3

Interlude: Results from Probability Theory

Primes play a game of chance.

M. Kac

In this chapter, we briefly present facts from probability theory which will be used later. These results can be found in the monographs of Billingsley [21, 22], Buldygin [45], Cramér and Leadbetter [64], Heyer [133], Laurinčikas [186], and Loève [226]. However, there are two exceptions in this crash course in probability theory. In Sect. 3.3 we present Denjoy's heuristic probabilistic argument for the truth of Riemann's hypothesis. Finally, in Sect. 3.7, we introduce the universe for our later studies on universality, the space of analytic functions, and state some of its properties, following Conway [62] and Laurinčikas [186].

3.1 Weak Convergence of Probability Measures

The notion of weak convergence of probability measures is a useful tool in investigations on the value-distribution of Dirichlet series. This powerful theory was initiated by Kolmogorov, Erdös and Kac and further developed by Doob, Prokhorov, Skorokhod and others. To present the main properties of weakly convergent probability measures we have to introduce the concept of σ-field and the axiomatic setting of probability measures.

Let Ω be a non-empty set. By $\mathcal{P}(\Omega)$ we denote the set of all subsets of Ω. A subset $\mathcal{F}$ of $\mathcal{P}(\Omega)$ is called a field (or algebra) if it satisfies the following axioms:

- $\emptyset, \Omega \in \mathcal{F}$;
- $A^{\mathrm{c}} \in \mathcal{F}$ for $A \in \mathcal{F}$, where A^{c} denotes the complement of A;
- $\mathcal{F}$ is closed under finite unions and finite intersections, i.e., if $A_1, \ldots, A_n \in \mathcal{F}$, then

$$\bigcup_{j=1}^{n} A_j \in \mathcal{F} \quad \text{and} \quad \bigcap_{j=1}^{n} A_j \in \mathcal{F}.$$

$\mathcal{F}$ is called a σ-field (or σ-algebra) if it satisfies the first two axioms above in addition with

- $\mathcal{F}$ is closed under countable unions and countable intersections, i.e., if $\{A_j\}$ is a countable sequence of events in $\mathcal{F}$, then

$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{F} \qquad \text{and} \qquad \bigcap_{j=1}^{\infty} A_j \in \mathcal{F}.$$

For $\mathcal{C} \subset \mathcal{P}(\Omega)$ we denote by $\sigma(\mathcal{C})$ the smallest σ-field containing $\mathcal{C}$. This σ-field is said to be generated by $\mathcal{C}$.

A non-negative function $\mathbf{P}$ defined on a σ-field $\mathcal{F}$ with the properties:

- $\mathbf{P}(\emptyset) = 0$ and $\mathbf{P}(\Omega) = 1$;
- For every countable sequence $\{A_j\}$ of pairwise disjoint elements of $\mathcal{F}$,

$$\mathbf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbf{P}(A_j),$$

is called a probability measure. The triple $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be a probability space. This setting of probability dates back to Kolmogorov who introduced it in the early 1930s.

Let S be a topological space and let $\mathcal{B}(S)$ denote the class of Borel sets of S, i.e., the σ-field generated by the system of all open subsets of the space S. Then each measure on $\mathcal{B}(S)$ is called Borel measure. Usually, we consider probability measures defined on the Borel sets $\mathcal{B}(S)$ of some metric space S. A class $\mathcal{A}$ of sets of S is said to be a determining class (also separating class) in case the measures $\mathbf{P}$ and $\mathbf{Q}$ on $(S, \mathcal{B}(S))$ coincide on the whole of S when $\mathbf{P}(A) = \mathbf{Q}(A)$ for all $A \in \mathcal{A}$.

Given two probability measures $\mathbf{P}_1$ and $\mathbf{P}_2$ on $(S_1, \mathcal{B}(S_1))$ and $(S_2, \mathcal{B}(S_2))$, respectively, there exists a unique measure $\mathbf{P}_1 \times \mathbf{P}_2$ such that

$$(\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times A_2) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2)$$

for $A_j \in \mathcal{B}(S_j)$. This measure is a probability measure on $(S, \mathcal{B}(S))$, where $S = S_1 \times S_2$ and $\mathcal{B}(S) = \mathcal{B}(S_1) \times \mathcal{B}(S_2)$, and is said to be the product measure of the measures $\mathbf{P}_1$ and $\mathbf{P}_2$.

In the sequel let $\mathbf{P}_n$ and $\mathbf{P}$ be probability measures on $(S, \mathcal{B}(S))$. We say that $\mathbf{P}_n$ converges weakly to $\mathbf{P}$ as n tends to infinity, and write $\mathbf{P}_n \Rightarrow \mathbf{P}$, if for all bounded continuous functions $f : S \to \mathbb{R}$

$$\lim_{n\to\infty} \int_S f \, \mathrm{d}\mathbf{P}_n = \int_S f \, \mathrm{d}\mathbf{P}.$$

Since the integrals on the right-hand side completely determine $\mathbf{P}$ (which is a consequence of Lebesgue's dominated convergence theorem), the sequence

$\{\mathbf{P}_n\}$ cannot converge weakly to two different limits at the same time. Further, note that weak convergence depends only on the topology of the underlying space S, not on the metric that generates it.

A set A in S whose boundary ∂A satisfies $\mathbf{P}(\partial A) = 0$ is called a continuity set of $\mathbf{P}$. The Portmanteau theorem provides useful conditions equivalent to weak convergence.

Theorem 3.1. *Let $\mathbf{P}_n$ and $\mathbf{P}$ be probability measures on $(S, \mathcal{B}(S))$. Then the following assertions are equivalent:*

- $\mathbf{P}_n \Rightarrow \mathbf{P}$*;*
- *For all open sets G,*
$$\liminf_{n\to\infty} \mathbf{P}_n(G) \geq \mathbf{P}(G);$$
- *For all continuity sets A of $\mathbf{P}$,*
$$\lim_{n\to\infty} \mathbf{P}_n(A) = \mathbf{P}(A).$$

This theorem is part of Theorem 2.1 in Billingsley [21]. Next we state a useful criterion for weak convergence.

Lemma 3.2. *We have $\mathbf{P}_n \Rightarrow \mathbf{P}$ if and only if every subsequence $\{\mathbf{P}_{n_k}\}$ contains a subsequence $\{\mathbf{P}_{n_{k_j}}\}$ such that $\mathbf{P}_{n_{k_j}} \Rightarrow \mathbf{P}$.*

This is Theorem 2.3 from Billingsley [21].

Now we consider continuous mappings between metric spaces S_1 and S_2. A function $h : S_1 \to S_2$ is said to be measurable if

$$h^{-1}(\mathcal{B}(S_2)) \subset \mathcal{B}(S_1).$$

Let $h : S_1 \to S_2$ be a measurable function. Then every probability measure $\mathbf{P}$ on $(S_1, \mathcal{B}(S_1))$ induces a probability measure $\mathbf{P}h^{-1}$ on $(S_2, \mathcal{B}(S_2))$ defined by

$$(\mathbf{P}h^{-1})(A) = \mathbf{P} \circ h^{-1}(A) = \mathbf{P}(h^{-1}(A)),$$

where $A \in \mathcal{B}(S_2)$. This measure is uniquely determined. A function $h : S_1 \to S_2$ is continuous if for every open set $G_2 \subset S_2$ the set $h^{-1}(G_2)$ is open in S_1. Continuous mappings transport the property of weak convergence.

Theorem 3.3. *Let $h : S_1 \to S_2$ be a continuous function. If $\mathbf{P}_n \Rightarrow \mathbf{P}$, then also $\mathbf{P}_n h^{-1} \Rightarrow \mathbf{P}h^{-1}$.*

This theorem is a particular case of Theorem 5.1 from Billingsley [21].

A family $\{\mathbf{P}_n\}$ of probability measures on $(S, \mathcal{B}(S))$ is said to be relatively compact if every sequence of elements of $\{\mathbf{P}_n\}$ contains a weakly convergent subsequence. A family $\{\mathbf{P}_n\}$ is called tight if for arbitrary $\varepsilon > 0$ there exists a compact set K such that $\mathbf{P}(K) > 1 - \varepsilon$ for all $\mathbf{P}$ from $\{\mathbf{P}_n\}$. Prokhorov's theorem is a powerful tool in the theory of weak convergence of probability measures; it is given below as Theorem 3.4, the direct half, and as Theorem 3.5, the converse half. These theorems connect relative compactness with the tightness of a family of probability measures.

Theorem 3.4. *If a family of probability measures is tight, then it is relatively compact.*

Theorem 3.5. *Let S be separable (i.e., S contains a countable dense subset) and complete. If a family of probability measures on $(S, \mathcal{B}(S))$ is relatively compact, then it is tight.*

These are Theorems 6.1 and 6.2 from Billingsley [21]. Note that a topological space is said to be separable if it contains a countable dense subset.

In the theory of Dirichlet series we investigate the weak convergence of probability measures $\mathbf{P}_T \Rightarrow \mathbf{P}$, where T is a continuous parameter which tends to infinity. As it is noted in [21], we have $\mathbf{P}_T \Rightarrow \mathbf{P}$, as $T \to \infty$, if and only if $\mathbf{P}_{T_n} \Rightarrow \mathbf{P}$, as $n \to \infty$, for every sequence $\{T_n\}$ with $\lim_{n\to\infty} T_n = \infty$. All theorems on weak convergence analoguous to those stated above remain valid in the case of continuous parameters.

3.2 Random Elements

The theory of weak convergence of probability measures can be paraphrased as the theory of convergence of random elements in distribution.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(S, \mathcal{B}(S))$ be a metric space with its class of Borel sets $\mathcal{B}(S)$, and $X : \Omega \to S$ a mapping. If

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$$

for every $A \in \mathcal{B}(S)$, then X is called an S-valued random element defined on Ω; if $S = \mathbb{R}$ we say that X is a random variable. The distribution of an S-valued random element X is the probability measure $\mathbf{P}_X$ on $(S, \mathcal{B}(S))$, given by

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}\{\omega \in \Omega : X(\omega) \in A\}$$

for arbitrary $A \in \mathcal{B}(S)$ (in the sequel we will often write $\mathbf{P}$ in place of $\mathbf{P}_X$). We say that a sequence $\{X_n\}$ of random elements converges in distribution to a random element X if the distributions $\mathbf{P}_n$ of the elements X_n converge weakly to the distribution of the element X, and in this case we write

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X;$$

if $\mathbf{P}_n \Rightarrow \mathbf{P}_X$, then we also write $X_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbf{P}_X$.

Let S be a metric space with metric ϱ, and let X_n, Y_n be S-valued random elements defined on $(\Omega, \mathcal{F}, \mathbf{P})$. If X_n and Y_n have a common domain, it makes sense to speak of the distance $\varrho(X_n(\omega), Y_n(\omega))$ for $\omega \in \Omega$. If S is separable, then $\varrho(X_n, Y_n)$ is a random variable. In this case, convergence in distribution of two sequences of random elements X_n and Y_n is related to the distribution of $\varrho(X_n, Y_n)$ (convergence in probability).

Theorem 3.6. *Let S be separable and, for $n \in \mathbb{N}$, let $Y_n, X_{1n}, X_{2n}, \ldots$ be S-valued random elements, all defined on $(\Omega, \mathcal{F}, \mathbf{P})$. Suppose that*

$$X_{kn} \xrightarrow[n\to\infty]{\mathcal{D}} X_k \quad \text{for each } k, \quad \text{and} \quad X_k \xrightarrow[k\to\infty]{\mathcal{D}} X.$$

If for any $\epsilon > 0$

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbf{P}\{\varrho(X_{kn}, Y_n) \geq \epsilon\} = 0,$$

then $Y_n \xrightarrow[n\to\infty]{\mathcal{D}} X$.

This is Theorem 4.2 of Billingsley [21].

The mean (expectation value) $\mathbf{E}X$ of a random element X is defined by

$$\mathbf{E}X = \int_\Omega X(\omega)\,\mathrm{d}\mathbf{P}$$

if the integral exists in the sense of Lebesgue. A simple but fundamental result on the deviation of a random variable from its expectation value is Chebyshev's (respectively, Markov's) inequality:

Lemma 3.7. *Let X be a real-valued random variable, $h : \mathbb{R} \to [0, \infty)$ be a non-negative function, and $a > 0$. Then*

$$\mathbf{P}\{\omega \in \Omega : h(X) \geq a\} \leq \frac{1}{a}\,\mathbf{E}h(X).$$

A proof can be found, for example, in Billingsley [22]. Taking $h(x) = x^2$, we deduce the classical Chebyshev's inequality:

$$\mathbf{P}\{|X| \geq a\} \leq \frac{1}{a^2}\,\mathbf{E}X^2.$$

We say that some property is valid almost surely if there exists a set $A \in \mathcal{F}$ with $\mathbf{P}(A) = 0$ such that this property is valid for every $\omega \in \Omega \setminus A$. Random variables X and Y are said to be orthogonal if $\mathbf{E}XY = 0$. An important result on almost sure convergence of series of orthogonal random variables is the following

Theorem 3.8. *Assume that the random variables $X_1, X_2, \ldots$ are orthogonal and that*

$$\sum_{n=1}^{\infty} \mathbf{E}|X_n|^2 (\log n)^2 < \infty.$$

Then the series $\sum_{n=1}^{\infty} X_n$ converges almost surely.

Further, we need a similar result for independent random variables. Random variables X and Y are said to be independent if for all $A, B \in \mathcal{B}(S)$

$$\mathbf{P}\{\omega \in \Omega : X \in A, Y \in B\} = \mathbf{P}\{\omega \in \Omega : X \in A\} \cdot \mathbf{P}\{\omega \in \Omega : Y \in B\}.$$

Usually, independence of random variables is defined via the σ-fields generated by the related events; however, for the sake of simplicity, we introduced this notion by the equivalent condition above.

Theorem 3.9. *Assume that the random variables $X_1, X_2, \dots$ are independent. If the series*

$$\sum_{n=1}^{\infty} \mathbf{E}X_n \quad \text{and} \quad \sum_{n=1}^{\infty} \mathbf{E}(X_n - \mathbf{E}X_n)^2$$

converge, then the series $\sum_{n=1}^{\infty} X_n$ converges almost surely.

Proofs of the last two theorems can be found in Loève [226].

Now we present a criterion on almost sure convergence of series of independent random variables for Hilbert spaces. Let $\mathcal{H}$ be a separable Hilbert space with norm $\|\cdot\|$ and define for an $\mathcal{H}$-valued random element X and a real number c the truncated function

$$X^{(c)} = \begin{cases} X & \text{if} \quad \|X\| \leq c, \\ 0 & \text{if} \quad \|X\| > c. \end{cases}$$

Then

Theorem 3.10. *Let $X_1, X_2, \dots$ be independent $\mathcal{H}$-valued random elements. If there is a constant $c > 0$ so that the series*

$$\sum_{n=1}^{\infty} \mathbf{E}\|X_n^{(c)} - \mathbf{E}X_n^{(c)}\|^2, \quad \sum_{n=1}^{\infty} \mathbf{E}X_n^{(c)}, \quad \textit{and} \quad \sum_{n=1}^{\infty} \mathbf{P}\{\|X_n\| > c\}$$

converge, then the series $\sum_{n=1}^{\infty} X_n$ converges in $\mathcal{H}$ almost surely.

A proof can be found in Buldygin [45].

3.3 Denjoy's Probabilistic Argument for Riemann's Hypothesis

At this point our survey of probability theory will be cut in order to give a heuristic probabilistic argument for the truth of Riemann's hypothesis. The Möbius μ-function is defined by $\mu(1) = 1,\ \mu(n) = 0$ if n has a quadratic divisor $\neq 1$, and $\mu(n) = (-1)^r$ if n is the product of r distinct primes. It is easily seen that $\mu(n)$ is multiplicative and appears as coefficients of the Dirichlet series representation of the reciprocal of the zeta-function:

$$\zeta(s)^{-1} = \prod_p \left(1 - \frac{1}{p^s}\right) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s},$$

valid for $\sigma > 1$. Riemann's hypothesis is equivalent to the estimate

$$M(x) := \sum_{n \leq x} \mu(n) \ll x^{(1/2)+\epsilon}.$$

This is related to (1.11); for a proof see, for example, Titchmarsh [353, Sect. 14.25].

Denjoy [68] argued as follows. Assume that $\{X_n\}$ is a sequence of random variables with distribution

$$\mathbf{P}(X_n = +1) = \mathbf{P}(X_n = -1) = \frac{1}{2}.$$

Define

$$S_0 = 0 \quad \text{and} \quad S_n = \sum_{j=1}^{n} X_j,$$

then $\{S_n\}$ is a symmetrical random walk in $\mathbb{Z}^2$ with starting point at 0. A simple application of Chebyshev's inequality yields, for any positive c,

$$\mathbf{P}\{|S_n| \geq cn^{\frac{1}{2}}\} \leq \frac{1}{2c^2},$$

which shows that *large* values for S_n are *rare* events. By the theorem of Moivre–Laplace ([22, Theorem 27.1]) this can be made more precise. It follows that:

$$\lim_{n\to\infty} \mathbf{P}\left\{|S_n| < cn^{\frac{1}{2}}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x.$$

Since the right-hand side above tends to 1 as $c \to \infty$, we obtain

$$\lim_{n\to\infty} \mathbf{P}\left\{|S_n| \ll n^{\frac{1}{2}+\epsilon}\right\} = 1$$

for every $\epsilon > 0$. We observe that this might be regarded as a model for the value-distribution of Möbius μ-function. To say it with the words of Edwards: "Thus these probabilistic assumptions about the values of $\mu(n)$ lead to the conclusion, ludicrous as it seems, that $M(x) = \mathrm{O}(x^{1/2+\epsilon})$ with probability one and hence that the Riemann hypothesis is true with probability one!" (cf. [74]). The law of the iterated logarithm [22, Theorem 9.5] would even gives the stronger estimate

$$\lim_{n\to\infty} \mathbf{P}\left\{|S_n| \ll (n \log\log n)^{1/2}\right\} = 1,$$

which suggests for $M(x)$ the upper bound $(x \log\log x)^{1/2}$. This estimate is pretty close to the so-called weak Mertens hypothesis which states

$$\int_1^X \left(\frac{M(x)}{x}\right)^2 \mathrm{d}x \ll \log X.$$

Note that this bound implies the Riemann hypothesis and the essential simplicity hypothesis. On the contrary, Odlyzko and te Riele [284] disproved the original Mertens hypothesis [248],

$$|M(x)| < x^{\frac{1}{2}},$$

by showing that

$$\liminf_{x\to\infty} \frac{M(x)}{x^{1/2}} < -1.009 \quad \text{and} \quad \limsup_{x\to\infty} \frac{M(x)}{x^{1/2}} > 1.06; \tag{3.1}$$

for more details see also the notes to Sect. 14 in Titchmarsh [353].

3.4 Characteristic Functions and Fourier Transforms

There is an intimate relationship between weak convergence and characteristic functions which makes characteristic functions very useful in studying limit distributions.

The characteristic function $\varphi(\tau)$ of a probability measure $\mathbf{P}$ on $(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r))$ is defined by

$$\varphi(\tau) = \int_{\mathbb{R}^r} \exp(\mathrm{i}\langle \tau, x\rangle)\mathbf{P}(\,\mathrm{d}x),$$

where $\langle \tau, x\rangle$ stands for the inner product of τ and $x \in \mathbb{R}^r$. Notice that the characteristic function uniquely determines the measure it comes from. Let $\{\mathbf{P}_n\}$ be a sequence of probability measures on $(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r))$ and let $\{\varphi_n(\tau)\}$ be the sequence of the corresponding characteristic functions. Suppose that

$$\lim_{n\to\infty} \varphi_n(\tau) = \varphi(\tau)$$

for all τ, and that $\varphi(\tau)$ is continuous at the point $\mathbf{0} = (0, \ldots, 0)$. Then Lévy's famous continuity theorem (see [22, Sect. 26]) yields the existence of a probability measure $\mathbf{P}$ on $(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r))$ such that $\mathbf{P}_n \Rightarrow \mathbf{P}$, and $\varphi(\tau)$ is the characteristic function of $\mathbf{P}$. However, later we shall deal with Fourier transforms instead of characteristic functions; their theory is quite similar to the theory of characteristic functions (for details we refer to the first chapter from [186]).

Let $\gamma = \{s \in \mathbb{C} : |s| = 1\}$ and denote by γ^m the cartesian product of m copies of γ. Further, let $\mathbf{P}$ be a probability measure on $(\gamma^m, \mathcal{B}(\gamma^m))$, then the Fourier transform $g(k_1, \ldots, k_m)$ of the measure $\mathbf{P}$ is defined by

$$g(k_1, \ldots, k_m) = \int_{\gamma^m} x_1^{k_1} \ldots x_m^{k_m} \,\mathrm{d}\mathbf{P},$$

where $k_j \in \mathbb{Z}$ and $x_j \in \gamma$ for $1 \le j \le m$. Similarly to the characteristic function, the measure $\mathbf{P}$ is uniquely determined by its Fourier transform. The role of Lévy's continuity theorem in the theory of Fourier transforms is played by

Theorem 3.11. *Let* $\{\mathbf{P}_n\}$ *be a sequence of probability measures on* $(\gamma^m, \mathcal{B}(\gamma^m))$ *and let* $\{g_n(k_1, \ldots, k_m)\}$ *be the sequence of the corresponding Fourier transforms. Suppose that for every vector* $(k_1, \ldots, k_m) \in \mathbb{Z}^m$ *the limit*

$$g(k_1, \ldots, k_m) = \lim_{n\to\infty} g_n(k_1, \ldots, k_m)$$

exists. Then there is a probability measure $\mathbf{P}$ *on* $(\gamma^m, \mathcal{B}(\gamma^m))$ *such that* $\mathbf{P}_n \Rightarrow \mathbf{P}$*, and* $g(k_1, \ldots, k_m)$ *is the Fourier transform of* $\mathbf{P}$.

Theorem 3.11 is a special case of a more general continuity theorem for probability measures on compact abelian groups. A proof of this result can be found in Heyer [133, Theorem 1.4.2].

3.5 Haar Measure and Characters

Let G be a set equipped with the structures of a group and of a topological space. If the function $h : \mathsf{G} \times \mathsf{G} \to \mathsf{G}$, defined by $h(x, y) = xy^{-1}$, is continuous, then G is called a topological group. A topological group is said to be compact if its topology is compact. In what follows, G is assumed to be compact.

A Borel measure $\mathbf{P}$ on a compact topological group G is said to be invariant if

$$\mathbf{P}(A) = \mathbf{P}(xA) = \mathbf{P}(Ax)$$

for all $A \in \mathcal{B}(\mathsf{G})$ and all $x \in \mathsf{G}$, where xA and Ax denote the sets $\{xy : y \in A\}$ and $\{yx : y \in A\}$, respectively. An invariant Borel measure on a compact topological group is called Haar measure.

Theorem 3.12. *On every compact topological group there exists a unique probability Haar measure.*

The uniqueness follows from $\mathrm{m}(\mathsf{G}) = 1$. For the proof see Hewitt and Ross [132, Chap. IV] or Theorem 5.14 in Rudin [313].

In the sequel, we denote the Haar measure associated with a compact topological group G simply by m; there will be no confusion about the underlying group.

Now assume further that G is a commutative group. A continuous homomorphism $\chi : \mathsf{G} \to \mathbb{C}$ is called a character of G. The character of G which is identically 1 is called trivial or principal, and we denote it by χ_0; other characters are said to be non-trivial or non-principal. The characters build up a group $\hat{\mathsf{G}}$, the character group. The Fourier transform of a function f defined on G is given by

$$\hat{f}(\chi) = \int_{\mathsf{G}} \chi(g) f(g) \, \mathrm{m}(\,\mathrm{d}g),$$

where χ is a character of G. Then $\hat{f}$ is a continuous map defined on $\hat{\mathsf{G}}$. The orthogonality relation for characters states

$$\int_{\mathsf{G}} \chi(g) \, \mathrm{m}(\,\mathrm{d}g) = \begin{cases} 1 & \text{if} \quad \chi = \chi_0, \\ 0 & \text{if} \quad \chi \neq \chi_0. \end{cases} \tag{3.2}$$

This generalizes the concept of Dirichlet characters.

3.6 Random Processes and Ergodic Theory

To be able to identify later the explicit form of the limit measure in limit theorems, we recall some facts from ergodic theory.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\mathcal{T}$ denote a parameter set. A finite real function $X(\tau, \omega)$ with $\tau \in \mathcal{T}$ and $\omega \in \Omega$ is said to be a random (or stochastic) process if $\omega \mapsto X(\tau, \omega)$ is a random variable for each fixed $\tau \in \mathcal{T}$. For fixed $\omega \in \Omega$, the function $\tau \mapsto X(\tau, \omega)$ is called a sample path of the random process. Let $\tau_1, \ldots, \tau_n$ be an arbitrary set of values of $\mathcal{T}$. Then the family of all common distributions of random variables $X(\tau_1, \omega), \ldots, X(\tau_n, \omega)$, i.e.,

$$\mathbf{P}\{X(\tau_1, \omega) < x_1, \ldots, X(\tau_n, \omega) < x_n\}$$

for all $n \in \mathbb{N}$ and all possible values of τ_j with $1 \leq j \leq n$, is called a family of finite-dimensional distributions of the process $X(\tau, \omega)$. Part of the structure of the random process is specified by its finite-dimensional distributions. However, they do not determine the character of the sample paths (see [22, Sect. 23], for a nice example). Kolmogorov's existence theorem (see Theorem 36.1 in [22]) states that if a family of finite dimensional distributions satisfies certain consistency conditions, then there exists on some probability space a random process having exactly the same finite-dimensional distributions. For instance, a special application of Kolmogorov's existence theorem yields a model for Brownian motion with continuous paths.

Let Y be the space of all finite real-valued functions $y(\tau)$ with $\tau \in \mathbb{R}$. In this case it is known that the family of finite-dimensional distributions of each random process determines a probability measure $\mathbf{P}$ on $(Y, \mathcal{B}(Y))$. Then, on the probability space $(Y, \mathcal{B}(Y), \mathbf{P})$, we define for real u the translation g_u which maps each function $y(\tau) \in Y$ to $y(\tau + u)$. It is easily seen that the translations g_u form a group. A random process $X(\tau, \omega)$ is said to be strongly stationary if all its finite-dimensional distributions are invariant under the translations by u. It is known that if a process $X(\tau, \omega)$ is strongly stationary, then the translation g_u is measure preserving, i.e., for any set $A \in \mathcal{B}(Y)$ and all $u \in \mathbb{R}$ the equality

$$\mathbf{P}(A) = \mathbf{P}(A_u), \quad \text{where} \quad A_u := g_u(A)$$

holds. A set $A \in \mathcal{B}(Y)$ is called an invariant set of the process $X(\tau, \omega)$ if for each u the sets A and A_u differ from each other by a set of zero $\mathbf{P}$-measure. In other words, $\mathbf{P}(A \Delta A_u) = 0$, where Δ denotes the symmetric difference of two sets A and B:

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

It is easy to see that all invariant sets of Y form a σ-field which is a sub-σ-field of $\mathcal{B}(Y)$. We say that a strongly stationary process $X(\tau, \omega)$ is ergodic if its σ-field of invariant sets consists only of sets having $\mathbf{P}$-measure equal to 0 or 1. For ergodic processes the Birkhoff–Khintchine theorem gives an expression for the expectation of $X(0, \omega)$ in terms of an integral taken over the sample paths.

Theorem 3.13. *Let $X(\tau, \omega)$ be an ergodic process with $\mathbf{E}|X(\tau, \omega)| < \infty$ and almost surely Riemann-integrable sample paths over every finite interval. Then*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T X(\tau, \omega)\,\mathrm{d}\tau = \mathbf{E}X(0, \omega)$$

almost surely.

A proof of this theorem can be found in Cramér and Leadbetter [64].

3.7 The Space of Analytic Functions

Let $\mathcal{G}$ be a simply connected region in the complex plane. We denote by $\mathcal{H}(\mathcal{G})$ the space of analytic functions f defined on $\mathcal{G}$ equipped with the topology of uniform convergence on compacta.

In order to introduce an appropriate metric on $\mathcal{H}(\mathcal{G})$ we note

Lemma 3.14. *For any open set $\mathcal{G}$ in the complex plane there exists a sequence of compact subsets K_j of $\mathcal{G}$ with the properties:*

- *$K_j \subset K_{j+1}$ for any $j \in \mathbb{N}$;*
- *If K is compact and $K \subset \mathcal{G}$, then $K \subset K_j$ for some $j \in \mathbb{N}$;*

such that

$$\mathcal{G} = \bigcup_{j=1}^{\infty} K_j$$

The proof is straightforward and can be found in Conway's book [62, Sect. VII.1].

Now, for $f, g \in \mathcal{H}(\mathcal{G})$ let

$$\varrho_j(f, g) = \max_{s \in K_j} |f(s) - g(s)|$$

and put

$$\varrho(f, g) = \sum_{j=1}^{\infty} 2^{-j} \frac{\varrho_j(f, g)}{1 + \varrho_j(f, g)}.$$

This defines a metric on $\mathcal{H}(\mathcal{G})$ which induces the desired topology; of course, the metric ϱ depends on the family $\{K_j\}$. Note that the series above is dominated by $\sum_j 2^{-j}$ and therefore convergent.

Theorem 3.15. *Let $\mathcal{G}$ be a simply connected region in the complex plane. Then $\mathcal{H}(\mathcal{G})$ is a complete separable metric space.*

In Conway [62, Sect. VII.1], it is shown that $\mathcal{H}(\mathcal{G})$ is a complete metric space; the separability, i.e., the existence of a countable dense subset, follows from

Runge's approximation theorem (see [312, Sect. 13]) which states that the set of polynomials is dense in $\mathcal{H}(\mathcal{G})$.

In our later studies, we deal with the supports of $\mathcal{H}(\mathcal{G})$-valued random elements. Let S be a separable metric space and let $\mathbf{P}$ be a probability measure on $(S, \mathcal{B}(S))$. The minimal closed set $S_{\mathbf{P}} \subseteq S$ with $\mathbf{P}(S_{\mathbf{P}}) = 1$ is called the support of $\mathbf{P}$. Note that $S_{\mathbf{P}}$ consists of all $x \in S$ such that for every neighbourhood U of x the inequality $\mathbf{P}(U) > 0$ is satisfied. Let X be a S-valued random element defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then the support of the distribution $\mathbf{P}(X \in A)$ for $A \in \mathcal{B}(S)$ is called the support of the random element X. We denote the support of X by S_X.

Theorem 3.16. *Let $\{X_n\}$ be a sequence of independent $\mathcal{H}(\mathcal{G})$-valued random elements, and suppose that the series $\sum_{n=1}^{\infty} X_n$ converges almost everywhere. Then the support of the sum of this series is the closure of the set of all $f \in \mathcal{H}(\mathcal{G})$ which may be written as a convergent series*

$$f = \sum_{n=1}^{\infty} f_n, \quad \textit{where} \quad f_n \in S_{X_n}.$$

This is Theorem 1.7.10 of Laurinčikas [186]. The proof follows the lines of an analogous statement for independent real variables due to Lukacs [231]. We only sketch the main ideas. Suppose that the random elements X_n are defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P}^*)$. Put

$$X := \sum_{n=1}^{\infty} X_n = L_N + R_N,$$

where

$$L_N := \sum_{n=1}^{N} X_n \quad \text{and} \quad R_N := \sum_{n=N+1}^{\infty} X_n.$$

Since the series $\sum_{n=1}^{\infty} X_n$ converges almost surely, we have

$$\lim_{N \to \infty} \mathbf{P}^*\{\omega \in \Omega : \varrho(R_N, 0) \geq \epsilon\} = 0.$$

Let

$$\mathbf{P}_N(A) = \mathbf{P}^*\{L_N \in A\} \quad \text{and} \quad \mathbf{P}(A) = \mathbf{P}^*\{X \in A\}$$

for $A \in \mathcal{B}(\mathcal{H}(\mathcal{G}))$. It follows that $\mathbf{P}_n \Rightarrow \mathbf{P}$, which implies

$$S_{\mathbf{P}} \subset \lim S_{\mathbf{P}_N}, \tag{3.3}$$

where $\lim S_{\mathbf{P}_N}$ denotes the set of all $f \in \mathcal{H}(\mathcal{G})$ such that any neighbourhood of f contains at least one g which belongs to $S_{\mathbf{P}_n}$ for almost all $n \in \mathbb{N}$.

To show the converse inclusion, put

$$\mathbf{Q}_N(A) = \mathbf{P}^*\{R_N \in A\}$$

for $A \in \mathcal{B}(\mathcal{H}(\mathcal{G}))$. The distribution of $X = L_N + R_N$ is given by the convolution $\mathbf{P}_N * \mathbf{Q}_N$, defined by

$$(\mathbf{P}_N * \mathbf{Q}_N)(A) = \int_{\mathcal{H}(\mathcal{G})} \mathbf{P}_N(A - g)\mathbf{Q}_N(\,\mathrm{d}g).$$

The support of $X = L_N + R_N$ is the closure of the set

$$\{f \in \mathcal{H}(\mathcal{G}) : f = f_1 + f_2, \text{ where } f_1 \in S_{L_N}, f_2 \in S_{R_N}\}. \tag{3.4}$$

For $g \in \lim S_{\mathbf{P}_N}$ let

$$A_\epsilon := \{f \in \mathcal{H}(\mathcal{G}) : \varrho(f, g) < \epsilon\}.$$

It follows that $\mathbf{P}_N(A_\epsilon) = \mathbf{P}^*(L_N \in A_\epsilon) > 0$ and $\mathbf{Q}_N(A_\epsilon) > 0$ for N large enough. This leads to

$$\mathbf{P}(A_{2\epsilon}) \geq \mathbf{P}_N(A_\epsilon)\mathbf{Q}_N(A_\epsilon) > 0.$$

This implies

$$S_X = S_{\mathbf{P}} \supset \lim S_{\mathbf{P}_N}.$$

By (3.3) it follows that the latter inclusion is an equality. In view of (3.4) the support of L_N is the set of all $g \in \mathcal{H}(\mathcal{G})$ which have a representation

$$g = \sum_{n=1}^{N} f_n,$$

where $f_n \in S_{X_n}$. From the definition of $\lim S_{\mathbf{P}_N}$ we deduce that for any $f \in S_X$ there exists a sequence of $g_N \in S_{L_N}$ which converges to f. This yields the assertion of the theorem.