C}_b(\mathbb{R}) : \quad \lim_{n\to\infty} \int f(x)\mathbb{P}\big(X_n \in \mathrm{d}x\big) = \int f(x)\mathbb{P}(X \in \mathrm{d}x)\,.$$

In functional analysis it is common to write $\lim_{n\to\infty} \left\langle f, \mathbb{P}_{X_n} \right\rangle = \langle f, \mathbb{P}_X \rangle$, where $\langle f, \mu \rangle = \int f \,\mathrm{d}\mu$.

2 Implications

Theorem 2.1. *Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$) and $1 \leqslant p \leqslant \infty$. Then*

(a) $X_n \xrightarrow{\mathsf{L}^p} X \Longrightarrow X_n \xrightarrow{\mathsf{L}^1} X$

(b) $X_n \xrightarrow{\mathsf{L}^1} X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X$

(c) $X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X$

Proof. (a) Let $1 \leqslant p < \infty$ and q the conjugate exponent $1/p + 1/q = 1$. By Hölder's inequality

$$\mathbb{E}\left|X_n - X\right| = \int \left|X_n - X\right| \cdot 1 \,\mathrm{d}\mathbb{P} \leqslant \left(\int \left|X_n - X\right|^p \mathrm{d}\mathbb{P}\right)^{1/p} \left(\int 1^q \,\mathrm{d}\mathbb{P}\right)^{1/q}$$
$$= \left(\mathbb{E}\left|X_n - X\right|^p\right)^{1/p} \underset{\text{assumption}}{\overset{n\uparrow\infty}{\longrightarrow}} 0.$$

The case $p = \infty$ follows in a completely analogue matter. In particular, we have shown that $\mathsf{L}^p(\mathbb{P}) \hookrightarrow \mathsf{L}^1(\mathbb{P})$ is a continuous embedding.

(b) By the Chebyshev–Markov inequality, we get

$$\forall \varepsilon > 0 : \quad \mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon}\mathbb{E}\left|X_n - X\right| \overset{n\uparrow\infty}{\longrightarrow} 0.$$

(c) For all $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) = \mathbb{E}\mathbb{1}_{\{|X_n - X| > \varepsilon\}} = \mathbb{E}\mathbb{1}_{[-\varepsilon,\varepsilon]^c}(X_n - X).$$

The random variables $Y_n := \mathbb{1}_{\{|X_n - X| > \varepsilon\}}$ are uniformly bounded by $\mathbf{1} \in \mathsf{L}^p(\mathbb{P})$ and $Y_n \overset{n\uparrow\infty}{\longrightarrow} 0$ a.s. Hence, by dominated convergence A.1, it follows

$$\forall \varepsilon > 0 : \quad \mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) = \mathbb{E}\mathbb{1}_{\{|X_n - X| > \varepsilon\}} \overset{n\uparrow\infty}{\longrightarrow} 0. \qquad \blacksquare$$

Theorem 2.2. *Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$) such that $X_n \xrightarrow{\mathbb{P}} X$. Then $X_n \xrightarrow{d} X$. Moreover, it holds $f(X_n) \xrightarrow{\mathsf{L}^1} f(X)$ for all $f \in \mathsf{C}_b(\mathbb{R})$.*

Proof. Let $f \in \mathsf{C}_b(\mathbb{R})$ and $\varepsilon > 0$ fixed. Since $\{|X| > k\} \downarrow \emptyset$ for $k \uparrow \infty$, by the $\emptyset$-continuity of the probability measure $\mathbb{P}$, we have

$$\exists N = N(\varepsilon) \in \mathbb{N}\ \forall k \geqslant N : \quad \mathbb{P}(|X| > k) < \varepsilon. \tag{2.1}$$

By assumption $|f| \leqslant M$, for some suitable constant M, and $f\big|_{[-(N+1),N+1]}$ is uniformly continuous, i.e.

$$\exists \delta = \delta(\varepsilon) \in (0,1)\ \forall\, |x|, |y| \leqslant N+1, |x-y| \leqslant \delta : \quad |f(x) - f(y)| \leqslant \varepsilon. \tag{2.2}$$

Hence, splitting the area of integration in clever way,

$$\begin{aligned}\left|\mathbb{E}\left[f(X_n)-f(X)\right]\right| &\leqslant \mathbb{E}\left|f(X_n)-f(X)\right| \\ &\leqslant \left(\int_{\{|X_n-X|\leqslant\delta\}\cap\{|X|\leqslant N\}} + \int_{\{|X_n-X|\leqslant\delta\}\cap\{|X|>N\}} + \int_{\{|X_n-X|>\delta\}}\right)\left|f(X_n)-f(X)\right| \mathrm{d}\mathbb{P} \\ &\leqslant \underbrace{\varepsilon\mathbb{P}\left(\left|X_n-X\right|\leqslant\delta, |X|\leqslant N\right)}_{|X_n|\leqslant|X_n-X|+|X|\leqslant\delta+N\leqslant 1+N} + \underbrace{2M\mathbb{P}(|X|>N)+2M\mathbb{P}\left(\left|X_n-X\right|>\delta\right)}_{|f(x)-f(y)|\leqslant 2\|f\|_\infty\leqslant 2M},\end{aligned}$$

where we used that $|f| \leqslant M$ together with (2.2) and since on the set $\left\{|X_n - X| \leqslant \delta\right\} \cap \{|X| \leqslant N\}$, we have $|X_n| \leqslant |X_n - X| + |X| \leqslant \delta + N \leqslant 1 + N$. Finally, using (2.1), we get

$$\mathbb{E}\left|f(X_n)-f(X)\right| \leqslant (2M+1)\varepsilon + 2M\mathbb{P}\left(\left|X_n-X\right|>\delta\right) \xrightarrow{n\uparrow\infty} (2M+1)\varepsilon \xrightarrow{\varepsilon\downarrow 0} 0,$$

thus $f(X_n) \to f(X)$ in L^1 and $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$. ■

3 «Subsequence» implications

Recall the definition of the limit superior of sets:

$$\omega \in \limsup A_n := \bigcap_{m\in\mathbb{N}}\left(\bigcup_{n\geqslant m} A_n\right) \iff \omega \text{ appears in infinitely many of the } A_n$$

Hence, we can justify the probabilistic terms

$$\limsup A_n = \left\{A_n \text{ for infinitely many } n\in\mathbb{N}\right\} = \left\{A_n \text{ infinitely often (i.o.)}\right\}.$$

Lemma 3.1 (Borel-Cantelli). *Let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$. Then*

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(A_n\right) < \infty \implies \mathbb{P}\left(A_n \textit{ for infinitely many } n\in\mathbb{N}\right) = 0.$$

Proof.

$$\omega \in \limsup A_n \iff \omega \text{ appears in infinitely many of the } A_n \iff \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega) = \infty.$$

By Beppo Levi, it follows that

$$\mathbb{E}\left(\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega)\right) = \sum_{n\in\mathbb{N}} \mathbb{E}\mathbb{1}_{A_n}(\omega) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(A_n\right) < \infty,$$

and we see that $\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega) < \infty$ a.s., hence $\mathbb{P}\left(A_n \text{ for infinitely many } n\in\mathbb{N}\right) = 0$. ■

Making the special choice $A_n := \left\{|X_n - X| > \varepsilon\right\}$ in 3.1, we can prove that having control of a fast rate of convergence of $\mathbb{P}\left(|X_n - X| > \varepsilon\right) \to 0$, we already get convergence almost surely.

Lemma 3.2. *Let $X, X_n : \Omega \to \mathbb{R}$ be random variables $(n \in \mathbb{N})$ with $X_n \xrightarrow{\mathbb{P}} X$ and a null sequence $\varepsilon_n \downarrow 0$ with $\mathbb{P}\left(|X_n - X| > \varepsilon_n\right) < \infty$. Then $X_n \xrightarrow{a.s.} X$.*

Proof. By the Borel-Cantelli lemma 3.1,

$$\begin{aligned}
&\mathbb{P}\big(A_n \text{ for infinitely many } n \in \mathbb{N}\big) = 0 \\
&\quad \iff \mathbb{P}\big(A_n \text{ for at most finitely many } n \in \mathbb{N}\big) = 1 \\
&\quad \iff \exists \Omega' \subset \Omega,\ \mathbb{P}\big(\Omega'\big) = 1 \quad \forall w \in \Omega'\ \exists N(\omega)\ \forall n \geqslant N(\omega) : \quad \big|X_n(\omega) - X(\omega)\big| < \varepsilon_n \\
&\quad \implies \forall w \in \Omega' : \quad \big|X_n(\omega) - X(\omega)\big| \xrightarrow{n\uparrow\infty} 0.
\end{aligned}$$

■

Corollary 3.3 ($\mathbb{P}$ convergence $\implies$ a.s. of subsequence). *Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$). Then*

$$X_n \xrightarrow{\mathbb{P}} X \quad \implies \quad \exists (X_{n(k)})_{k\in\mathbb{N}} : \quad X_{n(k)} \xrightarrow[a.s.]{k\uparrow\infty} X.$$

Proof. By assumption,

$$\forall k \in \mathbb{N}\ \forall \varepsilon > 0\ \exists N(k,\varepsilon) \in \mathbb{N}\ \forall n \geqslant N(k,\varepsilon) : \quad \mathbb{P}\big(\big|X_k - X\big| > \varepsilon\big) \leqslant 2^{-k}.$$

Choose $\varepsilon = 2^{-k}$ and $n(k) := N(k, 2^{-k})$. Hence,

$$\sum_{k\in\mathbb{N}} \mathbb{P}\big(\big|X_{n(k)} - X\big| > 2^{-k}\big) \leqslant \sum_{k\in\mathbb{N}} 2^{-k} < \infty.$$

and the assertion follows from Lemma 3.2. ■

Finally, we show a relation between convergence in distribution and convergence in probability.

Lemma 3.4. *Let $X, X_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$) defined on the same probability space. If $X \equiv c$ is constant a.s., then*

$$X_n \xrightarrow{\mathbb{P}} X \equiv c \quad \iff \quad X_n \xrightarrow{d} X \equiv c.$$

Proof. $\implies$ Clearly by 2.1.

$\impliedby$ We choose a cut-off function in a clever way: Fix $\varepsilon > 0$ and $\chi_\varepsilon \in C_b(\mathbb{R})$ with $\chi_\varepsilon(0) = 0$ and $\chi_\varepsilon \geqslant \mathbb{1}_{[-\varepsilon,\varepsilon]^c}$. Then $\chi_\varepsilon(\cdot - c) \in C_c(\mathbb{R})$ and we have

$$\mathbb{P}\big(\big|X_n - X\big| > \varepsilon\big) \leqslant \int \chi_\varepsilon(X_n - X)\mathrm{d}\mathbb{P} = \int \chi_\varepsilon(X_n - c)\mathrm{d}\mathbb{P} \xrightarrow[n\uparrow\infty]{\mathrm{w}} \int \chi_\varepsilon(X - c)\mathrm{d}\mathbb{P} = 0.$$

■

4 Weak convergence as convergence in distribution

The following theorem links convergence in distribution to a.s. convergence (on a different probability space). Note that the theorem also holds in metric spaces.

Theorem 4.1 (Skorokhod's representation theorem). *Let X_n, X be real random variables ($n \in \mathbb{N}$), not necessarily defined on the same probability space. If $X_n \xrightarrow{d} X$, then there is another probability space and random variables Y_n, Y on this probability space, such that*

$$X_n \sim Y_n, \quad X \sim Y \qquad \text{and} \qquad X_n \xrightarrow{a.s.} X.$$

Theorem 4.2 (CMT – continuous mapping theorem). *Let $X_n, X : \Omega \to \mathbb{R}^d$ be random variables with $X_n \xrightarrow[n\to\infty]{d} X$ and $g : \mathbb{R}^d \to \mathbb{R}^r$ continuous. Then $g(X_n) \xrightarrow[n\to\infty]{d} g(X)$ in $\mathbb{R}^r$.*

Proof. Let $f \in C_b(\mathbb{R}^r)$. Then $f \circ g \in C_b(\mathbb{R}^d)$ and a direct calculation shows

$$\mathbb{E}f(g(X_n)) = \mathbb{E}(f \circ g)(X_n) \xrightarrow[n\to\infty]{\mathrm{d}} \mathbb{E}(f \circ g)(X) = \mathbb{E}f(g(X)).$$

Hence, $g(X_n) \xrightarrow[n\to\infty]{\mathrm{d}} g(X)$ in $\mathbb{R}^r$. ■

Theorem 4.3 (Cramér). *Let $X_n, Y_n : \Omega \to \mathbb{R}^d$ be random variables. Then*

$$X_n - Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0 \qquad \Longrightarrow \qquad \left(X_n \xrightarrow[n\to\infty]{d} Z \quad \Longleftrightarrow \quad Y_n \xrightarrow[n\to\infty]{d} Z \right).$$

Such sequences (X_n) and (Y_n) are called **stochastically equivalent**.

Theorem 4.4 (Cramér-Slutsky). *Let $X_n : \Omega \to \mathbb{R}^r$ and $Y_n : \Omega \to \mathbb{R}^d$ be random variables with $X_n \xrightarrow[n\to\infty]{d} X$ in $\mathbb{R}^r$ and $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ in $\mathbb{R}^d$ for some constant c. Then*

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow[n\to\infty]{d} \begin{pmatrix} X \\ c \end{pmatrix} \qquad \text{in } \mathbb{R}^{r+d}.$$

By the CMT, Theorem 4.2 above, for any function g continuous on $\mathbb{R}^{r+d}$ or a.s. continuous in (X, c), moreover

$$g(X_n, Y_n) \xrightarrow[n\to\infty]{d} g(X, c).$$

Proof. Using the assumption $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} X_n \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ Y_n - c \end{pmatrix} \xrightarrow[n\to\infty]{\mathbb{P}} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Longrightarrow \qquad \begin{pmatrix} X_n \\ Y_n \end{pmatrix}, \begin{pmatrix} X_n \\ c \end{pmatrix} \text{ are stochastically equivalent.}$$

Let $f \in C_b(\mathbb{R}^{r+d})$. Then $f(\cdot, c) \in C_b(\mathbb{R}^r)$ for fixed $c \in \mathbb{R}^d$ and

$$\begin{aligned} &\Longrightarrow \qquad \mathbb{E}f\begin{pmatrix} X_n \\ c \end{pmatrix} \xrightarrow[n\to\infty]{} \mathbb{E}\begin{pmatrix} X \\ c \end{pmatrix} \\ &\Longrightarrow \qquad \mathbb{E}f(X_n, c) \xrightarrow[n\to\infty]{} \mathbb{E}f(X, c), \end{aligned}$$

as $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X$ by assumption. So the assertion follows. ■

We emphasise that the CMT, Cramér's theorem and the Cramér-Slutsky theorem also hold in a metric space setting.

Example 4.5. Let $X_n : \Omega \to \mathbb{R}^r$ and $Y_n : \Omega \to \mathbb{R}^d$ be random variables with $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X$ in $\mathbb{R}^r$ and $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ in $\mathbb{R}^d$ for some constant c. By applying the CMT, Theorem 4.2 to the functions $g(x, y) = x+y$, $g(x, y) = xy$, and $g(x, y) = xy^{-1}$, we get by Theorem 4.4: For $r = d$

(a) $X_n + Y_n \xrightarrow[n\to\infty]{\mathrm{d}} X + c$,

(b) $\langle X_n, Y_n\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \langle X, c\rangle$,

and for $d = 1$,

(c) $X_nY_n \xrightarrow[n\to\infty]{\mathrm{d}} Xc$,

(c) $\dfrac{X_n}{Y_n} \xrightarrow[n\to\infty]{\mathrm{d}} \dfrac{X}{c}$.

Next, we characterise convergence in distribution with help of the characteristic function $\varphi : \mathbb{R}^d \to \mathbb{E}$, $\varphi(\xi) = e^{i\langle\xi,X\rangle}$.

Theorem 4.6 (Lévy). *Let $X_n, X : \Omega \to \mathbb{R}^d$ be random variables ($n \in \mathbb{N}$). Then*

$$X_n \xrightarrow[n\to\infty]{d} X \qquad \Longleftrightarrow \qquad \forall \xi \in \mathbb{R}^d : \quad \varphi_{X_n} \xrightarrow[n\to\infty]{} \varphi_X(\xi).$$

The convergence of the characteristic function is locally uniformly.

We immediately get a characterisation of multidimensional convergence in distribution, known as *Cramér-Wold device.*

Corollary 4.7 (Cramér-Wold). *Let $X_n, X : \Omega \to \mathbb{R}^d$ be random variables ($n \in \mathbb{N}$). Then*

$$X_n \xrightarrow[n\to\infty]{d} X \qquad \Longleftrightarrow \qquad \forall \xi \in \mathbb{R}^d : \quad \langle \xi, X_n\rangle \xrightarrow[n\to\infty]{d} \langle \xi, X\rangle .$$

Finally, we give two result for the convergence in distribution of sums and vectors of random variables. Without further assumptions (in general) it is not possible to deduce that, if $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X$ and $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$, then also $X_n + Y_n \xrightarrow[n\to\infty]{\mathrm{d}} X + Y$ or $(X_n, Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} (X, Y)$, even if the random variables X_n, X, Y_n, Y are defined on the same probability space. But it holds

Lemma 4.8. *Let $X_n, Y_n : \Omega \to \mathbb{R}$ be random variables ($n \in \mathbb{N}$), not necessarily defined on the same probability space. Then*

$$(X_n, Y_n) \xrightarrow[n\to\infty]{d} (X, Y) \qquad \Longrightarrow \qquad \begin{cases} X_n \xrightarrow[n\to\infty]{d} X, \quad Y_n \xrightarrow[n\to\infty]{d} Y \qquad \textit{and} \\ X_n + Y_n \xrightarrow[n\to\infty]{d} X + Y. \end{cases}$$

Proof. We choose $d = 2$, and $\xi = (\tau, 0)$, $\xi = (0, \tau)$ and $\xi(\tau, \tau)$ in the Cramér-Wold device 4.7. ■

Theorem 4.9 (Slutsky). *Let $X_n, Y_n : \Omega \to \mathbb{R}^d$ be sequences of random variables ($n \in \mathbb{N}$) such that $X_n \xrightarrow{d} X$ and $X_n - Y_n \xrightarrow{\mathbb{P}} X$. Then $X_n \xrightarrow{d} X$.*

5 Counterexamples

Example 5.1. Let $(\Omega, \mathscr{A}, \mathbb{P}) = ([0,1), \mathscr{B}[0,1), \mathrm{Leb}\big|_{[0,1)} = \mathrm{d}\omega)$.

(a) L^1-convergence $\not\Longrightarrow$ L^p-convergence ($p > 1$)

$$X_n(\omega) = \mathbb{1}_{[1/n,1)}(\omega)\ \omega^{-1/p}, \quad X(\omega) = \omega^{-1/p}.$$

(b) L^1-convergence $\not\Longrightarrow$ a.s. convergence

$$X_{n,k}(\omega) = \mathbb{1}_{[k/n,(k+1)/n)}(\omega), \quad n \in \mathbb{N},\ k = 0, 1, \dots, n-1.$$

It is easy so see that $X(\omega) \equiv 0$ in L^1, but the sequence does not converge at any point $\omega \in [0,1)$.

(c) $\mathbb{P}$-convergence $\not\Longrightarrow$ L^1-convergence

$$X_n(\omega) = n\ \mathbb{1}_{[0,1/n]}(\omega), \quad X_n(\omega) \equiv 0.$$

(d) $\mathbb{P}$-convergence $\not\Longrightarrow$ a.s. convergence. Clear by part (b).

(e) w-convergence $\not\Longrightarrow$ $\mathbb{P}$-convergence (also if all random variables are defined on the same probability space). We define the so called «Rademacher» functions $R_1, R_2, R_3, \dots$ by $R_n \sim \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, i.e. the R_n are defined alternating on sets of the same length. Clearly, $\mathbb{E}f(R_n) = \frac{1}{2}f(1) + \frac{1}{2}f(-1) \longrightarrow \mathbb{E}f(R_1)$, i.e. $R_n \xrightarrow{\mathrm{w}} R_1$. On the other hand we see that the $R_n(\omega)$ cannot converge a.s., since

$$\liminf_{n\to\infty} R_n(\omega) = -1 < +1 = \limsup_{n\to\infty} R_n(\omega) \quad \forall \omega \in (0,1).$$

For symmetry reasons we also have that, for $k \neq n$,

$$R_n - R_k = \begin{cases} 0, & \{R_n = R_k\} & \text{in } \frac{1}{2} \text{ of all cases} \\ +2, & \{R_n = 1, R_k = -1\} & \text{in } \frac{1}{4} \text{ of all cases} \\ -2, & \{R_n = -1, R_k = 1\} & \text{in } \frac{1}{4} \text{ of all cases.} \end{cases}$$

Hence, it follows that $\mathbb{P}(|R_n - R_k| > \varepsilon) = \frac{1}{2}$ for all $\varepsilon < 2$. So, $(R_n)_{n\in\mathbb{N}}$ cannot be a $\mathbb{P}$-Cauchy sequence, and thus, also not stochastically convergent.

Example 5.2. $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X,\ Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \not\Longrightarrow (X_n, Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} (X, Y)$. Choose $X, Y \sim \mathrm{Ber}(1/2)$ iid Bernoulli and define $X_n := X + \frac{1}{n}$ and $Y_n := 1 - X_n$. Then $1 - X \sim \mathrm{Ber}(1/2) \sim Y$, $X_n \xrightarrow{\mathrm{d}} X$ and $Y_n \xrightarrow{\mathrm{d}} 1 - X \sim Y$. Assuming that, $(X_n, Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} (X, Y)$, then it follows that

$$X + Y \sim \frac{1}{2}\left(\delta_0 + \delta_1\right) \otimes \frac{1}{2}\left(\delta_0 + \delta_1\right)$$

by $X \perp\!\!\!\perp Y$. On the other hand, obviously $X_n + Y_n \equiv 1 \xrightarrow{\mathrm{d}} 1$.

A Appendix - some measure theory

Let E be a polish space, typically $E = \mathbb{R}, \mathbb{R}^n$, and $(E, \mathscr{A}, \mu)$ a measure space. Let $1 \leqslant p < \infty$. Recall that the spaces of **p-times integrable functions** are defined as

$$\mathscr{L}^p(E, \mathscr{A}, \mu) := \left\{ u : E \to \mathbb{R} : u \text{ measurable}, \int |u|(x)^p \mu(\mathrm{d}x) < \infty \right\}, \qquad (1 \leqslant p < \infty)$$

$$\mathscr{L}^\infty(E, \mathscr{A}, \mu) := \left\{ u : E \to \mathbb{R} : u \text{ measurable}, \ \exists c \geqslant 0, \mu\{|u| \geqslant c\} = 0 \right\}. \qquad (p = \infty)$$

Moreover, we define the norms $u \mapsto \|u\|_{\mathrm{L}^p}$ $(1 \leqslant p \leqslant \infty)$[1]

$$\|u\|_{\mathrm{L}^p} := \left(\int |u(x)|^p \, \mu(\mathrm{d}x) \right)^{1/p}, \qquad (1 \leqslant p < \infty)$$

$$\|u\|_\infty := \inf\{c \geqslant 0, \mu\{|u| \geqslant c\} = 0\}. \qquad (p = \infty)$$

$\mathscr{L}^p(\mu)$ is a quasi-normed vector space, since only $\|u\|_{\mathrm{L}^p} \iff u = 0$ *a.e.* and not necessarily $u \equiv 0$. But we can make $\mathscr{L}^p \rightsquigarrow \mathrm{L}^p$ to a normed space by a standard procedure:

- Define an equivalence relation by: $u, v \in \mathscr{L}^p(\mu)$, $u \sim v :\iff \mu(u \neq v) = 0$.
- Let $[u] := \left\{ v \in \mathscr{L}^p(\mu) : v \sim u \right\}$ the equivalence relation with representative u.
- Then $\|[u]\|_{\mathrm{L}^p} := \inf\left\{ \|v\|_{\mathrm{L}^p} : v \in [u] \right\} = \|u\|_{\mathrm{L}^p}$ ($u = v$ a.e.!).
- Define $\mathrm{L}^p(\mu) := \mathscr{L}^p(\mu)/_\sim \equiv \left\{ [u] : u \in \mathscr{L}^p(\mu) \right\}$.

Then $\mathrm{L}^p(\mu)$ is a vector space with (true) norm $\|[u]\|_{\mathrm{L}^p}$.

Caution

- One normally speaks of L^p-*functions*, where we just identify every $[u]$ with a «good» representative $u_0 \in [u]$. This is justified since $[u] = [u_0]$ ($u_0 \in [u]$) and hence every representative is unique only up to a null set.
- Expressions like $u = v$, $u \leqslant v$ are understood only up to null sets, i.e. $u = v$ a.e., $u \leqslant v$ a.e. etc.
- $u \in \mathscr{L}^p(\mu) \iff u$ measurable and $|u|^p \in \mathrm{L}^1(\mu)$.

Next, we state the **Dominated convergence theorem** or **Theorem of Lebesgue**. Its power and flexibility is one of the primary advantages of Lebesgue's integration theory over Riemmanian. It is heavily used in probability theory to prove the convergence of the expectation of random variables.

Theorem A.1 (Dominated convergence, Lebesgue). *Let $(u_n)_{n \in \mathbb{N}} \subset \mathscr{L}^p(\mu)$, $1 \leqslant p < \infty$, be a sequence of real-valued measurable functions on $(E, \mathscr{A}, \mu)$ with*

- $u_n(x) \xrightarrow{n \uparrow \infty} u(x)$ *for μ-almost all x,*
- $|u(x)| \leqslant w$ *for μ-almost all x and some positive $w \in \mathscr{L}^p(\mu)$ ($n \in \mathbb{N}$).*

Then $u \in \mathscr{L}^p(\mu)$ and it holds

(a) $\left\| u - u_n \right\|_{\mathrm{L}^p} \xrightarrow{n \uparrow \infty} 0,$

[1] To be precise, $u \mapsto \|u\|_{\mathrm{L}^p}$ $(1 \leqslant p \leqslant \infty)$ behaves almost like a norm, since we only have $\|u\|_{\mathrm{L}^p} \iff u = 0$ a.e.

(b) $\|u_n\|_{\mathrm{L}^p} \xrightarrow{n\uparrow\infty} \|u\|_{\mathrm{L}^p}$.

Mind that

$$\text{Convergence in } \mathrm{L}^p \lim_{n\to\infty} \|u-u_n\|_{\mathrm{L}^p} \quad \neq \quad \text{Convergence of } \mathrm{L}^p\text{-norms} \lim_{n\to\infty} \|u_n\|_{\mathrm{L}^p} = \lim_{n\to\infty} \|u\|_{\mathrm{L}^p}$$

This is reflected in the following theorem.

Theorem A.2 (Riesz). *Let $(u_n)_{n\in\mathbb{N}} \subset \mathscr{L}^p(\mu)$, $1 \leqslant p < \infty$. If $u_n(x) \xrightarrow{n\uparrow\infty} u(x)$ μ-a.e. and $u \in \mathscr{L}^p(\mu)$, then:*

$$\lim_{n\to\infty} \|u-u_n\|_{\mathrm{L}^p} \quad \Longleftrightarrow \quad \lim_{n\to\infty} \|u_n\|_{\mathrm{L}^p} = \lim_{n\to\infty} \|u\|_{\mathrm{L}^p}$$

Proof. $\Longrightarrow$ Δ-inequality backwards.

$\Longleftarrow$ Using $|u-u_n|^p \leqslant 2^p\left(|u_n|^p + |u|^p\right)$ & Fatou's lemma for $2^p\left(|u_n|^p+|u|^p\right) - |u-u_n|^p \geqslant 0$. ■

Theorem A.3 (Riesz-Fischer). *The space $\mathscr{L}^p(\mu)$, $1 \leqslant p < \infty$, is complete, i.e. every Cauchy sequence $(u_n)_{n\in\mathbb{N}} \subset \mathscr{L}^p(\mu)$ converges to some $u \in \mathscr{L}^p(\mu)$.*

Corollary A.4. *Let $(u_n)_{n\in\mathbb{N}} \subset \mathscr{L}^p(\mu)$, $1 \leqslant p \leqslant \infty$ with $u_n \xrightarrow{\mathrm{L}^p} u$, then there exists a subsequence $\left(u_{n(k)}\right)_{k\in\mathbb{N}}$ such that $u_{n(k)} \xrightarrow{n\uparrow\infty} u$ for almost all x.*

Limits in measure on a non-σ-finite measure space $(E, \mathscr{A}, \mu)$ need not be unique, but in probability we only work with finite measures of mass one. Vitali's theorem generalises Lebesgue's dominated convergence theorem.

Theorem A.5 (Vitali's theorem). *For $1 \leqslant p < \infty$, let $(X_n)_{n\in\mathbb{N}} \subset \mathrm{L}^p(\mathbb{P})$ a sequence of random variables with $X_n \xrightarrow{\mathbb{P}} X$. Then for all equivalent:*

(i) $X_n \xrightarrow{\mathrm{L}^p} X$

(ii) $\left(|X_n|^p\right)_{n\in\mathbb{N}}$ *is uniformly integrable*

(iii) $\mathbb{E}\,|X_n|^p \xrightarrow{n\uparrow\infty} \mathbb{E}\,|X|$

Remark A.6. Vitali's theorem A.5 still holds for measure spaces $(X, \mathscr{A}, \mu)$ which are not σ-finite. In this case, we can no longer identify the L^p-limit and the theorem reads: If $X_j \xrightarrow{\mathrm{w}} X$ (measurable function enough), then the following are equivalent:

(i) $\left(X_n\right)_{n\in\mathbb{N}}$ converges in L^p.

(ii) $\left(X_n\right)_{n\in\mathbb{N}}$ is uniformly integrable.

(iii) $\left(\|X_n\|_p\right)_{n\in\mathbb{N}}$ converges in $\mathbb{R}$.