Chapter 6
On Convergence of Random Variables and Distributions

Abstract In this chapter, several different types of convergence used in Probability Theory are defined and relationships between them are elucidated. Section 6.1 deals with convergence in probability and convergence with probability one (the almost sure convergence), presenting some criteria for them and, in particular, discussing the concept of Cauchy sequences (in probability and almost surely). Then the continuity theorem is established (convergence of functions of random variables) and the concept of uniform integrability is introduced and discussed, together with its consequences (in particular, for convergence in mean of suitable orders). Section 6.2 contains an extensive discussion of weak convergence of distributions. The chapter ends with Sect. 6.3 presenting criteria for weak convergence of distributions, including the concept of distribution determining classes of functions and that of tightness.

6.1 Convergence of Random Variables

In previous chapters we have already encountered several assertions which dealt with convergence, in some sense, of the distributions of random variables or of the random variables themselves. Now we will give definitions of different types of convergence and elucidate the relationships between them.

6.1.1 Types of Convergence

Let a sequence of random variables $\{\xi_n\}$ and a random variable ξ be given on a probability space $\langle\Omega, \mathfrak{F}, \mathbf{P}\rangle$.

Definition 6.1.1 The sequence $\{\xi_n\}$ *converges in probability*[1] to ξ if, for any $\varepsilon > 0$,

$$\mathbf{P}\big(|\xi_n - \xi| > \varepsilon\big) \to 0 \quad \text{as } n \to \infty.$$

[1] In the set-theoretic terminology, convergence in probability means convergence in measure.

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9_6,

One writes this as

$$S_n \xrightarrow{p} \xi \quad \text{as } n \to \infty.$$

In this notation, the assertion of the law of large numbers for the Bernoulli scheme could be written as

$$\frac{S_n}{n} \xrightarrow{p} p,$$

since S_n/n can be considered as a sequence of random variables given on a common probability space.

Definition 6.1.2 We will say that the *sequence* ξ_n *converges to* ξ *with probability* 1 (or *almost surely*: $\xi_n \to \xi$ a.s., $\xi_n \xrightarrow{a.s.} \xi$), if $\xi_n(\omega) \to \xi(\omega)$ as $n \to \infty$ for all $\omega \in \Omega$ except for ω from a set $N \subset \Omega$ of null probability: $\mathbf{P}(N) = 0$. This convergence can also be called convergence *almost everywhere* (a.e.) with respect to the measure $\mathbf{P}$.

Convergence $\xi_n \xrightarrow{a.s.} \xi$ implies convergence $\xi_n \xrightarrow{p} \xi$. Indeed, if we assume that the convergence in probability does not take place then there exist $\varepsilon > 0$, $\delta > 0$, and a sequence n_k such that, for the sequence of events $A_k = \{|\xi_{n_k} - \xi| > \varepsilon\}$, we have $\mathbf{P}(A_k) \ge \delta$ for all k. Let B consist of all elementary events belonging to infinitely many A_k, i.e. $B = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k$. Then, clearly for $\omega \in B$, the convergence $\xi_n(\omega) \to \xi(\omega)$ is impossible. But $B = \bigcap_{m=1}^{\infty} B_m$, where $B_m = \bigcup_{k \ge m} A_k$ are decreasing events ($B_{m+1} \subset B_m$), $\mathbf{P}(B_m) \ge \mathbf{P}(A_{n_m}) \ge \delta$ and, by the continuity axiom, $\mathbf{P}(B_m) \to \mathbf{P}(B)$ as $m \to \infty$. Therefore $\mathbf{P}(B) \ge \delta$ and a.s. convergence is impossible. The obtained contradiction proves the desired statement. □

The converse assertion, that convergence in probability implies a.s. convergence, is, generally speaking, not true, as we will see below. However in one important special case such a converse holds true.

Theorem 6.1.1 *If* ξ_n *is monotonically increasing or decreasing then convergence* $\xi_n \xrightarrow{p} \xi$ *implies that* $\xi_n \xrightarrow{a.s.} \xi$.

Proof Assume, without loss of generality, that $\xi \equiv 0$, $\xi_n \ge 0$, $\xi_n \downarrow$ and $\xi_n \xrightarrow{p} \xi$. If convergence $\xi_n \xrightarrow{a.s.} \xi$ did not hold, there would exist an $\varepsilon > 0$ and a set A with $\mathbf{P}(A) > \delta > 0$ such that $\sup_{k \ge n} \xi_k > \varepsilon$ for $\omega \in A$ and all n. But $\sup_{k \ge n} \xi_k = \xi_n$ and hence we have

$$\mathbf{P}(\xi_n > \varepsilon) \ge \mathbf{P}(A) > \delta > 0$$

for all n, which contradicts the assumed convergence $\xi_n \xrightarrow{p} 0$. □

Thus convergence in probability is determined by the behaviour of the numerical sequence $\mathbf{P}(|\xi_n - \xi| > \varepsilon)$. Is it possible to characterise convergence with probability 1 in a similar way? Set $\zeta_n := \sup_{k \ge n} |\xi_n - \xi|$.

Corollary 6.1.1 *$\xi_n \xrightarrow{a.s.} \xi$ if and only if $\zeta_n \xrightarrow{p} 0$, or, which is the same, when, for any $\varepsilon > 0$,*

$$\mathbf{P}\Big(\sup_{k\ge n} |\xi_k - \xi| > \varepsilon\Big) \to 0 \quad \textit{as } n \to \infty. \tag{6.1.1}$$

Proof Clearly $\xi_n \to \xi$ a.s. if and only if $\zeta_n \to 0$ a.s. But the sequence ζ_n decreases monotonically and it remains to make use of Theorem 6.1.1, which implies that $\zeta_n \xrightarrow{p} 0$ if and only if $\zeta_n \xrightarrow{a.s.} 0$. The corollary is proved. □

In the above argument, the random variables ξ_n and ξ could be improper, where the random variables ξ_n and ξ are only defined on a set B and $\mathbf{P}(B) \in (0, 1)$. (These random variables can take infinite values on $\Omega \setminus B$.) In this case, all the considerations concerning convergence are carried out on the set $B \subset \Omega$ only.

In the introduced terminology, the assertion of the strong law of large numbers for the Bernoulli scheme (Theorem 5.1.2) can be stated, by virtue of (6.1.1), as convergence $S_n/n \to p$ with probability 1.

We have already noted that convergence almost surely implies convergence in probability. Now we will give an example showing that the converse assertion is, generally speaking, not true. Let $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ be the unit circle with the σ-algebra of Borel sets and uniform distribution. Put $\xi(\omega) \equiv 1$, $\xi_n(\omega) = 2$ on the arc $[r(n), r(n) + 1/n]$ and $\xi_n(\omega) = 1$ outside the arc. Here $r(n) = \sum_{k=1}^n \frac{1}{k}$. It is obvious that $\xi_n \xrightarrow{p} \xi$. At the same time, $r(n) \to \infty$ as $n \to \infty$, and the set on which ξ_n converges to ξ is empty (we can find no ω for which $\xi_n(\omega) \to \xi(\omega)$).

However, if $\mathbf{P}(|\xi_n - \xi| > \varepsilon)$ decreases as $n \to \infty$ sufficiently fast, then convergence in probability will also become a.s. convergence. In particular, relation (6.1.1) gives the following sufficient condition for convergence with probability 1.

Theorem 6.1.2 *If the series $\sum_{k=1}^\infty \mathbf{P}(|\xi_n - \xi| > \varepsilon)$ converges for any $\varepsilon > 0$, then $\xi_n \to \xi$ a.s.*

Proof This assertion is obvious, for

$$\mathbf{P}\Big(\bigcup_{k\ge n}\big\{|\xi_k - \xi| > \varepsilon\big\}\Big) \le \sum_{k=n}^\infty \mathbf{P}\big(|\xi_n - \xi| > \varepsilon\big).$$

□

It is this criterion that has actually been used in proving the strong law of large numbers for the Bernoulli scheme.

One cannot deduce a converse assertion about the convergence rate to zero of the probability $\mathbf{P}(|\xi_n - \xi| > \varepsilon)$ from the a.s. convergence. The reader can easily construct an example where $\xi_n \to \xi$ a.s., while $\mathbf{P}(|\xi_n - \xi| > \varepsilon)$ converges to zero arbitrarily slowly.

Theorem 6.1.2 implies the following result.

Corollary 6.1.2 *If $\xi_n \xrightarrow{p} \xi$, then there exists a subsequence $\{n_k\}$ such that $\xi_{n_k} \to \xi$ a.s. as $k \to \infty$.*

Proof This assertion is also obvious since it suffices to take n_k such that $\mathbf{P}(|\xi_{n_k} - \xi| > \varepsilon) \le 1/k^2$ and then make use of Theorem 6.1.2. □

There is one more important special case where convergence in probability $\xi_n \xrightarrow{p} \xi$ implies convergence $\xi_n \to \xi$ a.s. This is the case when the ξ_n are sums of independent random variables. Namely, the following assertion is true. *If* $\xi_n = \sum_{k=1}^n \eta_k$, η_k *are independent, then convergence of* ξ_n *in probability implies convergence with probability* 1. This assertion will be proved in Sect. 11.2.

Finally we consider a third type of convergence of random variables.

Definition 6.1.3 We will say that ξ_n *converges to* ξ *in the* r*-th order mean* (*in mean* if $r = 1$; *in mean square* if $r = 2$) if, as $n \to \infty$,

$$\mathbf{E}|\xi_n - \xi|^r \to 0.$$

This convergence will be denoted by $\xi_n \xrightarrow{(r)} \xi$.

Clearly, by Chebyshev's inequality $\xi_n \xrightarrow{(r)} \xi$ implies that $\xi_n \xrightarrow{p} \xi$. On the other hand, convergence $\xrightarrow{(r)}$ does not follow from a.s. convergence (and all the more from convergence in probability). Thus convergence in probability is the weakest of the three types of convergence we have introduced.

Note that, under additional conditions, convergence $\xi_n \xrightarrow{p} \xi$ can imply that $\xi_n \xrightarrow{(r)} \xi$ (see Theorem 6.1.7 below). For example, it will be shown in Corollary 6.1.4 that if $\xi_n \xrightarrow{p} \xi$ and $\mathbf{E}|\xi_n|^{r+\alpha} < c$ for some $\alpha > 0$, $c < \infty$ and all n, then $\xi_n \xrightarrow{(r)} \xi$.

Definition 6.1.4 A sequence ξ_n is said to be a *Cauchy sequence in probability* (*a.s., in mean*) if, for any $\varepsilon > 0$,

$$\mathbf{P}\big(|\xi_n - \xi_m| > \varepsilon\big) \to 0$$
$$\Big(\mathbf{P}\Big(\sup_{n \ge m} |\xi_n - \xi_m| > \varepsilon\Big) \to 0,\ \mathbf{E}|\xi_n - \xi_m|^r \to 0\Big)$$

as $n \to \infty$ and $m \to \infty$.

Theorem 6.1.3 (Cauchy convergence test) $\xi_n \to \xi$ *in one of the senses* $\xrightarrow{p}$, $\xrightarrow{a.s.}$ *or* $\xrightarrow{(r)}$ *if and only if* ξ_n *is a Cauchy sequence in the respective sense.*

Proof That ξ_n is a Cauchy sequence follows from convergence by virtue of the inequalities

$$|\xi_n - \xi_m| \le |\xi_n - \xi| + |\xi_m - \xi|,$$
$$\sup_{n \ge m} |\xi_n - \xi_m| \le \sup_{n \ge m} |\xi_n - \xi| + |\xi_m - \xi| \le 2 \sup_{n \ge m} |\xi_n - \xi|,$$
$$|\xi_n - \xi_m|^r \le C_r\big(|\xi_n - \xi|^r + |\xi_m - \xi|^r\big)$$

for some C_r.

Now assume that ξ_n is a Cauchy sequence in probability. Choose a sequence $\{n_k\}$ such that

$$\mathbf{P}\big(|\xi_n - \xi_m| > 2^{-k}\big) < 2^{-k}$$

for $n \geq n_k$, $m \geq n_k$. Put

$$\xi'_k := \xi_{n_k}, \qquad A_k := \big\{|\xi'_k - \xi'_{k+1}| > 2^{-k}\big\}, \qquad \eta = \sum_{k=1}^{\infty} I(A_k).$$

Then $\mathbf{P}(A_k) \leq 2^{-k}$ and $\mathbf{E}\eta = \sum_{k=1}^{\infty} \mathbf{P}(A_k) \leq 1$. This means, of course, that the number of occurrences of the events A_k is a proper random variable: $\mathbf{P}(\eta < \infty) = 1$, and hence with probability 1 finitely many events A_k occur. This means that, for any ω for which $\eta(\omega) < \infty$, there exists a $k_0(\omega)$ such that $|\xi'_k(\omega) - \xi'_{k+1}(\omega)| \leq 2^{-k}$ for all $k \geq k_0(\omega)$. Therefore one has the inequality $|\xi'_k(\omega) - \xi'_l(\omega)| \leq 2^{-k+1}$ for all $k \geq k_0(\omega)$ and $l \geq k_0(\omega)$, which means that $\xi'_n(\omega)$ is a numerical Cauchy sequence and hence there exists a value $\xi(\omega)$ such that $|\xi'_k(\omega) - \xi(\omega)| \to 0$ as $k \to \infty$. This means, in turn, that $\xi'_k \xrightarrow{a.s.} \xi$ and hence

$$\mathbf{P}\big(|\xi_n - \xi| \geq \varepsilon\big) \leq \mathbf{P}\bigg(|\xi_n - \xi_{n_k}| \geq \frac{\varepsilon}{2}\bigg) + \mathbf{P}\bigg(|\xi_{n_k} - \xi| > \frac{\varepsilon}{2}\bigg) \to 0$$

as $n \to \infty$ and $k \to \infty$.

Now assume that ξ_n is a Cauchy sequence in mean. Then, by Chebyshev's inequality, it will be a Cauchy sequence in probability and hence, by Corollary 6.1.2, there will exist a random variable ξ and a subsequence $\{n_k\}$ such that $\xi_{n_k} \xrightarrow{a.s.} \xi$. Now we will show that $\mathbf{E}|\xi_n - \xi|^r \to 0$. For a given $\varepsilon > 0$, choose an n such that $\mathbf{E}|\xi_k - \xi_l|^r < \varepsilon$ for $k \geq n$ and $l \geq n$. Then, by Fatou's lemma (see Appendix 3),

$$\begin{aligned}\mathbf{E}|\xi_n - \xi|^r &= \mathbf{E}\lim_{n_k\to\infty} |\xi_n - \xi_{n_k}|^r \\ &= \mathbf{E}\liminf_{n_k\to\infty} |\xi_n - \xi_{n_k}|^r \leq \liminf_{n_k\to\infty} \mathbf{E}|\xi_n - \xi_{n_k}|^r \leq \varepsilon.\end{aligned}$$

This means that $\mathbf{E}|\xi_n - \xi|^r \to 0$ as $n \to \infty$.

It remains to verify the assertion of the theorem related to a.s. convergence. We already know that if ξ_n is a Cauchy sequence in probability (or a.s.) then there exist a ξ and a subsequence ξ_{n_k} such that $\xi_{n_k} \xrightarrow{a.s.} \xi$. Therefore, if we put $n_{k(n)} := \min\{n_k : n_k \geq n\}$, then

$$\mathbf{P}\Big(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\Big) \leq \mathbf{P}\Big(\sup_{k\geq n} |\xi_k - \xi_{n_{k(n)}}| \geq \varepsilon/2\Big) + \mathbf{P}\big(|\xi_{n_{k(n)}} - \xi| \geq \varepsilon/2\big) \to 0$$

as $n \to \infty$. The theorem is proved. □

Remark 6.1.1 If we introduce the space L_r of all random variables ξ on $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ for which $\mathbf{E}|\xi|^r < \infty$ and the norm $\|\xi\| = (\mathbf{E}|\xi|^r)^{1/r}$ on it (the triangle inequality $\|\xi_1 + \xi_2\| \leq \|\xi_1\| + \|\xi_2\|$ is then nothing else but Minkowski's inequality, see Theorem 4.7.2), then the assertion of Theorem 6.1.3 on convergence $\xrightarrow{(r)}$ (which is convergence in the norm of L_r, for we identify random variables ξ_1 and ξ_2 if $\|\xi_1 - \xi_2\| = 0$) means that L_r is complete and hence is a Banach space.

The space of *all* random variables on $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ can be metrised so that convergence in the metric will be equivalent to convergence in probability. For instance, one could put

$$\rho(\xi_1, \xi_2) := \mathbf{E}\frac{|\xi_1 - \xi_2|}{1 + |\xi_1 - \xi_2|}.$$

Since

$$\frac{|x+y|}{1+|x+y|} \le \frac{|x|}{1+|x|} + \frac{|y|}{1+|y|}$$

always holds, $\rho(\xi_1, \xi_2)$ satisfies all the axioms of a metric. It is not difficult to see that relations $\rho(\xi_1, \xi_2) \to 0$ and $\xi_n \xrightarrow{p} 0$ are equivalent. The assertion of Theorem 6.1.3 related to convergence $\xrightarrow{p}$ means that the metric space we introduced is complete.

6.1.2 The Continuity Theorem

Now we will derive the following "continuity theorem".

Theorem 6.1.4 *Let $\xi_n \xrightarrow{a.s.} \xi$ ($\xi_n \xrightarrow{p} \xi$) and $H(s)$ be a function continuous everywhere with respect to the distribution of the random variable ξ (i.e. $H(s)$ is continuous at each point of a set S such that $\mathbf{P}(\xi \in S) = 1$). Then*

$$H(\xi_n) \xrightarrow{a.s.} H(\xi) \qquad \big(H(\xi_n) \xrightarrow{p} H(\xi)\big).$$

Proof Let $\xi_n \xrightarrow{a.s.} \xi$. Since the sets $A = \{\omega : \xi_n(\omega) \to \xi(\omega)\}$ and $B = \{\omega : \xi(\omega) \in S\}$ are both of probability 1, $\mathbf{P}(AB) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B) = 1$. But one has $H(\xi_n) \to H(\xi)$ on the set AB. Convergence with probability 1 is proved.

Now let $\xi_n \xrightarrow{p} \xi$. If we assume that convergence $H(\xi_n) \xrightarrow{p} H(\xi)$ does not take place then there will exist $\varepsilon > 0$, $\delta > 0$ and a subsequence $\{n'\}$ such that

$$\mathbf{P}\big(\big|H(\xi_{n'}) - H(\xi)\big| > \varepsilon\big) > \delta.$$

But $\xi_{n'} \xrightarrow{p} \xi$ and hence there exists a subsequence $\{n''\}$ such that $\xi_{n''} \xrightarrow{a.s.} \xi$ and $H(\xi_{n''}) \xrightarrow{a.s.} H(\xi)$. This contradicts the assumption we made, for the latter implies that

$$\mathbf{P}\big(\big|H(\xi_{n''}) - H(\xi)\big| > \varepsilon\big) > \delta.$$

The theorem is proved. □

6.1.3 Uniform Integrability and Its Consequences

Now we will consider this question: in what cases does convergence in probability imply convergence in mean?

The main condition that ensures the transition from convergence in probability to convergence in mean is associated with the notion of uniform integrability.

Definition 6.1.5 A sequence $\{\xi_n\}$ is said to be *uniformly integrable* if

$$\sup_n \mathbf{E}\big(|\xi_n|;\ |\xi_n| > N\big) \to 0 \quad \text{as } N \to \infty.$$

A sequence of independent identically distributed random variables with finite mean is, clearly, uniformly integrable.

If $\{\xi_n\}$ is uniformly integrable then so are $\{c\xi_n\}$ and $\{\xi_n + c\}$, where $c = \text{const}$.

Let us present some further, less evident, properties of uniform integrability.

U1. *If the sequences* $\{\xi'_n\}$ *and* $\{\xi''_n\}$ *are uniformly integrable then the sequences defined by* $\zeta_n = \max(|\xi'_n|, |\xi''_n|)$ *and* $\zeta_n = \xi'_n + \xi''_n$ *are also uniformly integrable.*

Proof Indeed, for $\zeta_n = \max(|\xi'_n|, |\xi''_n|)$ we have

$$\begin{aligned}\mathbf{E}(\zeta_n;\ \zeta_n > N) &= \mathbf{E}\big(\zeta_n;\ \zeta_n > N, \big|\xi'_n\big| > \big|\xi''_n\big|\big) + \mathbf{E}\big(\zeta_n;\ \zeta_n > N, \big|\xi'_n\big| \le \big|\xi''_n\big|\big)\\ &\le \mathbf{E}\big(\big|\xi'_n\big|;\ \big|\xi'_n\big| > N\big) + \mathbf{E}\big(\big|\xi''_n\big|;\ \big|\xi''_n\big| \ge N\big) \to 0\end{aligned}$$

as $N \to \infty$.

Since

$$\big|\xi'_n + \xi''_n\big| \le \big|\xi'_n\big| + \big|\xi''_n\big| \le 2\max\big(\big|\xi'_n\big|, \big|\xi''_n\big|\big),$$

from the above it follows that the sequence defined by the sum $\zeta_n = \xi'_n + \xi''_n$ is also uniformly integrable. □

U2. *If* $\{\xi_n\}$ *is uniformly integrable then* $\sup_n \mathbf{E}|\xi_n| \le c < \infty$.

Proof Indeed, choose N so that

$$\sup_n \mathbf{E}\big(|\xi_n|;\ |\xi_n| > N\big) \le 1.$$

Then

$$\sup_n \mathbf{E}|\xi_n| = \sup_n \big[\mathbf{E}\big(|\xi_n|;\ |\xi_n| \le N\big) + \mathbf{E}\big(|\xi_n|;\ |\xi_n| > N\big)\big] \le N + 1. \qquad \square$$

The converse assertion is not true. For example, for a sequence

$$\xi_n : \mathbf{P}(\xi_n = n) = 1/n = 1 - \mathbf{P}(\xi_n = 0)$$

one has $\mathbf{E}|\xi_n| = 1$, but the sequence is not uniformly integrable.

If we somewhat strengthen the above statement U2, it becomes "characteristic" for uniform integrability.

Theorem 6.1.5 *For a sequence* $\{\xi_n\}$ *to be uniformly integrable, it is necessary and sufficient that there exists a function* $\psi(x)$ *such that*

$$\frac{\psi(x)}{x} \uparrow \infty \quad \textit{as } x \uparrow \infty, \qquad \sup_n \mathbf{E}\psi\big(|\xi_n|\big) < c < \infty. \tag{6.1.2}$$

In the necessity assertion one can choose a convex function ψ.

Proof Without loss of generality we can assume that $\xi_i \ge 0$.

The *sufficiency* is evident, since, putting $v(x) := \frac{\psi(x)}{x}$, we get

$$\mathbf{E}(\xi_n;\ \xi_n \ge N) \le \frac{1}{v(N)}\mathbf{E}\big(\xi_n v(\xi_n);\ \xi_n \ge N\big) \le \frac{c}{v(N)}.$$

To prove the *necessity*, put

$$\varepsilon(N) := \sup_n \mathbf{E}(\xi_n;\ \xi_n \ge N).$$

Then, by virtue of uniform integrability, $\varepsilon(N) \downarrow 0$ as $N \uparrow \infty$. Choose a sequence $N_k \uparrow \infty$ as $k \uparrow \infty$ such that

$$\sum_{k=1}^{\infty} \sqrt{\varepsilon(N_k)} < c_1 < \infty,$$

and put

$$g(x) = x\big(\varepsilon(N_k)\big)^{-1/2} \quad \text{for } x \in [N_k, N_{k+1}).$$

Since

$$\frac{g(N_k - 0)}{N_k} = \big(\varepsilon(N_{k-1})\big)^{-1/2} \le \big(\varepsilon(N_k)\big)^{-1/2} = \frac{g(N_k)}{N_k},$$

we have $\frac{g(x)}{x} \uparrow \infty$ as $x \to \infty$. Further,

$$\begin{aligned}\mathbf{E}g(\xi_n) &= \sum_k \mathbf{E}\big[g(\xi_n);\ \xi_n \in [N_k, N_{k+1})\big] \\ &= \sum_k \mathbf{E}\big[\xi_n\big(\varepsilon(N_k)\big)^{-1/2};\ \xi_n \in [N_k, N_{k+1})\big] \\ &\le \sum_k \big(\varepsilon(N_k)\big)^{-1/2}\varepsilon(N_k) = \sum_k \sqrt{\varepsilon(N_k)} < c_1,\end{aligned}$$

where the right-hand side does not depend on n. Therefore, to prove the theorem it is sufficient to construct a function $\psi \le g$ which is convex and such that $\frac{\psi(x)}{x} \uparrow \infty$ as $x \uparrow \infty$.

Define the function $\psi(x)$ as the continuous polygon with nodes $(N_k, g(N_k - 0))$. Since

$$\frac{g(N_k - 0)}{N_k} = \varepsilon(N_{k-1})^{-1/2}$$

monotonically increases as k grows, ψ is a lower envelope curve for the discontinuous function $g(x) \ge \psi(x)$. The monotonicity of $\frac{\psi(x)}{x}$ follows from the fact that, on the interval $[N_k, N_{k+1})$, this function can be represented as

$$\frac{\psi(x)}{x} = a_{k,\psi} - \frac{b_k}{x},$$

where $b_k > 0$, because the values $\psi(N_{k+1}-0)$ and $g(N_{k+1}-0)$ coincide, while the angular incline $a_{k,\psi}$ of the function ψ on the interval $[N_k, N_{k+1})$ is greater than the "radial" incline $a_{k,g}$ of the function g:

$$a_{k,g} = \frac{g(N_{k+1}-0) - g(N_k)}{N_{k+1} - N_k} < \frac{g(N_{k+1}-0) - g(N_k-0)}{N_{k+1} - N_k} = a_{k,\psi}.$$

It is clear that $\frac{\psi(x)}{x}$ increases unboundedly, for

$$\frac{\psi(N_k)}{N_k} = \frac{g(N_k-0)}{N_k} = \varepsilon(N_{k-1})^{-1/2} \uparrow \infty$$

as $k \to \infty$. The theorem is proved. □

In studying the mean values of sums of random variables, the following theorem on *uniform integrability of average values*, following from Theorem 6.1.5, plays an important role.

Theorem 6.1.6 *Let $\xi_1, \xi_2, \dots$ be an arbitrary uniformly integrable sequence of random variables,*

$$p_{i,n} \ge 0, \qquad \sum_{i=1}^{n} p_{i,n} = 1, \qquad \zeta_n = \sum_{k=1}^{n} |\xi_i| p_{i,n}.$$

Then the sequence $\{\zeta_n\}$ is uniformly integrable as well.

Proof Let $\psi(x)$ be the convex function from Theorem 6.1.5 satisfying properties (6.1.2). Then, by that theorem,

$$\mathbf{E}\psi(\zeta_n) = \mathbf{E}\,\psi\left(\sum_{i=1}^{n} p_{i,n}|\xi_i|\right) \le \mathbf{E}\sum_{i=1}^{n} p_{i,n}\psi\big(|\xi_i|\big) \le c.$$

It remains to make use of Theorem 6.1.5 again. □

Now we will show that convergence in probability together with uniform integrability imply convergence in mean.

Theorem 6.1.7 *Let $\xi_n \xrightarrow{p} \xi$ and $\{\xi_n\}$ be uniformly integrable. Then $\mathbf{E}|\xi|$ exists and, as $n \to \infty$,*

$$\mathbf{E}|\xi_n - \xi| \to 0.$$

If, moreover, $\{|\xi_n^r|\}$ is uniformly integrable then $\xi_n \xrightarrow{(r)} \xi$.

Conversely, if, for an $r \ge 1$, $\xi_n \xrightarrow{(r)} \xi$ and $\mathbf{E}|\xi|^r < \infty$, then $\{|\xi_n|^r\}$ is uniformly integrable.

In the law of large numbers for the Bernoulli scheme (see Theorem 5.1.1) we proved that the normed sum S_n/n converges to p in probability. Since $0 \le S_n/n \le 1$,

S_n/n is clearly uniformly integrable and the convergence in mean $\mathbf{E}|S_n/n - p|^r \to 0$ holds for any r. This fact can also be established directly. For a more substantiative example of application of Theorems 6.1.6 and 6.1.7, see Sect. 8.1.

Proof We show that $\mathbf{E}\xi$ exists. By the properties of integrals (see Lemma A3.2.3 in Appendix 3), if $\mathbf{E}|\zeta| < \infty$ then $\mathbf{E}(\zeta; A_n) \to 0$ as $\mathbf{P}(A_n) \to 0$. Since $\mathbf{E}\xi_n < \infty$, for any N and ε one has

$$\begin{aligned}\mathbf{E}\min\bigl(|\xi|, N\bigr) &= \lim_{n\to\infty}\bigl[\mathbf{E}\min\bigl(|\xi|, N\bigr);\ |\xi_n - \xi| < \varepsilon\bigr] \\ &\le \lim_{n\to\infty}\mathbf{E}\min\bigl(|\xi| + \varepsilon, N\bigr) \le c + \varepsilon.\end{aligned}$$

It follows that $\mathbf{E}|\xi| \le c$.

Further, for brevity, put $\eta_n = |\xi_n - \xi|$. Then $\eta_n \xrightarrow{p} 0$ and η_n are uniformly integrable together with ξ_n. For any N and ε, one has

$$\begin{aligned}\mathbf{E}\eta_n &= \mathbf{E}(\eta_n;\ \eta_n \le \varepsilon) + \mathbf{E}(\eta_n;\ N \ge \eta_n > \varepsilon) + \mathbf{E}(\eta_n;\ \eta_n \ge N) \\ &\le \varepsilon + N\mathbf{P}(\eta_n \ge \varepsilon) + \mathbf{E}(\eta_n;\ \eta_n > N).\end{aligned} \tag{6.1.3}$$

Choose N so that $\sup_n \mathbf{E}(\eta_n;\ \eta_n > N) \le \varepsilon$. Then, for such an N,

$$\limsup_{n\to\infty}\mathbf{E}\eta_n \le 2\varepsilon.$$

Since ε is arbitrary, $\mathbf{E}\eta_n \to 0$ as $n \to \infty$.

The relation $\mathbf{E}|\xi_n - \xi|^r \to 0$ can be proved in the same way as (6.1.3), since $\eta_n^r = |\xi_n - \xi|^r \xrightarrow{p} 0$ and η_n^r are uniformly integrable together with $|\xi_n|^r$.

Now we will prove the converse assertion. Let, for simplicity, $r = 1$. One has

$$\begin{aligned}\mathbf{E}\bigl(|\xi_n|;\ |\xi_n| > N\bigr) &\le \mathbf{E}\bigl(|\xi_n - \xi|;\ |\xi_n| > N\bigr) + \mathbf{E}\bigl(|\xi|;\ |\xi_n| > N\bigr) \\ &\le \mathbf{E}|\xi_n - \xi| + \mathbf{E}\bigl(|\xi|;\ |\xi_n| > N\bigr) \\ &\le \mathbf{E}|\xi_n - \xi| + \mathbf{E}\bigl(|\xi|;\ |\xi_n - \xi| > 1\bigr) + \mathbf{E}\bigl(|\xi|;\ |\xi| > N - 1\bigr).\end{aligned}$$

The first term on the right-hand side tends to zero by the assumption, and the second term, by Lemma A3.2.3 from Appendix 3, which we have just mentioned, and the fact that $\mathbf{P}(|\xi_n - \xi| > 1) \to 0$. The last term does not depend on n and can be made arbitrarily small by choosing N. Theorem 6.1.7 is proved. □

Now we can derive yet another *continuity theorem* which has the following form.

Theorem 6.1.8 *If* $\xi_n \xrightarrow{p} \xi$, $H(s)$ *satisfies the conditions of Theorem* 6.1.4, *and* $H(\xi_n)$ *is uniformly integrable, then, as* $n \to \infty$,

$$\mathbf{E}\bigl|H(\xi_n) - H(\xi)\bigr| \to 0$$

and, in particular, $\mathbf{E}H(\xi_n) \to \mathbf{E}H(\xi)$.

This assertion follows from Theorems 6.1.4 and 6.1.7, for $H(\xi_n) \xrightarrow{p} H(\xi)$ by Theorem 6.1.4.

Sometimes it is convenient to distinguish between *left* and *right* uniform integrability. We will say that a sequence $\{\xi_n\}$ is *right* (*left*) *uniformly integrable* if

$$\sup \mathbf{E}(\xi_n;\ \xi_n \ge N) \to 0 \qquad \bigl(\sup \mathbf{E}\bigl(|\xi_n|;\ \xi_n \le -N\bigr) \to 0\bigr)$$

as $N \to \infty$. It is evident that a sequence $\{\xi_n\}$ is uniformly integrable if and only if it is both right and left uniformly integrable.

Lemma 6.1.1 *A sequence $\{\xi_n\}$ is right uniform integrable if at least one of the following conditions is met*:

1. *For any sequence $N(n) \to \infty$ as $n \to \infty$, one has*

$$\mathbf{E}\bigl(\xi_n;\ \xi_n > N(n)\bigr) \to 0.$$

   (*This condition is clearly also necessary for uniform integrability.*)
2. $\xi_n \le \eta$, *where* $\mathbf{E}\eta < \infty$.
3. $\mathbf{E}(\xi_n^+)^{1+\alpha} < c < \infty$ *for some* $\alpha > 0$ (*here* $x^+ = \max(0, x)$).
4. ξ_n *is left uniformly integrable*, $\xi_n \xrightarrow{p} \xi$, *and* $\mathbf{E}\xi_n \to \mathbf{E}\xi < \infty$.

Proof

1. If the sequence $\{\xi_n\}$ were not right uniformly integrable, there would exist an $\varepsilon > 0$ and subsequences $n' \to \infty$ and $N' = N'(n') \to \infty$ such that $\mathbf{E}(\xi_{n'};\ \xi_{n'} > N') > \varepsilon$. But this contradicts condition 1.
2. $\mathbf{E}(\xi_n;\ \xi_n > N) \le \mathbf{E}(\eta;\ \eta > N) \to 0$ as $N \to \infty$.
3. $\mathbf{E}(\xi_n;\ \xi_n > N) \le \mathbf{E}(\xi_n^{1+\alpha} N^{-\alpha};\ \xi_n > N) \le N^{-\alpha} c \to 0$ as $N \to \infty$.
4. Without loss of generality, put $\xi := 0$. Then

$$\mathbf{E}(\xi_n;\ \xi_n > N) = \mathbf{E}\xi_n - \mathbf{E}(\xi_n;\ \xi_n < -N) - \mathbf{E}\bigl(\xi_n;\ |\xi_n| \le N\bigr).$$

The first two terms on the right-hand side vanish as $n \to \infty$ for any $N = N(n) \to \infty$. For the last term, for any $\varepsilon > 0$, one has

$$\begin{aligned}\bigl|\mathbf{E}\bigl(\xi_n;\ |\xi_n| \le N\bigr)\bigr| &\le \bigl|\mathbf{E}\bigl(\xi_n;\ |\xi_n| \le \varepsilon\bigr)\bigr| + \bigl|\mathbf{E}\bigl(\xi_n;\ \varepsilon < |\xi_n| \le N\bigr)\bigr| \\ &\le \varepsilon + N\mathbf{P}\bigl(|\xi_n| > \varepsilon\bigr).\end{aligned}$$

For any given $\varepsilon > 0$, choose an $n(\varepsilon)$ such that, for all $n \ge n(\varepsilon)$, we would have $\mathbf{P}(|\xi_n| > \varepsilon) < \varepsilon$, and put $N(\varepsilon) := \lfloor 1/\sqrt{\varepsilon} \rfloor$. This will mean that, for all $n \ge n(\varepsilon)$ and $N \le N(\varepsilon)$, one has $\mathbf{E}(\xi_n;\ |\xi_n| \le N) < \varepsilon + \sqrt{\varepsilon}$, and therefore condition 1 of the lemma holds for $\mathbf{E}(\xi_n;\ \xi_n > N)$. The lemma is proved. □

Now, based on the above, we can state three useful corollaries.

Corollary 6.1.3 (The dominated convergence theorem) *If $\xi_n \xrightarrow{p} \xi$, $|\xi_n| < \eta$, and $\mathbf{E}\eta < \infty$ then $\mathbf{E}\xi$ exists and $\mathbf{E}\xi_n \to \mathbf{E}\xi$.*

Corollary 6.1.4 *If* $\xi_n \xrightarrow{p} \xi$ *and* $\mathbf{E}|\xi_n|^{r+\alpha} < c < \infty$ *for some* $\alpha > 0$ *then* $\xi_n \xrightarrow{(r)} \xi$.

Corollary 6.1.5 *If* $\xi_n \xrightarrow{p} \xi$ *and* $H(x)$ *is a continuous bounded function, then* $\mathbf{E}|H(\xi_n) - \mathbf{E}H(\xi)| \to 0$ *as* $n \to \infty$.

In conclusion of the present section, we will derive one more auxiliary proposition that can be useful.

Lemma 6.1.2 (On integrals over sets of small probability) *If* $\{\xi_n\}$ *is a uniformly integrable sequence and* $\{A_n\}$ *is an arbitrary sequence of events such that* $\mathbf{P}(A_n) \to 0$, *then* $\mathbf{E}(|\xi_n|;\ A_n) \to 0$ *as* $n \to \infty$.

Proof Put $B_n := \{|\xi_n| \le N\}$. Then

$$\begin{aligned}\mathbf{E}\big(|\xi_n|;\ A_n\big) &= \mathbf{E}\big(|\xi_n|;\ A_n B_n\big) + \mathbf{E}\big(|\xi_n|;\ A_n \overline{B}_n\big)\\ &\le N\mathbf{P}(A_n) + \mathbf{E}\big(|\xi_n|;\ |\xi_n| > N\big).\end{aligned}$$

For a given $\varepsilon > 0$, first choose N so that the second summand on the right-hand side does not exceed $\varepsilon/2$ and then an n such that the first summand does not exceed $\varepsilon/2$. We obtain that, by choosing n large enough, we can make $\mathbf{E}(|\xi_n|;\ A_n)$ less than ε. The lemma is proved. □

6.2 Convergence of Distributions

In Sect. 6.1 we introduced three types of convergence which can be used to characterise the closeness of random variables given on a common probability space. But what can one do if random variables are given on different probability spaces (or if it is not known where they are given) which nevertheless have similar distributions? (Recall, for instance, the Poisson or de Moivre–Laplace theorems.) In such cases one should be able to characterise the closeness of the distributions themselves. Having found an apt definition for such a closeness, in many problems we will be able to approximate the required but hard to come by distributions by known and, as a rule, simpler distributions.

Now what distributions should be considered as close? We are clearly looking for a definition of convergence of a sequence of distribution functions $F_n(x)$ to a distribution function $F(x)$. It would be natural, for instance, that the distributions of the variables $\xi_n = \xi + 1/n$ should converge to that of ξ as $n \to \infty$. Therefore requiring in the definition of convergence that $\sup_x |F_n(x) - F(x)|$ is small would be unreasonable since this condition is not satisfied for the distributions of $\xi + 1/n$ and ξ if $F(x) = \mathbf{P}(\xi < x)$ has at least one point of discontinuity.

We will define the convergence of F_n to F as that which arises when one considers convergence in probability.

Definition 6.2.1 We will say that distribution functions F_n *converge weakly* to a distribution function F as $n \to \infty$, and denote this by $F_n \Rightarrow F$ if, for any continuous bounded function $f(x)$,

$$\int f(x)\,dF_n(x) \to \int f(x)\,dF(x). \tag{6.2.1}$$

Considering the distributions $\mathbf{F}_n(B)$ and $\mathbf{F}(B)$ (B are Borel sets) corresponding to F_n and F, we say that $\mathbf{F}_n$ converges weakly to $\mathbf{F}$ and write $\mathbf{F}_n \Rightarrow \mathbf{F}$. One can clearly re-write (6.2.1) as

$$\int f(x)\,\mathbf{F}_n(dx) \to \int f(x)\,\mathbf{F}(dx) \quad \text{or} \quad \mathbf{E} f(\xi_n) \to \mathbf{E} f(\xi) \tag{6.2.2}$$

(cf. Corollary 6.1.5), where $\xi_n \Subset \mathbf{F}_n$ and $\xi \Subset \mathbf{F}$.

Another possible definition of weak convergence follows from the next assertion.

Theorem 6.2.1 [2] $\mathbf{F}_n \Rightarrow \mathbf{F}$ *if and only if* $F_n(x) \to F(x)$ *at each point of continuity* x *of* F.

Proof Let (6.2.1) hold. Consider an $\varepsilon > 0$ and a continuous function $f_\varepsilon(t)$ which is equal to 1 for $t < x$ and to 0 for $t \ge x + \varepsilon$, and varies linearly on $[x, x+\varepsilon]$. Since

$$F_n(x) = \int_{-\infty}^{x} f_\varepsilon(t)\,dF_n(t) \le \int f_\varepsilon(t)\,dF_n(t),$$

by virtue of (6.2.1) one has

$$\limsup_{n\to\infty} F_n(x) \le \int f_\varepsilon(t)\,dF(t) \le F(x+\varepsilon).$$

If x is a point of continuity of F then

$$\limsup_{n\to\infty} F_n(x) \le F(x)$$

since ε is arbitrary.

In the same way, using the function $f_\varepsilon^*(t) = f_\varepsilon(t+\varepsilon)$, we obtain the inequality

$$\liminf_{n\to\infty} F_n(x) \ge F(x).$$

We now prove the converse assertion. Let $-M$ and N be points of continuity of F such that $F(-M) < \varepsilon/5$ and $1 - F(N) < \varepsilon/5$. Then $F_n(-M) < \varepsilon/4$ and $1 - F_n(N) < \varepsilon/4$ for all sufficiently large n. Therefore, assuming for simplicity that $|f| \le 1$, we obtain that

$$\int f\,dF_n \quad \text{and} \quad \int f\,dF \tag{6.2.3}$$

[2] In many texts on probability theory the condition of the theorem is given as the definition of weak convergence. However, the definition in terms of the relation (6.2.2) is apparently more appropriate for it continues to remain valid for distributions on arbitrary topological spaces (see, e.g. [1, 25]).

will differ from

$$\int_{-M}^{N} f\,dF_n \quad \text{and} \quad \int_{-M}^{N} f\,dF,$$

respectively, by less than $\varepsilon/2$. Construct on the semi-interval $(-M, N]$ a step function f_ε with jumps at the points of continuity of F which differs from f by less than $\varepsilon/2$. Outside $(-M, N]$ we set $f_\varepsilon := 0$. We can put, for instance,

$$f_\varepsilon(x) := \sum_{j=1}^{k} f(x_j)\delta_j(x),$$

where $x_0 = -M < x_1 < \cdots < x_k = N$ are appropriately chosen points of continuity of F, and $\delta_j(x)$ is the indicator function of the semi-interval $(x_{j-1}, x_j]$. Then $\int f_\varepsilon\,dF_n$ and $\int f_\varepsilon\,dF$ will differ from the respective integrals in (6.2.3), for sufficiently large n, by less than ε. At the same time,

$$\int f_\varepsilon\,dF_n = \sum_{j=1}^{k} f(x_j)\big[F_n(x_j) - F_n(x_{j-1})\big] \to \int f_\varepsilon\,dF.$$

Since $\varepsilon > 0$ is arbitrary, the last relation implies (6.2.1). (Indeed, one just has to make use of the inequality

$$\limsup \int f\,dF_n \le \varepsilon + \limsup \int f_\varepsilon\,dF_n = \varepsilon + \int f_\varepsilon\,dF \le 2\varepsilon + \int f\,dF$$

and a similar inequality for $\liminf \int f\,dF_n$.) The theorem is proved. □

For remarks on different and, in a certain sense, simpler proofs of the second assertion of Theorem 6.2.1, see the end of Sect. 6.3 and Sect. 7.4.

Remark 6.2.1 Repeating with obvious modifications the above-presented proof, we can get a somewhat different equivalent of convergence (4): *convergence of differences $F_n(y) - F_n(x) \to F(y) - F(x)$ for any points of continuity x and y of F.*

Remark 6.2.2 If $F(x)$ is continuous then convergence $\mathbf{F}_n \Rightarrow \mathbf{F}$ is equivalent to the *uniform* convergence $\sup_x |F_n(x) - F(x)| \to 0$.

We leave the proof of the last assertion to the reader. It follows from the fact that convergence $F_n(x) \to F(x)$ at any x implies, by virtue of the continuity of F, uniform convergence on any finite interval. The uniform smallness of $F_n(x) - F(x)$ on the "tails" is ensured by the smallness of $F(x)$ and $1 - F(x)$.

Remark 6.2.3 If distributions $\mathbf{F}_n$ and $\mathbf{F}$ are discrete and have jumps at the same points $x_1, x_2, \ldots$ then $\mathbf{F}_n \Rightarrow \mathbf{F}$ will clearly be equivalent to the convergence of the probabilities of the values $x_1, x_2, \ldots$ $(F_n(x_k + 0) - F_n(x_k) \to F(x_k + 0) - F(x_k))$.

We introduce some notation which will be convenient for the sequel. Let ξ_n and ξ be some random variables (given, generally speaking, on different probability spaces) such that $\xi_n \in \mathbf{F}_n$ and $\xi \in \mathbf{F}$.

Definition 6.2.2 If $\mathbf{F}_n \Rightarrow \mathbf{F}$ we will say that ξ_n *converges to* ξ *in distribution* and write $\xi_n \Rightarrow \xi$.

We used here the same symbol $\Rightarrow$ as for the weak convergence, but this leads to no confusion.

It is clear that $\xi_n \xrightarrow{p} \xi$ implies $\xi_n \Rightarrow \xi$, but not vice versa.

At the same time the following assertion holds true.

Lemma 6.2.1 *If* $\xi_n \Rightarrow \xi$ $(\mathbf{F}_n \Rightarrow \mathbf{F})$ *then one can construct random variables* ξ_n' *and* ξ' *on a common probability space so that* $\mathbf{P}(\xi_n' < x) = \mathbf{P}(\xi_n < x) = F_n(x)$, $\mathbf{P}(\xi' < x) = \mathbf{P}(\xi < x) = F(x)$, *and*

$$\xi_n' \xrightarrow{a.s.} \xi'.$$

Proof Define the quantile transforms (see Definition 3.2.6) by

$$F_n^{-1}(t) := \sup\{x : F_n(x) \le t\}, \qquad F^{-1}(t) := \sup\{x : F(x) \le t\}.$$

(If $F(x)$ is continuous and strictly increasing then $F^{-1}(t)$ coincides with the solution to the equation $F(v) = t$.) Let $\eta \in \mathbf{U}_{0,1}$. Put

$$\xi_n' := F_n^{-1}(\eta) \in F_n, \quad \xi' := F^{-1}(\eta) \in F$$

(cf. Theorem 3.2.2), and show that $\xi_n' \xrightarrow{a.s.} \xi'$. In order to do that, it suffices to prove that $F_n^{-1}(y) \to F^{-1}(y)$ for almost all $y \in [0, 1]$.

The functions F and F^{-1} are monotone and hence each of them has at most a countable set of discontinuity points. This means that, for all $y \in [0, 1]$ with the possible exclusion of the points from a countable set T, the function $F^{-1}(y)$ will be continuous.

So let y be a point of continuity of $F^{(-1)}$, and $F^{(-1)}(y) = x$.

For $t \le y$, choose a continuous strictly increasing function $G^{(-1)}(t)$ such that

$$G^{(-1)}(y) = F^{(-1)}(y), \qquad G^{(-1)}(t) \le F^{(-1)}(t) \quad \text{for } t \le y.$$

Denote by $G(v)$, $v \le x$, the function inverse to $G^{(-1)}(t)$. Clearly, $G(v)$ dominates the function $F(v)$ in the domain $v \le x$. By virtue of the continuity and strict monotonicity of the functions $G^{(-1)}$ and G (in the domain under consideration), for $\varepsilon > 0$ we have

$$G(x - \varepsilon) = y - \delta(\varepsilon),$$

where $\delta(\varepsilon) > 0$, $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Choose an ε such that $x - \varepsilon$ is a point of continuity of F. Then, for all n large enough,

$$F_n(x - \varepsilon) \le F(x - \varepsilon) + \frac{\delta(\varepsilon)}{2} \le G(x - \varepsilon) + \frac{\delta(\varepsilon)}{2} = y - \frac{\delta(\varepsilon)}{2}.$$

The opposite inequality can be proved in a similar way. Since ε can be arbitrarily small, we obtain that, for almost all y,

$$F_n^{-1}(y) \to F^{(-1)}(y) \quad \text{as } n \to \infty.$$

Hence $F_n^{(-1)}(\eta) \to F^{(-1)}(\eta)$ with probability 1 with respect to the distribution of η. The lemma is proved. □

Lemma 6.2.1 remains true for vector-valued random variables as well.

Sometimes it is also convenient to have a simple symbol for the relation "the distribution of ξ_n converges weakly to **F**". We will write this relation as

$$\xi_n \Subset\!\!\Rightarrow \mathbf{F}, \tag{6.2.4}$$

so that the symbol $\Subset\!\!\Rightarrow$ expresses the same fact as $\Rightarrow$ but relates objects of a different nature in the same way as the symbol $\Subset$ in the relation $\xi \Subset \mathbf{P}$ (on the left-hand side in (6.2.4) we have random variables, while on the right hand side there is a distribution).

In these terms, the assertion of the Poisson theorem could be written as $S_n \Subset\!\!\Rightarrow \mathbf{\Pi}_\mu$, while the statement of the law of large numbers for the Bernoulli scheme takes the form $S_n/n \Subset\!\!\Rightarrow \mathbf{I}_p$.

The coincidence of the distributions of ξ and η will be denoted by $\xi \stackrel{d}{=} \eta$.

Lemma 6.2.2 *If $\xi_n \Rightarrow \xi$ and $\varepsilon_n \xrightarrow{p} 0$ then $\xi_n + \varepsilon_n \Rightarrow \xi$.*
If $\xi_n \Rightarrow \xi$ and $\gamma_n \xrightarrow{p} 1$ then $\xi_n \gamma_n \Rightarrow \xi$.

Proof Let us prove the first assertion. For any t and $\delta > 0$ such that t and $t \pm \delta$ are points of continuity of $\mathbf{P}(\xi < t)$, one has

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}(\xi_n + \varepsilon_n < t) &= \limsup_{n\to\infty} \mathbf{P}(\xi_n + \varepsilon_n < t,\ \varepsilon_n > -\delta)\\ &\le \limsup_{n\to\infty} \mathbf{P}(\xi_n < t + \delta) = \mathbf{P}(\xi < t + \delta).\end{aligned}$$

Similarly,

$$\liminf_{n\to\infty} \mathbf{P}(\xi_n + \varepsilon_n < t) \ge \mathbf{P}(\xi < t - \delta).$$

Since $\mathbf{P}(\xi < t \pm \delta)$ can be chosen arbitrary close to $\mathbf{P}(\xi < t)$ by taking a sufficiently small δ, the required convergence follows.

The second assertion can be proved in the same way. The lemma is proved. □

Now we will give analogues of Theorems 6.1.4 and 6.1.7 in terms of distributions.

Theorem 6.2.2 *If $\xi_n \Rightarrow \xi$ and a function $H(s)$ satisfies the conditions of Theorem 6.1.4 then $H(\xi_n) \Rightarrow H(\xi)$.*

Theorem 6.2.3 *If* $\xi_n \Rightarrow \xi$ *and the sequence* $\{\xi_n\}$ *is uniformly integrable then* $\mathbf{E}\xi$ *exists and* $\mathbf{E}\xi_n \to \mathbf{E}\xi$.

Proof There are two ways of proving these theorems. One of them consists of reducing them to Theorems 6.1.4 and 6.1.7. To this end, one has to construct random variables ${\xi'}_n = F_n^{(-1)}(\eta)$ and $\xi' = F^{(-1)}(\eta)$, where $\eta \in \mathbf{U}_{0,1}$ and $F_n^{(-1)}$ and $F^{(-1)}$ are the quantile transforms of F_n and F, respectively, and prove that ${\xi'}_n \xrightarrow{p} \xi'$ (we already know that $F^{(-1)}(\eta) \in \mathbf{F}$; if F is discontinuous or not strictly increasing, then $F^{(-1)}$ should be defined as in Lemma 6.2.1).

Another approach is to prove the theorems anew using the language of distributions. Under inessential additional assumptions, such proofs are sometimes even simpler. To illustrate this, assume, for instance, in Theorem 6.2.3 that the function H is continuous. One has to prove that $\mathbf{E}g(H(\xi_n)) \to \mathbf{E}g(H(\xi))$ for any continuous bounded function g. But this is an immediate consequence of (6.2.1) and (6.2.2), for $f = g \circ H$ (f is the composition of the functions g and H).

In Theorem 6.2.3 assume that $\xi_n \ge 0$ (this does not restrict the generality). Then, integrating by parts, we get

$$\mathbf{E}\xi_n = -\int_0^\infty x\, d\mathbf{P}(\xi_n \ge x) = \int_0^\infty \mathbf{P}(\xi_n \ge x)\, dx. \tag{6.2.5}$$

Since by virtue of uniform integrability

$$\sup_n \int_N^\infty \mathbf{P}(\xi_n \ge x)\, dx \le \sup_n \mathbf{E}(\xi_n;\, \xi_n \ge N) \to 0$$

as $N \to \infty$, the integral in (6.2.5) is uniformly convergent. Moreover, $\mathbf{P}(\xi_n \ge x) \to \mathbf{P}(\xi \ge x)$ a.s., and therefore

$$\lim_{n\to\infty} \mathbf{E}\xi_n = \lim_{n\to\infty} \int_0^\infty \mathbf{P}(\xi_n \ge x)\, dx = \int_0^\infty \mathbf{P}(\xi \ge x)\, dx = \mathbf{E}\xi. \qquad \square$$

Conditions ensuring uniform integrability are contained in Lemma 6.1.1. Now we will give a modification of assertion 4 of this lemma for the case of weak convergence.

Lemma 6.2.3 *If* $\{\xi_n\}$ *is left uniformly integrable*, $\xi_n \Rightarrow \xi$ *and* $\mathbf{E}\xi_n \to \mathbf{E}\xi$ *then* $\{\xi_n\}$ *is uniformly integrable*.

We suggest to the reader to construct examples showing that all three conditions of the lemma are essential.

Lemma 6.2.3 implies, in particular, that *if* $\xi_n \ge 0$, $\xi_n \Rightarrow \xi$ *and* $\mathbf{E}\xi_n \to \mathbf{E}\xi$ *then* $\{\xi_n\}$ *is uniformly integrable*.

As for Theorems 6.2.2 and 6.2.3, two alternative ways to prove the result are possible here. One of them consists of using Lemma 6.1.1. We will present here a different, somewhat simpler, proof.

Proof of Lemma 6.2.3 For simplicity assume that $\xi_n \ge 0$. Suppose that the lemma is not valid. Then there exist an $\varepsilon > 0$ and subsequences $n' \to \infty$ and $N(n') \to \infty$ such that

$$\mathbf{E}\big(\xi_{n'};\ \xi_{n'} > N\big(n'\big)\big) > \varepsilon.$$

Since

$$\mathbf{E}\xi_{n'} = \mathbf{E}(\xi_{n'};\ \xi_{n'} \le N) + \mathbf{E}(\xi_{n'}; \xi_{n'} > N),$$

for any N that is a point of continuity of the distribution of ξ, one has

$$\mathbf{E}\xi = \lim_{n\to\infty} \xi_{n'} \ge \mathbf{E}(\xi;\ \xi \le N) + \varepsilon.$$

Choose an N such that the first summand on the right-hand side exceeds $\mathbf{E}\xi - \varepsilon/2$. Then we obtain the contradiction $\mathbf{E}\xi \ge \mathbf{E}\xi + \varepsilon/2$, which proves the lemma.

We leave it to the reader to extend the proof to the case of arbitrary left uniformly integrable $\{\xi_n\}$. □

The following theorem can also be useful.

Theorem 6.2.4 *Suppose that $\xi_n \Rightarrow \xi$, $H(s)$ is differentiable at a point a, and $b_n \to 0$ as $n \to \infty$. Then*

$$\frac{1}{b_n}\big(H(a + b_n\xi_n) - H(a)\big) \Rightarrow \xi H'(a).$$

If $H'(a) = 0$ and $H''(a)$ exists then

$$\frac{1}{b_n^2}\big(H(a + b_n\xi_n) - H(a)\big) \Rightarrow \frac{\xi^2}{2} H''(a).$$

Proof Consider the function

$$h(x) = \begin{cases} \frac{H(a+x)-H(a)}{x} & \text{if } x \ne 0, \\ H'(a) & \text{if } x = 0, \end{cases}$$

which is continuous at the point $x = 0$. Since $b_n\xi_n \Rightarrow 0$, by Theorem 6.2.2 one has $h(b_n\xi_n) \Rightarrow h(0) = H'(a)$. Using the theorem again (this time for two-dimensional distributions), we get

$$\frac{H(a + b_n\xi_n) - H(a)}{b_n} = h(b_n\xi_n)\xi_n \Rightarrow H'(a)\xi.$$

The second assertion is proved in the same way. □

A multivariate analogue of this theorem will look somewhat more complicated. The reader could obtain it himself, following the lines of the argument proving Theorem 6.2.4.

6.3 Conditions for Weak Convergence

Now we will return to the concept of weak convergence. We have two criteria for this convergence: relation (6.2.1) and Theorem 6.2.1. However, from the point of view of their possible applications (their verification in concrete problems) both these criteria are inconvenient. For instance, proving, say, convergence $\mathbf{E}f(\xi_n) \to \mathbf{E}f(\xi)$ not for all continuous bounded functions f but just for elements f of a certain rather narrow class of functions that has a simple and clear nature would be much easier. It is obvious, however, that such a class cannot be very narrow.

Before stating the basic assertions, we will introduce a few concepts.

Extend the class $\mathfrak{F}$ of all distribution functions to the class $\mathfrak{G}$ of all functions G satisfying conditions F1 and F2 from Sect. 3.2 and conditions $G(-\infty) \ge 0$, $G(\infty) \le 1$. Functions G from $\mathfrak{G}$ could be called generalised distribution functions. One can think of them as distribution functions of improper random variables assuming infinite values with positive probabilities, so that $G(-\infty) = \mathbf{P}(\xi = -\infty)$ and $1 - G(\infty) = \mathbf{P}(\xi = \infty)$. We will write $G_n \Rightarrow G$ for $G_n \in \mathfrak{G}$ and $G \in \mathfrak{G}$ if $G_n(x) \to G(x)$ at all points of continuity of $G(x)$.

Theorem 6.3.1 (Helly) *The class $\mathfrak{G}$ is compact with respect to convergence $\Rightarrow$, i.e. from any sequence $\{G_n\}$, $G_n \in \mathfrak{G}$, one can choose a convergent subsequence $G_{n_k} \Rightarrow G \in \mathfrak{G}$.*

For the proof of Theorem 6.3.1, see Appendix 4.

Corollary 6.3.1 *If each convergent subsequence $\{G_{n_k}\}$ of $\{G_n\}$ with $G_n \in \mathfrak{G}$ converges to G then $G_n \Rightarrow G$.*

Proof If $G_n \not\Rightarrow G$ then there exists a point of continuity x_0 of G such that $G_n(x_0) \not\to G(x_0)$. Since $G_n(x_0) \in [0, 1]$, there exists a convergent subsequence G_{n_k} such that $G_{n_k}(x_0) \to g \ne G(x_0)$. This, however, is impossible by our assumption, for $G_{n_k}(x_0) \to G(x_0)$. □

The reason for extending the class $\mathfrak{F}$ of all distribution functions is that it is not compact (in the sense of Theorem 6.3.1) and convergence $F_n \Rightarrow G$, $F_n \in \mathfrak{F}$, does not imply that $G \in \mathfrak{F}$. For example, the sequence

$$F_n(x) = \begin{cases} 0 & \text{if } x \le -n, \\ 1/2 & \text{if } -n < x \le n, \\ 1 & \text{if } x > n \end{cases} \tag{6.3.1}$$

converges everywhere to the function $G(x) \equiv 1/2 \notin \mathfrak{F}$ corresponding to an improper random variable taking the values $\pm\infty$ with probabilities $1/2$.

However, dealing with the class $\mathfrak{G}$ is also not very convenient. The fact is that convergence at points of continuity $G_n \Rightarrow G$ in the class $\mathfrak{G}$ is not equivalent to convergence

$$\int f\,dG_n \to \int f\,dG$$

(see example (6.3.1) for $f \equiv 1$), and the integrals $\int f\,dG$ do not specify G uniquely (they specify the *increments* of G, but not the values $G(-\infty)$ and $G(\infty)$). Now we will introduce two concepts that will help to avoid the above-mentioned inconvenience.

Definition 6.3.1 A sequence of distributions $\{\mathbf{F}_n\}$ (or distribution functions $\{F_n\}$) is said to be *tight* if, for any $\varepsilon > 0$, there exists an N such that

$$\inf_n \mathbf{F}_n\big([-N, N]\big) > 1 - \varepsilon. \tag{6.3.2}$$

Definition 6.3.2 A class $\mathcal{L}$ of continuous bounded functions is said to be *distribution determining* if the equality

$$\int f(x)\,dF(x) = \int f(x)\,dG(x), \quad F \in \mathcal{F},\ G \in \mathcal{G},$$

for all $f \in \mathcal{L}$ implies that $F = G$ (or, which is the same, if the relation $\mathbf{E}f(\xi) = \mathbf{E}f(\eta)$ for all $f \in \mathcal{L}$, where one of the random variables ξ and η is proper, implies that $\xi \overset{d}{=} \eta$).

The next theorem is the main result of the present section.

Theorem 6.3.2 *Let $\mathcal{L}$ be a distribution determining class and $\{F_n\}$ a sequence of distributions. For the existence of a distribution $F \in \mathcal{F}$ such that $F_n \Rightarrow F$ it is necessary and sufficient that:*[3]

(1) *the sequence $\{F_n\}$ is tight*; *and*
(2) $\lim_{n\to\infty} \int f\,dF_n$ *exists for all $f \in \mathcal{L}$.*

Proof The *necessary* part is obvious.

Sufficiency. By Theorem 6.3.1 there exists a subsequence $F_{n_k} \Rightarrow F \in \mathcal{G}$. But by condition (1) one has $F \in \mathcal{F}$. Indeed, if $x \geq N$ is a point of continuity of F then, by Definition 6.3.1, $F(x) = \lim F_{n_k}(x) \geq 1 - \varepsilon$. In a similar way we establish that for $x \leq -N$ one has $F(x) < \varepsilon$. Since ε is arbitrary, we have $F(-\infty) = 0$ and $F(\infty) = 1$.

Further, take another convergent subsequence $F_{n'_k} \Rightarrow G \in \mathcal{F}$. Then, for any $f \in \mathcal{L}$, one has

$$\lim \int f\,dF_{n_k} = \int f\,dF, \qquad \lim \int f\,dF_{n'_k} = \int f\,dG. \tag{6.3.3}$$

But, by condition (2),

$$\int f\,dF = \int f\,dG, \tag{6.3.4}$$

and hence $F = G$. The theorem is proved by virtue of Corollary 6.3.1. □

[3] In this form the theorem persists for spaces of a more general nature. The role of the segments $[-N, N]$ in (6.3.2) is played in that case by compact sets (cf. [1, 14, 25, 31]).

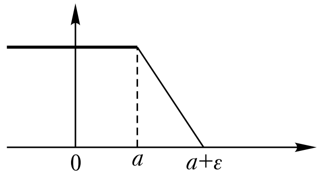

Fig. 6.1 The plot of the function $f_{a,\varepsilon}(x)$ from Example 6.3.1

If one needs to prove convergence to a "known" distribution $F \in \mathcal{F}$, the tightness condition in Theorem 6.3.2 becomes redundant.

Corollary 6.3.2 *Let $\mathcal{L}$ be a distribution determining class and*

$$\int f\,dF_n \to \int f\,dF, \quad F \in \mathcal{G}, \tag{6.3.5}$$

for any $f \in \mathcal{L}$. Moreover, assume that at least one of the following three conditions is met:

(1) *the sequence $\{F_n\}$ is tight*;
(2) $F \in \mathcal{F}$;
(3) $f \equiv 1 \in \mathcal{L}$ (*i.e.* (6.3.5) *holds for* $f \equiv 1$).

Then $F \in \mathcal{F}$ and $F_n \Rightarrow F$.

The proof of the corollary is almost next to obvious. Under condition (1) the assertion follows immediately from Theorem 6.3.2. Condition (3) and convergence (6.3.5) imply condition (2). If (2) holds, then $F \in \mathcal{F}$ in relations (6.3.3) and (6.3.4), and therefore $G = F$. □

Since, as a rule, at least one of conditions (1)–(3) is satisfied (as we will see below), the basic task is to verify convergence (6.3.5) for the class $\mathcal{L}$.

Note also that, in the case where one proves *convergence to a distribution $F \in \mathcal{F}$ "known" in advance, the whole arrangement of the argument can be different and simpler*. One such alternative approach is presented in Sect. 7.4.

Now we will give several examples of distributions determining classes $\mathcal{L}$.

Example 6.3.1 The class $\mathcal{L}_0$ of functions having the form

$$f_{a,\varepsilon}(x) = \begin{cases} 1 & \text{if } x \le a, \\ 0 & \text{if } x \ge a + \varepsilon. \end{cases}$$

On the segment $[a,\, a+\varepsilon]$ the functions $f_{a,\varepsilon}$ are defined to be linear and continuous (a plot of $f_{a,\varepsilon}(x)$ is given in Fig. 6.1). It is a two-parameter family of functions.

We show that $\mathcal{L}_0$ is a distribution determining class. Let

$$\int f\,dF = \int f\,dG$$

for all $f \in \mathcal{L}_0$. Then

$$F(a) \le \int f_{a,\varepsilon}\,dF = \int f_{a,\varepsilon}\,dG \le G(a+\varepsilon),$$

and, conversely,

$$G(a) \le F(a+\varepsilon)$$

for any $\varepsilon > 0$. Taking a to be a point of continuity of both F and G, we obtain that

$$F(a) = G(a).$$

Since this is valid for all points of continuity, we get $F = G$.

One can easily verify in a similar way that the class $\widehat{\mathcal{L}}_0$ of "trapezium-shaped" functions $f(x) = \min(f_{a,\varepsilon}, 1 - f_{b,\varepsilon})$, $a < b$, is also distribution determining.

Example 6.3.2 The class $\mathcal{L}_1$ of continuous bounded functions such that, for each $f \in \mathcal{L}_0$ (or $f \in \widehat{\mathcal{L}}_0$), there exists a sequence $f_n \in \mathcal{L}_1$, $\sup_x |f(x)| < M < \infty$, for which $\lim_{n\to\infty} f_n(x) = f(x)$ for each $x \in \mathbb{R}$.

Let

$$\int f\,dF = \int f\,dG$$

for all $f \in \mathcal{L}_1$. By the dominated convergence theorem,

$$\lim \int f_n\,dF = \int f\,dF, \qquad \lim \int f_n\,dG = \int f\,dG, \quad f \in \mathcal{L}_0.$$

Therefore

$$\int f\,dF = \int f\,dG, \quad f \in \mathcal{L}_0, \qquad \mathfrak{F} = \mathfrak{G}$$

and hence $\mathcal{L}_1$ is a distribution determining class.

Example 6.3.3 The class C_k of all bounded functions $f(x)$ having bounded uniformly continuous k-th derivatives $f^{(k)}(x)$ ($\sup_x |f^{(k)}(x)| < \infty$), $k \ge 1$.

It is evident that C_k is a distribution determining class for it is a special case of an $\mathcal{L}_1$ class.

In the same way one can see that the subclass $C_k^0 \subset C_k$ of functions having finite support (vanishing outside a finite interval) is also distribution determining. This follows from the fact that C_k^0 is an $\mathcal{L}_1$-class with respect to the class $\widehat{\mathcal{L}}_0$ of trapezium-shaped (and therefore having compact support) functions.

It is clear that the class C_k satisfies condition (3) from Corollary 6.3.2 ($f \equiv 1 \in C_k$). Therefore, to prove convergence $F_n \Rightarrow F \in \mathfrak{F}$ it suffices to verify convergence (6.3.5) for $f \in C_k$ only.

If one takes $\mathcal{L}$ to be the class C_k^0 of differentiable functions with finite support then relation (6.3.5) together with condition (2) of Corollary 6.3.2 could be re-written as

$$\int F_n f'\,dx \to \int F f'\,dx, \quad F \in \mathfrak{F}. \tag{6.3.6}$$

(One has to integrate (6.3.5) by parts and use the fact that f' also has a finite support.) The convergence criterion (6.3.6) is sometimes useful. It can be used to show, for example, that (6.3.5) follows from convergence $F_n(x) \to F(x)$ at all points of continuity of F (i.e. almost everywhere), since that convergence and the dominated convergence theorem imply (6.3.6) which is equivalent to (6.3.5).

Example 6.3.4 One of the most important distribution determining classes is the one-parameter family of complex-valued functions $\{e^{itx}\}$, $t \in \mathbb{R}$.

The next chapter will be devoted to studying the properties of $\int e^{itx} dF(x)$.

After obvious changes, all the material in the present chapter can be extended to the multivariate case.