7

An Outlook on Further Topics

Probability theory is, of course, much more than what one will find in this book (so far). In this chapter we provide an outlook on some extensions and further areas and concepts in probability theory. For more we refer to the more advanced literature cited in Appendix A.

We begin, in the first section, by presenting some extensions of the classical limit theorems, that is, the law of large numbers and the central limit theorem, to cases where one relaxes the assumptions of independence and equidistribution.

Another question in this context is whether there exist (other) limit distributions if the variance of the summands does not exist (is infinite). This leads, in the case of i.i.d. summands, to the class of stable distributions and their, what is called, *domains of attraction*. Sections 2 and 3 are devoted to this problem.

In connection with the convergence concepts in Section 6.3, it was mentioned that convergence in r-mean was, in general, not implied by the other convergence concepts. In Section 4 we define *uniform integrability*, which is the precise condition one needs in order to assure that moments converge whenever convergence almost surely, in probability, or in distribution holds. As a pleasant illustration we prove Stirling's formula with the aid of the exponential distribution.

There exists an abundance of situations where *extremes* rather than sums are relevant; earthquakes, floods, storms, and many others. Analogous to "limit theory for sums" there exists a "limit theory for extremes," that is for $Y_n = \max\{X_1, X_2, \ldots, X_n\}$, $n \geq 1$, where (in our case) $X_1, X_2, \ldots, X_n$ are i.i.d. random variables. Section 5 provides an introduction to the what is called *extreme value theory*. We also mention the closely related *records*, which are extremes at first appearance.

Section 7 introduces the Borel–Cantelli lemmas, which are a useful tool for studying the limit superior and limit inferior of sequences of events, and, as an extension, in order to decide whether some special event will occur infinitely many times or not. As a toy example we prove the intuitively obvious fact that

A. Gut, *An Intermediate course in Probabilty,* Springer Texts in Statistics,
DOI: 10.1007/978-1-4419-0162-0_7,

if one tosses a coin an infinite number of times there will appear infinitely many heads and infinitely many tails. For a fair coin this is trivial due to symmetry, but what about an unfair coin? We also revisit Examples 6.3.1 and 6.3.2, and introduce the concept of *complete convergence*.

The final section, preceding some problems for solution, serves as an introduction to one of the most central tools in probability theory and the theory of stochastic processes, namely the theory of *martingales*, which, as a very rough definition, may be thought of as an extension of the theory of sums of independent random variables with mean zero and of fair games. In order to fully appreciate the theory one needs to base it on measure theory. Nevertheless, the basic flavor of the topic can be understood with our more elementary approach.

1 Extensions of the Main Limit Theorems

Several generalizations of the central limit theorem seem natural, such as:

1. the summands have (somewhat) different distributions;
2. the summands are *not independent*;
3. the variance does not exist.

In the first two subsections we provide some hints on the law of large numbers and the central limit theorem for the case of independent but not identically distributed summands. In the third subsection a few comments are given in the case of dependent summands. Possible (other) limit theorems when the variance is infinite (does not exist) is a separate issue, to which we return in Sections 2 and 3 for a short introduction.

1.1 The Law of Large Numbers: The Non-i-i.d. Case

It is intuitively reasonable to expect that the law of large numbers remains valid if the summands have different distributions—within limits.

We begin by presenting two extensions of this result.

Theorem 1.1. *Let* $X_1, X_2, \ldots$ *be independent random variables with* $E\,X_k = \mu_k$ *and* $\operatorname{Var} X_k = \sigma_k^2$, *and suppose that*

$$\frac{1}{n}\sum_{k=1}^{n}\mu_k \to \mu \quad \text{and that} \quad \frac{1}{n}\sum_{k=1}^{n}\sigma_k^2 \to \sigma^2 \quad \text{as} \quad n \to \infty,$$

(where $|\mu| < \infty$ *and* $\sigma^2 < \infty$*). Then*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu \quad \text{as} \quad n \to \infty.$$

Proof. Set $S_n = \sum_{k=1}^n X_k$, $m_n = \sum_{k=1}^n \mu_k$, and $s_n^2 = \sum_{k=1}^n \sigma_k^2$, and let $\varepsilon > 0$. By Chebyshev's inequality we then have

$$P\Big(\Big|\frac{S_n - m_n}{n}\Big| > \varepsilon\Big) \le \frac{s_n^2}{n^2\varepsilon^2} \to 0 \quad \text{as} \quad n \to \infty\,,$$

which tells us that

$$\frac{S_n - m_n}{n} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

which implies that

$$\frac{S_n}{n} = \frac{S_n - m_n}{n} + \frac{m_n}{n} \xrightarrow{p} 0 + \mu = \mu \quad \text{as} \quad n \to \infty$$

via Theorem 6.6.2. □

The next result is an example of *the law of large numbers for weighted sums.*

Theorem 1.2. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with finite mean μ, and let $\{(a_{nk}, 1 \le k \le n), n \ge 1\}$ be "weights," that is, suppose that $a_{nk} \ge 0$ and $\sum_{k=1}^n a_{nk} = 1$ for $n = 1, 2, \ldots$. Suppose, in addition, that*

$$\max_{1\le k\le n} a_{nk} \le \frac{C}{n} \quad \textit{for all} \quad n,$$

for some positive constant C, and set

$$S_n = \sum_{k=1}^n a_{nk} X_k\,, \quad n = 1, 2, \ldots.$$

Then

$$S_n \xrightarrow{p} \mu \qquad \textit{as} \quad n \to \infty.$$

Proof. The proof follows very much the lines of the previous one. We first note that

$$E\,S_n = \mu \sum_{k=1}^n a_{nk} = \mu \quad \text{and that} \quad \operatorname{Var} S_n = \sigma^2 \sum_{k=1}^n a_{nk}^2 = \sigma^2 A_n\,,$$

where thus

$$A_n = \sum_{k=1}^n a_{nk}^2 \le \max_{1\le k\le n} a_{nk} \sum_{k=1}^n a_{nk} \le \frac{C}{n} \cdot 1 = \frac{C}{n}.$$

By Chebyshev's inequality we now obtain

$$P\big(|S_n - \mu| > \varepsilon\big) \le \frac{\operatorname{Var} S_n}{\varepsilon^2} = \frac{\sigma^2 A_n}{\varepsilon^2} \le \sigma^2 \frac{C}{n} \to 0 \quad \text{as} \quad n \to \infty\,,$$

and the conclusion follows. □

1.2 The Central Limit Theorem: The Non-i-i.d. Case

An important criterion pertaining to the central limit theorem is the *Lyapounov condition*. It should be said, however, that more than finite variance is necessary in order for the condition to apply. This is the price one pays for relaxing the assumption of equidistribution. For the proof we refer to the literature cited in Appendix A.

Theorem 1.3. *Suppose that $X_1, X_2, \ldots$ are independent random variables, set, for $k \geq 1$, $\mu_k = E\,X_k$ and $\sigma_k^2 = \operatorname{Var} X_k$, and suppose that $E|X_k|^r < \infty$ for all k and some $r > 2$. If*

$$\beta(n,r) = \frac{\sum_{k=1}^{n} E|X_k - \mu_k|^r}{\big(\sum_{k=1}^{n} \sigma_k^2\big)^{r/2}} \to 0 \qquad \text{as} \quad n \to \infty, \tag{1.1}$$

then

$$\frac{\sum_{k=1}^{n}(X_k - \mu_k)}{\sqrt{\sum_{k=1}^{n} \sigma_k^2}} \xrightarrow{d} N(0,1) \qquad \text{as} \quad n \to \infty. \qquad \square$$

If, in particular, $X_1, X_2, \ldots$ are identically distributed and, for simplicity, with mean zero, then Lyapounov's condition turns into

$$\beta(n,r) = \frac{nE|X_1|^r}{(n\sigma^2)^{r/2}} = \frac{E|X_1|^r}{\sigma^r} \cdot \frac{1}{n^{1-r/2}} \to 0 \qquad \text{as} \quad n \to \infty, \tag{1.2}$$

which proves the central limit theorem under this slightly stronger assumption.

1.3 Sums of Dependent Random Variables

There exist many notions of dependence. One of the first things one learns in probability theory is that the outcomes of repeated drawings of balls *with replacement* from an urn of balls with different colors are independent, whereas the drawings *without replacement* are not. *Markov dependence* means, vaguely speaking, that the future of a process depends on the past only through the present. Another important dependence concept is *martingale dependence*, which is the topic of Section 8. Generally speaking, a typical dependence concept is defined via some kind of decay, in the sense that the further two elements are apart in time or index, the weaker is the dependence.

A simple such concept is m-dependence.

Definition 1.1. *The sequence $X_1, X_2, \ldots$ is* m-dependent *if X_i and X_j are independent whenever $|i - j| > m$.* $\square$

Remark 1.1. Independence is the same as 0-dependence.[1]

[1] In Swedish this looks fancier: "Oberoende" is the same as "0-beroende."

Example 1.1. $Y_1, Y_2, \dots$ be i.i.d. random variables, and set

$$X_1 = Y_1 \cdot Y_2, \quad X_2 = Y_2 \cdot Y_3, \quad \dots, \quad X_k = Y_k \cdot Y_{k+1}, \quad \dots .$$

The sequence $X_1, X_2, \dots$ clearly is a 1-dependent sequence; neighboring X variables are dependent, but X_i and X_j with $|i-j| > 1$ are independent. □

A common example of m-dependent sequences are the so-called $(m+1)$-block factors defined by

$$X_n = g(Y_n, Y_{n+1}, \dots, Y_{n+m-1}, Y_{n+m}), \quad n \geq 1,$$

where $Y_1, Y_2, \dots$ are independent random variables, and $g : \mathbb{R}^{m+1} \to \mathbb{R}$. Note that our example is a 2-block factor with $g(y_1, y_2) = y_1 \cdot y_2$.

The law of large numbers and the central limit theorem are both valid in this setting. Following is the law of large numbers.

Theorem 1.4. *Suppose that $X_1, X_2, \dots$ is a sequence of m-dependent random variables with finite mean μ and set $S_n = \sum_{k=1}^{n} X_k$, $n \geq 1$. Then*

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

Proof. For simplicity we confine ourselves to proving the theorem for $m = 1$. We then separate S_n into the sums over the odd and even summands, respectively.

Since the even as well as the odd summands are independent, the law of large numbers for independent summands, Theorem 6.5.1 tells us that

$$\frac{\sum_{k=1}^{m} X_{2k}}{m} \xrightarrow{p} \mu \quad \text{and} \quad \frac{\sum_{k=1}^{m} X_{2k-1}}{m} \xrightarrow{p} \mu \quad \text{as} \quad m \to \infty,$$

so that an application of Theorem 6.6.2 yields

$$\frac{S_{2m}}{2m} = \frac{1}{2}\frac{\sum_{k=1}^{m} X_{2k-1}}{m} + \frac{1}{2}\frac{\sum_{k=1}^{m} X_{2k}}{m} \xrightarrow{p} \frac{1}{2}\mu + \frac{1}{2}\mu = \mu \quad \text{as} \quad m \to \infty,$$

when $n = 2m$ is even. For $n = 2m+1$ odd we similarly obtain

$$\begin{aligned} \frac{S_{2m+1}}{2m+1} &= \frac{m+1}{2m+1} \cdot \frac{\sum_{k=1}^{m+1} X_{2k-1}}{m+1} + \frac{m}{2m+1} \cdot \frac{\sum_{k=1}^{m} X_{2k}}{m} \\ &\xrightarrow{p} \frac{1}{2}\mu + \frac{1}{2}\mu = \mu \quad \text{as} \quad m \to \infty, \end{aligned}$$

which finishes the proof. □

Exercise 1.1. Complete the proof of the theorem for general m. □

In the m-dependent case the dependence stops abruptly. A natural generalization would be to allow the dependence to drop gradually. This introduces the concept of *mixing*. There are variations with different names. We refer to the more advanced literature for details.

2 Stable Distributions

Let X, X_1, X_2, ... be i.i.d. random variables with partial sums S_n, $n \geq 1$. The law of large numbers states that $S_n/n \stackrel{p}{\to} \mu$ as $n \to \infty$ if the mean μ is finite. The central limit theorem states that $(S_n - n\mu)/(\sigma\sqrt{n}) \stackrel{d}{\to} N(0,1)$ as $n \to \infty$, provided the mean μ and the variance σ^2 exist. A natural question is whether there exists something "in between," that is, can we obtain some (other) limit by normalizing with n to some other power than 1 or 1/2? In this section and the next one we provide a glimpse into more general limit theorems for sums of i.i.d. random variables.

Before addressing the question just raised, here is another observation. If, in particular, we assume that the random variables are $C(0,1)$-distributed, then we recall from Remark 6.5.2 that, for any $n \geq 1$,

$$\varphi_{\frac{S_n}{n}}(t) = \Big(\varphi_X\Big(\frac{t}{n}\Big)\Big)^n = \Big(e^{-|t/n|}\Big)^n = e^{-|t|} = \varphi_X(t),$$

that

$$\frac{S_n}{n} \stackrel{d}{=} X \quad \text{for all} \quad n,$$

and, hence, that law of large numbers does not hold, which was no contradiction, because the mean does not exist.

Now, if, instead the random variables are $N(0,\sigma^2)$-distributed, then the analogous computation shows that

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \Big(\varphi_X\Big(\frac{t}{\sqrt{n}}\Big)\Big)^n = \Big(\exp\Big\{-\frac{1}{2}\Big(\frac{t}{\sqrt{n}}\Big)^2\Big\}\Big)^n = e^{-t^2/2} = \varphi_X(t),$$

that is,

$$\frac{S_n}{\sqrt{n}} \stackrel{d}{=} X \quad \text{for all} \quad n,$$

in view of the uniqueness theorem for characteristic functions.

Returning to our question above it seems, with this in mind, reasonable to try a distribution whose characteristic function equals $\exp\{-|t|^\alpha\}$ for $\alpha > 0$ (provided this is really a characteristic function also when $\alpha \neq 1$ and $\neq 2$). By modifying the computations above we similarly find that

$$\frac{S_n}{n^{1/\alpha}} \stackrel{d}{=} X \quad \text{for all} \quad n, \tag{2.1}$$

where, thus, $\alpha = 1$ corresponds to the Cauchy distribution and $\alpha = 2$ to the normal distribution.

Distributions with a characteristic function of the form

$$\varphi(t) = e^{-c|t|^\alpha}, \quad \text{where } 0 < \alpha \leq 2 \text{ and } c > 0, \tag{2.2}$$

are called *symmetric stable*. However, φ as defined in (2.2) is *not* a characteristic function for any $\alpha > 2$.

The general definition of stable distributions, stated in terms of random variables is as follows.

Definition 2.1. *Let $X_1, X_2, \dots$ be i.i.d. random variables, and set $S_n = X_1 + X_2 + \cdots + X_n$. The distribution of the random variables is* stable (in the broad sense) *if there exist sequences $a_n > 0$ and b_n such that*

$$S_n \stackrel{d}{=} a_n X + b_n.$$

The distribution is strictly stable *if $b_n = 0$ for all n.* □

Remark 2.1. The stability pertains to the fact that the sum of any number of random variables has the same distribution as the individual summands themselves (after scaling and translation).

Remark 2.2. One can show that if X has a stable distribution, then, necessarily, $a_n = n^{1/\alpha}$ for some $\alpha > 0$, which means that our first attempt to investigate possible characteristic functions was exhaustive (except for symmetry) and that, once again, only the case $0 < \alpha \leq 2$ is possible. Moreover, α is called the *index*.

Exercise 2.1. Another fact is that if X has a stable distribution with index α, $0 < \alpha < 2$, then

$$E|X|^r \begin{cases} < \infty, & \text{for} \quad 0 < r < \alpha, \\ = \infty, & \text{for} \quad r \geq \alpha. \end{cases}$$

This implies, in particular, that the law of large numbers must hold for stable distributions with $\alpha > 1$. Prove directly via characteristic functions that this is the case. Recall also, from above, that the case $\alpha = 1$ corresponds to the Cauchy distribution for which the law of large numbers does not hold.

We close this section by mentioning that there exist characterizations in terms of characteristic functions for the general class of stable distributions (not just the symmetric ones), but that is beyond the present outlook.

3 Domains of Attraction

We now return to the question posed in the introduction of Section 2, namely whether there exist limit theorems "in between" the law of large numbers and the central limit theorem. With the previous section in mind it is natural to guess that the result is positive, that such results would be connected with the stable distributions, and that the variance is not necessarily assumed to exist.

In order to discuss this problem we introduce the notion of *domains of attraction*.

Definition 3.1. *Let X, X_1, X_2, ... be i.i.d. random variables with partial sums S_n, $n \geq 1$. We say that X, or, equivalently, the distribution F_X,* belongs to the domain of attraction of the (non-degenerate) distribution G *if there exist normalizing sequences $\{a_n > 0, n \geq 1\}$ and $\{b_n, n \geq 1\}$ such that*

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} G \quad \text{as} \quad n \to \infty.$$

The notation is $F_X \in \mathcal{D}(G)$; alternatively, $X \in \mathcal{D}(Z)$ if $Z \in G$. □

If $\operatorname{Var} X < \infty$, the central limit theorem tells us that X belongs to the domain of attraction of the normal distribution; choose $b_n = nE\,X$, and $a_n = \sqrt{n \operatorname{Var} X}$. In particular, the normal distribution belongs to its own domain of attraction. Recalling Section 2 we also note that the stable distributions belong to their own domain of attraction.

In fact, the stable distributions are the only possible limit distributions.

Theorem 3.1. *Only the stable distributions or random variables possess a domain of attraction.*

With this information the next problem of interest would be to exhibit criteria for a distribution to belong to the domain of attraction of some given (stable) distribution. In order to state such results we need some facts about what is called *regular* and *slow variation.*

Definition 3.2. *Let $a > 0$. A positive measurable function u on $[a, \infty)$* varies regularly *at infinity with exponent ρ, $-\infty < \rho < \infty$, denoted $u \in \mathcal{RV}(\rho)$, iff*

$$\frac{u(tx)}{u(t)} \to x^{\rho} \quad \text{as} \quad t \to \infty \quad \text{for all} \quad x > 0.$$

If $\rho = 0$ the function is slowly varying *at infinity; $u \in \mathcal{SV}$.* □

Typical examples of regularly varying functions are

$$x^{\rho}, \quad x^{\rho} \log^+ x, \quad x^{\rho} \log^+ \log^+ x, \quad x^{\rho} \frac{\log^+ x}{\log^+ \log^+ x}, \qquad \text{and so on.}$$

Typical slowly varying functions are the above when $\rho = 0$. Every positive function with a positive finite limit as $x \to \infty$ is slowly varying.

Exercise 3.1. Check that the typical functions behave as claimed.

Here is now the main theorem.

Theorem 3.2. *A random variable X with distribution function F belongs to the domain of attraction of a stable distribution iff there exists $L \in \mathcal{SV}$ such that*

$$U(x) = E\,X^2 I\{|X| \leq x\} \sim x^{2-\alpha} L(x) \quad \text{as} \quad x \to \infty, \tag{3.1}$$

and, moreover, for $\alpha \in (0,2)$, *that*

$$\frac{P(X > x)}{P(|X| > x)} \to p \quad \text{and} \quad \frac{P(X < -x)}{P(|X| > x)} \to 1 - p \quad \text{as} \quad x \to \infty. \tag{3.2}$$

By partial integration and properties of regularly varying functions one can show that (3.1) is equivalent to

$$\frac{x^2 P(|X| > x)}{U(x)} \to \frac{2-\alpha}{\alpha} \qquad \text{as} \quad x \to \infty, \qquad \text{for} \quad 0 < \alpha \le 2, \tag{3.3}$$

$$P(|X| > x) \sim \frac{2-\alpha}{\alpha} \cdot \frac{L(x)}{x^\alpha} \qquad \text{as} \quad x \to \infty, \qquad \text{for} \quad 0 < \alpha < 2, \tag{3.4}$$

which, in view of Theorem 3.1 yields the following alternative formulation of Theorem 3.2.

Theorem 3.3. *A random variable* X *with distribution function* F *belongs to the domain of attraction of*
(a) *the normal distribution iff* $U \in \mathcal{SV}$;
(b) *a stable distribution with index* $\alpha \in (0,2)$ *iff (3.4) and (3.2) hold.*

Let us, as a first illustration, look at the simplest example.

Example 3.1. Let $X, X_1, X_2, \dots$ be independent random variables with common density

$$f(x) = \begin{cases} \dfrac{1}{2x^2}, & \text{for} \quad |x| > 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the distribution is symmetric and that the mean is infinite.

Now, for $x > 1$,

$$P(X > x) = \frac{1}{2x}, \quad P(X < -x) = \frac{1}{2|x|}, \quad P(|X| > x) = \frac{1}{x}, \quad U(x) = x - 1,$$

so that (3.1)–(3.4) are satisfied ($p = 1/2$ and $L(x) = 1$).

Our second example is a boundary case in that the variance does not exist, but the asymptotic distribution is still the normal one.

Example 3.2. Suppose that $X, X_1, X_2, \dots$ are independent random variables with common density

$$f(x) = \begin{cases} \dfrac{1}{|x|^3}, & \text{for} \quad |x| > 1, \\ 0, & \text{otherwise.} \end{cases}$$

The distribution is symmetric again, the mean is finite and the variance is infinite—$\int_1^\infty (x^2/x^3)\,dx = +\infty$. As for (3.1) we find that

$$U(x) = \int_{|y|\leq x} y^2 f(y)\,dy = 2\int_1^x \frac{dy}{y} = 2\log x,$$

so that $U \in \mathcal{SV}$ as $x \to \infty$, that is, X belongs to the domain of attraction of the normal distribution.

This means that, for a suitable choice of normalizing constants $\{a_n,\ n \geq 1\}$ (no centering because of symmetry), we have

$$\frac{S_n}{a_n} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty.$$

More precisely, omitting all details, we just mention that one can show that, in fact,

$$\frac{S_n}{\sqrt{n \log n}} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty.$$

Remark 3.1. The object of Problem 6.8.50 was to prove this result with the aid of characteristic functions, that is, directly, without using the theory of domains of attraction.

4 Uniform Integrability

We found in Section 6.3 that convergence in probability does not necessarily imply convergence of moments. A natural question is whether there exists some condition that guarantees that a sequence that converges in probability (or almost surely or in distribution) also converges in r-mean. It turns out that *uniform integrability* is the adequate concept for this problem.

Definition 4.1. *A sequence* $X_1, X_2, \ldots$ *is called* uniformly integrable *if*

$$E|X_n|I\{|X_n| > a\} \to 0 \qquad \textit{as} \quad a \to \infty \qquad \text{uniformly } \textit{in} \quad n. \qquad \square$$

Remark 4.1. If, for example, all distributions involved are continuous, this is the same as

$$\int_{|x|>a} |x| f_{X_n}(x)\,dx \to 0 \quad \text{as} \quad a \to \infty \quad \text{uniformly in} \quad n. \qquad \square$$

The following result shows why uniform integrability is the correct concept. For a proof and much more on uniform integrability, we refer to the literature cited in Appendix A.

Theorem 4.1. *Let* $X, X_1, X_2, \ldots$ *be random variables such that* $X_n \xrightarrow{p} X$ *as* $n \to \infty$. *Let* $r > 0$, *and suppose that* $E|X_n|^r < \infty$ *for all* n. *The following are equivalent:*

(a) $\{|X_n|^r,\ n \geq 1\}$ *is uniformly integrable;*
(b) $X_n \xrightarrow{r} X$ *as* $n \to \infty$*;*
(c) $E|X_n|^r \to E|X|^r$ *as* $n \to \infty$. $\square$

The immediate application of the theorem is manifested in the following exercise.

Exercise 4.1. Show that if $X_n \xrightarrow{p} X$ as $n \to \infty$ and $X_1, X_2, \ldots$ is uniformly integrable, then $E\,X_n \to E\,X$ as $n \to \infty$. □

Example 4.1. A uniformly bounded sequence of random variables is uniformly integrable. Technically, if the random variables $X_1, X_2, \ldots$ are uniformly bounded, there exists some constant $A > 0$ such that $P(|X_n| \le A) = 1$ for all n. This implies that the expectation in the definition, in fact, *equals* zero as soon as $a > A$.

Example 4.2. In Example 6.3.1 we found that X_n converges in probability as $n \to \infty$ and that X_n converges in r-mean as $n \to \infty$ when $r < 1$ but not when $r \ge 1$. In view of Theorem 4.1 it must follow that $\{|X_n|^r, n \ge 1\}$ is uniformly integrable when $r < 1$ but not when $r \ge 1$.

Indeed, it follows from the definition that (for $a > 1$)

$$E|X_n|^r I\{|X_n| > a\} = n^r \cdot \frac{1}{n} \cdot I\{a < n\} \to 0 \quad \text{as} \quad a \to \infty$$

uniformly in n iff $r < 1$, which verifies the desired conclusion. □

Exercise 4.2. State and prove an analogous statement for Example 6.3.2.

Exercise 4.3. Consider the following modification of Example 6.3.1. Let $X_1, X_2, \ldots$ be random variables such that

$$P(X_n = 1) = 1 - \frac{1}{n} \quad \text{and} \quad P(X_n = 1000) = \frac{1}{n}, \quad n \ge 2.$$

Show that $X_n \xrightarrow{p} 1$ as $n \to \infty$, that $\{|X_n|^r, n \ge 1\}$ is uniformly integrable for all $r > 0$, and hence that $X_n \xrightarrow{r} 1$ as $n \to \infty$ for all $r > 0$. □

Remark 4.2. Since $X_1, X_2, \ldots$ are uniformly bounded, the latter part follows immediately from Example 4.1, but it is instructive to verify the conclusion directly via the definition.

Note also that the difference between Exercise 4.3 and Example 6.3.1 is that there the "rare" value n drifts off to infinity, whereas here it is a fixed constant (1000). □

It is frequently difficult to verify uniform integrability of a sequence directly. The following result provides a convenient sufficient criterion.

Theorem 4.2. *Let $X_1, X_2, \ldots$ be random variables, and suppose that*

$$\sup_n E|X_n|^r < \infty \quad \textit{for some} \quad r > 1.$$

Then $\{X_n, n \ge 1\}$ is uniformly integrable. In particular, this is the case if $\{|X_n|^r, n \ge 1\}$ is uniformly integrable for some $r > 1$.

Proof. We have

$$E|X_n|I\{|X_n|>a\} \le a^{1-r}E|X_n|^r I\{|X_n|>a\} \le a^{1-r}E|X_n|^r \\ \le a^{1-r}\sup_n E|X_n|^r \to 0 \quad \text{as} \quad a\to\infty,$$

independently, hence uniformly, in n.

The particular case is immediate since more is assumed. □

Remark 4.3. The typical case is when one wishes to prove convergence of the sequence of expected values and knows that the sequence of variances is uniformly bounded. □

We close this section with an illustration of how one can prove Stirling's formula via the central limit theorem with the aid of the exponential distribution and Theorems 4.1 and 4.2.

Example 4.3. Let $X_1, X_2, \ldots$ be independent Exp(1)-distributed random variables, and set $S_n = \sum_{k=1}^n X_k$, $n \ge 1$. From the central limit theorem we know that

$$\frac{S_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as} \quad n\to\infty,$$

and, since, for example, the variances of the normalized partial sums are equal to 1 for all n (so that the second moments are uniformly bounded), it follows from Theorems 4.2 and 4.1 that

$$\lim_{n\to\infty} E\left|\frac{S_n-n}{\sqrt{n}}\right| = E|N(0,1)| = \sqrt{\frac{2}{\pi}}. \tag{4.1}$$

Since we know that $S_n \in \Gamma(n,1)$ the expectation can be spelled out exactly and we can rewrite (4.1) as

$$\int_0^\infty \left|\frac{x-n}{\sqrt{n}}\right| \frac{1}{\Gamma(n)} x^{n-1}e^{-x}\,dx \to \sqrt{\frac{2}{\pi}} \quad \text{as} \quad n\to\infty. \tag{4.2}$$

By splitting the integral at $x=n$, and making the change of variable $u = x/n$ one arrives after some additional computations at the relation

$$\lim_{n\to\infty} \frac{\left(\frac{n}{e}\right)^n\sqrt{2n\pi}}{n!} = 1\,,$$

which is—Stirling's formula. □

Exercise 4.4. Carry out the details of the program. □

5 An Introduction to Extreme Value Theory

Suppose that $X_1, X_2, \ldots$ is a sequence of i.i.d. distributed random variables. What are the possible limit distributions of the normalized partial sums? If the variance is finite the answer is the normal distribution in view of the central limit theorem. In the general case, we found in Section 3 that the possible limit distributions are the stable distributions.

This section is devoted to the analogous problem for extremes. Thus, let, for $n \geq 1$,

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}.$$

What are the possible limit distributions of Y_n, after suitable normalization, as $n \to \infty$?

The following definition is the analog of Definition 3.1 (which concerned sums) for extremes.

Definition 5.1. *Let $X, X_1, X_2, \ldots$ be i.i.d. random variables, and set $Y_n = \max_{1 \leq k \leq n} X_k$, $n \geq 1$. We say that X, or, equivalently, the distribution function F_X,* belongs to the domain of attraction of the extremal distribution G *if there exist normalizing sequences $\{a_n > 0, n \geq 1\}$ and $\{b_n, n \geq 1\}$, such that*

$$\frac{Y_n - b_n}{a_n} \xrightarrow{d} G \quad \textit{as} \quad n \to \infty. \qquad \square$$

Example 5.1. Let $X_1, X_2, \ldots$ be independent Exp(1)-distributed random variables, and set $Y_n = \max\{X_1, X_2, \ldots, X_n\}$, $n \geq 1$. Then,

$$F(x) = 1 - e^{-x} \quad \text{for} \quad x > 0,$$

(and 0 otherwise), so that

$$P(Y_n \leq x) = \big(1 - e^{-x}\big)^n.$$

Aiming at something like $(1 - u/n)^n \to e^u$ as $n \to \infty$ suggests that we try $a_n = 1$ and $b_n = \log n$ to obtain

$$\begin{aligned} F_{Y_n - \log n}(x) &= P(Y_n \leq x + \log n) = \big(1 - e^{-x - \log n}\big)^n \\ &= \Big(1 - \frac{e^{-x}}{n}\Big)^n \to e^{-e^{-x}} \quad \text{as} \quad n \to \infty, \end{aligned}$$

for all $x \in \mathbb{R}$.

Example 5.2. Let $X_1, X_2, \ldots$ be independent Pa(β, α)-distributed random variables, and set $Y_n = \max\{X_1, X_2, \ldots, X_n\}$, $n \geq 1$. Then,

$$F(x) = \int_\beta^x \frac{\alpha \beta^\alpha}{y^{\alpha+1}}\, dy = 1 - \Big(\frac{\beta}{x}\Big)^\alpha \quad \text{for} \quad x > \beta,$$

(and 0 otherwise), so that

$$P(Y_n \le x) = \Big(1 - \Big(\frac{\beta}{x}\Big)^{\alpha}\Big)^n.$$

An inspection of this relation suggests the normalization $a_n = n^{1/\alpha}$ and $b_n = 0$, which, for $x > 0$ and n large, yields

$$\begin{aligned} F_{n^{-1/\alpha}Y_n}(x) &= P(Y_n \le xn^{1/\alpha}) = \Big(1 - \Big(\frac{\beta}{xn^{1/\alpha}}\Big)^{\alpha}\Big)^n \\ &= \Big(1 - \frac{(\beta/x)^{\alpha}}{n}\Big)^n \to e^{-(\beta/x)^{\alpha}} \quad \text{as} \quad n \to \infty\,. \end{aligned}$$

Remark 5.1. For $\beta = 1$ the example reduces to Example 6.1.2.

Example 5.3. Let $X_1, X_2, \dots$ be independent $U(0,\theta)$-distributed random variables ($\theta > 0$), and set $Y_n = \max\{X_1, X_2, \dots, X_n\}$, $n \ge 1$. Thus, $F(x) = x/\theta$ for $x \in (0,\theta)$ and 0 otherwise, so that,

$$P(Y_n \le x) = \Big(\frac{x}{\theta}\Big)^n.$$

Now, since $Y_n \xrightarrow{p} \theta$ as $n \to \infty$ (this is intuitively "obvious," but check Problem 6.8.1), it is more convenient to study $\theta - Y_n$, viz.,

$$P(\theta - Y_n \le x) = P(Y_n \ge \theta - x) = 1 - \Big(1 - \frac{x}{\theta}\Big)^n.$$

The usual approach now suggests $a_n = 1/n$ and $b_n = \theta$. Using this we obtain, for any $x < 0$,

$$\begin{aligned} P(n(Y_n - \theta) \le x) &= P\Big(\theta - Y_n \ge \frac{(-x)}{n}\Big) = \Big(1 - \frac{(-x)}{\theta n}\Big)^n \\ &\to e^{-(-x)/\theta} \quad \text{as} \quad n \to \infty\,. \end{aligned}$$ □

Looking back at the examples we note that the limit distributions have different expressions and that their domains vary; they are $x > 0$, $x \in \mathbb{R}$, and $x < 0$, respectively. It seems that the possible limits may be of at least three kinds. The following result tells us that this is indeed the case. More precisely, there are exactly three so-called *types*, meaning those mentioned in the theorem below, together with linear transformations of them.

Theorem 5.1. *There exist three* types *of extremal distributions:*

Fréchet: $\Phi_\alpha(x) = \begin{cases} 0, & \textit{for} \quad x < 0, \\ \exp\{-x^{-\alpha}\}, & \textit{for} \quad x \ge 0, \end{cases} \qquad \alpha > 0;$

Weibull: $\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & \textit{for} \quad x < 0, \\ 1, & \textit{for} \quad x \ge 0, \end{cases} \qquad \alpha > 0;$

Gumbel: $\Lambda(x) = \exp\{-e^{-x}\}, \qquad \textit{for} \quad x \in \mathbb{R}.$

The proof is beyond the scope of this book, let us just mention that the so-called convergence of types theorem is a crucial ingredient.

Remark 5.2. Just as the normal and stable distributions belong to their own domain of attraction (recall relation (2.1) above), it is natural to expect that the three extreme value distributions of the theorem belong to their domain of attraction. This is more formally spelled out in Problem 9.10 below.

6 Records

Let $X, X_1, X_2, \ldots$ be i.i.d. continuous random variables. The *record times* are $L(1) = 1$ and, recursively,

$$L(n) = \min\{k : X_k > X_{L(n-1)}\}, \quad n \geq 2,$$

and the *record values* are

$$X_{L(n)}, \quad n \geq 1.$$

The associated *counting process* $\{\mu(n), n \geq 1\}$ is defined by

$$\mu(n) = \#\,\text{records among } X_1, X_2, \ldots, X_n = \max\{k : L(k) \leq n\}.$$

The reason for assuming continuity is that we wish to avoid ties.

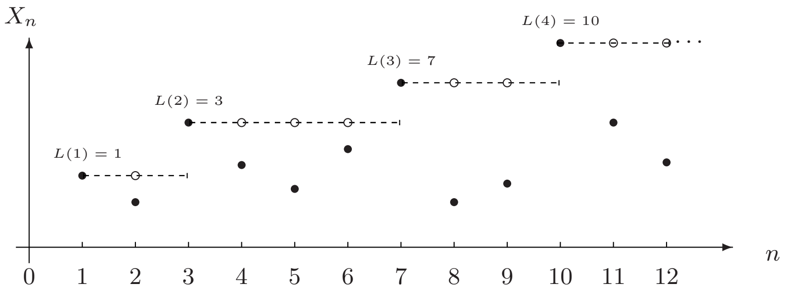

Fig. 7.1. Partial maxima ∘

Whereas the sequence of partial maxima, Y_n, $n \geq 1$, describe "the largest value so far," the record values pick these values the first time they appear. The sequence of record values thus constitutes a subsequence of the partial maxima. Otherwise put, the sequence of record values behaves like a compressed sequence of partial maxima, as is depicted in the above figure.

We begin by noticing that the record times and the number of records are distribution independent (under our continuity assumption). This is due to the fact that for a given random variable X with distribution function F, it follows that $F(X) \in U(0,1)$. This implies that there is a 1-to-1 map from every random variable to every other one, which preserves the record *times*, and therefore also the number of records—but not the record *values*.

Next, set

$$I_k = \begin{cases} 1, & \text{if } X_k \text{ is a record,} \\ 0, & \text{otherwise,} \end{cases}$$

so that $\mu(n) = \sum_{k=1}^{n} I_k$, $n \geq 1$.

By symmetry, all permutations between $X_1, X_2, \ldots, X_n$ are equally likely, from which we conclude that

$$P(I_k = 1) = 1 - P(I_k = 0) = \frac{1}{k}, \quad k = 1, 2, \ldots, n.$$

In addition one can show that the random variables $\{I_k,\, k \geq 1\}$ are independent. We collect these facts in the following result.

Theorem 6.1. *Let $X_1, X_2, \ldots, X_n$, $n \geq 1$, be i.i.d. continuous random variables. Then*

(a) *the indicators $I_1, I_2, \ldots, I_n$ are independent;*

(b) *$P(I_k = 1) = 1/k$ for $k = 1, 2, \ldots, n$.*

As a corollary it is now a simple task to compute the mean and the variance of $\mu(n)$ and their asymptotics.

Theorem 6.2. *Let $\gamma = 0.5772\ldots$ denote Euler's constant. We have*

$$m_n = E\,\mu(n) = \sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma + o(1) \quad \text{as} \quad n \to \infty;$$

$$\operatorname{Var} \mu(n) = \sum_{k=1}^{n} \frac{1}{k}\Big(1 - \frac{1}{k}\Big) = \log n + \gamma - \frac{\pi^2}{6} + o(1) \quad \text{as} \quad n \to \infty.$$

Proof. That $E\,\mu(n) = \sum_{k=1}^{n} \frac{1}{k}$, and that $\operatorname{Var} \mu(n) = \sum_{k=1}^{n} \frac{1}{k}(1 - \frac{1}{k})$, is clear. The remaining claims follow from the facts that

$$\sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma + o(1) \text{ as } n \to \infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \qquad \square$$

Next we present the weak laws of large numbers for the counting process.

Theorem 6.3. *We have*

$$\frac{\mu(n)}{\log n} \xrightarrow{p} 1 \quad \text{as} \quad n \to \infty.$$

Proof. Chebyshev's inequality together with Theorem 6.2 yields

$$P\Big(\frac{\mu(n) - E\,\mu(n)}{\operatorname{Var}(\mu(n))} > \varepsilon\Big) \leq \frac{1}{\varepsilon^2 \operatorname{Var}(\mu(n))} \to 0 \quad \text{as} \quad n \to \infty,$$

which tells us that

$$\frac{\mu(n) - E\,\mu(n)}{\mathrm{Var}\,(\mu(n))} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

Finally,

$$\frac{\mu(n)}{\log n} = \frac{\mu(n) - E\,\mu(n)}{\mathrm{Var}\,(\mu(n))} \cdot \frac{\mathrm{Var}\,(\mu(n))}{\log n} + \frac{E\,\mu(n)}{\log n} \xrightarrow{p} 0 \cdot 1 + 1 = 1 \quad \text{as} \quad n \to \infty,$$

in view of Theorem 6.2 (and Theorem 6.6.2). □

The central limit theorem for the counting process runs as follows.

Theorem 6.4. *We have*

$$\frac{\mu(n) - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1) \quad \textit{as} \quad n \to \infty.$$

Proof. We check the Lyapounov condition (1.1) with $r = 3$:

$$\begin{aligned} E|I_k - E\,I_k|^3 &= \left|0 - \frac{1}{k}\right|^3 \cdot \left(1 - \frac{1}{k}\right) + \left|1 - \frac{1}{k}\right|^3 \frac{1}{k} \\ &= \left(1 - \frac{1}{k}\right)\frac{1}{k} \cdot \left(\frac{1}{k^2} + \left(1 - \frac{1}{k}\right)^2\right) \leq 2\left(1 - \frac{1}{k}\right)\frac{1}{k}, \end{aligned}$$

so that

$$\begin{aligned} \beta(n,3) &= \frac{\sum_{k=1}^n E|X_k - \mu_k|^3}{\left(\sum_{k=1}^n \sigma_k^2\right)^{3/2}} \leq 2\frac{\sum_{k=1}^n \left(1 - \frac{1}{k}\right)\frac{1}{k}}{\left(\sum_{k=1}^n \left(1 - \frac{1}{k}\right)\frac{1}{k}\right)^{3/2}} \\ &= 2\left(\sum_{k=1}^n \left(1 - \frac{1}{k}\right)\frac{1}{k}\right)^{-1/2} \to 0 \quad \text{as} \quad n \to \infty, \end{aligned}$$

since

$$\sum_{k=1}^n \left(1 - \frac{1}{k}\right)\frac{1}{k} \geq \frac{1}{2}\sum_{k=2}^n \frac{1}{k} \to \infty \quad \text{as} \quad n \to \infty.$$ □

Exercise 6.1. Another way to prove this is via characteristic functions or moment generating functions; note, in particular, that $|I_k - \frac{1}{k}| \leq 1$ for all k. □

The analogous results for record times state that

$$\frac{\log L(n)}{n} \xrightarrow{p} 1 \quad \text{as} \quad n \to \infty,$$

$$\frac{\log L(n) - n}{\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty.$$

In the opening of this section we found that the *record values*, $\{X_{L(n)},\, n \geq 1\}$, seemed to behave like a compressed sequence of partial maxima, which makes it reasonable to believe that there exist three possible limit distributions for $X_{L(n)}$ as $n \to \infty$, which are somehow connected with the the three limit theorems for extremes. The following theorem shows that this is, indeed, the case.

Theorem 6.5. *Suppose that F is absolutely continuous. The possible* types *of limit distributions for record values are*

$$\Phi(-\log(-\log G(x))),$$

where G is an extremal distribution and Φ the distribution function of the standard normal distribution. More precisely, the three classes or types of limit distributions are

$$\Phi_\alpha^{(R)}(x) = \begin{cases} 0, & \text{for} \quad x < 0, \\ \Phi(\alpha \log x), & \text{for} \quad x \geq 0, \end{cases} \qquad \alpha > 0;$$

$$\Psi_\alpha^{(R)}(x) = \begin{cases} \Phi(-\alpha \log(-x)), & \text{for} \quad x < 0, \\ 1, & \text{for} \quad x \geq 0, \end{cases} \qquad \alpha > 0;$$

$$\Lambda^{(R)}(x) = \Phi(x), \qquad \text{for} \quad x \in \mathbb{R}.$$

7 The Borel–Cantelli Lemmas

The aim of this section is to provide some additional material on a.s. convergence. Although the reader cannot be expected to appreciate the concept fully at this level, we add here some additional facts and properties to shed somewhat light on it. The main results or tools are the Borel–Cantelli lemmas. We begin, however, with the following definition:

Definition 7.1. *Let $\{A_n,\ n \geq 1\}$ be a sequence of events (subsets of Ω). We define*

$$A_* = \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m,$$

$$A^* = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m. \qquad \square$$

Thus, if $\omega \in \Omega$ belongs to the set $\liminf_{n\to\infty} A_n$, then ω belongs to $\bigcap_{m=n}^{\infty} A_m$ for some n, that is, there exists an n such that $\omega \in A_m$ for *all* $m \geq n$. In particular, if A_n is the event that something special occurs at "time" n, then $\liminf_{n\to\infty} A_n^c$ means that from some n onward this property never occurs.

Similarly, if $\omega \in \Omega$ belongs to the set $\limsup_{n\to\infty} A_n$, then ω belongs to $\bigcup_{m=n}^{\infty} A_m$ for every n, that is, no matter how large we choose m there is always some $n \geq m$ such that $\omega \in A_n$, or, equivalently, $\omega \in A_n$ for infinitely many values of n or, equivalently, for arbitrarily large values of n. A convenient way to express this is

$$\omega \in \{A_n \text{ infinitely often (i.o.)}\} \quad \Longleftrightarrow \quad \omega \in A^*. \tag{7.1}$$

Example 7.1. Let $X_1, X_2, \ldots$ be a sequence of random variables and let $A_n = \{|X_n| > \varepsilon\}$, $n \geq 1$, $\varepsilon > 0$. Then $\omega \in \liminf_{n\to\infty} A_n^c$ means that ω is such that $|X_n(\omega)| \leq \varepsilon$ for all sufficiently large n, and $\omega \in \limsup_{n\to\infty} A_n$ means that ω is such that there exist arbitrarily large values of n such that $|X_n(\omega)| > \varepsilon$. In particular, every ω for which $X_n(\omega) \to 0$ as $n \to \infty$ must be such that, for every $\varepsilon > 0$, only finitely many of the real numbers $X_n(\omega)$ exceed ε in absolute value. Hence,

$$X_n \overset{a.s.}{\to} 0 \text{ as } n \to \infty \quad \Longleftrightarrow \quad P(|X_n| > \varepsilon \text{ i.o.}) = 0 \quad \text{for all} \quad \varepsilon > 0. \tag{7.2}$$

We shall return to this example later. □

Here is the first Borel–Cantelli lemma.

Theorem 7.1. *Let $\{A_n, n \geq 1\}$ be arbitrary events. Then*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Longrightarrow \quad P(A_n \ i.o.) = 0.$$

Proof. We have

$$\begin{aligned} P(A_n \text{ i.o.}) &= P(\limsup_{n\to\infty} A_n) = P(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m) \\ &\leq P(\bigcup_{m=n}^{\infty} A_m) \leq \sum_{m=n}^{\infty} P(A_m) \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$
□

The converse does not hold in general—one example is given at the very end of this section. However, with an additional assumption of independence, the following, second Borel–Cantelli lemma, holds true.

Theorem 7.2. *Let $\{A_n, n \geq 1\}$ be independent events. Then*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \Longrightarrow \quad P(A_n \ i.o.) = 1.$$

Proof. By the De Morgan formula and independence we obtain

$$\begin{aligned} P(A_n \text{ i.o.}) &= P\Big(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big) = 1 - P\Big(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c\Big) \\ &= 1 - \lim_{n\to\infty} P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = 1 - \lim_{n\to\infty}\lim_{N\to\infty} P\Big(\bigcap_{m=n}^{N} A_m^c\Big) \\ &= 1 - \lim_{n\to\infty}\lim_{N\to\infty} \prod_{m=n}^{N} \big(1 - P(A_m)\big). \end{aligned}$$

Now, since for $0 < x < 1$ we have $e^{-x} \geq 1 - x$, it follows that

$$\prod_{m=n}^{N} \big(1 - P(A_m)\big) \leq \exp\Big\{ -\sum_{m=n}^{N} P(A_m)\Big\} \to 0 \quad \text{as} \quad N \to \infty$$

for every n, since, by assumption, $\sum_{m=1}^{\infty} P(A_m) = \infty$. □

Remark 7.1. There exist more general versions of this result that allow for some dependence between the events (i.e., independence is not necessary for the converse to hold). □

As a first application, let us reconsider Examples 6.3.1 and 6.3.2.

Example 7.2. Thus, $X_2, X_3, \ldots$ is a sequence of random variables such that

$$P(X_n = 1) = 1 - \frac{1}{n^\alpha} \quad \text{and} \quad P(X_n = n) = \frac{1}{n^\alpha}, \quad n \geq 2,$$

where α is some positive number. Under the additional assumption that the random variables are independent, it was claimed in Remark 6.3.5 that $X_n \stackrel{a.s.}{\to} 1$ as $n \to \infty$ when $\alpha = 2$ and proved in Example 6.3.1 that this is not the case when $\alpha = 1$.

Now, in view of the first Borel–Cantelli lemma, it follows immediately that $X_n \stackrel{a.s.}{\to} 1$ as $n \to \infty$ for *all* $\alpha > 1$, even without any assumption about independence! To see this we first recall Example 7.1, according to which

$$X_n \stackrel{a.s.}{\to} 1 \quad \text{as} \quad n \to \infty \quad \Longleftrightarrow \quad P(|X_n - 1| > \varepsilon \text{ i.o.}) = 0 \quad \text{for all} \quad \varepsilon > 0.$$

The desired conclusion now follows from Theorem 7.1 since, for $\alpha > 1$,

$$\sum_{n=1}^{\infty} P(|X_n - 1| > \varepsilon) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty \quad \text{for all} \quad \varepsilon > 0.$$

It follows, moreover, from the second Borel–Cantelli lemma that if, in addition, we assume that $X_1, X_2, \ldots$ are independent, then we do not have almost-sure convergence for any $\alpha \leq 1$. In particular, almost-sure convergence holds if and only if $\alpha > 1$ in that case. □

A second look at the arguments above shows (please check!) that, in fact, the following, more general result holds true.

Theorem 7.3. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables. Then*

$$X_n \stackrel{a.s.}{\to} 0 \text{ as } n \to \infty \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} P(|X_n| > \varepsilon) < \infty \quad \text{for all} \quad \varepsilon > 0.$$ □

Let us now comment on formula(s) (6.3.1) (and (6.3.2)), which were presented before without proof, and show, at least, that almost-sure convergence implies their validity. Toward this end, let $X_1, X_2, \ldots$ be a sequence of random variables and $A = \{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$ for some random variable X. Then (why?)

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} \Big\{|X_i - X| \le \frac{1}{n}\Big\}. \tag{7.3}$$

Thus, assuming that almost-sure convergence holds, we have $P(A) = 1$, from which it follows that

$$P\Big(\bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} \Big\{|X_i - X| \le \frac{1}{n}\Big\}\Big) = 1$$

for all n. Furthermore, the sets $\{\bigcap_{i=m}^{\infty}\{|X_i - X| \le 1/n\},\, m \ge 1\}$ are monotone increasing as $m \to \infty$, which, in view of Lemma 6.3.1, implies that, for all n,

$$\lim_{m\to\infty} P\Big(\bigcap_{i=m}^{\infty} \Big\{|X_i - X| \le \frac{1}{n}\Big\}\Big) = P\Big(\bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} \Big\{|X_i - X| \le \frac{1}{n}\Big\}\Big).$$

However, the latter probability was just seen to equal 1, from which it follows that $P(\bigcap_{i=m}^{\infty}\{|X_i - X| \le 1/n\})$ can be made arbitrary close to 1 by choosing m large enough. Therefore, since n was arbitrary we have shown (why?) that if $X_n \overset{a.s.}{\to} X$ as $n \to \infty$ then, for every $\varepsilon > 0$ and δ, $0 < \delta < 1$, there exists m_0 such that for all $m > m_0$ we have

$$P\Big(\bigcap_{i=m}^{\infty} \{|X_i - X| < \varepsilon\}\Big) > 1 - \delta,$$

which is exactly (6.3.1) (which was equivalent to (6.3.2)).

7.1 Patterns

We begin with an example of a different and simpler nature.

Example 7.3. Toss a regular coin repeatedly (independent tosses) and let $A_n = \{\text{the } n\text{th toss yields a head}\}$ for $n \ge 1$. Then

$$P(A_n \text{ i.o.}) = 1.$$

To see this we note that $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} 1/2 = \infty$, and the conclusion follows from Theorem 7.2.

In words, if we toss a regular coin repeatedly, we obtain only finitely many heads with probability zero. Intuitively, this is obvious since, by symmetry, if this were not true, the same would not be true for tails either, which is impossible, since at least one of them must appear infinitely often.

However, for a biased coin, one could imagine that if the probability of obtaining heads is "very small," then it might happen that, with some "very small" probability, only finitely many heads appear. To treat that case, suppose that $P(\text{heads}) = p$, where $0 < p < 1$. Then $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} p = \infty$. We thus conclude, from the second Borel–Cantelli lemma, that $P(A_n \text{ i.o.}) = 1$ for *any* coin (unless it has two heads and no tails, or vice versa). □

The following exercise can be solved similarly, but a little more care is required, since the corresponding events are no longer independent; recalling Subsection 1.3 we find that the events form a 1-dependent sequence.

Exercise 7.1. Toss a coin repeatedly as before and let $A_n = \{\text{the } (n-1)\text{th and the } n\text{th toss both yield a head}\}$ for $n \geq 2$. Then

$$P(A_n \text{ i.o.}) = 1.$$

In other words, the event "two heads in a row" will occur infinitely often with probability 1.

Exercise 7.2. Toss another coin as above. Show that any finite pattern occurs infinitely often with probability 1. □

Remark 7.2. There exists a theorem, called Kolmogorov's 0-1 law, according to which, for independent events $\{A_n, n \geq 1\}$, the probability $P(A_n \text{ i.o.})$ can only assume the values 0 or 1. Example 7.3 above is of this kind, and, by exploiting the fact that the events $\{A_{2n}, n \geq 1\}$ are independent, one can show that the law also applies to Exercise 7.1. The problem is, of course, to decide which of the values is the true one for the problem at hand. □

The previous problem may serve as an introduction to *patterns*. In some vague sense we may formulate this by stating that given a finite alphabet, any finite sequence of letters, such that the letters are selected uniformly at random, will appear infinitely often with probability 1. A natural question is to ask how long one has to wait for the appearance of a given sequence. That this problem is more sophisticated than one might think at first glance is illustrated by the following example.

Example 7.4. Let $X, X_1, X_2, \ldots$ be i.i.d. random variables, such that $P(X = 0) = P(X = 1) = 1/2$.

(a) Let N_1 be the number of 0's and 1's until the first appearance of the pattern 10. Find $E\,N_1$.
(b) Let N_2 be the number of 0's and 1's until the first appearance of the pattern 11. Find $E\,N_2$.

Before we try to solve this problem it seems pretty obvious that the answers are the same for (a) and (b). However, this is not true!
(a) Let N_1 be the required number. A realization of the game would run as follows: We start off with a random number of 0's (possibly none) which at some point are followed by a 1, after which we are done as soon as a 0 appears. Technically, the pattern 10 appears after the following sequence

$$\underbrace{000\ldots0001}_{M_1}\underbrace{111\ldots1110}_{M_2},$$

where thus M_1 and M_2 are independent $\text{Fs}(1/2)$-distributed random variables, which implies that

$$E\,N_1 = E(M_1 + M_2) = E\,M_1 + E\,M_2 = 2 + 2 = 4.$$

(b) Let N_2 be the required number. This case is different, because when the first 1 has appeared we are done *only if the next digit equals* 1. If this is not the case we start over again. This means that there will be a geometric number of M_1 blocks followed by 0, after which the sequence is finished off with another M_1 block followed by 1:

$$\underbrace{000\ldots0001}_{M_1(1)}0\,\underbrace{000\ldots0001}_{M_1(2)}0\ldots\underbrace{000\ldots0001}_{M_1(Y)}0\,\underbrace{000\ldots0001}_{M_1^*}1\,,$$

that is,

$$N_2 = \sum_{k=1}^{Y}(M_1(k) + 1) + (M_1^* + 1),$$

where, thus $Y \in \text{Ge}(1/2)$, $M_1(k)$, and M_1^* all are distributed as M_1 and all random variables are independent. Thus,

$$E\,N_2 = E\,(Y + 1) \cdot E(M_1 + 1) = (1 + 1) \cdot (2 + 1) = 6.$$

Alternatively, and as the mathematics reveals, we may consider the experiment as consisting of $Z\,(= Y + 1)$ blocks of size $M_1 + 1$, where the last block is a success and the previous ones are failures. With this viewpoint we obtain

$$N_2 = \sum_{k=1}^{Z}(M_1(k) + 1),$$

and the expected value turns out the same as before, since $Z \in \text{Fs}(1/2)$.

Another solution that we include because of its beauty is to condition on the outcome of the first digit(s) and see how the process evolves after that using the law of total probability. A similar kind of argument was used in the early part of the proof of Theorem 3.7.3 concerning the probability of extinction in a branching process.

There are three ways to start off:

1. the first digit is a 0, after which we start from scratch;
2. the first two digits are 10, after which we start from scratch;
3. the first two digits are 11, after which we are done.

It follows that

$$N_2 = \frac{1}{2}(1 + N_2') + \frac{1}{4}(2 + N_2'') + \frac{1}{4} \cdot 2,$$

where N_2' and N_2'' are distributed as N_2. Taking expectation yields

$$EN_2 = \frac{1}{2} \cdot (1 + E\,N_2) + \frac{1}{4} \cdot (2 + E\,N_2) + \frac{1}{4} \cdot 2 = \frac{3}{2} + \frac{3}{4} E\,N_2,$$

from which we conclude that $E\,N_2 = 6$.

To summarize, for the sequence "10" the expected number was 4 and for the sequence "11" it was 6. By symmetry it follows that for "01" and "00" the answers must also be 4 and 6, respectively.

The reason for the different answers is that beginning and end are overlapping in 11 and 00, but not in 10 and 01. The overlapping makes it harder to obtain the desired sequence. This may also be observed in the different solutions. Whereas in (a) once the first 1 has appeared we simply have to wait for a 0, in (b) the 0 *must appear immediately* after the 1, otherwise we start from scratch again. Note how this is reflected in the last solution of (b).

7.2 Records Revisited

For another application of the Borel–Cantelli lemmas we recall the records from Section 6. For a sequence $X_1, X_2, \ldots$ of i.i.d. continuous random variables the record times were $L(1) = 1$ and $L(n) = \min\{k : X_k > X_{L(n-1)}\}$ for $n \geq 2$. We also introduced the indicator variables $\{I_k,\, k \geq 1\}$, which equal 1 if a record is observed and 0 otherwise, and the counting process $\{\mu(n),\, n \geq 1\}$ is defined by

$$\mu(n) = \sum_{k=1}^{n} I_k = \#\,\text{records among } X_1, X_2, \ldots, X_n = \max\{k : L(k) \leq n\}.$$

Since $P(I_k = 1) = 1/k$ for all k we conclude that

$$\sum_{n=1}^{\infty} P(I_k = 1) = \infty,$$

so that, because of the independence of the indicators, the second Borel–Cantelli lemma tells us that there will be infinitely many records with probability 1. This is not surprising, since, intuitively, there is always room for a new observation that is bigger than all others so far.

After this it is tempting to introduce *double records*, which appear whenever there are two records immediately following each other. Intuition this time might suggest once more that there is always room for two records in a row. So, let us check this.

Let $D_n = 1$ if X_n produces a double record, that is, if X_{n-1} and X_n both are records, and let $D_n = 0$ otherwise. Then, for $n \geq 2$,

$$P(D_n = 1) = P(I_n = 1, I_{n-1} = 1) = P(I_n = 1) \cdot P(I_{n-1} = 1) = \frac{1}{n} \cdot \frac{1}{n-1}\,.$$

We also note that the random variables $\{D_n,\, n \geq 2\}$ are *not* independent (more precisely, they are 1-dependent), which causes no problem. Namely,

$$\sum_{n=2}^{\infty} P(D_n = 1) = \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = \lim_{m\to\infty} \sum_{n=2}^{m} \Big(\frac{1}{n-1} - \frac{1}{n}\Big) = \lim_{m\to\infty} (1 - \frac{1}{m}) = 1,$$

so that by the first Borel–Cantelli lemma—which does not require independence—we conclude that

$$P(D_n = 1 \text{ i.o.}) = 0,$$

that is, the probability of infinitely many double records is equal to zero.

Moreover, the expected number of double records is

$$E \sum_{n=2}^{\infty} D_n = \sum_{n=2}^{\infty} E\, D_n = \sum_{n=2}^{\infty} P(D_n = 1) = 1;$$

in other words, we can expect *one* double record. A detailed analysis shows that, in fact, the total number of double records is

$$\sum_{n=2}^{\infty} D_n \in \text{Po}(1).$$

7.3 Complete Convergence

We close this section by introducing another convergence concept, which, as will be seen, is closely related to the Borel–Cantelli lemmas.

Definition 7.2. *A sequence* $\{X_n,\, n \geq 1\}$ *of random variables* converges completely *to the constant* θ *if*

$$\sum_{n=1}^{\infty} P(|X_n - \theta| > \varepsilon) < \infty \quad \text{for all} \quad \varepsilon > 0. \qquad \square$$

Two immediate observations are that complete convergence always implies a.s. convergence in view of the first Borel–Cantelli lemma and that complete convergence and almost-sure convergence are equivalent for sequences of independent random variables.

Theorem 7.4. *Let $X_1, X_2, \ldots$ be random variables and θ be some constant. The following implications hold as $n \to \infty$:*

$$X_n \to \theta \text{ completely} \quad \Longrightarrow \quad X_n \stackrel{a.s.}{\to} \theta.$$

If, in addition, $X_1, X_2, \ldots$ are independent, then

$$X_n \to \theta \text{ completely} \quad \Longleftrightarrow \quad X_n \stackrel{a.s.}{\to} \theta. \qquad \square$$

Example 7.5. Another inspection of Example 6.3.1 tells us that it follows immediately from the definition of complete convergence that $X_n \to 1$ completely as $n \to \infty$ when $\alpha > 1$ and that complete convergence does not hold if $X_1, X_2, \ldots$ are independent and $\alpha \le 1$. □

The concept was introduced in the late 1940s in connection with the following result:

Theorem 7.5. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, and set $S_n = \sum_{k=1}^{n} X_k$, $n \ge 1$. Then*

$$\frac{S_n}{n} \to 0 \text{ completely as } n \to \infty \quad \Longleftrightarrow \quad E\,X = 0 \text{ and } E\,X^2 < \infty\,,$$

or, equivalently,

$$\sum_{n=1}^{\infty} P(|S_n| > n\varepsilon) < \infty \text{ for all } \varepsilon > 0 \quad \Longleftrightarrow \quad E\,X = 0 \text{ and } E\,X^2 < \infty\,. \quad \square$$

Remark 7.3. A first naive attempt to prove the sufficiency would be to use Chebyshev's inequality. The attack fails, however, since the harmonic series diverges; more sophisticated tools are required. □

We mentioned in Remark 6.5.1 that the so-called strong law of large numbers, which states that S_n/n converges almost surely as $n \to \infty$, is equivalent to the existence of the mean, $E\,X$. Consequently, if the mean exists and/but the variance (or any moment of higher order than the first one) does not exist, then almost-sure convergence holds. In particular, if the mean equals 0, then

$$P(|S_n| > n\varepsilon \text{ i.o.}) = 0 \quad \text{for all} \quad \varepsilon > 0\,,$$

whereas Theorem 7.5 tells us that the corresponding Borel–Cantelli sum *diverges* in this case. This is the example we promised just before stating Theorem 7.2. Note also that the events $\{|S_n| > n\varepsilon,\, n \ge 1\}$ are definitely not independent.

8 Martingales

One of the most important modern concepts in probability is the concept of martingales. A rigorous treatment is beyond the scope of this book. The purpose of this section is to give the reader a flavor of martingale theory in a slightly simplified way.

Definition 8.1. *Let* X_1, X_2, ... *be a sequence of random variables with finite expectations. We call* X_1, X_2, ... *a* martingale *if*

$$E(X_{n+1} \mid X_1, X_2, \ldots, X_n) = X_n \quad \textit{for all} \quad n \geq 1. \qquad \square$$

The term *martingale* originates in gambling theory. The famous game *double or nothing*, in which the gambling strategy is to double one's stake as long as one loses and leave as soon as one wins, is called a "martingale." That it is, indeed, a martingale in the sense of our definition will be seen below.

Exercise 8.1. Use Theorem 2.2.1 to show that X_1, X_2, ... is a martingale if and only if

$$E(X_n \mid X_1, X_2, \ldots, X_m) = X_m \quad \text{for all} \quad n \geq m \geq 1. \qquad \square$$

In general, consider a game such that X_n is the gambler's fortune after n plays, $n \geq 1$. If the game satisfies the martingale property, it means that the expected fortune of the player, given the history of the game, equals the current fortune. Such games may be considered to be fair, since on average neither the player nor the bank loses any money.

Example 8.1. The canonical example of a martingale is a sequence of partial sums of independent random variables with mean zero. Namely, let Y_1, Y_2, ... be independent random variables with mean zero, and set

$$X_n = Y_1 + Y_2 + \cdots + Y_n, \quad n \geq 1.$$

Then

$$\begin{aligned} E(X_{n+1} \mid X_1, X_2, \ldots, X_n) &= E(X_n + Y_{n+1} \mid X_1, X_2, \ldots, X_n) \\ &= X_n + E(Y_{n+1} \mid X_1, X_2, \ldots, X_n) \\ &= X_n + E(Y_{n+1} \mid Y_1, Y_2, \ldots, Y_n) \\ &= X_n + 0 = X_n, \end{aligned}$$

as claimed. For the second equality we used Theorem 2.2.2(a), and for the third one we used the fact that knowledge of $X_1, X_2, \ldots, X_n$ is equivalent to knowledge of $Y_1, Y_2, \ldots, Y_n$. The last equality follows from the independence of the summands; recall Theorem 2.2.2(b). $\square$

Another example is a sequence of products of independent random variables with mean 1.

Example 8.2. Suppose that $Y_1, Y_2, \ldots$ are independent random variables with mean 1, and set $X_n = \prod_{k=1}^{n} Y_k$, $n \geq 1$ (with $Y_0 = X_0 = 1$). Then

$$\begin{aligned} E(X_{n+1} \mid X_1, X_2, \ldots, X_n) &= E(X_n \cdot Y_{n+1} \mid X_1, X_2, \ldots, X_n) \\ &= X_n \cdot E(Y_{n+1} \mid X_1, X_2, \ldots, X_n) \\ &= X_n \cdot 1 = X_n, \end{aligned}$$

which verifies the martingale property of $\{X_n, n \geq 1\}$.

One application of this example is the game "double or nothing" mentioned above. To see this, set $X_0 = 1$ and, recursively,

$$X_{n+1} = \begin{cases} 2X_n, & \text{with probability } \frac{1}{2}, \\ 0, & \text{with probability } \frac{1}{2}, \end{cases}$$

or, equivalently,

$$P(X_n = 2^n) = \frac{1}{2^n}, \qquad P(X_n = 0) = 1 - \frac{1}{2^n} \quad \text{for} \quad n \geq 1.$$

Since

$$X_n = \prod_{k=1}^{n} Y_k,$$

where $Y_1, Y_2, \ldots$ are i.i.d. random variables such that $P(Y_k = 0) = P(Y_k = 2) = 1/2$ for all $k \geq 1$, it follows that X_n equals a product of i.i.d. random variables with mean 1, so that $\{X_n, n \geq 1\}$ is a martingale.

A problem with this game is that the expected money spent when the game is over is infinite. Namely, suppose that the initial stake is 1 euro. If the gambler wins at the nth game, she or he has spent $1+2+4+\cdots+2^{n-1} = 2^n-1$ euros and won 2^n euros, for a total net of 1 euro. The total number of games is Fs(1/2)-distributed. This implies on the one hand that, on average, a success or win occurs after two games, and on the other hand that, on average, the gambler will have spent an amount of

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \big(2^n - 1\big) = \infty$$

euros in order to achieve this. In practice this is therefore an impossible game. A truncated version would be to use the same strategy but to leave the game no matter what happens after (at most) a fixed number of games (to be decided before the game starts).

Another example is related to the likelihood ratio test. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with common density f and some characterizing parameter θ of interest. In order to test the null hypothesis $H_0 : \theta = \theta_0$

against the alternative $H_1 : \theta = \theta_1$, the *Neyman–Pearson lemma* in statistics tells us that such a test should be based on the likelihood ratio statistic

$$L_n = \prod_{k=1}^{n} \frac{f(X_k;\theta_1)}{f(X_k;\theta_0)},$$

where f_{θ_0} and f_{θ_1} are the densities under the null and alternative hypotheses, respectively.

Now, the factors $f(X_k;\theta_1)/f(X_k;\theta_0)$ are i.i.d. random variables, and, under the null hypothesis, the mean equals

$$E_0\Big(\frac{f(X_k;\theta_1)}{f(X_k;\theta_0)}\Big) = \int_{-\infty}^{\infty} \frac{f(x;\theta_1)}{f(x;\theta_0)} f(x;\theta_0)\,dx = \int_{-\infty}^{\infty} f(x;\theta_1)\,dx = 1,$$

that is, L_n is made up as a product of i.i.d. random variables with mean 1, from which we immediately conclude that $\{L_n,\, n \geq 1\}$ is a martingale.

We also remark that if $=$ in the definition is replaced by $\geq$ then $X_1, X_2, \ldots$ is called a *submartingale*, and if it is replaced by $\leq$ it is called a *supermartingale.* As a typical example one can show that if $\{X_n,\, n \geq 1\}$ is a martingale and $E|X_n|^r < \infty$ for all $n \geq 1$ and some $r \geq 1$, then $\{|X_n|^r,\, n \geq 1\}$ is a submartingale.

Applying this to the martingale in Example 8.1 tells us that whereas the sums $\{X_n,\, n \geq 1\}$ of independent random variables with mean zero constitute a martingale, such is not the case with the sequence of sums of squares $\{X_n^2,\, n \geq 1\}$ (provided the variances are finite); that sequence is a submartingale. However by centering the sequence one obtains a martingale. This is the topic of Problems 9.11 and 9.12.

There also exist so-called *reversed martingales.* If we interpret n as time, then "reversing" means reversing time. Traditionally one defines reversed martingales via the relation

$$X_n = E(X_m \mid X_{n+1}, X_{n+2}, X_{n+3}, \ldots) \quad \text{for all} \quad m < n\,,$$

which means that one conditions on "the future." The more modern way is to let the index set be the negative integers as follows.

Definition 8.2. *Let* $\ldots, X_{-3}, X_{-2}, X_{-1}$ *be a sequence of random variables with finite expectations. We call* $\ldots, X_{-3}, X_{-2}, X_{-1}$ *a* reversed martingale *if*

$$E(X_{n+1} \mid \ldots, X_{n-3}, X_{n-2}, X_{n-1}, X_n) = X_n \quad \textit{for all} \quad n \leq -1.$$ □

The obvious parallel to Exercise 8.1 is next.

Exercise 8.2. Use Theorem 2.2.1 to show that $\ldots, X_{-3}, X_{-2}, X_{-1}$ is a reversed martingale if and only if

$$E(X_n \mid \ldots, X_{m-3}, X_{m-2}, X_{m-1}, X_m) = X_m \quad \text{for all} \quad m \leq n \leq 0.$$

In particular, $\ldots, X_{-3}, X_{-2}, X_{-1}$ is a reversed martingale if and only if,

$$E(X_{-1} \mid \ldots, X_{m-3}, X_{m-2}, X_{m-1}, X_m) = X_m \quad \text{for all} \quad m \leq -1.$$ □

Just as the sequence of sums of independent random variables with mean zero constitutes the generic example of a martingale it turns out that the sequence of arithmetic means of i.i.d. random variables with finite mean (not necessarily equal to zero) constitutes the generic example of a reversed martingale.

To see this, suppose that $Y_1, Y_2, \ldots$ are i.i.d. random variables with finite mean μ, set $S_n = \sum_{k=1}^{n} Y_k$, $n \geq 1$, and

$$X_{-n} = \frac{S_n}{n} \quad \text{for} \quad n \geq 1.$$

We wish to show that

$$\{X_n,\, n \leq -1\} \quad \text{is a martingale.} \tag{8.1}$$

Now, knowing the arithmetic means when $k \geq n$ is the same as knowing S_n and Y_k, $k > n$, so that, due to independence,

$$\begin{aligned} E\big(X_{-n} \mid X_k,\, k \leq n-1\big) &= E\Big(\frac{S_n}{n} \mid S_{n+1},\, Y_{n+2},\, Y_{n+3}\,, \ldots\Big) \\ &= E\Big(\frac{S_n}{n} \mid S_{n+1}\Big) = \frac{1}{n}\sum_{k=1}^{n} E(Y_k \mid S_{n+1}) \\ &= \frac{1}{n}\sum_{k=1}^{n} \frac{S_{n+1}}{n+1} = \frac{S_{n+1}}{n+1} = X_{-n-1}, \end{aligned}$$

where, in the third to last equality we exploited the symmetry, which in turn, is due to the equidistribution.

We have thus established relation (8.1) as desired.

Remark 8.1. Reversed submartingales and *reversed supermartingales* may be defined "the obvious way." □

Exercise 8.3. Define them! □

We close this introduction to the theory of martingales by stating (without proof) the main convergence results. Analogous, although slightly different, results also hold for submartingales and supermartingales.

Theorem 8.1. *Suppose that* $\{X_n,\, n \geq 1\}$ *is a martingale. If*

$$\sup_n E \max\{X_n, 0\} < \infty,$$

then X_n *converges almost surely as* $n \to \infty$*. Moreover, the following are equivalent:*

(a) $\{X_n,\, n \geq 1\}$ *is uniformly integrable;*
(b) X_n *converges in* 1*-mean;*

(c) $X_n \stackrel{a.s.}{\to} X_\infty$ *as* $n \to \infty$, *where* $E|X_\infty| < \infty$, *and* X_∞ closes *the sequence, that is,* $\{X_n, n = 1, 2, \ldots, \infty\}$ *is a martingale;*
(d) *there exists a random variable* Y *with finite mean such that*

$$X_n = E(Y \mid X_1, X_2, \ldots, X_n) \quad \textit{for all} \quad n \geq 1. \qquad \square$$

The analog for reversed martingales runs as follows.

Theorem 8.2. *Suppose that* $\{X_n, n \leq -1\}$ *is a reversed martingale. Then*

(a) $\{X_n, n \leq -1\}$ *is uniformly integrable;*
(b) $X_n \to X_{-\infty}$ *a.s. and in* 1*-mean as* $n \to -\infty$*;*
(c) $\{X_n, -\infty \leq n \leq -1\}$ *is a martingale.* □

Note that the results differ somewhat. This is due to the fact that whereas ordinary, forward martingales always have a first element, but not necessarily a last element (which would correspond to X_∞), reversed martingales always have a last element, namely X_{-1}, but not necessarily a first element (which would correspond to $X_{-\infty}$). This, in turn, has the effect that reversed martingales "automatically" are uniformly integrable, as a consequence of which conclusions (a)–(c) are "automatic" for reversed martingales, but only hold under somewhat stronger assumptions for (forward) martingales.

Note also that the generic martingale, the sum of independent random variables with mean zero, need not be convergent at all. This is, in particular, the case if the summands are equidistributed with finite variance σ^2, in which case the sum S_n behaves, asymptotically, like $\sigma\sqrt{n} \cdot N(0,1)$, where $N(0,1)$ is a standard normal random variable.

9 Problems

1. Let $X_1, X_2, \ldots$ be independent, equidistributed random variables, and set $S_n = X_1 + \cdots + X_n$, $n \geq 1$. The sequence $\{S_n, n \geq 0\}$ (where $S_0 = 0$) is called a *random walk*. Consider the following "perturbed" random walk. Let $\{\varepsilon_n, n \geq 1\}$ be a sequence of random variables such that, for some fixed $A > 0$, we have $P(|\varepsilon_n| \leq A) = 1$ for all n, and set

$$T_n = S_n + \varepsilon_n, \quad n = 1, 2, \ldots.$$

Suppose that $E\,X_1 = \mu$ exists. Show that the law of large numbers holds for the perturbed random walk $\{T_n, n \geq 1\}$.

2. In a game of dice one wishes to use one of two dice A and B. A has two white and four red faces and B has two red and four white faces. A coin is tossed in order to decide which die is to be used and that die is then used throughout. Let $\{X_k, k \geq 1\}$ be a sequence of random variables defined as follows:

$$X_k = \begin{cases} 1, & \text{if red is obtained,} \\ 0, & \text{if white is obtained} \end{cases}$$

at the kth roll of the die. Show that the law of large numbers does not hold for the sequence $\{X_k,\, k \geq 1\}$. Why is this the case?

3. Suppose that $X_1, X_2, \ldots$ are independent random variables such that $X_k \in \text{Be}(p_k)$, $k \geq 1$, and set $S_n = \sum_{k=1}^n X_k$, $m_n = \sum_{k=1}^n p_k$, and $s_n^2 = \sum_{k=1}^n p_k(1-p_k)$, $n \geq 1$. Show that if

$$\sum_{k=1}^{\infty} p_k(1-p_k) = +\infty, \tag{9.1}$$

then

$$\frac{S_n - m_n}{s_n} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty.$$

Remark 1. The case $p_k = 1/k$, $k \geq 1$, corresponds to the record times, and we rediscover Theorem 6.4.

Remark 2. One can show that the assumption (9.1) is necessary for the conclusion to hold.

4. Prove the following central limit theorem for a sum of independent (not identically distributed) random variables: Suppose that $X_1, X_2, \ldots$ are independent random variables such that $X_k \in U(-k, k)$, and set $S_n = \sum_{k=1}^n X_k$, $n \geq 1$. Show that

$$\frac{S_n}{n^{3/2}} \xrightarrow{d} N(\mu, \sigma^2) \quad \text{as} \quad n \to \infty,$$

and determine μ and σ^2.

Remark. Note that the normalization is not proportional to $\sqrt{n}$; rather, it is asymptotically proportional to $\sqrt{\text{Var}\, S_n}$.

5. Let $X_1, X_2, \ldots$ be independent, $U(0,1)$-distributed random variables. We say that there is a *peak* at X_k if X_{k-1} and X_{k+1} are both smaller than X_k, $k \geq 2$. What is the probability of a peak at
(a) X_2?
(b) X_3?
(c) X_2 and X_3?
(d) X_2 and X_4?
(e) X_2 and X_5?
(f) X_i and X_j, $i, j \geq 2$?

Remark. Letting $I_k = 1$ if there is a peak at X_k and 0 otherwise, the sequence $\{I_k,\, k \geq 1\}$ forms a 2-dependent sequence of random variables.

6. Verify formula (2.1), i.e., that if $X, X_1, X_2, \ldots$ are i.i.d. symmetric stable random variables, then

$$\frac{S_n}{n^{1/\alpha}} \stackrel{d}{=} X \quad \text{for all} \quad n.$$

7. Prove that the law of large numbers holds for symmetric, stable distributions with index α, $1 < \alpha \leq 2$.
8. Let $0 < \alpha < 2$ and suppose that $X, X_1, X_2, \ldots$ are independent random variables with common (two-sided Pareto) density

$$f(x) = \begin{cases} \dfrac{\alpha}{2|x|^{\alpha+1}}, & \text{for} \quad |x| > 1, \\ 0, & \text{otherwise.} \end{cases}$$

Show that the distribution belongs to the domain of attraction of a symmetric stable distribution with index α; in other words, that the sums $S_n = \sum_{k=1}^{n} X_k$, suitably normalized, converge in distribution to a symmetric stable distribution with index α.

Remark 1. More precisely, one can show that $S_n/n^{1/\alpha}$ converges in distribution to a symmetric stable law with index α.

Remark 2. This problem generalizes Examples 3.1 and 3.2.

9. The same problem as the previous one, but for the density

$$f(x) = \begin{cases} \dfrac{c \log|x|}{|x|^{\alpha+1}}, & \text{for} \quad |x| > 1, \\ 0, & \text{otherwise,} \end{cases}$$

where c is an appropriate normalizing constant.

Remark. In this case one can show that $S_n/(n \log n)^{1/\alpha}$ converges in distribution to a symmetric stable law with index α.

10. Show that the extremal distributions belong to their own domain of attraction. More precisely, let $X, X_1, X_2, \ldots$ be i.i.d. random variables, and set

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}, \quad n \geq 1.$$

Show that,

(a) if X has a Fréchet distribution, then

$$\frac{Y_n}{n^{1/\alpha}} \stackrel{d}{=} X;$$

(b) if X has a Weibull distribution, then

$$n^{1/\alpha} Y_n \stackrel{d}{=} X;$$

(c) if X has a Gumbel distribution, then

$$Y_n - \log n \stackrel{d}{=} X.$$

11. Let $Y_1, Y_2, \ldots$ be independent random variables with mean zero and finite variances $\operatorname{Var} Y_k = \sigma_k^2$. Set

$$X_n = \Big(\sum_{k=1}^{n} Y_k\Big)^2 - \sum_{k=1}^{n} \sigma_k^2, \quad n \geq 1.$$

Show that $X_1, X_2, \ldots$ is a martingale.

12. Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with finite mean μ, and finite variance σ^2, and let S_n, $n \geq 1$, denote their partial sums. Set

$$X_n = (S_n - n\mu)^2 - n\sigma^2\,, \quad n \geq 1.$$

Show that $X_1, X_2, \ldots$ is a martingale.

13. Let $X(n)$ be the number of individuals in the nth generation of a branching process ($X(0) = 1$) with reproduction mean m ($= E\,X(1)$). Set

$$U_n = \frac{X(n)}{m^n}\,, \quad n \geq 1.$$

Show that $U_1, U_2, \ldots$ is a martingale.

14. Let $Y_1, Y_2, \ldots$ are i.i.d. random variables with a finite moment generating function ψ, set $S_n = \sum_{k=1}^{n} Y_k$, $n \geq 1$, with $S_0 = 0$, and

$$X_n = \frac{e^{tS_n}}{(\psi(t))^n}, \quad n \geq 1.$$

(a) Show that $\{X_n,\, n \geq 1\}$ is a martingale (which is frequently called *the exponential martingale*).
(b) Find the relevant martingale if the common distribution is the standard normal one.