Math Semesterber (2021) 68:69–104
https://doi.org/10.1007/s00591-020-00287-z

MATHEMATIK IN FORSCHUNG UND ANWENDUNG

Statistical independence in mathematics–the key to a Gaussian law

Gunther Leobacher · Joscha Prochno

Received: 27 January 2020 / Accepted: 26 August 2020 / Published online: 1 October 2020

Abstract In this manuscript we discuss the notion of (statistical) independence embedded in its historical context. We focus in particular on its appearance and role in number theory, concomitantly exploring the intimate connection of independence and the famous Gaussian law of errors. As we shall see, this at times requires us to go adrift from the celebrated Kolmogorov axioms, which give the appearance of being ultimate ever since they have been introduced in the 1930s. While these insights are known to many a mathematician, we feel it is time for both a reminder and renewed awareness. Among other things, we present the independence of the coefficients in a binary expansion together with a central limit theorem for the sum-of-digits function as well as the independence of divisibility by primes and the resulting, famous central limit theorem of Paul Erdős and Mark Kac on the number of different prime factors of a number $n \in \mathbb{N}$. We shall also present some of the (modern) developments in the framework of lacunary series that have its origin in a work of Raphaël Salem and Antoni Zygmund.

Keywords Central limit theorem · Gaussian law · Independence · Relative measure · Lacunary series

Mathematical Subject Classification 60F05 · 60G50 · 42A55 · 42A61

G. Leobacher (✉) · J. Prochno
Institute of Mathematics & Scientific Computing, University of Graz, Graz, Austria
E-Mail: gunther.leobacher@uni-graz.at

J. Prochno
E-Mail: joscha.prochno@uni-graz.at

1 Introduction

One of the most famous graphs, not only among mathematicians and scientists, is the probability density function of the (standard) normal distribution (see Fig. 1), which has adorned the 10 Mark note of the former German currency for many years. Although already taking a central role in a work of Abraham de Moivre (26. May 1667 in Vitry-le-Francois; 27. November 1754 in London) from 1718, this curve only earned its enduring fame through the work of famous German mathematician Carl Friedrich Gauß (30. April 1777 in Braunschweig; 23. February 1855 in Göttingen), who used it in the approximation of orbits by ellipsoids when developing the least squares method, nowadays a standard approach in regression analysis. More precisely, Gauß conceived this method to master the random errors, i.e., those which fluctuate due to the unpredictability or uncertainty inherent in the measuring process, that occur when one tries to measure orbits of celestial bodies. The strength of this method became apparent when he used it to predict the future location of the newly discovered asteroid Ceres. Ever since, this curve seems to be the key to the mysterious world of chance and still the myth holds on that wherever this curve appears, randomness is at play.

With this article we seek to address mathematicians as well as a mathematically educated audience alike. One can say that the goal of this manuscript is 3-fold. First, for those less familiar with it we want to undo the fetters that connect chance and the Gaussian curve so onesidedly. Second, we want to recall the deep and intimate connection of the notion of statistical independence and the Gaussian law of errors beyond classical probability theory, which, thirdly, demonstrates that occasionally one is obliged to step aside from its seemingly ultimate form in terms of the Kolmogorov axioms and work with notions having its roots in earlier foundations of probability theory.

To achieve this goal we shall, partially embedded in a historic context, present and discuss several results from mathematics where, once an appropriate form of statistical independence has been established, the Gaussian curve emerges naturally. In more modern language this means that central limit theorems describe the fluctuations of mathematical quantities in different contexts. Our focus shall be on results that nowadays are considered to be part of probabilistic number theory. At the very heart of this development lies the true comprehension and appreciation of independence by Polish mathematician Mark Kac (3. August 1914 in Kremenez; 26. October 1984 in California). His pioneering works and insights, especially his

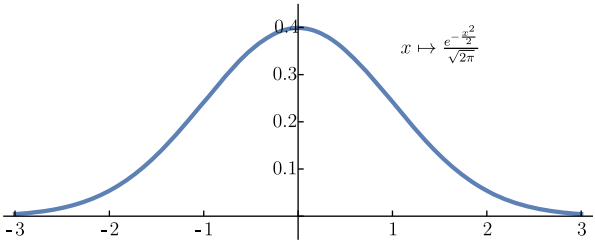

Fig. 1 The Gaussian curve

collaboration with Hugo Steinhaus (14. January 1887 in Jasło; 25. February 1972 in Wrocław) and famous mathematician Paul Erdős (26. March 1913 in Budapest; 20. September 1996 in Warsaw), have revolutionized our understanding and formed the development of probabilistic number theory for many years with lasting influence. We refer the reader to [10, 11, 49, 50] for general literature on the subject.

2 The classical central limit theorems and independence–a refresher

In this section we start with two fundamental results of probability theory and the notion of independence. These considerations form the starting point for future deliberations.

2.1 The notion of independence

Independence is one of the central notions in probability theory. It is hard to imagine today that this, for us so seemingly elementary and simple concept, has only been used vaguely and intuitively for hundreds of years without a formal definition underlying this notion. Implicitly this concept can be traced back to the works of Jakob Bernoulli (6. January 1655 in Basel; 16. August 1705 in Basel) and evolved in the capable hands of Abraham de Moivre. In his famous oeuvre "The Doctrine of Chances" [15] he wrote:

> "*...if a Fraction expresses the Probability of an Event, and another Fraction the Probability of another Event, and those two Events are independent; the Probability that both those Events will Happen, will be the Product of those Fractions.*"

It is to be noted that, even though this definition matches the modern one, neither the notion "Probability" nor "Event" had been introduced in an axiomatic way. It seems that the first formal definition of independence goes back to the year 1900 and the work [12] of German mathematician Georg Bohlmann (23. April 1869 in Berlin; 25. April 1928 in Berlin)[1]. In fact, long before Andrei Nikolajewitsch Kolmogorov (25. April 1903 in Tambow; 20. October 1987 in Moscow) proposed his axioms that today form the foundation of probability theory, Bohlmann had presented an axiomatization–but without asking for σ-additivity. For a detailed exposition of the historical development and the work of Bohlmann, we refer the reader to an article of Ulrich Krengel [37].

Roughly speaking, two events are considered to be independent if the occurrence of one does not affect the probability of occurrence of the other, see also Remark 3. We now continue with the formal definition of independence as it is used today. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space consisting of a non-empty set Ω (the sample space),

[1] It was decades later that Hugo Steinhaus and Mark Kac rediscovered this concept independently of the other mathematicians [34]. They were unaware of the previous works.

a σ-Algebra (the set of events) on Ω, and a probability measure $\mathbb{P} : \mathcal{A} \to [0,1]$. We then say that two events $A, B \in \mathcal{A}$ are *(statistically) independent* if and only if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

In other words two events are independent if their joint probability equals the product of their probabilities. This extends to any collection $(A_i)_{i \in I}$ of events, which is said to be independent if and only if for every $n \in \mathbb{N}$, $n \geq 2$, and all subsets $J \subseteq I$ of cardinality n,

$$\mathbb{P}\Big[\bigcap_{i \in J} A_i\Big] = \prod_{i \in J} \mathbb{P}[A_i].$$

It is important to note that in this case we ask for way more than just

$$\mathbb{P}\Big[\bigcap_{i \in I} A_i\Big] = \prod_{i \in I} \mathbb{P}[A_i].$$

and still much more than pairwise independence. Consequently, we also have to verify much more: the number of conditions to be verified to show that n given events are independent is exactly

$$\binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n} = 2^n - (n+1).$$

Having this notion of independence at hand, we define independent random variables. If $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are two random variables, then we say they are independent if and only if for all measurable subsets $A, B \subseteq \mathbb{R}$,

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B].$$

We use the standard notation $\{X \in A\}$ for $\{\omega \in \Omega : X(\omega) \in A\}$, $\mathbb{P}[X \in A]$ for $\mathbb{P}[\{X \in A\}]$, and $\mathbb{P}[X \in A, Y \in B]$ for $\mathbb{P}[\{X \in A\} \cap \{Y \in B\}]$.

This means that the random variables X and Y are independent if and only if for all measurable subsets $A, B \subseteq \mathbb{R}$ the events $\{X \in A\} \in \mathcal{A}$ and $\{Y \in B\} \in \mathcal{A}$ are independent. Again, a sequence $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ of random variables is said to be independent if and only if for every $n \in \mathbb{N}$, $n \geq 2$, any subset $I \subseteq \mathbb{N}$ of cardinality n, and all measurable sets $A_i \subseteq \mathbb{R}$, $i \in I$,

$$\mathbb{P}\Big[\bigcap_{i \in I} \{X_i \in A_i\}\Big] = \prod_{i \in I} \mathbb{P}[X_i \in A_i].$$

2.2 The central limit theorems of de Moivre-Laplace and Lindeberg

The history of the central limit theorem starts with the work of French mathematician Abraham de Moivre, who, around the year 1730, proved a central limit

theorem for standardized sums of independent random variables following a symmetric Bernoulli distribution [16].[2] It was not before 1812 that Pierre-Simon Laplace (28. March 1749 in Beaumont-en-Auge; 5. March 1827 in Paris) generalized this result to the asymmetric case [39]. However, a central limit theorem for standardized sums of independent random variables together with a rigorous proof only appeared much later in a work of Russian mathematician Alexander Michailowitsch Ljapunov (06. June 1857 in Jaroslawl; 03. November 1918 in Odessa) from 1901 [43]. Jarl Waldemar Lindeberg (04. August 1876 in Helsinki; 12. December 1932 Helsinki) published his works on the central limit theorem, in which he developed his famous and ingenious method of proof (today known as Lindeberg method), in 1922 [41, 42]. While in a certain sense elementary, this technique can be applied in various ways. A very nice exposition on Lindeberg's method can be found in the survey article [20] of Peter Eichelsbacher and Matthias Löwe. For an exhaustive presentation on the history of the central limit theorem we warmly recommend the monograph of Hans Fischer [24].

Let us start with the classical central limit theorem of de Moivre, hence restricting ourselves to the symmetric case $p = \frac{1}{2}$ in the Bernoulli distribution.

Theorem 1 (De Moivre, 1730) Let $X_1, X_2, X_3, \ldots$ be a sequence of independent random variables with a symmetric Bernoulli distribution. Then, for all $a, b \in \mathbb{R}$ with $a < b$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[a \leq \frac{\sum_{k=1}^{n} X_k - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \leq b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \, \mathrm{d}x \,.$$

The theorem of de Moivre, when discussed in school for instance, can be nicely depicted using the Galton Board (also known as bean machine). Let us consider the experiment of throwing an ideal and fair coin n-times (i.e., head shows up with probability $1/2$). The single throws are regarded to be independent as none of them influences the other. The number k of heads showing up in that experiment is a number between 0 and n. The probability that we see heads exactly k-times is described by a binomial distribution. Now de Moivre's theorem says that, for a large number n of tosses tending to infinity, the form of a suitably standardized histogramm approaches the Gaussian curve.

We have already mentioned at the beginning of this section that under suitable conditions a central limit theorem for general independent random variables may be obtained, not only those describing or modeling a coin toss.

We formulate Lindeberg's central limit theorem. In what follows, we shall denote by 1_A the indicator function of the set A, i.e., $1_A(x) \in \{0, 1\}$ with $1_A(x) = 1$ if and only if $x \in A$. The *expectation* of a random variable X with respect to the probability measure $\mathbb{P}$ is defined as $\mathbb{E}[X] := \int_\Omega X \mathrm{d}\mathbb{P}$, if this integral is defined. X

[2] A random variable X is Bernoulli distributed if and only if $\mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 1$. Here $p = \mathbb{P}(X = 1)$ is the parameter of the Bernoulli distribution and in the case where $p = \frac{1}{2}$, we call the distribution "symmetric". In his paper de Moivre did not call them Bernoulli random variables, but spoke of the probability distribution of the number of heads in coin toss.

is called *centered* if and only if $\mathbb{E}[X] = 0$. If $\mathbb{E}[|X|] < \infty$ we define the variance by $\mathrm{Var}[X] := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$.

Theorem 2 (Lindeberg CLT, 1922) Let $X_1, X_2, X_3, \ldots$ be a sequence of independent, centered, and square integrable random variables. Assume that for each $\varepsilon \in (0, \infty)$,

$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[X_k^2 \, 1_{\{|X_k| > \varepsilon s_n\}}] \overset{n\to\infty}{\longrightarrow} 0 \qquad \text{(Lindeberg condition)},$$

where $s_n^2 := \sum_{k=1}^{n} \mathrm{Var}[X_k]$. Then, for all $a, b \in \mathbb{R}$ with $a < b$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[a \le \frac{\sum_{k=1}^{n} X_k}{s_n} \le b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \, \mathrm{d}x \, .$$

Lindeberg's condition guarantees that no single random variable has too much influence. This immediately becomes apparent when looking at the Feller condition, which is implied by Lindeberg's condition. We refrain from discussing or presenting the details and refer again to [20].

Remark 1 Let us assume that the random variables in Theorem 2 are identically distributed and have variance $\mathrm{Var}[X_k] = \sigma^2 \in (0, \infty)$ for all $k \in \mathbb{N}$. Then Lindeberg's condition is automatically satisfied:

$$s_n^2 = \sum_{k=1}^{n} \mathrm{Var}[X_k] = n\sigma^2$$

and therefore, since the random variables X_k are identically distributed, we obtain for any $\varepsilon > 0$ that

$$\begin{aligned} L_n(\varepsilon) &= \frac{1}{n\sigma^2} \sum_{k=1}^{n} \mathbb{E}\Big[X_k^2 \, 1\!\!1_{\{|X_k| > \varepsilon s_n\}} \Big] = \frac{1}{n\sigma^2} \sum_{k=1}^{n} \mathbb{E}\Big[X_1^2 \, 1\!\!1_{\{|X_1| > \varepsilon \sqrt{n}\sigma\}} \Big] \\ &= \frac{1}{\sigma^2} \mathbb{E}\Big[X_1^2 \, 1\!\!1_{\{|X_1| > \varepsilon \sqrt{n}\sigma\}} \Big] \overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

where the convergence to 0 is a consequence of the Beppo Levi Theorem[3].

The previous remark immediately implies the classical central limit theorem for independent and identically distributed random variables.

[3] Which is a version of the monotone convergence theorem.

Corollary 1 *Let $X_1, X_2, X_3, ...$ be a sequence of independent and identically distributed random variables with $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}[X_1] = \sigma^2 \in (0, \infty)$. Then, for all $a, b \in \mathbb{R}$ with $a < b$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[a \leq \frac{\sum_{k=1}^{n} X_k}{\sqrt{n\sigma^2}} \leq b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, \mathrm{d}x\,.$$

One thing we immediately notice in the general version of Lindeberg's central limit theorem is the universality towards the underlying distribution of the random variables. Hence, the distribution seems to be irrelevant. On the other hand, in both the central limit theorem of de Moivre and the one of Lindeberg, we require the random variables to be independent. Could it be that independence is the key to a Gaussian law of errors? If so, does this connection go deeper and beyond a purely probabilistic framework? In the remaining parts of this work we want to get to the bottom of those questions.

2.3 Binary expansion and independence

In this section we will present a first example which a priori is non probabilistic. It has to do with intervals corresponding to binary expansions of real numbers $x \in [0,1]$ and a corresponding product rule for their lengths.

For simplicity, we start by reminding the reader of the decimal expansion of a number $x \in [0,1)$. One can prove that each number $x \in [0,1)$ has a non-terminating and unique decimal expansion (see, e.g., [8]). For example,

$$\frac{2}{7} = 0.285714285714...$$

and this expression is merely a short way for writing

$$\frac{2}{7} = \frac{2}{10} + \frac{8}{10^2} + \frac{5}{10^3} + \frac{7}{10^4} + \ldots.$$

Generally, for each $x \in [0,1)$ there exist unique numbers $d_1(x), d_2(x), d_3(x), ...$ in $\{0, 1, ..., 9\}$ such that

$$x = \frac{d_1(x)}{10} + \frac{d_2(x)}{10^2} + \frac{d_2(x)}{10^3} + \ldots.$$

Analogous to the decimal expansion, each number $x \in [0,1)$ has a binary expansion (also known as dyadic expansion), i.e., there are unique numbers $b_1(x), b_2(x), b_3(x), ...$ in the set $\{0,1\}$ such that

$$x = \frac{b_1(x)}{2} + \frac{b_2(x)}{2^2} + \frac{b_3(x)}{2^3} + \ldots. \tag{1}$$

For instance, we can write

$$\frac{2}{7} = \frac{0}{2} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{0}{2^4} + \frac{1}{2^5} + \frac{0}{2^6} + \ldots.$$

To guarantee uniqueness in the expansion, we agree to write the expansion in such a way that infinitely many of the binary digits are zero. As already indicated by the way we write it, the binary digits are functions in the variable we denoted by x, i.e.,

$$b_k : [0,1) \to \{0,1\}, \qquad x \mapsto b_k(x).$$

Sometimes these functions are called Rademacher functions, although Hans Rademacher (3. April 1892 in Wandsbek; 7. February 1969 in Haverford) defined a slightly different version [45]. The value that b_k takes at x not only provides information about the k-th binary digit of x, but also about x itself. Obviously, if $b_1(x) = 1$, then $x \in [1/2,1)$ or if $b_2(x) = 0$, then $x \in [0,1/4) \cup [1/2,3/4)$. More generally, if we define for each $k \in \mathbb{N}$ the set

$$B_k := \bigcup_{j=1}^{2^{k-1}} \Big[\frac{2j-2}{2^k}, \frac{2j-1}{2^k}\Big),$$

then

$$b_k(x) = 1_{[0,1)\setminus B_k}(x) = \begin{cases} 0 & : x \in B_k \\ 1 & : x \in [0,1) \setminus B_k \,. \end{cases}$$

These considerations yield the following: if $n \in \mathbb{N}$, $k_1, ..., k_n \in \mathbb{N}$, and $\varepsilon_1, ...\varepsilon_n \in \{0,1\}$, then

$$\begin{aligned} \lambda\left(\bigcap_{i=1}^{n} b_{k_i}^{-1}(\varepsilon_i)\right) &= \lambda\big(\{x \in [0,1) : b_{k_1} = \varepsilon_1, ..., b_{k_n} = \varepsilon_n\}\big) \\ &= \Big(\frac{1}{2}\Big)^n = \prod_{i=1}^{n} \lambda\big(\{x \in [0,1) : b_{k_i} = \varepsilon_i\}\big), \end{aligned}$$

where λ denotes the 1-dimensional Lebesgue measure (which in this case simply assigns the length to an interval). This implies that the binary coefficients as functions in $x \in [0,1)$, are independent; a result seemingly discovered by French mathematician Émile Borel (7. January 1871 in Saint-Affrique; 3. February 1956 in Paris) in 1909 [13]. In particular, the random variables $X_k = b_k$ satisfy the assumptions of de Moivre's theorem (Theorem 1) and so we obtain a central limit theorem for binary expansions b_k. Probability in the sense of coin tosses or events has not played any role in our arguments. (Nevertheless, technically the X_k's are bona-fide random variables on the probability space $\big([0,1), \mathcal{B}([0,1)), \lambda\big)$.)

2.4 Prime factors and independence

We shall now consider a fundamentally different example of independence in mathematics. Take a sufficiently large natural number $N \in \mathbb{N}$. We note that roughly half of the numbers between 1 and N are divisible by the prime number 2, namely $2, 4, 6$

and so on. In the same way, roughly one third of the numbers between 1 and N are divisible by the prime number 3, namely $3, 6, 9$ and so on. If we now consider the numbers between 1 and N which are divisible by 6, then this is again roughly one sixth. However, divisibility by 6 is equivalent to both divisibility by 2 *and* 3 and we can write this as

$$\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$$

for the corresponding fractions of numbers between 1 and N. But this reminds us of the multiplication of probabilities–as occurring in the concept of independence! Of course, the same argument applies for divisibility by general distinct primes p and q as well as by any finite number of primes. We can say, in this sense, that divisibility of a number by distinct primes is independent.

Apparently, every second natural number is divisible by 2, so that the numbers with this property constitute one half of all natural numbers. One could thus think that a randomly chosen natural number is divisible by 2 with probability $\frac{1}{2}$. In the same way, this number would be divisible by 3 with probability $\frac{1}{3}$, and an analog statement would hold for divisibility by every natural number.

It turns out that this notion, although intuitive, is incompatible with Kolmogorov's concept of probability in that no probability measure on the naturals with the above property exists (And, as a consequence, it is impossible to define a uniform measure on any countably infinite set).

To see this, define, for every pair of numbers n, k with $n \in \mathbb{N}$ and $k \in \{1, ..., n\}$ the set $A_{n,k} := \{jn + k : j \in \mathbb{N} \cup \{0\}\}$. For $k \neq n$, $A_{n,k}$ consists of all natural numbers which yield remainder k after division by n, while for $k = n$ we have $A_{n,k} = A_{n,n}$, which is the set of all natural numbers that are divisible by n. We denote by $\mathcal{P}(\mathbb{N})$ the set of all subsets of $\mathbb{N}$.

Lemma 1 *Let μ be a finite measure on the set $\mathcal{P}(\mathbb{N})$, which satisfies*

$$\mu(A_{p,k}) = \mu(A_{p,p}) \tag{2}$$

for every prime number p and every $k \in \{1, ..., p\}$. Then $\mu(\{m\}) = 0$ for every $m \in \mathbb{N}$, and therefore $\mu(A) = 0$ for all $A \subseteq \mathbb{N}$.

Proof First note that (2) implies $\mu(A_{p,k}) = \mu(\mathbb{N})/p$ for every prime number p and all $k \in \{1, ..., p\}$: indeed, if p is a prime number, then

$$\mu(\mathbb{N}) = \mu\Big(\bigcup_{k\in\{1,...,p\}} A_{p,k}\Big) = \sum_{k\in\{1,...,p\}} \mu(A_{p,k}) \overset{(2)}{=} p\mu(A_{p,p}),$$

where we used the finite additivity of μ to obtain the second equality. Combining $\mu(\mathbb{N}) = p\mu(A_{p,p})$ with (2) gives $\mu(A_{p,k}) = \mu(\mathbb{N})/p$ for all $k \in \{1, ..., p\}$.

Now fix $m \in \mathbb{N}$. For every prime number p there exist numbers $j \in \mathbb{N} \cup \{0\}$ and $k \in \{1, ..., p\}$ such that $m = jp+k$. Thus $m \in A_{p,k}$. From our earlier considerations it follows

$$\mu(\{m\}) \leq \mu(A_{p,k}) = \mu(\mathbb{N})/p\,.$$

Since $\mu(\mathbb{N}) < \infty$ by assumption, and since there are arbitrarily large primes, it follows that $\mu(\{m\}) = 0$. But since μ is a measure, and thus is σ-additive, we get $\mu(A) = \sum_{m\in A}\mu(\{m\}) = 0$ for every $A \subseteq \mathbb{N}$. □

So there exists no measure on $\mathcal{P}(\mathbb{N})$ having the desired property (2). But could it be that we have chosen the domain of μ too large? The next proposition shows that there is no smaller domain containing all $A_{p,k}$.

Proposition 1 *We have* $\sigma\big(\{A_{p,k} : p \text{ prime}, k \in \{1, ..., p\}\}\big) = \mathcal{P}(\mathbb{N})$.

Proof We define the set $\Sigma := \sigma\big(\{A_{p,k} : p \text{ prime}, k \in \{1, ..., p\}\}\big)$. It is sufficient to show that $\{m\} \in \Sigma$ for all $m \in \mathbb{N}$. To this end fix $m \in \mathbb{N}$. For every prime $p > m$ we have $m \in A_{p,m}$, since $m = 0 \cdot p + m$. Therefore,

$$m \in \bigcap_{p \text{ prime}, p>m} A_{p,m}\,.$$

Let $\ell \in \bigcap_{p \text{ prime}, p>m} A_{p,m}$. Then there exists a prime p with $p > \ell$ so that, since $\ell \in A_{p,m}$, $\ell = 0 \cdot p + m = m$. Thus, $\{m\} = \bigcap_{p \text{ prime}, p>m} A_{p,m} \in \Sigma$. □

Remark 2 Eq. (2) in Lemma 1 formalizes our earlier intuition that if $\mu(A_{p,p})$ is the fraction of numbers divisible by p then this should equal the fraction of numbers giving remainder 1 and so on. The Lemma shows us that there cannot be a non-trivial finite measure μ with this property and therefore we cannot assign meaningful probabilities to those subsets in the framework of Kolmogorov's theory. In contrast to the independence of distinct binary digits of a number in [0,1), we cannot cover the independence of divisibility by distinct primes of a number in $\mathbb{N}$ using Kolmogorov's notion of independence of random variables.

3 Relative Measures

A possible remedy is a notion related to one of the earlier approaches to probability theory going back at least to Richard von Mises (19. April 1883 in Lviv; 14. Juli 1953 in Boston) and can be found in early work of Kac and Steinhaus. However, we were unable to trace the original source. In any case, this approach has to a large extent been replaced by Kolmogorov's axiomatization of probability.

One of the central notions in this manuscript shall be referred to as relative measure and its definition and properties be discussed in the following section.

3.1 Relative measurable subsets of $\mathbb{N}$

Definition 1 (Relatively measurable subsets of $\mathbb{N}$ and relative measure) We say that a subset $A \subseteq \mathbb{N}$ is *relatively measurable* if and only if the limit

$$\lim_{N\to\infty} \frac{|A \cap \{1, ..., N\}|}{N},$$

exists. In that case we define the *relative measure* μ_R of A as exactly this limit,

$$\mu_R(A) := \lim_{N\to\infty} \frac{|A \cap \{1, ..., N\}|}{N}.$$

It is easy to see that the collection of relatively measurable subsets of $\mathbb{N}$ forms an algebra and that μ_R is a non-negative and (finitely-)additive set function on it. Moreover, it is obvious that every finite subset of $\mathbb{N}$ is relatively measurable with relative measure 0.

The sets $A_{n,k}$, $n \in \mathbb{N}$ and $k \in \{0, ..., n-1\}$ defined in Sect. 2.4 are relatively measurable with

$$\mu_R(A_{n,k}) = \frac{1}{n}.$$

It is a direct consequence of Lemma 1 that μ_R cannot be σ-additive. Indeed,

$$\mu_R\Big(\bigcup_{i\in\mathbb{N}} \{i\}\Big) = \mu_R(\mathbb{N}) = 1 \neq 0 = \sum_{i\in\mathbb{N}} \mu_R(\{i\}).$$

On the other hand, we can construct sets which are *not* relatively measurable.

Example 1 Let $a_1 = 0$ and define

$$a_k := \begin{cases} 0 & : 2^{2m} < k \leq 2^{2m+1} \text{ for some } m \in \mathbb{N}_0 \\ 1 & : 2^{2m+1} < k \leq 2^{2m+2} \text{ for some } m \in \mathbb{N}_0. \end{cases}$$

Consider the level set $A := \{k \in \mathbb{N} : a_k = 1\}$. Then A is not relatively measurable because

$$2^{-(2m+2)}|A \cap \{1, ..., 2^{2m+2}\}| = 2^{-(2m+2)}2(1 + 2^2 + ... + 2^{2m+1}) =$$
$$2^{-(2m+1)}\frac{2^{2m+2} - 1}{3} \to \frac{2}{3}$$
$$2^{-(2m+1)}|A \cap \{1, ..., 2^{2m+1}\}| = 2^{-(2m+1)}2(1 + 2^2 + ... + 2^{2m-1}) =$$
$$2^{-(2m)}\frac{2^{2m} - 1}{3} \to \frac{1}{3}.$$

The relative measure allows us to conceive and show the independence of divisibility by different primes in a formal way. In this regard this notion is superior to a measure in the sense of Kolmogorov. We are now going to prove the indepen-

dence of $A_{p,p}$ and $A_{q,q}$ for different primes p and q. By the fundamental theorem of arithmetic a number is divisible by p as well as q if and only if it is divisible by their product pq, and so $A_{p,p} \cap A_{q,q} = A_{pq,pq}$. Therefore, we obtain

$$\mu_R(A_{p,p} \cap A_{q,q}) = \mu_R(A_{pq,pq}) = \frac{1}{p \cdot q} = \frac{1}{p} \cdot \frac{1}{q} = \mu_R(A_{p,p})\,\mu_R(A_{q,q}),$$

which is the product rule so characteristic for independence. Similarly, one can show this property for each finite collection of different primes $p_1, ..., p_m$.

The following lemma shows that if the indicator function of a subset of the natural numbers is eventually periodic, then the relative measure of that set is equal to the average over the period. We shall leave the proof to the reader.

Lemma 2 *Consider a set $A \subseteq \mathbb{N}$. If there exist $k \in \mathbb{N}$ and $n_0 \in \mathbb{N}$ such that*

$$\forall n \geq n_0 : 1_A(n+k) = 1_A(n),$$

then A is relatively measurable and

$$\mu_R(A) = \frac{|A \cap \{n_0+1, ..., n_0+k\}|}{k}.$$

Remark 3 (Independence and information) One important property of statistical independence is that knowledge of one event, say B, does not present any information about an independent event A: for independent A, B we have $\mathbb{P}(A|B) = \mathbb{P}(A)$.

A similar situation occurs with numbers: knowledge about divisibility by one prime does not tell us anything about divisibility by another one. This holds also true for the digits considered earlier: if we know the k-th digit of a number $x \in [0,1)$ this does not tell us anything about its ℓ-th digit.

Consider now, for every $j \in \mathbb{N}$ the function $\beta_j : \mathbb{N} \to \{0,1\}$ defined by

$$\beta_j(n) := \begin{cases} 0 & : \lfloor \frac{n}{2^{j-1}} \rfloor \text{ is even} \\ 1 & : \lfloor \frac{n}{2^{j-1}} \rfloor \text{ is odd} \end{cases}, \tag{3}$$

such that $\beta_j(n)$ is the j-th binary digit of n, and

$$n = \sum_{j=1}^{\infty} \beta_j(n)\, 2^{j-1} = \sum_{j=1}^{\lfloor \log_2(n) \rfloor + 1} \beta_j(n)\, 2^{j-1}.$$

To every $j \in \mathbb{N}$ assign the set $B_j := \{n \in \mathbb{N} : \beta_j(n) = 1\}$, i.e. the set of all natural numbers for which the j-th binary digit equals 1.

It follows from the definition of binary digits that for each $j \in \mathbb{N}$

$$B_j = \bigcup_{m \in \mathbb{N} \cup \{0\}} \{2^{j-1}(2m+1), ..., 2^{j-1}(2m+1) + 2^{j-1} - 1\},$$

which means that

$$B_j^c = \bigcup_{m\in\mathbb{N}\cup\{0\}} \{2^j m, ..., 2^j m + 2^{j-1} - 1\},$$

and so $\mu_R(B_j) = \frac{1}{2}$. Moreover, for every choice of $j, k \in \mathbb{N}$ with $j < k$, we have $\mu_R(B_j \cap B_k) = \mu_R(B_j)\mu_R(B_k)$, which can be proven using Lemma 2.

Definition 2 Let $(A_j)_{j\in J}$ be a family of relatively measurable subsets of $\mathbb{N}$. We say that $(A_j)_{j\in J}$ are *independent* if and only if for every $m \in \mathbb{N}$ and every subset I of cardinality m

$$\mu_R\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I} \mu_R(A_i).$$

Summarizing the preceding thoughts, we obtain the following result.

Proposition 2

1. *For* $n \in \mathbb{N}$ *and* $k \in \{1, ..., n\}$, *let* $A_{n,k} := \{jn + k : j \in \mathbb{N} \cup \{0\}\}$. *Then the family* $\left(A_{p,p}\right)_{p\in\mathbb{N}, p \text{ prime}}$ *is independent.*
2. *For every* $j \in \mathbb{N}$ *let* $B_j = \bigcup_{m\in\mathbb{N}\cup\{0\}} \{2^{j-1}(2m + 1), ..., 2^{j-1}(2m + 1) + 2^{j-1}-1\}$. *Then the family* $\left(B_j\right)_{j\in\mathbb{N}}$ *is independent.*

It is quite interesting that similar results to the ones for expansions of real numbers in [0,1) with respect to the Lebesgue measure can be obtained for the expansion of natural numbers with respect to the relative measure on $\mathbb{N}$.

3.2 Relatively measurable sequences and their distribution

In this subsection we shall introduce the notion of a relatively measurable sequence and, in broad similarity to the way independence is defined in the sense of Kolmogorov, we introduce the notion of relatively independent sequences $x, y : \mathbb{N} \to \mathbb{R}$ and define a distribution function with respect to relative measures. As we shall see, such a distribution function does not possess all the properties that—coming from probability theory—we might expect it to have.

Definition 3 (Relatively measurable sequence) A sequence $x : \mathbb{N} \to \mathbb{R}$ is said to be *relatively measurable* if and only if the pre-image

$$x^{-1}(I) := \{n \in \mathbb{N} : x_n \in I\}$$

of each interval $I \subseteq \mathbb{R}$ under x is a relatively measurable subset of $\mathbb{N}$.

To us an interval means a convex subset of $\mathbb{R}$, in particular singleton sets are intervals. Natural examples of measurable sequences are indicator functions of relatively measurable sets and their finite sums.

We shall now introduce what it means for two sequences to be independent with respect to a relative measure. This is again done via a product rule.

Definition 4 (Independent sequences) Two relatively measurable sequences $x, y : \mathbb{N} \to \mathbb{R}$ are said to be μ_R-independent if and only if for any two intervals $I, J \subseteq \mathbb{R}$ we have

$$\mu_R\big(x^{-1}(I) \cap y^{-1}(J)\big) = \mu_R\big(x^{-1}(I)\big)\,\mu_R\big(y^{-1}(J)\big)\,.$$

This definition can be generalized in an obvious way to any finite number of relatively measurable sequences.

We now turn to the definition of a (relative) distribution function of a relatively measurable sequence.

Definition 5 (Distribution function) Let $x : \mathbb{N} \to \mathbb{R}$ be a relatively measurable sequence. Then the function

$$F_x : \mathbb{R} \to [0,1], \quad F_x(z) := \mu_R\Big(\big\{n \in \mathbb{N} : x_n \in (-\infty, z]\big\}\Big)$$

is called the *(relative) distribution function* of x.

By its very definition such a distribution function resembles a classical distribution function we know from probability theory. In particular, it is immediately clear that it is non-decreasing. However, in general not all properties we may expect from a relative distribution function have to hold.

Example 2 Consider the sequence $x : \mathbb{N} \to \mathbb{R}$ given by

$$x_n := \begin{cases} -n & \text{if } n = 4k \text{ for some } k \in \mathbb{N}_0 \\ 0 & \text{if } n = 4k+1 \text{ for some } k \in \mathbb{N}_0 \\ \frac{1}{n} & \text{if } n = 4k+2 \text{ for some } k \in \mathbb{N}_0 \\ n & \text{if } n = 4k+3 \text{ for some } k \in \mathbb{N}_0\,. \end{cases}$$

Then it is easy to see that x is relatively measurable and that its relative distribution function is given by

$$F_x(z) = \frac{1}{4}1_{(-\infty,0)}(z) + \frac{2}{4}1_{\{0\}}(z) + \frac{3}{4}1_{(0,\infty)}(z)\,.$$

Hence, F_x is neither left nor right continuous, and we have

$$\lim_{z\to-\infty} F_x(z) > 0 \qquad \text{and} \qquad \lim_{z\to-\infty} F_x(z) < 1\,.$$

Note however that for every bounded relatively measurable sequence x

$$\lim_{z\to-\infty} F_x(z) = 0 \qquad \text{and} \qquad \lim_{z\to-\infty} F_x(z) = 1\,.$$

Next we introduce and study the notion of an average of a relatively measurable sequence.

Definition 6 (Relative average) Let $x : \mathbb{N} \to \mathbb{R}$ be a relatively measurable sequence. Then we define the *relative average* of x by

$$M(x) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n \,,$$

whenever this limit exists.

The following theorem shows that the relative average of a relatively measurable and bounded sequence can be written in terms of a Stieltjes integral with respect to the relative distribution function.

Theorem 1 Let $x : \mathbb{N} \to \mathbb{R}$ be a relatively measurable and bounded sequence. Then $M(x)$ exists and

$$M(x) = \int_{-\infty}^{\infty} z \, \mathrm{d}F_x(z) \,. \tag{4}$$

Proof In this proof we simply write F instead of F_x. By assumption there exists some $K \in (0, \infty)$ such that $-K+1 \le x_n \le K$ for every $n \in \mathbb{N}$. The Stieltjes integral exists since the function $\mathrm{id} : [-K, K] \to \mathbb{R}, z \mapsto z$ is continuous and F is monotone on $[-K, K]$ and constant on the intervals $(-\infty, -K]$ and $[K, \infty)$. Therefore, given $\varepsilon > 0$ there exists a decomposition $Z = \{-K = t_0 < t_1 < \ldots < t_m = K\}$ of $[-K, K]$ such that $O(\mathrm{id}, F, Z) - U(\mathrm{id}, F, Z) < \varepsilon$, where U and O denote upper and lower Riemann-Stieltjes sums, i.e.,

$$U(\mathrm{id}, F, Z) = \sum_{k=1}^{m} t_{k-1}\big(F(t_k) - F(t_{k-1})\big) \quad \text{and}$$
$$O(\mathrm{id}, F, Z) = \sum_{k=1}^{m} t_k\big(F(t_k) - F(t_{k-1})\big) \,.$$

We observe that

$$\begin{aligned}
\mathrm{limsup}_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n &= \sum_{k=1}^{m} \mathrm{limsup}_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{(t_{k-1},t_k]}(x_n) x_n \\
&\le \sum_{k=1}^{m} \mathrm{limsup}_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{(t_{k-1},t_k]}(x_n) t_k \\
&\le \sum_{k=1}^{m} t_k\big(F(t_k) - F(t_{k-1})\big) = O(\mathrm{id}, F, Z) \,.
\end{aligned}$$

Similarly one can show that $\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}x_n \geq U(\mathrm{id}, F, Z)$, which then proves the assertion. □

It follows from the properties of Riemann-Stieltjes integrals that

$$M(x) = \int_{-\infty}^{\infty} z\, F_x'(z)\,\mathrm{d}z\,, \tag{5}$$

whenever F_x is differentiable on $\mathbb{R}$ with $F_x' = f_x$ outside some at most finite subset of $\mathbb{R}$.

Remark 4 We see that measurable sequences behave in many ways like random variables, and indeed a measurable sequence can be taken as a mathematical model for a "random number". As noted before, this kind of model has been put forward by Austrian mathematician Richard von Mises in the first half of the 20th century. This model was–at least among the vast majority of probabilists–replaced by Kolmogorov's approach, mainly because of the potent tools from Lebesgue's measure theory and the accompanied clean and simple concepts and theorems of convergence.

Nevertheless there is a certain appeal to the alternative, in particular its slim theoretical foundation. Within this approach one can simply state that a *real* number is a Cauchy sequence of rational numbers and a *random* number is a relatively measurable sequence of real numbers.

We now assign to every $\mathbb{Z}$-valued and relatively measurable sequence x a function $\rho_x : \mathbb{Z} \to [0,1]$ via

$$\rho_x(k) := \mu_R\big(\{n \in \mathbb{N} : x_n = k\}\big)\,.$$

Then for bounded, $\mathbb{Z}$-valued and relatively measurable sequences we have $\sum_{k\in\mathbb{Z}}\rho_x(k) = 1$ and the well-known convolution formula:

Proposition 3 *Let $x, y : \mathbb{N} \to \mathbb{R}$ be bounded and relatively measurable sequences taking values in $\mathbb{Z}$. If x and y are μ_R-independent, then $\rho_{x+y} = \rho_x * \rho_y$, where*

$$\rho_x * \rho_y(k) := \sum_{j\in\mathbb{Z}} \rho_x(j)\rho_y(k-j)\,, \quad k \in \mathbb{Z}\,.$$

All in all, we can say that relatively measurable sequences behave in many ways like random variables. For instance, the indicator functions of the sets B_j introduced after Remark 3 form an independent, relatively measurable, bounded, and $\mathbb{Z}$-valued sequence. Therefore, their sums satisfy

$$\rho_{1_{B_1}+\ldots+1_{B_m}}(k) = \binom{m}{k} 2^{-m}\,, \quad k \in \mathbb{Z}\,.$$

This means that the partial sums of the indicator functions of the sets B_j satisfy the central limit theorem of de Moivre (Theorem 1), i.e., for any $a, b \in \mathbb{R}$ with $a < b$,

$$\begin{aligned} &\lim_{m\to\infty} \mu_R\left(\left\{n \in \mathbb{N} : a \le \frac{\sum_{j=1}^m 1_{B_j}(n) - \frac{m}{2}}{\sqrt{\frac{m}{4}}} \le b\right\}\right) \\ &= \lim_{m\to\infty} \sum_{k=0}^{m} \binom{m}{k} 2^{-m} 1_{[a,b]}\left(\frac{k-\frac{m}{2}}{\sqrt{\frac{m}{4}}}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, \mathrm{d}x\,. \end{aligned} \tag{6}$$

Note again that the set considered above is indeed relatively measurable. To see this, we note that, as was argued before, the sets B_j are all relatively measurable and hence, because the collection of relatively measurable sets forms an algebra, so are their complements B_j^c. This immediately implies that the sequences $(1_{B_j}(n))_{n\in\mathbb{N}}$ and their finite sums are relatively measurable.

Thus, for the binary expansion of natural numbers we have the same central limit theorem as for the binary expansion of real numbers in [0,1). In fact, we can now formulate a quite interesting version of this, which can be found, for example, in [18]. Contrary to almost all numbers in [0,1), every natural number has a finite expansion and hence it is reasonable to define for $n \in \mathbb{N}$ its sum-of-digits function with respect to the binary expansion,

$$s_2(n) := \sum_{j=1}^{\lfloor \log_2(n) \rfloor + 1} 1_{B_j}(n) = \sum_{j=1}^{\infty} 1_{B_j}(n)\,, \quad n \in \mathbb{N}\,.$$

The following result describes the Gaussian fluctuations of the sum-of-digits function.

Theorem 2 (Central limit theorem for the sum-of-digits function) For all $b \in \mathbb{R}$, we have

$$\mu_R\left(\left\{n \in \mathbb{N} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{x^2}{2}}\, \mathrm{d}x\,.$$

We recall the following lemma from probability theory.

Lemma 3 *Let $F : \mathbb{R} \to [0,1]$ be a continuous cumulative distribution function and let $(F_n)_{n\in\mathbb{N}}$ be a sequence of non-decreasing functions $F_n : \mathbb{R} \to [0,1]$ with $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$. Then $F_n \to F$ uniformly on $\mathbb{R}$.*

We are now able to prove the central limit theorem for the sum-of-digits function.

Proof (Proof of Theorem 2) Let $\varepsilon \in (0,\infty)$. For $b \in \mathbb{R}$ let us write $\Phi(b) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b} e^{-\frac{x^2}{2}}\,\mathrm{d}x$. It follows from de Moivre's central limit theorem (see Eq. (6)) and Lemma 3 that there exists $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$ and every $b \in \mathbb{R}$,

$$-\frac{\varepsilon}{6} < \mu_R\left(\left\{n \in \mathbb{N} : \frac{\sum_{j=1}^{m} 1_{B_j}(n) - \frac{m}{2}}{\sqrt{\frac{m}{4}}} \leq b\right\}\right) - \Phi(b) < \frac{\varepsilon}{6}.$$

Moreover, for each $m \geq m_0$, we have

$$\left|\left\{0 \leq n < 2^m : \sum_{j=1}^{m} s_2(n) = k\right\}\right| = \left|\left\{0 \leq n < 2^m : \sum_{j=1}^{m} 1_{B_j}(n) = k\right\}\right| = \binom{m}{k}$$

and therefore,

$$2^{-m}\left|\left\{0 \leq n < 2^m : s_2(n) \leq b\sqrt{\frac{1}{4}m} + \frac{1}{2}m\right\}\right| \in \left(\Phi(b) - \frac{\varepsilon}{6}, \Phi(b) + \frac{\varepsilon}{6}\right).$$

Now let $\ell \in \mathbb{N}$ with $2^{-\ell} < \frac{\varepsilon}{3}$ and $j \in \{1, ..., 2^\ell\}$. For every $m \geq \ell + m_0$,

$$\begin{aligned}
&\frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \leq n < 2^m + j\,2^{m-\ell} : s_2(n) \leq b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\geq \frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \leq n < 2^m + j\,2^{m-\ell} : s_2(n) \leq b\sqrt{\frac{1}{4}m} + \frac{1}{2}m\right\}\right| \\
&= \sum_{i=1}^{j} \frac{2^{m-\ell}}{j\,2^{m-\ell}} 2^{-(m-\ell)}\left|\left\{0 \leq n < 2^{m-\ell} : s_2(n) \leq b\sqrt{\frac{m}{4}} + \frac{m}{2} - s_2(i)\right\}\right|.
\end{aligned}$$

Since $m - \ell \geq m_0$,

$$\begin{aligned}
&2^{-(m-\ell)}\left|\left\{0 \leq n < 2^{m-\ell} : s_2(n) \leq b\sqrt{\frac{m}{4}} + \frac{m}{2} - s_2(i)\right\}\right| \\
&\geq \Phi\Big(b\sqrt{\frac{m}{m-\ell}} + \frac{\ell - 2s_2(i)}{\sqrt{m-\ell}}\Big) - \frac{\varepsilon}{6} \geq \Phi\Big(b\sqrt{\frac{m}{m-\ell}} - \frac{\ell}{\sqrt{m-\ell}}\Big) - \frac{\varepsilon}{6}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \leq n < 2^m + j\,2^{m-\ell} : s_2(n) \leq b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\geq \Phi\Big(b\sqrt{\frac{m}{m-\ell}} - \frac{\ell}{\sqrt{m-\ell}}\Big) - \frac{\varepsilon}{6},
\end{aligned}$$

and in the same way,

$$\begin{aligned}
&\frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \le n < 2^m + j\,2^{m-\ell} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\quad = \frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \le n < 2^m + j\,2^{m-\ell} : s_2(n) \le b\sqrt{\frac{1}{4}(m+1)} + \frac{1}{2}(m+1)\right\}\right| \\
&\quad \le \Phi\Big(b\sqrt{\frac{m+1}{m-\ell}} + \frac{1+\ell}{\sqrt{m-\ell}}\Big) + \frac{\varepsilon}{6}.
\end{aligned}$$

Now for fixed $b \in \mathbb{R}$ there exists $m_1 \in \mathbb{N}$ with $m_1 \ge m_0 + \ell$ such that for all $m \ge m_1$

$$\begin{aligned}
\Phi(b) - \frac{\varepsilon}{3} &< \frac{1}{j\,2^{m-\ell}}\left|\left\{2^m \le n < 2^m + j\,2^{m-\ell} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&< \Phi(b) + \frac{\varepsilon}{3}.
\end{aligned}$$

Note that this equation holds in particular for $j = 2^\ell$, so that

$$\Phi(b) - \frac{\varepsilon}{3} < \frac{1}{2^m}\left|\left\{2^m \le n < 2^{m+1} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| < \Phi(b) + \frac{\varepsilon}{3}.$$

Now let $N > 2^{m_1}\frac{3}{\varepsilon}$, and let $m = \lfloor \log_2(N) \rfloor$. Then $2^m + (j-1)2^{m-\ell} \le N < 2^m + j2^{m-\ell}$ for some $j \in \{1, ..., 2^\ell\}$. Then,

$$
\begin{aligned}
&\frac{1}{N}\left|\left\{0 \le n < N : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&= \frac{1}{N}\left|\left\{0 \le n < 2^{m_1} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\quad + \sum_{k=m_1}^{m-1} \frac{2^k}{N}\frac{1}{2^k}\left|\left\{2^k \le n < 2^{k+1} : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\quad + 1_{\{j\neq -1\}}\frac{(j-1)2^{m-\ell}}{N}\frac{1}{(j-1)2^{m-\ell}}\left|\left\{2^m \le n < 2^m + (j-1)2^{m-\ell} : s_2(n)\right.\right. \\
&\qquad \left.\left. \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\quad + \frac{1}{N}\left|\left\{2^m + (j-1)2^{m-\ell} \le n < N : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\le \frac{\varepsilon}{3} + \frac{1}{N}\sum_{k=0}^{m-1} 2^k\left(\Phi(b) + \frac{\varepsilon}{3}\right) + 1_{\{j\neq -1\}}\frac{(j-1)2^{m-\ell}}{N}\left(\Phi(b) + \frac{\varepsilon}{3}\right) + 2^{m-\ell}\frac{1}{N} \\
&< \frac{\varepsilon}{3} + \frac{2^m + (j-1)2^{m-\ell}}{N}\left(\Phi(b) + \frac{\varepsilon}{3}\right) + \frac{\varepsilon}{3} \le \Phi(b) + \varepsilon,
\end{aligned}
$$

where we have used that since $2^{-\ell} < \frac{\varepsilon}{3}$, we also have $2^{m-\ell}\frac{1}{N} \le 2^{m-\ell}\frac{1}{2^m} < \frac{\varepsilon}{3}$. In the same way we get

$$
\begin{aligned}
&\frac{1}{N}\left|\left\{0 \le n < N : s_2(n) \le b\sqrt{\frac{1}{4}\log_2(n)} + \frac{1}{2}\log_2(n)\right\}\right| \\
&\ge \frac{1}{N}\sum_{k=m_1}^{m-1} 2^k\left(\Phi(b) - \frac{\varepsilon}{3}\right) + 1_{\{j\neq -1\}}\frac{(j-1)2^{m-\ell}}{N}\left(\Phi(b) - \frac{\varepsilon}{3}\right) \\
&= \frac{2^m - 2^{m_1} + (j-1)2^{m-\ell}}{N}\left(\Phi(b) - \frac{\varepsilon}{3}\right) \\
&= \left(1 - \frac{N - 2^m + 2^{m_1} - (j-1)2^{m-\ell}}{N}\right)\left(\Phi(b) - \frac{\varepsilon}{3}\right) \\
&= \Phi(b) - \frac{2^{m_1}}{N} - \frac{\varepsilon}{3} - \frac{N - 2^m - (j-1)2^{m-\ell}}{N} \\
&> \Phi(b) - 2\frac{\varepsilon}{3} - \frac{2^{m-\ell}}{N} > \Phi(b) - \varepsilon,
\end{aligned}
$$

which proves the result. □

3.3 Uniform distribution mod 1 and Weyl's theorem

In this section we address a famous theorem of Hermann Weyl (9. November 1885 in Elmshorn; 8. December 1955 in Zürich). Before we start, let us remind the reader that the fractional part of a number $x \in \mathbb{R}$ is defined as

$$\{x\} := x - \lfloor x \rfloor$$

where

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\}.$$

If we are given a sequence $x : \mathbb{N} \to \mathbb{R}$ and a set $B \subseteq [0,1)$, then we define another set by setting

$$A_{x,B} := \big\{n \in \mathbb{N} : \{x_n\} \in B\big\}.$$

The sequence $x = (x_n)_{n\in\mathbb{N}}$ is said to be uniformly distributed modulo 1 (we simply write mod 1) if and only if for all $a, b \in \mathbb{R}$ with $0 \leq a < b \leq 1$, we have

$$\mu_R\big(A_{x,[a,b)}\big) = b - a\,.$$

In particular, this means that for each uniformly distributed sequence $(x_n)_{n\in\mathbb{N}}$ the sequence $(\{x_n\})_{n\in\mathbb{N}}$ is relatively measurable.

Weyl's theorem [51, 52], also known as Weyl's criterion, says that a sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is uniformly distributed mod 1 if and only if for every $h \in \mathbb{Z} \setminus \{0\}$ the following condition is satisfied,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0\,.$$

In an extended and multivariate version this theorem reads as follows.

Theorem 3 Let $m \in \mathbb{N}$ and consider sequences $x^1, ..., x^m : \mathbb{N} \to \mathbb{R}$. Then the following are equivalent:

1. Every sequence $x^k, k \in \{1, ..., m\}$ is uniformly distributed mod 1 and $\{x^1\}, ..., \{x^m\}$ are μ_R-independent;
2. For each m-tuple $(h_1, ..., h_m) \in \mathbb{Z}^m \setminus \{0\}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (h_1 x_n^1 + ... + h_m x_n^m)} = 0;$$

3. For every continuous function $\psi : [0,1]^m \to \mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \psi(\{x_n^1\}, ..., \{x_n^m\}) = \int_{[0,1]^m} \psi(z_1, ..., z_m)\, \mathrm{d}z_1 ... \mathrm{d}z_m\,;$$

4. For every Riemann integrable function $\psi : [0,1]^m \to \mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \psi(\{x_n^1\}, ..., \{x_n^m\}) = \int_{[0,1]^m} \psi(z_1, ..., z_m)\, \mathrm{d}z_1 ... \mathrm{d}z_m\,.$$

An important consequence is that for each $\alpha \in \mathbb{R}$ the sequence $(n\alpha)_{n\in\mathbb{N}}$ is uniformly distributed mod 1 if and only if α is irrational, and that for $\alpha_1, ..., \alpha_m \in \mathbb{R}$ the sequences $\{\alpha_1 n\}_{n\in\mathbb{N}}, ..., \{\alpha_m n\}_{n\in\mathbb{N}}$ are uniformly distributed mod 1 and μ_R-independent if and only if $1, \alpha_1, ..., \alpha_m$ are linearly independent over $\mathbb{Q}$.

Remark 5 Theorem 3 is also of practical interest, as it provides us with a method for numerical integration of a Riemann integrable function ψ on $[0,1]^m$. Note that, if we only know that the coordinate sequences are uniformly distributed mod 1 and μ_R-independent, we cannot say anything about the speed of convergence of the sums towards the integral.

The concept of discrepany of a sequence measures the speed with which a sequence in $[0,1)^m$ approaches the uniform distribution on $[0,1)^m$. Sequences with a "high" speed of convergence are informally called low-discrepancy sequences and give rise to a class of numerical integration algorithms called quasi-Monte Carlo methods. For more information about these sequences and algorithms see [17, 19, 38, 40].

Definition 7 (Finitely measurable function) We say that a function $g : I \to \mathbb{R}$ is *finitely measurable* if and only if the pre-image of each interval $J \subset \mathbb{R}$ under g can be written as the union of finitely many subintervals, i.e., there exists $k \in \mathbb{N}$ and subintervals $I_1, ..., I_k$ of I such that

$$g^{-1}(J) = I_1 \cup ... \cup I_k\,.$$

Examples of finitely measurable functions are the monotone functions and the functions g with the following so-called Dirichlet property:

A function $g : [a, b] \to \mathbb{R}$ is said to have the Dirichlet property if and only if it is continuous on $[a, b]$ and has only finitely many local extreme points.

A concrete example of a finitely measurable function thus is $\cos(2\pi\cdot) : [0,1] \to \mathbb{R}$, $z \mapsto \cos(2\pi z)$.

Proposition 4 *Let $m \in \mathbb{N}$ and $x^1, ..., x^m : \mathbb{N} \to \mathbb{R}$ be sequences. Consider finitely measurable functions $g^1, ..., g^m : \mathbb{R} \to \mathbb{R}$. If $x^1, ..., x^m$ are relatively measurable and μ_R-independent, then the sequences $g^1(x^1), ..., g^m(x^m)$ are relatively measurable and μ_R-independent.*

The previous result, whose proof is left to the reader, has the following interesting corollary.

Corollary 2 *Let* $1, \alpha_1, ..., \alpha_m \in \mathbb{R}$ *be linearly independent over* $\mathbb{Q}$. *Then the sequences* $\big(\cos(2\pi\alpha_1 n)\big)_{n\in\mathbb{N}}, ..., \big(\cos(2\pi\alpha_m n)\big)_{n\in\mathbb{N}}$ *are relatively measurable and* μ_R*-independent.*

Proof We have already concluded, as a consequence of Weyl's theorem, that the sequences $\{\alpha_1 n\}_{n\in\mathbb{N}}, ..., \{\alpha_m n\}_{n\in\mathbb{N}}$ are uniformly distributed mod 1 and μ_R-independent. Hence, by Proposition 4 the sequences

$$\big(\cos(2\pi\{\alpha_1 n\})\big)_{n\in\mathbb{N}}, ..., \big(\cos(2\pi\{\alpha_m n\})\big)_{n\in\mathbb{N}}$$

are μ_R-independent as well and thus the sequences

$$\big(\cos(2\pi\alpha_1 n)\big)_{n\in\mathbb{N}}, ..., \big(\cos(2\pi\alpha_m n)\big)_{n\in\mathbb{N}}.$$

□

Proposition 5 *Let* $x, y : \mathbb{N} \to \mathbb{R}$ *be bounded and relatively measurable sequences with continuous and increasing distribution functions* F_x *and* F_y *respectively. If* x *and* y *are* μ_R*-independent, then the distribution function* F_{x+y} *of* $x + y$ *is given by the convolution of* F_x *and* F_y*, i.e.,*

$$F_{x+y}(z) = F_x * F_y(z) = \int_{-\infty}^{\infty} F_x(z-\eta)\mathrm{d}F_y(\eta) = \int_{-\infty}^{\infty} F_y(z-\xi)\mathrm{d}F_x(\xi).$$

Proof It is comparably easy to see that the sequences $(F_x(x_n))_{n\in\mathbb{N}}$ and $(F_y(y_n))_{n\in\mathbb{N}}$ are uniformly distributed mod 1. Proposition 4 implies that they are μ_R-independent. Observe that the restriction of F_x to the closure of $\{t \in \mathbb{R} : F_x(t) \in (0,1)\}$ is continuous and increasing and therefore has an inverse, which we denote by G_x. Denote by G_y the corresponding inverse function of F_y. We have

$$\begin{aligned}
\mu_R(x + y \le z) &= \lim_{N\to\infty} \sum_{n=1}^{N} 1_{(-\infty,z]}(x_n + y_n) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} 1_{(-\infty,z]}\Big(G_x\big(F_x(x_n)\big) + G_y\big(F_y(y_n)\big)\Big) \\
&\overset{(*)}{=} \int_{[0,1]^2} 1_{(-\infty,z]}\big(G_x(\xi) + G_y(\eta)\big)\mathrm{d}\xi\,\mathrm{d}\eta \\
&= \int_{\mathbb{R}^2} 1_{(-\infty,z]}(\xi+\eta)\mathrm{d}F_x(\xi)\mathrm{d}F_y(\eta) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-\eta} \mathrm{d}F_x(\xi)\mathrm{d}F_y(\eta) = \int_{-\infty}^{\infty} F_x(z-\eta)\mathrm{d}F_y(\eta),
\end{aligned}$$

where we have used in $(*)$ that $(F_x(x_n))_{n\in\mathbb{N}}$ and $(F_y(y_n))_{n\in\mathbb{N}}$ are uniformly distributed mod 1 and independent. □

If we consider, for instance, the sequence $x = \big(\cos(2\pi\alpha n)\big)_{n\in\mathbb{N}}$ with irrational α, then, since $(\alpha n)_{n\in\mathbb{N}}$ is uniformly distributed mod 1,

$$\begin{aligned} F_x(z) = \mu_R(x \le z) &= \lim_{N->\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{(-\infty,z]}\big(\cos(2\pi\alpha n)\big) \\ &= \int_0^1 1_{(-\infty,z]}\big(\cos(2\pi\xi)\big)\mathrm{d}\xi \\ &= 2\int_0^{\frac{1}{2}} 1_{(-\infty,z]}\big(\cos(2\pi\xi)\big)\mathrm{d}\xi = \frac{1}{\pi}\int_1^{-1} 1_{(-\infty,z]}(\eta)\arccos'(\eta)\mathrm{d}\eta \\ &= \frac{1}{\pi}\int_{-1}^{1} 1_{(-\infty,z]}(\eta)\arcsin'(\eta)\mathrm{d}\eta = 1_{[-1,1]}(z)\frac{1}{\pi}\arcsin(z) + 1_{(1,\infty)}(z)\,. \end{aligned}$$

This means that the distribution function of the sequence $\Big(\cos(2\pi\alpha_1 n) + \ldots + \cos(2\pi\alpha_m n)\Big)_{n\in\mathbb{N}}$ is given by F_x^{*m}. Therefore, we obtain a central limit theorem for partial sums of cosines with linearly independent frequencies, i.e., with $1, \alpha_1, \alpha_2, \ldots$ linearly independent over $\mathbb{Q}$,

$$\begin{aligned} &\lim_{m\to\infty} \mu_R\left(\left\{n \in \mathbb{N} : a \le \frac{\cos(2\pi\alpha_1 n) + \ldots + \cos(2\pi\alpha_m n)}{\sqrt{m/2}} \le b\right\}\right) \\ &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{\xi^2}{2}}\mathrm{d}\xi\,. \end{aligned}$$

3.4 Relatively measurable subsets of $(0,\infty)$–the continuous setting

The deliberations of the previous subsection can quite effortlessly be lifted to a continuous setting. A continuous version of a relative measure on Lebesgue measurable subsets of $\mathbb{R}$ can be defined as the limit

$$\mu_R(A) := \lim_{T\to\infty} \frac{1}{T}\int_0^T 1_A(x)\,\mathrm{d}x$$

if it exists. In analogy to the case of sequences, one obtains a continuous version of Weyl's theorem (see also [38, Chap. 9]) and thus the independence of functions of uniformly distributed functions. An example is again given by the cosines with lin-

early independent frequencies (cf. [34]), i.e., if $1, \alpha_1, \alpha_2, ...$ are linearly independent over $\mathbb{Q}$, then for all $m \in \mathbb{N}$ and all $s_1, ..., s_m \in \mathbb{R}$,

$$\mu_R\Big(\Big\{t \in (0,\infty) : \cos(2\pi\alpha_1 t) \le s_1, \cdots, \cos(2\pi\alpha_m t) \le s_m\Big\}\Big)$$
$$= \prod_{j=1}^{m} \mu_R\Big(\big\{t \in (0,\infty) : \cos(2\pi\alpha_j t) \le s_j\big\}\Big).$$

Those considerations then yield a central limit theorem of the form

$$\lim_{m\to\infty}\mu_R\left(\left\{t \in (0,\infty) : a \le \frac{\cos(2\pi\alpha_1 t) + ... + \cos(2\pi\alpha_m t)}{\sqrt{m/2}} \le b\right\}\right)$$
$$= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{\xi^2}{2}}\,\mathrm{d}\xi\,.$$

The original approach to this result is, as we find, more complicated and can be found in [34]. The latter is presented in a more accessible way in [32, Chap. 3].

4 The Erdős-Kac Theorem

This section is devoted to a famous theorem of Paul Erdős and Mark Kac. One can say that this result marks the birth of what is today known as probabilistic number theory. The close link between probability theory and number theory illustrated by this theorem can hardly be overrated and turned out to be extremely fruitful.

We shall start with the original heuristics of Mark Kac, which led him to conjecture the result he later proved together with Paul Erdős.

4.1 Heuristics–Independence & CLT

A guiding idea of Mark Kac has been that if there is some sort of independence, then there is the Gaussian law of errors at play. Exactly this maxim underlies the Erdős-Kac theorem. The object of interest is the number of different prime factors of a given number.

Let us consider the following indicator functions. For each prime number p and every $n \in \mathbb{N}$, we define

$$I_p(n) = \begin{cases} 1 & : \text{if } p \text{ divides } n \\ 0 & : \text{if } p \text{ does not divide } n\,. \end{cases}$$

Given a natural number $n \in \mathbb{N}$, we denote by $\omega(n)$ the number of different prime factors of n. The indicator functions allow us to express $\omega(n)$ as follows,

$$\omega(n) = \sum_{p \text{ prime}} I_p(n).$$

From Sect. 2.4 we already know that this collection of indicator functions is μ_R-independent. We now want to provide a plausibility argument, and here we follow Mark Kac's original heuristics, that suggests these indicator functions also satisfy Lindeberg's condition. In analogy to the central limit theorem of Lindeberg, this suggests that the properly normalized sum of indicator functions follows a Gaussian law of errors. For this we note first that for all $x \in \mathbb{R}$ with $x \geq 2$ we have

$$\sum_{\substack{p \text{ prime}, \\ p \leq x}} \frac{1}{p} > \ln \ln x - \frac{1}{2}, \tag{7}$$

see [28, Kap. 3]. As we already explained in the first part of Sect. 2.4, essentially a fraction of $1/p$ of the numbers is divisible by the prime p, i.e., we may say that a number $n \in \mathbb{N}$ is divisible by p with probability $1/p$. In other words, the indicator functions $I_p(n)$ behave like Bernoulli random variables with parameter $1/p$ and are independent. But then the expectation is $1/p$ and the variance $1/p(1-1/p)$. What does it mean for Lindeberg's condition? Well, using the notation of Theorem 2, we have for all $n \geq 2$

$$\begin{aligned} s_n &= \sqrt{\sum_{\substack{p \text{ prime} \\ p \leq n}} \operatorname{Var}[I_p(n)]} = \sqrt{\sum_{\substack{p \text{ prime} \\ p \leq n}} \frac{1}{p}\left(1 - \frac{1}{p}\right)} \\ &\geq \frac{1}{\sqrt{2}} \sqrt{\sum_{\substack{p \text{ prime} \\ p \leq n}} \frac{1}{p}} \overset{(7)}{\geq} \frac{1}{\sqrt{2}} \sqrt{\ln \ln n - \frac{1}{2}}. \end{aligned}$$

So if $\varepsilon \in (0, \infty)$, then for sufficiently large $n \in \mathbb{N}$, we have

$$\begin{aligned} \mathbb{E}\Big[I_p(n)^2 \, \mathbb{1}_{\{|I_p(n)| > \varepsilon s_n\}} \Big] &\leq \mathbb{P}\big[I_p(n) > \varepsilon s_n \big] \\ &\leq \mathbb{P}\Big[I_p(n) > \frac{\varepsilon}{\sqrt{2}} \sqrt{\ln \ln n - 1/2} \Big] = 0. \end{aligned}$$

The latter holds since $I_p(n)$ only takes the values 0 and 1. Therefore, Lindeberg's condition in Theorem 2 is satisfied. Together with the independence of the indicator functions $I_p(n)$, p prime as well as property (7), this suggests that the sequence

$$\frac{\omega(n) - \ln \ln n}{\sqrt{\ln \ln n}}, \qquad n \in \mathbb{N}$$

satisfies a central limit theorem. Indeed, for every $m \in \mathbb{N}$ let $c_m = \sum_{p \text{ prime}, p \le m} \frac{1}{p}$ and $d_m^2 = \sum_{p \text{ prime}, p \le m} \frac{1}{p}\left(1 - \frac{1}{p}\right)$. Further let, $\omega_m(n) := \sum_{p \text{ prime}, p \le m} I_p(n)$. Then for every $a, b \in \mathbb{R}$ with $a < b$,

$$\lim_{m\to\infty} \mu_R\left(\left\{n \in \mathbb{N} : a \le \frac{\omega_m(n) - c_m}{d_m} \le b\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, \mathrm{d}x\,,$$

which appears as Lemma 1 in [22]. This means that

$$\lim_{m\to\infty} \lim_{N\to\infty} \frac{1}{N} \left|\left\{n \in \{1, ..., N\} : a \le \frac{\omega_m(n) - c_m}{d_m} \le b\right\}\right| = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, \mathrm{d}x\,.$$

If one could show that the two limits may be taken simultaneously, then we would obtain

$$\lim_{N\to\infty} \frac{1}{N} \left|\left\{n \in \{1, ..., N\} : a \le \frac{\omega_N(n) - c_N}{d_N} \le b\right\}\right| = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, \mathrm{d}x\,.$$

Together with the (proper) asymptotics for $\omega_N(n), c_N, d_N$, this would give

$$\lim_{N\to\infty} \frac{1}{N} \left|\left\{n \in \{1, ..., N\} : a \le \frac{\omega(n) - \ln\ln N}{\sqrt{\ln\ln N}} \le b\right\}\right| = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, \mathrm{d}x\,.$$

Of course, this is merely a heuristic argument, not a proof. In any case, the heuristic and conjecture just presented leads us in the following subsection to the ingenious and famous central limit theorem of Erdős-Kac [22].

4.2 The CLT of Erdős-Kac

After having presented the heuristic of Mark Kac, let us tell the anecdote about the origin of the Erdős-Kac theorem as described by Mark Kac himself in his autobiography [33].

> "*I knew very little number theory at the time, and I tried to find a proof along purely probabilistic lines but to no avail. In March 1939 I journeyed from Baltimore to Princeton to give a talk. Erdős, who was spending the year at the Institute for Advanced Study, was in the audience but he half-dozed through most of my lecture; the subject matter was too far removed from his interests. Toward the end I described briefly my difficulties with the number of prime divisors. At the mention of number theory Erdős perked up and asked me to explain once again what the difficulty was. Within the next few minutes, even before the lecture was over, he interrupted to announce that he had the solution.*"

When once asked about their famous result, Mark Kac replied the following (see [14] and [33]):

"*It took what looks now like a miraculous confluence of circumstances to produce our result.... It would not have been enough, certainly not in 1939, to bring a number theorist and a probabilist together. It had to be Erdős and me: Erdős because he was almost unique in his knowledge and understanding of the number theoretic method of Viggo Brun,... and me because I could see independence and the normal law through the eyes of Steinhaus.*"

We will now formulate the central limit theorem of Erdős and Kac.

Theorem 1 (Erdős-Kac, 1940) Let $a, b \in \mathbb{R}$ with $a < b$. Then

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ n \in \{1, ..., N\} : a \leq \frac{\omega(n) - \ln\ln N}{\sqrt{\ln\ln N}} \leq b \right\} \right| = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, \mathrm{d}x \, .$$

In other words, for large $N \in \mathbb{N}$ the proportion of natural numbers in the set $\{1, ..., N\}$ for which the suitably normalized number of different prime factors is between a and b is close to a Gaussian integral from a to b. In short: the number of prime factors of a large, suitably normalized number follow a Gaussian curve.

Providing a formal proof for Theorem 1 would go beyond the scope of this paper. The original argument of Erdős and Kac use number theoretic methods of sieve theory (more precisely Brun's sieve). Another proof is due to Alfréd Rényi (20. March 1921 in Budapest; 1. February 1970 Budapest) and Pál Turán (18. August 1910 in Budapest; 26. September 1976 Budapest) and can be found in [46]. Let us mention that Godfrey Harold Hardy (7. February 1877 in Cranleigh; 1. December 1947 in Cambridge) and Srinivasa Ramanujan (22. December 1887 in Erode; 26. April 1920 in Kumbakonam) prove in their paper [27] from 1917 that for all $\varepsilon \in (0, \infty)$

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ n \in \{1, ..., N\} : \left| \frac{\omega(n)}{\ln\ln N} - 1 \right| \geq \varepsilon \right\} \right| = 0 \, .$$

This means that for large $N \in \mathbb{N}$ if we pick a number $n \in \{1, ..., N\}$ at random (with respect to the uniform distribution), then the number $\omega(n)$ of different prime factors is of order $\ln\ln N$.

Remark 6 Even though Pál Turán already noticed that the result of Hardy and Ramanujan can be obtained from an inequality for the second moment of $\omega(n)$ together with an application of Chebychev's inequality [9], one can say that the Erdős-Kac Theorem marks the beginning of probabilistic number theory. Also the work [23] of Paul Erdős and Aurel Wintner (8. April 1903 in Budapest; 15. January 1958 in Baltimore) has been one of the pioneering contributions to this complex of problems.

We close this section with the statement of a corollary that gives a different version of the Erdős-Kac theorem, in which N in the loglog terms is replaced by n,

which looks more natural in our setup, because it directly states that the distribution function of the sequence $\left(\frac{\omega(n)-\ln\ln n}{\sqrt{\ln\ln n}}\right)_{n\in\mathbb{N}}$ is that of the standard normal one.

Corollary 3 *Let $a, b \in \mathbb{R}$ with $a < b$. Then*

$$\lim_{N\to\infty} \frac{1}{N}\left|\left\{n \in \{1,...,N\} : a \le \frac{\omega(n)-\ln\ln n}{\sqrt{\ln\ln n}} \le b\right\}\right| = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,\mathrm{d}x\,.$$

Proof Clearly, for every $b \in \mathbb{R}$, we have

$$\begin{aligned}&\limsup_{N\to\infty} \frac{1}{N}\left|\left\{n \in \{1,...,N\} : \frac{\omega(n)-\ln\ln n}{\sqrt{\ln\ln n}} \le b\right\}\right| \\ &\le \lim_{N\to\infty} \frac{1}{N}\left|\left\{n \in \{1,...,N\} : \frac{\omega(n)-\ln\ln N}{\sqrt{\ln\ln N}} \le b\right\}\right| = \Phi(b)\,,\end{aligned}$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-x^2/2}\,\mathrm{d}x$ for all $t \in \mathbb{R}$ as before. First note that, by Theorem 1, the distribution functions F_N with $F_N(t) := \frac{1}{N}|\{1 \le n \le N : \omega(n) \le t\sqrt{\ln\ln N} + \ln\ln N\}|$ converge pointwise to Φ, and therefore also uniformly on $\mathbb{R}$, by Lemma 3.

Now fix $b \in \mathbb{R}$ and let $K \in (0,\infty)$ be such that $e^{-\frac{K}{2}} < \frac{\varepsilon}{3}$. Let $N_0 \in \mathbb{N}$ be such that for all $N \ge N_0$ and all $t \in \mathbb{R}$ we have $F_N(t) \in (\Phi(t) - \frac{\varepsilon}{3}, \Phi(t) + \frac{\varepsilon}{3})$, $\Phi(b-\frac{K}{\ln\ln N}) > \Phi(b)-\frac{\varepsilon}{3}$, $\sqrt{\ln\ln N} > b$, and $\ln\ln N > 0$. With this

$$\begin{aligned}&\frac{1}{N}\left|\left\{n \in \{1,...,N\} : \omega(n) \le b\sqrt{\ln\ln N} + \ln\ln N - K\right\}\right| \\ &\ge \Phi\Big(b - \frac{K}{\ln\ln N}\Big) - \frac{\varepsilon}{3} > \Phi(b) - \frac{2\varepsilon}{3}\end{aligned}$$

If we denote $N_1 := \sup\{n \in \mathbb{N} : b\sqrt{\ln\ln N} + \ln\ln N - K > b\sqrt{\ln\ln n} + \ln\ln n\}$, then

$$\begin{aligned}&\frac{1}{N}\left|\left\{n \in \{1,...,N\} : \omega(n) \le b\sqrt{\ln\ln n} + \ln\ln n\right\}\right| \\ &\ge \frac{1}{N}\left|\left\{n \in \{N_1+1,...,N\} : \omega(n) \le b\sqrt{\ln\ln n} + \ln\ln n\right\}\right| \\ &\ge \frac{1}{N}\left|\left\{n \in \{N_1+1,...,N\} : \omega(n) \le b\sqrt{\ln\ln N} + \ln\ln N - K\right\}\right| \\ &\ge \frac{1}{N}\left|\left\{n \in \{1,...,N\} : \omega(n) \le b\sqrt{\ln\ln N} + \ln\ln N - K\right\}\right| - \frac{N_1}{N} \\ &> \Phi(b) - \frac{2\varepsilon}{3} - \frac{N_1}{N}\,.\end{aligned}$$

Now, let us compare N and N_1. We observe that if

$$b(\sqrt{\ln\ln N}-\sqrt{\ln\ln N_1}) + \ln\ln N - \ln\ln N_1 > K\,,$$

then

$$(\sqrt{\ln\ln N} + \sqrt{\ln\ln N_1})(\sqrt{\ln\ln N} - \sqrt{\ln\ln N_1}) + \ln\ln N - \ln\ln N_1 > K\,,$$

which implies that

$$2(\ln\ln N - \ln\ln N_1) > K\,.$$

Hence, we have

$$\ln\ln N - \frac{K}{2} > \ln\ln N_1$$

and so $N^{e^{-\frac{K}{2}}} > N_1$. Therefore,

$$\frac{N_1}{N} < N^{e^{-\frac{K}{2}}-1} < N^{-K/2} < e^{-K/2} < \frac{\varepsilon}{3}\,,$$

which completes the proof. □

A similar calculation shows that the two formulations of the Erdős-Kac theorem are actually equivalent.

5 Some complementary considerations–The case of lacunary series

What we have seen so far shows the power of the concept of relative measure in number theory and how it can naturally (in large parts along the lines of classical probability theory) lead us to central limit theorems for number theoretic quantities, even where the axiomatic framework of Kolmogorov is not applicable. On the other hand, we have seen, when studying binary expansions, that Kolmogorov's theory is a powerful tool as well and allows us to obtain information about the Gaussian fluctuations of number theoretic quantities. A common spirit of both, and eventually a key to a Gaussian law, has always been a notion of independence.

In what follows, we complement the previous considerations by showing that lacunary series, for instance those that are formed with functions $\cos(2\pi n_k \cdot) : [0,1] \to \mathbb{R}$ and quickly increasing gap sequence $(n_k)_{k\in\mathbb{N}}$, behave in many ways like *independent* random variables, and that this almost-independence or weak form of independence may still lead to fascinating results within the axiomatic theory of Kolmogorov.

Already in Sect. 2.3 on binary expansions we noted that Hans Rademacher introduced in [45] what is known today as Rademacher functions. Those functions are defined in the following way,

$$r_k(t) = \operatorname{sign}(\sin(2^k \pi t)), \qquad t \in [0,1],\ k \in \mathbb{N}\,,$$

where for $x \in \mathbb{R}$,

$$\operatorname{sign}(x) := \begin{cases} -1 & : x < 0 \\ 0 & : x = 0 \\ +1 & : x > 0. \end{cases}$$

Rademacher studied the convergence behavior of series

$$\sum_{k=1}^{\infty} a_k r_k(t), \qquad t \in [0,1],\ (a_k)_{k=1}^{\infty} \in \mathbb{R}^{\mathbb{N}}, \tag{8}$$

and proved that such series converge for almost all $t \in [0,1]$ if

$$\sum_{k=1}^{\infty} a_k^2 < +\infty. \tag{9}$$

The necessity of square integrability was obtained by Alexander Khintchine (19. July 1894 in Kondyrjowo; 18. November 1959 in Moscow) and Andrei Kolmogorov in their 1925 paper [35], showing that if

$$\sum_{k=1}^{\infty} a_k^2 = +\infty, \tag{10}$$

then the series (8) diverges for almost all $t \in [0,1]$.

Starting in the 1920s, Stefan Banach (30. March 1892 in Krakow; 31. August 1945 in Lviv), Andrei Kolmogorov, Raymond Paley (7. January 1907 in Bournemouth; 7. April 1933 near Banff), Antoni Zygmund (25. December 1900 in Warsaw; 30. May 1992 in Chicago) and others studied the convergence behavior of trigonometric series

$$\sum_{k=1}^{\infty} a_k \cos(2\pi n_k t), \qquad t \in [0,1],\ (a_k)_{k=1}^{\infty} \in \mathbb{R}^{\mathbb{N}}, \tag{11}$$

where the sequence $(n_k)_{k=1}^{\infty}$ satisfies the Hadamard gap condition

$$\frac{n_{k+1}}{n_k} > q > 1$$

for all $k \in \mathbb{N}$ (see [7, 36, 44, 53]). For such series one can obtain results similar to those for Rademacher series (8). Kolmogorov could prove in [36] that the square summability condition (9) is also sufficient for almost everywhere convergence of lacunary series. The necessity of (9) has been shown by Zygmund in [53].

An important analogy between Rademacher series and lacunary series, in particular in view of our article, remained unnoticed for a long time. In Sect. 2.3 we proved that the Rademacher functions (more precisely a version of them) are independent. In particular, given any sequence $(a_k)_{k=1}^{\infty}$ of real numbers, the functions

$a_k r_k$, $k \in \mathbb{N}$ are independent (but no longer identically distributed), and we have for all $k \in \mathbb{N}$ that

$$\mathbb{E}[a_k r_k] = 0 \qquad \text{and} \qquad \mathrm{Var}[a_k r_k] = a_k^2 .$$

Using the notation from Lindeberg's theorem (see Theorem 2), we see that

$$s_n^2 = \sum_{k=1}^{n} \mathrm{Var}[a_k r_k] = \sum_{k=1}^{n} a_k^2 .$$

But this means that for $\varepsilon \in (0,\infty)$, Lindeberg's condition for the weighted Rademacher functions reads as follows,

$$\frac{1}{\sum\limits_{k=1}^{n} a_k^2} \sum_{k=1}^{n} \mathbb{E}\left[(a_k r_k)^2 \mathbb{1}_{\left\{ |a_k r_k| \geq \varepsilon \sqrt{\sum_{k=1}^{n} a_k^2} \right\}} \right]$$
$$= \frac{1}{\sum\limits_{k=1}^{n} a_k^2} \sum_{k=1}^{n} a_k^2 \, \mathbb{P}\left[|a_k| \geq \varepsilon \sqrt{\sum_{k=1}^{n} a_k^2} \right].$$

For Lindeberg's condition to be satisfied, we require the right-hand side to converge to 0 as $n \to \infty$. A moment's thought, however, reveals that this is the case whenever

$$\sum_{k=1}^{\infty} a_k^2 = +\infty \qquad \text{and} \qquad \max_{1 \leq k \leq n} |a_k| = o\left(\sqrt{\sum_{k=1}^{n} a_k^2} \right). \tag{12}$$

Therefore, under condition (12), we obtain that, for all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \lambda\left(\left\{ x \in [0,1] : \sum_{k=1}^{n} a_k r_k(x) \leq t \sqrt{\sum_{k=1}^{n} a_k^2} \right\} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} \, \mathrm{d}y .$$

It was not before 1947 that Raphaël Salem (7. November 1898 in Saloniki; 20. June 1963 in Paris) and Antoni Zygmund proved in [47] that for Hadamard gap sequences the functions $\big(\cos(2\pi n_k \cdot)\big)_{k\in\mathbb{N}}$ follow a central limit theorem, i.e., for all $t \in \mathbb{R}$,

$$\lim_{N\to\infty} \lambda\left(\left\{ x \in (0,1) : \sum_{k=1}^{N} \cos(2\pi n_k x) \leq t \sqrt{N/2} \right\} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} \, \mathrm{d}y .$$

For sequences with very large gaps, i.e., those satisfying the stronger condition

$$\frac{n_{k+1}}{n_k} \overset{k\to\infty}{\longrightarrow} +\infty ,$$

such a central limit theorem had been obtained in 1939 by Mark Kac in [29].

Around the same time as Salem and Zygmund, Mark Kac [30] (see also [31, 32] and the references therein) obtained a central limit theorem for functions $f : \mathbb{R} \to \mathbb{R}$ of bounded variation on [0,1] satisfying

$$f(t+1) = f(t) \qquad \text{and} \qquad \int_0^1 f(t)\,\mathrm{d}t = 0\,.$$

He showed that for such functions

$$\lim_{N\to\infty} \lambda\left(\left\{x \in (0,1) : \sum_{k=1}^{N} f(2^k x) \le t\sigma\sqrt{N}\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}}\,\mathrm{d}y$$

whenever

$$\sigma^2 := \int_0^1 f(t)^2\,\mathrm{d}t + 2\sum_{k=1}^{\infty} \int_0^1 f(t) f(2^k t)\,\mathrm{d}t \neq 0\,. \tag{13}$$

This already indicates that the functions $f(2^k \cdot)$, $k \in \mathbb{N}$ do not behave like independent random variables. In fact, in that case we would expect something like

$$\sigma^2 = \int_0^1 f(t)^2\,\mathrm{d}t \neq 0$$

rather than condition (13). After further progress had been made by Gapoškin [25] and Takahashi [48], Gapoškin eventually discovered a deep connection between the validity of a central limit theorem and the number of solutions of a certain Diophantine equation [26], i.e., whether a central limit theorem holds or not depends not only on the growth rate of the sequence $(n_k)_{k\in\mathbb{N}}$, but also critically on its number theoretic properties. In 2010 Christoph Aistleitner and István Berkes presented a paper in which they obtained both necessary and sufficient conditions under which a sequence $f(n_k \cdot)_{k\in\mathbb{N}}$ follows a Gaussian law of errors [1].

Please note that the preceding paragraph is not intended to be exhaustive. Still it indicates the development of the subject, highlights some fascinating results, and shows how analytic, probabilistic, and number theoretic arguments and properties intertwine.

Remark 7 The results presented in this final section are not restricted to central limit phenomena. Beyond the normal fluctuations one can also prove laws of the iterated logarithm for lacunary series and we refer the reader to the work of Erdős and Gál [21], Aistleitner and Fukuyama [4, 5], Aistleitner, Berkes, and Tichy [2, 3], and the references cited therein. The study of large deviation principles for lacunary sums has recently been initiated by Aistleitner, Gantert, Kabluchko, Prochno, and Ramanan in [6].

Acknowledgements GL is supported by the Austrian Science Fund (FWF) Project F5508-N26, which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications". JP is supported by the Austrian Science Fund (FWF) Project P32405 "Asymptotic Geometric Analysis and

Applications" as well as a visiting professorship from Ruhr University Bochum and its Research School PLUS.

We thank Christoph Aistleitner, Jordan Stoyanov and an anonymous referee for valuable comments and suggestions.

Funding Open access funding provided by University of Graz.

References

1. Aistleitner, C., Berkes, I.: On the central limit theorem for f(n_{k}x). Probab. Theory Relat. Fields **146**(1-2), 267–289 (2010)
2. Aistleitner, C., Berkes, I., Tichy, R.: On permutations of lacunary series. In Functions in number theory and their probabilistic aspects. RIMS Kôkyûroku Bessatsu **B34**, 1–25 (2012)
3. Aistleitner, C., Berkes, I., Tichy, R.: On the law of the iterated logarithm for permuted lacunary sequences. Tr. Mat. Inst. Steklova **276**, 9–26 (2012)
4. Aistleitner, C., Fukuyama, K.: On the law of the iterated logarithm for trigonometric series with bounded gaps. Probab. Theory Relat. Fields **154**(3-4), 607–620 (2012)
5. Aistleitner, C., Fukuyama, K.: On the law of the iterated logarithm for trigonometric series with bounded gaps II. J. Théor. Nombres Bordeaux **28**(2), 391–416 (2016)
6. Aistleitner, C., Gantert, N., Kabluchko, Z., Prochno, J., Ramanan, K.: Large deviation principles for lacunary sums. preprint (2020)
7. Banach, S.: Über einige Eigenschaften der lakunären trigonometrischen Reihen. Studia Math. **2**(1), 207–220 (1930)
8. Behrends, E.: Analysis Band 1: Ein Lernbuch für den sanften Wechsel von der Schule zur Uni. BVL-Reporte, vol. 1. Vieweg + Teubner, Wiesbaden (2009)
9. Berkes, I., Dehling, H.: Dependence in probability, analysis and number theory: the mathematical work of Walter Philipp (1936–2006). In: Dependence in probability, analysis and number theory, pp. 1–19. Kendrick Press, Heber City (2010)
10. Billingsley, P.: Prime numbers and Brownian motion. Amer. Math. Mon. **80**, 1099–1115 (1973)
11. Billingsley, P.: The probability theory of additive arithmetic functions. Ann. Probab. **2**, 749–791 (1974)
12. Bohlmann, G.: Lebensversicherungsmathematik. Encykl. Math. Wissenschaften **I**(2), 852–917 (1900)
13. Borel, É.: Les probabilités dénombrables et leurs applications arithmétiques. Rendiconti Del Circolo Matematico Di Palermo (1884-1940) **27**(1), 247–271 (1909)
14. Cohen, J.E.: A life of the immeasurable mind. Ann. Probab. **14**(4), 1139–1148 (1986)
15. de Moivre, A.: The doctrine of chances: a method of calculating the probabilities of events in play. Chelsea Publishing Co, New York (1967)
16. de Moivre, A.: The doctrine of chances or, a method of calculating the probabilities of events in play. New impression of the second edition, with additional material. Cass Library of Science Classics, vol. 1. Frank Cass & Co., Ltd, London (1967)
17. Dick, J., Pillichshammer, F.: Digital nets and sequences: discrepancy theory and quasi-Monte Carlo integration. Cambridge University Press, Cambridge (2010)
18. Drmota, M., Gajdosik, J.: The distribution of the sum-of-digits function. J. Théor. Nombres Bordeaux **10**(1), 17–32 (1998)
19. Drmota, M., Tichy, R.F.: Sequences, discrepancies and applications. Lecture Notes in Mathematics, vol. 1651. Springer, Berlin (1997)
20. Eichelsbacher, P., Löwe, M.: 90 Jahre Lindeberg-Methode. Math Semesterber **61**(1), 7–34 (2014)
21. Erdős, P., Gál, I.S.: On the law of the iterated logarithm. I, II. Nederl. Akad. Wetensch. Proc. Ser. A. 58. Indag. Math. **17**(65–76), 77–84 (1955)

22. Erdős, P., Kac, M.: The Gaussian law of errors in the theory of additive number theoretic functions. Amer. J. Math. **62**, 738–742 (1940)
23. Erdős, P., Wintner, A.: Additive arithmetical functions and statistical independence. Amer. J. Math. **61**, 713–721 (1939)
24. Fischer, H.: A history of the central limit theorem. Sources and studies in the history of mathematics and physical sciences. Springer, New York (2011). From classical to modern probability theory
25. Gapoškin, V.F.: Lacunary series and independent functions. Uspehi Mat. Nauk **21**(6), 3–82 (1966)
26. Gapoškin, V.F.: The central limit theorem for certain weakly dependent sequences. Teor. Verojatnost. I Primenen. **15**, 666–684 (1970)
27. Hardy, G.H., Ramanujan, S.: The normal number of prime factors of a number n [Quart. J. Math. 48 (1917), 76–92]. In: Collected papers of Srinivasa Ramanujan, pp. 262–275. AMS Chelsea Publ, Providence (2000)
28. Ischebeck, F.: Einladung zur Zahlentheorie. BI-Wiss.-Verlag, Mannheim (1992)
29. Kac, M.: Note on power series with big gaps. Amer. J. Math. **61**(2), 473–476 (1939)
30. Kac, M.: On the distribution of values of sums of the type f(2^{k}t). Ann. Math. **47**(2), 33–49 (1946)
31. Kac, M.: Probability methods in some problems of analysis and number theory. Bull. Amer. Math. Soc. **55**, 641–665 (1949)
32. Kac, M.: Statistical independence in probability, analysis and number theory. The Carus mathematical monographs, vol. 12. Mathematical Association of America, New York (1959). Distributed by John Wiley and Sons, Inc.
33. Kac, M.: Enigmas of chance. Alfred P. Sloan Foundation. Harper & Row, New York (1985). An autobiography
34. Kac, M., Steinhaus, H.: Sur les fonctions indendantes (iv) (intervalle infini). Studia Math. **7**(1), 1–15 (1938)
35. Khintchine, A., Kolmogorov, A.: Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Rec. Math. Moscou **32**, 668–677 (1925)
36. Kolmogoroff, A.: Une contribution à l'étude de la convergence des séries de fourier. Fund. Math. **5**(1), 96–97 (1924)
37. Krengel, U.: On the contributions of Georg Bohlmann to probability theory. J. Électron. Hist. Probab. Stat. **7**(1), 13 (2011)
38. Kuipers, L., Niederreiter, H.: Uniform distribution of sequences. John Wiley & Sons, New York-London-Sydney (1974). Pure and Applied Mathematics
39. Laplace, P.-S.: Théorie analytique des probabilités vol. I. Éditions Jacques Gabay, Paris (1995). Introduction: Essai philosophique sur les probabilités. [Introduction: Philosophical essay on probabilities], Livre I: Du calcul des fonctions génératrices. [Book I: On the calculus of generating functions], Reprint of the 1819 fourth edition (Introduction) and the 1820 third edition (Book I)
40. Leobacher, G., Pillichshammer, F.: Introduction to quasi-Monte Carlo integration and applications. Compact textbooks in mathematics. Birkhäuser, Basel (2014)
41. Lindeberg, J.W.: Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Math. Z. **15**(1), 211–225 (1922)
42. Lindeberg, J.W.: Über das Gauss'sche Fehlergesetz. Skand. Aktuarietidskr. **5**, 217–234 (1922)
43. Ljapunov, A.M.: Nouvelle forme du théorème sur la limite de probabilité. (Mémoires de l'Académie Impériale d. sciences de St.-Pétersbourg). Acad. Imp. d. Sciences, St. Petersburg (1901)
44. Paley, R.E.A.C., Zygmund, A.: On some series of functions, (1). Proc. Camb. Philos. Soc. **26**(3), 337–357 (1930)
45. Rademacher, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. Math. Ann. **87**, 112–138 (1922)
46. Rényi, A., Turán, P.: On a theorem of Erdős-Kac. Acta Arith. **4**, 71–84 (1958)
47. Salem, R., Zygmund, A.: On lacunary trigonometric series. Proc. Natl. Acad. Sci. U. S. A. **33**, 333–338 (1947)
48. Takahashi, S.: A gap sequence with gaps bigger than the Hadamards. Tohoku Math. J. **2**(13), 105–111 (1961)
49. Tenenbaum, G.: Introduction à la théorie analytique et probabiliste des nombres, 2nd edn. Cours Spécialisés [Specialized Courses], vol. 1. Société Mathématique de France, Paris (1995)
50. Tenenbaum, G.: Introduction to analytic and probabilistic number theory, 3rd edn. Graduate studies in mathematics, vol. 163. American Mathematical Society, Providence, RI (2015). Translated from the 2008 French edition by Patrick D. F. Ion
51. Weyl, H.: Über ein Problem aus dem Gebiet der Diophantischen Approximationen. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Math.-Phys. Kl. **1914**, 234–244 (1914)

52. Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann. **77**(3), 313–352 (1916)
53. Zygmund, A.: On the convergence of lacunary trigonometric series. Fund. Math. **16**(1), 90–107 (1930)