2

Random variables in Banach spaces

In this lecture we take up the study of random variables with values in a Banach space E. The main result is the Itô-Nisio theorem (Theorem 2.17), which asserts that various modes of convergence of sums of independent symmetric E-valued random variables are equivalent. This result gives us a powerful tool to check the almost sure convergence of sums of independent symmetric random variables and will play an important role in the forthcoming lectures. The proof of the Itô-Nisio theorem is based on a uniqueness property of Fourier transforms (Theorem 2.8).

From this lecture onwards, we shall always assume that all spaces are real. This assumption is convenient when dealing with Fourier transforms and, in later lectures, when using the Riesz representation theorem to identify Hilbert spaces and their duals. However, much of the theory also works for complex scalars and can in fact be deduced from the real case. For some results it suffices to note that every complex vector space is a real space (by restricting the scalar multiplication to the reals); in others one proceeds by considering real and imaginary parts separately. We leave it to the interested reader to verify this in particular instances.

2.1 Random variables

A *probability space* is a triple $(\Omega, \mathscr{F}, \mathbb{P})$, where $\mathbb{P}$ is a *probability measure* on a measurable space $(\Omega, \mathscr{F})$, that is, $\mathbb{P}$ is a non-negative measure on $(\Omega, \mathscr{F})$ satisfying $\mathbb{P}(\Omega) = 1$.

Definition 2.1. *An E-valued* random variable *is an E-valued strongly $\mathbb{P}$-measurable function X defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$.*

We think of X as a 'random' element x of E, which explains the choice of the letter 'X'.

The underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$ will always be considered as fixed, and the prefix '$\mathbb{P}$-' will be omitted from our terminology unless confusion may arise. For instance, 'strongly measurable' means 'strongly $\mathbb{P}$-measurable' and 'almost surely' means '$\mathbb{P}$-almost surely', which is used synonymously with '$\mathbb{P}$-almost everywhere'. All integrals of E-valued random variables will be Bochner integrals unless stated otherwise, and the prefix 'Bochner' will usually be omitted.

The integral of an integrable random variable X is called its *mean value* or *expectation* and is denoted by

$$\mathbb{E}X := \int_{\Omega} X \, d\mathbb{P}.$$

If X is an E-valued random variable, then by Proposition 1.10 X has a strongly $\mathscr{F}$-measurable version $\widetilde{X}$ and by Proposition 1.8 the event

$$\{\widetilde{X} \in B\} := \{\omega \in \Omega : \ \widetilde{X}(\omega) \in B\}$$

belongs to $\mathscr{F}$ for all $B \in \mathscr{B}(E)$. The probability $\mathbb{P}\{\widetilde{X} \in B\}$ does not depend on the particular choice of the $\mathscr{F}$-measurable version $\widetilde{X}$, a fact which justifies the notation

$$\mathbb{P}\{X \in B\} := \mathbb{P}\{\widetilde{X} \in B\}$$

which will be used in the sequel without further notice.

Definition 2.2. *The* distribution *of an E-valued random variable X is the Borel probability measure μ_X on E defined by*

$$\mu_X(B) := \mathbb{P}\{X \in B\}, \qquad B \in \mathscr{B}(E).$$

Random variables having the same distribution are said to be identically distributed.

In the second part of this definition we allow the random variables to be defined on different probability spaces. If X and Y are identically distributed E-valued random variables and $f : E \to F$ is a Borel function, then $f(X)$ and $f(Y)$ are identically distributed. For example, for $1 \leqslant p < \infty$ it follows that

$$\mathbb{E}\|X\|^p = \mathbb{E}\|Y\|^p$$

if at least one (and then both) of these expectations are finite.

The next proposition shows that every E-valued random variable is *tight*:

Proposition 2.3. *If X is a random variable in E, then for every $\varepsilon > 0$ there exists a compact set K in E such that $\mathbb{P}\{X \notin K\} < \varepsilon$.*

Proof. Since X is separably valued outside some null set, we may assume that E is separable. Let $(x_n)_{n=1}^{\infty}$ be a dense sequence in E and fix $\varepsilon > 0$. For each integer $k \geqslant 1$ the closed balls $B(x_n, \frac{1}{k})$ cover E, and therefore there exists an index $N_k \geqslant 1$ such that

$$\mathbb{P}\Big\{X \in \bigcup_{n=1}^{N_k} B\big(x_n, \frac{1}{k}\big)\Big\} > 1 - \frac{\varepsilon}{2^k}.$$

The set $K := \bigcap_{k\geqslant 1} \bigcup_{n=1}^{N_k} B\big(x_n, \frac{1}{k}\big)$ is closed and totally bounded. Since E is complete, K is compact. Moreover,

$$\mathbb{P}\{X \notin K\} < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon. \qquad \square$$

This result motivates the following definition.

Definition 2.4. *A family $\mathscr{X}$ of random variables in E is* uniformly tight *if for every $\varepsilon > 0$ there exists a compact set K in E such that*

$$\mathbb{P}\{X \notin K\} < \varepsilon \qquad \forall X \in \mathscr{X}.$$

The following lemma will be useful in the proof of the Itô-Nisio theorem.

Lemma 2.5. *If $\mathscr{X}$ is uniformly tight, then $\mathscr{X} - \mathscr{X} = \{X_1 - X_2 : \ X_1, X_2 \in \mathscr{X}\}$ is uniformly tight.*

Proof. Let $\varepsilon > 0$ be arbitrary and fixed. Choose a compact set K in E such that $\mathbb{P}\{X \in K\} \geqslant 1 - \varepsilon$ for all $X \in \mathscr{X}$. The set $L = \{x - y : \ x, y \in K\}$ is compact, being the image of the compact set $K \times K$ under the continuous map $(x, y) \mapsto x - y$. Since $X_1(\omega), X_2(\omega) \in K$ implies $X_1(\omega) - X_2(\omega) \in L$,

$$\mathbb{P}\{X_1 - X_2 \notin L\} \leqslant \mathbb{P}\{X_1 \notin K\} + \mathbb{P}\{X_2 \notin K\} < 2\varepsilon. \qquad \square$$

2.2 Fourier transforms

We begin with a definition.

Definition 2.6. *The* Fourier transform *of a Borel probability measure μ on E is the function $\widehat{\mu} : E^* \to \mathbb{C}$ defined by*

$$\widehat{\mu}(x^*) := \int_E \exp(-i\langle x, x^*\rangle)\, d\mu(x).$$

The Fourier transform *of a random variable $X : \Omega \to E$ is the Fourier transform of its distribution μ_X.*

Note that the above integral converges absolutely, as $|\exp(-i\langle x, x^*\rangle)| = 1$ for all $x \in E$ since we are assuming that E is a real Banach space. By a change of variable, the Fourier transform of a random variable X on E is given by

$$\widehat{X}(x^*) := \mathbb{E}\exp(-i\langle X, x^*\rangle) = \int_E \exp(-i\langle x, x^*\rangle)\, d\mu_X(x).$$

The proof of the next theorem is based upon a uniqueness result known as *Dynkin's lemma*. It states that two probability measures agree if they agree on a sufficiently rich family of sets.

Lemma 2.7 (Dynkin). *Let μ_1 and μ_2 be two probability measures defined on a measurable space $(\Omega, \mathscr{F})$. Let $\mathscr{A} \subseteq \mathscr{F}$ be a collection of sets with the following properties:*

(1) *$\mathscr{A}$ is closed under finite intersections;*
(2) *$\sigma(\mathscr{A})$, the σ-algebra generated by $\mathscr{A}$, equals $\mathscr{F}$.*

If $\mu_1(A) = \mu_2(A)$ for all $A \in \mathscr{A}$, then $\mu_1 = \mu_2$.

Proof. Let $\mathscr{D}$ denote the collection of all sets $D \in \mathscr{F}$ with $\mu_1(D) = \mu_2(D)$. Then $\mathscr{A} \subseteq \mathscr{D}$ and $\mathscr{D}$ is a *Dynkin system*, that is,

- $\Omega \in \mathscr{D}$;
- if $D_1 \subseteq D_2$ with $D_1, D_2 \in \mathscr{D}$, then also $D_2 \setminus D_1 \in \mathscr{D}$;
- if $D_1 \subseteq D_2 \subseteq \dots$ with all $D_n \in \mathscr{D}$, then also $\bigcup_{n\geqslant 1} D_n \in \mathscr{D}$.

By assumption we have $\mathscr{D} \subseteq \mathscr{F} = \sigma(\mathscr{A})$; we will show that $\sigma(\mathscr{A}) \subseteq \mathscr{D}$. To this end let $\mathscr{D}_0$ denote the smallest Dynkin system in $\mathscr{F}$ containing $\mathscr{A}$. We will show that $\sigma(\mathscr{A}) \subseteq \mathscr{D}_0$. In view of $\mathscr{D}_0 \subseteq \mathscr{D}$, this will prove the lemma.

Let $\mathscr{C} = \{D_0 \in \mathscr{D}_0 : \ D_0 \cap A \in \mathscr{D}_0 \text{ for all } A \in \mathscr{A}\}$. Then $\mathscr{C}$ is a Dynkin system and $\mathscr{A} \subseteq \mathscr{C}$ since $\mathscr{A}$ is closed under taking finite intersections. It follows that $\mathscr{D}_0 \subseteq \mathscr{C}$, since $\mathscr{D}_0$ is the smallest Dynkin system containing $\mathscr{A}$. But obviously, $\mathscr{C} \subseteq \mathscr{D}_0$, and therefore $\mathscr{C} = \mathscr{D}_0$.

Now let $\mathscr{C}' = \{D_0 \in \mathscr{D}_0 : \ D_0 \cap D \in \mathscr{D}_0 \text{ for all } D \in \mathscr{D}_0\}$. Then $\mathscr{C}'$ is a Dynkin system and the fact that $\mathscr{C} = \mathscr{D}_0$ implies that $\mathscr{A} \subseteq \mathscr{C}'$. Hence $\mathscr{D}_0 \subseteq \mathscr{C}'$, since $\mathscr{D}_0$ is the smallest Dynkin system containing $\mathscr{A}$. But obviously, $\mathscr{C}' \subseteq \mathscr{D}_0$, and therefore $\mathscr{C}' = \mathscr{D}_0$.

It follows that $\mathscr{D}_0$ is closed under taking finite intersections. But a Dynkin system with this property is a σ-algebra. Thus, $\mathscr{D}_0$ is a σ-algebra, and now $\mathscr{A} \subseteq \mathscr{D}_0$ implies that also $\sigma(\mathscr{A}) \subseteq \mathscr{D}_0$. ☐

Theorem 2.8 (Uniqueness of the Fourier transform). *Let X_1 and X_2 be E-valued random variables whose Fourier transforms are equal:*

$$\widehat{X_1}(x^*) = \widehat{X_2}(x^*) \qquad \forall x^* \in E^*.$$

Then X_1 and X_2 are identically distributed.

Proof. Since X_1 and X_2 are μ-separably valued there is no loss of generality in assuming that E is separable.

Step 1 - First we prove: if λ_1 and λ_2 are Borel probability measures on $\mathbb{R}^d$ with the property that $\widehat{\lambda_1}(t) = \widehat{\lambda_2}(t)$ for all $t \in \mathbb{R}^d$, then $\lambda_1 = \lambda_2$. By Dynkin's lemma, for the latter it suffices to prove that $\lambda_1(K) = \lambda_2(K)$ for all compact subsets K of $\mathbb{R}^d$. By the dominated convergence theorem, for the latter suffices to prove that

$$\int_{-\infty}^{\infty} f(\xi)\,d\lambda_1(\xi) = \int_{-\infty}^{\infty} f(\xi)\,d\lambda_2(\xi) \qquad \forall f \in C_c(\mathbb{R}^d), \tag{2.1}$$

where $C_c(\mathbb{R}^d)$ denote the space of all compactly supported continuous functions on $\mathbb{R}^d$.

Let $\varepsilon > 0$ be arbitrary and fix an $f \in C_c(\mathbb{R}^d)$. We may assume that $\|f\|_\infty \leqslant 1$. Let $r > 0$ be so large that the support of f is contained in $[-r,r]^d$ and such that $\lambda_j\big(\complement[-r,r]^d\big) \leqslant \varepsilon$ for $j = 1, 2$. By the Stone-Weierstrass theorem there exists a trigonometric polynomial $p : \mathbb{R}^d \to \mathbb{C}$ of period $2r$ such that $\sup_{t\in[-r,r]^d} |f(t) - p(t)| \leqslant \varepsilon$. Then,

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^d} f(\xi)\,d\lambda_1(\xi) - \int_{\mathbb{R}^d} f(\xi)\,d\lambda_2(\xi)\Big| \\
&\qquad \leqslant 4\varepsilon + 2(1+\varepsilon)\varepsilon + \Big|\int_{\mathbb{R}^d} p(\xi)\,d\lambda_1(\xi) - \int_{\mathbb{R}^d} p(\xi)\,d\lambda_2(\xi)\Big| \\
&\qquad = 4\varepsilon + 2(1+\varepsilon)\varepsilon,
\end{aligned}$$

where the terms $2(1+\varepsilon)\varepsilon$ come from the estimate $\|p\|_\infty \leqslant 1 + \varepsilon$ and the last equality follows from the equality of the Fourier transforms of λ_1 and λ_2. Since $\varepsilon > 0$ was arbitrary, this proves (2.1).

Step 2 - If μ is any Borel probability measure on E, then for all $d \geqslant 1$ and all $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ and $x_1^*, \dots, x_d^* \in E^*$ we have

$$\widehat{\mu}\Big(\sum_{j=1}^d t_j x_j^*\Big) = \int_E e^{-i\sum_{j=1}^d \langle x, t_j x_j^*\rangle}\,d\mu(x) = \int_{\mathbb{R}^d} e^{-i\langle t,\xi\rangle}\,d(T\mu)(\xi) = \widehat{T\mu}(t),$$

where $T\mu$ denotes Borel probability measure on $\mathbb{R}^d$ obtained as the image measure of μ under the map $T : E \to \mathbb{R}^d$, $x \mapsto (\langle x, x_1^*\rangle, \dots, \langle x, x_d^*\rangle)$, that is,

$$T\mu(B) := \mu\big\{x \in E :\ (\langle x, x_1^*\rangle,\ \dots,\ \langle x, x_d^*\rangle) \in B\big\}.$$

Step 3 - Applying Step 2 to the measures μ_{X_1} and μ_{X_2} it follows that $\widehat{T\mu_{X_1}}(t) = \widehat{T\mu_{X_2}}(t)$ for all $t \in \mathbb{R}^d$. By Step 1, $T\mu_{X_1} = T\mu_{X_2}$. Hence μ_{X_1} and μ_{X_2} agree on the collection $\mathscr{C}(E)$ consisting of all Borel sets in E of the form

$$\big\{x \in E :\ (\langle x, x_1^*\rangle,\ \dots,\ \langle x, x_d^*\rangle) \in B\big\}$$

with $d \geqslant 1$, $x_1^*, \dots, x_d^* \in E^*$ and $B \in \mathscr{B}(\mathbb{R}^d)$. Since E is separable, every closed ball $\{x \in E :\ \|x - x_0\| \leqslant r\}$ can be written as a countable intersection of sets in $\mathscr{C}(E)$ (see Exercise 1.1). Thus the family $\mathscr{C}(E)$ generates the Borel σ-algebra $\mathscr{B}(E)$ and $\mu_{X_1} = \mu_{X_2}$ by Dynkin's Lemma. $\square$

2.3 Convergence in probability

In the absence of integrability conditions the following definition for convergence of random variables is often very useful.

Definition 2.9. *A sequence $(X_n)_{n=1}^\infty$ of E-valued random variables* converges in probability *to an E-valued random variable X if for all $r > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\{\|X_n - X\| > r\} = 0.$$

If $\lim_{n\to\infty} X_n = X$ in $L^p(\Omega; E)$ for some $1 \leqslant p < \infty$, then $\lim_{n\to\infty} X_n = X$ in probability. This follows from *Chebyshev's inequality*, which states that if $\xi \in L^p(\Omega)$, then for all $r > 0$ we have

$$\mathbb{P}\{|\xi| \geqslant r\} \leqslant \frac{1}{r^p}\mathbb{E}|\xi|^p.$$

The proof is simple:

$$\mathbb{P}\{|\xi| \geqslant r\} = \frac{1}{r^p}\int_{\{|\xi|^p \geqslant r^p\}} r^p \, d\mathbb{P} \leqslant \frac{1}{r^p}\int_{\{|\xi|^p \geqslant r^p\}} |\xi|^p \, d\mathbb{P} \leqslant \frac{1}{r^p}\mathbb{E}|\xi|^p.$$

Our first aim is to show that if $(X_n)_{n=1}^\infty$ converges in probability, then some subsequence converges almost surely. For this we need a lemma which is known as the *Borel-Cantelli lemma*.

Lemma 2.10 (Borel-Cantelli). *If $(A, \mathscr{A}, \mu)$ is a measure space and $(A_n)_{n=1}^\infty$ is a sequence in $\mathscr{A}$ satisfying $\sum_{n=1}^\infty \mu(A_n) < \infty$, then*

$$\mu\Big(\bigcap_{k\geqslant 1}\bigcup_{n\geqslant k} A_n\Big) = 0.$$

Proof. Let $k_0 \geqslant 1$. Then,

$$\mu\Big(\bigcap_{k\geqslant 1}\bigcup_{n\geqslant k} A_n\Big) \leqslant \mu\Big(\bigcup_{n\geqslant k_0} A_n\Big) \leqslant \sum_{n=k_0}^\infty \mu(A_n),$$

and the right hand side tends to 0 as $k_0 \to \infty$. □

Note that $\omega \in \bigcap_{k\geqslant 1}\bigcup_{n\geqslant k} A_n$ if and only if $\omega \in A_n$ for infinitely many indices n.

Proposition 2.11. *If a sequence $(X_n)_{n=1}^\infty$ of E-valued random variables converges in probability, then it has an almost surely convergent subsequence $(X_{n_k})_{k=1}^\infty$.*

Proof. Let $\lim_{n\to\infty} X_n = X$ in probability. Choose an increasing sequence of indices $n_1 < n_2 < \dots$ satisfying

$$\mathbb{P}\Big\{\|X_{n_k} - X\| > \frac{1}{k}\Big\} < \frac{1}{2^k} \qquad \forall k \geqslant 1.$$

By the Borel-Cantelli lemma,

$$\mathbb{P}\Big\{\|X_{n_k} - X\| > \frac{1}{k} \text{ for infinitely many } k \geqslant 1\Big\} = 0.$$

Outside this null set we have $\lim_{k\to\infty} X_{n_k} = X$ pointwise. $\square$

2.4 Independence

Next we recall the notion of independence. The reader who is already familiar with it may safely skip this section.

Definition 2.12. *A family of random variables $(X_i)_{i\in I}$, where I is some index set and each X_i takes values in a Banach space E_i, is* independent *if for all choices of distinct indices $i_1, \dots, i_N \in I$ and all Borel sets $B_1, \dots, B_N$ in $E_{i_1}, \dots, E_{i_N}$ we have*

$$\mathbb{P}\{X_{i_1} \in B_1,\ \dots,\ X_{i_N} \in B_N\} = \prod_{n=1}^{N} \mathbb{P}\{X_{i_n} \in B_n\}.$$

Note that $(X_i)_{i\in I}$ is independent if and only if every finite subfamily of $(X_i)_{i\in I}$ is independent. Thus, in order to check independence of a given family of random variables it suffices to consider its finite subfamilies.

We assume that the reader is familiar with the elementary properties of independent real-valued random variables such as covered in a standard course on probability. Here we content ourselves recalling that if η and ξ are real-valued random variables which are integrable and independent, then their product $\eta\xi$ is integrable and $\mathbb{E}(\eta\xi) = \mathbb{E}\eta\,\mathbb{E}\xi$.

In the next two propositions, $X_1, \dots, X_N$ are random variables with values in the Banach spaces $E_1, \dots, E_N$, respectively. If $\nu_1, \dots, \nu_n$ are probability measures, we denote by $\nu_1 \times \dots \times \nu_n$ their product measure. The distribution of the E^N-valued random variable $(X_1, \dots, X_N)$ is denoted by $\mu_{(X_1,\dots,X_N)}$.

Proposition 2.13. *The random variables $X_1, \dots, X_N$ are independent if and only if*

$$\mu_{(X_1,\dots,X_N)} = \mu_{X_1} \times \dots \times \mu_{X_N}.$$

Proof. By definition, the random variables $X_1, \dots, X_N$ are independent if and only if $\mu_{(X_1,\dots,X_N)}$ and $\mu_{X_1} \times \dots \times \mu_{X_N}$ agree on all Borel rectangles $B_1 \times \dots \times B_N$ in $E_1 \times \dots \times E_N$. By Dynkin's lemma this happens if and only if $\mu_{(X_1,\dots,X_N)} = \mu_{X_1} \times \dots \times \mu_{X_N}$. $\square$

We record two corollaries.

Proposition 2.14. *If* $\lim_{n\to\infty} X_n = X$ *and* $\lim_{n\to\infty} Y_n = Y$ *in probability and each* X_n *is independent of* Y_n*, then* X *and* Y *are independent.*

Proof. By passing to a subsequence we may assume that $\lim_{n\to\infty} X_n = X$ and $\lim_{n\to\infty} Y_n = Y$ almost surely. We consider the $E \times E$-valued random variables $Z_n = (X_n, Y_n)$ and $Z = (X, Y)$. Identifying the dual of $E \times E$ with $E^* \times E^*$, by dominated convergence we obtain

$$\begin{aligned}
\widehat{\mu}_Z(x^*, y^*) &= \mathbb{E}\exp(-i(\langle X, x^*\rangle + \langle Y, y^*\rangle)) \\
&= \lim_{n\to\infty} \mathbb{E}\exp(-i(\langle X_n, x^*\rangle + \langle Y_n, y^*\rangle)) \\
&= \lim_{n\to\infty} \mathbb{E}\exp(-i\langle X_n, x^*\rangle)\mathbb{E}\exp(-i\langle Y_n, y^*\rangle) \\
&= \mathbb{E}\exp(-i\langle X, x^*\rangle)\mathbb{E}\exp(-i\langle Y, y^*\rangle) \\
&= \widehat{\mu_X}(x^*)\widehat{\mu_Y}(y^*) = \widehat{\mu_X \times \mu_Y}(x^*, y^*).
\end{aligned}$$

From Theorem 2.8 we conclude that $\mu_Z = \mu_X \times \mu_Y$. Now the result follows from Proposition 2.13. $\square$

Definition 2.15. *An* E*-valued random variable* X *is called* symmetric *if* X *and* $-X$ *are identically distributed.*

Proposition 2.16. *If* X *is symmetric and independent of* Y*, then for all* $1 \leqslant p < \infty$ *we have*

$$\mathbb{E}\|X\|^p \leqslant \mathbb{E}\|X + Y\|^p.$$

Proof. The symmetry of X and the independence of X and Y imply that $X + Y$ and $-X + Y$ are identically distributed, and therefore

$$\begin{aligned}
\big(\mathbb{E}\|X\|^p\big)^{\frac{1}{p}} &= \tfrac{1}{2}\big(\mathbb{E}\|(X+Y) + (X-Y)\|^p\big)^{\frac{1}{p}} \\
&\leqslant \tfrac{1}{2}\big(\mathbb{E}\|X+Y\|^p\big)^{\frac{1}{p}} + \tfrac{1}{2}\big(\mathbb{E}\|X-Y\|^p\big)^{\frac{1}{p}} = \big(\mathbb{E}\|X+Y\|^p\big)^{\frac{1}{p}}. \quad \square
\end{aligned}$$

2.5 The Itô-Nisio theorem

In this section we prove a celebrated result, due to Itô and Nisio, which states that a sum of symmetric and independent E-valued random variables converges (weakly) almost surely if and only if it converges in probability.

Here is the precise statement of the theorem:

Theorem 2.17 (Itô-Nisio). *Let* $X_n : \Omega \to E$, $n \geqslant 1$, *be independent symmetric random variables, put* $S_n := \sum_{j=1}^n X_j$*, and let* $S : \Omega \to E$ *be a random variable. The following assertions are equivalent:*

(1) *for all* $x^* \in E^*$ *we have* $\lim_{n\to\infty}\langle S_n, x^*\rangle = \langle S, x^*\rangle$ *almost surely;*
(2) *for all* $x^* \in E^*$ *we have* $\lim_{n\to\infty}\langle S_n, x^*\rangle = \langle S, x^*\rangle$ *in probability;*
(3) *we have* $\lim_{n\to\infty} S_n = S$ *almost surely;*
(4) *we have* $\lim_{n\to\infty} S_n = S$ *in probability.*

If these equivalent conditions hold and $\mathbb{E}\|S\|^p < \infty$ *for some* $1 \leqslant p < \infty$*, then*

$$\lim_{n\to\infty} \mathbb{E}\|S_n - S\|^p = 0.$$

We begin with a tail estimate known as *Lévy's inequality.*

Lemma 2.18. *Let* $X_1, \dots, X_n$ *be independent symmetric* E*-valued random variables, and put* $S_k := \sum_{j=1}^k X_j$ *for* $k = 1, \dots, n$*. Then for all* $r \geqslant 0$ *we have*

$$\mathbb{P}\Big\{\max_{1\leqslant k\leqslant n} \|S_k\| > r\Big\} \leqslant 2\mathbb{P}\{\|S_n\| > r\}.$$

Proof. Put

$$\begin{aligned} A &:= \Big\{\max_{1\leqslant k\leqslant n} \|S_k\| > r\Big\},\\ A_k &:= \{\|S_1\| \leqslant r, \dots, \|S_{k-1}\| \leqslant r, \|S_k\| > r\}; \quad k = 1, \dots, n. \end{aligned}$$

The sets $A_1, \dots, A_n$ are disjoint and $\bigcup_{k=1}^n A_k = \Big\{\max_{1\leqslant k\leqslant n} \|S_k\| > r\Big\}$.

The identity $S_k = \frac{1}{2}(S_n + (2S_k - S_n))$ implies that

$$\{\|S_k\| > r\} \subseteq \{\|S_n\| > r\} \cup \{\|2S_k - S_n\| > r\}.$$

We also note $(X_1, \dots, X_n)$ and $(X_1, \dots, X_k, -X_{k+1}, \dots, -X_n)$ are identically distributed (see Exercise 2), which, in view of the identities

$$S_n = S_k + X_{k+1} + \dots + X_n, \qquad 2S_k - S_n = S_k - X_{k+1} - \dots - X_n,$$

implies that $(X_1, \dots, X_k, S_n)$ and $(X_1, \dots, X_k, 2S_k - S_n)$ are identically distributed. Hence,

$$\begin{aligned} \mathbb{P}(A_k) &\leqslant \mathbb{P}(A_k \cap \{\|S_n\| > r\}) + \mathbb{P}(A_k \cap \{\|2S_k - S_n\| > r\})\\ &= 2\mathbb{P}(A_k \cap \{\|S_n\| > r\}). \end{aligned}$$

Summing over k we obtain

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A_k) \leqslant 2\sum_{k=1}^n \mathbb{P}(A_k \cap \{\|S_n\| > r\}) = 2\mathbb{P}\{\|S_n\| > r\}. \qquad \square$$

Proof (Proof of Theorem 2.17). We prove the implications (2)⇒(4)⇒(3), the implications (3)⇒(1)⇒(2) being clear.

(2)⇒(4): We split this proof into two steps.

Step 1 – In this step we prove that the sequence $(S_n)_{n\geqslant 1}$ is uniformly tight.

For all $m \geqslant n$ and $x^* \in E^*$ the random variables $\langle S_m - S_n, x^*\rangle$ and $\pm\langle S_n, x^*\rangle$ are independent. Hence by Proposition 2.14, $\langle S - S_n, x^*\rangle$ and $\pm\langle S_n, x^*\rangle$ are independent. Next we claim that S and $S - 2S_n$ are identically distributed. Indeed, denote their distributions by μ and λ_n, respectively. By the independence of $\langle S - S_n, x^*\rangle$ and $\pm\langle S_n, x^*\rangle$ and the symmetry of S_n, for all $x^* \in E^*$ we have

$$\begin{aligned}\widehat{\mu}(x^*) = \mathbb{E}\left(e^{-i\langle S,x^*\rangle}\right) &= \mathbb{E}\left(e^{-i\langle S-S_n,x^*\rangle}\right)\cdot \mathbb{E}\left(e^{-i\langle S_n,x^*\rangle}\right)\\ &= \mathbb{E}\left(e^{-i\langle S-S_n,x^*\rangle}\right)\cdot \mathbb{E}\left(e^{-i\langle -S_n,x^*\rangle}\right)\\ &= \mathbb{E}\left(e^{-i\langle S-2S_n,x^*\rangle}\right) = \widehat{\lambda_n}(x^*).\end{aligned}$$

By Theorem 2.8, this shows that $\mu = \lambda_n$ and the claim is proved.

Given $\varepsilon > 0$ we can find a compact set $K \subseteq E$ with $\mu(K) = \mathbb{P}\{S \in K\} > 1 - \varepsilon$. The set $L := \frac{1}{2}(K - K)$ is compact as well, and arguing as in the proof of Lemma 2.5 we have

$$\mathbb{P}\{S_n \notin L\} \leqslant \mathbb{P}\{S \notin K\} + \mathbb{P}\{S - 2S_n \notin K\} = 2\mathbb{P}\{S \notin K\} < 2\varepsilon.$$

It follows that $\mathbb{P}\{S_n \in L\} > 1 - 2\varepsilon$ for all $n \geqslant 1$, and therefore the sequence $(S_n)_{n=1}^{\infty}$ is uniformly tight.

Step 2 – By Lemma 2.5, the sequence $(S_n - S)_{n\geqslant 1}$ is uniformly tight. Let ν_n denote the distribution of $S_n - S$. We need to prove that for all $\varepsilon > 0$ and $r > 0$ there exists an index $N \geqslant 1$ such that

$$\mathbb{P}\{\|S_n - S\| \geqslant r\} = \nu_n(\complement B(0,r)) < \varepsilon \qquad \forall n \geqslant N.$$

Suppose, for a contradiction, that such an N does not exist for some $\varepsilon > 0$ and $r > 0$. Then there exists a subsequence $(S_{n_k})_{k\geqslant 1}$ such that

$$\nu_{n_k}(\complement B(0,r)) \geqslant \varepsilon, \qquad k \geqslant 1.$$

On the other hand, by uniform tightness we find a compact set K such that $\nu_{n_k}(K) \geqslant 1 - \frac{1}{2}\varepsilon$ for all $k \geqslant 1$. It follows that

$$\nu_{n_k}(K \cap \complement B(0,r)) \geqslant \tfrac{1}{2}\varepsilon, \qquad k \geqslant 1.$$

By covering the compact set $K \cap \complement B(0,r)$ with open balls of radius $\frac{1}{2}r$ and passing to a subsequence, we find a ball B not containing 0 and a number $\delta > 0$ such that

$$\nu_{n_{k_j}}(K \cap B) = \mathbb{P}\{S_{n_{k_j}} - S \in K \cap B\} \geqslant \delta, \qquad j \geqslant 1.$$

By the Hahn-Banach separation theorem, there is a functional $x^* \in E^*$ such that $\langle x, x^* \rangle \geqslant 1$ for all $x \in B$. For all $\omega \in \{S_{n_k} - S \in K \cap B\}$ it follows that $\langle S_{n_{k_j}}(\omega) - S(\omega), x^* \rangle \geqslant 1$. Thus, $\langle S_{n_k}, x^* \rangle$ fails to converge to $\langle S, x^* \rangle$ in probability. This contradiction concludes the proof.

(4)⇒(3): Assume that $\lim_{n\to\infty} S_n = S$ in probability for some random variable S. By Proposition 2.11 there is a subsequence $(S_{n_k})_{k=1}^{\infty}$ converging almost surely to S. Fix k and let $m > n_k$. Then by Lévy's inequality,

$$\begin{aligned}\mathbb{P}\Big\{ \sup_{n_k \leqslant j \leqslant m} \|S_j - S_{n_k}\| \geqslant r \Big\} &\leqslant 2\mathbb{P}(\|S_m - S_{n_k}\| \geqslant r) \\ &\leqslant 2\mathbb{P}\Big\{\|S_m - S\| \geqslant \frac{r}{2}\Big\} + 2\mathbb{P}\Big\{\|S - S_{n_k}\| \geqslant \frac{r}{2}\Big\}.\end{aligned}$$

Letting $m \to \infty$ we find

$$\mathbb{P}\Big\{ \sup_{j \geqslant n_k} \|S_j - S_{n_k}\| \geqslant r \Big\} \leqslant 2\mathbb{P}\Big\{\|S - S_{n_k}\| \geqslant \frac{r}{2}\Big\},$$

and hence, upon letting $k \to \infty$,

$$\lim_{k\to\infty} \mathbb{P}\Big\{ \sup_{j \geqslant n_k} \|S_j - S_{n_k}\| \geqslant r \Big\} = 0.$$

Since $S_{n_k} \to S$ pointwise a.e., it follows that

$$\begin{aligned}\mathbb{P}\Big\{ \lim_{k\to\infty} \sup_{j \geqslant n_k} \|S_j - S\| \geqslant 2r \Big\} &\leqslant \lim_{k\to\infty} \mathbb{P}\Big\{ \sup_{j \geqslant n_k} \|S_j - S\| \geqslant 2r \Big\} \\ &\leqslant \lim_{k\to\infty} \mathbb{P}\Big\{ \sup_{j \geqslant n_k} \|S_j - S_{n_k}\| \geqslant r \Big\} + \lim_{k\to\infty} \mathbb{P}\Big\{ \sup_{j \geqslant n_k} \|S_{n_k} - S\| \geqslant r \Big\} = 0.\end{aligned}$$

It remains to prove the assertion about L^p-convergence. First we note that $S = S_n + (S - S_n)$ with S_n and $S - S_n$ independent (by the independence of S_n and $S_m - S_n$ for $m \geqslant n$ and Proposition 2.14), and therefore $\mathbb{E}\|S_n\|^p \leqslant \mathbb{E}\|S\|^p$ by Proposition 2.16. Hence by an integration by parts (see Exercise 1) and Lévy inequality,

$$\begin{aligned}\mathbb{E} \sup_{1 \leqslant k \leqslant n} \|S_k\|^p &= \int_0^\infty p r^{p-1} \mathbb{P}\Big\{ \sup_{1 \leqslant k \leqslant n} \|S_k\| > r \Big\}\, dr \\ &\leqslant 2 \int_0^\infty p r^{p-1} \mathbb{P}\{\|S_n\| > r\}\, dr = 2\mathbb{E}\|S_n\|^p \leqslant 2\mathbb{E}\|S\|^p.\end{aligned}$$

Hence $\mathbb{E}\sup_{k \geqslant 1} \|S_k\|^p \leqslant 2\mathbb{E}\|S\|^p$ by the monotone convergence theorem. Now $\lim_{n\to\infty} \|S_n - S\|^p = 0$ follows from the dominated convergence theorem. □

2.6 Exercises

1. (!) Let ξ be a non-negative random variable and let $1 \leqslant p < \infty$. Prove the *integration by parts formula*

$$\mathbb{E}\xi^p = \int_0^\infty p\lambda^{p-1}\mathbb{P}\{\xi > \lambda\}\,d\lambda.$$

Hint: Write $\mathbb{P}\{\xi > \lambda\} = \mathbb{E}1_{\{\xi>\lambda\}}$ and apply Fubini's theorem.

2. (!) Let $X_1, \dots, X_N$ be independent symmetric E-valued random variables. Show that for all choices of $\varepsilon_1, \dots, \varepsilon_N \in \{-1, +1\}$ the E^N-valued random variables $(X_1, \dots, X_N)$ and $(\varepsilon_1 X_1, \dots, \varepsilon_N X_N)$ are identically distributed.
3. (!) Define the *convolution* of two Borel measures μ and ν on E by
$$\mu * \nu(B) := \int_E \int_E 1_B(x+y)\,d\mu(x)\,d\nu(y), \quad B \in \mathscr{B}(E).$$
Prove that for all $x^* \in E^*$ we have $\widehat{\mu * \nu}(x^*) = \widehat{\mu}(x^*)\widehat{\nu}(x^*)$.
4. A sequence of E-valued random variables $(X_n)_{n=1}^\infty$ is *Cauchy in probability* if for all $\varepsilon > 0$ and $r > 0$ there exists an index $N \geqslant 1$ such that
$$\mathbb{P}\{\|X_n - X_m\| > r\} < \varepsilon \qquad \forall m, n \geqslant N.$$
Show that $(X_n)_{n=1}^\infty$ is Cauchy in probability if and only if $(X_n)_{n=1}^\infty$ converges in probability.
Hint: For the 'if' part, first show that some subsequence of $(X_n)_{n=1}^\infty$ converges almost surely.
5. Let $(X_n)_{n=1}^\infty$ be a sequence of E-valued random variables. Prove that if $\lim_{n\to\infty} X_n = X$ in probability, then $(X_n)_{n=1}^\infty$ is uniformly tight.

Notes. There are many excellent introductory texts on Probability Theory, among them the classic by Chung [21]. The more analytically inclined reader might consult Stromberg [101]. A comprehensive treatment of modern Probability Theory is offered by Kallenberg [55].

Thorough discussions of Banach space-valued random variables can be found in the monographs by Kwapień and Woyczyński [65], Ledoux and Talagrand [69], and Vakhania, Tarieladze, and Chobanyan [105].

The Itô-Nisio theorem was proved by Itô and Nisio in their beautiful paper [52] which we recommend for further reading. The usual proofs of this theorem are based upon the following celebrated and non-trivial compactness theorem due to Prokhorov:

Theorem 2.19 (Prokhorov). *For a family $\mathscr{M}$ of Borel probability measures on a separable complete metric space M the following assertions are equivalent:*

(1) *$\mathscr{M}$ is uniformly tight;*
(2) *Every sequence $(\mu_n)_{n=1}^\infty$ in $\mathscr{M}$ has a weakly convergent subsequence.*

Here, (1) means that for all $\varepsilon > 0$ there exists a compact set K in M such that $\mu(\complement K) < \varepsilon$ for all $\mu \in \mathscr{M}$, and (2) means that there exist a subsequence $(\mu_{n_k})_{k\geqslant 1}$ and a Borel probability measure μ such that

$$\lim_{k\to\infty} \int_M f \, d\mu_{n_k} = \int_M f \, d\mu$$

for all bounded continuous functions $f : M \to \mathbb{R}$. This theorem is the starting point of measure theory on metric spaces. Expositions of this subject can be found in the monographs by Billingsley [7] and Parthasarathy [88], as well as in the recent two-volume treatise on measure theory by Bogachev [9]. Readers familiar with it will have noticed that some of the results which we have stated for E-valued random variables, such as Proposition 2.3 and Theorem 2.8, could just as well be stated for probability measures on E.