2.4. CONVERGENCE OF RANDOM VARIABLES

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and random variables $\{X_n\}_{n\in\mathbf{N}}$, X on it. We have already encountered many different notions for the convergence of the sequence $\{X_n\}_{n\in\mathbf{N}}$ to the random variable X, and we shall need a few more. The purpose of this section is to present all these notions together, at the expense perhaps of some minor repetition, and to discuss in a systematic way how they are interrelated.

• **Pointwise Convergence:** This means that the sequence of real numbers $\{X_n(\omega)\}_{n\in\mathbf{N}}$ converges to the real number $X(\omega)$, for every $\omega \in \Omega$. It is the simplest notion of convergence.

• **Almost-Everywhere Convergence:** This is the most familiar notion of convergence, meaning that there exists a set $\Omega^* \in \mathcal{F}$ with $\mathbf{P}(\Omega^*) = 1$ and such that the sequence of real numbers $\{X_n(\omega)\}_{n\in\mathbf{N}}$ converges to the real number $X(\omega)$ for every $\omega \in \Omega^*$.

We write then $X_n \longrightarrow X$, a.e. This is the notion of convergence in the strong law of large numbers.

Similarly, we say that the sequence $\{X_n\}_{n\in\mathbf{N}}$ is *Cauchy a.e.* if the sequence of real numbers $\{X_n(\omega)\}_{n\in\mathbf{N}}$ is Cauchy for every $\omega \in \Omega^*$: $\lim_{n\to\infty,\, m\to\infty} |X_n(\omega) - X_m(\omega)| = 0$.

• **Convergence in Probability:** The sequence $\{X_n\}_{n\in\mathbf{N}}$ converges to X *in probability*, if for every $\varepsilon > 0$ we have $\mathbf{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$. This is the notion of convergence in the weak law of large numbers.

By the same token, we say that the sequence $\{X_n\}_{n\in\mathbf{N}}$ is *Cauchy in probability,* if $\lim_{m,n\to\infty} \mathbf{P}(|X_n - X_m| > \varepsilon) = 0$ holds for every $\varepsilon > 0$.

• **Convergence in $\mathbf{L}^p$:** The sequence $\{X_n\}_{n\in\mathbf{N}} \subset \mathbf{L}^p$ *converges in* $\mathbf{L}^p$ to $X \in \mathbf{L}^p$ for some $p \in (0, \infty)$, if $\mathbf{E}(|X_n - X|^p) \to 0$ as $n \to \infty$.

We say that $\{X_n\}_{n\in\mathbf{N}}$ is *Cauchy in* $\mathbf{L}^p$, if we have $\lim_{m,n\to\infty} \mathbf{E}(|X_n - X_m|^p) = 0$.

• **Weak Convergence in $\mathbf{L}^p$:** The sequence $\{X_n\}_{n\in\mathbf{N}} \subset \mathbf{L}^p$ converges to $X \in \mathbf{L}^p$ *weakly in* $\mathbf{L}^p$ for some $p \in [1, \infty]$ if $\lim_{n\to\infty} \mathbf{E}(X_n Y) = \mathbf{E}(XY)$ holds for every $Y \in \mathbf{L}^q$ with the usual notation $(1/p) + (1/q) = 1$. This is the mode of convergence introduced in (1.8.7).

• **Convergence in Distribution:** The sequence $\{X_n\}_{n\in\mathbf{N}}$ converges to the random variable X *in distribution*, if the sequence of distribution functions $\{F_{X_n}(\cdot)\}_{n\in\mathbf{N}}$ converges to the distribution function $F_X(\cdot)$ at every continuity-point x of $F_X(\cdot)$.

This is the notion of convergence in the DeMoivre/Laplace Theorem of Exercise 2.5 and in its generalization, the Central Limit Theorem.

• **Vague Convergence:** The sequence $\{X_n\}_{n\in\mathbf{N}}$ converges to X *vaguely*, if we have $\mathbf{E}[\Phi(X_n)] \to \mathbf{E}[\Phi(X)]$ as $n \to \infty$, for any bounded, continuous function $\Phi : \mathbf{R} \to \mathbf{R}$.

Similarly, we say that a sequence $\{\mu_n\}_{n\in\mathbf{N}}$ of probability measures on $\mathcal{B}(\mathbf{R})$ converges *vaguely* to a probability measure μ on $\mathcal{B}(\mathbf{R})$, if

$$\lim_{n\to\infty}\int \Phi\, d\mu_n \;=\; \int \Phi\, d\mu \quad \text{holds for any bounded, continuous function } \;\Phi:\mathbf{R}\to\mathbf{R}\,.$$

Of course, the sequence $\{X_n\}_{n\in\mathbf{N}}$ converges to X vaguely, if and only if the corresponding sequence $\{\mu_n\}_{n\in\mathbf{N}}$ of induced measures $\mu_n = \mathbf{P}\circ X_n^{-1}$ converges vaguely to the induced measure $\mu = \mathbf{P}\circ X^{-1}$.

We shall see in Theorem 4.3 below that the last two notions of convergence (in distribution, and vague) are equivalent; note also that for these two notions the random variables $\{X_n\}_{n\in\mathbf{N}}$, X need *not* be defined on the same probability space. From Exercise 1.4, if $\{X_n\}_{n\in\mathbf{N}}$ converges vaguely to both X and Y, then these two random variables must have the same distribution.

4.1 THEOREM : Relations Among Different Modes of Convergence. *We have the following implications*:

(i) Convergence $\mathbf{P}-$a.e. $\Rightarrow$ Convergence in Probability $\Rightarrow$ Convergence in Distribution.

(ii) $\sum_{n\in\mathbf{N}}\mathbf{P}(|X_n - X|>\varepsilon)<\infty\,,\ \forall\ \varepsilon>0 \implies \mathbf{P}(|X_n - X|>\varepsilon,\ \text{i.o.})=0\,,\ \forall\ \varepsilon>0$
$\iff X_n\to X\,,\ \mathbf{P}-\text{a.e.} \iff \lim_{k\to\infty}\mathbf{P}\big[\sup_{n\ge k}|X_n - X|>\varepsilon\big]=0\,,\ \forall\ \varepsilon>0.$

(iii) $\sum_{n\in\mathbf{N}}\mathbf{E}(|X_n - X|^p)<\infty$, for some $p\in(0,\infty) \implies X_n\to X\,,\ \mathbf{P}-$a.e.

(iv) Convergence in $\mathbf{L}^p$, for some $p\in(0,\infty) \implies$ Convergence in Probability
$\implies$ Convergence $\mathbf{P}-$a.e. along some subsequence $\{X_{n_k}\}_{k\in\mathbf{N}}$.

(v) $\mathbf{P}\big(\lim_{n\to\infty}X_n \text{ exists in } \mathbf{R}\big) = 1 \iff \{X_n\}_{n\in\mathbf{N}}$ is Cauchy a.e.
$\iff \lim_{n\to\infty}\mathbf{P}\big(\sup_{k\ge1}|X_{n+k}-X_n|>\varepsilon\big) = 0$ holds for every $\varepsilon>0$.

PROOF : First, we recall the definition of convergence of $\{X_n(\omega)\}_{n\in\mathbf{N}}$ to $X(\omega)$: for all $\varepsilon>0$, there exists $k\in\mathbf{N}$, such that $|X_n(\omega)-X(\omega)|\le\varepsilon$ holds for all $n\ge k$. This translates into

$$\{X_n\to X\} := \Big\{\omega\in\Omega\,\Big|\, \lim_{n\to\infty}X_n(\omega)=X(\omega)\Big\}$$
$$= \bigcap_{\varepsilon>0}\bigcup_{k\in\mathbf{N}}\bigcap_{n\ge k}\{|X_n - X|\le\varepsilon\} \;=\; \bigcap_{\varepsilon>0}C(\varepsilon)\,, \qquad (4.1)$$

where we set

$$B_k(\varepsilon) := \bigcap_{n\ge k}\{|X_n - X|\le\varepsilon\} \qquad \text{and} \qquad C(\varepsilon) := \bigcup_{k\in\mathbf{N}}B_k(\varepsilon) = \{|X_n - X|>\varepsilon\,,\ \text{i.o.}\}^c.$$

Note that $\left(C(\varepsilon)\right)^c = \bigcap_{k\in\mathbf{N}} \bigcup_{n\geq k} \left\{|X_n - X| > \varepsilon\right\} \equiv \{|X_n - X| > \varepsilon, \text{ i.o.}\}$.

Now $X_n \to X$ a.e. $\Leftrightarrow$ $\mathbf{P}(X_n \to X) = 1$ $\Leftrightarrow$ $\mathbf{P}(C(\varepsilon)) = 1$ for all $\varepsilon > 0$, so we obtain the following characterization of $\mathbf{P}-$a.e. convergence:

$$X_n \to X \quad \text{a.e.} \qquad \Longleftrightarrow \qquad \mathbf{P}(|X_n - X| > \varepsilon\,, \text{ i.o.}) = 0\,, \quad \forall\, \varepsilon > 0\,. \tag{4.2}$$

Since $\mathbf{P}(|X_n - X| > \varepsilon\,, \text{ i.o.}) = 0$ if $\sum_{n\in\mathbf{N}} \mathbf{P}(|X_n - X| > \varepsilon) < \infty$ from the Borel-Cantelli lemma, we have also established (ii).

- Returning to (4.1), we note that $\mathbf{P}(C(\varepsilon)) = 1 \iff \lim_{k\to\infty}\mathbf{P}(B_k(\varepsilon)) = 1$, for all $\varepsilon > 0$. This gives another characterization of $\mathbf{P}-$a.e. convergence:

$$\begin{aligned} X_n \to X\,,\ \mathbf{P}-\text{a.e.} \iff & \ \forall\,\varepsilon > 0\,,\ \ \lim_{k\to\infty} \mathbf{P}\big(|X_n - X| \leq \varepsilon,\ \forall\, n \geq k\big) = 1 \\ \iff & \ \forall\,\varepsilon > 0\,,\ \ \lim_{k\to\infty} \mathbf{P}\big(\sup_{n\geq k}|X_n - X| > \varepsilon\big) = 0. \end{aligned} \tag{4.3}$$

It also establishes the last equivalence in (ii), as well as the first implication in (i), since $\mathbf{P}(|X_k - X| > \varepsilon) \leq \mathbf{P}(\sup_{n\geq k}|X_n - X| > \varepsilon)$.

- A more direct way of showing that convergence $\mathbf{P}-$a.e. implies convergence in probability, proceeds as follows: the random variables $Y_n = \chi_{\{|X_n - X| > \varepsilon\}}$ take values in $[0,1]$ and converge a.e. to zero, so the Dominated Convergence Theorem (applicable because we are working on a finite measure space!) gives $\mathbf{P}(|X_n - X| > \varepsilon) = \mathbf{E}(Y_n) \to 0$ as $n \to \infty$.
- The characterization (4.3) also leads easily to an example, showing that *convergence in probability does not imply a.e. convergence.* It suffices to exhibit a sequence $\{X_n\}_{n\in\mathbf{N}}$ with

$$\lim_{n\to\infty} \mathbf{P}\big(|X_n| > \varepsilon\big) = 0\,, \quad \text{but} \quad \mathbf{P}\big(\sup_{k\geq n}|X_k| > \varepsilon\big) = 1\,, \quad \forall\, n \in \mathbf{N}\,.$$

Such a sequence is given by $\Omega = [0,1)$ with Lebesgue measure, $X_m^k(\omega) = 1$ if $\omega \in [\frac{k-1}{2^m}, \frac{k}{2^m})$ and $X_m^k(\omega) = 0$ otherwise, for $1 \leq k \leq 2^m$, $1 \leq m < \infty$.

- For the second implication in (iv), we argue in a manner quite similar to the proof of Theorem 1.5.1. Suppose that the sequence $\{X_n\}_{n\in\mathbf{N}}$ is **Cauchy in probability**, i.e., for every $\varepsilon > 0$ we have: $\mathbf{P}\left[\,|X_n - X_m| > \varepsilon\,\right] \longrightarrow 0$, as $n \to \infty$, $m \to \infty$.

This is certainly the case, if $\{X_n\}_{n\in\mathbf{N}}$ converges in probability. Then we can select a subsequence $\{Y_k\}_{k\in\mathbf{N}} \equiv \{X_{n_k}\}_{k\in\mathbf{N}} \subseteq \{X_n\}_{n\in\mathbf{N}}$ satisfying

$$\mathbf{P}(E_k) \leq \frac{1}{2^k}\,, \quad \forall\, k \in \mathbf{N}\,, \quad \text{where } E_k := \left\{|Y_{k+1} - Y_k| > 2^{-k}\right\} = \left\{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\right\}\,;$$

to wit, $\mathbf{P}(E_k)$ converges to zero "very fast". Now notice that the event $F_m := \bigcup_{k\geq m} E_k$ has $\mathbf{P}(F_m) \leq \sum_{k\geq m} \mathbf{P}(E_k) \leq 2^{-m+1}$.

In particular, for every $\omega \in \Omega \setminus F_m$:

$$|Y_\ell(\omega) - Y_k(\omega)| \le \sum_{j=\ell}^{k-1} |Y_{j+1}(\omega) - Y_j(\omega)| \le \sum_{j=\ell}^{k-1} 2^{-j} \le 2^{-\ell+1}, \quad \forall\, k > \ell > m\,.$$

Thus $\{Y_k(\omega)\}_{k\in\mathbf{N}}$ is a Cauchy sequence of real numbers, and $Y(\omega) := \lim_{k\to\infty} Y_k(\omega)$ exists in $\mathbf{R}$ for every $\omega \in \bigcup_{m\in\mathbf{N}}(\Omega \setminus F_m) = \Omega \setminus F$, where now $F := \bigcap_{m\in\mathbf{N}} F_m = \limsup_{k\to\infty} E_k$ has $\mathbf{P}(F) \le \mathbf{P}(F_m) \le 2^{-m+1}$ for all $m \in \mathbf{N}$, thus $\mathbf{P}(F) = 0$. Defining $Y(\omega) \equiv 0$ on F we see that: $Y_k = X_{n_k} \to Y$, $\mathbf{P}-$a.e.

- Assume now that $X_n \to X$ in probability. Then

$$\begin{aligned}\mathbf{P}(X_n \le x) &= \mathbf{P}(X_n \le x, |X_n - X| \le \varepsilon) + \mathbf{P}(X_n \le x, |X_n - X| > \varepsilon)\\ &\le \mathbf{P}(X \le x + \varepsilon) + \mathbf{P}(|X_n - X| > \varepsilon)\,,\\ \mathbf{P}(X \le x - \varepsilon) &\le \mathbf{P}(X_n \le x) + \mathbf{P}(|X_n - X| > \varepsilon)\,,\end{aligned} \tag{4.4}$$

so that letting $n \to \infty$ gives

$$F_X(x-\varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x+\varepsilon) \le F_X(x+\varepsilon)\,. \tag{4.5}$$

If x is a point of continuity of $F_X(\cdot)$, we can let $\varepsilon \to 0$ to get $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. This completes the proof of (i).

Remark: We note that nothing can be said about $\lim_{n\to\infty} F_{X_n}(x)$ if $F_X(\cdot)$ is discontinuous at x, even if we make the stronger assumption that $X_n \to X$ $\mathbf{P}-$a.e. In fact, we have in general

$$\{X < x\} \subseteq \liminf_{n\to\infty}\{X_n < x\} \subseteq \limsup_{n\to\infty}\{X_n < x\} \subseteq \{X \le x\} \tag{4.6}$$

so difficulties arise whenever $\mathbf{P}(X = x) > 0$. For instance, if we set $\Omega = [0,1)$ with Lebesgue measure, $X_n = \frac{1}{n} \to X = 0$, then 0 is a discontinuity point for $F_X(\cdot)$, $F_X(0) = 1$, while $F_{X_n}(0) = 0$ for all $n \in \mathbf{N}$.

- We observe that (iii) has already been proved as part of the proof of completeness of $\mathbf{L}^p$ spaces, when $p \ge 1$. Here we point out that, for probability spaces and any $p \in (0,\infty)$, the claim (iii) follows easily from (ii) and the Čebyšev inequality $\sum_{n\in\mathbf{N}} \mathbf{P}(|X_n - X| > \varepsilon) \le \varepsilon^{-p} \sum_{n\in\mathbf{N}} \mathbf{E}(|X_n - X|^p)$. The first part of (iv) is also a simple consequence of Čebyšev's inequality. ◇

We are now in a position to provide Etemadi's (1981) proof of the Kolmogorov Strong Law of Large Numbers.

PROOF OF THEOREM 3.1 : The random variables in each of the sequences $\{X_n^\pm\}_{n\in\mathbf{N}}$ of positive and negative parts, are still pairwise independent and identically distributed, so it suffices to prove the theorem with X_n replaced by $X_n^{\pm}$. Thus, we may assume that $X_n \geq 0$. We use the *method of truncation* as in the proof of Theorem 3.2, and set $Z_n = X_n$ when $X_n \leq n$, $Z_n = 0$ otherwise. The main step is to show that for $\mathbf{P}-$a.e. $\in \Omega$ we have

$$\frac{1}{n}\sum_{k=1}^{n} Z_k \longrightarrow \mathbf{E}(X_1) \qquad \text{as } n \to \infty\,. \tag{4.7}$$

For this, we begin by proving this kind of convergence along a suitable subsequence. More precisely, let $\alpha > 1$ be arbitrary, and set $k_n = [\alpha^n]$ (this defines an increasing sequence for n sufficiently large). Then, just as in the proof of the Weak Law of Large Numbers (Theorem 3.2), we obtain

$$\varepsilon^2 \sum_{n\in\mathbf{N}} \mathbf{P}\left(\sum_{j=1}^{k_n} |Z_j - \mathbf{E}(Z_j)| > \varepsilon k_n\right) \leq \sum_{n\in\mathbf{N}} \frac{1}{k_n^2}\sum_{j=1}^{k_n}\mathbf{E}(Z_j^2) \leq \text{const.} \sum_{j\in\mathbf{N}} \mathbf{E}(Z_j^2) \sum_{\substack{n\in\mathbf{N}\\ [\alpha^n]\geq j}} \frac{1}{[\alpha^n]^2}. \tag{4.8}$$

But this last sum is essentially a convergent geometric series, so its size is that of its largest term j^{-2}. Hence the double sum is dominated by a constant, times

$$\sum_{j\in\mathbf{N}} \frac{1}{j^2}\,\mathbf{E}(Z_j^2) \;=\; \sum_{j\in\mathbf{N}} \frac{1}{j^2}\,\mathbf{E}\big[\,(X_j)^2\,\chi_{\{X_j\leq j\}}\,\big] \;=\; \sum_{j\in\mathbf{N}} \frac{1}{j^2}\,\mathbf{E}\big[\,(X_1)^2\,\chi_{\{X_1\leq j\}}\,\big]\,.$$

This last series can be re-expressed as

$$\begin{aligned}
\sum_{j\in\mathbf{N}} \frac{1}{j^2}\sum_{l=0}^{j-1}\mathbf{E}\big[\,(X_1)^2\,\chi_{\{l\leq X_1<l+1\}}\,\big] &= \sum_{l\in\mathbf{N}_0}\mathbf{E}\big[\,(X_1)^2\,\chi_{\{l\leq X_1<l+1\}}\,\big]\left(\sum_{j+1}\frac{1}{j^2}\right)\\
&= \text{const.}\sum_{l\in\mathbf{N}_0}\frac{1}{l+1}\,\mathbf{E}\big[\,(X_1)^2\,\chi_{\{l\leq X_1<l+1\}}\,\big]\\
&\leq \text{const.}\sum_{l\in\mathbf{N}_0}\mathbf{E}\big[\,X_1\cdot\chi_{\{l\leq X_1<l+1\}}\,\big] \;=\; \mathbf{E}(X_1) < \infty.
\end{aligned}$$

Thus, the left-hand side in (4.8) is finite. This finiteness implies, by Theorem 4.1 (iii), that we have the convergence $\lim_n \frac{1}{k_n}\sum_{j=1}^{k_n}(Z_j - \mathbf{E}(Z_j)) = 0$, a.e. But

$$\lim_n \frac{1}{k_n}\sum_{j=1}^{k_n}\mathbf{E}(Z_j) \;=\; \lim_n \mathbf{E}(Z_n) \;=\; \lim_n \mathbf{E}\big(X_1\,\chi_{\{X_1\leq n\}}\big) \;=\; \mathbf{E}(X_1)$$

and then $\lim_n \frac{1}{k_n}\sum_{j=1}^{k_n} Z_j = \mathbf{E}(X_1)$, a.e.

To obtain convergence for the *entire sequence,* we show convergence for *any* subsequence $\{m_j\}$. For every $j \in \mathbf{N}$, define n_j by $[\alpha^{n_j}] \leq m_j < [\alpha^{n_j+1}]$. Since $Z_n \geq 0$, we have

$$\frac{\sum_{j=1}^{[\alpha^{n_k}]} Z_j}{\alpha^2[\alpha^{n_k}]} \leq \frac{\sum_{j=1}^{[\alpha^{n_k}]} Z_j}{[\alpha^{n_k+1}]} \leq \frac{\sum_{j=1}^{m_k} Z_j}{m_k} \leq \frac{\sum_{j=1}^{[\alpha^{n_k}]+1} Z_j}{[\alpha^{n_k}]} \leq \frac{\alpha^2\,\sum_{j=1}^{[\alpha^{n_k}]+1} Z_j}{[\alpha^{n_k}]}\,.$$

Letting $k \to \infty$ yields

$$\frac{\mathbf{E}(X_1)}{\alpha^2} \le \text{liminf}_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Z_j \le \text{limsup}_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Z_j \le \alpha^2\, \mathbf{E}(X_1)\,;$$

and letting $\alpha \to 1$, we obtain (4.7). Finally, we need to replace the truncated variables $\{Z_n\}$ in (4.7), by the original variables $\{X_n\}$; but this is done in exactly the same way, as in the proof of Theorem 3.2. $\diamond$

In Theorem 4.1 there is no mention of whether convergence in distribution can provide any information on any other kind of convergence. This is to be expected, of course, since distinct random variables can have the same distribution! Perhaps paradoxically, however, convergence in distribution actually implies convergence pointwise a.e., if we can *choose both* the representative random variables *and* the probability spaces they are defined on. In fact, we have the following result, which complements Proposition 1.1.

4.2 Theorem : Skorokhod Representation Theorem. *Let F, $\{F_n\}_{n\in\mathbf{N}}$ be probability distribution functions on the real line, and suppose that $\lim_n F_n(x) = F(x)$ at any continuity point x of $F(\cdot)$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and real-valued random variables X, $\{X_n\}_{n\in\mathbf{N}}$ on it, such that $F_X = F$, $F_{X_n} = F_n$ for all $n \in \mathbf{N}$, and $X_n \to X$, $\mathbf{P}-$a.e.*

Proof : Take $\Omega = [0,1]$ with its Borel sets and Lebesgue measure $\mathbf{P} \equiv \lambda$, and recall the Skorohod construction $X_n^+(\omega) = \inf\{x \,:\, F_n(x) > \omega\} \ \ge\ X_n^-(\omega) = \inf\{x \,:\, F_n(x) \ge \omega\}$ of Proposition 1.1. We have $F_{X_n^\pm} \equiv F_n$, $\mathbf{P}(X_n^+ = X_n^-) = 1$ for every $n \in \mathbf{N}$, and $F_{X^\pm} \equiv F$, $\mathbf{P}(X^+ = X^-) = 1$. Denote by $\mathcal{D}$ the set of discontinuity points of F.

Fix now $\omega \in \Omega$, and take $x \in (X^+(\omega), \infty) \cap \mathcal{D}^c$. We have then $x > X^+(\omega)$, so $F(x) > \omega$, and consequently $F_n(x) > \omega$ (hence also $X_n^+(\omega) \le x$) for all n sufficiently large; thus, it develops that $\limsup_n X_n^+(\omega) \le x$. Letting $x \downarrow X^+(\omega)$ along $\mathcal{D}^c$ (which is possible, because $\mathcal{D}$ is at most countable) we obtain $\limsup_n X_n^+(\omega) \le X^+(\omega)$. Similarly, we obtain $\liminf_n X_n^+(\omega) \ \ge\ \liminf_n X_n^-(\omega) \ \ge\ X^-(\omega)$. Since $\mathbf{P}(X^+ = X^-) = 1$, the result follows.

4.3 Theorem : Equivalence of Vague and Distributional Convergence. *Let $\{\mu_n\}$, μ be probability measures on the real line, and $\{F_n\}$, F their corresponding distribution functions. Then $F_n(x) \to F(x)$ at all points of continuity of F, if and only if $\int \Phi\, d\mu_n \to \int \Phi\, d\mu$ for all bounded and continuous functions Φ.*

In particular, if X and $\{X_n\}_{n\in\mathbf{N}}$ are random variables, then $X_n \to X$ vaguely, if and only if $X_n \to X$ in distribution.

Proof: By the Skorokhod representation theorem, we can view $\{F_n\}$, F as the distribution functions $\{F_{X_n}\}$, F_X of random variables $\{X_n\}$, X such that $X_n(\omega) \to X(\omega)$ for $\mathbf{P}-$a.e.

$\omega \in \Omega$; then $\{\mu_n\}$, μ are the corresponding induced measures. Then for $\Phi : \mathbf{R} \to \mathbf{R}$ bounded and continuous, we have $\mathbf{E}[\Phi(X_n)] = \int \Phi \, d\mu_n \to \int \Phi \, d\mu = \mathbf{E}[\Phi(X)]$ by the Lebesgue Dominated Convergence Theorem. This shows that convergence in distribution implies vague convergence.

To prove the converse, assume that $\int \Phi \, d\mu_n \to \int \Phi \, d\mu$ for all bounded, continuous Φ (equivalently, that $X_n \to X$ vaguely). Let $x \in \mathbf{R}$, $\delta > 0$, and choose $h : \mathbf{R} \to [0,1]$ to be continuous, equal to $h(y) = 1$ for $y \le x$, and equal to $h(y) = 0$ for $y > x + \delta$. Evidently,

$$F_n(x) \le \mathbf{E}[\,h(X_n)\,] = \int h \, d\mu_n \le F_n(x+\delta), \qquad F(x) \le \mathbf{E}[\,h(X)\,] = \int h \, d\mu \le F(x+\delta)\,.$$

The vague convergence implies

$$\limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} \int h \, d\mu_n = \int h \, d\mu \le F(x+\delta),$$

and hence, using the fact that F is right continuous, $\limsup_{n\to\infty} F_n(x) \le F(x)$. Now let $g(\cdot) = h(\cdot + \delta)$ and observe that we have $F_n(x-\delta) \le \int g \, d\mu_n \le F_n(x)$, as well as $F(x-\delta) \le \int g \, d\mu \le F(x)$, whence

$$\liminf_{n\to\infty} F_n(x) \ge \lim_{n\to\infty} \int g \, d\mu_n = \int g \, d\mu \ge F(x-\delta), \qquad \liminf_{n\to\infty} F_n(x) \ge F(x-).$$

When x is a point of continuity for $F(\cdot)$, we have actually equality everywhere. $\diamond$

A: MISCELLANEOUS EXERCISES

4.1 Exercise: (i) Show that convergence in distribution to a *degenerate* random variable X (with $\mathbf{P}(X = a) = 1$ for some $a \in \mathbf{R}$) implies convergence in probability.

(ii) Show that $\lim_{n\to\infty} Y_n = Y$ in distribution, and $\lim_{n\to\infty} X_n = x$ in probability for some $x \in \mathbf{R}$, imply: $\lim_{n\to\infty} (X_n + Y_n) = x + Y$ in distribution.

(iii) Show that $\lim_{n\to\infty} Y_n = Y$ in distribution, and $\lim_{n\to\infty} X_n = x$ in probability for some $x \in \mathbf{R}$, imply: $\lim_{n\to\infty} (X_n Y_n) = xY$ in distribution.
(*Hint:* Deal with the case $X_n \ge 0$ and $x > 0$ first.)

(iv) *Here is a setting where a.e. convergence implies convergence in* $\mathbf{L}^1$: Suppose the random variables $X, X_1, X_2, \cdots$ are integrable, non-negative, and we have $X_n \longrightarrow X$ a.e., as well as $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$. Then $\lim_{n\to\infty} \mathbf{E}(|\,X_n - X|) = 0$.

(v) Show that a sequence of random variables $\{X_n\}_{n\in\mathbf{N}}$ converges to another random variable X vaguely, if and only if $\lim_{n\to\infty} \mathbf{E}[\Phi(X_n)] = \mathbf{E}[\Phi(X)]$ holds for any bounded, continuous function $\Phi : \mathbf{R} \to \mathbf{R}$ with compact support.

4.2 Exercise: Consider a sequence $\xi_1, \xi_2, \cdots$ of identically distributed random variables, and let $X_n := \max_{1\le i\le n} |\xi_i|$, $S_n = \sum_{i=1}^n \xi_i$, $n \in \mathbf{N}$. Show that

(i) $\lim_{x\to\infty} x\,\mathbf{P}(|\xi_i| > x) = 0 \implies (X_n / n) \to 0$ in probability, as $n \to \infty$.

(ii) The reverse implication in (i) is also true, if the variables $\xi_1, \xi_2, \cdots$ are independent.

(iii) If $\mathbf{E}(|\xi_1|) < \infty$, then $(\xi_n / n) \to 0$ and $(X_n / n) \to 0$, $\mathbf{P}-$a.e. as $n \to \infty$

(iv) If $0 < \mathbf{E}(|\xi_1|) < \infty$ and the variables $\xi_1, \xi_2, \cdots$ are independent, then we have $(X_n / S_n) \to 0$, $\mathbf{P}-$a.e.

4.3 Exercise : Markov's Strong Law of Large Numbers. Consider independent random variables $X_1, X_2, \cdots$ with common distribution that satisfies $\mathbf{E}(X_1^4) < \infty$. Use Theorem 4.1(iii) and the Čebyšev inequality, to provide an 'elementary' proof of the strong law of large numbers (3.3) in this case.

4.4 Exercise: If $X_1, X_2, \cdots$ are pairwise uncorrelated random variables and $K := \sup_{n\in\mathbf{N}} \mathbf{E}(X_n^2) < \infty$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \big(X_j - \mathbf{E}(X_j)\big) = 0, \qquad \text{a.e.}$$

(*Hint:* Use Čebyšev, Borel-Cantelli and Theorem 4.1, to establish the result along the subsequence $k_n = n^2$; then argue that "nothing bad happens" between n^2 and $(n+1)^2$.)

4.5 Exercise: (a): Cantelli's Strong Law of Large Numbers. Consider independent random variables $X_1, X_2, \cdots$ with $\sup_{n\in\mathbf{N}} \mathbf{E}(X_n^4) < \infty$. Show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \big(X_j - \mathbf{E}(X_j)\big) = 0, \qquad \text{holds a.e.}$$

(*Hint:* Proceed as in the Hint for Exercise 4.4, working now with fourth instead of second powers. Note also that the full strength of independence is not necessary.)

(b): A strengthening of Borel-Cantelli. If $A_1, A_2, \cdots$ are pairwise-independent events with $\sum_{n\in\mathbf{N}} \mathbf{P}(A_n) = \infty$, then for

$$Q_n := \frac{\sum_{j=1}^n \chi_{A_j}}{\sum_{j=1}^n \mathbf{P}(A_j)} \quad \text{we have} \quad \lim_{n\to\infty} Q_n = 1 \quad \text{a.e.}$$

(*Hint:* Establish convergence in probability first; then show almost-everywhere convergence – first for the subsequence $n_k = \inf\{n \ge 1 : \sum_{j=1}^n \mathbf{P}(A_j) \ge k^2\}$, $k \in \mathbf{N}$, then for the entire sequence.)

4.6 Exercise: Let $X_1, X_2, \cdots$ be independent random variables with $\mathbf{P}(X_n = 1) = p_n \in (0,1)$ and $\mathbf{P}(X_n = 0) = 1 - p_n$. Show that $\lim_{n\to\infty} X_n = 0$ holds

- in probability, if and only if $\lim_{n\to\infty} p_n = 0$;
- a.e., if and only if $\sum_{n\in\mathbf{N}} p_n < \infty$.

4.7 Exercise: Let $X_1, X_2, \cdots$ be identically distributed (but not necessarily independent) random variables with $\mathbf{P}(X_1 \geq 0) = 1$ and $0 < \mathbf{E}(X_1) < \infty$. Define $M_n := \max_{1\leq j\leq n} X_j$, $S_n := \sum_{j=1}^n X_j$ for $n \in \mathbf{N}$, and show that:

(i) $\lim_{n\to\infty} (X_n/n) = 0$, a.e.

(ii) $\lim_{n\to\infty} (M_n/n) = 0$, a.e. as well as in $\mathbf{L}^1$.

(iii) $\lim_{n\to\infty} (M_n/S_n) = 0$ a.e., if $X_1, X_2, \cdots$ are independent.

4.8 Exercise: In the context and with the notation of Exercise 1.6.5, show that a sequence of probability measures $\{\mu_n\}_{n\in\mathbf{N}}$ on $\mathcal{B}(\mathbf{R})$ converges vaguely to the probability measure μ, if the corresponding potentials converge pointwise: $(\mathcal{P}\mu_n)(x) \longrightarrow (\mathcal{P}\mu)(x)$, $\forall\, x \in \mathbf{R}$.

(*Hint:* Argue by contradiction, and use the following result known as the *Helly-Bray* Lemma 3.2.1: For any sequence $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ of probability distribution functions, there exists a subsequence $\{F_{n_k}(\cdot)\}_{k\in\mathbf{N}}$ and a right-continuous, increasing function $F(\cdot)$, such that $F_{n_k}(x) \to F(x)$ as $k \to \infty$, at all continuity points x of $F(\cdot)$.)

4.9 Exercise: Let (Ω, ρ) be a metric space and μ, $\{\mu_n\}_{n\in N}$ be probability measures on the σ-algebra $\mathcal{F} = \sigma(\mathcal{O})$ generated by the open sets, such that $\{\mu_n\}$ *converges vaguely* to μ: namely,

$$\lim_{n\to\infty} \int_\Omega f\, d\mu_n = \int_\Omega f\, d\mu \qquad \text{holds for any bounded, continuous } f : \Omega \to \mathbf{R}.$$

(i) Show that we have: $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ for any closed subset C of Ω.

(ii) Suppose that $F : \Omega \to \mathbf{R}$ is bounded and lower-semicontinuous: $\limsup_{y\to x} F(y) \leq F(x)$, for all $x \in \Omega$. Show that

$$\limsup_{n\to\infty} \int_\Omega F\, d\mu_n \leq \int_\Omega F\, d\mu.$$

4.10 Exercise: Let $\{X_n\}_{n\in\mathbf{N}}$ be a sequence of random variables in $\mathbf{L}^p$ with $1 \leq p < \infty$. If $X_n \to X$ weakly in $\mathbf{L}^p$, show that: $\mathbf{E}(|X|) \leq \liminf_{n\to\infty} \mathbf{E}(|X_n|)$.

4.11 Exercise: Show that the definition of vague convergence is equivalent to the requirement $\lim_{n\to\infty} \mathbf{E}\big(\Phi(X_n)\big) = \mathbf{E}\big(\Phi(X)\big)$ for every bounded, continuous $\Phi : \mathbf{R} \to \mathbf{R}$ with compact support.

4.12 Exercise: Show that a sequence $\{X_n\}_{n\in\mathbf{N}}$ of random variables converges in probability, if and only if every subsequence $\{X_{n_k}\}_{k\in\mathbf{N}}$ contains a further sub-subsequence $\{X_{n_{k_j}}\}_{j\in\mathbf{N}}$ which converges a.e.

4.13 Exercise: A sequence of random variables converges in probability, if and only if it is Cauchy in probability.

4.14 Exercise: If a sequence $\{X_n\}_{n\in\mathbf{N}}$ of random variables is Cauchy in probability, it contains a subsequence $\{X_{n_k}\}_{k\in\mathbf{N}}$ which is Cauchy a.e.

4.15 Exercise: A sequence of random variables $\{X_n\}_{n\in\mathbf{N}}$ is Cauchy a.e., if and only if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}\big(\sup_{k\geq 0} |X_{n+k}-X_n| > \varepsilon\big) = 0\,, \quad \text{or equivalently} \quad \lim_{n\to\infty} \mathbf{P}\Big(\sup_{\ell\geq n,\, m\geq n} |X_\ell - X_m| > \varepsilon\Big) = 0\,.$$

4.16 Exercise: Show that there exists no metric on the space $\mathbf{L}^0$ of measurable, real-valued random variables, such that convergence in this metric is equivalent to convergence a.e.

4.17 Exercise: Suppose that the sequence of random variables $\{X_n\}_{n\in\mathbf{N}}$ is bounded in $\mathbf{L}^1$, that is $M := \sup_{n\in\mathbf{N}} \mathbf{E}(|X_n|) < \infty$; suppose also that *for each* $k \in \mathbf{N}$ the 'truncated' variables $Y_n^{(k)} := X_n \cdot \chi_{\{|X_n|\leq k\}}$, $n \in \mathbf{N}$ converge to some random variable Y_k weakly in $\mathbf{L}^p$, for some $1 \leq p < \infty$.

Show that there exists then a random variable Y such that $\lim_{k\to\infty} Y_k = Y$ both a.e. and in $\mathbf{L}^1$. (*Hint:* With the help of Exercise 4.10, argue that $\sum_{k\in\mathbf{N}} \mathbf{E}\big(|Y_{k+1}-Y_k|\big) < \infty$.)

B: A TECHNICAL RESULT

We shall close this section with a rather technical result. Apart from its intrinsic probabilistic interest, it will be used in a crucial manner when we prove the Komlós Theorem 1.8.5 in Chapter 4. The reader can safely skip or skim it at first reading.

4.1 Proposition: *Suppose* $\{\xi_n\}_{n\in\mathbf{N}}$ *is a sequence of random variables with* $M := \sup_{n\in\mathbf{N}} \mathbf{E}(|\xi_n|) < \infty$ *(bounded in* $\mathbf{L}^1$*). Show that such a sequence contains a subsequence* $\{\Xi_n\}_{n\in\mathbf{N}}$ *such that, for every further subsequence* $\{Z_n\}_{n\in\mathbf{N}}$ *of* $\{\Xi_n\}_{n\in\mathbf{N}}$ *and setting* $Y_n := Z_n \cdot \chi_{\{|Z_n|\leq k\}}$, $n \in \mathbf{N}$, *we have the following properties:*

$$\textit{The sequence } \{Y_n\}_{n\in\mathbf{N}} \textit{ is uniformly integrable;} \tag{4.8}$$

$$\sum_{n\in\mathbf{N}} \mathbf{P}(Y_n \neq Z_n) \,<\, \infty\,; \;\; \textit{and} \tag{4.9}$$

$$\sum_{n\in\mathbf{N}} \frac{1}{n^{1+\varepsilon}}\, \mathbf{E}\left(|Y_n|^{1+\varepsilon}\right) \,<\, \infty \quad \textit{holds for every } \varepsilon > 0\,. \tag{4.10}$$

Proof: Because the sequence $\left(\xi_n \cdot \chi_{\{|\xi_n|\le k\}}\right)_{n\in\mathbf{N}}$ is bounded, there exists a subsequence $\{\xi_n^{(1)}\}_{n\in\mathbf{N}}$ of $\{\xi_n\}_{n\in\mathbf{N}} =: \{\xi_n^{(0)}\}_{n\in\mathbf{N}}$ that converges to some η_1 weakly in $\mathbf{L}^2$. Proceeding inductively, we see that for evey integer $k \ge 1$ there exists a subsequence $\{\xi_n^{(k)}\}_{n\in\mathbf{N}}$ of $\{\xi_n^{(k-1)}\}_{n\in\mathbf{N}} \subset \{\xi_n\}_{n\in\mathbf{N}}$ such that $\left(\xi_n^{(k)} \cdot \chi_{\{|\xi_n^{(k)}|\le k\}}\right)_{n\in\mathbf{N}}$ converges to some random variable η_k weakly in $\mathbf{L}^2$.

Consider the *diagonal sequence* $X_n := \xi_n^{(n)}$, $n \in \mathbf{N}$ and observe that for every $k \in \mathbf{N}$ the sequence

$$\left(X_n \cdot \chi_{\{|X_n|\le k\}}\right)_{n\in\mathbf{N}} \quad \text{converges to the random variable } \eta_k \text{ weakly in } \mathbf{L}^2\,. \tag{4.11}$$

• Now let us look at the bounded sequence of real numbers $\{\,\mathbf{P}(k-1 \le |X_n| < k)\}_{k\in\mathbf{N}}$; proceeding inductively once again, we construct for each integer $k \ge 1$ a subsequence $\{X_n^{(k)}\}_{n\in\mathbf{N}}$ of $\{X_n^{(k-1)}\}_{n\in\mathbf{N}} \subset \{X_n\}_{n\in\mathbf{N}} =: \{X_n^{(0)}\}_{n\in\mathbf{N}}$ such that

$$p_k \,:=\, \lim_{n\to\infty} \mathbf{P}\left(k-1 \le |X_n^{(k)}| < k\right) \,\in\, [0,1]$$

exists and satisfies

$$\frac{1}{2}\cdot p_k \,\le\, \mathbf{P}\left(k-1 \le |X_n^{(k)}| < k\right) \,\le\, p_k + \frac{1}{k^3}\,, \quad \forall\, n \in \mathbf{N}\,.$$

Diagonalizing once again, we set

$$\Xi_n := X_n^{(n^2)}\,, \quad n \in \mathbf{N} \tag{4.12}$$

and observe that for this new sequence $\{\,\Xi_n\,\}_{n\in\mathbf{N}}$ and all its subsequences we have

$$\lim_{n\to\infty} \mathbf{P}\left(k-1 \le |\Xi_n| < k\right) \,=\, p_k$$

and

$$\frac{1}{2}\cdot p_k \,\le\, \mathbf{P}\left(k-1 \le |\Xi_n| < k\right) \,\le\, p_k + \frac{1}{k^3}\,, \quad \forall\, n \in \mathbf{N}\,,\; n^2 > k\,. \tag{4.13}$$

From $\sum_{k=1}^{\infty} \mathbf{P}\left(k-1 \le |\Xi_n| < k\right) = 1$ and (4.13) we see $\sum_{k=1}^{n^2} p_k \le 2$ for all $n \in \mathbf{N}$, thus $\sum_{k=1}^{\infty} p_k \le 2$. By the same token, the comparisons

$$\sum_{k=1}^{\infty} (k-1)\cdot \mathbf{P}\left(k-1 \le |\Xi_n| < k\right) \,\le\, \mathbf{E}(|\Xi_n|) \,\le M < \infty$$

and

$$\sum_{k=1}^{\infty} k\cdot \mathbf{P}\left(k-1 \le |\Xi_n| < k\right) \,\le\, 1 + \mathbf{E}(|\Xi_n|) \,\le M + 1$$

imply $\sum_{k=1}^{\infty} k\, p_k \le 2(M+1) < \infty$.

- Finally, we take an *arbitrary* subsequence $\{Z_n\}_{n\in\mathbf{N}}$ of $\{\Xi_n\}_{n\in\mathbf{N}}$ and define

$$Y_n \;:=\; Z_n \cdot \chi_{\{|Z_n|\le n\}}\,, \quad n \in \mathbf{N}\,. \tag{4.14}$$

It is then seen that this sequence satisfies the required properties. $\diamond$

4.18 Exercise: Show that the sequence $\mathcal{Y}$ of (4.14) satisfies the properties (4.8)-(4.10).