Appendix 1
Extension of a Probability Measure

In this appendix we will prove Carathéodory's theorem, which was used in Sect. 2.1.

Let $\mathcal{A}$ be an algebra of subsets of Ω on which a probability measure $\mathbf{P}$, i.e., a real-valued function satisfying conditions P1–P3 of Chap. 2, is given. Let $\mathcal{P}$ denote the class of all subsets of Ω. For any $A \in \mathcal{P}$, there always exists a sequence $\{A_n\}_{n=1}^{\infty}$ of disjoint sets from $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \supset A$ (it suffices to take $A_1 = \Omega$ and $A_n = \varnothing$, $n \ge 2$). Denote by $\gamma(A)$ the class of all such sequences and introduce on $\mathcal{P}$ the real-valued function

$$\mathbf{P}^*(A) := \inf\left\{\sum_{n=1}^{\infty} \mathbf{P}(A_n);\ \{A_n\} \in \gamma(A)\right\}.$$

This function (the outer measure on $\mathcal{P}$ induced by the measure $\mathbf{P}$ on $\mathcal{A}$) has the following properties:

(1) $\mathbf{P}^*(A) \le \mathbf{P}^*(B) \le 1$ if $A \subset B$.
(2) $\mathbf{P}^*(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$ if the sets $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$, are disjoint.
(3) $\mathbf{P}^*(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \mathbf{P}^*(A_n)$ for any $A_1, A_2, \ldots \in \mathcal{P}$.

Property (1) is obvious. Property (2) is established by the following argument. Let $\{B_n\}$ be any sequence from $\gamma(A)$, where $A = \bigcup_{n=1}^{\infty} A_n$. Since $\bigcup_{m=1}^{\infty} A_n B_m = A_n \in \mathcal{A}$, one has $\mathbf{P}(A_n) = \sum_{m=1}^{\infty} \mathbf{P}(A_n B_m)$. Therefore,

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n}\sum_{m} \mathbf{P}(A_n B_m) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \mathbf{P}(A_n B_m).$$

But, for each $N < \infty$,

$$\sum_{n=1}^{N} \mathbf{P}(A_n B_m) \le \mathbf{P}(B_m).$$

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013

Hence this equality holds for $N=\infty$ as well and, for any sequence $\{B_m\}_{m=1}^{\infty} \in \gamma(A)$,

$$\sum_{n=1}^{N} \mathbf{P}(A_n) \le \mathbf{P}(B_m).$$

This implies that $\mathbf{P}^*(A) \ge \sum_{n=1}^{\infty} \mathbf{P}(A_n)$. Because the converse inequality is obvious, we have $\mathbf{P}^*(A) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$.

Proof of property (3) Consider, for some $\varepsilon > 0$, sequences $\{A_{nk}\}_{k=1}^{\infty} \in \gamma(A_n)$ such that

$$\sum_{k=1}^{\infty} \mathbf{P}(A_{nk}) \le \mathbf{P}^*(A_n) + \frac{\varepsilon}{2^n}.$$

The sequence of sets $\{A_{nk}\}_{n,k=1}^{\infty}$ clearly contains $\bigcup A_n$ and therefore

$$\mathbf{P}^*\Big(\bigcup A_n\Big) \le \sum_n \sum_k \mathbf{P}(A_{nk}) \le \sum_{n=1}^{\infty} \mathbf{P}^*(A_n) + \varepsilon.$$

Since ε is arbitrary, property (3) is proved. □

Introduce now the binary operation of *symmetric difference* $\oplus$ on arbitrary sets A and B from $\mathcal{P}$ by means of the equality

$$A \oplus B := A\overline{B} \cup \overline{A}B.$$

It is not hard to see that

$$A \oplus B = B \oplus A = \overline{A} \oplus \overline{B} \subset A \cup B, \qquad A \oplus A = \varnothing,$$
$$A \oplus \varnothing = A, \qquad (A \oplus B) \oplus C = A \oplus (B \oplus C).$$

With the help of this operation and the function $\mathbf{P}^*$, we introduce on $\mathcal{P}$ a distance ρ by putting, for any $A, B \in \mathcal{P}$,

$$\rho(A, B) := \mathbf{P}^*(A \oplus B).$$

This construction is quite similar to the one used in Sect. 3.4 (we considered there the distance $d(A, B) = \mathbf{P}(A \oplus B)$ between measurable sets A and B). The properties of the distance ρ are the same as in (3.4.2). We will need the following properties:

(1) $\rho(A, B) = \rho(B, A) \ge 0$, $\rho(A, A) = 0$,
(2) $\rho(\overline{A}, \overline{B}) = \rho(A, B)$,
(3) $\rho(AB, CD) \le \rho(A, C) + \rho(B, D)$,
(4) $\rho(\bigcup A_k, \bigcup B_k) \le \sum_k \rho(A_k, B_k)$.

We also note that

(5) $|\mathbf{P}^*(A) - \mathbf{P}^*(B)| \le \rho(A, B)$, and therefore $\mathbf{P}^*(\cdot)$ is a uniformly continuous function with respect to ρ.

Properties (1)–(3) were listed in (3.4.2); in the present context, they are proved in exactly the same way based on the properties of the measure $\mathbf{P}^*$. Property (4) follows from property (3) of the measure $\mathbf{P}^*$ and the relation (we put here $A = \bigcup A_n$ and $B = \bigcup B_n$)

$$A \oplus B \subset \bigcup (A_n \oplus B_n),$$

because

$$\begin{aligned} A \oplus B &= \left[\left(\bigcup A_n \right) \cap \left(\bigcap \overline{B}_n \right) \right] \cup \left[\left(\bigcap \overline{A}_n \right) \cap \left(\bigcup B_n \right) \right] \\ &\subset \left[\bigcup A_n \overline{B}_n \right] \cup \left[\bigcup B_n \overline{A}_n \right] = \bigcup (A_n \overline{B}_n \cup \overline{A}_n B_n) = \bigcup (A_n \oplus B_n). \end{aligned}$$

Property (5) follows from the fact that

$$A \subset B \cup (A \oplus B), \qquad B \subset A \cup (A \oplus B) \tag{A1.1}$$

and therefore

$$\begin{aligned} \mathbf{P}^*(A) - \mathbf{P}^*(B) &\le \mathbf{P}^*(A \oplus B) = \rho(A, B), \\ \mathbf{P}^*(B) - \mathbf{P}^*(A) &\le \mathbf{P}^*(A \oplus B) = \rho(A, B). \end{aligned}$$

Similarly to the terminology adopted in Sect. 3.4 we call a set $A \in \mathcal{P}$ *approximable* if there exists a sequence $A_n \in \mathcal{A}$ for which $\rho(A, A_n) \to 0$. The totality of all approximable sets we denote by $\mathfrak{A}$. This is clearly the closure of $\mathcal{A}$ with respect to ρ.

Lemma A1.1 *$\mathfrak{A}$ is a σ-algebra.*

Proof We verify that $\mathfrak{A}$ satisfies properties A1, A2′ and A3 of σ-algebras of Chap. 2. Property A1: $\Omega \in \mathfrak{A}$ is obvious, for $\mathcal{A} \in \mathfrak{A}$. Property A3 ($\overline{A} \in \mathfrak{A}$ if $A \in \mathfrak{A}$) follows from the fact that, for $A \in \mathfrak{A}$, there exist $A_n \in \mathcal{A}$ such that, as $n \to \infty$,

$$\rho(A, A_n) \to 0, \quad \rho(\overline{A}, \overline{A}_n) = \rho(A, A_n) \to 0.$$

Finally, consider property A2′. We show first that if $A_n \in \mathcal{A}$, then $A = \bigcup A_n \in \mathfrak{A}$.

Indeed, we can assume without loss of generality that the A_n are disjoint. Then, by virtue of the properties of the measure $\mathbf{P}^*$, for any $\varepsilon > 0$,

$$\sum \mathbf{P}(A_k) \le \mathbf{P}^*(\Omega) = 1,$$

$$\rho\left(A, \bigcup_{k=1}^{n} A_k\right) = \mathbf{P}^*\left(\bigcup_{k=n+1}^{\infty} A_k\right) = \sum_{k=n+1}^{\infty} \mathbf{P}(A_k) < \varepsilon$$

for n large enough.

Now let $A_n \in \mathfrak{A}$. We have to show that

$$A = \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}.$$

Let $\{B_n\}$ be a sequence of sets from $\mathcal{A}$ such that $\rho(A_n, B_n) < \varepsilon/2^n$. Then one has $B = \bigcup B_n \in \mathfrak{A}$ and, by property (4) of the distance ρ,

$$\rho(A, B) \le \sum_{n=1}^{\infty} \rho(A_n, B_n) < \varepsilon.$$

The lemma is proved. □

Now we can prove the main assertion.[1]

Theorem A1.1 *The probability* $\mathbf{P}$ *can be extended from the algebra* $\mathcal{A}$ *to some probability* $\overline{\mathbf{P}}$ *given on the* σ*-algebra* $\mathfrak{A}$.

Proof For $A \in \mathfrak{A}$, put

$$\overline{\mathbf{P}}(A) := \mathbf{P}^*(A).$$

It is evident that $\overline{\mathbf{P}}(A) = \mathbf{P}(A)$ for $A \in \mathcal{A}$, and $\overline{\mathbf{P}}(\Omega) = 1$. To verify that $\overline{\mathbf{P}}$ is a probability we just have to prove the countable additivity of $\overline{\mathbf{P}}$. We first prove the finite additivity. It suffices to prove it for two sets:

$$\mathbf{P}^*(A \cup B) = \mathbf{P}^*(A) + \mathbf{P}^*(B), \tag{A1.2}$$

where $A, B \in \mathfrak{A}$ and $A \cap B = \varnothing$. Let $A_n \in \mathcal{A}$ and $B_n \in \mathcal{A}$ be such that $\rho(A, A_n) \to 0$ and $\rho(B, B_n) \to 0$ as $n \to \infty$. Then

$$\left|\mathbf{P}^*(A \cup B) - \mathbf{P}^*(A_n \cup B_n)\right| \le \rho(A \cup B, A_n \cup B_n) \le \rho(A, A_n) + \rho(B, B_n) \to 0,$$

$$\mathbf{P}^*(A_n \cup B_n) = \mathbf{P}(A_n \cup B_n) = \mathbf{P}(A_n) + \mathbf{P}(B_n) - \mathbf{P}(A_n B_n). \tag{A1.3}$$

Here

$$\mathbf{P}(A_n) \to \mathbf{P}^*(A), \qquad \mathbf{P}(B_n) \to \mathbf{P}^*(B),$$

[1]The theorem on the extension of a measure to the minimum σ-algebra containing $\mathcal{A}$ was obtained by C. Carathéodory. The metrisation of normed Boolean algebras $\mathcal{A}$ by the distance $\rho(A, B) = \mathbf{P}(A \oplus B)$ was used by many authors (see, e.g., the talk by A.N. Kolmogorov at the 6th Polish Mathematical Congress in 1948 and Halmos [19]).

It was L.Ya. Savel'ev who suggested the use of the continuity properties of the measure with respect to the distance $\rho(A, B) = \mathbf{P}^*(A \oplus B)$ in order to extend it.

$$\mathbf{P}(A_n B_n) \le \mathbf{P}^*(A_n B) + \mathbf{P}^*(B_n \overline{B})$$
$$\le \mathbf{P}^*(A_n \overline{A}) + \mathbf{P}^*(B_n \overline{B}) \le \rho(A, A_n) + \rho(B, B_n) \to 0.$$

Hence (A1.3) implies (A1.2).

We now prove countable additivity. Let $A_n \in \mathfrak{A}$ be disjoint. Then, putting

$$A = \bigcup_{n=1}^{\infty} A_n,$$

we obtain from the finite additivity of $\overline{\mathbf{P}}$ that

$$\overline{\mathbf{P}}(A) = \sum_{k=1}^{n} \overline{\mathbf{P}}(A_k) + \overline{\mathbf{P}}\left(\bigcup_{k=n+1}^{\infty} A_k \right).$$

Therefore

$$\overline{\mathbf{P}}(A) \ge \sum_{k=1}^{\infty} \overline{\mathbf{P}}(A_k).$$

On the other hand,

$$\overline{\mathbf{P}}(A) = \mathbf{P}^*(A) \le \sum_{k=1}^{\infty} \mathbf{P}^*(A_k) = \sum_{k=1}^{\infty} \overline{\mathbf{P}}(A_k).$$

The theorem is proved. □

Theorem A1.2 *The extension of the probability* $\mathbf{P}$ *from the algebra* $\mathcal{A}$ *to the* σ-*algebra* $\mathfrak{A}$ *is unique.*

Proof Assume that there exists another probability $\mathbf{P}_1$ on $\mathfrak{A}$, which coincides with $\mathbf{P}$ on $\mathcal{A}$ and is such that, for some $A \in \mathfrak{A}$,

$$\mathbf{P}_1(A) \ne \overline{\mathbf{P}}(A).$$

Suppose first that $\varepsilon = \mathbf{P}_1(A) - \overline{\mathbf{P}}(A) > 0$. Consider a sequence $\{B_n\} \in \gamma(A)$ such that

$$\sum_{n=1}^{\infty} \mathbf{P}(B_n) - \overline{\mathbf{P}}(A) < \frac{\varepsilon}{2}.$$

Then

$$\mathbf{P}_1(A) = \overline{\mathbf{P}}(A) + \varepsilon \ge \sum_{n=1}^{\infty} \mathbf{P}(B_n) + \varepsilon/2$$

which contradicts the assumption that $A \subset \bigcup_{n=1}^{\infty} B_n$. Therefore

$$\mathbf{P}_1(A) \le \overline{\mathbf{P}}(A), \quad A \in \mathfrak{A}.$$

Since $\overline{\mathbf{P}}$ is ρ-continuous at the point $\varnothing$, it follows that $\mathbf{P}_1$ is also ρ-continuous at the point $\varnothing$, and hence at any "point" $A \in \mathfrak{A}$. Indeed, by virtue of (A1.1),

$$\left|\mathbf{P}_1(A) - \mathbf{P}_1(B)\right| \le \mathbf{P}_1(A \oplus B) \le \overline{\mathbf{P}}(A \oplus B) \to 0$$

if only $\rho(A, B) = \overline{\mathbf{P}}(A \oplus B) \to 0$. Hence, for $A \in \mathfrak{A}$,

$$\overline{\mathbf{P}}(A) = \lim_{\substack{B \to A \\ B \in \mathfrak{A}}} \mathbf{P}(B) = \lim_{\substack{B \to A \\ B \in \mathfrak{A}}} \mathbf{P}_1(B) = \mathbf{P}_1(A).$$

The theorem is proved. □

Let $\mathfrak{A}^* = \sigma(\mathcal{A})$ be the σ-algebra generated by $\mathcal{A}$. Since $\mathcal{A} \subset \mathfrak{A}$, we have $\mathfrak{A}^* \in \mathfrak{A}$, and the next statement follows in an obvious way from the above assertions.

Corollary A1.1 *The probability* $\mathbf{P}$ *can be uniquely extended from the algebra* $\mathcal{A}$ *to the* σ*-algebra* $\mathfrak{A}^*$ *generated by* $\mathcal{A}$.

Remark A1.1 The σ-algebra $\mathfrak{A}$ defined above as the closure of the algebra $\mathcal{A}$ with respect to the introduced distance ρ is in many cases wider than the σ-algebra $\mathfrak{A}^* = \sigma(\mathcal{A})$ generated by $\mathcal{A}$. This fact is closely related to the concept of the *completion* of a measure. To explain the concept, we assume from the very beginning that $\mathcal{A} = \mathfrak{F}$ is a σ-algebra. Then the measure $\overline{\mathbf{P}}$ can be constructed in a rather simple way. To do this we extend the measure $\mathbf{P}$ from $\langle \Omega, \mathfrak{F} \rangle$ to a σ-algebra which is wider than $\mathfrak{F}$ and is constructed as follows. We will say that a subset N of Ω belongs to the class $\mathcal{N}$ if there exists an $A = A(N) \in \mathfrak{F}$ such that $N \subset A$ and $\mathbf{P}(A) = 0$. It is not hard to see that the class of all sets of the form $B \cup N$, where $B \in \mathfrak{F}$ and $N \in \mathcal{N}$, also forms a σ-algebra. Denote it by $\mathfrak{F}_{\mathcal{N}}$. Putting $\mathbf{P}(B \cup N) := \mathbf{P}(B)$ we obtain an extension of $\mathbf{P}$ to $\langle \Omega, \mathfrak{F}_{\mathcal{N}} \rangle$. Such a measure is said to be *complete*, and the above operation itself is called the *completion of the measure* $\mathbf{P}$.

Now we can say that the measure $\overline{\mathbf{P}}$ constructed in Theorem A1.1 is complete, and the σ-algebra $\mathfrak{A}$ coincides with $\mathfrak{F}_{\mathcal{N}}$.

If, for example, $\Omega = [0, 1]$ and $\mathcal{A}$ is the algebra generated by the intervals, then $\mathfrak{A}^* = \sigma(\mathcal{A})$ will, as we already know, be the Borel σ-algebra, and $\mathfrak{A}$ will be the Lebesgue extension of $\mathfrak{A}^*$ consisting of all "Lebesgue measurable" sets.

Appendix 2
Kolmogorov's Theorem on Consistent Distributions

In this appendix we will prove the Kolmogorov theorem asserting that consistent distributions define a unique probability measure such that the consistent distributions are its projections. We used this theorem in Sect. 5.5 and in some other places, where distributions on infinite-dimensional spaces were considered.

Let T be an index set and, for each $t \in T$, $\mathbb{R}_t$ be the real line $(-\infty, \infty)$. Let $N \in T$ be a finite subset of T. Then the product space

$$\prod_{t \in T} \mathbb{R}_t = \mathbb{R}^N$$

is a Euclidean space of dimension equal to the number n of elements in N, spanned on n axes of the space

$$\mathbb{R}^T = \prod_{t \in T} \mathbb{R}_t.$$

Assume that, for any finite subset $N \subset T$, a probability measure $\mathbf{P}_N$ is given on $\langle \mathbb{R}^N, \mathfrak{B}^N \rangle$, where $\mathfrak{B}^N$ is the σ-algebra of Borel subsets of $\mathbb{R}^N$. Thereby a family of measures is given on $\mathbb{R}^T$. The family is said to be *consistent* if, for any $L \subset N$ and any Borel set B from $\mathbb{R}^L$,

$$\mathbf{P}_L(B) = \mathbf{P}_N\big(B \times \mathbb{R}^{N-L}\big).$$

The measure $\mathbf{P}_L$ is said to be the *projection* of $\mathbf{P}_N$ onto $\mathbb{R}^L$. A set from $\mathbb{R}^T$ that can be represented in the form $B \times \mathbb{R}^{T-N}$, where $B \in \mathfrak{B}^N$ and N is a finite set, is called a *cylinder set* in $\mathbb{R}^T$. The set B is said to be the *base* of the cylinder.

Denote by $\mathfrak{B}^T$ the σ-algebra of sets from $\mathbb{R}^T$ generated by all cylinder sets.

Theorem A2.1 (Kolmogorov) *If a consistent family of probability measures is given on $\mathbb{R}^T$, then there exists a unique probability measure $\mathbf{P}$ on $\langle \mathbb{R}^T, \mathfrak{B}^T \rangle$ such that, for any N, the measure $\mathbf{P}_N$ coincides with the projection of $\mathbf{P}$ onto $\mathbb{R}^N$.*

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

Proof The cylinder subsets of $\mathbb{R}^T$ form an algebra. We show that, for $B \in \mathfrak{B}^N$, the relations

$$\mathbf{P}\big(B \times \mathbb{R}^{T-N}\big) = \mathbf{P}_N(B) \tag{A2.1}$$

define a measure on this algebra. First of all, by consistency of the measures $\mathbf{P}_N$, this definition of probability on cylinder sets is consistent (we mean the cases when $B = B_1 \times \mathbb{R}^{N-L}$ for $B_1 \in \mathfrak{B}^L$; then the left-hand side of (A2.1) will also be equal to $\mathbf{P}(B_1 \times \mathbb{R}^{T-L})$). Further, the thus defined probability is additive. Indeed, let $B_1 \times \mathbb{R}^{T-N_1}$ and $B_2 \times \mathbb{R}^{T-N_2}$ be two disjoint cylinder sets. Then, putting $N = N_1 \cup N_2$, we will have

$$\begin{aligned}
&\mathbf{P}\big(\big(B_1 \times \mathbb{R}^{T-N_1}\big) \cup \big(B_2 \times \mathbb{R}^{T-N_2}\big)\big)\\
&\quad = \mathbf{P}\big(\big\{\big(B_1 \times \mathbb{R}^{N-N_1}\big) \cup \big(B_2 \times \mathbb{R}^{N-N_2}\big)\big\} \times \mathbb{R}^{T-N}\big)\\
&\quad = \mathbf{P}_N\big(\big\{\big(B_1 \times \mathbb{R}^{N-N_1}\big) \cup \big(B_2 \times \mathbb{R}^{N-N_2}\big)\big\}\big)\\
&\quad = \mathbf{P}_N\big(B_1 \times \mathbb{R}^{N-N_1}\big) + \mathbf{P}_N\big(B_2 \times \mathbb{R}^{N-N_2}\big)\\
&\quad = \mathbf{P}\big(B_1 \times \mathbb{R}^{T-N_1}\big) + \mathbf{P}\big(B_2 \times \mathbb{R}^{T-N_2}\big).
\end{aligned}$$

To verify that $\mathbf{P}$ is countably additive, we make use of the equivalence of properties P3 and P3′ (see Chap. 2). By this equivalence, it suffices to show that if $\mathcal{B}_n$, $n = 1, 2, \ldots$, is a decreasing sequence of cylinder sets and, for some $\varepsilon > 0$, we have $\mathbf{P}(\mathcal{B}) > \varepsilon$, $n = 1, 2, \ldots$, then $\mathcal{B} = \bigcap_{n=1}^{\infty} \mathcal{B}_n$ is not empty. Since the $\mathcal{B}_n$ are enclosed in all the preceding sets, in the representation $\mathcal{B}_n = B_n \times \mathbb{R}^{T-N_n}$ one has $N_n \subset N_{n+1}$ and $B_{n+1} \cap \mathbb{R}^{N_n} \subset B_n$. Without loss of generality, we will assume that the number of elements in the set $N_n = \{t_1, \ldots, t_n\}$ is equal to n, and denote by x_i (with various superscripts) the coordinates in the space $\mathbb{R}_{t_i}$.

Thus, let

$$\mathbf{P}(\mathcal{B}_n) = \mathbf{P}_{N_n}(B_n) \geq \varepsilon > 0.$$

We prove that the intersection

$$\mathcal{B} = \bigcap_{n=1}^{\infty} \mathcal{B}_n$$

is non-empty. For any Borel set $B_n \subset \mathbb{R}^{N_n}$, there exists a compactum K_n such that

$$K_n \subset B_n, \quad \mathbf{P}_{N_n}(B_n - K_n) < \frac{\varepsilon}{2^{n+1}}.$$

Setting $\mathcal{K}_n := K_n \times \mathbb{R}^{T-N_n}$, we obtain

$$\mathbf{P}(\mathcal{B}_n - \mathcal{K}_n) = \mathbf{P}_{N_n}(B_n - K_n) < \frac{\varepsilon}{2^{n+1}}.$$

Introduce the sets $\mathcal{D}_n := \bigcap_{k=1}^n \mathcal{K}_k$. It is easy to see that $\mathcal{D}_n \subset \mathcal{B}_n$ are also cylinders. Because

$$\mathcal{B}_n - \bigcap_{k=1}^{n} \mathcal{K}_k \subset \bigcap_{k=1}^{n} (\mathcal{B}_k - \mathcal{K}_k),$$

we have

$$\mathbf{P}(\mathcal{B}_n - \mathcal{D}_n) \le \mathbf{P}\left(\bigcap_{k=1}^{n} (\mathcal{B}_k - \mathcal{K}_k)\right) \le \sum_{k=1}^{n} \mathbf{P}(\mathcal{B}_k - \mathcal{K}_k) \le \frac{\varepsilon}{2};$$

$$\mathbf{P}(\mathcal{D}_n) \ge \mathbf{P}(\mathcal{B}_n) - \frac{\varepsilon}{2} \ge \frac{\varepsilon}{2}.$$

It follows that $\mathcal{D}_n$ is a decreasing sequence of non-empty cylinder sets. Denote by $X^n = (x_1^n, x_2^n, \ldots, x_n^n)$ an arbitrary point of the base

$$D_n = \bigcap_{k=1}^{n} K_k \times \mathbb{R}^{N_n - N_k}$$

of the cylinder $\mathcal{D}_n$. The point specifies a cylinder subset $\mathcal{X}$ of $\mathbb{R}^T$. Since the sets $\mathcal{D}_n$ decrease, we have $(x_1^{n+r}, x_2^{n+r}, \ldots, x_n^{n+r}) \in K_n$ for any $r \ge 0$. By compactness of K_n, we can choose a subsequence n_{1k} such that $x_1^{n_{1k}} \to x_1$ as $k \to \infty$. From this subsequence, one can choose a subsequence n_{2k} such that $x_2^{n_{2k}} \to x_2$, and so on.

Now consider the diagonal sequence of the points (or, more precisely, cylinder sets) $X^{n_{kk}} = (x_1^{n_{kk}}, x_2^{n_{kk}}, \ldots, x_{n_{kk}}^{n_{kk}})$. It is clear that

$$X^{n_{kk}} \to X = (x_1, x_2, \ldots)$$

(component-wise) as $k \to \infty$, and that

$$\left(x_1^{n_{kk}}, x_2^{n_{kk}}, \ldots, x_m^{n_{kk}}\right) \to (x_1, \ldots, x_m) \in K_m$$

for any m. This means that, for the set $\mathcal{X}$ corresponding to the point X, one has

$$\mathcal{X} := \left\{ y(t) \in \mathbb{R}^T : y(t_1) = x_1, y(t_2) = x_2, \ldots \right\} \subset K_m \subset \mathcal{B}_m$$

for any m, and therefore

$$\mathcal{X} \subset \bigcap_{m=1}^{\infty} \mathcal{B}_m.$$

Thus $\mathcal{B}$ is non-empty, and the countable additivity of $\mathbf{P}$ on the algebra of cylinder sets is proved. Hence $\mathbf{P}$ is a measure, and it remains to make use of the theorem on the extension of a measure from an algebra to the σ-algebra generated by that algebra.

The theorem is proved. □

Appendix 3
Elements of Measure Theory and Integration

In this appendix, the properties of integrals with respect to a measure are presented in more detail than in Chaps. 4 and 6. We also prove the basic theorems on decomposition of measure and on convergence of sequences of measures.

3.1 Measure Spaces

Let $\langle \Omega, \mathfrak{F} \rangle$ be a measurable space. We will say that a *measure space* $\langle \Omega, \mathfrak{F}, \boldsymbol{\mu} \rangle$ is given if $\boldsymbol{\mu}$ is a nonnegative countably additive set function on $\mathfrak{F}$, i.e. a function having the following properties:

(1) $\boldsymbol{\mu}(\bigcup_j A_j) = \sum_j \boldsymbol{\mu}(A_j)$ for any countable collection of disjoint sets $A_j \in \mathfrak{F}$ (σ-additivity);
(2) $\boldsymbol{\mu}(A) \ge 0$ for any $A \in \mathfrak{F}$;
(3) $\boldsymbol{\mu}(\varnothing) = 0$, where $\varnothing$ is the empty set.

The value $\boldsymbol{\mu}(A)$ is called the *measure* of the set A. We will only consider *finite* and σ*-finite* measures. In the former case one assumes that $\boldsymbol{\mu}(\Omega) < \infty$. In the latter case there exists a partition of Ω into countably many sets A_j such that $\boldsymbol{\mu}(A_j) < \infty$.

A probability space is an example of a space with a finite (unit) measure. The space $\langle \mathbb{R}, \mathfrak{B}, \boldsymbol{\mu} \rangle$, where $\mathbb{R}$ is the real line, $\mathfrak{B}$ is the σ-algebra of Borel sets, and $\boldsymbol{\mu}$ is the Lebesgue measure, is an example of a space with a σ-finite measure.

We can also consider such set functions $\boldsymbol{\mu}(A)$ that satisfy conditions (1) and (3) only, but are not necessarily nonnegative. Such functions are called *signed measures*. Any finite signed measure (i.e., such that $\sup_A \boldsymbol{\mu}(A) < \infty$ and $\inf_A \boldsymbol{\mu}(A) > -\infty$) can be represented as a difference of two nonnegative measures (the Hahn decomposition theorem, see Sect. 3.5 of the present appendix). We will need signed measures in Sect. 3.5 only. Everywhere else, unless otherwise specified, by measures we will understand set functions possessing properties (1)–(3).

In the same manner as when establishing the simplest properties of probability, one easily establishes the following properties of measures:

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013

(1) $\boldsymbol{\mu}(A) \le \boldsymbol{\mu}(B)$ if $A \subset B$,
(2) $\boldsymbol{\mu}(\bigcup_j A_j) \le \sum_j \boldsymbol{\mu}(A_j)$ for any A_j,
(3) if $A_n \subset A_{n+1}$ and $\bigcup_n A_n = A$ then $\boldsymbol{\mu}(A_n) \to \boldsymbol{\mu}(A)$, or, which is the same,
(3′) if $A_n \supset A_{n+1}$, $\bigcap_n A_n = A$, and $\boldsymbol{\mu}(A_1) < \infty$ then $\boldsymbol{\mu}(A_n) \to \boldsymbol{\mu}(A)$.

Consider further measurable functions on $\langle \Omega, \mathfrak{F} \rangle$, i.e., functions $\xi(\omega)$ having the property $\{\omega : \xi(\omega) \in B\} \in \mathfrak{F}$ for any Borel subset B of the real line.

The notions of *convergence in measure* and *convergence almost everywhere* are introduced similarly to the case of probability measure.

We will say that a sequence of measurable functions ξ_n *converges to* ξ *almost everywhere* (a.e.): $\xi_n \xrightarrow{a.e.} \xi$ as $n \to \infty$ if $\xi_n(\omega) \to \xi(\omega)$ for all ω except from a set of measure 0.

We will say that the ξ_n *converge to* ξ *in measure*: $\xi_n \xrightarrow{\mu} \xi$ if, for any $\varepsilon > 0$, as $n \to \infty$,

$$\boldsymbol{\mu}\big(\{|\xi_n - \xi| > \varepsilon\}\big) \to 0.$$

Now we turn to the construction of integrals and the study of their properties. First we consider *finite measures* assuming them without loss of generality to be *probability measures*. In that case we will write $\mathbf{P}(A)$ instead of $\boldsymbol{\mu}(A)$. We will turn to integrals with respect to arbitrary measures in Sect. 3.4.

3.2 The Integral with Respect to a Probability Measure

3.2.1 The Integrals of a Simple Function

A measurable function $\xi(\omega)$ is said to be *simple* if its range is finite. The *indicator* of a set $F \in \mathfrak{F}$ is the simple function

$$\mathrm{I}_F(\omega) = \begin{cases} 1, & \text{if } \omega \in F, \\ 0, & \text{if } \omega \notin F. \end{cases}$$

Clearly, any simple function $\xi(\omega)$ can be written in the form

$$\xi(\omega) = \sum_{k=1}^{n} x_k \mathrm{I}_{F_k}(\omega),$$

where x_k, $k = 1, 2, \ldots, n$, are values assumed by ξ, and $F_k = \{\omega : \xi(\omega) = x_k\}$. The sets $F_k \in \mathfrak{F}$ are disjoint, and $\bigcup_{k=1}^{n} F_k = \Omega$. The *integral* of the simple function $\xi(\omega)$ with respect to a measure $\mathbf{P}$ is defined as the quantity

$$\int \xi \, d\mathbf{P} = \int \xi(\omega) \, d\mathbf{P}(\omega) = \sum_{k=1}^{n} x_k \mathbf{P}(F_k) = \mathbf{E}\xi.$$

The *integral* of the simple function $\xi(\omega)$ *over a set* $A \in \mathfrak{F}$ is defined as

$$\int_A \xi \, d\mathbf{P} = \int \xi(\omega) \mathrm{I}_A(\omega) \, d\mathbf{P}(\omega).$$

That these definitions are consistent (the partitions into sets F_k may be different) can be verified in an obvious way.

3.2.2 The Integrals of an Arbitrary Function

Lemma A3.2.1 *Let* $\xi(\omega) > 0$. *There exists a sequence* $\xi_n(\omega)$ *of simple functions such that* $\xi_n(\omega) \uparrow \xi(\omega)$ *as* $n \to \infty$ *for all* $\omega \in \Omega$.

Proof Partition the segment $[0, n]$ into $n2^n$ equal intervals. Let

$$x_0 = 0, \quad x_1 = 2^{-n}, \quad \ldots, \quad x_{n2^n} = n,$$

denote the partition points, so that $x_{i+1} - x_i = 2^{-n}$. Put

$$F_i := \{\omega : x_i \le \xi(\omega) < x_{i+1}\}, \quad i = 1, 2, \ldots, n2^n - 1;$$

$$F_0 := \{0 \le \xi(\omega) < x_1\} \cup \{\xi(\omega) \ge n\}, \qquad \xi_n(\omega) := \sum_{i=0}^{n2^n-1} x_i \mathrm{I}_{F_i}(\omega) \le \xi(\omega).$$

The function $\xi_n(\omega)$ is clearly simple, $\xi_n(\omega) \le \xi_{n+1}(\omega) \le \xi(\omega)$ for all ω, and has the property that if $n > \xi(\omega)$ at a point $\omega \in \Omega$ then

$$0 \le \xi(\omega) - \xi_n(\omega) \le \frac{1}{2^n}.$$

The lemma is proved. □

Lemma A3.2.2 *Let* $\xi_n \uparrow \xi \ge 0$ *and* $\eta_n \uparrow \xi \ge 0$ *be sequences of simple functions. Then*

$$\lim_{n\to\infty} \int \xi_n \, d\mathbf{P} = \lim_{n\to\infty} \int \eta_n \, d\mathbf{P}.$$

Proof We verify that, for any m,

$$\int \xi_m \, d\mathbf{P} \le \lim_{n\to\infty} \int \eta_n \, d\mathbf{P}.$$

The function ξ_n is simple. Therefore it is bounded by some constant: $\xi_m \le c_m$. Hence, for any integer n and $\varepsilon > 0$,

$$\xi_m - \eta_n \le c_m \cdot \mathrm{I}_{\{\xi_m \ge \eta_n + \varepsilon\}} + \varepsilon.$$

This implies that

$$\mathbf{E}\xi_m \le c_m \mathbf{P}\{\xi_m \ge \eta_n + \varepsilon\} + \varepsilon + \mathbf{E}\eta_n.$$

The probability on the right-hand side vanishes as $n \to \infty$:

$$\mathbf{P}\{\xi_m \ge \eta_n + \varepsilon\} \le \mathbf{P}\{\xi \ge \eta_n + \varepsilon\} \to 0,$$

because η_n converges almost surely (and hence in probability) to ξ. Therefore $\mathbf{E}\xi_m \le \varepsilon + \lim_{n\to\infty} \mathbf{E}\eta_n$. Since ε is arbitrary,

$$\lim_{n\to\infty} \mathbf{E}\xi_n \le \lim_{n\to\infty} \mathbf{E}\eta_n.$$

Swapping $\{\xi_n\}$ and $\{\eta_n\}$, we obtain the converse inequality.

The lemma is proved. □

The assertions of Lemmas A3.2.1 and A3.2.2 make the following definitions consistent.

The *integral of a nonnegative measurable function* $\xi(\omega)$ (with respect to measure $\mathbf{P}$) is the quantity

$$\int \xi \, d\mathbf{P} = \lim_{n\to\infty} \int \xi_n \, d\mathbf{P}, \tag{A3.2.1}$$

where ξ_n is a sequence of simple functions such that $\xi_n \uparrow \xi$ as $n \to \infty$.

The integral $\int \xi \, d\mathbf{P}$ will also be denoted by $\mathbf{E}\xi$. We will say that the integral $\int \xi \, d\mathbf{P}$ *exists* and ξ is *integrable* if $\mathbf{E}\xi < \infty$.

The *integral of an arbitrary function* (assuming values of both signs) $\xi(\omega)$ (with respect to measure $\mathbf{P}$) is the quantity

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^-, \quad \xi^\pm := \max(0, \pm\xi),$$

which is defined when at least one of the values $\mathbf{E}\xi^\pm$ is finite. Otherwise $\mathbf{E}\xi$ is undefined. The integral $\mathbf{E}\xi$ *exists if and only if* $\mathbf{E}|\xi| < \infty$ *exists* (for $|\xi| = \xi^+ + \xi^-$). If $\mathbf{E}\xi$ exists then

$$\mathbf{E}(\xi; A) := \int_A \xi \, d\mathbf{P} = \mathbf{E}\xi \mathrm{I}_A$$

exists for any $A \in \mathfrak{F}$ as well.

Lemma A3.2.3 *If* $\mathbf{E}\xi$ *exists and* $B_n \in \mathfrak{F}$ *is a sequence of sets such that* $\mathbf{P}(B_n) \to 0$ *as* $n \to \infty$, *then*

$$\mathbf{E}(\xi; B_n) \to 0.$$

Proof For any sequence $|\xi_m| \uparrow |\xi|$ of simple functions and $A_m := \{|\xi| \le m\}$ one has

$$\mathbf{E}|\xi| \ge \lim_{m\to\infty} \mathbf{E}|\xi|\mathrm{I}_{A_m} \ge \lim_{m\to\infty} \mathbf{E}|\xi_m|\mathrm{I}_{A_m} = \mathbf{E}|\xi|,$$

since $|\xi_m|\mathrm{I}_{A_m} \uparrow |\xi|$. This implies that

$$\mathbf{E}|\xi| = \lim_{m\to\infty} \mathbf{E}|\xi|\mathrm{I}_{A_m} = \lim_{m\to\infty} \mathbf{E}\big(|\xi|;\, |\xi| \le m\big),$$

and hence, for any $\varepsilon > 0$, there exists an $m(\varepsilon)$ such that

$$\mathbf{E}|\xi| - \mathbf{E}\big(|\xi|;\, |\xi| \le m\big) < \varepsilon$$

for $m > m(\varepsilon)$. Consequently, for such m, one has

$$\mathbf{E}\big(|\xi|;\, B_n\big) = \mathbf{E}\big(|\xi|;\, \big\{|\xi| \le m\big\}B_n\big) + \mathbf{E}\big(|\xi|;\, \big\{|\xi| > m\big\}B_n\big) \le m\mathbf{P}(B_n) + \varepsilon,$$

and hence

$$\limsup_{n\to\infty} \mathbf{E}\big(|\xi|;\, B_n\big) \le \varepsilon.$$

The lemma is proved. □

Note that Lemma 6.1.2 somewhat extends Lemma A3.2.3.

Corollary A3.2.1 *If* $\mathbf{E}\xi$ *is well-defined* (*the values* $\pm\infty$ *not being excluded*) *and* $B_n \in \mathfrak{F}$ *is a sequence of sets such that* $\mathbf{P}(B_n) \to 1$ *as* $n \to \infty$, *then*

$$\mathbf{E}(\xi;\, B_n) \to \mathbf{E}\xi.$$

Proof If $\mathbf{E}\xi$ exists then the required assertion follows from Lemma A3.2.3.

Now let $\mathbf{E}\xi = \infty$. Then $\mathbf{E}\xi^- < \infty$ and $\mathbf{E}\xi^+ = \infty$, where $\xi^\pm = \max(0, \pm\xi)$. It follows that $\mathbf{E}(\xi^-;\, B_n) \to \mathbf{E}\xi^-$ as $n \to \infty$. We show that

$$\mathbf{E}\big(\xi^+;\, B_n\big) \to \infty. \tag{A3.2.2}$$

Let $A_k := \{\xi \in [2^{k-1}, 2^k)\}$, $k = 1, 2, \ldots;$ $p_k := \mathbf{P}(A_k)$. We can assume without loss of generality that all $p_k > 0$ (if this is not the case we can consider a subsequence k_j such that all $p_{k_j} > 0$). Since $\mathbf{E}\xi^+ \le 1 + \sum_{k=1}^\infty 2^k p_k$, we have $\sum_{k=1}^\infty 2^k p_k = \infty$. For a given $N > 1$, choose n large enough such that $\mathbf{P}(B_n A_k) > p_k/2$ for all $k \le N$. Then

$$\mathbf{E}\big(\xi^+;\, B_n\big) \ge \sum_{k=1}^N 2^{k-2} p_k,$$

where the right-hand side can be made arbitrarily large by an appropriate choice of N. This proves (A3.2.2). Since $\xi = \xi^+ - \xi^-$, the required convergence is proved.

The case $\mathbf{E}\xi = -\infty$ can be dealt with in the same way. The corollary is proved. □

3.2.3 Properties of Integrals

I1. *If sets* $A_j \in \mathfrak{F}$ *are disjoint and* $\bigcup_j A_j = \Omega$ *then*

$$\int \xi \, d\mathbf{P} = \sum_j \int_{A_j} \xi \, d\mathbf{P}. \tag{A3.2.3}$$

Proof It suffices to prove this relation for $\xi(\omega) \ge 0$. For simple functions equality (A3.2.3) is obvious, because

$$\int \xi \, d\mathbf{P} = \sum_k x_k \mathbf{P}(\xi = x_k) = \sum_j \sum_k x_k \mathbf{P}(\xi = x_k;\, A_j).$$

In the general case, using definition (A3.2.1) one gets

$$\begin{aligned} \int \xi \, d\mathbf{P} &= \lim_{n\to\infty} \int \xi_n \, d\mathbf{P} = \lim_{n\to\infty} \sum_j \int_{A_j} \xi_n \, d\mathbf{P} \\ &= \sum_j \lim_{n\to\infty} \int_{A_j} \xi_n \, d\mathbf{P} = \sum_j \int_{A_j} \xi \, d\mathbf{P}. \end{aligned} \tag{A3.2.4}$$

Swapping summation and passage to the limit is justified here, for by Lemma A3.2.3

$$\sum_{j=N}^{\infty} \int_{A_j} \xi_n \, d\mathbf{P} = \mathbf{E}\left(\xi_n;\, \bigcup_{j=N}^{\infty} A_j\right) \le \mathbf{E}\left(\xi;\, \bigcup_{j=N}^{\infty} A_j\right) \to 0$$

as $N \to \infty$ uniformly in n. □

I2.

$$\int (\xi + \eta) \, d\mathbf{P} = \int \xi \, d\mathbf{P} + \int \eta \, d\mathbf{P}.$$

Proof For simple functions this property is obvious. Hence, for $\xi \ge 0$ and $\eta \ge 0$, this property follows from the additivity of the limit.

In the general case we have ($\xi^\pm$ and $\eta^\pm$ are defined here as before)

$$\begin{aligned} \int (\xi + \eta) \, d\mathbf{P} &= \int \left(\xi^+ + \eta^+\right) d\mathbf{P} - \int \left(\xi^- + \eta^-\right) d\mathbf{P} \\ &= \int \xi^+ \, d\mathbf{P} - \int \xi^- \, d\mathbf{P} + \int \eta^+ \, d\mathbf{P} - \int \eta^- \, d\mathbf{P} = \int \xi \, d\mathbf{P} + \int \eta \, d\mathbf{P}. \end{aligned}$$ □

I3. *If* c *is an arbitrary constant, then*

$$\int c\xi \, d\mathbf{P} = c \int \xi \, d\mathbf{P}.$$

I4. *If* $\xi \le \eta$, *then* $\int \xi\, d\mathbf{P} \le \int \eta\, d\mathbf{P}$.

The proof of properties I3 and I4 is obvious. Since

$$\int \xi\, d\mathbf{P} = \mathbf{E}\xi,$$

we can write down properties I1–I4 in terms of expectations as follows:

I1. $\mathbf{E}\xi = \sum_j \mathbf{E}(\xi; A_j)$ *if* A_j *are disjoint and* $\bigcup_j A_j = \Omega$.
I2. $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$.
I3. $\mathbf{E}a\xi = a\mathbf{E}\xi$.
I4. $\mathbf{E}\xi \le \mathbf{E}\eta$, if $\xi \le \eta$.

Note also the following properties of integrals which easily follow from I1–I4.

I5. $|\mathbf{E}\xi| \le \mathbf{E}|\xi|$.
I6. *If* $c_1 \le \xi \le c_2$, *then* $c_1 \le \mathbf{E}\xi \le c_2$.
I7. *If* $\xi \ge 0$ *and* $\mathbf{E}\xi = 0$, *then* $\mathbf{P}(\xi = 0) = 1$.

This property follows from the Chebyshev inequality: $\mathbf{P}(\xi \ge \varepsilon) \le \mathbf{E}\xi/\varepsilon = 0$ for any $\varepsilon > 0$.

I8. *If* $\mathbf{P}(\xi = \eta) = 1$ *and* $\mathbf{E}\xi$ *exists then* $\mathbf{E}\xi = \mathbf{E}\eta$.

Indeed,

$$\mathbf{E}\eta = \lim_{n\to\infty} \mathbf{E}\big(\eta; |\eta| \le n\big) = \lim_{n\to\infty} \mathbf{E}\big(\xi; |\xi| \le n\big) = \mathbf{E}\xi.$$

3.3 Further Properties of Integrals

3.3.1 Convergence Theorems

A number of convergence theorems were proved in Sect. 6.1. One of them was the dominated convergence theorem (Corollary 6.1.3):

If $\xi_n \xrightarrow{p} \xi$ *as* $n \to \infty$ *and* $|\xi_n| \le \eta$, $\mathbf{E}\eta < \infty$, *then the expectation* $\mathbf{E}\xi$ *exists and* $\mathbf{E}\xi_n \to \mathbf{E}\xi$.

Now we will present some further useful assertions concerning convergence of integrals.

Theorem A3.3.1 (Monotone convergence) *If* $0 \le \xi_n \uparrow \xi$, *then* $\mathbf{E}\xi = \lim_{n\to\infty} \mathbf{E}\xi_n$.

Proof In addition to Corollary 6.1.3, here we only need to prove that $\mathbf{E}\xi_n \to \infty$ if $\mathbf{E}\xi = \infty$. Put $\xi_n^N := \min(\xi_n, N)$ and $\xi^N := \min(\xi, N)$. Then clearly $\xi_n^N \uparrow \xi^N$ as $n \to \infty$, and $\mathbf{E}\xi_n^N \uparrow \mathbf{E}\xi^N$. Therefore the value $\mathbf{E}\xi_n^N \le \mathbf{E}\xi_n$ can be made arbitrarily large by choosing appropriate n and N. The theorem is proved. □

These theorems can be generalised in the following way. To make the extension of the convergence theorems to the case of integrals with respect to signed measures in Sect. 3.4 more convenient, we will now write $\mathbf{E}\xi$ in the form of the integral $\int \xi\, d\mathbf{P}$.

Theorem A3.3.2 (Fatou–Lebesgue) *Let η and ζ be integrable. If $\xi_n \le \eta$ then*

$$\limsup_{n\to\infty} \int \xi_n\, d\mathbf{P} \le \int \limsup_{n\to\infty} \xi_n\, d\mathbf{P}. \qquad \text{(A3.3.1)}$$

If $\xi_n \ge \zeta$ then

$$\liminf_{n\to\infty} \int \xi_n\, d\mathbf{P} \ge \int \liminf_{n\to\infty} \xi_n\, d\mathbf{P}. \qquad \text{(A3.3.2)}$$

If $\xi_n \uparrow \xi$ and $\xi_n \ge \zeta$, or $\xi_n \xrightarrow{a.e.} \xi$ and $\zeta \le \xi_n \le \eta$, then

$$\lim_{n\to\infty} \int \xi_n\, d\mathbf{P} = \int \xi\, d\mathbf{P}. \qquad \text{(A3.3.3)}$$

Proof We prove for instance (A3.3.2). Assume without loss of generality that $\zeta \equiv 0$. In this case, as $n \to \infty$,

$$\xi \ge \eta_n := \inf_{k\ge n} \xi_k \uparrow \liminf_{k\to\infty} \xi_k, \qquad \eta_n \ge 0,$$

and by the monotone convergence theorem

$$\liminf_{n\to\infty} \int \xi_n\, d\mathbf{P} \ge \lim_{n\to\infty} \int \eta_n\, d\mathbf{P} = \int \liminf_{n\to\infty} \xi_n\, d\mathbf{P}.$$

Applying (A3.3.2) to the sequence $\eta - \xi_n$ we obtain (A3.3.1); (A3.3.3) follows from the previous theorems. The theorem is proved. □

3.3.2 Connection to Integration with Respect to a Measure on the Real Line

Let $g(x)$ be a Borel function given on the real line $\mathbb{R}$ (if $\mathfrak{B}$ is the σ-algebra of Borel sets on the line and $B \in \mathfrak{B}$, then $\{x : g(x) \in B\} \in \mathfrak{B}$). If ξ is a random variable then $\eta := g(\xi(\omega))$ will clearly also be a random variable. As we saw in Sect. 3.2, a random variable ξ induces the probability space $\langle \mathbb{R}, \mathfrak{B}, \mathbf{F}_\xi \rangle$ with measure $\mathbf{F}_\xi$ on the line such that $\mathbf{F}_\xi(B) = \mathbf{P}(\xi \in B)$. Therefore one can speak about integrals with respect to that measure.

Theorem A3.3.3 *If $\eta = g(\xi(\omega))$ and $\mathbf{E}\eta$ exists, then*

$$\mathbf{E}\eta = \int_\Omega \eta\, d\mathbf{P} = \int_{\mathbb{R}} g(x) \mathbf{F}_\xi(dx)$$

(*on the right-hand side we used a somewhat different notation for $\int g\, d\mathbf{F}_\xi$*).

Proof Let first $g(x) = \mathrm{I}_B(x)$ be the indicator of a set $B \in \mathfrak{B}$. Then $\eta = g(\xi(\omega)) = \mathrm{I}_{\{\xi \in B\}}(\omega)$ and $\mathbf{E}\eta = \mathbf{P}(\xi \in B)$. Therefore

$$\int g(x)\mathbf{F}_\xi(dx) = \int \mathrm{I}_B(x)\mathbf{F}_\xi(dx) = \mathbf{F}_\xi(B) = \mathbf{P}(\xi \in B) = \mathbf{E}\eta.$$

Using the properties of the integral it is easy to establish that the assertion of the theorem holds for simple functions g. Passing to the limit extends that assertion to bounded functions. Now let $g \geq 0$. If the function $g(\xi)\mathrm{I}_B(\xi) = \eta(\omega)\mathrm{I}_{\{\xi \in B\}}(\omega)$ is bounded, then

$$\int_B g(x)\,\mathbf{F}_\xi(dx) = \mathbf{E}(\eta;\, \xi \in B).$$

Therefore

$$\int_{\{g \leq n\}} g\,d\mathbf{F}_\xi = \mathbf{E}(\eta;\, \eta \leq n).$$

Passing to the limit as $n \to \infty$ we get the assertion of the theorem. Considering the case when g takes values of both signs does not create any difficulties. The theorem is proved. □

Introducing the notation

$$F_\xi(x) = \mathbf{P}(\xi < x),$$

we can also consider, along with the integral just discussed,

$$\int_{\mathbb{R}} g(x)\mathbf{F}_\xi(dx), \tag{A3.3.4}$$

the Riemann–Stieltjes integral

$$\int g(x)\,dF_\xi(x), \tag{A3.3.5}$$

the definition of which was given in Sect. 3.6. It was also shown there that, for *continuous* functions $g(x)$, these integrals coincide. Moreover, we discussed in Sect. 3.6 some other conditions for these integrals to coincide.

Also recall that if

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,dt$$

and the functions $g(x)$ and $f_\xi(x)$ are Riemann integrable, then integrals (A3.3.4) and (A3.3.5) coincide with the Riemann integral

$$\int g(x) f_\xi(x)\,dx.$$

3.3.3 Product Measures and Iterated Integrals

Consider a two-dimensional random variable $\zeta = (\xi, \eta)$ given on $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$. The random variables ξ and η induce a sample probability space $\langle \mathbb{R}^2, \mathfrak{B}^2, \mathbf{F}_{\xi,\eta} \rangle$ with the measure $\mathbf{F}_{\xi,\eta}$ given on elements of the σ-algebra $\mathfrak{B}^2$ of Borel sets on the plane (the σ-algebra generated by rectangles) and such that

$$\mathbf{F}_{\xi,\eta}(A \times B) = \mathbf{P}(\xi \in A, \eta \in B).$$

Here $A \times B$ is the set of points (x, y) for which $x \in A$ and $y \in B$. If $g(x, y)$ is a Borel function ($\{(x, y) : g(x, y) \in B\} \in \mathfrak{B}^2$ for each $B \in \mathfrak{B}$), then it easily follows from the above that

$$\mathbf{E}g(\xi, \eta) = \int_{\mathbb{R}^2} g(x, y)\mathbf{F}_{\xi,\eta}(dx\,dy), \qquad \text{(A3.3.6)}$$

since both integrals are equal to $\int_{\mathbb{R}} x\mathbf{F}_\theta(dx)$ for $\theta = g(\xi, \eta)$.

Now let ξ and η be independent random variables, i.e.

$$\mathbf{P}(\xi \in A,\ \eta \in B) = \mathbf{P}(\xi \in A)\mathbf{P}(\eta \in B)$$

for any $A, B \in \mathfrak{B}$.

Theorem A3.3.4 (Fubini's theorem on iterated integrals) *If $g(x, y) \ge 0$ is a Borel function and ξ and η are independent, then*

$$\mathbf{E}g(\xi, \eta) = \mathbf{E}\big[\mathbf{E}g(x, \eta)|_{x=\xi}\big].$$

For arbitrary Borel functions $g(x, y)$ the above equality holds if $\mathbf{E}g(\xi, \eta)$ exists.

This very assertion we stated in Chap. 3 in the form

$$\int g(x, y)\mathbf{F}_{\xi,\eta}(dx\,dy) = \int \bigg[\int g(x, y)\mathbf{F}_\eta(dy)\bigg]\mathbf{F}_\xi(dx). \qquad \text{(A3.3.7)}$$

We will need the following.

Lemma A3.3.1 1. *The section*

$$B_x := \big\{y : (x, y) \in B\big\}$$

of any set $B \in \mathfrak{B}^2$ is measurable: $B_x \in \mathfrak{B}$.

2. *The section $g_x(y) = g(x, y)$ of any Borel function g ($\mathfrak{B}^2$-measurable) is a Borel function.*

3. *The integral*

$$\int g(x, y)\mathbf{F}_\eta(dy) \qquad \text{(A3.3.8)}$$

of a Borel function g is a Borel function of x.

Proof 1. Let $\mathcal{K}_1$ be the class of all sets from $\mathfrak{B}^2$ of which all x-sections are measurable. It is evident that $\mathcal{K}_1$ contains all rectangles $B = B_{(1)} \times B_{(2)}$, where $B_{(1)} \in \mathfrak{B}$ and $B_{(2)} \in \mathfrak{B}$. Moreover, $\mathcal{K}_1$ is a σ-algebra. Indeed, consider for example the set $B = \bigcup_k B^{(k)}$ where $B^{(k)} \in \mathcal{K}_1$. The operation $\bigcup$ on the sets $B^{(k)}$ leads to the same operation on their sections, so that $B_x = \bigcup_k B_x^{(k)} \in \mathfrak{B}$. For the other operations ($\cap$ and taking complements) the situation is similar. Thus, $\mathcal{K}_1$ is a σ-algebra containing all rectangles. This means that $\mathfrak{B}^2 \subset \mathcal{K}_1$.

2. For $B \in \mathfrak{B}$, one has

$$\begin{aligned} g_x^{-1}(B) &= \{y : g_x(y) \in B\} = \{y : g(x, y) \in B\} \\ &= \{y : (x, y) \in g^{-1}(B)\} = [g^{-1}(B)]_x \in \mathfrak{B}. \end{aligned}$$

3. Integral (A3.3.8) is, as a function of x, the result of passing to the limit in a sequence of measurable functions, and hence is measurable itself. The lemma is proved. □

Proof of Theorem A3.3.4 First we prove (A3.3.7) in the case where $g(x, y) = \mathrm{I}_B(x, y)$, so that the theorem turns into the formula for consecutive computation of the measure of the set $B \in \mathfrak{B}^2$:

$$\mathbf{P}\big((\xi, \eta) \in B\big) = \int \mathbf{F}_\eta\big((x, y) \in B\big)\mathbf{F}_\xi(dx) = \int \mathbf{F}_\eta(B_x)\mathbf{F}_\xi(dx). \tag{A3.3.9}$$

We introduce the set function

$$\mathbf{Q}(B) := \int \mathbf{F}_\eta(B_x)\mathbf{F}_\xi(dx).$$

Clearly, $\mathbf{Q}(B) \ge 0$ and $\mathbf{Q}(\varnothing) = 0$. Further, if $B = \bigcup_k B^{(k)}$ and $B^{(k)}$ are disjoint, then $B_x = \bigcup_k B_x^{(k)}$ and $B_x^{(k)}$ are also disjoint, and

$$\mathbf{Q}(B) = \int \mathbf{F}_\eta\Big(\bigcup_k B_x^{(k)}\Big)\mathbf{F}_\xi(dx) = \sum_k \int \mathbf{F}_\eta\big(B_x^{(k)}\big)\mathbf{F}_\xi(dx) = \sum_k \mathbf{Q}\big(B^{(k)}\big).$$

This means that $\mathbf{Q}(B)$ is a measure.

The measure $\mathbf{Q}(B)$ coincides with $\mathbf{F}_{\xi,\eta}(B) = \mathbf{P}((\xi, \eta) \in B)$ on rectangles $B = B_{(1)} \times B_{(2)}$. Indeed, for rectangles,

$$B_x = \begin{cases} B_{(2)} & \text{for } x \in B_{(1)}, \\ \varnothing & \text{for } x \notin B_{(1)}, \end{cases}$$

and

$$\begin{aligned} \mathbf{P}\big((\xi, \eta) \in B\big) &= \mathbf{F}_\xi(B_{(1)})\mathbf{F}_\eta(B_{(2)}) \\ &= \int_{B_{(1)}} \mathbf{F}_\eta(B_{(2)})\mathbf{F}_\xi(dx) = \int \mathbf{F}_\eta(B_x)\mathbf{F}_\xi(dx) = \mathbf{Q}(B). \end{aligned}$$

This means that the measures $\mathbf{Q}$ and $\mathbf{F}_{\xi,\eta}$ coincide on the algebra generated by rectangles. By the measure extension theorem we obtain that $\mathbf{Q} = \mathbf{F}_{\xi,\eta}$.

We have proved (A3.3.9). This implies that Fubini's theorem holds for simple functions $g_N = \sum_{j=1}^{N} c_j \mathrm{I}_{A_j}$, because

$$\begin{aligned}\mathbf{E}g_N(\xi,\eta) &= \sum_{j=1}^{N} c_j \mathbf{E}\mathrm{I}_{A_j}(\xi,\eta) \\ &= \sum_{j=1}^{N} c_j \int \mathbf{E}\mathrm{I}_{A_j}(x,\eta)\mathbf{F}_\xi(dx) = \int \mathbf{E}g_N(x,\eta)\mathbf{F}_\xi(dx)\end{aligned} \quad \text{(A3.3.10)}$$

Now if $g \ge 0$ is an arbitrary Borel function then there exists a sequence of simple functions $g_N \uparrow g$ and, as in (A3.2.1), it remains to pass to the limit:

$$\begin{aligned}\mathbf{E}g(\xi,\eta) &= \lim_{N\to\infty} \mathbf{E}g_N(\xi,\eta) \\ &= \lim_{N\to\infty} \int \mathbf{E}g_N(\xi,\eta)\mathbf{F}_\xi(dx) = \int \mathbf{E}g_N(\xi,\eta)\mathbf{F}_\xi(dx).\end{aligned}$$

For an arbitrary function g one has to use the representation $g = g^+ - g^-$, $g^+ \ge 0$, $g^- \ge 0$. The theorem is proved. □

Remark A3.1 We see from the proof of the theorem that the random variables ξ and η do not need to be scalar. The assertion remains true in a more general form (see property 5A in Sect. 4.8) and, in particular, for vector-valued ξ and η.

3.4 The Integral with Respect to an Arbitrary Measure

If μ is a finite measure on $\langle \Omega, \mathfrak{F}\rangle$, $\boldsymbol{\mu}(\Omega) < \infty$, then the definition of the integral $\int \xi\, d\boldsymbol{\mu}$ with respect to the measure $\boldsymbol{\mu}$ does not differ from that of the integral with respect to a probability measure (one could just put $\int_A \xi\, d\boldsymbol{\mu} = \boldsymbol{\mu}(\Omega)\int_A \xi\, d\mathbf{P}$, where $\mathbf{P}(B) = \boldsymbol{\mu}(B)/\boldsymbol{\mu}(\Omega)$ is a probability distribution). If $\boldsymbol{\mu}$ is σ-finite and $\boldsymbol{\mu}(\Omega) = \infty$, then the situation is somewhat more complicated, although it can still be reduced to the already used constructions. First we will make several preliminary remarks.

Let $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ be a probability space and $f = f(\omega) \ge 0$ an a.e. finite nonnegative measurable function (i.e., a random variable). Consider the set function

$$\boldsymbol{\mu}(A) := \int_A f\, d\mathbf{P}. \quad \text{(A3.4.1)}$$

If f is integrable ($\boldsymbol{\mu}(\Omega) < \infty$) then $\boldsymbol{\mu}(A)$ is a finite σ-additive measure (see property I1) satisfying conditions (1)–(3) of Sect. 3.1 of the present appendix. In other words, $\boldsymbol{\mu}$ is a finite measure on $\langle \Omega, \mathfrak{F} \rangle$. But if f is not integrable, then $\boldsymbol{\mu}$ is a σ-finite measure, which immediately follows from the representation

$$\boldsymbol{\mu}(A) = \sum_{k=1}^{\infty} \int_{A \cap \{k-1 \le f < k\}} f \, d\mathbf{P}$$

(the integrals in the sum that are equal to $\int_A f \mathrm{I}_{(k-1 \le f < k)} \, d\mathbf{P}$ are clearly finite measures).

Thus, the integral of the form (A3.4.1) is a measure for any distribution $\mathbf{P}$ and function $f \ge 0$. It turns out that the following assertion, converse in a certain sense to the above, also holds.

Lemma A3.4.1 *For any measure* $\boldsymbol{\mu}$ *on* $\langle \Omega, \mathfrak{F} \rangle$, *there exists a distribution* $\mathbf{P}$ *on that space and a measurable function* $f > 0$ *such that representation* (A3.4.1) *holds.*

Thus, any measure can be represented as an integral with respect to a probability measure (i.e., in the form $\mathbf{E}(f; A)$ for the respective function f and distribution $\mathbf{P}$).

Proof Let $\boldsymbol{\mu}$ be a σ-finite measure on $\langle \Omega, \mathfrak{F} \rangle$, and sets $B_j \in \mathfrak{F}$, $j = 1, 2, \ldots$, possess the properties $\bigcup_{j=1}^{\infty} B_j = \Omega$, $B_i B_j = \varnothing$ for $i \ne j$, and $\boldsymbol{\mu}(B_j) < \infty$. Put

$$\mathbf{P}(A) := \sum_{k=1}^{\infty} \frac{\boldsymbol{\mu}(AB_k)}{2^k \boldsymbol{\mu}(B_k)}. \tag{A3.4.2}$$

Obviously, $\mathbf{P}(\Omega) = 1$ and $\mathbf{P}$ is a measure. Further, if $A \subset B_k$ then

$$\boldsymbol{\mu}(A) = 2^k \boldsymbol{\mu}(B_k) \mathbf{P}(A).$$

This means that we should put $f(\omega) := 2^k \boldsymbol{\mu}(B_k)$ for $\omega \in B_k$. Then the set function

$$\boldsymbol{\lambda}(A) := \int_A f \, d\mathbf{P} = \int_{\Omega} f \mathrm{I}_A \, d\mathbf{P}$$

will coincide with $\boldsymbol{\mu}(A)$:

$$\begin{aligned} \boldsymbol{\lambda}(A) &= \sum_{k=1}^{\infty} 2^k \boldsymbol{\mu}(B_k) \mathbf{P}(AB_k) \\ &= \sum_{k=1}^{\infty} 2^k \boldsymbol{\mu}(B_k) \sum_{j=1}^{\infty} \frac{\boldsymbol{\mu}(AB_k B_j)}{2^j \boldsymbol{\mu}(B_j)} = \sum_{k=1}^{\infty} \boldsymbol{\mu}(AB_k) = \boldsymbol{\mu}(A). \end{aligned}$$

The lemma is proved. □

Besides the required assertion, we also obtain that in representation (A3.4.1) the range of values of the function f can be assumed without loss of generality to be countable.

The function f for which equality (A3.4.1) holds is called the *density* of the measure $\boldsymbol{\mu}$ with respect to $\mathbf{P}$ (or *Radon–Nikodym derivative* of the measure $\boldsymbol{\mu}$ with respect to $\mathbf{P}$) and is denoted by $d\boldsymbol{\mu}/d\mathbf{P}$. It is evident that alteration of the function $f = d\boldsymbol{\mu}/d\mathbf{P}$ on a set of zero $\mathbf{P}$-measure leaves the equality (A3.4.1) unchanged.

Now let $\boldsymbol{\mu}$ and $\mathbf{P}$ be two *given arbitrary measures*. The question of under what conditions these two measures $\boldsymbol{\mu}$ and $\mathbf{P}$ could be related by (A3.4.1) and whether the function f is determined uniquely thereby (up to values on a set of zero $\mathbf{P}$-measure) is rather important for probability theory. (We stress that, in the preceding considerations, the measure $\mathbf{P}$ was constructed in a special way from the measure $\boldsymbol{\mu}$, or vice versa.) Answers to these questions are given by the Radon–Nikodym theorem to be discussed in the next section.

Now, using the simple assertion of Lemma A3.4.1 we have just proved, we will give the definition of the integral with respect to an arbitrary measure $\boldsymbol{\mu}$.

Let $\boldsymbol{\mu}$ be an arbitrary σ-finite measure on $\langle \Omega, \mathfrak{F} \rangle$ and $\xi \geq 0$ a $\mathfrak{F}$-measurable function.

The *integral* $\int_A \xi \, d\boldsymbol{\mu}$ over a set $A \in \mathfrak{F}$ of the function $\xi \geq 0$ with respect to the measure $\boldsymbol{\mu}$ is the integral

$$\int_A \xi \, d\boldsymbol{\mu} = \int_A \left(\xi \frac{d\boldsymbol{\mu}}{d\mathbf{P}} \right) d\mathbf{P} \tag{A3.4.3}$$

with respect to any distribution $\mathbf{P}$ satisfying equality (A3.4.1) (for example, with respect to measure (A3.4.2)).

This definition is *consistent* because it does not depend on the choice of $\mathbf{P}$. Indeed, for simple functions ξ ($\xi(\omega) = x_k$ for $\omega \in F_k$),

$$\int_A \xi \, d\boldsymbol{\mu} = \sum_k x_k \int_A \frac{d\boldsymbol{\mu}}{d\mathbf{P}} \mathrm{I}_{B_k} \, d\mathbf{P} = \sum_k x_k \int_{AB_k} \frac{d\boldsymbol{\mu}}{d\mathbf{P}} \, d\mathbf{P} = \sum_k x_k \boldsymbol{\mu}(AB_k).$$

If now $\xi \geq 0$ is an arbitrary function, then by the monotone convergence theorem the integral $\int_A \xi \, d\boldsymbol{\mu}$ is equal to

$$\lim_{n\to\infty} \int_A \xi^{(n)} \frac{d\boldsymbol{\mu}}{d\mathbf{P}} \, d\mathbf{P} = \lim_{n\to\infty} \int_A \xi^{(n)} d\boldsymbol{\mu},$$

where $\xi^{(n)} \uparrow \xi$ is a sequence of simple functions which converge monotonically to ξ (see Lemma A3.2.1). In both cases, the result does not depend on the choice of $\mathbf{P}$.

The integral

$$\int_A \xi \, d\boldsymbol{\mu}$$

of an arbitrary measurable function ξ is defined by

$$\int_A \xi \, d\boldsymbol{\mu} = \int_A \xi^+ d\boldsymbol{\mu} - \int_A \xi^- d\boldsymbol{\mu},$$

when both expressions on the right-hand side are finite. (In that case one says that the integral $\int_A \xi\, d\boldsymbol{\mu}$ exists.) Here, as before, $\xi^+ = \max(0, \xi) \geq 0$ and $\xi^- = \max(0, -\xi) \geq 0$, so that $\xi = \xi^+ - \xi^-$.

Thus we see that the above definition of the integral with respect to an arbitrary measure is essentially equivalent to the construction used in Sect. 3.2 of the present appendix. However, the definition in the form (A3.4.3) saves us from the necessity of repeating what we have already done (and now in a more complex setting) and enables one to transfer all the properties of the integrals $\int \xi\, d\mathbf{P}$ to the general case. We will list the basic properties preserving the existing numeration.

I1. $\int \xi\, d\boldsymbol{\mu} = \sum_j \int_{A_j} \xi\, d\boldsymbol{\mu}$ if A_j are disjoint and $\bigcup_j A_j = \Omega$.
I2. $\int (\xi + \eta)\, d\boldsymbol{\mu} = \int \xi\, d\boldsymbol{\mu} + \int \eta\, d\boldsymbol{\mu}$.
I3. $\int a\xi\, d\boldsymbol{\mu} = a \int \xi\, d\boldsymbol{\mu}$.
I4. $\xi\, d\boldsymbol{\mu} \leq \int \eta\, d\boldsymbol{\mu}$ if $\xi \leq \eta$.
I5. $|\int \xi\, d\boldsymbol{\mu}| \leq \int |\xi|\, d\boldsymbol{\mu}$.
I6. If $c_1 \leq \xi(\omega) \leq c_2$ for $\omega \in A$, then $c_1 \boldsymbol{\mu}(A) \leq \int_A \xi\, d\boldsymbol{\mu} \leq c_2 \boldsymbol{\mu}(A)$.
I7. If $\xi \geq 0$ and $\int \xi\, d\boldsymbol{\mu} = 0$, then $\boldsymbol{\mu}(\xi > 0) = 0$.
I8. If $\boldsymbol{\mu}(\xi \neq \eta) = 0$, then $\int \xi\, d\boldsymbol{\mu} = \int \eta\, d\boldsymbol{\mu}$.

It is clear that all the convergence theorems remain valid as well.

Theorem A3.4.1 (The dominated convergence theorem) *Let* $|\xi_n| \leq \eta$ *and* $\int \eta\, d\boldsymbol{\mu}$ *exist. If* $\xi_n \xrightarrow{\boldsymbol{\mu}}$ *or* $\xi_n \to \xi$ *a.e. as* $n \to \infty$ *then*

$$\int \xi_n\, d\boldsymbol{\mu} \to \int \xi\, d\boldsymbol{\mu}.$$

Theorem A3.4.2 (The monotone convergence theorem) *If* $0 \leq \xi_n \uparrow \xi$ *as* $n \to \infty$ *then*

$$\int \xi_n\, d\boldsymbol{\mu} \to \int \xi\, d\boldsymbol{\mu}.$$

Theorem A3.4.3 (Fatou–Lebesgue) *The statement and proof of this theorem is obtained from those of Theorem A3.3.2 by replacing* $\mathbf{P}$ *with* $\boldsymbol{\mu}$.

In conclusion we note that if $\Omega = \mathbb{R} = (-\infty, \infty)$, $\mathfrak{F} = \mathfrak{B}$ is the σ-algebra of Borel sets, $\boldsymbol{\mu}$ is the Lebesgue measure, and the function $g(x)$ is continuous, then the integral $\int_{[a,b]} g(x)\, d\boldsymbol{\mu}(x)$ coincides with the Riemann integral $\int_a^b g(x)\, dx$. This follows from the preceding remarks in part 2 of Sect. 3.3 of this appendix.

3.5 The Lebesgue Decomposition Theorem and the Radon–Nikodym Theorem

We return to a question that has already been asked in the previous section. Under what conditions on measures $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ given on $\langle \Omega, \mathfrak{F} \rangle$ can the measure $\boldsymbol{\mu}$ be represented as

$$\boldsymbol{\mu}(A) = \int_A f\, d\boldsymbol{\lambda}?$$

We do not assume here that $\boldsymbol{\lambda}$ is a probability measure.

Definition A3.5.1 A measure $\boldsymbol{\mu}$ is said to be *absolutely continuous* with respect to a measure $\boldsymbol{\lambda}$ (we write $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$) if, for any A such that $\boldsymbol{\lambda}(A) = 0$, one has $\boldsymbol{\mu}(A) = 0$.

Definition A3.5.2 A set $N_{\boldsymbol{\mu}}$ is said to be a *support*[1] of measure $\boldsymbol{\mu}$ if $\boldsymbol{\mu}(\Omega - N_{\boldsymbol{\mu}}) = 0$.

Definition A3.5.2 specifies a rather wide class of sets which can be called the support of the measure $\boldsymbol{\mu}$ when $\boldsymbol{\mu}$ is concentrated on a part of the space Ω. If $\Omega = \mathbb{R}$ is the real line (and in some other cases as well), one can use another definition which specifies a unique set for each measure. Consider the collection of all intervals $(a,b) \subset \mathbb{R}$ with rational endpoints a and b. This collection is countable. Remove from $\Omega = \mathbb{R}$ all such intervals for which $\boldsymbol{\mu}((a,b)) = 0$. The remaining set (which is measurable) is called the *support of the measure* $\boldsymbol{\mu}$.

Definition A3.5.3 One says that a measure $\boldsymbol{\mu}$ is *singular*with respect to $\boldsymbol{\lambda}$ if there exists a support $N_{\boldsymbol{\lambda}}$ of the measure $\boldsymbol{\lambda}$ such that $\boldsymbol{\mu}(N_{\boldsymbol{\lambda}}) = 0$. Or, which is the same, if there exists a support $N_{\boldsymbol{\mu}}$ of the measure $\boldsymbol{\mu}$ such that $\boldsymbol{\lambda}(N_{\boldsymbol{\mu}}) = 0$.

The last definition, in contrast to Definition A3.5.1, is symmetric, so one can speak about *mutually singular measures* $\boldsymbol{\mu}$ *and* $\boldsymbol{\lambda}$ (this relation is often written as $\boldsymbol{\mu} \perp \boldsymbol{\lambda}$).

Theorem A3.5.1 (Radon–Nikodym) *A necessary and sufficient condition for the absolute continuity* $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$ *is that there exists a function* f *unique up to* $\boldsymbol{\lambda}$-*equivalence* (*i.e.*, *up to values on a set of zero* $\boldsymbol{\lambda}$-*measure*) *such that*[2]

$$\boldsymbol{\mu}(A) = \int_A f\, d\boldsymbol{\lambda}.$$

As we have already noted, the function f is called the *Radon–Nikodym derivative* $d\boldsymbol{\mu}/d\boldsymbol{\lambda}$ of the measure $\boldsymbol{\mu}$ with respect to $\boldsymbol{\lambda}$ (or *density* of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\lambda}$).

Since sufficiency in the assertion of the theorem is obvious, we will obtain the Radon–Nikodym theorem as a consequence of the following Lebesgue decomposition theorem.

[1]The conventional definition of support refers to the case when Ω is a topological space. Then the support of $\boldsymbol{\mu}$ is the smallest closed set such that its complement is of $\boldsymbol{\mu}$-measure zero.

[2]This equality is sometimes adopted as a definition of absolute continuity.

Theorem A3.5.2 (Lebesgue) *Let $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ be two σ-finite measures given on $\langle \Omega, \mathfrak{F} \rangle$. There exists a unique decomposition of the measure $\boldsymbol{\mu}$ into two components $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_s$ such that*

$$\boldsymbol{\mu}_a \prec \boldsymbol{\lambda}, \qquad \boldsymbol{\mu}_s \perp \boldsymbol{\lambda}.$$

Moreover, there exists a function f, unique up to $\boldsymbol{\lambda}$-equivalence, such that

$$\boldsymbol{\mu}_a(A) = \int_A f \, d\boldsymbol{\lambda}.$$

It is obvious that if $\boldsymbol{\mu} \prec \boldsymbol{\lambda}$ then $\boldsymbol{\mu}_s = 0$, and the Lebesgue theorem then implies the Radon–Nikodym theorem.

Proof Since $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are σ-finite, there exist increasing sequences of sets $\Omega_n^{\boldsymbol{\mu}}$ and $\Omega_n^{\boldsymbol{\lambda}}$ such that

$$\boldsymbol{\mu}(\Omega_n^{\boldsymbol{\mu}}) < \infty, \qquad \boldsymbol{\lambda}(\Omega_n^{\boldsymbol{\lambda}}) < \infty, \qquad \bigcup_n \Omega_n^{\boldsymbol{\mu}} = \Omega, \qquad \bigcup_n \Omega_n^{\boldsymbol{\lambda}} = \Omega.$$

Putting $\Omega_n := \Omega_n^{\boldsymbol{\mu}} \cap \Omega_n^{\boldsymbol{\lambda}}$ we obtain a sequence of sets increasing to Ω for which

$$\boldsymbol{\mu}(\Omega_n) < \infty, \qquad \boldsymbol{\lambda}(\Omega_n) < \infty.$$

If we prove the decomposition theorem for restrictions of the measures $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ to $\langle B_n, \mathfrak{F}_n \rangle$, where $B_n = \Omega_{n+1} - \Omega_n$ and $\mathfrak{F}_n$ is formed by sets $B_n A$, $A \in \mathfrak{F}$, we will thereby prove it for the whole Ω. It will suffice to take $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_s$ to be the sums of the respective components for each of the restrictions. This remark means that we can consider the case of finite measures only.

Thus let $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ be finite measures.

(a) Let $\mathcal{F}$ be the class of functions $f \geq 0$ such that

$$\int_A f \, d\boldsymbol{\lambda} \leq \boldsymbol{\mu}(A) \quad \text{for all } A \in \mathfrak{F} \tag{A3.5.1}$$

(the class $\mathcal{F}$ is non-empty, for the function $f = 0$ belongs to $\mathcal{F}$). Set

$$\alpha := \sup_{f \in \mathcal{F}} \int_\Omega f \, d\boldsymbol{\lambda} \leq \boldsymbol{\mu}(\Omega) < \infty$$

and choose a sequence f_n such that, as $n \to \infty$,

$$\int f_n \, d\boldsymbol{\lambda} \to \alpha.$$

Put $\widehat{f}_n := \max(f_1, \ldots, f_n)$. Then clearly $\widehat{f}_n \uparrow \widehat{f} := \sup f_n$ and by the monotone convergence theorem

$$\int_A \widehat{f}_n \, d\boldsymbol{\lambda} \to \int_A \widehat{f} \, d\boldsymbol{\lambda}. \tag{A3.5.2}$$

We now show that $\widehat{f} \in \mathcal{F}$, i.e., that (A3.5.1) holds for $\widehat{f}$. To this end, it suffices by virtue of (A3.5.2) to notice that $\widehat{f_n} \in \mathcal{F}$. Let A_k, $k = 1, \ldots, n$ be disjoint sets on which $\widehat{f_n} = f_k$. Then $\bigcup_{k=1}^n A_k = \Omega$ and

$$\int_A \widehat{f_n}\, d\boldsymbol{\lambda} = \sum_{k=1}^n \int_{AA_k} f_k\, d\boldsymbol{\lambda} \le \sum_{k=1}^n \boldsymbol{\mu}(AA_k) = \boldsymbol{\mu}(A).$$

Thus, for the "maximum" element f' of $\mathcal{F}$, (A3.5.1) also holds.

(b) Putting

$$\boldsymbol{\mu}_a(A) := \int_A \widehat{f}\, d\boldsymbol{\lambda}, \qquad \boldsymbol{\mu}_s = \boldsymbol{\mu} - \boldsymbol{\mu}_a \tag{A3.5.3}$$

we prove that $\boldsymbol{\mu}_s$ is singular with respect to $\boldsymbol{\lambda}$. We will need the following assertion about the decomposition of an arbitrary signed measure (for the definition, see Sect. 3.3.1 of this appendix).

Theorem A3.5.3 (The Hahn theorem on decomposition of a measure) *For any signed finite measure $\boldsymbol{\gamma}$, there exist disjoint sets $D^+ \in \mathfrak{F}$ and $D^- \in \mathfrak{F}$ such that, for any $A \in \mathfrak{F}$,*

$$\boldsymbol{\gamma}\big(AD^+\big) \ge 0, \qquad \boldsymbol{\gamma}\big(AD^-\big) \le 0.$$

Proof We first prove that there exists a set $D \in \mathfrak{F}$ on which $\boldsymbol{\gamma}(A)$ attains its upper bound.

Let $B_n \in \mathfrak{F}$ be a sequence such that $\boldsymbol{\gamma}(B_n) \to \Gamma = \sup_A \boldsymbol{\gamma}(A)$ as $n \to \infty$. Put $B := \bigcup_k B_k$ and consider, for a fixed n, the decomposition of T into 2^n sets $B_{n,m}$, $m = 1, \ldots, 2^n$, of the form $\bigcap_{k=1}^n B_k'$, where $B_k' = B_k$ or $B - B_k$, $k \le n$. For $n < N$, each $B_{n,m}$ is a finite union of sets $B_{N,M}$, $1 \le M \le 2^N$. Denote by D_n the sum of all $B_{n,m}$ for which $\boldsymbol{\gamma}(B_{n,m}) \ge 0$. Then $\boldsymbol{\gamma}(B_n) \le \boldsymbol{\gamma}(D_n)$.

On the other hand, for $N > n$, each $B_{N,M}$ either belongs to D_n or is disjoint with it. Therefore

$$\boldsymbol{\gamma}(D_n) \le \boldsymbol{\gamma}(D_n + D_{n+1} + \cdots + D_N).$$

This implies that, for the set $D = \lim \bigcup_{k=n}^\infty D_k$, one has $\boldsymbol{\gamma}(B_n) \le \boldsymbol{\gamma}(D)$, $\Gamma \le \boldsymbol{\gamma}(D)$. Recalling the definition of Γ, we obtain that $\boldsymbol{\gamma}(D) = \Gamma$.

Thus we have proved the existence of a set D on which $\boldsymbol{\gamma}(D)$ attains its maximum. We now show that, for any $A \in \mathfrak{F}$, one has $\boldsymbol{\gamma}(A\,D) \ge 0$ and $\boldsymbol{\gamma}(A\,\overline{D}) \le 0$, where $\overline{D} = \Omega - D$. Indeed, assuming, for instance, that $\boldsymbol{\gamma}(AD) < 0$, we come to a contradiction, for in that case

$$\boldsymbol{\gamma}(D - AD) = \boldsymbol{\gamma}(D) - \boldsymbol{\gamma}(AD) > \boldsymbol{\gamma}(D).$$

Similarly, assuming that $\boldsymbol{\gamma}(A\overline{D}) > 0$, we would get

$$\boldsymbol{\gamma}(D + A\overline{D}) = \boldsymbol{\gamma}(D) + \boldsymbol{\gamma}(A\overline{D}) > \boldsymbol{\gamma}(D).$$

It remains to put $D^+ := D$, $D^- := \overline{D}$. The theorem is proved. □

Corollary A3.5.1 *Any finite signed measure $\boldsymbol{\gamma}$ can be represented as $\boldsymbol{\gamma} = \boldsymbol{\gamma}^+ - \boldsymbol{\gamma}^-$, where $\boldsymbol{\gamma}^\pm$ are finite nonnegative measures.*

To prove the corollary, it suffices to put

$$\boldsymbol{\gamma}^\pm(A) := \pm\boldsymbol{\gamma}\big(A \cap D^\pm\big),$$

where $D^\pm$ are the sets from the Hahn decomposition theorem. □

We return to the proof of the fact that the measure $\boldsymbol{\mu}_s$ in equality (A3.5.3) is singular. Let D_n^+ be the set in the Hahn decomposition of the signed measure

$$\boldsymbol{\nu}_n = \boldsymbol{\mu}_s - \frac{1}{n}\boldsymbol{\lambda}.$$

Put $N = \bigcap_n D_n^-$. Then $\overline{N} = \bigcup_n D_n^+$ and, for all n and $A \in \mathfrak{F}$,

$$0 \le \boldsymbol{\mu}_s(AN) \le \frac{1}{n}\boldsymbol{\lambda}(AN).$$

From here, letting $n \to \infty$, we obtain $\boldsymbol{\mu}_s(AN) = 0$ and hence $\boldsymbol{\mu}_s(A) = \boldsymbol{\mu}_s(A\overline{N})$. That is, the set $\overline{N}$ is a support of the measure $\boldsymbol{\mu}_s$.

Further, because

$$\boldsymbol{\mu}_a(A) = \boldsymbol{\mu}(A) - \boldsymbol{\mu}_s(A\overline{N}) \le \boldsymbol{\mu}(A) - \boldsymbol{\mu}_s\big(AD_n^+\big),$$

we have

$$\int_A \left(\widehat{f} + \frac{1}{n}\mathrm{I}_{D_n^+}\right) d\boldsymbol{\lambda} = \boldsymbol{\mu}_a(A) + \frac{1}{n}\boldsymbol{\lambda}\big(AD_n^+\big) \le \boldsymbol{\mu}(A) - \boldsymbol{\nu}_n\big(AD_n^+\big) \le \boldsymbol{\mu}(A).$$

This means that $\widehat{f} + \frac{1}{n}\mathrm{I}_{D_n^+} \in \mathcal{F}$ and hence

$$\alpha \ge \int \left(\widehat{f} + \frac{1}{n}\mathrm{I}_{D_n^+}\right) d\boldsymbol{\lambda} = \alpha + \frac{1}{n}\boldsymbol{\lambda}\big(D_n^+\big).$$

This implies $\boldsymbol{\lambda}(D_n^+) = 0$ and $\boldsymbol{\lambda}(\overline{N}) = 0$, so that $\boldsymbol{\mu}_s$ is singular with respect to $\boldsymbol{\lambda}$ since $\overline{N}$ is a support of $\boldsymbol{\mu}_s$. □

Uniqueness of the decomposition $\boldsymbol{\mu} = \boldsymbol{\mu}_a + \boldsymbol{\mu}_s$ can be established as follows. Assume that $\boldsymbol{\mu} = \boldsymbol{\mu}_a' + \boldsymbol{\mu}_s'$ is another decomposition. Then $\boldsymbol{\gamma} := \boldsymbol{\mu}_a' - \boldsymbol{\mu}_a = \boldsymbol{\mu}_s - \boldsymbol{\mu}_s'$. By singularity, there exist sets N and N' such that $\boldsymbol{\mu}_s(\overline{N}) = 0$, $\boldsymbol{\lambda}(N) = 0$, $\boldsymbol{\mu}_s'(\overline{N}') = 0$, and $\boldsymbol{\lambda}(N') = 0$. Clearly, $\boldsymbol{\lambda}(D) = 0$, where $D = N \cup N'$. If we assumed that $\boldsymbol{\gamma} = \boldsymbol{\mu}_a' - \boldsymbol{\mu}_a = \boldsymbol{\mu}_s - \boldsymbol{\mu}_s' \ne 0$, then there would exist an $A \in \mathfrak{F}$ such that $\boldsymbol{\gamma}(A) \ne 0$. Therefore, either $\boldsymbol{\gamma}(AD) \ne 0$ or $\boldsymbol{\gamma}(A\overline{D}) \ne 0$. However, the former is impossible, for $\boldsymbol{\lambda}(D) = 0$ implies $\boldsymbol{\mu}_a'(D) = \boldsymbol{\mu}_a(D) = 0$. The latter is also impossible, since $\overline{D} = \overline{N}\,\overline{N}'$ and hence $\boldsymbol{\mu}_s(\overline{D}) = \boldsymbol{\mu}_s'(\overline{D}) = 0$.

Uniqueness of the function f (up to $\boldsymbol{\lambda}$-equivalence) follows from the observation that the equalities

$$\boldsymbol{\mu}_a(A) = \int_A f\,d\boldsymbol{\lambda} = \int_A f' d\boldsymbol{\lambda}, \qquad \int_A (f - f')\,d\boldsymbol{\lambda} = 0$$

for all A imply the equality $f - f' = 0$ a.e. Assuming, say, that $\boldsymbol{\lambda}(A) > 0$ for $A = \{\omega : f - f' > \varepsilon\}$ would yield for such A the relation $\int_A (f - f')\,d\boldsymbol{\lambda} > 0$. The theorem is proved. □

One of the most important applications of the Radon–Nikodym theorem is the proof of *existence and uniqueness of conditional expectations*.

Proof Let $\mathfrak{F}_0$ be a σ-subalgebra of $\mathfrak{F}$ and ξ a random variable on $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ such that $\mathbf{E}\xi$ exists. In Sect. 4.8 we defined the conditional expectation $\mathbf{E}(\xi \mid \mathfrak{F}_0)$ of the variable ξ given $\mathfrak{F}_0$ as an $\mathfrak{F}_0$-measurable random variable η for which

$$\mathbf{E}(\eta;\, B) = \mathbf{E}(\xi;\, B) \tag{A3.5.4}$$

for any $B \in \mathfrak{F}$. We can assume without loss of generality that $\xi \geq 0$ (an arbitrary function ξ can be represented as a difference of two positive functions). Then the right-hand side of (A3.5.4) will be a measure on $\langle \Omega, \mathfrak{F}_0\rangle$. Since $\mathbf{E}(\xi;\, B) = 0$ if $\mathbf{P}(B) = 0$, this measure will be absolutely continuous with respect to $\mathbf{P}$. This implies, by the Radon–Nikodym theorem, the existence of a unique (up to $\mathbf{P}$-equivalence) measurable function η on $\langle \Omega, \mathfrak{F}_0\rangle$ such that, for any $B \in \mathfrak{F}_0$,

$$\mathbf{E}(\xi;\, B) = \int_B \eta\, d\mathbf{P}.$$

This relation is clearly equivalent to (A3.5.4). It establishes the required existence and uniqueness of the conditional expectation. □

Another consequence of the assertions proved in the present section was mentioned in Sect. 3.6 and is related to the Lebesgue theorem stating that any distribution $\mathbf{P}$ on the real line $\mathbb{R} = (-\infty, \infty)$ (or the respective distribution function) has a unique representation as a sum of the three components $\mathbf{P} = \mathbf{P}_a + \mathbf{P}_s + \mathbf{P}_\partial$, where the component $\mathbf{P}_a$ is absolutely continuous with respect to Lebesgue measure:

$$\mathbf{P}_a(A) = \int_A f(x)\,dx;$$

$\mathbf{P}_\partial$ is the discrete component concentrated on an at most countable set of points $x_1, x_2, \ldots$ such that $\mathbf{P}(\{x_k\}) > 0$, and the component $\mathbf{P}_s$ has a support of Lebesgue measure zero and a continuous distribution function. This is an immediate consequence of the Lebesgue decomposition theorem. One just has to extract the discrete part from the singular (with respect to Lebesgue measure $\boldsymbol{\lambda}$) component of $\mathbf{P}$, first removing all the points x for which $\mathbf{P}(\{x\}) \geq 1/2$, then all points x for which

$\mathbf{P}(\{x\}) \geq 1/3$, and so on. It is clear that in this way we will get at most a countable set of xs, and that this process determines uniquely the discrete component $\mathbf{P}_\partial$.

All the aforesaid clearly also applies to distributions in n-dimensional Euclidean spaces $\mathbb{R}^n$.

3.6 Weak Convergence and Convergence in Total Variation of Distributions in Arbitrary Spaces

3.6.1 Weak Convergence

In Sects. 6.2 and 7.6 we studied weak convergence of distributions of random variables and vectors, i.e., weak convergence of distributions in $\mathbb{R}^k$, $k \geq 1$. Now we want to introduce the notion of weak convergence in more general spaces $\mathcal{X}$. As the definitions given in Sect. 6.2 show, we will need continuous functions $f(x)$ on $\mathcal{X}$. This is possible only if the space $\mathcal{X}$ is endowed with a topology. For simplicity's sake, we restrict ourselves to the case where the space $\mathcal{X}$ is endowed with a metric $\rho(x, y)$. Thus, assume we are given a measurable space $\langle \mathcal{X}, \mathfrak{B} \rangle$ with a metric ρ which is "consistent" with the σ-algebra $\mathfrak{B}$, i.e., all open (with respect to the metric ρ) sets from $\mathcal{X}$ belong to $\mathfrak{B}$ (cf. Sect. 16.1), so that any continuous (with respect to ρ) functional will be $\mathfrak{B}$-measurable. This means that if a distribution $\mathbf{Q}$ is given on $\langle \mathcal{X}, \mathfrak{B} \rangle$ (i.e., a probability space $\langle \mathcal{X}, \mathfrak{B}, \mathbf{Q} \rangle$ is given), then $\{x : f(x) < t\} \in \mathfrak{B}$ for any t, and the probabilities of these sets are defined.

Now let $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ be the basic probability space. A measurable mapping $\xi = \xi(\omega)$ of the space $\langle \Omega, \mathfrak{F} \rangle$ to $\langle \mathcal{X}, \mathfrak{B} \rangle$ is called an $\mathcal{X}$-*valued random element*. If $\langle \Omega, \mathfrak{F} \rangle = \langle \mathcal{X}, \mathfrak{B} \rangle$, the mapping ξ may be the identity mapping. The space $\langle \mathcal{X}, \mathfrak{B} \rangle$ is said to be the *sample* or *state space* of the random element ξ. When a functional f is continuous, $f(\xi)$ is a random variable in $\langle \mathbb{R}, \mathfrak{B} \rangle$.

Definition A3.6.1 Let a distribution $\mathbf{P}$ and a sequence of distributions $\mathbf{P}_n$ be given on the space $\langle \mathcal{X}, \mathfrak{B} \rangle$. The sequence $\mathbf{P}_n$ is said to *converge weakly* to $\mathbf{P}$: $\mathbf{P}_n \Rightarrow \mathbf{P}$ as $n \to \infty$ if, for any bounded continuous functional f ($f \in C_b(\mathcal{X})$),

$$\int f(x)\, d\mathbf{P}_n(x) \to \int f(x)\, d\mathbf{P}(x). \qquad \text{(A3.6.1)}$$

If ξ_n and ξ are random elements having the distributions $\mathbf{P}_n$ and $\mathbf{P}$, respectively, then (A3.6.1) is equivalent to

$$\mathbf{E} f(\xi_n) \to \mathbf{E} f(\xi). \qquad \text{(A3.6.2)}$$

This, in turn, for any continuous functional f ($f \in C(\mathcal{X})$), is equivalent to

$$f(\xi_n) \Rightarrow f(\xi). \qquad \text{(A3.6.3)}$$

Indeed, (A3.6.3) means that, for any bounded continuous function g ($g \in C_b(\mathbb{R})$),

$$\mathbf{E}g\big(f(\xi_n)\big) \to \mathbf{E}g\big(f(\xi)\big), \tag{A3.6.4}$$

which is equivalent to (A3.6.2).

If $\mathfrak{X} = \mathfrak{X}(T)$ is the space of real-valued functions $x(t)$, $t \in T$, given on a parametric set T, and a measurable mapping $\xi(\omega)$ of the basic probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ into $\langle \mathfrak{X}, \mathfrak{B}\rangle$ is given, then the random element $\xi(\omega) = \xi(\omega, t)$ will be a *random process* (see Sect. 18.1) if $\{x : x(t) < u\} \in \mathfrak{B}$ for all t, u. In that case (A3.6.1)–(A3.6.4) will refer to the weak convergence of the distributions of random processes which has already been studied in Chap. 20.

In the metric space $\mathfrak{X}$, for any $A \in \mathfrak{X}$, one can define its boundary

$$\partial A = [A] - (A),$$

where $[A]$ is the closure of A, (A) being its interior ($(A) = \mathfrak{X} - [\overline{A}]$, where $\overline{A}$ is the complement of A).

Definition A3.6.2 A set A is said to be a *continuity set* of the distribution $\mathbf{P}$ (or $\mathbf{P}$-continuous set) if $\mathbf{P}(\partial A) = 0$. We will denote the class of all $\mathbf{P}$-continuous sets by $\mathfrak{D}_P$.

The following criterion of weak convergence of distributions holds true.

Theorem A3.6.1 *The following four conditions are equivalent*:

(i) $\mathbf{P}_n \Rightarrow \mathbf{P}$,
(ii) $\lim_{n\to\infty} \mathbf{P}_n(A) = \mathbf{P}(A)$ *for all* $A \in \mathfrak{D}_P$,
(iii) $\limsup_{n\to\infty} \mathbf{P}_n(F) \le \mathbf{P}(F)$ *for all closed* $F \subset \mathfrak{X}$,
(iv) $\liminf_{n\to\infty} \mathbf{P}_n(G) \ge \mathbf{P}(G)$ *for all open* $G \subset \mathfrak{X}$.

Observe that if $\mathbf{P}_n \Rightarrow \mathbf{P}$, then convergence (A3.6.1)–(A3.6.3) takes place for a wider class of functionals than $C_b(\mathfrak{X})$ ($C(\mathfrak{X})$), namely, for the so-called $\mathbf{P}$-continuous functionals (or functionals continuous with $\mathbf{P}$-probability 1). We will call so the functionals f for which $f(x_n) \to f(x)$ as $\rho(x_n, x) \to 0$ not for all $x \in \mathfrak{X}$, but only for $x \in A$, $\mathbf{P}(A) = 1$. The class of $\mathbf{P}$-continuous functionals will be denoted by $C_P(\mathfrak{X})$.

The classes $\mathfrak{D}_P$ and $C_P(\mathfrak{X})$, and also the classes of all closed and open sets participating in Theorem A3.6.1, are very wide which makes verifying the conditions of Theorem A3.6.1 rather difficult and cumbersome. These classes can be substantially restricted if we consider not arbitrary but only *relatively compact* sequences $\mathbf{P}_n$ (from any subsequence $\mathbf{P}'_n$ one can choose a convergent subsequence; this approach was already used in Sect. 6.3).

Definition A3.6.3 A class $\mathfrak{D}$ of sets from $\mathfrak{B}$ is said to *determine the measure* $\mathbf{P}$ if, for a measure $\mathbf{Q}$, the equalities $\mathbf{P}(A) = \mathbf{Q}(A)$ for all $A \in \mathfrak{D}\mathfrak{D}_P$ imply $\mathbf{Q} = \mathbf{P}$.

A class $\mathfrak{D}$ determines the measure $\mathbf{P}$ if $\mathfrak{D}$ is an algebra and $\sigma(\mathfrak{D}\mathfrak{D}_P) = \mathfrak{B}_\mathfrak{X}$ (condition $\sigma(\mathfrak{D}) = \mathfrak{B}_\mathfrak{X}$ is insufficient (see e.g. [4])).

In a similar way we introduce the class $\mathfrak{F}$ of functionals f determining the distribution $\mathbf{P}$ of a random element $\xi = \xi^P$: for any $\mathbf{Q}$, the coincidence of the distributions of $f(\xi^P)$ and $f(\xi^Q)$ for all $f \in \mathfrak{F}C_P(X)$ implies $\mathbf{P} = \mathbf{Q}$.

Theorem A3.6.2 *A necessary and sufficient condition for convergence $\mathbf{P}_n \Rightarrow \mathbf{P}$ is that*:

(1) *the sequence $\mathbf{P}_n$ is relatively compact*; *and*
(2) *there exists a class of sets $\mathfrak{D} \subset \mathfrak{B}_\mathfrak{X}$ determining the measure $\mathbf{P}$ and such that $\mathbf{P}_n(A) \to \mathbf{P}(A)$ for any $A \in \mathfrak{D}\mathfrak{D}_P$.*

An alternative to condition (2) is the existence of a class of functionals $\mathfrak{F}$ which determines the measure $\mathbf{P}$ and is such that $\mathbf{P}(f(\xi_n) < t) \Rightarrow \mathbf{P}(f(\xi) < t)$ for all $f \in \mathfrak{F}C_P(X)$.

The following notion of tightness plays an important role in establishing the compactness of $\{\mathbf{P}_n\}$.

Definition A3.6.4 A family of distributions $\{\mathbf{P}_n\}$ on $\langle\mathfrak{X}, \mathfrak{B}\rangle$ is said to be *tight* if, for any $\varepsilon > 0$, there exists a compact set $K = K_\varepsilon \subset \mathfrak{X}$ such that $\mathbf{P}_n(K) > 1 - \varepsilon$ for all n.

Theorem A3.6.3 (Prokhorov) *If $\{\mathbf{P}_n\}$ is a tight family of distributions then it is relatively compact*. *If $\mathfrak{X}$ is a complete separable space*, *the converse assertion is also true.*

Since, for many functional spaces (in particular, for spaces $C(0, T)$ and $D(0, T)$), there exist simple explicit criteria for compactness of sets, one can now establish conditions ensuring convergence $\mathbf{P}_n \Rightarrow \mathbf{P}$ in these spaces. It is well known, for example, that in the above-mentioned spaces compacta are, roughly speaking, of the form $\{x : \omega_\triangle(x) \le \varepsilon(\triangle)\}$, where $\omega_\triangle(x)$ is the so-called "modulus of continuity" (in the space C or D, respectively) of the element x, and $\varepsilon(\triangle) \ge 0$ is an arbitrary function vanishing as $\triangle \downarrow 0$.

The proofs of Theorems A3.6.1–A3.6.3 can be found, for example, in [1]. We do not present them here as they are somewhat beyond the scope of this book and, on the other hand, the theorems themselves are not used in the body of the text. We use only the special cases of these theorems given in Sects. 6.2 and 6.3.

The invariance principle of Sect. 20.1 is a theorem about weak convergence of distributions in the space $C(0, 1)$. In order to use Theorems A3.6.2 and A3.6.3 to prove this result, one has to choose the class $\mathfrak{D}$ to be the class of cylinder sets. Convergence of $\mathbf{P}_n$ to $\mathbf{P}$ on sets from this class $\mathfrak{D}$ is the convergence of finite-dimensional distributions of processes $s_n(t)$ generated by sums of random variables (see Sect. 20.1). Since the increments of $s_n(t)$ are essentially independent, the demonstration of that part of the theorem reduces to proving asymptotic normality of these increments, which follows immediately from the central limit theorem.

The condition of compactness of the family of distributions in $C(0,1)$ requires, according to Theorem A3.6.3, a proof that the modulus of continuity of the trajectory $s_n(t)$ converges to zero in probability (for more details, see e.g. [1]). This could be proved using the Kolmogorov inequality from Corollary 11.2.1.

3.6.2 Convergence in Total Variation

So, to consider weak convergence of distributions in spaces $\langle \mathcal{X}, \mathfrak{B}\rangle$ of a general nature, one has to introduce a topology in the space, which is not always convenient and feasible. There exists another type of convergence of distributions on $\langle \mathcal{X}, \mathfrak{B}\rangle$ which does not require the introduction of topologies. This is convergence in total variation.

Definition A3.6.5 Let $\boldsymbol{\gamma}$ be a finite signed measure on $\langle \mathcal{X}, \mathfrak{B}\rangle$. The total variation of $\boldsymbol{\gamma}$ (or the total variation norm $\|\boldsymbol{\gamma}\|$) is the quantity

$$\|\boldsymbol{\gamma}\| = \sup_{f:|f|\le 1} \left| \int f(x)\, d\boldsymbol{\gamma}(x) \right|, \tag{A3.6.5}$$

where the supremum is taken over the class of all $\mathfrak{B}$-measurable functions $f(x)$ such that $|f(x)| \le 1$ for all $x \in \mathcal{X}$.

The supremum in (A3.6.5) is clearly attained on such functions f for which, roughly speaking, $f(x) = 1$ at points x such that $d\boldsymbol{\gamma}(x) > 0$, and $f(x) = -1$ at points x for which $d\boldsymbol{\gamma}(x) < 0$. Therefore (A3.6.5) can be written in the form

$$\|\boldsymbol{\gamma}\| = \int \big|d\boldsymbol{\gamma}(x)\big|. \tag{A3.6.6}$$

An exact meaning to this expression can be given using the Hahn decomposition theorem (see Corollary A3.5.1), which implies

$$\|\boldsymbol{\gamma}\| = \boldsymbol{\gamma}^+(\mathcal{X}) + \boldsymbol{\gamma}^-(\mathcal{X}). \tag{A3.6.7}$$

The right-hand side of this equality may be taken as a definition of $\int |d\boldsymbol{\gamma}(x)|$.

Lemma A3.6.2 *If* $\boldsymbol{\gamma}(\mathcal{X}) = 0$, *then* $\|\boldsymbol{\gamma}\| = 2\sup_{B\in\mathfrak{B}} \boldsymbol{\gamma}(B)$.

Proof From (A3.6.5) it follows that, for any B ($\overline{B}$ is the complement of B, $\boldsymbol{\gamma}(B) \cup \boldsymbol{\gamma}(\overline{B})$=0),

$$\|\boldsymbol{\gamma}\| \ge \big|\boldsymbol{\gamma}(B)\big| + \big|\boldsymbol{\gamma}(\overline{B})\big| = 2\big|\boldsymbol{\gamma}(B)\big|.$$

Therefore $\|\boldsymbol{\gamma}\| \ge 2\sup_{B\in\mathfrak{B}} |\boldsymbol{\gamma}(B)|$.

To obtain the converse inequality, we will make use of Corollary A3.5.1 of the Hahn decomposition theorem. As we have already noted, according to that theorem (for the definition of the set $D^{\pm}$ see the Hahn theorem),

$$\begin{aligned}\|\boldsymbol{\gamma}\| &= \boldsymbol{\gamma}^{+}(\mathcal{X}) + \boldsymbol{\gamma}^{-}(\mathcal{X}) = \boldsymbol{\gamma}^{+}\big(D^{+}\big) + \boldsymbol{\gamma}^{-}\big(\overline{D^{+}}\big) \\ &= \boldsymbol{\gamma}\big(D^{+}\big) - \boldsymbol{\gamma}\big(\overline{D^{+}}\big) = 2\boldsymbol{\gamma}\big(D^{+}\big) \le 2 \sup_{B\in\mathfrak{B}} \boldsymbol{\gamma}(B).\end{aligned}$$

The lemma is proved. □

Definition A3.6.6 Let $\mathbf{P}$ be a distribution and $\mathbf{P}_n$, $n = 1, 2, \ldots$, a sequence of distributions given on $\langle\mathcal{X}, \mathfrak{B}\rangle$. We will say that $\mathbf{P}_n$ *converges to* $\mathbf{P}$ *in total variation*: $\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$, if $\|\mathbf{P}_n - \mathbf{P}\| \to 0$ as $n \to \infty$.

Convergence in total variation is a very strong form of convergence. If $\langle\mathcal{X}, \mathfrak{B}\rangle$ is a metric space and $\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$, then $\mathbf{P}_n \Rightarrow \mathbf{P}$. Indeed, since any functional $f \in C_b(\mathcal{X})$ is bounded: $|f(x)| < b$, we have

$$\left|\int f(d\mathbf{P}_n - d\mathbf{P})\right| \le b \int \big|d(\mathbf{P}_n - \mathbf{P})\big| = b\|\mathbf{P}_n - \mathbf{P}\| \to 0.$$

Thus in that case

$$\int f\, d\mathbf{P}_n \to \int f\, d\mathbf{P}$$

even without assuming the continuity of f.

The converse assertion about convergence $\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$ if $\mathbf{P}_n \Rightarrow \mathbf{P}$ is not true. Let, for example, $\mathcal{X} = [0, 1]$, $\mathbf{P}_n$ be the uniform distribution on the set of $n + 1$ points $\{0, 1/n, \ldots, n/n\}$, and $\mathbf{P} = \mathbf{U}_{0,1}$. It is clear that all $\mathbf{P}_n$ are concentrated on the countable set $\mathcal{N}$ of all rational numbers. Therefore $\mathbf{P}_n(\mathcal{N}) = 1$, $\mathbf{P}(\mathcal{N}) = 0$, and $\|\mathbf{P}_n - \mathbf{P}\| = \mathbf{P}_n(\mathcal{N}) + \mathbf{P}(\mathcal{X} \setminus \mathcal{N}) = 2$. At the same time, clearly $\mathbf{P}_n \Rightarrow \mathbf{P}$.

Now let the distribution $\mathbf{P}$ have a density p with respect to a measure $\boldsymbol{\mu}$ (one could take, in particular, $\boldsymbol{\mu} = \mathbf{P}$, in which case $p(x) \equiv 1$). Denote by p_n the density (with respect to $\boldsymbol{\mu}$) of the absolutely continuous (with respect to $\boldsymbol{\mu}$) component $\mathbf{P}_n^a$ of the distribution $\mathbf{P}_n$.

Theorem A3.6.4 *A necessary and sufficient condition for convergence* $\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$ *is that* p_n *converges to* p *in measure* $\boldsymbol{\mu}$, *i.e., for any* $\varepsilon > 0$,

$$\boldsymbol{\mu}\big\{x : \big|p_n(x) - p(x)\big| > \varepsilon\big\} \to 0 \quad \text{as } n \to \infty.$$

Proof We have

$$\int \big|d(\mathbf{P}_n - \mathbf{P})\big| = \int \big|p_n(x) - p(x)\big|\boldsymbol{\mu}(dx) + \big\|\mathbf{P}_n^s\big\|,$$

where $\mathbf{P}_n^s$ is the singular component of $\mathbf{P}_n$ with respect to the measure $\boldsymbol{\mu}$.

Let $\|\mathbf{P}_n - \mathbf{P}\| \to 0$. Then

$$\int |p_n - p| \, d\boldsymbol{\mu} \to 0, \tag{A3.6.8}$$

and hence

$$\boldsymbol{\mu}\{x : |p_n(x) - p(x)| > \varepsilon\} \le \varepsilon^{-1} \int |p_n - p| \, d\boldsymbol{\mu} \to 0.$$

Now let $p_n \xrightarrow{\boldsymbol{\mu}} p$. Put

$$B_\varepsilon = \{x : p(x) \ge \varepsilon\}, \qquad A_{n,\varepsilon} = \{x : |p_n(x) - p(x)| \le \varepsilon^2\}.$$

Then

$$1 \ge \int_{B_\varepsilon} p \, d\boldsymbol{\mu} \ge \varepsilon \boldsymbol{\mu}(B_\varepsilon), \quad \boldsymbol{\mu}(B_\varepsilon) \le \frac{1}{\varepsilon}.$$

Consider

$$\int |p_n - p| \, d\boldsymbol{\mu} = \int_{B_\varepsilon A_{n,\varepsilon}} + \int_{\overline{B_\varepsilon A_{n,\varepsilon}}}. \tag{A3.6.9}$$

Here the first integral on the right-hand side does not exceed ε. Since

$$\lim_{\varepsilon \to 0} \int_{B_\varepsilon} p \, d\boldsymbol{\mu} \to 1,$$

we will have, for a given $\delta > 0$ and sufficiently small ε, the inequality

$$\int_{B_\varepsilon} p \, d\boldsymbol{\mu} > 1 - \delta$$

and, for n large enough,

$$\int_{B_\varepsilon A_{n,\varepsilon}} p \, d\boldsymbol{\mu} > 1 - 2\delta, \quad \int_{B_n A_{n,\varepsilon}} p_n \, d\boldsymbol{\mu} > 1 - 3\delta. \tag{A3.6.10}$$

It follows from these two inequalities that the second integral in (A3.6.9) does not exceed 5δ, which proves (A3.6.8). Furthermore, (A3.6.10) implies that $\|\mathbf{P}_n^a\| > 1 - 3\delta$ and $\|\mathbf{P}_n^s\| < 3\delta$. The theorem is proved. □

The theorem implies that if $\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$ then the absolutely continuous with respect to $\boldsymbol{\mu} = \mathbf{P}$ component $\mathbf{P}_n^a$ of the distribution $\mathbf{P}_n$ has a density $p_n(x) \xrightarrow{p} 1$, $\mathbf{P}_n^a(\mathcal{X}) \to 1$.

Appendix 4
The Helly and Arzelà–Ascoli Theorems

In this appendix we will prove Helly's theorem and the Arzelà–Ascoli theorem. The former theorem was used in Sect. 6.3, and both theorems will be used in the proof of the main theorem of Appendix 9.

Let $\mathcal{F}$ be the class of all distribution functions, and $\mathcal{G}$ the class of functions G possessing properties F1 and F2 from Sect. 3.2 (monotonicity and left continuity), and the properties $G(-\infty) \ge 0$ and $G(\infty) \le 1$. We will write $G_n \Rightarrow G$ as $n \to \infty$, $G \in \mathcal{G}$, if $G_n(x) \to G(x)$ at all points of continuity of the function G.

Theorem A4.1 (Helly) *Any sequence $F_n \in \mathcal{F}$ contains a convergent subsequence $F_{nn} \Rightarrow F \in \mathcal{G}$.*

We will need the following.

Lemma A4.1 *A sufficient condition for convergence $F_n \Rightarrow F \in \mathcal{G}$ is that*

$$F_n(x) \to F(x), \quad x \in D,$$

as $n \to \infty$ on some everywhere dense set D of the reals.

Proof Let x be an arbitrary point of continuity of $F(x)$. For arbitrary $x', x'' \in D$ such that $x' \le x \le x''$, one has

$$F_n(x') \le F_n(x) \le F_n(x'').$$

Consequently,

$$\lim_{n\to\infty} F_n(x') \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} F_n(x'').$$

From here and the conditions of the lemma we obtain

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

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013

Letting $x' \uparrow x$ and $x'' \downarrow x$ along the set D and taking into account that x is a point of continuity of F, we get

$$\lim_{n\to\infty} F_n(x) = F(x).$$

The lemma is proved. □

Proof of Theorem A4.1 Let $D = \{x_n\}$ be an arbitrary countable everywhere dense set of real numbers. The numerical sequence $\{F_n(x_1)\}$ is bounded and hence contains a convergent sequence $\{F_{1n}(x_1)\}$. Denote the limit of this sequence by $F(x_1)$. Consider now the numerical sequence $\{F_{1n}(x_2)\}$. It also contains a convergent subsequence $\{F_{2n}(x_2)\}$ with a limit $F(x_2)$. Moreover,

$$\lim_{n\to\infty} F_{2n}(x_1) = F(x_1).$$

Continuing this process, we will obtain, for any number k, k sequences

$$\big\{F_{kn}(x_i)\big\}, \quad i = 1, \dots, k,$$

such that $\lim_{n\to\infty} F_{kn}(x_i) = F(x_i)$.

Consider the diagonal sequence of the distribution functions $\{F_{nn}(x)\}$. For any $x_k \in D$, only $k - 1$ first elements of the numerical sequence $\{F_{nn}(x_k)\}$ may not belong to the sequence $F_{kn}(x_k)$. Therefore

$$\lim_{n\to\infty} F_{nn}(x_k) = F(x_k).$$

It is clear that $F(x)$ is a non-decreasing bounded function given on D. It can easily be extended by continuity from the left to a non-decreasing function on the whole real line. Now we see that the sequence $\{F_{nn}\}$ and the function F satisfy the conditions of Lemma A4.1. The theorem is proved. □

The conditions of Helly's theorem can be weakened. Namely, instead of $\mathcal{F}$ we could consider a wider class $\mathcal{H}$ of non-decreasing left continuous (i.e., satisfying properties F1 and F3) functions H majorised by a fixed function: for any x, $|H(x)| \le N(x) < \infty$, where N is a given function characterising the class $\mathcal{H}$. We do not exclude the case when $|H(x)|$ (or $N(x)$) grow unboundedly as $|x| \to \infty$. The following generalised version of Helly's theorem is true.

Theorem A4.2 (Generalised Helly theorem) *Any sequence $H_n \in \mathcal{F}$ contains a subsequence H_{nn} which converges to a function $H \in \mathcal{H}$ at each point of continuity of H.*

The Proof repeats the above proof of Helly's theorem. □

To each function $H_n \in \mathcal{H}$ we can associate a measure $\boldsymbol{\mu}_n$ by putting

$$\boldsymbol{\mu}_n\big([a, b)\big) = H_n(b) - H_n(a).$$

The generalised Helly theorem will then mean that, for any sequence of measures $\boldsymbol{\mu}_n$ generated by functions from $\mathcal{H}$, there exists a subsequence $\boldsymbol{\mu}_{nn}$ converging weakly on each finite interval of which the endpoints are not atoms of the limiting measure $\boldsymbol{\mu}_n$.

We give one more analogue of Helly's theorem which refers to a collection of equicontinuous functions g_n. Recall that a sequence of functions $\{g_n\}$ is said to be *equicontinuous* if, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|x_1 - x_2| < \delta$ implies $|g_n(x_1) - g_n(x_2)| < \varepsilon$ for all n.

Theorem A4.3 (Arzelà–Ascoli) *Let $\{g_n\}$ be a sequence of uniformly bounded and equicontinuous functions of a real variable. Then there exists a subsequence g_{n_k} converging to a continuous limit g uniformly on each finite interval.*

Proof Choose again a countable everywhere dense subset $\{x_n\}$ of the real line, and a subsequence $\{g_{n_k}\}$ converging at the points $x_1, x_2, \ldots$ Denote its limit at the point x_j by $g(x_j)$. We have

$$\begin{aligned}\left|g_{n_k}(x) - g_{n_r}(x)\right| \le \left|g_{n_k}(x) - g_{n_k}(x_j)\right| + \left|g_{n_r}(x) - g_{n_r}(x_j)\right| \\ + \left|g_{n_k}(x_j) - g_{n_r}(x_j)\right|.\end{aligned} \tag{A4.1}$$

By assumption, the last term on the right-hand side tends to 0 as $n_k \to \infty$, $n_r \to \infty$. By virtue of equicontinuity, for any point x there exists a point x_j such that, for all n,

$$\left|g_n(x) - g_n(x_j)\right| < \varepsilon. \tag{A4.2}$$

In any given finite interval I there exists a finite collection of points x_j such that (A4.2) will hold for all points $x_j \in I$. This implies that the right-hand side of (A4.1) will be less than 3ε for all sufficiently large n_k, n_r uniformly over $x_j \in I$. Thus there exists the limit $g(x) = \lim g_{n_k}(x)$, for which by (A4.2) we have $|g(x) - g(x_j)| \le \varepsilon$, which implies that g is continuous. The theorem is proved. □

Appendix 5
The Proof of the Berry–Esseen Theorem

In this appendix we prove the following assertion stated in Sect. 8.5.

Theorem A5.1 (Berry–Esseen) *Let ξ_k be independent identically distributed random variables*,

$$\mathbf{E}\xi_k = 0, \qquad \operatorname{Var}(\xi_k) = 1, \qquad \mu = \mathbf{E}|\xi_k|^3 < \infty, \qquad S_n = \sum_{k=1}^{n} \xi_k, \qquad \zeta_n = \frac{S_n}{\sqrt{n}}.$$

Then, for all n,

$$\Delta_n := \sup_x \bigl|\mathbf{P}(\zeta_n < x) - \Phi(x)\bigr| < \frac{c\mu}{\sqrt{n}},$$

where Φ is the standard normal distribution function and c is an absolute constant.

Proof We will make use of the composition method. As in Sect. 8.5, we will bound Δ_n based on estimates of smallness of $\mathbf{E}g(\zeta_n) - \mathbf{E}g(\eta)$, $\eta \in\!\!\!= \mathbf{\Phi}_{0,1}$, for smooth g. To get a bound for Δ_n in Sect. 8.5, we chose g to be a function constant outside a small interval. The next lemma shows that such a choice is not obligatory. Let G be a distribution function and $\gamma \in\!\!\!= G$ be independent of ζ_n and η. Put

$$g(z) := G\left(\frac{x-z}{\varepsilon}\right),$$

so that

$$\mathbf{E}g(\zeta_n) = \mathbf{E}G\left(\frac{x-\zeta_n}{\varepsilon}\right) = \mathbf{P}\left(\gamma < \frac{x-\zeta_n}{\varepsilon}\right) = \mathbf{P}(\zeta_n + \varepsilon\gamma < x),$$
$$\mathbf{E}g(\eta) = \mathbf{P}(\eta + \varepsilon\gamma < x).$$

Set

$$\Delta_{n,\varepsilon} := \sup_x \left|\mathbf{E}G\left(\frac{x-\zeta_n}{\varepsilon}\right) - \mathbf{E}G\left(\frac{x-\eta}{\varepsilon}\right)\right|$$

DOI 10.1007/978-1-4471-5201-9,

$$= \sup_x \big| \mathbf{P}(\zeta_n + \varepsilon\gamma < x) - \mathbf{P}(\eta + \varepsilon\gamma < x) \big|$$
$$= \sup_x \left| \int dG(y) \big[\mathbf{P}(\zeta_n < x - \varepsilon y) - \mathbf{P}(\eta < x - \varepsilon y) \big] \right|.$$

Clearly, $\Delta_{n,\varepsilon} \le \Delta_n$. Our aim will be to obtain a converse inequality for Δ_n.

Lemma A5.1 *Let $v > 0$ be such that $G(v) - G(-v) \ge 3/4$. Then, for any $\varepsilon > 0$,*

$$\Delta_n \le 2\Delta_{n,\varepsilon} + \frac{3v\varepsilon}{\sqrt{2\pi}}.$$

Proof Assume that the $\sup_x$ in the definition of Δ_n is attained on a positive value $\Delta_n(x) := F_n(x) - \Phi(x)$ (the case of a negative value $\Delta_n(x)$ is similar) and that, for a given $\delta > 0$, the value x_δ is such that

$$\Delta_n(x_\delta) = F_n(x_\delta) - \Phi(x_\delta) \ge \Delta_n - \delta,$$

where F_n is the distribution function of ζ_n. When the argument increases, the value of $\Delta_n(x_\delta)$ varies little in the following sense. Let $|y| < v$. Then $v - y > 0$ and

$$\Delta_n\big(x_\delta + \varepsilon(v - y)\big) = F_n\big(x_\delta + \varepsilon(v - y)\big) - \Phi\big(x_\delta + \varepsilon(v - y)\big)$$
$$\ge F_n(x_\delta) - \Phi(x_\delta) - \big[\Phi\big(x_\delta + \varepsilon(v - y)\big) - \Phi(x_\delta)\big].$$

Here the difference in the brackets does not exceed $\varepsilon(v - y)\Phi'(0) \le 2v\varepsilon/\sqrt{2\pi}$, and hence

$$\Delta_n\big(x_\delta + \varepsilon(v - y)\big) \ge \Delta_n - \delta - \frac{2v\varepsilon}{\sqrt{2\pi}}.$$

Therefore

$$\Delta_{n,\varepsilon} \ge \int dG(y) \Delta_n(x_\delta + \varepsilon v - \varepsilon y) = \int_{|y|<v} + \int_{|y|\ge v}$$
$$\ge \frac{3}{4}\left(\Delta_n - \delta - \frac{2v\varepsilon}{\sqrt{2\pi}}\right) - \frac{1}{4}\Delta_n = \frac{\Delta_n}{2} - \frac{3}{4}\left(\delta + \frac{2v\varepsilon}{\sqrt{2\pi}}\right).$$

Since δ is arbitrary, the assertion of the lemma follows. □

Corollary A5.1 *For $G = \Phi (\gamma \in \mathbf{\Phi}_{0,1})$ the value $v = 6/5$ satisfies the condition of Lemma A5.1, and*

$$\Delta_n \le 2(\Delta_{n,\varepsilon} + \varepsilon). \tag{A5.1}$$

At the next stage of the proof we bound $\Delta_{n,\varepsilon}$, and it is at that stage where the composition method will be used. Put

$$u(n) := \max_{k \le n} \Delta_k \frac{\sqrt{k}}{\mu}, \qquad \alpha^2 := \varepsilon^2 n.$$

By letters c (with or without indices) we will denote absolute constants, not necessarily the same ones.

Lemma A5.2 *For* $\alpha \ge 1$,

$$\Delta_{n,\varepsilon} \le c\mu\left(\frac{1}{\sqrt{n}} + \frac{\mu u(n-1)}{\alpha\sqrt{n}}\right). \tag{A5.2}$$

Proof Set $H_n := \sum_{k=1}^{n} \eta_k$, where $\eta_k \Subset \mathbf{\Phi}_{0,1}$ are independent of each other and of H_n and γ. The composition method is based on the following identity (cf. Theorem 8.5.1 and identity (8.5.3), $\eta \in \mathbf{\Phi}_{0,1}$):

$$\begin{aligned}
&\mathbf{P}(\zeta_n + \varepsilon\gamma < x) - \mathbf{P}(\eta + \varepsilon\gamma < x) = \mathbf{P}(S_n + \alpha\gamma < x\sqrt{n}) - \mathbf{P}(H_n + \alpha\gamma < x\sqrt{n}) \\
&\quad = \sum_{m=1}^{n} \big[\mathbf{P}\big(S_{m-1} + (H_n - H_m) + \xi_m + \alpha\gamma < x\sqrt{n}\big) \\
&\qquad - \mathbf{P}\big(S_{m-1} + (H_n - H_m) + \eta_m + \alpha\gamma < x\sqrt{n}\big)\big].
\end{aligned}$$

Since for $\gamma \Subset \mathbf{\Phi}_{0,1}$ one has $H_n - H_m + \alpha\gamma \Subset \mathbf{\Phi}_{0,n-m+\alpha^2}$, the last sum is equal to $\sum_{m=1}^{n} D_m$, where

$$\begin{aligned}
D_m &:= \mathbf{E}\left[\Phi\left(\frac{x\sqrt{n} - S_{m-1} - \xi_m}{d_m}\right) - \Phi\left(\frac{x\sqrt{n} - S_{m-1} - \eta_m}{d_m}\right)\right] \\
&= \mathbf{E}\left[\Phi\left(T_m - \frac{\xi_m}{d_m}\right) - \Phi\left(T_m - \frac{\eta_m}{d_m}\right)\right], \\
d_m^2 &:= n - m + \alpha^2, \qquad T_m := \frac{x\sqrt{n} - S_{m-1}}{d_m}.
\end{aligned}$$

To bound D_m we will adopt the same approach as in Lemma 8.5.1. Because the first two moments of ξ_m and η_m coincide, expanding Φ into a series yields

$$|D_m| \le \frac{2\mu}{d_m^3} \sup_t \mathbf{E}\phi''(T_m + t),$$

where $\phi(x) = \Phi'(x)$ and $\phi'' = \Phi'''$. Since the function ϕ'' is bounded,

$$|D_m| \le \frac{c\mu}{d_m^3}. \tag{A5.3}$$

We will also need another bound for D_m. To obtain it, consider the quantity

$$\begin{aligned}
R_m &:= \sup_t \big|\mathbf{E}\phi''(T_m + t)\big| \\
&\le \sup_t \big|\mathbf{E}\big[\phi''(T_m + t) - \phi''(V_m + t)\big]\big| + \sup_t \big|\mathbf{E}\phi''(V_m + t)\big|,
\end{aligned} \tag{A5.4}$$

where V_m is defined in the same way as T_m but with S_{m-1} replaced by H_{m-1}. Integrating by parts yields

$$\begin{aligned}\left|\mathbf{E}\left[\phi''(T_m+t)-\phi''(V_m+t)\right]\right| &= \left|\int \phi''(u+t)\,d\left[\mathbf{P}(T_m<u)-\mathbf{P}(V_m<u)\right]\right| \\ &= \left|\int \phi'''(u+t)\left[\mathbf{P}(T_m<u)-\mathbf{P}(V_m<u)\right]du\right| \\ &\le \varDelta_{m-1}\int\left|\phi'''(u)\right|du = c\varDelta_{m-1},\end{aligned}$$

since $|\mathbf{P}(T_m<u)-\mathbf{P}(V_m<u)|\le\varDelta_{m-1}$ (the variables T_m and V_m are obtained from $S_{m-1}/\sqrt{m-1}$ and $H_{m-1}/\sqrt{m-1}$, respectively, by one and the same linear transformation).

To bound the second summand on the right-hand side of (A5.4), note that

$$\mathbf{E}\phi''(V_m+t)=\int\phi''(u+t)\,\frac{1}{r_m}\,\phi\left(\frac{u-a_m}{r_m}\right)du, \tag{A5.5}$$

where

$$a_m = x\frac{\sqrt{n}}{d_m}, \qquad r_m=\sqrt{\frac{m-1}{n-m+\alpha^2}},$$

so that $\frac{1}{r_m}\phi(\dfrac{u-a_m}{r_m})$ is the density of $V_m=\frac{(x\sqrt{n}-H_{m-1})}{d_m}$. Integrating the right-hand side of (A5.5) twice by parts, we obtain

$$\left|\mathbf{E}\phi''(V_m+t)\right|=\frac{1}{r_m^3}\left|\int\phi(u+t)\phi''\left(\frac{u-a_m}{r_m}\right)du\right|\le\frac{c}{r_m^3}.$$

Thus,

$$R_m\le c\left(\varDelta_{m-1}+\frac{1}{r_m^3}\right), \qquad D_m\le c\mu\left(\frac{\varDelta_{m-1}}{d_m^3}+\frac{1}{(m-1)^{3/2}}\right).$$

The bounds derived for D_m do not depend on x. Therefore, using the bound just obtained for $m>n/2$, and bound (A5.3) for $m\le n/2$ (the latter bound implies then that $|D_m|\le c\mu/n^{3/2}$), we get

$$\varDelta_{n,\varepsilon}\le c\mu\left[\sum_{m\le n/2}n^{-3/2}+\sum_{m>n/2}^{n}\frac{\varDelta_{m-1}}{d_m^3}+\sum_{m>n/2}^{n}\frac{1}{(m-1)^{3/2}}\right]. \tag{A5.6}$$

Here the first sum does not exceed $(n/2)n^{-3/2}=1/(2\sqrt{n})$ and the last sum does not exceed

$$\int_{n/2-1}^{n}\frac{ds}{s^{3/2}}\le\frac{c}{\sqrt{n}}.$$

It remains to bound the middle sum. Setting $(u(n) := \max_{k\le n}(\Delta_k\sqrt{k})/\mu)$, we have

$$\sum_{m>n/2}^{n}\frac{\Delta_{m-1}}{d_m^3}\le \mu u(n-1)\sqrt{\frac{2}{n}}\sum_{m>n/2}^{n}\frac{1}{(n-m+\alpha^2)^{3/2}}.$$

The last sum does not exceed

$$\sum_{k=0}^{\infty}\frac{1}{(k+\alpha^2)^{3/2}}\le\frac{1}{\alpha^3}+\int_0^\infty\frac{dt}{(t+\alpha^2)^{3/2}}=\frac{1}{\alpha^3}+\frac{1}{2\alpha}\le\frac{3}{2\alpha},$$

provided that $\alpha\ge 1$. Collecting (A5.6) and the above estimates together, we obtain the assertion of the lemma. □

We now turn directly to the proof of the theorem. By virtue of (A5.1) and (A5.2),

$$v(n):=\frac{\Delta_n\sqrt{n}}{\mu}\le\frac{2}{\mu}\sqrt{n}\Delta_{n,\varepsilon}+\frac{2\alpha}{\mu}\le 2c+\frac{2c\mu u(n-1)}{\alpha}+\frac{2\alpha}{\mu}.$$

Put here $\alpha := \max(4c\mu, 1)$. Then $(\mu\ge 1)$

$$v(n)\le c_1+\frac{u(n+1)}{2}.$$

This implies that $u(n)\le 2c_1$ for all n. To verify this, we make use of induction. Clearly, $u(1)=v(1)\le 1\le 2c_1$. Let $u(n-1)\le 2c_1$. Then $v(n)\le 2c_1$ and $u(n)=\max(v(n),u(n-1))\le 2c_1$. The theorem is proved. □

Appendix 6
The Basic Properties of Regularly Varying Functions and Subexponential Distributions

The properties of regularly varying functions and subexponential distributions were used in Sects. 8.8, 9.4–9.6 and 12.7 and will be used in Appendices 7 and 8.

6.1 General Properties of Regularly Varying Functions

Definition A6.1.1 A positive measurable function $L(t)$ is called a *slowly varying function* (s.v.f.) as $t \to \infty$ if, for any fixed $v > 0$,

$$\frac{L(vt)}{L(t)} \to 1 \quad \text{as } t \to \infty. \tag{A6.1.1}$$

A function $V(t)$ is called a *regularly varying function* (r.v.f.) (with exponent $-\beta \in \mathbb{R}$) as $t \to \infty$ if it can be represented as

$$V(t) = t^{-\beta} L(t), \tag{A6.1.2}$$

where $L(t)$ is an s.v.f. as $t \to \infty$. We will denote the class of all r.v.f.s by $\mathfrak{R}$.

The definitions of an s.v.f. and r.v.f. as $t \downarrow 0$ are quite similar. In what follows, the term s.v.f. (r.v.f.) will (unless specified otherwise) always refer to a slowly (regularly) varying function at infinity.

It is easy to see that, similarly to (A6.1.1), a characteristic property of regularly varying functions is the convergence, for any fixed $v > 0$,

$$\frac{V(vt)}{V(t)} \to v^{-\beta} \quad \text{as } t \to \infty. \tag{A6.1.3}$$

Thus, an s.v.f. is an r.v.f. with exponent zero.

Typical representatives of the class of s.v.f.s are the logarithmic function and its powers $\ln^\gamma t$, $\gamma \in \mathbb{R}$, their linear combinations, multiple logarithms, functions with

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013

the property $L(t) \to L = \text{const} \neq 0$ as $t \to \infty$, etc. As an example of a *bounded oscillating* s.v.f. one can give

$$L_0(t) = 2 + \sin(\ln \ln t), \quad t > 1.$$

We will need the following two basic properties of s.v.f.s.

Theorem A6.1.1 (Uniform convergence theorem) *If $L(t)$ is an s.v.f. as $t \to \infty$ then convergence* (A6.1.1) *holds uniformly in v on any segment* $[v_1, v_2]$, $0 < v_1 < v_2 < \infty$.

The theorem implies that the uniform convergence (A6.1.1) on the segment $[1/M, M]$ also takes place in the case when, as $t \to \infty$, the quantity $M = M(t)$ grows unboundedly slowly enough.

Theorem A6.1.2 (Integral representation) *A function $L(t)$ is an s.v.f. as $t \to \infty$ if and only if, for some $t_0 > 0$, one has*

$$L(t) = c(t) \exp\left\{ \int_{t_0}^{t} \frac{\varepsilon(u)}{u} \, du \right\}, \quad t \geq t_0, \tag{A6.1.4}$$

where the functions $c(t)$ and $\varepsilon(t)$ are measurable and such that $c(t) \to c \in (0, \infty)$ and $\varepsilon(t) \to 0$ as $t \to \infty$.

For instance, for $L(t) = \ln t$ representation (A6.1.4) is valid with $c(t) = 1$, $t_0 = e$ and $\varepsilon(t) = (\ln t)^{-1}$.

Proof of Theorem A6.1.1 Put

$$h(x) := \ln L\big(e^x\big). \tag{A6.1.5}$$

Then property (A6.1.1) of s.v.f.s is equivalent, for each $u \in \mathbb{R}$, to the condition that the convergence

$$h(x+u) - h(x) \to 0 \tag{A6.1.6}$$

takes place as $x \to \infty$. To prove the theorem, we need to show that this convergence is uniform in $u \in [u_1, u_2]$ for any fixed $u_i \in \mathbb{R}$. In order to do that, it suffices to verify that convergence (A6.1.6) is uniform on the segment [0, 1]. Indeed, from the obvious inequality

$$\big|h(x+u_1+u_2) - h(x)\big| \leq \big|h(x+u_1+u_2) - h(x+u_1)\big| + \big|h(x+u_1) - h(x)\big| \tag{A6.1.7}$$

we have

$$\big|h(x+u) - h(x)\big| \leq (u_2 - u_1 + 1) \sup_{y \in [0,1]} \big|h(x+y) - h(x)\big|, \quad u \in [u_1, u_2].$$

For a given $\varepsilon \in (0, 1)$ and an $x > 0$, set $I_x := [x, x + 2]$,

$$I_x^* := \{u \in I_x : |h(u) - h(x)| \geq \varepsilon/2\}, \quad I_{0,x}^* := \{u \in I_0 : |h(x+u) - h(x)| \geq \varepsilon/2\}.$$

Clearly, the sets I_x^* and $I_{0,x}^*$ are measurable and differ from each other by a translation by x, so that $\boldsymbol{\mu}(I_x^*) = \boldsymbol{\mu}(I_{0,x}^*)$, where $\boldsymbol{\mu}$ is the Lebesgue measure. By (A6.1.6) the indicator function of the set $I_{0,x}^*$ converges, at each point $u \in I_0$, to 0 as $x \to \infty$. Therefore, by the dominated convergence theorem, the integral of this function, being equal to $\boldsymbol{\mu}(I_{0,x}^*)$, converges to 0, so that $\boldsymbol{\mu}(I_x^*) < \varepsilon/2$ for $x \geq x_0$, where x_0 is large enough.

Further, for $s \in [0, 1]$, the segment $I_x \cap I_{x+s} = [x+s, x+2]$ has length $2-s \geq 1$, so that, for $x \geq x_0$, the set

$$(I_x \cap I_{x+s}) \setminus (I_x^* \cup I_{x+s}^*)$$

has measure $\geq 1 - \varepsilon > 0$ and hence is non-empty. Let y be a point from this set. Then

$$|h(x+s) - h(x)| \leq |h(x+s) - h(y)| + |h(y) - h(x)| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

for $x \geq x_0$, which proves the required uniformity on $[0, 1]$ and hence on any fixed segment. The theorem is proved. □

Proof of Theorem A6.1.2 The fact that the right-hand side of (A6.1.4) is an s.v.f. is almost obvious: for any fixed $v \neq 1$,

$$\frac{L(vt)}{L(t)} = \frac{c(vt)}{c(t)} \exp\left\{\int_t^{vt} \frac{\varepsilon(u)}{u} du\right\}, \tag{A6.1.8}$$

where $c(vt)/c(t) \to c/c = 1$ and, as $t \to \infty$,

$$\int_t^{vt} \frac{\varepsilon(u)}{u} du = o\left(\int_t^{vt} \frac{du}{u}\right) = o(\ln v) = o(1). \tag{A6.1.9}$$

We now prove that any s.v.f. admits the representation (A6.1.4). The required representation in terms of the function (A6.1.5) is equivalent (after substituting $t = e^x$) to the relation

$$h(x) = d(x) + \int_{x_0}^x \delta(y)\, dy, \tag{A6.1.10}$$

where $d(x) = \ln c(e^x) \to d \in \mathbb{R}$ and $\delta(x) = \varepsilon(e^x) \to 0$ as $x \to \infty$, $x_0 = \ln t_0$. Therefore it suffices to establish representation (A6.1.10) for the function $h(x)$.

First of all note that $h(x)$ (as well as $L(t)$) is a "locally bounded" function. Indeed, Theorem A6.1.1 implies that, for x_0 large enough and all $x \geq x_0$,

$$\sup_{0 \leq y \leq 1} |h(x+y) - h(x)| < 1.$$

Hence, for any $x > x_0$, we have by virtue of (A6.1.7) the bound

$$\bigl|h(x) - h(x_0)\bigr| \le x - x_0 + 1.$$

Further, the local boundedness and measurability of the function h mean that it is locally integrable on $[x_0, \infty)$ and hence can be represented for $x \ge x_0$ as

$$h(x) = \int_{x_0}^{x_0+1} h(y)\,dy + \int_0^1 \bigl(h(x) - h(x+y)\bigr)\,dy + \int_{x_0}^{x} \bigl(h(y+1) - h(y)\bigr)\,dy. \tag{A6.1.11}$$

The first integral in (A6.1.11) is a constant, which will be denoted by d. The second integral, by virtue of Theorem A6.1.1, converges to zero as $x \to \infty$, so that

$$d(x) := d + \int_0^1 \bigl(h(x) - h(x+y)\bigr)\,dy \to d, \quad x \to \infty.$$

As for the third integral in (A6.1.11), by the definition of an s.v.f., the integrand satisfies

$$\delta(y) := h(y+1) - h(y) \to 0$$

as $y \to \infty$, which completes the proof of representation (A6.1.10). □

6.2 The Basic Asymptotic Properties

In this section we will obtain a number of consequences of Theorems A6.1.1 and A6.1.2 that are related to the asymptotic behaviour of s.v.f.s and r.v.f.s.

Theorem A6.2.1 (i) *If L_1 and L_2 are s.v.f.s then $L_1 + L_2$, $L_1 L_2$, L_1^b and $L(t) := L_1(at+b)$, where $a \ge 0$ and $b \in \mathbb{R}$, are also s.v.f.s*

(ii) *If L is an s.v.f. then, for any $\delta > 0$, there exists a $t_\delta > 0$ such that*

$$t^{-\delta} \le L(t) \le t^{\delta} \quad \text{for all } t \ge t_\delta. \tag{A6.2.1}$$

In other words, $L(t) = t^{o(1)}$ as $t \to \infty$.

(iii) *If L is an s.v.f. then, for any $\delta > 0$ and $v_0 > 1$, there exists a $t_\delta > 0$ such that, for all $v \ge v_0$ and $t \ge t_\delta$,*

$$v^{-\delta} \le \frac{L(vt)}{L(t)} \le v^{\delta}. \tag{A6.2.2}$$

(iv) (*Karamata's theorem*) *If an r.v.f. V in (A6.1.2) has exponent $-\beta$, $\beta > 1$, then*

$$V^I(t) := \int_t^\infty V(u)\,du \sim \frac{tV(t)}{\beta - 1} \quad \text{as } t \to \infty. \tag{A6.2.3}$$

If $\beta < 1$ *then*

$$V_I(t) := \int_0^t V(u)\,du \sim \frac{tV(t)}{1-\beta} \quad \text{as } t \to \infty. \tag{A6.2.4}$$

If $\beta = 1$ *then*

$$V_I(t) = tV(t)L_1(t) \tag{A6.2.5}$$

and

$$V^I(t) = tV(t)L_2(t) \quad \text{if } \int_0^\infty V(u)\,du < \infty, \tag{A6.2.6}$$

where $L_i(t) \to \infty$ *as* $t \to \infty$, $i = 1, 2$, *are s.v.f.s.*

(v) *For an r.v.f.* V *with exponent* $-\beta < 0$, *put*

$$b(t) := V^{(-1)}(1/t) = \inf\{u : V(u) < 1/t\}.$$

Then $b(t)$ *is an r.v.f. with exponent* $1/\beta$:

$$b(t) = t^{1/\beta} L_b(t), \tag{A6.2.7}$$

where L_b *is an s.v.f. If the function* L *possesses the property*

$$L\big(tL^{1/\beta}(t)\big) \sim L(t) \tag{A6.2.8}$$

as $t \to \infty$ *then*

$$L_b(t) \sim L^{1/\beta}\big(t^{1/\beta}\big). \tag{A6.2.9}$$

Similar assertions hold for functions slowly/regularly varying as $t \downarrow 0$.

Note that Theorem A6.1.1 and inequality (A6.2.2) imply the following property of s.v.f.s: *for any* $\delta > 0$ *there exists a* $t_\delta > 0$ *such that, for all* t *and* v *satisfying the inequalities* $t \ge t_\delta$ *and* $vt \ge t_\delta$, *we have*

$$(1-\delta)\min\{v^\delta, v^{-\delta}\} \le \frac{L(vt)}{L(t)} \le (1+\delta)\max\{v^\delta, v^{-\delta}\}. \tag{A6.2.10}$$

Proof of Theorem A6.2.1 Assertion (i) is evident (just note that, in order to prove the last part of (i), one needs Theorem A6.1.1).

(ii) This property follows immediately from representation (A6.1.4) and the bound

$$\left|\int_{t_0}^t \frac{\varepsilon(u)}{u}\,du\right| = \left|\int_{t_0}^{\ln t} + \int_{\ln t}^t\right| = O\left(\int_{t_0}^{\ln t} \frac{du}{u}\right) + o\left(\int_{\ln t}^t \frac{du}{u}\right) = o(\ln t)$$

as $t \to \infty$.

(iii) In order to prove this property, notice that on the right-hand side of (A6.1.8), for any fixed $\delta > 0$ and $v_0 > 1$ and all t large enough, we have

$$v^{-\delta/2} \le v_0^{-\delta/2} \le \frac{c(vt)}{c(t)} \le v_0^{\delta/2} \le v^{\delta/2}, \quad v \ge v_0,$$

and

$$\left| \int_t^{vt} \frac{\varepsilon(u)}{u} du \right| \le \frac{\delta}{2} \ln v$$

(by virtue of (A6.1.9)). This implies (A6.2.2).

(iv) By the dominated convergence theorem, we can choose an $M = M(t) \to \infty$ as $t \to \infty$ such that the convergence in (A6.1.1) will be uniform in $v \in [1, M]$. Changing the variable $u = vt$, we obtain

$$V^I(t) = t^{-\beta+1} L(t) \int_1^\infty v^{-\beta} \frac{L(vt)}{L(t)} dv = t^{-\beta+1} L(t) \left[\int_1^M + \int_M^\infty \right]. \tag{A6.2.11}$$

If $\beta > 1$ then, as $t \to \infty$,

$$\int_1^M \sim \int_1^M v^{-\beta} dv \to \frac{1}{\beta - 1},$$

whereas by property (iii), for $\delta = (\beta - 1)/2$, we have

$$\int_M^\infty < \int_M^\infty v^{-\beta+\delta} dv = \int_M^\infty v^{-(\beta+1)/2} dv \to 0.$$

These relations together imply

$$V^I(t) \sim \frac{t^{-\beta+1}}{\beta - 1} L(t) = \frac{tV(t)}{\beta - 1}.$$

The case $\beta < 1$ can be treated quite similarly, but taking into account the uniform in $v \in [1/M, 1]$ convergence in (A6.1.1) and the equality

$$\int_0^1 v^{-\beta} dv = \frac{1}{1 - \beta}.$$

If $\beta = 1$ then the first integral on the right-hand side of (A6.2.11) is

$$\int_1^M \sim \int_1^M v^{-1} dv = \ln M,$$

so that if

$$\int_0^\infty V(u)\, du < \infty \tag{A6.2.12}$$

then

$$V^I(t) \ge \big(1 + o(1)\big) L(t) \ln M \gg L(t) \tag{A6.2.13}$$

and hence

$$L_2(t) := \frac{V^I(t)}{tV(t)} = \frac{V^I(t)}{L(t)} \to \infty \quad \text{as } t \to \infty.$$

Note now that, by property (i), the function L_2 will be an s.v.f. whenever $V^I(t)$ is an s.v.f. But, for $v > 1$,

$$V^I(t) = V^I(vt) + \int_t^{vt} V(u)\,du,$$

where the last integral clearly does not exceed $(v-1)L(t)(1+o(1))$. By (A6.2.13) this implies that $V^I(vt)/V^I(t) \to 1$ as $t \to \infty$, which completes the proof of (A6.2.6).

That relation (A6.2.5) is true in the subcase when (A6.2.12) holds is almost obvious, since

$$V_I(t) = tV(t)L_1(t) = L(t)L_1(t) = \int_0^t V(u)\,du \to \int_0^\infty V(u)\,du,$$

so that, firstly, L_1 is an s.v.f. by property (i) and, secondly, $L_1(t) \to \infty$ because $L(t) \to 0$ by (A6.2.13).

Now let $\beta = 1$ and $\int_0^\infty V(u)\,du = \infty$. Then, as $M = M(t) \to \infty$ slowly enough, similarly to (A6.2.11) and (A6.2.13), by the uniform convergence theorem we have

$$V_I(t) = \int_0^1 v^{-1}L(vt)\,dv \ge \int_{1/M}^1 v^{-1}L(vt)\,dv \sim L(t)\ln M \gg L(t).$$

Therefore $L_1(t) := V_I(t)/L(t) \to \infty$ as $t \to \infty$. Further, also similarly to the above, we have, as $v \in (0, 1)$,

$$V_I(t) = V_I(vt) + \int_{vt}^t V(u)\,du,$$

where the last integral does not exceed $(1-v)L(t)(1+o(1)) \ll V_I(t)$, so that $V_I(t)$ (as well as $L_1(t)$ by virtue of property (i)) is an s.v.f. This completes the proof of property (iv).

(v) Clearly, by the uniform convergence theorem the quantity $b = b(t)$ is a solution to the "asymptotic equation"

$$V(b) \sim \frac{1}{t} \quad \text{as } t \to \infty \tag{A6.2.14}$$

(where the symbol $\sim$ can be replaced by the equality sign if the function V is continuous and monotonically decreasing). Substituting $t^{1/\beta}L_b(t)$ for b, we obtain an equivalent relation

$$L_b^{-\beta}L\big(t^{1/\beta}L_b\big) \sim 1, \tag{A6.2.15}$$

where clearly

$$t^{1/\beta}L_b \to \infty \quad \text{as } t \to \infty. \tag{A6.2.16}$$

Fix an arbitrary $v > 0$. Substituting vt for t in (A6.2.15) and setting, for brevity's sake, $L_2 = L_2(t) := L_b(vt)$, we get the relation

$$L_2^{-\beta} L\big(t^{1/\beta} L_2\big) \sim 1, \tag{A6.2.17}$$

since $L(v^{1/\beta} t^{1/\beta} L_2) \sim L(t^{1/\beta} L_2)$ by virtue of (A6.2.16) (with L_b replaced with L_2). Now we will show by contradiction that (A6.2.15)–(A6.2.17) imply that $L_b \sim L_2$ as $t \to \infty$, which obviously means that L_b is an s.v.f.

Indeed, the contrary assumption means that there exist $v_0 > 1$ and a sequence $t_n \to \infty$ such that

$$u_n := L_2(t_n)/L_b(t_n) > v_0, \quad n = 1, 2, \ldots \tag{A6.2.18}$$

(the possible alternative case can be dealt with in the same way). Clearly, $t_n^* := t_n^{1/\beta} L_b(t_n) \to \infty$ by (A6.2.16), so we obtain from (A6.2.15)–(A6.2.16) and property (iii) with $\delta = \beta/2$ that

$$1 \sim \frac{L_2^{-\beta}(t_n) L(t_n^{1/\beta} L_2(t_n))}{L_b^{-\beta}(t_n) L(t_n^{1/\beta} L_b(t_n))} = u_n^{-\beta} \frac{L(u_n t_n^*)}{L(t_n^*)} \le u_n^{-\beta/2} < v_0^{-\beta/2} < 1.$$

We get a contradiction.

Note that the above argument proves the uniqueness (up to asymptotic equivalence) of the solution to Eq. (A6.2.14).

Finally, relation (A6.2.9) can be proved by a direct verification of (A6.2.14) for $b := t^{1/\beta} L^{1/\beta}(t^{1/\beta})$: using (A6.2.8), we have

$$V(b) = b^{-\beta} L(b) = \frac{L(t^{1/\beta} L^{1/\beta}(t^{1/\beta}))}{t L(t^{1/\beta})} \sim \frac{L(t^{1/\beta})}{t L(t^{1/\beta})} = \frac{1}{t}.$$

The required assertion follows now by the aforementioned uniqueness of the solution to the asymptotic equation (A6.2.14). Theorem A6.2.1 is proved. □

6.3 The Asymptotic Properties of the Transforms of R.V.F.s (Abel-Type Theorems)

For an r.v.f. $V(t)$, its Laplace transform

$$\psi(\lambda) := \int_0^\infty e^{-\lambda t} V(t)\, dt < \infty$$

is defined for all $\lambda > 0$. The following asymptotic relations hold true for the transform.

Theorem A6.3.1 *Assume that $V(t) \in \mathfrak{R}$ (i.e. $V(t)$ has the form (A6.1.2)).*

(i) *If* $\beta \in [0, 1)$ *then*

$$\psi(\lambda) \sim \frac{\Gamma(1-\beta)}{\lambda} V(1/\lambda) \quad \text{as } \lambda \downarrow 0. \tag{A6.3.1}$$

(ii) *If* $\beta = 1$ *and* $\int_0^\infty V(t)\,dt = \infty$ *then*

$$\psi(\lambda) \sim V_I(1/\lambda) \quad \text{as } \lambda \downarrow 0, \tag{A6.3.2}$$

where $V_I(t) = \int_0^t V(u)\,du \to \infty$ *is an s.v.f. such that* $V_I(t) \gg L(t)$ *as* $t \to \infty$.

(iii) *In any case,* $\psi(\lambda) \uparrow V_I(\infty) = \int_0^\infty V(t)\,dt \le \infty$ *as* $\lambda \downarrow 0$.

Assertions (i) and (ii) are called *Abelian* theorems.

If we resolve relation (A6.3.1) for V then we obtain

$$V(t) \sim \frac{\psi(1/t)}{t\Gamma(1-\beta)} \quad \text{as } t \to \infty.$$

Relations of this type will also be valid in the case when, instead of the regularity of the function V, one requires the monotonicity of V and assumes that $\psi(\lambda)$ is an r.v.f. as $\lambda \downarrow 0$. Statements of such type are called *Tauberian* theorems. We will not need these theorems and so will not dwell on them.

Proof of Theorem A6.3.1 (i) For any fixed $\varepsilon > 0$ we have

$$\psi(\lambda) = \int_0^{\varepsilon/\lambda} + \int_{\varepsilon/\lambda}^\infty,$$

where, for the first integral on the right-hand side, for $\beta < 1$, by virtue of (A6.2.4) we have the following relation

$$\int_0^{\varepsilon/\lambda} e^{-\lambda t} V(t)\,dt \le \int_0^{\varepsilon/\lambda} V(t)\,dt \sim \frac{\varepsilon V(\varepsilon/\lambda)}{\lambda(1-\beta)} \quad \text{as } \lambda \downarrow 0. \tag{A6.3.3}$$

Changing the variable $\lambda t = u$, we can rewrite the second integral in the above representation for $\psi(\lambda)$ as

$$\int_{\varepsilon/\lambda}^\infty = \frac{V(1/\lambda)}{\lambda} \int_\varepsilon^\infty e^{-u} u^{-\beta} \frac{L(u/\lambda)}{L(1/\lambda)}\,du = \frac{V(1/\lambda)}{\lambda}\left[\int_\varepsilon^2 + \int_2^\infty\right]. \tag{A6.3.4}$$

Each of the integrals on the right-hand side converges, as $\lambda \downarrow 0$, to the corresponding integral of $e^{-u}u^{-\beta}$: the former integral converges by the uniform convergence theorem (the convergence $L(u/\lambda)/L(1/\lambda) \to 1$ is uniform in $u \in [\varepsilon, 2]$), and the latter converges by virtue of (A6.1.1) and the dominated convergence theorem, since by Theorem A6.2.1(iii), for all λ small enough, we have $L(u/\lambda)/L(1/\lambda) < u$ for $u \ge 2$. Therefore,

$$\int_{\varepsilon/\lambda}^\infty \sim \frac{V(1/\lambda)}{\lambda} \int_\varepsilon^\infty u^{-\beta} e^{-u}\,du. \tag{A6.3.5}$$

Now note that, as $\lambda \downarrow 0$,

$$\frac{\varepsilon V(\varepsilon/\lambda)}{\lambda} \Big/ \frac{V(1/\lambda)}{\lambda} = \varepsilon^{1-\beta} \frac{L(\varepsilon/\lambda)}{L(1/\lambda)} \to \varepsilon^{1-\beta}.$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, this relation together with (A6.3.3) and (A6.3.5) completes the proof of (A6.3.1).

(ii) Integrating by parts and changing the variable $\lambda t = u$, we obtain, for $\beta = 1$ and $M > 0$, that

$$\begin{aligned}\psi(\lambda) &= \int_0^\infty e^{-\lambda t} dV_I(t) = -\int_0^\infty V_I(t)\, de^{-\lambda t} \\ &= \int_0^\infty V_I(u/\lambda)e^{-u} du = \int_0^{1/M} + \int_{1/M}^M + \int_M^\infty .\end{aligned} \tag{A6.3.6}$$

By Theorem A6.2.1(iv), $V_I(t) \gg L(t)$ is an s.v.f. as $t \to \infty$. Therefore, by the uniform convergence theorem, for $M = M(\lambda) \to \infty$ slowly enough as $\lambda \to 0$, the middle integral on the right-hand side of (A6.3.6) is

$$V_I(1/\lambda) \int_{1/M}^M \frac{V_I(u/\lambda)}{V_I(1/\lambda)} e^{-u} du \sim V_I(1/\lambda) \int_{1/M}^M e^{-u} du \sim V_I(1/\lambda).$$

The remaining two integrals are negligibly small: since $V_I(t)$ is an increasing function, the first integral does not exceed $V_I(1/\lambda M)/M = o(V_I(1/\lambda))$, while for the last integral we have by Theorem A6.2.1(iii) that

$$V_I(1/\lambda) \int_M^\infty \frac{V_I(u/\lambda)}{V_I(1/\lambda)} e^{-u} du \le V_I(1/\lambda) \int_M^\infty u e^{-u} du = o\big(V_I(1/\lambda)\big).$$

Hence (ii) is proved. Assertion (iii) is evident. □

6.4 Subexponential Distributions and Their Properties

Let $\xi, \xi_1, \xi_2, \ldots$ be independent identically distributed random variables with distribution $\mathbf{F}$, and let the *right tail of this distribution*

$$F_+(t) := \mathbf{F}\big([t, \infty)\big) = \mathbf{P}(\xi \ge t), \quad t \in \mathbb{R},$$

be an r.v.f. as $t \to \infty$ of the form (A6.1.2), which we will denote by $V(t)$. Recall that *we denoted the class of all such distributions by* $\mathcal{R}$.

In this section we will introduce one more class of distributions, which is substantially wider than $\mathcal{R}$.

Let $\zeta \in \mathbb{R}$ be a random variable with distribution $\mathbf{G}$: $\mathbf{G}(B) = \mathbf{P}(\zeta \in B)$ for any Borel set B (recall that in this case we write $\zeta \Subset \mathbf{G}$). Denote by $G(t)$ the right tail of the distribution of the random variable ζ:

$$G(t) := \mathbf{P}(\zeta \ge t), \quad t \in \mathbb{R}.$$

The convolution of tails $G_1(t)$ and $G_2(t)$ is the function

$$G_1 * G_2(t) := -\int G(t-y)dG_2(y) = \int G_1(t-y)\mathbf{G}_2(dy) = \mathbf{P}(Z_2 \ge t),$$

where $Z_2 = \zeta_1 + \zeta_2$ is the sum of independent random variables $\zeta_i \Subset \mathbf{G}_i$, $i = 1, 2$. Clearly, $G_1 * G_2(t) = G_2 * G_1(t)$. Denote by $G^{2*}(t) := G * G(t)$ the convolution of the tail $G(t)$ with itself and put $G^{(n+1)*}(t) := G * G^{n*}(t)$, $n \ge 2$.

Definition A6.4.1 A distribution $\mathbf{G}$ on $[0, \infty)$ belongs to the class $\mathcal{S}_+$ of *subexponential distributions on the positive half-line* if

$$G^{2*}(t) \sim 2G(t) \quad \text{as } t \to \infty. \tag{A6.4.1}$$

A distribution $\mathbf{G}$ on the whole line $(-\infty, \infty)$ belongs to the class $\mathcal{S}$ of *subexponential distributions* if the distribution $\mathbf{G}^+$ of the positive part $\zeta^+ = \max\{0, \zeta\}$ of the random variable $\zeta \Subset \mathbf{G}$ belongs to $\mathcal{S}_+$. A random variable is called subexponential if its distribution is subexponential.

As we will see below (Theorem A6.4.3), the subexponentiality property of a distribution $\mathbf{G}$ is essentially the property of the asymptotics of the tail $G(t)$ as $t \to \infty$. Therefore we can also speak about *subexponential functions*.

A nondecreasing function $G_1(t)$ on $(0, \infty)$ is called *subexponential* if a distribution $\mathbf{G}$ with the tail $G(t) \sim cG_1(t)$ as $t \to \infty$ with some $c > 0$ is subexponential. (For example, distributions with the tails $G(t) = G_1(t)/G_1(0)$ or $G(t) = \min(1, G_1(t))$).

Remark A6.4.1 Since we obviously always have

$$\begin{aligned}
\left(G^+\right)^{2*}(t) &= \mathbf{P}\left(\zeta_1^+ + \zeta_2^+ \ge t\right) \ge \mathbf{P}\left(\left\{\zeta_1^+ \ge t\right\} \cup \left\{\zeta_2^+ \ge t\right\}\right) \\
&= \mathbf{P}(\zeta_1 \ge t) + \mathbf{P}(\zeta_2 \ge t) - \mathbf{P}(\zeta_1 \ge t, \zeta_2 \ge t) \\
&= 2G(t) - G^2(t) = 2G^+(t)\left(1 + o(1)\right)
\end{aligned}$$

as $t \to \infty$, subexponentiality is equivalent to the following property:

$$\limsup_{t \to \infty} \frac{(G^+)^{2*}(t)}{G^+(t)} \le 2. \tag{A6.4.2}$$

Note also that, since relation (A6.4.1) makes sense only when $G(t) > 0$ for all $t \in \mathbb{R}$, the support of any subexponential distribution is unbounded from the right.

We show that regularly varying distributions are subexponential, i.e., that $\mathcal{R} \subset \mathcal{S}$. Let $\mathbf{F} \in \mathcal{R}$ and $\mathbf{P}(\xi \ge t) = V(t)$ be r.v.f.s. We need to show that

$$\begin{aligned}
\mathbf{P}(\xi_1 + \xi_2 \ge x) &= V^{2*}(x) := V * V(x) \\
&= -\int_{-\infty}^{\infty} V(x-t)\, dV(t) \sim 2V(x).
\end{aligned} \tag{A6.4.3}$$

In order to do that, we introduce events $A := \{\xi_1 + \xi_2 \ge x\}$ and $B_i := \{\xi_i < x/2\}$, $i = 1, 2$. Clearly,

$$\mathbf{P}(A) = \mathbf{P}(AB_1) + \mathbf{P}(AB_2) - \mathbf{P}(AB_1B_2) + \mathbf{P}(A\overline{B}_1\overline{B}_2),$$

where $\mathbf{P}(AB_1B_2) = 0$, $\mathbf{P}(A\overline{B}_1\overline{B}_2) = \mathbf{P}(\overline{B}_1\overline{B}_2) = V^2(x/2)$ (here and in what follows, $\overline{B}$ denotes the event complementary to B) and

$$\mathbf{P}(AB_1) = \mathbf{P}(AB_2) = \int_{-\infty}^{x/2} V(x-t)\,\mathbf{F}(dt).$$

Therefore

$$V^{2*}(x) = 2\int_{-\infty}^{x/2} V(x-t)\,\mathbf{F}(dt) + V^2(x/2). \tag{A6.4.4}$$

(The same result can be obtained by integrating the convolution in (A6.4.3) by parts.) It remains to note that $V^2(x/2) = o(V(x))$ and

$$\int_{-\infty}^{x/2} V(x-t)\,\mathbf{F}(dt) = \int_{-\infty}^{-M} + \int_{-M}^{M} + \int_{M}^{x/2}, \tag{A6.4.5}$$

where, as one can easily see, for any $M = M(x) \to \infty$ as $x \to \infty$ such that $M = o(x)$, we have

$$\int_{-M}^{M} \sim V(x) \quad \text{and} \quad \int_{-\infty}^{-M} + \int_{M}^{x/2} = o\big(V(x)\big),$$

which proves (A6.4.3).

The same argument is valid for distributions with a right tail of the form

$$F_+(t) = e^{-t^\beta L(t)}, \quad \beta \in (0,1), \tag{A6.4.6}$$

where $L(t)$ is an s.v.f. as $t \to \infty$ satisfying a certain smoothness condition (for instance, that L is differentiable with $L'(t) = o(L(t)/t)$ as $t \to \infty$).

One of the basic properties of subexponential distributions $\mathbf{G}$ is that their tails $G(t)$ are asymptotically locally constant in the following sense.

Definition A6.4.2 We will call a function $G(t) > 0$ (asymptotically) *locally constant* (l.c.) if, for any fixed v,

$$\frac{G(t+v)}{G(t)} \to 1 \quad \text{as } t \to \infty. \tag{A6.4.7}$$

In the literature, distributions with l.c. tails are often referred to as long-tailed distributions; however, we feel that the term "locally constant function" better reflects the meaning of the concept. Denote the class of all distributions $\mathbf{G}$ with l.c. tails $G(t)$ by $\mathcal{L}$.

For future reference, we will state the basic properties of l.c. functions as a separate theorem.

Theorem A6.4.1 (i) *For an l.c. function $G(t)$ the convergence in* (A6.4.7) *is uniform in v on any fixed finite interval.*

(ii) *A function $G(t)$ is l.c. if and only if, for some $t_0 > 0$, it admits a representation of the form*

$$G(t) = c(t)\exp\left\{\int_{t_0}^{t} \varepsilon(u)\,du\right\}, \quad t \ge t_0, \tag{A6.4.8}$$

where the functions $c(t)$ and $\varepsilon(t)$ are measurable and such that $c(t) \to c \in (0,\infty)$ and $\varepsilon(t) \to 0$ as $t \to \infty$.

(iii) *If $G_1(t)$ and $G_2(t)$ are l.c. functions then $G_1(t)+G_2(t)$, $G_1(t)G_2(t)$, $G_1^b(t)$, and $G(t) := G_1(at+b)$, where $a \ge 0$ and $b \in \mathbb{R}$, are also l.c.*

(iv) *If $G(t)$ is an l.c. function then, for any $\varepsilon > 0$,*

$$e^{\varepsilon t}G(t) \to \infty \quad \text{as } t \to \infty.$$

In other words, any l.c. function $G(t)$ can be represented as

$$G(t) = e^{-l(t)}, \quad l(t) = o(t) \quad \text{as } t \to \infty. \tag{A6.4.9}$$

(v) *Let*

$$G^I(t) := \int_t^\infty G(u)\,du < \infty$$

and at least one of the following conditions be satisfied:

(a) *$G(t)$ is an l.c. function*; *or*
(b) *$G^I(t)$ is an l.c. function and $G(t)$ is monotone.*

Then

$$G(t) = o\big(G^I(t)\big) \quad \text{as } t \to \infty. \tag{A6.4.10}$$

(vi) *If $\mathbf{G} \in \mathcal{L}$ then $G^{2*}(t) \sim (G^+)^{2*}(t)$ as $t \to \infty$.*

Remark A6.4.2 Assertion (i) of the theorem implies that the uniform convergence in (A6.4.7) on the interval $[-M, M]$ persists in the case when, as $t \to \infty$, $M = M(t)$ grows unboundedly slowly enough.

Proof of Theorem A6.4.1, (i)–(iii) It is clear from Definitions A6.4.1 and A6.4.2 that $G(t)$ is l.c. if and only if $L(t) := G(\ln t)$ is an s.v.f. Having made this observation, assertion (i) follows directly from Theorem A6.1.1 (on uniform convergence of s.v.f.s), while assertions (ii) and (iii) follow from Theorems A6.1.2 and A6.2.1(i), respectively.

Assertion (iv) follows from the integral representation (A6.4.8).

(v) If (a) holds then, for any $M > 0$ and all t large enough,

$$G^I(t) > \int_t^{t+M} G(u)\,du > \frac{1}{2}MG(t).$$

Since M is arbitrary, $G^I(t) \gg G(t)$. Further, if (b) holds then

$$\frac{G(t)}{G^I(t)} \le \frac{1}{G^I(t)} \int_{t-1}^{t} G(u)\,du = \frac{G^I(t-1)}{G^I(t)} - 1 \to 0$$

as $t \to \infty$.

(vi) Let ζ_1 and ζ_2 be independent copies of a random variable ζ, $Z_2 := \zeta_1 + \zeta_2$, $Z_2^{(+)} := \zeta_1^+ + \zeta_2^+$. Clearly, $\zeta_i \le \zeta_i^+$, so that

$$G^{2*}(t) = \mathbf{P}(Z_2 \ge t) \le \mathbf{P}\big(Z_2^{(+)} \ge t\big) = \big(G^+\big)^{2*}(t). \tag{A6.4.11}$$

On the other hand, for any $M > 0$,

$$G^{2*}(t) \ge \mathbf{P}(Z_2 \ge t, \zeta_1 > 0, \zeta_2 > 0) + \sum_{i=1}^{2} \mathbf{P}\big(Z_2 \ge t, \zeta_i \in [-M, 0]\big),$$

where the first term on the right-hand side is equal to $\mathbf{P}(Z_2^{(+)} \ge t,\ \zeta_1^+ > 0,\ \zeta_2^+ > 0)$, and the last two terms can be bounded as follows: since $\mathbf{G} \in \mathcal{L}$, then, for any $\varepsilon > 0$ and M and t large enough,

$$\begin{aligned}
\mathbf{P}\big(Z_2 \ge t, \zeta_1 \in [-M, 0]\big) &\ge \mathbf{P}\big(\zeta_2 \ge t + M,\ \zeta_1 \in [-M, 0]\big) \\
&= G(t)\frac{G(t+M)}{G(t)}\big[\mathbf{P}(\zeta_1 \le 0) - \mathbf{P}(\zeta_1 < -M)\big] \\
&\ge (1-\varepsilon)G(t)\mathbf{P}\big(\zeta_1^+ = 0\big) = (1-\varepsilon)\mathbf{P}\big(Z_2^{(+)} \ge t,\ \zeta_1^+ = 0\big).
\end{aligned}$$

Thus we obtain for $G^{2*}(t)$ the lower bound

$$\begin{aligned}
G^{2*}(t) &\ge \mathbf{P}\big(Z_2^{(+)} \ge t, \zeta_1^+ > 0, \zeta_2^+ > 0\big) + (1-\varepsilon)\sum_{i=1}^{2} \mathbf{P}\big(Z_2^{(+)} \ge t, \zeta_i^+ = 0\big) \\
&\ge (1-\varepsilon)\mathbf{P}\big(Z_2^{(+)} \ge t\big) = (1-\varepsilon)\big(G^*\big)^{2*}(t).
\end{aligned}$$

Therefore (vi) is proved, as ε can be arbitrarily small. The theorem is proved. □

We return now to our discussion of subexponential distributions. First of all, we turn to the relationship between the classes $\mathcal{S}$ and $\mathcal{L}$.

Theorem A6.4.2 *We have $\mathcal{S} \subset \mathcal{L}$, and hence all the assertions of Theorem A6.4.1 are valid for subexponential distributions as well.*

Remark A6.4.3 The coinage of the term "subexponential distribution" was apparently due mostly to the fact that the tail of such a distribution decreases as $t \to \infty$ slower than any exponential function $e^{-\varepsilon t}$, as shown in Theorems A6.4.1(iv) and A6.4.2.

Remark A6.4.4 In the case when the distribution $\mathbf{G}$ is *not concentrated* on $[0,\infty)$, the tails' additivity condition (A6.4.1) alone is not sufficient for the function $G(t)$ to be l.c. (and hence for ensuring the "subexponential decay" of the distribution tail, cf. Remark A6.4.3). This explains the necessity of defining subexponentiality in the general case in terms of condition (A6.4.1) on the distribution $\mathbf{G}^+$ of the random variable ζ^+. Actually, as we will see below (Corollary A6.4.1), the subexponentiality of a distribution $\mathbf{G}$ on $\mathbb{R}$ *is equivalent* to the combination of conditions (A6.4.1) (on $\mathbf{G}$ itself) and $\mathbf{G} \in \mathcal{L}$.

The next example shows that, for random variables *assuming values of both signs*, condition (A6.4.1), generally speaking, does not imply the subexponential behaviour of $G(t)$.

Example A6.4.1 Let $\mu > 0$ be fixed and the right tail of the distribution $\mathbf{G}$ have the form

$$G(t) = e^{-\mu t} V(t), \tag{A6.4.12}$$

where $V(t)$ is an r.v.f. vanishing as $t \to \infty$ and such that

$$g(\mu) := \int_{-\infty}^{\infty} e^{\mu y} \mathbf{G}(dy) < \infty.$$

Similarly to (A6.4.4) and (A6.4.5), we have

$$G^{2*}(t) = 2\int_{-\infty}^{t/2} G(t-y)\,\mathbf{G}(dy) + G^2(t/2),$$

where

$$\begin{aligned}\int_{-\infty}^{t/2} G(t-y)\,\mathbf{G}(dy) &= e^{-\mu t}\int_{-\infty}^{t/2} e^{\mu y} V(t-y)\,\mathbf{G}(dy)\\ &= e^{-\mu t}\left[\int_{-\infty}^{-M} + \int_{-M}^{M} + \int_{M}^{t/2}\right].\end{aligned}$$

One can easily see that, for $M = M(t) \to \infty$ slowly enough as $t \to \infty$, we have

$$\int_{-M}^{M} e^{\mu y} V(t-y)\,\mathbf{G}(dy) \sim g(\mu)V(t), \qquad \int_{-\infty}^{-M} + \int_{M}^{t/2} = o\big(G(t)\big),$$

while

$$G^2(t/2) = e^{-\mu t} V^2(t/2) \le c e^{-\mu t} V^2(t) = o\big(G(t)\big).$$

Thus, we obtain

$$G^{2*}(t) \sim 2g(\mu) e^{-\mu t} V(t) = 2g(\mu)G(t), \tag{A6.4.13}$$

and it is clear that we can always find a distribution $\mathbf{G}$ (with a negative mean) such that $g(\mu) = 1$. In that case relation (A6.4.1) from the definition of subexponentiality will be satisfied, although $G(t)$ decreases exponentially fast and hence is not an l.c. function.

On the other hand, note that the class of distributions satisfying relation (A6.4.1) only is an extension of the class $\mathcal{S}$. Distributions in the former class possess many of the properties of distributions from $\mathcal{S}$.

Proof of Theorem A6.4.2 We have to prove that $\mathcal{S} \subset \mathcal{L}$. Since the definitions of both classes are given in terms of the right distribution tails, we can assume without loss of generality, that $\mathbf{G} \in \mathcal{S}_+$ (or just consider the distribution $\mathbf{G}^+$). For independent (nonnegative) $\zeta_i \Subset \mathbf{G}$ we have, for $t > 0$,

$$G^{2*}(t) = \mathbf{P}(\zeta_1 + \zeta_2 \geq t) = \mathbf{P}(\zeta_1 \geq t) + \mathbf{P}(\zeta_1 + \zeta_2 \geq t, \zeta_1 < t)$$
$$= G(t) + \int_0^t G(t-y)\,\mathbf{G}(dy). \tag{A6.4.14}$$

Since $G(t)$ is non-increasing and $G(0) = 1$, it follows that, for $t > v > 0$,

$$\frac{G^{2*}(t)}{G(t)} = 1 + \int_0^v \frac{G(t-y)}{G(t)}\,\mathbf{G}(dy) + \int_v^t \frac{G(t-y)}{G(t)}\,\mathbf{G}(dy)$$
$$\geq 1 + \big[1 - G(v)\big] + \frac{G(t-v)}{G(t)}\big[G(v) - G(t)\big].$$

Therefore, for t large enough (such that $G(v) - G(t) > 0$),

$$1 \leq \frac{G(t-v)}{G(t)} \leq \frac{1}{G(v) - G(t)}\left[\frac{G^{2*}(t)}{G(t)} - 2 + G(v)\right].$$

Since $\mathbf{G} \in \mathcal{S}_+$, the right-hand side of the last formula converges as $t \to \infty$ to the quantity $G(v)/G(v) = 1$ and hence $\mathbf{G} \in \mathcal{L}$. The theorem is proved. □

The next theorem contains several important properties of subexponential distributions.

Theorem A6.4.3 *Let* $\mathbf{G} \in \mathcal{S}$.

(i) *If* $G_i(t)/G(t) \to c_i$ *as* $t \to \infty$, $c_i \geq 0$, $i = 1, 2$, $c_1 + c_2 > 0$, *then*

$$G_1 * G_2(t) \sim G_1(t) + G_2(t) \sim (c_1 + c_2)G(t).$$

(ii) *If* $G_0(t) \sim cG(t)$ *as* $t \to \infty$, $c > 0$, *then* $\mathbf{G}_0 \in \mathcal{S}$.

(iii) *For any fixed* $n \geq 2$,

$$G^{n*}(t) \sim nG(t) \quad \text{as } t \to \infty. \tag{A6.4.15}$$

(iv) *For any* $\varepsilon > 0$ *there exists a* $b = b(\varepsilon) < \infty$ *such that*

$$\frac{G^{n*}(t)}{G(t)} \leq b(1+\varepsilon)^n$$

for all $n \geq 2$ *and* t.

In addition to assertions (i) and (ii) of the theorem, we can also show that if $\mathbf{G} \in \mathcal{S}$ and the function $m(t) \in \mathcal{L}$ possesses the property

$$0 < m_1 \le m(t) \le m_2 < \infty$$

then $G_1(t) = m(t)G(t) \in \mathcal{S}$.

Theorems A6.4.1(vi), A6.4.2 and A6.4.3(iii) imply the following simple statement elucidating the subexponentiality condition for random variables taking values of both signs.

Corollary A6.4.1 *A distribution* $\mathbf{G}$ *belongs to* $\mathcal{S}$ *if and only if* $\mathbf{G} \in \mathcal{L}$ *and* $G^{2*}(t) \sim 2G(t)$ *as* $t \to \infty$.

Remark A6.4.5 Evidently the asymptotic relation $G_1(t) \sim G_2(t)$ as $t \to \infty$ is an *equivalence relation* on the set of distributions on $\mathbb{R}$. Theorem A6.4.3(ii) means that the class $\mathcal{S}$ *is closed with respect to that equivalence*. One can easily see that in each of the equivalence subclasses of the class $\mathcal{S}$ with respect to this relation *there is always a distribution with an arbitrarily smooth tail* $G(t)$.

Indeed, let $p(t)$ be an infinitely differentiable probability density on $\mathbb{R}$ vanishing outside [0, 1] (we can take, e.g., $p(x) = c \cdot e^{-1/(x(1-x))}$ if $x \in (0, 1)$ and $p(x) = 0$ if $x \notin (0, 1)$). Now we "smooth" the function $l(t) := -\ln G(t)$, $\mathbf{G} \in \mathcal{S}$, putting

$$l_0(t) := \int p(t-u)l(u)\,du, \quad \text{and let } G_0(t) := e^{-l_0(t)}. \tag{A6.4.16}$$

Clearly, $G_0(t)$ is an infinitely differentiable function and, since $l(t)$ is nondecreasing and we actually integrate over $[t-1, t]$ only, one has $l(t-1) \le l_0(t) \le l(t)$ and hence by Theorem A6.4.2

$$1 \le \frac{G_0(t)}{G(t)} \le \frac{G(t-1)}{G(t)} \to 1 \quad \text{as } t \to \infty.$$

Thus, the distribution $\mathbf{G}_0$ is equivalent to the original $\mathbf{G}$. A simpler smoothing procedure leading to a less smooth asymptotically equivalent tail consists of replacing the function $l(t)$ with its linear interpolation with nodes at points $(k, l(k))$, k being an integer.

Therefore, *up to a summand* $o(1)$, we can always assume the function $l(t) = -\ln G(t)$, $\mathbf{G} \in \mathcal{S}$, to be arbitrarily smooth.

The aforesaid is clearly applicable to the class $\mathcal{L}$ as well: it is also closed with respect to the introduced equivalence, and each of its equivalence subclass contains arbitrarily smooth representatives.

Remark A6.4.6 Theorem A6.4.3(ii) and (iii) immediately implies that if $\mathbf{G} \in \mathcal{S}$ then also $\mathbf{G}^{n*} \in \mathcal{S}$, $n = 2, 3, \ldots$. Moreover, if we denote by $\mathbf{G}^{n\vee}$ the distribution of the maximum of independent identically distributed random variables $\zeta_1, \dots, \zeta_n \Subset \mathbf{G}$, then the evident relation

$$G^{n\vee}(t) = 1 - \big(1 - G(t)\big)^n \sim nG(t) \quad \text{as } t \to \infty \tag{A6.4.17}$$

and Theorem A6.4.3(ii) imply that $\mathbf{G}^{n\vee}$ also belongs to $\mathcal{S}$.

Relations (A6.4.17) and (A6.4.15) show that, in the case of a subexponential $\mathbf{G}$, the tail $G^{n*}(t)$ of the distribution of the sum of a fixed number n of independent identically distributed random variables $\zeta_i \Subset \mathbf{G}$ is asymptotically equivalent (as $t \to \infty$) to the tail $G^{n\vee}(t)$ of the maximum of these random variables, i.e., the "large" values of this sum are mainly due to by the presence of one "large" term ζ_i in the sum. It is easy to see that this property is characteristic of subexponentiality.

Remark A6.4.7 Note also that an assertion converse to what was stated at the beginning of Remark A6.4.6 is also valid: if $\mathbf{G}^{n*} \in \mathcal{S}$ for some $n \geq 2$ then $\mathbf{G} \in \mathcal{S}$ as well. That $\mathbf{G}^{n\vee} \in \mathcal{S}$ implies $\mathbf{G} \in \mathcal{S}$ evidently follows from (A6.4.17) and Theorem A6.4.3(ii).

Proof of Theorem A6.4.3 (i) First assume that $c_1 c_2 > 0$ and that both distributions $\mathbf{G}_i$ are concentrated on $[0, \infty)$. Fix an arbitrary $\varepsilon > 0$ and choose M large enough to have $G_i(M) < \varepsilon$, $i = 1, 2$, and $G(M) < \varepsilon$, and such that, for $t > M$,

$$(1-\varepsilon)c_i < \frac{G_i(t)}{G(t)} < (1+\varepsilon)c_i, \quad i = 1, 2, \qquad 1 - \varepsilon < \frac{G(t-M)}{G(t)} < 1 + \varepsilon \tag{A6.4.18}$$

(the last inequality holds by virtue of Theorem A6.4.2).

Let $\zeta \Subset \mathbf{G}$ and $\zeta_i \Subset \mathbf{G}_i$, $i = 1, 2$, be independent random variables. Then, for $t > 2M$, we have the representation

$$G_1 * G_2(t) = P_1 + P_2 + P_3 + P_4, \tag{A6.4.19}$$

where

$$\begin{aligned}
P_1 &:= \mathbf{P}\big(\zeta_1 \geq t - \zeta_2, \zeta_2 \in [0, M)\big), \\
P_2 &:= \mathbf{P}\big(\zeta_2 \geq t - \zeta_1,\ \zeta_1 \in [0, M)\big), \\
P_3 &:= \mathbf{P}\big(\zeta_2 \geq t - \zeta_1,\ \zeta_1 \in [M, t - M)\big), \\
P_4 &:= \mathbf{P}(\zeta_2 \geq M, \zeta_1 \geq t - M)
\end{aligned}$$

(see Fig. A.1).

We show that the first two terms on the right-hand side of (A6.4.19) are asymptotically equivalent to $c_1 G(t)$ and $c_2 G(t)$, respectively, while the last two terms are negligibly small compared with $G(t)$. Indeed, for P_1 we have the obvious two-sided bounds

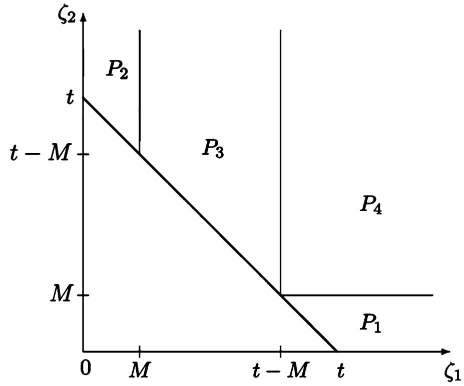

Fig. A.1 Illustration to the proof of Theorem A6.4.3, showing the regions P_i, $i = 1, 2, 3, 4$

$$(1-\varepsilon)^2 c_1 G(t) < G_1(t)\big(1-G_2(M)\big) = \mathbf{P}\big(\zeta_1 \ge t,\ \zeta_2 \in [0,M)\big)$$
$$\le P_1 \le \mathbf{P}(\zeta_1 \ge t-M) = G_1(t-M) \le (1+\varepsilon)^2 c_1 G(t)$$

by (A6.4.18); the term P_2 can be bounded in a similar way. Further,

$$P_4 = \mathbf{P}(\zeta_2 \ge M,\ \zeta_1 \ge t-M) = G_2(M)G_1(t-M) < \varepsilon(1+\varepsilon)^2 c_2 G(t).$$

It remains to estimate P_3 (note that it is here that we will need the condition $\mathbf{G} \in \mathcal{S}$; so far we have only used the fact that $\mathbf{G} \in \mathcal{L}$). We have

$$P_3 = \int_{[M,t-M)} G_2(t-y)\,\mathbf{G}_1(dy) \le (1+\varepsilon)c_2 \int_{[M,t-M)} G(t-y)\,\mathbf{G}_1(dy), \tag{A6.4.20}$$

where it is clear that, by (A6.4.18), the last integral is equal to

$$\begin{aligned}
&\mathbf{P}\big(\zeta + \zeta_1 \ge t,\ \zeta_1 \in [M, t-M)\big)\\
&\quad = \mathbf{P}\big(\zeta \ge t-M,\ \zeta_1 \in [M,t-M)\big) + \mathbf{P}\big(\zeta+\zeta_1 \ge t, \zeta \in [M,t-M)\big)\\
&\quad = G(t-M)\,G_1\big([M,t-M)\big) + \int_{[M,t-M)} G_1(t-y)\,\mathbf{G}(dy)\\
&\quad \le \varepsilon(1+\varepsilon)G(t) + (1+\varepsilon)c_1 \int_{[M,t-M)} G(t-y)\mathbf{G}(dy).
\end{aligned} \tag{A6.4.21}$$

Now note that similarly to the above argument we can easily obtain (setting $G_1 = G_2 = G$) that

$$G^{2*}(t) = (1+\theta_1\varepsilon)2G(t) + \int_{[M,t-M)} G(t-y)\mathbf{G}(dy) + \varepsilon(1+\theta_2\varepsilon)G(t),$$

where $|\theta_i| \le 1$, $i = 1, 2$. Since $G^{2*}(t) \sim 2G(t)$ by virtue of $\mathbf{G} \in \mathcal{S}_+$, this equality means that the integral on the right-hand side is $o(G(t))$. Now (A6.4.21) immediately implies that also $P_3 = o(G(t))$, and hence the required assertion is established for the case $\mathbf{G} \in \mathcal{S}_+$.

To extend the desired result to the case of distributions G_i on $\mathbb{R}$, it suffices to repeat the argument from the proof of Theorem A6.4.1(vi).

The case when one of the c_i can be zero can be reduced to the case $c_1c_2 > 0$, which has already been considered. If, say, $c_1 = 0$ and $c_2 > 0$, then we can introduce the distribution $\widetilde{\mathbf{G}}_1 := (\mathbf{G}_1 + \mathbf{G})/2$, for which clearly $\widetilde{G}_1(t)/G(t) \to \widetilde{c}_1 = 1/2$, and hence by the already proved assertion, as $t \to \infty$,

$$\frac{1}{2} + c_2 \sim \frac{\widetilde{G}_1 * G_2(t)}{G(t)} = \frac{G_1 * G_2(t) + G * G_2(t)}{2G(t)}$$
$$= \frac{G_1 * G_2(t)}{2G(t)} + \big(1 + o(1)\big)\frac{1 + c_2}{2},$$

so that $G_1 * G_2(t)/G(t) \to c_2 = c_1 + c_2$.

(ii) Denote by $\mathbf{G}_0^+$ the distribution of the random variable ζ_0^+, where $\zeta_0 \Subset \mathbf{G}_0$. Since $G_0^+(t) = G_0(t)$ for $t > 0$, it follows immediately from (i) with $\mathbf{G}_1 = \mathbf{G}_2 = \mathbf{G}_0^+$ that $(G_0^+)^{2*}(t) \sim 2G_0^+(t)$, i.e. $\mathbf{G}_0 \in \mathcal{S}$.

(iii) If $\mathbf{G} \in \mathcal{S}$ then by Theorems A6.4.1(vi) and A6.4.2 we have, as $t \to \infty$,

$$G^{2*}(t) \sim \big(G^+\big)^{2*}(t) \sim 2G(t).$$

Now relation (A6.4.15) follows immediately from (i) by induction.

(iv) Similarly to (A6.4.11), we have $G^{n*}(t) \le G_+^{n*}(t)$, $n \ge 1$. Therefore it is clear that it suffices to consider the case $\mathbf{G} \in \mathcal{S}_+$. Put

$$\alpha_n := \sup_{t \ge 0} \frac{G^{n*}(t)}{G(t)}.$$

Similarly to (A6.4.14), for $n \ge 2$, we have

$$G^{n*}(t) = G(t) + \int_0^t G^{(n-1)*}(t - y)\mathbf{G}(dy),$$

and hence, for each $M > 0$,

$$\alpha_n \le 1 + \sup_{0 \le t \le M} \int_0^t \frac{G^{(n-1)*}(t-y)}{G(t)}\mathbf{G}(dy)$$
$$+ \sup_{t > M} \int_0^t \frac{G^{(n-1)*}(t-y)}{G(t-y)} \frac{G(t-y)}{G(t)}\mathbf{G}(dy)$$
$$\le 1 + \frac{1}{G(M)} + \alpha_{n-1} \sup_{t > M} \frac{G^{2*}(t) - G(t)}{G(t)}.$$

Since $\mathbf{G} \in \mathcal{S}$, for any $\varepsilon > 0$ there exists an $M = M(\varepsilon)$ such that

$$\sup_{t > M} \frac{G^{2*}(t) - G(t)}{G(t)} < 1 + \varepsilon$$

and hence

$$\alpha_n \le b_0 + \alpha_{n-1}(1+\varepsilon), \quad b_0 := 1 + 1/G(M), \quad \alpha_1 = 1.$$

This recurrently implies

$$\alpha_n \le b_0 + b_0(1+\varepsilon) + \alpha_{n-2}(1+\varepsilon)^2 \le \cdots \le b_0 \sum_{j=0}^{n-1} (1+\varepsilon)^j \le \frac{b_0}{\varepsilon}\,(1+\varepsilon)^n.$$

The theorem is proved. □

Appendix 7
The Proofs of Theorems on Convergence to Stable Laws

In this appendix we will prove Theorems 8.8.1–8.8.4.

7.1 The Integral Limit Theorem

In this section we will prove Theorem 8.8.1 on convergence of the distributions of normalised sums $S_n = \sum_{k=1}^{n} \xi_k$ to stable laws. Recall the basic notation:

$$F_+(t) := \mathbf{P}(\xi \ge t), \qquad F_-(t) := \mathbf{P}(\xi < -t),$$
$$F_0(t) := F_+(t) + F_-(t) = \mathbf{P}\big(\xi \notin [-t, t)\big).$$

The main condition used in the theorem has this form:

$[\mathbf{R}_{\beta,\rho}]$ *The total tail* $F_0(x) = F_-(x) + F_+(x)$ *is a r.v.f. as* $x \to \infty$, *i.e., can be represented as*

$$F_0(x) = t^{-\beta} L_{F_0}(x), \quad \beta \in (0, 2], \tag{A7.1.1}$$

where $L_{F_0}(x)$ *is an s.v.f., and there exists the limit*

$$\rho_+ := \lim_{x\to\infty} \frac{F_+(x)}{F_0(x)} \in [0, 1], \qquad \rho := 2\rho_+ - 1. \tag{A7.1.2}$$

In the case $\beta < 2$ we put

$$b(n) := F_0^{(-1)}(1/n), \tag{A7.1.3}$$

while for $\beta = 2$ we set

$$b(n) := Y^{(-1)}(1/n), \tag{A7.1.4}$$

where

$$Y(t) := 2t^{-2} \int_0^t y F_0(y)\, dy = t^{-2}\mathbf{E}\big(\xi^2; -t \le \xi < t\big) = t^{-2} L_Y(t), \tag{A7.1.5}$$

DOI 10.1007/978-1-4471-5201-9,

$L_Y(t)$ is an s.v.f., so that (see Theorem A6.2.1(v) of Appendix 6)

$$b(n) = n^{1/\alpha} L_b(n), \quad L_b \text{ is an s.v.f.}$$

In the case when $F_+(t)$ and $F_-(t)$ are regularly varying functions (for instance, when condition $[\mathbf{R}_{\beta,\rho}]$ is satisfied and $\rho = 0$), we will denote these functions by $V(t)$ and $W(t)$, respectively, and put

$$V_I(t) := \int_0^t V(y)\,dy, \qquad V^I(t) := \int_t^\infty V(y)\,dy;$$

the same notational convention will be used for W.

If $F_+(t) = o(F_0(t))$ as $t \to \infty$ ($\rho = -1$), then $F_+(t)$ is not necessarily a regularly varying function, but everything we say below regarding the sums $V(t) + W(t)$ and $V^I(t) + W^I(t)$ remains valid if we understand by their first summands quantities negligibly small compared to the second summands (the first summands can also be replaced by zeros). This is also true for the sums $V_I(t) + W_I(t)$, except for the case when $\mathbf{E}\max(0, \xi)$ exists and $V_I(t)$ has to be replaced by $\mathbf{E}(\xi;\ \xi \ge 0) + o(1)$.

Theorem A7.1.1 *Let condition $[\mathbf{R}_{\beta,\rho}]$ be satisfied and $\zeta_n := \frac{S_n}{b(n)}$.*

(i) *For $\beta \in (0, 2)$, $\beta \ne 1$, and scaling factor* (A7.1.3), *as $n \to \infty$,*

$$\zeta_n \Rightarrow \zeta^{(\beta,\rho)}, \tag{A7.1.6}$$

where the distribution $\mathbf{F}_{\beta,\rho}$ of the random variable $\zeta^{(\beta,\rho)}$ depends on the parameters β and ρ only and has ch.f.

$$\varphi^{(\beta,\rho)}(t) := \mathbf{E}e^{it\zeta^{(\beta,\rho)}} = \exp\big\{|t|^\beta B(\beta, \rho, \vartheta)\big\}, \tag{A7.1.7}$$

where $\vartheta := \operatorname{sign} t$,

$$B(\beta, \rho, \vartheta) := \Gamma(1-\beta)\left[i\rho\vartheta \sin\frac{\beta\pi}{2} - \cos\frac{\beta\pi}{2}\right] \tag{A7.1.8}$$

and, for $\beta \in (1, 2)$, we assume that $\Gamma(1-\beta) = \Gamma(2-\beta)/(1-\beta)$.

(ii) *When $\beta = 1$, for the sequence ζ_n with scaling factor* (A7.1.3) *to converge to a limiting law the former, generally speaking, needs to be centred. More precisely, we have, as $n \to \infty$,*

$$\zeta_n - A_n \Rightarrow \zeta^{(1,\rho)}, \tag{A7.1.9}$$

where

$$A_n := \frac{n}{b(n)}\big[V_I\big(b(n)\big) - W_I\big(b(n)\big)\big] - \rho C, \tag{A7.1.10}$$

$C \approx 0.5772$ is the Euler constant, and

$$\varphi^{(1,\rho)}(t) := \mathbf{E}e^{it\zeta^{(1,\rho)}} = \exp\left\{-\frac{\pi|t|}{2} - i\rho t \ln|t|\right\}. \tag{A7.1.11}$$

If $n[V_I(b(n)) - W_I(b(n))] = o(b(n))$, *then* $\rho = 0$ *and we can put* $A_n = 0$.
If there exists $\mathbf{E}\xi = 0$ *then*

$$A_n = \frac{n}{b(n)}\big[W^I\big(b(n)\big) - V^I\big(b(n)\big)\big] - \rho C.$$

If $\mathbf{E}\xi = 0$ *and* $\rho \neq 0$ *then* $\rho A_n \to -\infty$ *as* $n \to \infty$.
(iii) *For* $\beta = 2$ *and scaling factor* (A7.1.4), *as* $n \to \infty$,

$$\zeta_n \Rightarrow \zeta^{(2,\rho)}, \qquad \varphi^{(2,\rho)}(t) := \mathbf{E}e^{it\zeta} = e^{-t^2/2},$$

so that $\zeta^{(2,\rho)}$ *has the standard normal distribution which does not depend on* ρ.

Proof We will use the same approach as in the proof of the central limit theorem using relation (8.8.1). We will study the asymptotic properties of the ch.f. $\varphi(t) = \mathbf{E}e^{it\xi}$ in the vicinity of zero (more precisely, the asymptotics of

$$\varphi\left(\frac{\mu}{b(n)}\right) - 1 \to 0$$

as $b(n) \to \infty$) and show that, under condition $[\mathbf{R}_{\beta,\rho}]$, for each $\mu \in \mathbb{R}$, we have

$$n\left(\varphi\left(\frac{\mu}{b(n)}\right) - 1\right) \to \ln\varphi^{(\beta,\rho)}(\mu) \quad \text{as } n \to \infty \tag{A7.1.12}$$

(or some modification of this relation, see (A7.1.48)). This will imply that, for $\zeta_n = S(n)/b(n)$, as $n \to \infty$, there holds the relation (cf. Lemma 8.3.2)

$$\varphi_{\zeta_n}(\mu) \to \varphi^{(\beta,\rho)}(\mu). \tag{A7.1.13}$$

Indeed,

$$\varphi_{\zeta_n}(\mu) = \varphi^n\left(\frac{\mu}{b(n)}\right).$$

Since $\varphi(t) \to 1$ as $t \to 0$, one has

$$\begin{aligned}\ln\varphi_{\zeta_n}(\mu) &= n\ln\varphi\left(\frac{\mu}{b(n)}\right)\\ &= n\ln\left[1 + \left(\varphi\left(\frac{\mu}{b(n)}\right) - 1\right)\right] = n\left[\varphi\left(\frac{\mu}{b(n)}\right) - 1\right] + R_n,\end{aligned}$$

where $|R_n| \le n|\varphi(\mu/b(n)) - 1|^2$ for all n large enough, and hence $R_n \to 0$ by virtue of (A7.1.12). It follows that (A7.1.12) implies (A7.1.13).

So first we will study the asymptotics of $\varphi(t)$ as $t \to 0$ and then establish (A7.1.12).

(i) *First let* $\beta \in (0, 1)$. We have

$$\varphi(t) = -\int_0^\infty e^{itx}\, dV(x) - \int_0^\infty e^{-itx}\, dW(x). \tag{A7.1.14}$$

Consider the former integral

$$-\int_0^\infty e^{itx}\, dV(x) = V(0) + it\int_0^\infty e^{itx} V(x)\, dx, \tag{A7.1.15}$$

where the substitution $|t|x = y$, $|t| = 1/m$ yields

$$I_+(t) := it\int_0^\infty e^{itx} V(x)\, dx = i\vartheta \int_0^\infty e^{i\vartheta y} V(my)\, dy, \tag{A7.1.16}$$

$\vartheta = \operatorname{sign} t$ (we will henceforth exclude the trivial case $t = 0$).

Assume for the present that $\rho_+ > 0$. Then $V(x)$ is an r.v.f. as $x \to \infty$ and, for each y, by virtue of the properties of s.v.f.s we have, as $|m| \to 0$,

$$V(my) \sim y^{-\beta} V(m).$$

Therefore it is natural to expect that, as $|t| \to 0$,

$$I_+(t) \sim i\vartheta V(m) \int_0^\infty e^{i\vartheta y} y^{-\beta}\, dy = i\vartheta V(m) A(\beta, \vartheta), \tag{A7.1.17}$$

where

$$A(\beta, \vartheta) := \int_0^\infty e^{i\vartheta y} y^{-\beta}\, dy. \tag{A7.1.18}$$

Assume that relation (A7.1.17) holds and similarly (in the case when $\rho_- > 0$)

$$-\int_0^\infty e^{-itx}\, dW(x) = W(0) + I_-(t), \tag{A7.1.19}$$

where

$$\begin{aligned} I_-(t) &:= -it\int_0^\infty e^{-itx} W(x)\, dx \sim -i\vartheta W(m) \int_0^\infty e^{-i\vartheta y} y^{-\beta}\, dy \\ &= -i\vartheta W(m) A(\beta, -\vartheta). \end{aligned} \tag{A7.1.20}$$

Since $V(0) + W(0) = 1$, relations (A7.1.14)–(A7.1.20) mean that, as $t \to 0$,

$$\varphi(t) = 1 + F_0(m) i\vartheta \left[\rho_+ A(\beta, \vartheta) - \rho_- A(\beta, -\vartheta)\right]\big(1 + o(1)\big). \tag{A7.1.21}$$

We can find an explicit form of the integral $A(\beta, \vartheta)$. Observe that the integral along the boundary of the positive quadrant (closed as a contour) in the complex plane of the function $e^{iz}z^{-\beta}$, which, as $|t| \to 0$, is equal to zero. From this it is not hard to obtain that

$$A(\beta, \vartheta) = \Gamma(1-\beta)e^{i\vartheta(1-\beta)\pi/2}, \quad \beta > 0. \tag{A7.1.22}$$

(Note also that (A7.1.18) is a table integral and its value can be found in handbooks, see, e.g., integrals 3.761.4 and 3.761.9 in [18].)

Thus, in (A7.1.21) one has

$$\begin{aligned} i\vartheta\left[\rho_+ A(\beta,\vartheta) - \rho_- A(\beta,-\vartheta)\right] &= i\vartheta\, \Gamma(1-\beta)\biggl[\rho_+ \cos\frac{(1-\beta)\pi}{2} \\ &\quad + i\vartheta\rho_+ \sin\frac{(1-\beta)\pi}{2} - \rho_- \cos\frac{(1-\beta)\pi}{2} + i\vartheta\rho_- \sin\frac{(1-\beta)\pi}{2}\biggr] \\ &= \Gamma(1-\beta)\biggl[i\vartheta(\rho_+ - \rho_-)\cos\frac{(1-\beta)\pi}{2} - \sin\frac{(1-\beta)\pi}{2}\biggr] \\ &= \Gamma(1-\beta)\biggl[i\vartheta\rho \sin\frac{\beta\pi}{2} - \cos\frac{\beta\pi}{2}\biggr] = B(\beta,\rho,\vartheta), \end{aligned}$$

where $B(\beta, \rho, \vartheta)$ is defined in (A7.1.8). Hence, as $t \to 0$,

$$\varphi(t) - 1 = F_0(m)B(\beta,\rho,\vartheta)\bigl(1+o(1)\bigr). \tag{A7.1.23}$$

Putting $t = \mu/b(n)$ (so that $m = b(n)/|\mu|$), where $b(n)$ is defined in (A7.1.3), and taking into account that $F_0(b(n)) \sim 1/n$, we obtain

$$n\left[\varphi\left(\frac{\mu}{b(n)}\right) - 1\right] = nF_0\left(\frac{b(n)}{|\mu|}\right)B(\beta,\rho,\vartheta)\bigl(1+o(1)\bigr) \sim |\mu|^\beta B(\beta,\rho,\vartheta). \tag{A7.1.24}$$

We have established the validity of (A7.1.12) and therefore that of assertion (i) of the theorem in the case $\beta < 1$, $\rho_+ > 0$.

If $\rho_+ = 0$ ($\rho_- = 0$) then, as was already mentioned, the above argument remains valid if we replace $V(m)$ ($W(m)$) by zero. This follows from the fact that in this case $F_+(t)$ ($F_-(t)$) admits a regularly varying majorant $V^*(t) = o(W(t))$ ($W^*(t) = o(V(t))$).

It remains only to justify the asymptotic equivalence in (A7.1.17). To do that, it is sufficient to verify that the integrals

$$\int_0^\varepsilon e^{i\vartheta y}V(my)\,dy, \qquad \int_M^\infty e^{i\vartheta y}V(my)\,dy \tag{A7.1.25}$$

can be made arbitrarily small compared to $V(m)$ by choosing appropriate ε and M. Note first that by Theorem A6.2.1(iii) of Appendix 6 (see (A6.1.2) in Appendix 6), for any $\delta > 0$, there exists an $x_\delta > 0$ such that, for all $v \le 1$ and $vx \ge x_\delta$, we have

$$\frac{V(vx)}{V(x)} \le (1+\delta)v^{-\beta-\delta}.$$

Therefore, for $\delta < 1-\beta$ and $x > x_\delta$,

$$\int_0^x V(u)\,du \le x_\delta + \int_{x_\delta}^x V(u)\,du = x_\delta + xV(x)\int_{x_\delta/x}^1 \frac{V(vx)}{V(x)}\,dv$$
$$\le x_\delta + xV(x)(1+\delta)\int_0^1 v^{-\beta-\delta}dv$$
$$= x_\delta + \frac{xV(x)(1+\delta)}{1-\beta-\delta} \le cxV(x) \qquad \text{(A7.1.26)}$$

since $xV(x) \to \infty$ as $x \to \infty$. It follows that

$$\left|\int_0^\varepsilon e^{i\vartheta y}V(my)\,dy\right| \le \frac{1}{m}\int_0^{\varepsilon m} V(u)\,du \le c\varepsilon V(\varepsilon m) \sim c\varepsilon^{1-\beta}V(m).$$

Since $\varepsilon^{1-\beta} \to 0$ as $\varepsilon \to 0$, the first assertion in (A7.1.25) is proved. The second integral in (A7.1.25) is equal to

$$\int_M^\infty e^{i\vartheta y}V(my)\,dy = \frac{1}{i\vartheta}e^{i\vartheta y}V(my)\Big|_M^\infty - \frac{1}{i\vartheta}\int_M^\infty e^{i\vartheta y}dV(my)$$
$$= -\frac{1}{i\vartheta}e^{i\vartheta M}V(mM) - \frac{1}{i\vartheta}\int_{mM}^\infty e^{i\vartheta u/m}dV(u),$$

so its absolute value does not exceed

$$2V(mM) \sim 2M^{-\beta}V(m) \qquad \text{(A7.1.27)}$$

as $m \to \infty$. Hence the value of the second integral in (A7.1.25) can also be made arbitrarily small compared to $V(m)$ by choosing an appropriate M. Relation (A7.1.17) together with the assertion of the theorem in the case $\beta < 1$ are proved.

Let now $\beta \in (1,2)$ and hence there exist a finite expectation $\mathbf{E}\xi$ which, according to our condition, will be assumed to be equal to zero. In this case,

$$\varphi(t) - 1 = \vartheta\int_0^{|t|}\varphi'(\vartheta u)\,du, \quad \vartheta = \operatorname{sign} t, \qquad \text{(A7.1.28)}$$

and we have to find the asymptotic behaviour of

$$\varphi'(t) = -i\int_0^\infty xe^{itx}dV(x) + i\int_0^\infty xe^{-itx}dW(x) =: I_+^{(1)}(t) + I_-^{(1)}(t) \quad \text{(A7.1.29)}$$

as $t \to 0$. Since $x\,dV(x) = d(xV(x)) - V(x)\,dx$, integration by parts yields

$$I_+^{(1)}(t) := -i\int_0^\infty xe^{itx}dV(x) = -i\int_0^\infty e^{itx}d\big(xV(x)\big) + i\int_0^\infty e^{itx}V(x)\,dx$$
$$= -t\int_0^\infty xV(x)e^{itx}\,dx + iV^I(0) - t\int_0^\infty V^I(x)e^{itx}\,dx$$

$$= iV^I(0) - t\int_0^\infty \widetilde{V}(x)e^{itx}\,dx, \qquad \text{(A7.1.30)}$$

where, by Theorem A6.2.1(iv) of Appendix 6, both functions

$$V^I(x) := \int_x^\infty V(u)\,du \sim \frac{xV(x)}{\beta-1} \quad \text{as } x\to\infty, \quad V^I(0)<\infty,$$

and

$$\widetilde{V}(x) := xV(x) + V^I(x) \sim \frac{\beta x V(x)}{\beta-1}$$

are regularly varying.

Letting, as before, $m = 1/|t|$, $m\to\infty$ (cf. (A7.1.16), (A7.1.17)), we get

$$-t\int_0^\infty \widetilde{V}(x)e^{itx}\,dx = -\vartheta\widetilde{V}(m)\int_0^\infty \widetilde{V}(my)e^{i\vartheta y}\,dy$$
$$\sim -\vartheta\int_0^\infty y^{-\beta+1}e^{i\vartheta y}\,dy = -\frac{\beta V(m)}{t(\beta-1)}A(\beta-1,\vartheta),$$

$$I_+^{(1)}(t) = iV^I(0) - \frac{\beta\rho_+F_0(m)}{t(\beta-1)}A(\beta-1,\vartheta)\big(1+o(1)\big), \qquad \text{(A7.1.31)}$$

where the function $A(\beta,\vartheta)$ defined in (A7.1.18) is equal to (A7.1.22).

Similarly,

$$I_-^{(1)}(t) := i\int_0^\infty te^{-itx}\,dW(x)$$
$$= -t\int_0^\infty xW(x)e^{-itx}\,dx - iW^I(0) - t\int_0^\infty W^I(x)e^{-itx}\,dx$$
$$= -iW^I(0) - t\int_0^\infty \widetilde{W}(x)e^{-itx}\,dx,$$

where

$$W^I(x) := \int_x^\infty W(u)\,du, \qquad \widetilde{W}(x) := xW(x) + W^I(x) \sim \frac{\beta xW(x)}{\beta-1},$$

and

$$-t\int_0^\infty \widetilde{W}(x)\,e^{-itx}\,dx \sim -\frac{\beta W(m)}{t(\beta-1)}A(\beta-1,-\vartheta).$$

Therefore

$$I_-^{(1)}(t) = iW^I(0) - \frac{\beta\rho_-F_0(m)}{t(\beta-1)}A(\beta-1,-\vartheta)\big(1+o(1)\big)$$

and hence, by virtue of (A7.1.29), (A7.1.31), and the equality $V^I(0) - W^I(0) = \mathbf{E}\xi = 0$, we have

$$\varphi'(t) = -\frac{\beta F_0(m)}{t(\beta-1)}\big[\rho_+ A(\beta-1,\vartheta) + \rho_- A(\beta-1,-\vartheta)\big]\big(1+o(1)\big).$$

We return now to relation (A7.1.28). Since

$$\int_0^{|t|} u^{-1} F_0\big(u^{-1}\big)\,du \sim \beta^{-1} F_0\big(|t|^{-1}\big) = \beta^{-1} F_0(m)$$

(see Theorem A6.2.1(iii) of Appendix 6), we obtain, again using (A7.1.22) and an argument similar to the one in the proof for the case $\beta < 1$, that

$$\begin{aligned}
\varphi(t) - 1 &= -\frac{1}{\beta-1} F_0(m)\big[\rho_+ A(\beta-1,\vartheta) + \rho_- A(\beta-1,-\vartheta)\big]\big(1+o(1)\big) \\
&= -\frac{\Gamma(2-\beta)}{\beta-1} F_0(m)\bigg[\rho_+\bigg(\cos\frac{(2-\beta)\pi}{2} + i\vartheta\sin\frac{(2-\beta)\pi}{2}\bigg) \\
&\quad + \rho_-\bigg(\cos\frac{(2-\beta)\pi}{2} - i\vartheta\sin\frac{(2-\beta)\pi}{2}\bigg)\bigg]\big(1+o(1)\big) \\
&= \frac{\Gamma(2-\beta)}{\beta-1} F_0(m)\bigg[\cos\frac{\beta\pi}{2} - i\vartheta\rho\sin\frac{\beta\pi}{2}\bigg]\big(1+o(1)\big) \\
&= F_0(m) B(\beta,\rho,\vartheta)\big(1+o(1)\big). \qquad \text{(A7.1.32)}
\end{aligned}$$

We arrive once again at relation (A7.1.23) which, by virtue of (A7.1.24), implies the assertion of the theorem for $\beta \in (1,2)$.

(ii) *Case* $\beta = 1$. In this case, the computation is somewhat more complicated. We again follow relations (A7.1.14)–(A7.1.16), according to which

$$\varphi(t) = 1 + I_+(t) + I_-(t). \qquad \text{(A7.1.33)}$$

Rewrite expression (A7.1.16) for $I_+(t)$ as

$$I_+(x) = i\vartheta\int_0^\infty e^{i\vartheta y} V(my)\,dy = i\vartheta\int_0^\infty V(my)\cos y\,dy - \int_0^\infty V(my)\sin y\,dy, \qquad \text{(A7.1.34)}$$

where the first integral on the right-hand side can be represented as the sum of two integrals:

$$\int_0^1 V(my)\,dy + \int_0^\infty g(y) V(my)\,dy, \qquad \text{(A7.1.35)}$$

$$g(y) = \begin{cases} \cos y - 1 & \text{if } y \le 1, \\ \cos y & \text{if } y > 1. \end{cases} \qquad \text{(A7.1.36)}$$

Note that (see, e.g., integral 3.782 in [18]) the value of the integral

$$-\int_0^\infty g(y)y^{-1}\,dy = C \approx 0.5772 \tag{A7.1.37}$$

is the Euler constant. Since $V(ym)/V(m) \to y^{-1}$ as $m \to \infty$, similarly to the above argument we obtain for the second integral in (A7.1.35) the relation

$$\int_0^\infty g(y)V(my)\,dy \sim -CV(m). \tag{A7.1.38}$$

Consider now the first integral in (A7.1.35):

$$\int_0^1 V(my)\,dy = m^{-1}\int_0^m V(u)\,du = m^{-1}V_I(m), \tag{A7.1.39}$$

where

$$V_I(x) := \int_0^t V(u)\,du \tag{A7.1.40}$$

can easily be seen to be an s.v.f. in the case $\beta = 1$ (see Theorem A6.2.1(iv) of Appendix 6). Here if $\mathbf{E}|\xi| = \infty$ then $V_I(x) \to \infty$ as $x \to \infty$, and if $\mathbf{E}|\xi| < \infty$ then $V_I(x) \to V_I(\infty) < \infty$.

Thus, for the first term on the right-hand side of (A7.1.34) we have

$$\operatorname{Im}\, I_+(t) = \vartheta\big(-CV(m) + m^{-1}V_I(m)\big) + o\big(V(m)\big). \tag{A7.1.41}$$

Now we will determine how $V_I(vx)$ depends on v as $x \to \infty$. For any fixed $v > 0$,

$$V_I(vx) = V_I(x) + \int_x^{vx} V(u)\,du = V_I(x) + xV(x)\int_1^v \frac{V(yx)}{V(x)}\,dy.$$

By Theorem A6.2.1 of Appendix 6,

$$\int_1^v \frac{V(yx)}{V(x)}\,dy \sim \int_1^v \frac{dy}{y} = \ln v,$$

so that

$$V_I(vx) = V_I(x) + \big(1+o(1)\big)xV(x)\ln v =: A_V(v,x) + xV(x)\ln v, \tag{A7.1.42}$$

where evidently

$$A_V(v,x) = V_I(x) + o\big(xV(x)\big) \quad \text{as } x \to \infty \tag{A7.1.43}$$

and $V_I(x) \gg xV(x)$ by Theorem A6.2.1(iv) of Appendix 6.

Therefore, for $t = \mu/b(n)$ (so that $m = b(n)/|\mu|$ and hence $V(m) \sim \rho_+|\mu|/n$), we obtain from (A7.1.41) and (A7.1.42) (where one has to put $x = b(n)$, $v = 1/|\mu|$) that the following representation is valid as $n \to \infty$:

$$\begin{aligned}\operatorname{Im} I_+(t) &= -C\frac{\rho_+\mu}{n} + \frac{\mu}{b(n)}\left[A_V\left(|\mu|^{-1}, b(n)\right) - \frac{\rho_+\mu}{n}\ln|\mu|\right] + o\left(n^{-1}\right) \\ &= \frac{\mu}{b(n)}A_V\left(|\mu|^{-1}, b(n)\right) - \frac{\rho_+\mu}{n}\left(C + \ln|\mu|\right) + o\left(n^{-1}\right). \qquad \text{(A7.1.44)}\end{aligned}$$

For the second term on the right-hand side of (A7.1.34) we have

$$\operatorname{Re} I_+(t) = -\int_0^\infty V(my)\sin y\,dy \sim -V(m)\int_0^\infty y^{-1}\sin y\,dy.$$

Because $\sin y \sim y$ as $y \to 0$, the last integral converges. Since $\Gamma(\gamma) \sim 1/\gamma$ as $\gamma \to 0$, the value of this integral can be found to be (see (A7.1.22) and (A7.1.22))

$$\lim_{\gamma\to 0}\Gamma(\gamma)\sin\frac{\gamma\pi}{2} = \frac{\pi}{2}. \qquad \text{(A7.1.45)}$$

Thus, for $t = \mu/b(n)$,

$$\operatorname{Re} I_+(t) = -\frac{\pi|\mu|}{2n} + o\left(n^{-1}\right). \qquad \text{(A7.1.46)}$$

In a similar way we can find an asymptotic representation for the integral $I_-(t)$ (see (A7.1.14)–(A7.1.20)):

$$\begin{aligned}I_-(t) &:= -i\vartheta\int_0^\infty W(my)e^{-i\vartheta y}\,dy \\ &= -i\vartheta\int_0^\infty W(my)\cos y\,dy - \int_0^\infty W(my)\sin y\,dy.\end{aligned}$$

Comparing this with (A7.1.34) and the subsequent computation of $I_+(t)$, we can immediately conclude that, for $t = \mu/b(n)$ (cf. (A7.1.44), (A7.1.46)),

$$\begin{aligned}\operatorname{Im} I_-(t) &= -\frac{-\mu A_W(|\mu|^{-1}, b(n))}{b(n)} + \frac{\rho_-\mu}{n}\left(C + \ln|\mu|\right) + o\left(n^{-1}\right), \\ \operatorname{Re} I_-(t) &= -\frac{\pi|\mu|\rho_-}{2n} + o\left(n^{-1}\right).\end{aligned} \qquad \text{(A7.1.47)}$$

Thus we obtain from (A7.1.33), (A7.1.44) and (A7.1.46) that (A7.1.47) imply

$$\begin{aligned}\varphi\left(\frac{\mu}{b(n)}\right) - 1 &= -\frac{\pi|\mu|}{n} - \frac{i\rho\mu}{n}\left(C + \ln|\mu|\right) \\ &\quad + \frac{i\mu}{b(n)}\left[A_V\left(|\mu|^{-1}, b(n)\right) - A_W\left(|\mu|^{-1}, b(n)\right)\right] + o\left(n^{-1}\right).\end{aligned}$$

It follows from (A7.1.43) that the penultimate term here is equal to

$$\frac{i\mu}{b(n)}\big[V_I\big(b(n)\big) - W_I\big(b(n)\big)\big] + o\big(n^{-1}\big),$$

so that finally,

$$\varphi\left(\frac{\mu}{b(n)}\right) - 1 = -\frac{\pi|\mu|}{2n} - \frac{i\rho\mu}{n}\ln|\mu| + i\mu\frac{A_n}{n} + o\big(n^{-1}\big), \tag{A7.1.48}$$

where

$$A_n = \frac{n}{b(n)}\big[V_I\big(b(n)\big) - W_I\big(b(n)\big)\big] - \rho C.$$

Therefore, similarly to (A7.1.12) and (A7.1.13), we obtain

$$\varphi_{\zeta_n - A_n}(\mu) = e^{-i\mu A_n}\varphi^n\left(\frac{\mu}{b(n)}\right) = \exp\left\{-i\mu A_n + n\ln\left[1 + \left(\varphi\left(\frac{\mu}{b(n)}\right) - 1\right)\right]\right\}$$
$$= \exp\left\{-i\mu A_n + n\left(\varphi\left(\frac{\mu}{b(n)}\right) - 1\right) + nO\left(\left|\varphi\left(\frac{\mu}{b(n)}\right) - 1\right|^2\right)\right\}.$$

As, for $\beta = 1$, by Theorem A6.2.1(iv) of Appendix 6, the functions V_I and W_I are slowly varying, by (A7.1.48) one has

$$n\left|\varphi\left(\frac{\mu}{b(n)}\right) - 1\right|^2 \le c\left(\frac{1}{n} + \frac{A_n^2}{n}\right) \le c_1\left(\frac{1}{n} + \frac{1}{b(n)}\big[V_I\big(b(n)\big)^2 + W_I\big(b(n)\big)^2\big]\right) \to 0.$$

Since clearly

$$-i\mu A_n + n\left(\varphi\left(\frac{\mu}{b(n)}\right) - 1\right) \to -\frac{\pi|\mu|}{2} - i\rho\mu\ln|\mu|,$$

we have

$$\varphi_{\zeta_n - A_n}(\mu) \to \exp\left\{-\frac{\pi|\mu|}{2} - i\rho\mu\ln|\mu|\right\},$$

so relation (A7.1.9) is proved. The subsequent assertions regarding the centring sequence $\{A_n\}$ are evident. □

(iii) It remains to consider the case $\beta = 2$. We will follow representations (A7.1.28)–(A7.1.30), according to which we have to find, as $m = 1/|t| \to \infty$, the asymptotics of

$$\varphi'(t) = I_+^{(1)}(t) + I_-^{(1)}(t), \tag{A7.1.49}$$

where

$$I_+^{(1)}(t) := iV^I(0) - t\int_0^\infty \widetilde{V}(x)e^{itx}\,dx = iV^I(0) - \vartheta\int_0^\infty \widetilde{V}(my)e^{i\vartheta y}\,dy \tag{A7.1.50}$$

and, by Theorem A6.2.1(iv) of Appendix 6,

$$V^I(x) = \int_x^\infty V(y)\,dy \sim xV(x), \quad \widetilde{V}(x) = xV(x) + V^I(x) \sim 2xV(x) \quad \text{(A7.1.51)}$$

as $x \to \infty$. Further,

$$\int_0^\infty \widetilde{V}(my)\,e^{i\vartheta y}dy = \int_0^\infty \widetilde{V}(my)\cos y\,dy + \vartheta \int_0^\infty \widetilde{V}(my)\sin y\,dy. \quad \text{(A7.1.52)}$$

Here the second integral on the right-hand side is asymptotically equivalent, as $m \to \infty$, to (see (A7.1.45))

$$\widetilde{V}(m)\int_0^\infty y^{-1}\sin y\,dy = \frac{\pi}{2}\,\widetilde{V}(m).$$

The first integral on the right-hand side of (A7.1.52) is equal to

$$\int_0^1 \widetilde{V}(my)\,dy + \int_0^\infty g(y)\widetilde{V}(my)\,dy,$$

where the function $g(y)$ was defined in (A7.1.35), and

$$\int_0^1 \widetilde{V}(my)\,dy = \frac{1}{m}\int_0^m \widetilde{V}(u)\,du = \frac{1}{m}\,\widetilde{V}_I(m),$$

$\widetilde{V}_I(x) := \int_0^x \widetilde{V}(u)\,du$ being an s.v.f. by (A7.1.51). Since

$$\int_0^x uV(u)\,du = \frac{x^2V(x)}{2} - \frac{1}{2}\int_0^x u^2 dV(u),$$
$$\int_0^x V^I(u)\,du = xV^I(x) + \int_0^x uV(u)\,du$$

and $V^I(x) \sim xV(x)$, we have

$$\begin{aligned}\widetilde{V}_I(x) &= \int_0^x \big(uV(u) + V^I(u)\big)\,du \\ &= xV^I(x) + x^2V(x) - \int_0^x u^2\,dV(u) \\ &= -\int_0^x u^2\,dV(y) + O\big(x^2V(x)\big), \end{aligned} \quad \text{(A7.1.53)}$$

where the last term is negligibly small, because

$$\int_0^x uV(u)\,du \gg x^2V(x)$$

(see Theorem A6.2.1(iv) of Appendix 6).

It is also clear that, as $x \to \infty$,

$$\widetilde{V}_I(x) \to \widetilde{V}_I(\infty) = \mathbf{E}\big(\xi^2; \xi > 0\big) \in (0, \infty].$$

As a result, we obtain (see also (A7.1.38))

$$\begin{aligned} I_+^{(1)}(t) &= iV^I(0) - \frac{i\pi}{2}\widetilde{V}(m) - t\widetilde{V}_I(m) + \vartheta C\widetilde{V}(m) + o\big(\widetilde{V}(m)\big) \\ &= iV^I(0) - t\widetilde{V}_I(m)\big(1 + o(1)\big) \end{aligned}$$

since $\widetilde{V}_I(x) \gg t\widetilde{V}(x)$.

Quite similarly we get

$$I_-^{(1)}(t) = -iW^I(0) - t\widetilde{W}_I(m)\big(1 + o(1)\big),$$

where $\widetilde{W}_I$ is an s.v.f. which is obtained from the function W in the same way as $\widetilde{V}_I$ from V. Since $V^I(0) = W^I(0)$, relation (A7.1.49) now yields that

$$\varphi'(t) = -t\big[\widetilde{V}_I(m) + \widetilde{W}_I(m)\big]\big(1 + o(1)\big).$$

Hence from (A7.1.28) we obtain the representation

$$\begin{aligned} \varphi(t) - 1 &= \vartheta \int_0^{1/m} \varphi'(\vartheta u)\,du = -\int_0^{1/m} u\big[\widetilde{V}_I(1/u) + \widetilde{W}_I(1/u)\big]\,du \\ &\sim -\frac{1}{2m^2}\big[\widetilde{V}_I(m) + \widetilde{W}_I(m)\big] \sim -\frac{1}{2m^2}\mathbf{E}\big(\xi^2;\ -m \le \xi < m\big) \end{aligned}$$

by virtue of (A7.1.53) and a similar relation for $\widetilde{W}_I$. Turning now to the definition of the function $Y(x) = x^{-2}L_Y(x)$ in (A7.1.5) and putting

$$b(n) := Y^{(-1)}(1/n), \qquad t = \mu/b(n),$$

we get

$$n\big(\varphi(t) - 1\big) \sim -\frac{n}{2}Y\big(b(n)/|\mu|\big) \sim -\frac{n\mu^2}{2}Y\big(b(n)\big) \to -\frac{\mu^2}{2}.$$

The theorem is proved. □

7.2 The Integro-Local and Local Limit Theorems

In this section we will prove Theorems 8.8.2–8.8.4. We will begin with the integro-local theorem.

Theorem A7.2.1 (Integro-local Stone's theorem) *Let ξ be a non-lattice random variable and the conditions of Theorem A7.1.1 be satisfied. Then, for each fixed $\Delta > 0$,*

$$\mathbf{P}\big(S_n \in \Delta[x)\big) = \frac{\Delta}{b(n)} f^{(\beta,\rho)}\left(\frac{x}{b(n)}\right) + o\left(\frac{1}{b(n)}\right) \quad \text{as } n \to \infty,$$

where the remainder term $o(\frac{1}{b(n)})$ is uniform in x.

Proof of Theorem A7.2.1 The Proof is analogous to the proof of Theorem 8.7.1. We will again use the smoothing approach and consider, along with the sums S_n, the sums

$$Z_n = S_n + \theta\eta,$$

where $\theta = \text{const}$ and η is chosen so that its ch.f. is equal to 0 outside a finite interval. For instance, we can choose η as in Sect. 8.7.3, i.e., with the ch.f. $\varphi_\eta(t) = \max(0,\ 1 - |t|)$. Then equality (8.7.19) will still be valid with the same decomposition of the integral on its right-hand side into the subintegral I_1 over the domain $|t| < \gamma$ and I_2 over the domain $\gamma \le |t| \le 1/\theta$. Here estimating I_2 can be done in the same way as in Theorem 8.7.1.

For the sake of brevity, put $\widehat{\varphi}(t) := \varphi_{\eta_\Delta}(t)\varphi_{\theta\eta}(t)$. Then, for the integral I_1 with $x = vb(n)$, we have

$$I_1 = \int_{|t|<\gamma} e^{-itx}\varphi^n(t)\widehat{\varphi}(t)\,dt = \frac{1}{b(n)}\int_{|u|<\gamma b(n)} e^{-iuv}\varphi^n\left(\frac{u}{b(n)}\right)\widehat{\varphi}\left(\frac{u}{b(n)}\right)du. \tag{A7.2.1}$$

As was shown in the proof of Theorem 8.1.1, for each u we have

$$\varphi^n\left(\frac{u}{b(n)}\right) \to \varphi^{(\beta,\rho)}(u) \quad \text{as } n \to \infty,$$

and, moreover, for some $c > 0$ and $\gamma > 0$ small enough, by, virtue of, say, (A7.1.23) and (A7.1.32), we have

$$\mathrm{Re}\big(\varphi(t) - 1\big) \le -c\,F_0\left(\frac{1}{|t|}\right),$$

and, for any $\varepsilon > 0$ and all n large enough,

$$n\,\mathrm{Re}\left(\varphi\left(\frac{u}{b(n)}\right) - 1\right) \le -cnF_0\left(\frac{b(n)}{|u|}\right) \le -c\,|u|^{\beta-\varepsilon}.$$

Here we used the properties of the r.v.f. F_0. Moreover,

$$\widehat{\varphi}\left(\frac{u}{b(n)}\right) \to 1 \quad \text{as } n \to \infty, \qquad \big|\widehat{\varphi}(u)/b(n)\big| \le 1.$$

The above also implies that, for all u such that $|u| < \gamma b(n)$,

$$\left|\varphi^n\left(\frac{u}{b(n)}\right)\right| \le e^{-c\,|u|^{\beta-\varepsilon}}. \tag{A7.2.2}$$

The obtained relations mean that we can use the dominated convergence theorem in (A7.2.1) which implies

$$\lim_{n\to\infty} b(n)I_1 = \int e^{-iuv}\varphi^{(\beta,\rho)}(u)\,du \tag{A7.2.3}$$

uniformly in v, since the right-hand side of (A7.2.1) is uniformly continuous in v. On the right-hand side of (A7.2.3) is the result of the application of the inversion formula (up to the factor $1/2\pi$) to the ch.f. $\varphi^{(\alpha,\rho)}$. This means that

$$\lim_{n\to\infty} b(n)I_1 = 2\pi f^{(\beta,\rho)}(v).$$

We have established that, for $x = vb(n)$, as $n \to \infty$,

$$\mathbf{P}\big(Z_n \in \Delta[x)\big) = \frac{\Delta}{b(n)}\, f^{(\beta,\rho)}\left(\frac{x}{b(n)}\right) + o\left(\frac{1}{b(n)}\right)$$

uniformly in v (and hence in x).

To prove the theorem it remains to use Lemma 8.7.1.

The theorem is proved. □

The proofs of the local Theorems 8.8.3 and 8.8.4 can be obtained by an obvious similar modification of the proofs of Theorems 8.7.2 and 8.7.3 under the conditions of Theorem 8.8.1.

Appendix 8
Upper and Lower Bounds for the Distributions of the Sums and the Maxima of the Sums of Independent Random Variables

Let $\xi, \xi_1, \xi_2, \dots$ be independent identically distributed random variables,

$$S_n = \sum_{i=1}^n \xi_i, \qquad \overline{S}_n = \max_{1 \le k \le n} S_k.$$

The main goal of this appendix is to obtain upper and lower bounds for the probabilities $\mathbf{P}(S_n \ge x)$ and $\mathbf{P}(\overline{S}_n \ge x)$. These bounds were used in Sect. 9.5 to find the asymptotics of the probabilities of large deviations for S_n and $\overline{S}_n$.

8.1 Upper Bounds Under the Cramér Condition

In this section we will assume that the following one-sided Cramér condition is met:

[C] *There exists a* $\lambda > 0$ *such that*

$$\psi(\lambda) = \mathbf{E}e^{\lambda\xi} < \infty. \tag{A8.1.1}$$

The following analogue of the exponential Chebyshev inequality holds true for $\mathbf{P}(S_n \ge x)$.

Theorem A8.1.1 *For all* $n \ge 1$, $x \ge 0$ *and* $\lambda \ge 0$, *we have*

$$\mathbf{P}(\overline{S}_n \ge x) \le e^{-\lambda x} \max\bigl(1, \psi^n(\lambda)\bigr). \tag{A8.1.2}$$

Proof As $\eta(x) := \inf\{k \ge 1 : S_k \ge x\} \le \infty$ is a Markov time, the event $\{\eta(x) = k\}$ is independent of the random variables $S_n - S_k$. Therefore

$$\psi^n(\lambda) = \mathbf{E}\, e^{\lambda S_n} \ge \sum_{k=1}^n \mathbf{E}\left(e^{\lambda S_n}; \eta(x) = k\right) \ge \sum_{k=1}^n \mathbf{E}\left(e^{\lambda(x+S_n-S_k)}; \eta(x) = k\right)$$

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

$$= e^{\lambda x} \sum_{k=1}^{n} \psi^{n-k}(\lambda)\mathbf{P}\big(\eta(x)=k\big) \geq e^{\lambda x} \min\big(1, \psi^n(\lambda)\big)\mathbf{P}(\overline{S}_n \geq x).$$

This immediately implies (A8.1.2). The theorem is proved. □

If $\psi(\lambda) \geq 1$ for $\lambda \geq 0$ (this is always the case if there exists $\mathbf{E}\xi \geq 0$) then the right-hand side of (A8.1.2) is equal to $e^{-\lambda x}\psi^n(\lambda)$, and the equality (A8.1.2) itself can also be obtained as a consequence of the well-known Kolmogorov–Doob inequality for submartingales (see Theorem 15.3.4, where one has to put $X_n := S_n$).

Thus, if $\mathbf{E}\xi \geq 0$ then

$$\mathbf{P}(\overline{S}_n \geq x) \leq e^{-\lambda x + n \ln \psi(\lambda)}.$$

Choosing the best possible value of λ we obtain the following inequality.

Corollary A8.1.1 *If* $\mathbf{E}\xi \geq 0$ *then, for all* $n \geq 1$ *and* $x \geq 0$, *we have*

$$\mathbf{P}(\overline{S}_n \geq x) \leq e^{-n\Lambda(\alpha)},$$

where

$$\alpha := \frac{x}{n}, \qquad \Lambda(\alpha) := \sup_{\lambda}\big(\lambda\alpha - \ln \psi(\lambda)\big).$$

The function $\Lambda(\alpha)$ is the rate function introduced in Sect. 9.1. Its basic properties were stated in that section. In particular, for $\mathbf{E}\xi = 0$ and $\mathbf{E}\xi^2 = \sigma^2 < \infty$, the asymptotic equivalence $\Lambda(\alpha) \sim \frac{\alpha^2}{2\sigma^2}$ as $\alpha \to 0$ takes place, which yields that, for $x = o(n)$,

$$\mathbf{P}(\overline{S}_n \geq x) \leq \exp\left\{-\frac{x^2}{2n\sigma^2}\big(1+o(1)\big)\right\}. \tag{A8.1.3}$$

8.2 Upper Bounds when the Cramér Condition Is Not Met

In this section we will assume that

$$\mathbf{E}\xi = 0, \qquad \mathbf{E}\xi^2 = \sigma^2 < \infty. \tag{A8.2.1}$$

For simplicity's sake, without losing generality, in what follows we will put $\sigma = 1$. The bounds will be obtained for the deviation zone $x > \sqrt{n}$ which is adjacent to the zone of "normal deviations" where

$$\mathbf{P}(S_n \geq x) \sim 1 - \Phi\left(\frac{x}{\sqrt{n}}\right) \tag{A8.2.2}$$

(uniformly in $x \in (0, N_n\sqrt{n})$, where $N_n \to \infty$ slowly enough as $n \to \infty$; see Sect. 8.2). Moreover, it was established in Sect. 19.1 that, in the normal deviations zone,

$$\mathbf{P}(\overline{S}_n \ge x) \sim 2\left(1 - \Phi\left(\frac{x}{\sqrt{n}}\right)\right). \tag{A8.2.3}$$

To derive upper bounds in the zone $x > \sqrt{n}$ when the Cramér condition **[C]** is not met, we will need additional conditions on the behaviour of the right tail $F_+(t) = \mathbf{P}(\xi \ge t)$ of the distribution $\mathbf{F}$.

Namely, we will assume that the following condition is satisfied.

[<] *For the right tail* $F_+(t) = \mathbf{P}(\xi \ge t)$ *there exists a regularly varying* (*see Appendix* 6) *majorant* $V(t)$:

$$F_+(t) \le V(t) := t^{-\beta} L(t) \quad \text{for all } t > 0,$$

where $\beta > 2$ *and* L *is a slowly varying function* (*s.v.f., see Appendix* 6).

By virtue of (A8.2.2) and (A8.2.3), for deviations $x < N_n\sqrt{n}$, $n \to \infty$, it would be natural to expect upper bounds with an exponential right-hand side $e^{-x^2/(2n)}$ (cf. (A8.1.3)). On the other hand, Theorem A6.4.3(iii) of Appendix 6 implies that, for $F_+(t) = V(t) \in \mathfrak{R}$ and any fixed n we have, as $x \to \infty$,

$$\mathbf{P}(S_n \ge x) \sim nV(x). \tag{A8.2.4}$$

This relation clearly holds true if $n \to \infty$ slowly enough (as $x \to \infty$).

The asymptotics (A8.2.2) and (A8.2.4) merge with each other remarkably as follows:

$$\mathbf{P}(S_n \ge x) \sim \left(1 - \Phi\left(\frac{x}{\sqrt{n}}\right)\right) + nV(x) \tag{A8.2.5}$$

as $n \to \infty$ for all $x > \sqrt{n}$ (for more details see, e.g., [8] and the bibliography therein). Relation (A8.2.5) allows us to "guess" the threshold values of $x = b(n)$ for which asymptotics (A8.2.2) changes to asymptotics (A8.2.4). To find such x it suffices to equate the logarithms of the right-hand sides of (A8.2.2) and (A8.2.4):

$$-\frac{x^2}{2n} = \ln nV(x) = \ln n - \beta \ln x + o(\ln x).$$

The main part $b(n)$ of the solution to this equation, as it is not hard to see, has the form

$$b(n) = \sqrt{(\beta - 2)n \ln n}$$

(we exclude the trivial case $n = 1$).

In what follows, we will represent deviations x as $x = sb(n)$. Based on the above, it is natural to expect (and it can be easily verified) that the first term will dominate on the right-hand side of (A8.2.5) if $s < 1$, while the second will dominate if $s > 1$. Accordingly, for small s (but such that $x > \sqrt{n}$), we will have the above-mentioned exponential bounds for $\mathbf{P}(S_n \ge x)$, while for large s there will hold bounds of the form $nV(x)$ (note that $nV(x) \to 0$ for $x > b(n)$ and $\beta > 2$).

The above claim is confirmed by the assertions below. Along with x introduce deviations

$$y = \frac{x}{r},$$

where $r \ge 1$ is fixed, and put

$$B_j := \{\xi_j < y\}, \qquad B := \bigcap_{j=1}^{n} B_j.$$

Theorem A8.2.1 *Let conditions* (A8.2.1) *and* [**<**] *be satisfied.*

(1) *For any fixed* $h > 1$, $s_0 > 0$, $x = sb(n)$, $s \ge s_0$ *and all* $\Pi := nV(x)$ *small enough, we have*

$$P := \mathbf{P}(\overline{S}_n \ge x;\, B) \le e^r \left(\frac{\Pi(y)}{r}\right)^{r-\theta}, \tag{A8.2.6}$$

where

$$\Pi(y) := nV(y), \qquad \theta := \frac{hr^2}{4s^2}\left(1 + b\frac{\ln s}{\ln n}\right), \qquad b := \frac{2\beta}{\beta - 2}.$$

(2) *For any fixed* $h > 1$, $\tau > 0$, *for* $x = sb(n) > \sqrt{n}$, $s^2 < (h-\tau)/2$, *and all* n *large enough, we have*

$$P \le e^{-x^2/(2nh)}. \tag{A8.2.7}$$

Corollary A8.2.1 (a) *If* $s \to \infty$ *then*

$$\mathbf{P}(\overline{S}_n \ge x) \le nV(x)\big(1 + o(1)\big). \tag{A8.2.8}$$

(b) *If* $s^2 \ge s_0^2$ *for some fixed* $s_0 > 1$ *then, for all* $nV(x)$ *small enough,*

$$\mathbf{P}(\overline{S}_n \ge x) \le cnV(x), \qquad c = \text{const}. \tag{A8.2.9}$$

(c) *For any fixed* $h > 1$, $\tau > 0$, *for* $s^2 < (h-\tau)/2$, $x > \sqrt{n}$, *and all* n *large enough,*

$$\mathbf{P}(\overline{S}_n \ge x) \le e^{-x^2/(2nh)}. \tag{A8.2.10}$$

Remark A8.2.1 It is not hard to verify (see the proofs of Theorem A8.2.1 and Corollary A8.2.1) that there exists a function $\varepsilon(t) \downarrow 0$ as $t \uparrow \infty$ such that one has, along with (A8.2.8), the relation

$$\sup_{x:s\ge t}\frac{\mathbf{P}(\overline{S}_n > x)}{nV(x)} \le 1+\varepsilon(t).$$

Proof of Corollary A8.2.1 The proof is based on the inequality

$$\mathbf{P}(\overline{S}_n > x) \le \mathbf{P}(\overline{B}) + \mathbf{P}(\overline{S}_n \ge x;\ B) \le nV(y) + P. \tag{A8.2.11}$$

Since $\theta \to 0$ as $s \to \infty$, we see that, for any fixed $\varepsilon > 0$ and all $\varPi = nV(x)$ small enough, we have $P \le c(nV(y))^{r-\varepsilon}$. Putting $r := 1+2\varepsilon$, we obtain from (A8.2.11) and (A8.2.6) that

$$\mathbf{P}(\overline{S}_n \ge x) \le nV(y) + c\big(nV(y)\big)^{1+\varepsilon} \sim n(1+2\varepsilon)^{-\beta}V(x).$$

Since the left-hand side of this inequality does not depend on ε, relation (A8.2.8) follows.

We now prove (b). If $s \to \infty$ then (b) follows from (a). If s is bounded then necessarily $n \to \infty$ (since $nV(x) \to 0$) and hence

$$r - \theta = r - \frac{hr^2}{4s^2}\left(1 + b\,\frac{\ln s}{\ln n}\right) = \psi(r,s) + o(1),$$

where the function

$$\psi(r,s) := r - \frac{hr^2}{4s^2}$$

attains its maximum $\psi(r_0,s) = s^2/h$ in r at the point $r_0 = 2s^2/h$. Moreover, $\psi(r,s)$ strictly decreases in s. Therefore, for $r_0 = 2s^2/h$, we obtain

$$\psi(r_0,s) = \frac{s^2}{h}, \tag{A8.2.12}$$

$$r_0 - \theta \ge \frac{s^2}{h} + o(1) \quad \text{as } n \to \infty. \tag{A8.2.13}$$

Choose h so close to 1 and $\tau > 0$ so small that $h + \tau \le s_0^2$. Putting $r := r_0$, for $s^2 \ge s_0^2 \ge h + \tau$ and as $n \to \infty$, we get from (A8.2.6), (A8.2.12) and (A8.2.13) that

$$\mathbf{P}(\overline{S}_n \ge x) \le nV(y) + c\big(nV(y)\big)^{1+\tau/2} \sim nV\left(\frac{x}{r_0}\right) \sim r_0^{\beta}nV(x).$$

This proves (b).

Relation (c) for $y = x$ follows from the inequality (see (A8.2.7) and (A8.2.11))

$$\mathbf{P}(\overline{S}_n \ge x) \le nV(x) + e^{-x^2/(2nh)}, \tag{A8.2.14}$$

where, for $s^2 < (h-\tau)/2$,

$$e^{-x^2(2nh)} > \exp\left\{-\frac{(h-\tau)}{2}\frac{(\beta-2)n\ln n}{2nh}\right\} > n^{-(\beta-2)/4}.$$

On the other hand, we have $x > \sqrt{n}$,

$$nV(x) \le nV(\sqrt{n}) = n^{-(\beta-2)/2}L^*(n),$$

where L^* is a s.v.f. Therefore the second term dominates on the right-hand side of (A8.2.14). Slightly changing h if necessary, we obtain (c). Corollary A8.2.1 is proved. □

Remark A8.2.2 One can see from the proof of the corollary that the main contribution to the bound for the probability $\mathbf{P}(\overline{S}_n \ge x)$ under the conditions of assertions (a) and (b) comes from the event $\overline{B} = \{\max_{j\le n}\xi_j \ge y\}$ with y close to x, so that the most probable trajectory of $\{S_k\}_{k=1}^n$ that reaches the level x contains at least one jump ξ_j of size comparable to x.

Proof of Theorem A8.2.1 In our case, the Cramér condition **[C]** is not met. In order to use Theorem A8.1.1 in such a situation, we introduce "truncated" random variables with distributions that coincide with the conditional distribution of ξ given $\{\xi < y\}$ for some level y the choice of which will be at our disposal. Namely, we introduce independent identically distributed random variables $\xi_j^{(y)}$, $j = 1, 2, \ldots$, with the distribution function

$$\mathbf{P}\big(\xi_j^{(y)} < t\big) = \mathbf{P}(\xi < t \,|\, \xi < y) = \frac{\mathbf{P}(\xi < t)}{\mathbf{P}(\xi < y)}, \quad t \le y,$$

and put

$$S_n^{(y)} := \sum_{j=1}^n \xi_j^{(y)}, \qquad \overline{S}_n^{(y)} := \max_{k\le n} S_k^{(y)}.$$

Then

$$P = \mathbf{P}(\overline{S}_n \ge x, B) = \big(\mathbf{P}(\xi < y)\big)^n \mathbf{P}\big(\overline{S}_n^{(y)} \ge x\big). \tag{A8.2.15}$$

Applying Theorem A8.1.1 to the variables $\xi_j^{(y)}$, we obtain that, for any $\lambda \ge 0$,

$$\mathbf{P}\big(\overline{S}_n^{(y)} \ge x\big) \le e^{-\lambda x}\big[\max\big\{1, \mathbf{E}\,e^{\lambda\xi^{(y)}}\big\}\big]^n.$$

Since

$$\mathbf{E}e^{\lambda\xi^{(y)}} = \frac{R(\lambda, y)}{F(y)}, \quad \text{where } R(\lambda, y) := \int_\infty^y e^{\lambda t}\mathbf{F}(dt),$$

we arrive at the following basic inequality. For $x, y, \lambda \ge 0$,

$$P = \mathbf{P}(\overline{S}_n \ge x, B) \le e^{-\lambda x}\big[\max\big\{\mathbf{P}(\xi < y), R(\lambda, y)\big\}\big]^n$$

$$\le e^{-\lambda x} \max\bigl\{1, R^n(\lambda, y)\bigr\}. \tag{A8.2.16}$$

Thus, the main problem is to bound the integral $R(\lambda, y)$. Put

$$M(v) := \frac{v}{\lambda}$$

and represent $R(\lambda, y)$ as

$$R(\lambda, y) = I_1 + I_2,$$

where, for a fixed $\varepsilon > 0$,

$$I_1 := \int_{-\infty}^{M(\varepsilon)} e^{\lambda t}\mathbf{F}(dt) = \int_{-\infty}^{M(\varepsilon)} \left(1 + \lambda t + \frac{\lambda^2 t^2}{2} e^{\lambda\theta(t)}\right)\mathbf{F}(dt), \quad 0 \le \frac{\theta(t)}{t} \le 1. \tag{A8.2.17}$$

Here

$$\int_{-\infty}^{M(\varepsilon)} \mathbf{F}(dt) = 1 - V\bigl(M(\varepsilon)\bigr) \le 1,$$

$$\int_{-\infty}^{M(\varepsilon)} t\mathbf{F}(dt) = -\int_{M(\varepsilon)}^{\infty} t\mathbf{F}(dt) \le 0, \tag{A8.2.18}$$

$$\int_{-\infty}^{M(\varepsilon)} t^2 e^{\lambda\theta(t)}\mathbf{F}(dt) \le e^{\varepsilon} \int_{-\infty}^{M(\varepsilon)} t^2\mathbf{F}(dt) \le e^{\varepsilon} =: h. \tag{A8.2.19}$$

Therefore,

$$I_1 \le 1 + \frac{\lambda^2 h}{2}. \tag{A8.2.20}$$

Estimate now

$$I_2 := -\int_{M(\varepsilon)}^{y} e^{\lambda t}\, dF_+(t) \le V\bigl(M(\varepsilon)\bigr)e^{\varepsilon} + \lambda \int_{M(\varepsilon)}^{y} V(t)e^{\lambda t}\, dt. \tag{A8.2.21}$$

First consider, for $M(\varepsilon) < M(2\beta) < y$, the subintegral

$$I_{2,1} := \lambda \int_{M(\varepsilon)}^{M(2\beta)} V(t)e^{\lambda t}\, dt.$$

For $t = v/\lambda$, as $\lambda \to 0$, we have

$$V(t)e^{\lambda t} = V\left(\frac{v}{\lambda}\right)e^{v} \sim V\left(\frac{1}{\lambda}\right)f(v), \tag{A8.2.22}$$

where the function

$$f(v) := v^{-\beta}e^{v}$$

is convex on $(0, \infty)$. Therefore

$$I_{2,1} \le \frac{\lambda}{2}\big(M(2\beta) - M(\varepsilon)\big)V\left(\frac{1}{\lambda}\right)\big(f(\varepsilon) + f(2\beta)\big)\big(1 + o(1)\big) \le cV\left(\frac{1}{\lambda}\right). \quad \text{(A8.2.23)}$$

We now proceed to estimating the remaining subintegral

$$I_{2,2} := \lambda \int_{M(2\beta)}^{y} V(t)e^{\lambda t}\,dt.$$

For brevity's sake, put $M(2\beta) =: M$. We will choose λ so that

$$\mu = \lambda y \to \infty \quad (y \gg 1/\lambda) \qquad \text{(A8.2.24)}$$

as $x \to \infty$. Substituting the variable $(y - t)\lambda =: u$ we obtain

$$\lambda I_{2,2} = e^{\lambda y} V(y) \int_0^{(y-M)\lambda} V\left(y - \frac{u}{\lambda}\right) V^{-1}(y) e^{-u}\,du. \qquad \text{(A8.2.25)}$$

Consider the integral on the right-hand side of (A8.2.25). Since $1/\lambda \ll y$, the integrand

$$r_{y,\lambda}(u) := \frac{V(y - u/\lambda)}{V(y)}$$

converges to 1 for each fixed u. In order to use the dominated convergence theorem which implies that the integral on the right-hand side of (A8.2.25) converges, as $y \to \infty$, to

$$\int_0^\infty e^{-u}\,du = 1, \qquad \text{(A8.2.26)}$$

it remains to estimate the growth rate of the function $r_{y,\lambda}(u)$ as u increases. By the properties of r.v.f.s (see Theorem A6.2.1(iii) in Appendix 6), for all λ small enough (or M large enough; recall that $y - u/\lambda \ge M$ in the integrand in (A8.2.25)), we have

$$r_{y,\lambda}(u) \le \left(1 - \frac{u}{\lambda y}\right)^{-2\beta/2} =: g(u).$$

Since $g(0) = 1$ and $\lambda y - u \ge M\lambda = 2\beta$, in this domain

$$\big(\ln g(u)\big)' = \frac{3\beta}{2(\lambda y - u)} \le \frac{3\beta}{4\beta} = \frac{3}{4},$$

$$\ln g(u) \le \frac{3u}{4}, \qquad r_{y,\lambda}(u) \le e^{3u/4}.$$

This means that the integrand in (A8.2.25) is dominated by the exponential $e^{-u/4}$, and the use of the dominated convergence theorem is justified. Therefore, due to the convergence of the integral in (A8.2.25) to the limit (A8.2.26), we obtain

$$\lambda I_{2,2} \sim e^{\lambda y} V(y) \int_0^\infty e^u \, du = e^\mu V(y),$$

and it is not hard to find a function $\varepsilon(\mu) \downarrow 0$ as $\mu \uparrow \infty$ such that

$$\lambda I_{2,2} \le e^{\lambda y} V(y)\bigl(1 + \varepsilon(\mu)\bigr). \tag{A8.2.27}$$

Summarising (A8.2.20)–(A8.2.23) and (A8.2.27), we obtain

$$R(\lambda, y) \le 1 + \frac{\lambda^2 h}{2} + cV\left(\frac{1}{\lambda}\right) + V(y)e^{\lambda y}\bigl(1 + \varepsilon(\mu)\bigr), \tag{A8.2.28}$$

$$R^n(\lambda, y) \le \exp\left\{\frac{n\lambda^2 h}{2} + cnV\left(\frac{1}{\lambda}\right) + nV(y)e^{\lambda y}\bigl(1 + \varepsilon(\mu)\bigr)\right\}. \tag{A8.2.29}$$

First take λ to be the value

$$\lambda = \frac{1}{y} \ln T$$

that "almost minimises" the function $-\lambda x + nV(y)e^{\lambda y}$, where $T := \frac{r}{nV(y)}$, so that $\mu = \ln T$. Note that, for such a choice of μ (or of $\lambda = y^{-1} \ln(r/\varPi(y))$), for $\varPi(y) \to 0$ we have that $\mu = \lambda y \sim -\ln \varPi(y) \to \infty$ and hence that the assumption $y \gg 1/\lambda$ we made in (A8.2.24) holds true. For such λ,

$$R^n(\lambda, y) \le \exp\left\{\frac{n\lambda^2 h}{2} + cnV\left(\frac{1}{\lambda}\right) + r\bigl(1 + \varepsilon(\mu)\bigr)\right\}, \tag{A8.2.30}$$

where, by the properties of r.v.f.s,

$$nV\left(\frac{1}{\mu}\right) \sim nV\left(\frac{y}{\ln T}\right) \sim cnV\left(\frac{y}{|\ln nV(y)|}\right) \le cnV(y)\bigl|\ln nV(y)\bigr|^{\beta+\delta} \to 0,$$
$$\delta > 0, \tag{A8.2.31}$$

as $nV(y) \to 0$. Therefore

$$\begin{aligned}\ln P &\le -r \ln T + r + \frac{nh}{2y^2} \ln^2 T + \varepsilon_1(T) \\ &= \left[-r + \frac{nh}{2y^2} \ln T\right] \ln T + r + \varepsilon_1(T),\end{aligned} \tag{A8.2.32}$$

where $\varepsilon_1(T) \downarrow 0$ as $T \uparrow \infty$. If $x = sb(n)$, $b(n) = \sqrt{(\beta - 2)n \ln n}$, and $nV(x) \to 0$ then

$$\ln T = -\ln nV(x) + O(1) = -\ln n + \beta \ln s + \frac{\beta}{2} \ln n + O\bigl(\ln L\bigl(s\sigma(n)\bigr)\bigr) + O(1)$$

$$= \frac{\beta - 2}{2} \ln n \left[1 + b \frac{\ln s}{\ln n} \right] \bigl(1 + o(1)\bigr), \tag{A8.2.33}$$

where $b = \frac{2\beta}{\beta - 2}$ (the term $o(1)$ in the last equality appears because in our case either $n \to \infty$ or $s \to \infty$.) Hence, by (A8.2.32),

$$\frac{nh}{2y^2} \ln T = \frac{hr^2}{4s^2} \left[1 + b \frac{\ln s}{\ln n} \right] \bigl(1 + o(1)\bigr),$$

$$\ln P \le r - \left[r - \frac{h' r^2}{4s^2} \left(1 + b \frac{\ln s}{\ln n} \right) \right] \ln T$$

for any $h' > h > 1$ and $nV(x)$ small enough. This proves the first assertion of the theorem.

We now prove the second assertion of the theorem for "small" values of s such that, for some $\tau > 0$,

$$s^2 < \frac{h - \tau}{2}.$$

Since we always assume that $x > \sqrt{n}$, we also have

$$s = \frac{x}{b(n)} > \frac{1}{\sqrt{(\beta - 2) \ln n}}$$

and we can assume that $s^2 \ge n^{-\gamma}$ for some $\gamma > 0$ to be chosen below. This corresponds to the following domain of the values of x^2:

$$cn^{1-\gamma} \ln n < x^2 < \frac{(h - \tau)(\beta - 2)}{2} n \ln n. \tag{A8.2.34}$$

For such x, as will be shown below, the main contribution to the exponent on the right-hand side of (A8.2.29) comes from the quadratic term $n\lambda^2 h/2$, and we will set

$$\lambda := \frac{x}{nh}.$$

Then, for $y = x$ ($r = 1$, $\mu = x^2/(nh)$),

$$\begin{aligned} \ln P &\le -\lambda x + \frac{n\lambda^2 h}{2} + cnV\left(\frac{1}{\lambda}\right) + nV(y) e^{\lambda y} \bigl(1 + \varepsilon(\mu)\bigr) \\ &= -\frac{x^2}{2nh} + cnV\left(\frac{nh}{x}\right) + nV(x) e^{\frac{x^2}{nh}} \bigl(1 + \varepsilon(\mu)\bigr). \end{aligned} \tag{A8.2.35}$$

We show that the last two terms on the right-hand side are negligibly small as $n \to \infty$. Indeed, by the second inequality in (A8.2.34),

$$nV\left(\frac{nh}{x}\right) \le cnV\left(\sqrt{\frac{n}{\ln n}}\right) \to 0 \quad \text{as } n \to \infty.$$

Further, by the first inequality in (A8.2.34),

$$nV(x) \le n^{(2-\beta)/2+\gamma'},$$

where we can choose γ'. Moreover, by (A8.2.34),

$$\frac{x^2}{nh} \le \frac{(h-\tau)(\beta-2)\ln n}{2h} = \left[\frac{\beta-2}{2} - \frac{\tau(\beta-2)}{2h}\right]\ln n.$$

Therefore

$$nV(x)e^{x^2/(nh)} \le n^{-\tau(\beta-2)/(2h)+\gamma'} \to 0$$

for $\gamma' < \frac{\tau(\beta-2)}{2h}$ as $n \to \infty$.

Thus,

$$\ln P \le -\frac{x^2}{2nh} + o(1).$$

Since $x^2/n > 1$, the term $o(1)$ in the last relation can be omitted by slightly changing $h > 1$. (Formally, we proved that, for any $h > 1$ and all n large enough, inequality (A8.2.7) is valid with the h on its right-hand side replaced with $h' > h$, where we can take, for instance, $h' = h + (h-1)/2$. Since $h' > 1$ can also be made arbitrarily close to 1 by the choice of h, the obtained relation is equivalent to the one from Theorem A8.2.1.) This proves (A8.2.7).

The theorem is proved. □

Comparing the assertions of Theorem A8.2.1 and Corollary A8.1.1, we see that, roughly speaking, for $s < 1/2$ and for $s > 1$ one can obtain quite satisfactory and, in a certain sense, unimprovable upper bounds for the probabilities P and $\mathbf{P}(\overline{S}_n > x)$.

8.3 Lower Bounds

In this section we will again assume that conditions (A8.2.1) are satisfied. The lower bounds for $\mathbf{P}(S_n \ge x)$ (they will clearly hold for $\mathbf{P}(\overline{S}_n \ge x)$ as well) can be obtained in a much simpler way than the upper bounds and need essentially no assumptions.

Theorem A8.3.1 *Let* $\mathbf{E}\xi_j = 0$ *and* $\mathbf{E}\xi_j^2 = 1$. *Then, for* $y = x + t\sqrt{n-1}$,

$$\mathbf{P}(S_n \ge x) \ge nF_+(y)\left[1 - t^{-2} - \frac{n-1}{2}F_+(y)\right]. \tag{A8.3.1}$$

Proof Put $G_n := \{S_n \ge x\}$ and $B_j := \{\xi_j < y\}$. Then

$$\mathbf{P}(S_n \ge x) \ge \mathbf{P}\left(G_n; \bigcup_{j=1}^n \overline{B}_j\right) \ge \sum_{j=1}^n \mathbf{P}(G_n\overline{B}_j) - \sum_{i<j\le n} \mathbf{P}(G_n\overline{B}_i\overline{B}_j)$$

$$\geq \sum_{j=1}^{n} \mathbf{P}(G_n \overline{B}_j) - \frac{n(n-1)}{2} F_+^2(y).$$

Here, for $y = x + t\sqrt{n-1}$,

$$\mathbf{P}(G_n \overline{B}_j) = \int_y^\infty \mathbf{P}(S_{n-1} \geq x - u)\mathbf{F}(du) \geq \mathbf{P}(S_{n-1} \geq x - y)F_+(y)$$
$$= \mathbf{P}(S_{n-1} \geq -t\sqrt{n-1})F_+(y) = \big(1 - \mathbf{P}(S_{n-1} < -t\sqrt{n-1})\big)F_+(t),$$

where, by the Chebyshev inequality,

$$\mathbf{P}(S_{n-1} < -t\sqrt{n-1}) \leq t^{-2}.$$

As a result we get

$$\mathbf{P}(S_n \geq x) \geq nF_+(t)\big(1 - t^{-2}\big) - \frac{n(n-1)}{2} F_+^2(t),$$

which is equivalent to (A8.3.1).

The theorem is proved. □

Corollary A8.3.1 *If $x \to \infty$ and $x \gg \sqrt{n}$ then, as $t \to \infty$,*

$$\mathbf{P}(S_n \geq x) \geq nF_+(y)\big(1 + o(1)\big). \tag{A8.3.2}$$

If, moreover, $F_+(u) \geq V(u) \in \mathcal{R}$ then

$$\mathbf{P}(S_n \geq x) \geq nV(x)\big(1 + o(1)\big).$$

Proof Since $y \geq x$, we have

$$nF_+(y) \ll ny^{-2} < nx^{-2} = o(1).$$

This together with (A8.3.1) implies the first assertion of the corollary as $t \to \infty$. To obtain the second one, in (A8.3.2) one should take $t \to \infty$ such that $t = o(x/\sqrt{n})$. Then $y \sim x$ and $V(y) \sim V(x)$.

The corollary is proved. □

Appendix 9
Renewal Theorems

The main goal of the present section is to prove Theorem 10.4.1, the key renewal theorem in the non-arithmetic case (in the terminology of Chap. 10). We will also consider some refinements and extensions of the theorem.

First consider *positive* independent identically distributed random variables $\tau_j \overset{d}{=} \tau$ with distribution function F and finite mean $a := \mathbf{E}\tau < \infty$. Here it will be more convenient to understand by the renewal function its left-continuous version

$$H(t) := \sum_{k=0}^{\infty} F^{*k}(t), \qquad t \ge 0,$$

where F^{*k} is the k-fold convolution of the distribution F with itself, which is the distribution function of the sum $T_k = \tau_1 + \cdots + \tau_k$. We first prove the following key assertion.

Theorem A9.1 *If g is a directly integrable function and τ_j are non-arithmetic* (*see Chap*. 10) *then*, *as $t \to \infty$*,

$$\int_0^t g(t-u)\, dH(u) \to \frac{1}{a} \int_0^\infty g(u)\, du.$$

The proof of the theorem mostly follows the argument suggested in [13] and will need several auxiliary assertions.

Lemma A9.1 *Let g be a bounded measurable function. The integral*

$$G(t) = \int_0^t g(t-u)\, dH(u) =: g * H(t) \tag{A9.1}$$

is the unique solution of the equation

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

$$G(t)=g(t)+\int_0^t G(t-u)\,dF(u)=g(t)+G*F(t) \qquad \text{(A9.2)}$$

in the class of functions bounded on finite intervals.

The function $G=H$ is the solution of (A9.2) *when $g\equiv 1$. The function $G\equiv 1$ is the solution of* (A9.2) *when $g=1-F$.*

Equation (A9.2) is called the *renewal equation*.

As we already noted in Theorem 10.4.1, one can associate, in an obvious way, measures **H** and **F** with the functions H and F, and write the integrals in (A9.1) and (A9.2) as integrals with respect to the measures:

$$\int_0^t g(t-u)\mathbf{H}(du) \quad \text{and} \quad \int_0^t G(t-u)\mathbf{F}(du),$$

respectively.

Proof of Lemma A9.1 Put

$$H_n(t):=\sum_{k=0}^{n} F^{*k}(t).$$

The functions $G_n=g*H_n$ satisfy the equation $G_{n+1}=g+G_n*F$ and form an increasing sequence $G_n\uparrow$ which is bounded by Lemma 10.2.3. Therefore $G_n\uparrow G$, and passing to the limit in the equation for G_n we obtain that G satisfies (A9.1). To prove uniqueness note that the difference $V=G^{(1)}-G^{(2)}$ of two solutions $G^{(1)}$ and $G^{(2)}$ must satisfy the homogeneous equation $V=V*F$ and therefore also the relations $V=V*(F^{k*})$ or, which is the same,

$$V(t)=\int_0^t V(t-u)\,dF^{*k}(u).$$

But $F^{*k}(u)\to 0$ as $k\to\infty$ for $u\in[0,t]$. Since by the assumption $|V(u)|<c$ on $[0,t]$, we have $V(t)\to 0$ as $k\to\infty$. But V does not depend on k, so that $V(t)\equiv 0$. The last assertion of the lemma can be verified directly. The lemma is proved. □

Note that if we considered functions g of bounded variation, the assertion of Lemma A9.1 would immediately follow from the equation for the Laplace–Stieltjes transform $\widetilde{G}(\lambda)=\int_0^\infty e^{-\lambda t}\,dG(t)$ of G which follows from (A9.2):

$$\widetilde{G}(\lambda)=\widetilde{g}(\lambda)+\widetilde{G}(\lambda)\psi(\lambda), \qquad \text{(A9.3)}$$

where

$$\widetilde{g}(\lambda):=\int_0^\infty e^{-\lambda t}\,dg(t), \qquad \psi(\lambda):=\int_0^\infty e^{-\lambda t}\,dF(t).$$

Indeed, it follows from (A9.3) that

$$\widetilde{G}(\lambda) = \frac{\widetilde{g}(\lambda)}{1 - \psi(\lambda)},$$

which is equivalent to (A9.1).

A point t is said to be a *point of growth* of the distribution function F provided that $F(t+\varepsilon) - F(t) > 0$ for any $\varepsilon > 0$.

Lemma A9.2 *Let the distribution F be non-arithmetic and Z be the set of all points of growth of H, i.e. points of growth of the functions $F, F^{*2}, F^{*3}, \ldots$. Then Z is "asymptotically dense at infinity", i.e., for any given $\varepsilon > 0$ and all x large enough, the intersection $(x, x+\varepsilon) \cap Z$ is non-empty.*

Proof Observe first that if t_1 is a point of growth of the distribution F_1 of a random variable τ, and t_2 is a point of growth of the distribution F_2 of a random variable ζ which is independent of τ, then $t = t_1 + t_2$ will be a point of growth of the distribution $F_1 * F_2$ of the variable $\tau + \zeta$. Indeed,

$$\mathbf{P}(t \le \tau + \zeta < t + \varepsilon) \ge \mathbf{P}\left(t_1 \le \tau < t_1 + \frac{\varepsilon}{2}\right)\mathbf{P}\left(t_2 \le \zeta < t_2 + \frac{\varepsilon}{2}\right).$$

Let, further, $x < y$ be two points of the set Z, and $\Delta := y - x$. The following alternative takes place: either

(1) for any $\varepsilon > 0$ there exist x and y such that $\Delta < \varepsilon$, or

(2) there exists a $\delta > 0$ such that $\Delta \ge \delta$ for all x and y from Z.

Put $I_n := [xn, yn]$. If $n\Delta > x$ then that interval contains $[nx, (n+1)x]$ as a subset, and therefore any point $v > v_0 = x^2/\Delta$ belongs to at least one of the intervals $I_1, I_2, \ldots$.

By virtue of the above observation, the $n+1$ points $nx + k\Delta = (n-k)x + ky$, $k = 0, \ldots, n$, belong to Z and divide I_n into n subintervals of length Δ. This means that, for any point $v > v_0$, the distance between v and the points from Z is at most $\Delta/2$.

This implies the assertion of the lemma when (1) holds.

If (2) is true, we can assume that x and y are chosen so that $\Delta < 2\delta$. Then the points of the form $nx + k\Delta$ exhaust all the points from Z lying inside I_n. Since the point $(n+1)x$ is among these points, the value x is a multiple of Δ, and all the points of Z lying inside I_n are multiples of Δ. Now let z be an arbitrary point of growth of F. For sufficiently large n, the interval I_n contains a point of the form $z + k\Delta$, and since the latter belongs to Z, the value z is also a multiple of Δ. Thus F is an arithmetic distribution, so that case (2) cannot take place. The lemma is proved. □

Lemma A9.3 *Let $q(x)$ be a bounded uniformly continuous function given on $(-\infty, \infty)$ such that, for all x, $q(x) \le q(0)$ for all x, and*

$$q(x) = \int_0^\infty q(x-y)\, dF(y). \tag{A9.4}$$

Then $q(x) \equiv q(0)$.

Proof Equation (A9.4) means that $q = q * F = \cdots = q * F^{*k}$ for all $k \geq 1$. The right-hand side of (A9.4) does not exceed $q(0)$, and hence, for $x = 0$, the equality (A9.4) is only possible if $q(-y) = q(0)$ for all $y \in Z_k$, where Z_k is the set of points of growth of F^{*k}, and therefore $q(-y) = q(0)$ for all $y \in Z$. By Lemma A9.2 and the uniform continuity of q this means that $q(-y) \to q(0)$ as $y \to \infty$. Further, for an arbitrarily large N we can choose k such that $q(x)$ will be arbitrarily close to $\int_N^\infty q(x-y)\,dF^{*k}(y)$, since $F^{*k}(N) \to 0$ as $k \to \infty$. This means, in turn, that $q(x)$ will be close to $q(0)$. Since $q(x)$ depends neither on N nor on k, we have $q(x) = q(0)$. The lemma is proved. □

Lemma A9.4 *Let g be a continuous function vanishing outside segment* $[0, b]$. *Then the solution G of the renewal equation* (A9.2) *is uniformly continuous and, for any u,*

$$G(x+u) - G(x) \to 0 \tag{A9.5}$$

as $x \to \infty$.

Proof By virtue of Lemma 10.2.3,

$$\begin{aligned} \bigl|G(x+\delta) - G(x)\bigr| &= \left|\int_{x-b}^{x+\delta} \bigl(g(x+\delta-y) - g(x-y)\bigr)\,dH(y)\right| \\ &\leq \max_{0 \leq x \leq b+\delta} \bigl|g(x+\delta) - g(x)\bigr|\bigl(c_1 + c_2(b+\delta)\bigr). \end{aligned} \tag{A9.6}$$

This means that the uniform continuity of g implies that of G.

Now assume that g has a continuous derivative g'. Then G' exists and satisfies the renewal equation

$$G'(x) = g'(x) + \int_0^x G'(x-y)\,dF(y).$$

Therefore the derivative G' is bounded and uniformly continuous. Let

$$\limsup_{x \to \infty} G'(x) = s.$$

Choose a sequence $t_n \to \infty$ such that $G'(t_n) \to s$. The family of functions q_n defined by the equalities

$$q_n(x) = G'(t_n + x)$$

is equicontinuous, and

$$q_n(x) = g'(t_n + x) + \int_0^{x+t_n} q_n(x-y)\,dF(y) = g'(t_n + x) + \int_0^\infty q_n(x-y)\,dF(y). \tag{A9.7}$$

By the Arzelà–Ascoli theorem (see Appendix 4) there exists a subsequence t_{n_r} such that q_{n_r} converges to a limit q. From (A9.7) it follows that this limit satisfies the conditions of Lemma A9.3, and therefore $q(x) = q(0) = s$ for all x. Thus $G'(t_{n_r}+x) \to s$ for all x, and hence

$$G(t_{n_r} + x) - G(t_{n_r}) \to sx.$$

Since the last relation holds for any x and the function g is bounded, we get $s = 0$.

We have proved the lemma for continuously differentiable g. But an arbitrary continuous function g vanishing outside $[0, b]$ can be approximated by a continuously differentiable function g_1 which also vanishes outside that interval. Let G_1 be the solution of the renewal equation corresponding to the function g_1. Then $|g - g_1| < \varepsilon$ implies $|G - G_1| < c\varepsilon$, $c = c_1 + c_2 b$ (see Lemma 10.2.3), and therefore

$$\bigl|G(x+u) - G(x)\bigr| < (2c+1)\varepsilon$$

for all sufficiently large x. This proves (A9.5) for arbitrary continuous functions g. The lemma is proved. □

Proof of Theorem A9.1 Consider an arbitrary sequence $t_n \to \infty$ and the measures $\boldsymbol{\mu}_n$ generated by the functions

$$H_{(n)}(u) = H(t_n + u) - H(t_n) \qquad \bigl(\boldsymbol{\mu}_n\bigl([u, v)\bigr) = H_{(n)}(v) - H_{(n)}(u)\bigr).$$

These functions satisfy the conditions of the generalised Helly theorem (see Appendix 4). Therefore there exists a subsequence t_{nn}, the respective subsequence of measures $\boldsymbol{\mu}_{nn}$, and the limiting measure $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}_{nn}$ converges weakly to $\boldsymbol{\mu}$ on any finite interval as $n \to \infty$.

Now let g be a continuous function vanishing outside $[0, b]$. Then

$$\begin{aligned} G(t_{nn} + x) &= \int_{-b}^{0} g(-u)\,dH(t_{nn} + x + u) \\ &= \int_{-b}^{0} g(-u)\,d\bigl(H(t_{nn} + x + u) - H(t_{nn})\bigr) \to \int_{0}^{b} g(u)\boldsymbol{\mu}(x + du). \end{aligned}$$

By Lemma A9.4, the sequence $G(t_{nn} + y)$ will have the same limit. This means that the measure $\boldsymbol{\mu}(x+du)$ does not depend on x, and therefore $\boldsymbol{\mu}([u, v))$ is proportional to the length of the interval (u, v):

$$\boldsymbol{\mu}\bigl((u, v)\bigr) = c(v - u), \quad \boldsymbol{\mu}(du) = c\,du.$$

Thus, we have proved that

$$G(t_{nn} + x) \to c\int_{0}^{\infty} g(u)\,du \tag{A9.8}$$

for any continuous function g vanishing outside $[0, b]$. But for any Riemann integrable function g on $[0, b]$ and given $\varepsilon > 0$ there exist continuous functions g_1 and g_2, $g_1 < g < g_2$, which are equal to 0 outside $[0, b+1]$ and such that

$$\int_0^b (g_2 - g_1)\, du < \varepsilon.$$

This means that convergence (A9.8) also holds for any Riemann integrable function vanishing outside $[0, b]$.

Now consider an arbitrary directly integrable function g. By property (2) of such functions (see Definition 10.4.1) one can choose a $b > 0$ such that for the function

$$g_{(b)}(u) = \begin{cases} g(u) & \text{if } u \le b, \\ 0 & \text{if } u > b, \end{cases}$$

the left- and right-hand sides of (A9.8) will be arbitrarily close to the respective expressions corresponding to the original function g (for the right-hand side it is obvious, while for the left-hand side it follows from the convergence

$$\left| \int_0^t g(t-s)\, dH(s) - \int g_{(b)}(t-s)\, dH(s) \right|$$
$$= \left| \int_0^{t-b} g(t-s)\, dH(s) \right| \le \sum_{k > b-1} (c_1 + c_2) g_k \to 0$$

as $b \to \infty$ (see Lemma 10.2.3)). Therefore (A9.8) is proved for any directly integrable function g. Putting $g := 1 - F$ we obtain from Lemma A9.1

$$1 = c \int_0^\infty \big(1 - F(u)\big)\, du = ac, \quad c = \frac{1}{a}.$$

Thus the limit in (A9.8) is one and the same for any initial sequence t_n. From this it follows that, as $t \to \infty$,

$$G(t) \to \frac{1}{a} \int_0^\infty g(u)\, du.$$

The theorem is proved. □

Theorem 10.4.1 is a simple consequence of Theorem A9.1 and the argument used in the proof of Theorem 10.2.3 that extends the key renewal theorem in the arithmetic case was extended to the setting where τ_j, $j \ge 2$, can assume values of different signs, while τ_1 is arbitrary. We will leave it to the reader to apply the argument in the non-arithmetic case.

Now we will give several further consequences of Theorem A9.1. In Sect. 10.4 we obtained a refinement of the renewal theorem in the case when $m_2 := \mathbf{E}\tau_j^2 < \infty$. Approaches developed while proving Theorem A9.1 enable one to obtain an alternative proof of the following assertion coinciding with Theorem 10.4.4.

Theorem A9.2 *Let the conditions of Theorem A9.1 be met and $m_2 < \infty$. Then*

$$0 \le H(t) - \frac{t}{a} \to \frac{m_2}{2a^2} \quad \text{as } t \to \infty.$$

Proof The function $G(t) := H(t) - t/a$ is the solution of the renewal equation (A9.2) corresponding to the function

$$g(t) := \frac{1}{a} \int_t^\infty \big(1 - F(u)\big)\, du.$$

Since g is directly integrable, we have

$$G(t) \to \frac{1}{a} \int_0^\infty \int_v^\infty \big(1 - F(u)\big)\, du\, dv = \frac{m_2}{2a^2}.$$

The theorem is proved. □

Theorem A9.3 (The local renewal theorem for densities) *Assume that F has a density $f = F'$ and this density is directly integrable. Then H has a density $h = H'$, and*

$$h(t) \to \frac{1}{a} \quad \text{as } t \to \infty.$$

Proof Denote by $f_n(x)$ the density of the sum $T_n = \tau_1 + \cdots + \tau_n$. We have

$$h(t) = H'(t) = \sum_{n=1}^\infty f_n(t) = f(t) + \int h(t-u) f(u)\, du = f(t) + h * F(t).$$

This means that $h(t)$ satisfies the renewal equation with the function $g = f$. Therefore by Theorem A9.1,

$$h(t) \to \frac{1}{a} \int_0^\infty f(u)\, du = \frac{1}{a}.$$

The theorem is proved. □

Consider now some extensions of Theorem A9.1. A function g given on the whole line $(-\infty, \infty)$ is said to be directly integrable if both functions $g(t)$ and $g(-t)$, $t \ge 0$, are directly integrable.

Theorem A9.4 *If the conditions of Theorem A9.1 are met and g is directly integrable, then*

$$G(t) = \int_0^\infty g(t-u)\mathbf{H}(du) \to \frac{1}{a} \int_{-\infty}^\infty g(u)\, du \quad \text{as } t \to \infty.$$

The Proof can be obtained by making several small and quite obvious modifications to the argument in the demonstration of Theorem A9.1. The main change is that instead of functions g vanishing outside $[0, b]$ one should now consider functions vanishing outside $[-b, b]$.

Another extension refers to the second version of the renewal function

$$U(t) := \sum_{k=0}^{\infty} F^{*k}(t), \quad -\infty < t < \infty,$$

in the case when τ_j can assume values of different signs.

Theorem A9.5 *If g is directly integrable and* $\mathbf{E}\tau_j = a > 0$, *then*

$$G(t) = \int_{-\infty}^{\infty} g(t-u)\mathbf{U}(du) \to \frac{1}{a}\int_{-\infty}^{\infty} g(u)\,du \quad \text{as } t \to \infty,$$

and, for any fixed u, $U(t+u) - U(t) \to 0$ *as* $t \to \infty$.

The proof is also obtained by modifying the argument proving Theorem A9.1 (see [13]).

References

1. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
2. Billingsley, P.: Probability and Measure. Anniversary edn. Wiley, Hoboken (2012)
3. Borovkov, A.A.: Stochastic Processes in Queueing Theory. Springer, New York (1976)
4. Borovkov, A.A.: Convergence of measures and random processes. Russ. Math. Surv. **31**, 1–69 (1976)
5. Borovkov, A.A.: Probability Theory. Gordon & Breach, Amsterdam (1998)
6. Borovkov, A.A.: Ergodicity and Stability of Stochastic Processes. Wiley, Chichester (1998)
7. Borovkov, A.A.: Mathematical Statistics. Gordon & Breach, Amsterdam (1998)
8. Borovkov, A.A., Borovkov, K.A.: Asymptotic Analysis of Random Walks. Heavy-Tailed Distributions. Cambridge University Press, Cambridge (2008)
9. Cramér, H., Leadbetter, M.R.: Stationary and Related Stochastic Processes. Willey, New York (1967)
10. Dudley, R.M.: Real Analysis and Probability. Cambridge University Press, Cambridge (2002)
11. Feinstein, A.: Foundations of Information Theory. McGraw-Hill, New York (1995)
12. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1. Wiley, New York (1968)
13. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (1971)
14. Gikhman, I.I., Skorokhod, A.V.: Introduction to the Theory of Random Processes. Saunders, Philadelphia (1969)
15. Gnedenko, B.V.: The Theory of Probability. Chelsea, New York (1962)
16. Gnedenko, B.V., Kolmogorov, A.N.: Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Reading (1968)
17. Gradsteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products. Academic Press, New York (1965). Fourth edition prepared by Ju.V. Geronimus and M.Ju. Ceitlin. Translated from the Russian by Scripta Technica, Inc. Translation edited by Alan Jeffrey
18. Grenander, U.: Probabilities on Algebraic Structures. Almqvist & Wiskel, Stockholm (1963)
19. Halmos, P.R.: Measure Theory. Van Nostrand, New York (1950)
20. Ibragimov, I.A., Linnik, Yu.V.: Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Croningen (1971)
21. Khinchin, A.Ya.: Ponyatie entropii v teorii veroyatnostei (The concept of entropy in the theory probability). Usp. Mat. Nauk **8**, 3–20 (1953) (in Russian)
22. Kifer, Yu.: Ergodic Theory of Random Transformations. Birkhäuser, Boston (1986)
23. Kolmogorov, A.N.: Markov chains with a countable number of possible states. In: Shiryaev, A.N. (ed.) Selected Works of A.N. Kolmogorov, vol. 2, pp. 193–208. Kluwer Academic, Dordrecht (1986)

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

24. Kolmogorov, A.N.: The theory of probability. In: Aleksandrov, A.D., et al. (eds.) Mathematics, Its Content, Methods, and Meaning, vol. 2, pp. 229–264. MIT Press, Cambridge (1963)
25. Lamperti, J.: Probability: A Survey of the Mathematical Theory. Wiley, New York (1996)
26. Loeve, M.: Probability Theory. Springer, New York (1977)
27. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer, New York (1993)
28. Natanson, I.P.: Theory of Functions of a Real Variable. Ungar, New York (1961)
29. Nummelin, E.: General Irreducible Markov Chains and Nonnegative Operators. Cambridge University Press, New York (1984)
30. Petrov, V.V.: Sums of Independent Random Variables. Springer, New York (1975)
31. Shiryaev, A.N.: Probability. Springer, New York (1984)
32. Skorokhod, A.V.: Random Processes with Independent Increments. Kluwer Academic, Dordrecht (1991)
33. Tyurin, I.S.: An improvement of the residual in the Lyapunov theorem. Theory Probab. Appl. **56**(4) (2011)

Index of Basic Notation

Spaces and σ-algebras

Distributions[1]

[1](All distributions and measures are denoted by bold letters).

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013

Relations

$:=$ means that the left-hand side is defined by the right-hand side, xi
$=:$ means that the right-hand side is defined by the left-hand side, xi
$\sim$ notation $a_n \sim b_n$ $(a(x) \sim b(x))$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$ $(\lim_{x\to\infty} \frac{a(x)}{b(x)} = 1)$, 109
$\xrightarrow{p}$—convergence of random variables in probability, 129
$\xrightarrow{a.s.}$—almost sure convergence of random variables, 130
$\xrightarrow{(r)}$—convergence of random variables in the mean, 132
$\stackrel{d}{=}$—notation $\xi \stackrel{d}{=} \eta$ means that the distributions of ξ and η coincide, 144
$\stackrel{d}{\le}$—relation $\xi \stackrel{d}{\le} \gamma$ means that $\mathbf{P}(\xi \ge t) \le \mathbf{P}(\gamma \ge t)$ for all t, 302
$\stackrel{d}{>}$—relation $\xi \underset{d}{>} \eta$ means that $\mathbf{P}(\xi > t) \ge \mathbf{P}(\eta > t)$ for all t, 302
$\Subset$—notation $\xi_n \Subset \mathbf{F}$ means that ξ has the distribution $\mathbf{F}$, 36
$\xi_n \Subset\!\!\Rightarrow \mathbf{F}$ means that the distribution of ξ_n converges weakly to $\mathbf{F}$, 144
$\Rightarrow$—relation $\mathbf{F}_n \Rightarrow \mathbf{F}$ means weak convergence of the distributions $\mathbf{F}_n$ to $\mathbf{F}$, 141, for random variables $\xi_n \Rightarrow \xi$ means that $\mathbf{F}_n \Rightarrow \mathbf{F}$, where $\xi_n \Subset \mathbf{F}_n$, $\xi \Subset \mathbf{F}$, 143

Conditions

[**C**]—the Cramér condition, 240
[$\mathbf{R}_{\beta,\rho}$]—conditions of convergence to the stable law $\mathbf{F}_{\beta,\rho}$, 229

Subject Index

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9, © Springer-Verlag London 2013