Arithmetic Randonnée
An introduction to probabilistic number theory

E. Kowalski

ETH Zürich – Fall Semester 2015
Version of December 29, 2015
kowalski@math.ethz.ch

Contents

CHAPTER 1

Introduction

1.1. Presentation

Different authors might define "probabilistic number theory" in different ways. Our point of view will be to see it as *the study of the asymptotic behavior of arithmetically-defined sequences of probability measures.* Thus the content of these notes is based on examples of situations where we can say interesting things concerning such sequences. However, we will quickly say a few words in Section 1.5 on some other topics that might quite legitimately be seen as part of probabilistic number theory in a broader sense.

To illustrate what we have in mind, the most natural starting point is a famous result of Erdős and Kac.

THEOREM 1.1.1 (the Erdős-Kac Theorem). *For any positive integer $n \geqslant 1$, let $\omega(n)$ denote the number of prime divisors of n, counted without multiplicity. Then for any real numbers $a < b$, we have*

$$\lim_{N\to+\infty} \frac{1}{N}\left|\left\{1 \leqslant n \leqslant N \mid a \leqslant \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}} \leqslant b\right\}\right| = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

To spell out the connection between this statement and our slogan, one *sequence of probability measures* involved here is the sequence $(\mu_N)_{N\geqslant 1}$ defined as the uniform probability measure supported on the finite set $\Omega_N = \{1,\dots,N\}$. This sequence is *defined arithmetically*, because the study of integers is part of arithmetic. The *asymptotic behavior* is revealed by the statement. Namely, consider the sequence of random variables

$$X_N(n) = \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}}$$

(defined on Ω_N for $N \geqslant 3$), and the sequence (ν_N) of their probability distributions, which are (Borel) probability measures on $\mathbf{R}$ defined by

$$\nu_N(A) = \mu_N(X_N \in A) = \frac{1}{N}\left|\left\{1 \leqslant n \leqslant N \mid \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}} \in A\right\}\right|$$

for any measurable set $A \subset \mathbf{R}$. These form another *arithmetically-defined sequence of probability measures*, since prime factorizations are definitely concepts of arithmetic nature. Theorem 1.1.1 is, by basic probability theory, equivalent to the fact that the sequence (ν_N) converges in law to a standard normal random variable as $N \to +\infty$.

The Erdős-Kac Theorem is probably the simplest case where a natural deterministic arithmetic quantity (the number of prime factors of an integer), individually very hard to grasp, nevertheless exhibits a probabilistic behavior that is the same as a very common probability distribution. This is the prototype of the kinds of statements we will discuss.

We will prove Theorem 1.1.1 in the next chapter. Before we do this, we will begin in this chapter with a much more elementary result that may, with hindsight, be considered as the simplest case of the type of results we want to describe.

1.2. Integers in arithmetic progressions

As mentioned in the previous section, we begin with a result that is so easy that it is usually not specifically presented as a separate statement (let alone as a theorem!). Nevertheless,

as we will see, it is the basic ingredient (and explanation) for the Erdős-Kac Theorem, and generalizations of it become quite quickly very deep.

THEOREM 1.2.1. *For $N \geqslant 1$, let $\Omega_N = \{1, \dots, N\}$ with the uniform probability measure $\mathbf{P}_N$. Fix an integer $q \geqslant 1$, and denote by $\pi_q : \mathbf{Z} \longrightarrow \mathbf{Z}/q\mathbf{Z}$ the reduction modulo q map. Let X_N be the random variables given by $\mathsf{X}_N(n) = \pi_q(n)$ for $n \in \Omega_N$.*

As $N \to +\infty$, the random variables X converge in law to the uniform probability measure μ_q on $\mathbf{Z}/q\mathbf{Z}$. In fact, for any function

$$f : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C},$$

we have

$$\left|\mathbf{E}(f(\mathsf{X}_N)) - \mathbf{E}(f)\right| \leqslant \frac{2}{N}\|f\|_1, \tag{1.1}$$

where

$$\|f\|_1 = \sum_{a \in \mathbf{Z}/q\mathbf{Z}} |f(a)|.$$

PROOF. It is enough to prove (1.1), which gives the convergence in law by letting $N \to +\infty$. This is quite simple. By definition, we have

$$\mathbf{E}(f(\mathsf{X}_N)) = \frac{1}{N} \sum_{1 \leqslant n \leqslant N} f(\pi_q(n)),$$

and

$$\mathbf{E}(f) = \frac{1}{q} \sum_{a \in \mathbf{Z}/q\mathbf{Z}} f(a).$$

The idea is then clear: among the integers $1 \leqslant n \leqslant N$, roughly N/q should be in any given residue class $a \pmod q$, and if we use this approximation in the first formula, we obtain precisely the second.

To do this in detail, we gather the integers in the sum according to their residue class a modulo q. This gives

$$\frac{1}{N} \sum_{1 \leqslant n \leqslant N} f(\pi_q(n)) = \sum_{a \in \mathbf{Z}/q\mathbf{Z}} f(a) \times \frac{1}{N} \sum_{\substack{1 \leqslant n \leqslant N \\ n \equiv a \pmod q}} 1.$$

The inner sum, for each a, counts the number of integers n in the interval $1 \leqslant n \leqslant N$ such that the remainder under division by q is a. These integers n can be written $n = mq + a$ for some $m \in \mathbf{Z}$, if we view a as an actual integer, and therefore it is enough to count those integers $m \in \mathbf{Z}$ for which $1 \leqslant mq + a \leqslant N$. The condition translates to

$$\frac{1 - a}{q} \leqslant m \leqslant \frac{N - a}{q},$$

and therefore we are reduced to counting integers *in an interval*. This is not difficult, although since the bounds of the interval are not necessarily integers, we have to be careful with boundary terms. Using the usual rounding functions, the number of relevant m is

$$\left\lfloor \frac{N - a}{q} \right\rfloor - \left\lceil \frac{1 - a}{q} \right\rceil + 1.$$

There would be quite a bit to say about trying to continue with this exact identity (many problems would naturally lead us to expand in Fourier series this type of expressions to detect the tiny fluctuations of the fractional parts of the ratio $(N - a)/q$ as a and q vary), but for our purposes we may be quite blunt, and observe that since

$$x \leqslant \lceil x \rceil \leqslant x + 1, \qquad x - 1 \leqslant \lfloor x \rfloor \leqslant x,$$

the number N_a of values of m satisfies

$$\Big(\frac{N-a}{q}-1\Big)-\Big(\frac{1-a}{q}+1\Big)+1\leqslant N_a\leqslant\Big(\frac{N-a}{q}\Big)-\Big(\frac{1-a}{q}\Big),$$

or in other words

$$\Big|N_a-\frac{N}{q}\Big|\leqslant 1+\frac{1}{q}.$$

By summing over a in $\mathbf{Z}/q\mathbf{Z}$, we deduce now that

$$\begin{aligned}\Big|\frac{1}{N}\sum_{1\leqslant n\leqslant N}f(\pi_q(n))-\frac{1}{q}\sum_{a\in\mathbf{Z}/q\mathbf{Z}}f(a)\Big|&=\Big|\sum_{a\in\mathbf{Z}/q\mathbf{Z}}f(a)\Big(\frac{N_a}{N}-\frac{1}{q}\Big)\Big|\\&\leqslant\frac{1+q^{-1}}{N}\sum_{a\in\mathbf{Z}/q\mathbf{Z}}|f(a)|\leqslant\frac{2}{N}\|f\|_1.\end{aligned}$$

□

Despite its simplicity, this result already brings up a number of important features that will occur extensively in later chapters. As a matter of notation, we will usually often simply denote the random variables X_N by π_q, with the value of N made clear by the context, frequently because of its appearance in an expression involving $\mathbf{P}_N(\cdot)$ or $\mathbf{E}_N(\cdot)$, which refers to the probability and expectation on Ω_N.

We begin with the remark that what we actually proved is much stronger than the statement of convergence in law: the bound (1.1) gives a rather precise estimate of the speed of convergence of expectations (or probabilities) computed using the law of X_N to those computed using the limit uniform distribution μ_q. Most importantly, as we will see shortly in our second remark, these estimates are uniform in terms of q, and give us information on convergence, or more properly speaking on "distance" between the law of X_N and μ_q even if q depends on N in some way.

To be more precise, take f to be the characteristic function of a residue class $a\in\mathbf{Z}/q\mathbf{Z}$. Then since $\mathbf{E}(f)=1/q$, we get

$$\Big|\mathbf{P}(\pi_q(n)=a)-\frac{1}{q}\Big|\leqslant\frac{2}{N}.$$

This is non-trivial information as long as q is a bit smaller than N. Thus, this states that the probability that $n\leqslant N$ is congruent to a modulo q is close to the intuitive probability $1/q$ uniformly for all q just a bit smaller than N, and also uniformly for all residue classes. We will see, both below and in many applications and similar situations, that this uniformity features are essential in applications.

Our second remark concerns the interpretation of the result. Theorem 1.2.1 can explain what is meant by such intuitive statements as: "the probability that an integer is divisible by 2 is $1/2$". Namely, this is the probability, according to the uniform measure on $\mathbf{Z}/2\mathbf{Z}$, of the set $\{0\}$, and this is simply the limit given by the convergence in law of the variables $\pi_2(n)$ defined on Ω_N to the uniform measure μ_2.

This idea applies to many other similar-sounding problems. The most elementary among these can often be solved using Theorem 1.2.1. For instance: what is the "probability" that an integer $n\geqslant 1$ is squarefree, which means that n is *not* divisible by a square m^2 for some integer $m\geqslant 2$? Here the interpretation is that this probability should be

$$\lim_{N\to+\infty}\frac{1}{N}|\{1\leqslant n\leqslant N\mid\ \ n\text{ is squarefree}\}|.$$

If we wanted to speak of sequences of random variables here, we would take the sequence of Bernoulli variables B_N defined by

$$\mathbf{P}(\mathsf{B}_N=1)=\frac{1}{N}|\{1\leqslant n\leqslant N\mid\ \ n\text{ is squarefree}\}|,$$

and ask about the limit in law of (B_N). The answer is as follows:

PROPOSITION 1.2.2. *The sequence (B_N) converges in law to a Bernoulli random variable B with $\mathbf{P}(B=1)=\frac{6}{\pi^2}$. In other words, the "probability" that an integer n is squarefree, in the interpretation discussed above, is $6/\pi^2$.*

PROOF. The idea is to use inclusion-exclusion: to say that n is squarefree means that it is not divisible by the square p^2 of any prime number. Thus, if we denote by $\mathbf{P}_N$ the probability measure on Ω_N, we have

$$\mathbf{P}_N(n \text{ is squarefree}) = \mathbf{P}_N\Big(\bigcap_{p \text{ prime}} \{p^2 \text{ does not divide } n\}\Big).$$

There is one key step now that is (in some sense) obvious but crucial: because of the nature of Ω_N, the infinite intersection may be replaced by the intersection over primes $p \leqslant \sqrt{N}$, since all integers in Ω_N are $\leqslant N$. Applying the inclusion-exclusion formula, we obtain

$$\mathbf{P}_N\Big(\bigcap_{p\leqslant N^{1/2}} \{p^2 \text{ does not divide } n\}\Big) = \sum_I (-1)^{|I|}\,\mathbf{P}_N\Big(\bigcap_{p\in I}\{p^2 \text{ divides } n\}\Big) \tag{1.2}$$

where I runs over the set of subsets of the set $\{p \leqslant N^{1/2}\}$ of primes $\leqslant N^{1/2}$, and $|I|$ is the cardinality of I. But, by the Chinese Remainder Theorem, we have

$$\bigcap_{p\in I}\{p^2 \text{ divides } n\} = \{d_I^2 \text{ divides } n\}$$

where d_I is the product of the primes in I. Once more, note that this set is empty if $d_I^2 > N$. Moreover, the fundamental theorem of arithmetic shows that $I \mapsto d_I$ is injective, and we can recover $|I|$ also from d_I as the number of prime factors of d_I. Therefore, we get

$$\mathbf{P}_N(n \text{ is squarefree}) = \sum_{d\leqslant N^{1/2}} \mu(d)\,\mathbf{P}_N(d^2 \text{ divides } n)$$

where $\mu(d)$ is the Möbius function, defined for integers $d \geqslant 1$ by

$$\mu(d) = \begin{cases} 0 & \text{if } d \text{ is not squarefree,} \\ (-1)^k & \text{if } d = p_1\dots p_k \text{ with } p_i \text{ distinct primes.} \end{cases}$$

But d^2 divides n if and only if the image of n by reduction modulo d^2 is 0. By Theorem 1.2.1 applied with $q=d^2$ for all $d \leqslant N^{1/2}$, with f the characteristic function of the 0 residue class, we get

$$\mathbf{P}_N(d^2 \text{ divides } n) = \frac{1}{d^2} + O(N^{-1})$$

for all d, where the implied constant in the $O(\cdot)$ symbol is independent of d (in fact, it is at most 2). Note in passing how we use crucially here the fact that Theorem 1.2.1 was uniform and explicit with respect to the parameter q.

Summing the last formula over $d \leqslant N^{1/2}$, we deduce

$$\mathbf{P}_N(n \text{ is squarefree}) = \sum_{d\leqslant n^{1/2}} \frac{\mu(d)}{d^2} + O\Big(\frac{1}{\sqrt{N}}\Big).$$

Since the series with terms $1/d^2$ converges, this shows the existence of the limit, and that (B_N) converges in law as $N \to +\infty$ to a Bernoulli random variable with success probability

$$\sum_{d\geqslant 1} \frac{\mu(d)}{d^2}.$$

It is a well-known fact (the "Basel problem", first solved by Euler) that

$$\sum_{d\geqslant 1}\frac{1}{d^2}=\frac{\pi^2}{6},.$$

Moreover, a basic property of the Möbius function states that

$$\sum_{d\geqslant 1}\frac{\mu(d)}{d^s}=\frac{1}{\zeta(s)}$$

for any complex number s with $\mathrm{Re}(s)>1$, where

$$\zeta(s)=\sum_{d\geqslant 1}\frac{1}{d^s},$$

and hence we get

$$\sum_{d\geqslant 1}\frac{\mu(d)}{d^2}=\frac{6}{\pi^2}.$$

□

This proof above was written in probabilistic style, emphasizing the connection with Theorem 1.2.1. It can be expressed more straightforwardly as a sequence of manipulation with sums, using the formula

$$\sum_{d^2|n}\mu(d)=\begin{cases}1 & \text{if } n \text{ is squarefree}\\ 0 & \text{otherwise,}\end{cases} \tag{1.3}$$

for $n\geqslant 1$ (which is implicit in our discussion) and the approximation

$$\sum_{\substack{1\leqslant n\leqslant N\\ d|n}}1=\frac{N}{d}+O(1)$$

for the number of integers in an interval which are divisible by some $d\geqslant 1$. This goes as follows:

$$\begin{aligned}\sum_{\substack{n\leqslant N\\ n \text{ squarefree}}}1=\sum_{n\leqslant N}\sum_{d^2|n}\mu(d)&=\sum_{d\leqslant\sqrt{N}}\mu(d)\sum_{\substack{n\leqslant N\\ d^2|n}}1\\ &=\sum_{d\leqslant\sqrt{N}}\mu(d)\Big(\frac{N}{d^2}+O(1)\Big)=N\sum_{d}\frac{\mu(d)}{d^2}+O(\sqrt{N}).\end{aligned}$$

Obviously, this is much shorter, although one needs to know the formula (1.3), which was implicitly derived in the previous proof.[1] But there is something quite important to be gained from the probabilistic viewpoint, which might be missed by reading too quickly the second proof. Indeed, in formulas like (1.2) (or many others), the precise nature of the underlying probability space Ω_N is quite hidden – as is customary in probability where this is often not really relevant. In our situation, this suggests naturally to study similar problems for *different* integer-valued random variables (or different probability measures on the integers) than the random variables n on Ω_N.

This has indeed been done, and in many different ways. But even before looking at any example, we can predict that some new – interesting – phenomena will arise when doing so. Indeed, even if our first proof of Proposition 1.2.2 was written in a very general probabilistic language, it did use one special feature of Ω_N: it only contains integers $n\leqslant N$, and even more particularly, it does not contain any element divisible by d^2 for d larger than $\sqrt{N}$. (More probabilistically, the probability $\mathbf{P}_N(d^2 \text{ divides } n)$ is zero).

[1] Readers who are already well-versed in analytic number theory might find it useful to translate back and forth various estimates written in probabilistic style in these notes.

Now consider the following extension of the problem, which is certainly one of the first that may come to mind beyond our initial setting: we fix a polynomial $P \in \mathbf{Z}[X]$, say irreducible of degree $m \geqslant 1$, and consider – instead of Ω_N and its uniform probability measure – the set $\Omega_{P,N} = P(\Omega_N)$ of values of P over the integers in Ω_N, together with the image $\mathbf{P}_{P,N}$ of the uniform probability measure $\mathbf{P}_N$: we have

$$\mathbf{P}_{P,N}(r) = \frac{1}{N}|\{1 \leqslant n \leqslant N \mid P(n) = r\}|$$

for all $r \in \mathbf{Z}$. Asking about the "probability" that r is squarefree, when r is taken according to $\mathbf{P}_{P,N}$ and $N \to \infty$, is pretty much the same as asking about squarefree values of the polynomial P. But although we know an analogue of Theorem 1.2.1, it is easy to see that this does *not* give enough control of

$$\mathbf{P}_{P,N}(d^2 \text{ divides } r)$$

when d is large compared with N. And this explains partly why, in fact, *there is no single irreducible polynomial $P \in \mathbf{Z}[X]$ of degree 4 or higher for which we know that $P(n)$ is squarefree infinitely often.*

Exercise 1.2.3. (1) Let $k \geqslant 2$ be an integer. Compute the "probability", in the same sense as in Proposition 1.2.2, that an integer n is k-free, i.e., that there is no integer $m \geqslant 2$ such that m^k divides n.

(2) Compute the "probability" that two integers n_1 and n_2 are coprime, in the sense of taking the corresponding Bernoulli random variables on $\Omega_N \times \Omega_N$ and their limit as $N \to +\infty$.

Exercise 1.2.4. Let $P \in \mathbf{Z}[X]$ be an irreducible polynomial of degree m and consider the probability spaces $(\Omega_{P,N}, \mathbf{P}_{P,N})$ as above.

(1) Show that for any $q \geqslant 1$, the random variables $\mathsf{X}_N(r) = \pi_q(r)$, where $\pi_q : \mathbf{Z} \longrightarrow \mathbf{Z}/q\mathbf{Z}$ is the projection, converge in law to a probability measure $\mu_{P,q}$ on $\mathbf{Z}/q\mathbf{Z}$. Is $\mu_{P,q}$ uniform?

(2) Find the largest parameter T, depending on N, such that you can prove that

$$\mathbf{P}_N(r \text{ is not divisible by } p^2 \text{ for } p \leqslant T) > 0$$

for all N large enough. For which value of T would one need to prove this in order to deduce straightforwardly that the set

$$\{n \geqslant 1 \mid P(n) \text{ is squarefree}\}$$

is infinite?

(3) Prove that the set

$$\{n \geqslant 1 \mid P(n) \text{ is } (m+1)\text{-free}\}$$

is infinite.

Finally, there is one last feature of Theorem 1.2.1 that deserves mention because of its probabilistic flavor, and that has to do with independence. If q_1 and q_2 are positive integers which are coprime, then the Chinese Remainder Theorem implies that the map

$$\begin{cases} \mathbf{Z}/q_1q_2\mathbf{Z} \longrightarrow \mathbf{Z}/q_1\mathbf{Z} \times \mathbf{Z}/q_2\mathbf{Z} \\ x \mapsto (x \,(\mathrm{mod}\, q_1), x \,(\mathrm{mod}\, q_2)) \end{cases}$$

is a bijection (in fact, a ring isomorphism). Under this bijection, the uniform measure $\mu_{q_1q_2}$ on $\mathbf{Z}/q_1q_2\mathbf{Z}$ corresponds to the product measure $\mu_{q_1} \otimes \mu_{q_2}$. In particular, the random variables $x \mapsto x \,(\mathrm{mod}\, q_1)$ and $x \mapsto x \,(\mathrm{mod}\, q_2)$ on $\mathbf{Z}/q_1q_2\mathbf{Z}$ are independent.

The interpretation of this is that the random variables π_{q_1} and π_{q_2} on Ω_N are *asymptotically independent* as $N \to +\infty$, in the sense that

$$\lim_{N\to+\infty} \mathbf{P}_N(\pi_{q_1}(n) = a \text{ and } \pi_{q_2}(n) = b) = \frac{1}{q_1q_2} = \Big(\lim_{N\to+\infty} \mathbf{P}_N(\pi_{q_1}(n) = a)\Big) \times \Big(\lim_{N\to+\infty} \mathbf{P}_N(\pi_{q_2}(n) = b)\Big)$$

for all $(a,b) \in \mathbf{Z}^2$. Intuitively, one would say that "divisibility by q_1 and q_2 are independent", and especially that "divisibility by distinct primes are independent events". We summarize this in the following useful proposition:

PROPOSITION 1.2.5. *For $N \geqslant 1$, let $\Omega_N = \{1, \dots, N\}$ with the uniform probability measure $\mathbf{P}_N$. Let $k \geqslant 1$ be an integer, and fix $q_1 \geqslant 1$, ..., $q_k \geqslant 1$ a family of coprime integers.*

As $N \to +\infty$, the vector

$$(\pi_{q_1}, \dots, \pi_{q_k}) : n \mapsto (\pi_{q_1}(n), \dots, \pi_{q_k}(n)),$$

seen as random vector on Ω_N with values in $\mathbf{Z}/q_1\mathbf{Z} \times \cdots \times \mathbf{Z}/q_k\mathbf{Z}$, converges in law to the product of uniform probability measures μ_{q_i}. In fact, for any function

$$f : \mathbf{Z}/q_1\mathbf{Z} \times \cdots \times \mathbf{Z}/q_k\mathbf{Z} \longrightarrow \mathbf{C}$$

we have

$$\left|\mathbf{E}(f(\pi_{q_1}(n), \dots, \pi_{q_k}(n))) - \mathbf{E}(f)\right| \leqslant \frac{2}{N}\|f\|_1. \tag{1.4}$$

PROOF. This is just an elaboration of the previous discussion: let $q = q_1 \cdots q_k$ be the product of the moduli involved. Then the Chinese Remainder Theorem gives a ring-isomorphism

$$\mathbf{Z}/q_1\mathbf{Z} \times \cdots \times \mathbf{Z}/q_k\mathbf{Z} \longrightarrow \mathbf{Z}/q\mathbf{Z}$$

such that the uniform measure μ_q on the right-hand side corresponds to the product measure $\mu_{q_1} \otimes \mu_{q_k}$ on the left-hand side. Thus f corresponds to a function $g : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C}$, and its expectation to the expectation of g according to μ_q. By Theorem 1.2.1, we get

$$\left|\mathbf{E}(f(\pi_{q_1}(n), \dots, \pi_{q_k}(n))) - \mathbf{E}(f)\right| = \left|\mathbf{E}(g(\pi_q(n))) - \mathbf{E}(g)\right| \leqslant \frac{2\|g\|_1}{N},$$

which is the desired result since f and g have also the same ℓ^1 norm. □

REMARK 1.2.6. It is also interesting to observe that the random variables obtained by reduction modulo two coprime integers are not exactly independent: it is not true that

$$\mathbf{P}_N(\pi_{q_1}(n) = a \text{ and } \pi_{q_2}(n) = b) = \mathbf{P}_N(\pi_{q_1}(n) = a)\,\mathbf{P}_N(\pi_{q_2}(n) = b).$$

This fact is the source of many interesting aspects of probabilistic number theory where classical ideas and concepts of probability for sequences of independent random variables are generalized or "tested" in a context where independence only holds in an asymptotic or approximate sense.

1.3. Further topics

Theorem 1.2.1 and Proposition 1.2.5 are obviously very simple statements. However, they should not be disregarded as trivial (and our careful presentation should – maybe – not be considered as overly pedantic). Indeed, if one extends the question to other sequences of probability measures on the integers instead of the uniform measures on $\{1, \dots, N\}$, one quickly encounters very delicate questions, and indeed fundamental open problems.

We have already mentioned the generalization related to polynomial values $P(n)$ for some fixed polynomial $P \in \mathbf{Z}[X]$. Here are some other natural sequences of measures that have been studied:

1.3.1. Primes. Maybe the most important variant consists in replacing all integers $n \leqslant N$ by the subset Π_N of prime numbers $p \leqslant N$ (with the uniform probability measure on these finite sets). According to the Prime Number Theorem, there are about $N/(\log N)$ primes in Π_N. In this case, the qualitative analogue of Theorem 1.2.1 is given by Dirichlet's theorem on primes in arithmetic progressions,[2] which implies that, for any fixed $q \geqslant 1$, the random variables π_q on Π_N converge in law to the probability measure on $\mathbf{Z}/q\mathbf{Z}$ which is the uniform measure on the subset $(\mathbf{Z}/q\mathbf{Z})^\times$ of invertible residue classes (this change of the measure compared with the case of integers is simply due to the obvious fact that at most one prime may be divisible by q).

[2] More precisely, by its strong variant.

It is *expected* that a bound similar to (1.1) should be true. More precisely, there *should* exist a constant $C \geqslant 0$ such that

$$\left|\mathbf{E}_{\Pi_N}(f(\pi_q)) - \mathbf{E}(f)\right| \leqslant \frac{C(\log qN)^2}{\sqrt{N}}\|f\|_1, \tag{1.5}$$

but that statement is very close to the Generalized Riemann Hypothesis for Dirichlet L-functions.[3] Even a similar bound with $\sqrt{N}$ replaced by N^θ for any fixed $\theta > 0$ is not known, and would be a sensational breakthrough. Note that here the function f is defined on $(\mathbf{Z}/q\mathbf{Z})^\times$ and we have

$$\mathbf{E}(f) = \frac{1}{\varphi(q)} \sum_{a \in (\mathbf{Z}/q\mathbf{Z})^\times} f(a),$$

with $\varphi(q) = |(\mathbf{Z}/q\mathbf{Z})^\times|$ denoting the Euler function.

However, weaker versions of (1.5), amounting roughly to a version valid on average over $q \leqslant \sqrt{N}$, are known: the Bombieri-Vinogradov Theorem states that, for any constant $A > 0$, there exists $B > 0$ such that we have

$$\sum_{q \leqslant \sqrt{N}/(\log N)^B} \max_{a \in (\mathbf{Z}/q\mathbf{Z})^\times} \left|\mathbf{P}_{\Pi_N}(\pi_q = a) - \frac{1}{\varphi(q)}\right| \ll \frac{1}{(\log N)^A}, \tag{1.6}$$

where the implied constant depends only on A. In many applications, this is essentially as useful as (1.5).

Exercise 1.3.1. Compute the "probability" that $p-1$ be squarefree, for p prime. (This can be done using the Bombieri-Vinogradov theorem, but in fact also using the weaker Siegel-Walfisz Theorem).

[**Further references:** Friedlander and Iwaniec [10]; Iwaniec and Kowalski [15].]

1.3.2. Random walks. A more recent (and extremely interesting) type of problem arises from taking measures on $\mathbf{Z}$ derived from *random walks* on certain discrete groups. For simplicity, we only consider a special case. Let $m \geqslant 2$ be an integer, and let $G = \mathrm{SL}_m(\mathbf{Z})$ be the group of $n \times n$ matrices with integral coefficients and determinant 1. This is a complicated infinite (countable) group, but it is known to have finite generating sets. We fix one such set S, and assume that $1 \in S$ and $S = S^{-1}$ for convenience. (A well-known example is the set S consisting of 1, the elementary matrices $1 + E_{i,j}$ for $1 \leqslant i \neq j \leqslant m$, where $E_{i,j}$ is the matrix where only the (i,j)-th coefficient is non-zero, and equal to 1, and their inverses $1 - E_{i,j}$).

The generating set S defines then a random walk $(\gamma_n)_{n\geqslant 0}$ on G: let $(\xi_n)_{n\geqslant 1}$ be a sequence of independent S-valued random variables (defined on some probability space Ω) such that $\mathbf{P}(\xi_n = s) = 1/|S|$ for all n and all $s \in S$. Then we let

$$\gamma_0 = 1, \qquad \gamma_{n+1} = \gamma_n \xi_{n+1}.$$

Fix some (non-constant) polynomial function F of the coefficients of an element $g \in G$ (so $F \in \mathbf{Z}[(g_{i,j})]$), for instance $F(g) = (g_{1,1})$, or $F(g) = \mathrm{Tr}(g)$ for $g = (g_{i,j})$ in G. We can then study the analogue of Theorem 1.2.1 when applied to the random variables $\pi_q(F(\gamma_n))$ as $n \to +\infty$, or in other words, the distribution of $F(g)$ modulo q, as g varies in G according to the distribution of the random walk.

Let $G_q = \mathrm{SL}_m(\mathbf{Z}/q\mathbf{Z})$ be the finite special linear group. It is an elementary exercise, using finite Markov chains and the surjectivity of the projection map $G \longrightarrow G_q$, to check that the sequence of random variables $(\pi_q(F(\gamma_n)))_{n\geqslant 0}$ converges in law as $n \to +\infty$. Indeed, its limit is a random variable F_q on $\mathbf{Z}/q\mathbf{Z}$ defined by

$$\mathbf{P}(F_q = x) = \frac{1}{|G_q|}|\{g \in G_q \mid F(g) = x\}|,$$

[3] It implies it for non-trivial Dirichlet characters.

for all $x \in \mathbf{Z}/q\mathbf{Z}$, where we view F as also defining a function $F : G_q \longrightarrow \mathbf{Z}/q\mathbf{Z}$ modulo q. In other words, F_q is distributed like the direct image under F of the uniform measure on G_q.

In fact, elementary Markov chain theory (or direct computations) shows that there exists a constant $c_q > 1$ such that for any function $f : G_q \longrightarrow \mathbf{C}$, we have

$$\left|\mathbf{E}(f(\pi_q(\gamma_n)) - \mathbf{E}(f)\right| \leqslant \frac{\|f\|_1}{c_q^n}, \tag{1.7}$$

in analogy with (1.1), with

$$\|f\|_1 = \sum_{g \in G_q} |f(g)|.$$

This is a very good result for a fixed q (note that the number of elements reached by the random walk after n steps also grows exponentially with n). For appplications, our previous discussion already shows that it will be important to exploit (1.7) for q varying with n, and uniformly over a wide range of q. This requires an understanding of the variation of the constant c_q with q. It is a rather deep fact (Property (τ) of Lubotzky for $\mathrm{SL}_2(\mathbf{Z})$, and Property (T) of Kazhdan for $\mathrm{SL}_m(\mathbf{Z})$ if $m \geqslant 3$) that there exists $c > 1$, depending only on m, such that $c_q \geqslant c$ for all $q \geqslant 1$. Thus we do get a uniform bound

$$\left|\mathbf{E}(f(\pi_q(\gamma_n)) - \mathbf{E}(f)\right| \leqslant \frac{\|f\|_1}{c^n}$$

valid for all $n \geqslant 1$ and all $q \geqslant 1$. This is related to the theory (and applications) of *expander graphs*.

[**Further references:** Breuillard and Oh [6], Kowalski [18], [20].]

1.4. Outline of the notes

Here is now a quick outline of the main results that we will prove in the text. For detailed statements, we refer to the introductory sections of the corresponding chapters.

Chapter 2 presents different aspects of the Erdős-Kac Theorem. This is a good example to begin with because it is the most natural starting point for probabilistic number theory, and it remains quite a lively topic of contemporary research. This leads to natural appearances of the normal distribution as well as the Poisson distribution.

Chapter 3 is concerned with the distribution of values of the Riemann zeta function. We discuss results outside of the critical line (due to Bohr-Jessen, Bagchi and Voronin) as well as on the critical line (due to Selberg), and again attempt to view these in a consistent manner. The limit theorems one obtains have often quite unorthodox limiting distributions (random Euler products, sometimes viewed as random functions, and – conjecturally – also eigenvalues of random unitary matrices of large size).

In Chapter 4, we consider the distribution, in the complex plane, of polygonal paths joining partial sums of Kloosterman sums, following recent work of the author and W. Sawin [24]. Here we will use convergence in law in Banach spaces and some elementary probability in Banach spaces, and the limit object that arises will be a very special random Fourier series.

In all of these chapters, we discuss in detail a specific example of fairly general settings or theories: just the additive function $\omega(n)$ instead of more general additive functions, just the Riemann zeta function instead of more general L-functions, and specific families of exponential sums. However, we mention some of the natural generalizations of the results presented, as in Section 1.3 in this chapter.

In Chapter 5, we survey more briefly, and without full proofs, some additional natural instances of probabilistic number theory. This includes a discussion of the distribution of gaps between primes (where the Poisson distribution appears again, as well as some particular random Euler products), and more "algebraic" questions, for instance concerning the behavior of "random" number fields, or "random" algebraic curves.

1.5. What we do not talk about

There are many more interactions between probability theory and number theory than what we have space to discuss. Here are some examples, with references where interested readers may find out more about them. We order them, roughly speaking, in terms of how far they may look from our perspective.

- Application of limit theorems of the type we discuss to other problems of analytic number theory. We will give a few examples, but this is not our main concern.
- Using probabilistic ideas to *model* arithmetic objects, and make conjectures or prove theorems concerning those; in contrast we our point of view, it is not expected in such cases that there exist actual limit theorems comparing the model with the actual arithmetic phenomena. A typical example is the co-called Cramer model for the distribution of primes.
- Using number theoretic ideas to *derandomize* certain constructions or algorithms. There are indeed a number of very interesting results that use the randomness of specific arithmetic objects to give deterministic constructions, or deterministic proofs of existence, for mathematical objects that might have first been shown to exist using probabilistic ideas. Examples include the construction of expander graphs by Margulis, or of Ramanujan graphs by Lubotzky, Phillips and Sarnak, or in different vein, the construction of explicit "ultraflat" trigonometric polynomials (in the sense of Kahane) by Bombieri and Bourgain.

Prerequisites and notation

The basic requirements for most of this text are standard introductory graduate courses in algebra, analysis (including Lebesgue integration and complex analysis) and probability. Of course, knowledge and familiarity with basic number theory (for instance, the distribution of primes up to the Bombieri-Vinogradov Theorem) are helpful, but we review in Appendix C all the basic results we use. Similarly, Appendix B summarizes the notation and facts from probability theory which are the most important for us.

We will use the following notation:

(1) A compact topological space is always assumed to be separated.
(2) For a set X, $|X| \in [0, +\infty]$ denotes its cardinal, with $|X| = \infty$ if X is infinite. There is no distinction in this text between the various infinite cardinals.
(3) If X is a set and f, g two complex-valued functions on X, then we write synonymously $f = O(g)$ or $f \ll g$ to say that there exists a constant $C \geqslant 0$ (sometimes called an "implied constant") such that $|f(x)| \leqslant g(x)$ for all $x \in X$. Note that this implies that in fact $g \geqslant 0$. We also write $f \asymp g$ to indicate that $f \ll g$ and $g \ll f$.
(4) If X is a topological space, $x_0 \in X$ and f and g are functions defined on a neighborhood of x_0, with $g(x) \neq 0$ for x in a neighborhood of x_0, then we say that $f(x) = o(g(x))$ as $x \to x_0$ if $f(x)/g(x) \to 0$ as $x \to x_0$, and that $f(x) \sim g(x)$ as $x \to x_0$ if $f(x)/g(x) \to 1$.
(5) We write $a \mid b$ for the divisibility relation "a divides b".
(6) We denote by $\mathbf{F}_p$ the finite field $\mathbf{Z}/p\mathbf{Z}$, for p prime, and more generally by $\mathbf{F}_q$ a finite field with q elements, where $q = p^n$, $n \geqslant 1$, is a power of p. We will recall the properties of finite fields when we require them.
(7) For a complex number z, we write $e(z) = e^{2i\pi z}$. If $q \geqslant 1$ and $x \in \mathbf{Z}/q\mathbf{Z}$, then $e(x/q)$ is then well-defined by taking any representative of x in $\mathbf{Z}$ to compute the exponential.
(8) If $q \geqslant 1$ and $x \in \mathbf{Z}$ (or $x \in \mathbf{Z}/q\mathbf{Z}$) is an integer which is coprime to q (or a residue class invertible modulo q), we sometimes denote by $\bar{q}$ the inverse class such that $x\bar{x} = 1$ in $\mathbf{Z}/q\mathbf{Z}$. This will always be done in such a way that the modulus q is clear from context, in the case where x is an integer.
(9) Given a probability space $(\Omega, \Sigma, \mathbf{P})$, we denote by $\mathbf{E}(\cdot)$ (resp. $\mathbf{V}(\cdot)$) the expectation (resp. the variance) computed with respect to $\mathbf{P}$. It will often happen (as already

above) that we have a sequence $(\Omega_N, \Sigma_N, \mathbf{P}_N)$ of probability spaces; we will then denote by $\mathbf{E}_N$ or $\mathbf{V}_N$ the respective expectation and variance with respect to $\mathbf{P}_N$.

(10) Given a measure space (Ω, Σ, μ) (not necessarily a probability space), a set Y with a σ-algebra Σ' and a measurable map $f : \Omega \longrightarrow Y$, we denote by $f_*(\mu)$ (or sometimes $f(\mu)$) the image measure on Y; in the case of a probability space, so that f is seen as a random variable on Ω, this is the probability law of f seen as a "random Y-valued element". If the set Y is given without specifiying a σ-algebra, we will view it usually as given with the σ-algebra generated by sets $Z \subset Y$ such that $f^{-1}(Z)$ belongs to Σ.

(11) As a typographical convention, we will often use sans-serif fonts like X to denote arithmetically defined random variables, and standard font X for an "abstract" random variable that is often somehow related.

Acknowledgments. The first version of these notes were prepared for a course "Introduction to probabilistic number theory" that I taught at ETH Zürich during the Fall Semester 2015. Thanks to the students of the course for their interest, in particular to M. Gerspach for sending corrections, and to B. Löffel for organizing and writing the exercise sessions.

CHAPTER 2

The Erdős-Kac principle

2.1. The basic Erdős-Kac Theorem

We begin by recalling the statement (see Theorem 1.1.1), in its probabilistic phrasing:

THEOREM 2.1.1 (Erdős-Kac Theorem). *For $N \geqslant 1$, let $\Omega_N = \{1, \dots, N\}$ with the uniform probability measure $\mathbf{P}_N$. Let X_N be the random variable*

$$n \mapsto \frac{\omega(n) - \log\log N}{\sqrt{\log\log N}}$$

on Ω_N for $N \geqslant 3$. Then $(\mathsf{X}_N)_{N\geqslant 3}$ converges in law to a standard normal random variable, i.e., to a normal random variable with expectation 0 and variance 1.

Figure 2.1 shows a plot of the density of X_N for $N = 10^{10}$: one can see something that could be the shape of the gaussian density appearing, but the fit is very far from perfect.

The original proof is due to Erdős and Kac in 1939 [7]. We will explain a proof following the work of Granville and Soundararajan [13] and of Billingsley [3, p. 394]. The presentation emphasizes the general probabilistic nature of the argument, so that generalizations will be easily derived in the next section.

We will prove convergence in law using the method of moments, as explained in Section B.2 of Appendix B, specifically in Theorem B.3.5 and Remark B.3.8.

We first outline the different steps of the proof and the intuition behind them:

(1) We will show, using Theorem 1.2.1, that for any fixed integer $k \geqslant 0$, we have

$$\mathbf{E}_N(\mathsf{X}_N^k) = \mathbf{E}(X_N^k) + o(1),$$

where X_N is a sequence of normalized random variables of the form

$$X_N = \frac{Z_N - \mathbf{E}(Z_N)}{\sqrt{\mathbf{V}(Z_N)}}$$

with

$$Z_N = \sum_{p\leqslant N} B_p, \tag{2.1}$$

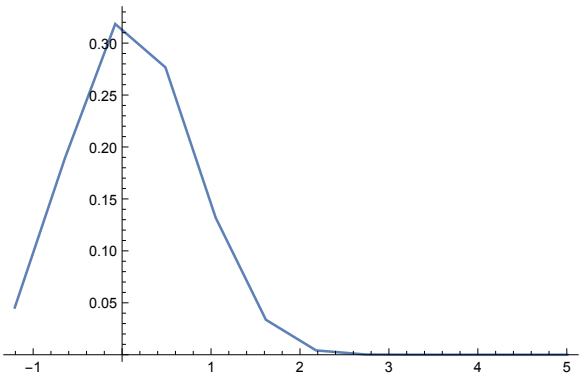

FIGURE 2.1. The normalized number of prime divisors for $n \leqslant 10^{10}$.

for some *independent* Bernoulli random variables $(B_p)_p$ indexed by primes and defined on some auxiliary probability space. The intuition behind this fact is quite simple: by definition, we have

$$\omega(n) = \sum_{p|n} 1 = \sum_p \mathsf{B}_p(n) \tag{2.2}$$

where $\mathsf{B}_p : \mathbf{Z} \longrightarrow \{0,1\}$ is the characteristic function of the multiples of p. In other words, B_p is obtained by composing the reduction map $\pi_p : \mathbf{Z} \longrightarrow \mathbf{Z}/p\mathbf{Z}$ with the characteristic function of the residue class $\{0\} \subset \mathbf{Z}/p\mathbf{Z}$. For $n \in \Omega_N$, it is clear that $\mathsf{B}_p(n) = 0$ unless $p \leqslant N$, and on the other hand, Proposition 1.2.5 (applied to the different primes $p \leqslant N$) suggests that B_p, viewed as random variables on Ω_N, are "almost" independent, while for N large, each B_p is "close" to a Bernoulli random variable B_p with

$$\mathbf{P}(B_p = 1) = \frac{1}{p}, \qquad \mathbf{P}(B_p = 0) = 1 - \frac{1}{p}.$$

(2) The Central Limit Theorem applies to the sequence (X_N), and shows that it converges in law to a standard normal random variable $\mathcal{N}$. This is not quite the most standard form of the Central Limit Theorem, since the summands B_p defining X_N are not identically distributed, but it is nevertheless a very simple and well-known case (see Theorem B.5.1).

(3) It follows that

$$\lim_{N\to+\infty} \mathbf{E}_N(\mathsf{X}_N^k) = \mathbf{E}(\mathcal{N}^k),$$

and hence, by the method of moments (Theorem B.3.5), we conclude that X_N converges in law to $\mathcal{N}$. (Interestingly, we do not need to know the value of the moments $\mathbf{E}(\mathcal{N}^k)$ for this argument to apply.)

This sketch shows that the Erdős-Kac Theorem is really a result of very general nature. Note that only Step 1 has real arithmetic content. As we will see, that arithmetic content is concentrated on two results: Theorem 1.2.1, which makes the link with probability theory, and the basic Mertens estimate from prime number theory

$$\sum_{p\leqslant N} \frac{1}{p} = \log\log N + O(1)$$

for $N \geqslant 3$ (see Proposition C.1.1 in Appendix C). Indeed, this is where the normalization of the arithmetic function $\omega(n)$ comes from, and one could essentially dispense with this ingredient by replacing the factors $\log\log N$ in Theorem 2.1.1 by

$$\sum_{p\leqslant N} \frac{1}{p}.$$

In order to prove such a statement, one only needs to know that this quantity tends to ∞ as $N \to +\infty$, which follows from the most basic statements concerning the distribution of primes, due to Chebychev.

We now implement those steps. As will be seen, some tweaks will be required. (The reader is invited to check that omitting those tweaks leads, at the very least, to a much more complicated-looking problem!).

Step 1. (Truncation) This is a classical technique that applies here, and is used to shorten and simplify the sum in (2.1), in order to control the error terms in Step 2. We consider the random variables B_p on Ω_N as above, i.e., $\mathsf{B}_p(n) = 1$ if p divides n and $\mathsf{B}_p(n) = 0$ otherwise. Let

$$\sigma_N = \sum_{p\leqslant N} \frac{1}{p}.$$

We only need recall at this point that $\sigma_N \to +\infty$ as $N \to +\infty$. We then define

$$Q = N^{1/(\log\log N)^{1/3}} \tag{2.3}$$

and

$$\tilde{\omega}(n) = \sum_{\substack{p\mid n\\ p\leqslant Q}} 1 = \sum_{p\leqslant Q} \mathsf{B}_p(n), \qquad \tilde{\omega}_0(n) = \sum_{p\leqslant Q}\Big(\mathsf{B}_p(n) - \frac{1}{p}\Big),$$

viewed as random variables on Ω_N. The point of this truncation is the following: first, for $n \in \Omega_N$, we have

$$\tilde{\omega}(n) \leqslant \omega(n) \leqslant \tilde{\omega}(n) + (\log\log N)^{1/3},$$

simply because if $\alpha > 0$ and if $p_1, \dots, p_m$ are primes $\geqslant N^\alpha$ dividing $n \leqslant N$, then we get

$$N^{m\alpha} \leqslant p_1 \cdots p_m \leqslant N,$$

and hence $m \leqslant \alpha^{-1}$. Second, for any $N \geqslant 1$ and any $n \in \Omega_N$, we get by definition of σ_N the identity

$$\begin{aligned}\tilde{\omega}_0(n) &= \tilde{\omega}(n) - \sum_{p\leqslant Q}\frac{1}{p}\\ &= \omega(n) - \sigma_N + O((\log\log N)^{1/3})\end{aligned} \tag{2.4}$$

because the Mertens formula

$$\sum_{p\leqslant x}\frac{1}{p} = \log\log x + O(1),$$

(see Proposition C.1.1) and the definition of σ_N show that

$$\sum_{p\leqslant Q}\frac{1}{p} = \sum_{p\leqslant N}\frac{1}{p} + O(\log\log\log N) = \sigma_N + O(\log\log\log N).$$

Now define

$$\tilde{\mathsf{X}}_N(n) = \frac{\tilde{\omega}_0(n)}{\sqrt{\sigma_N}}$$

as random variables on Ω_N. We will prove that $\tilde{\mathsf{X}}_N$ converges in law to $\mathcal{N}$. The elementary Lemma B.3.3 of Appendix B (applied using (2.4)) then shows that the random variables

$$n \mapsto \frac{\omega(n) - \sigma_N}{\sqrt{\sigma_N}}$$

converge in law to $\mathcal{N}$. Finally, applying the same lemma one more time using the Mertens formula we obtain the Erdős-Kac Theorem.

It remains to prove the convergence of $\tilde{\mathsf{X}}_N$. We fix a non-negative integer k, and our target is to prove the the moment limit

$$\mathbf{E}_N(\tilde{\mathsf{X}}_N^k) \to \mathbf{E}(\mathcal{N}^k) \tag{2.5}$$

as $N \to +\infty$. Once this is proved for all k, then the method of moments shows that (X_N) converges in law to the standard normal random variable $\mathcal{N}$.

Remark 2.1.2. We might also have chosen to perform a truncation at $p \leqslant N^\alpha$ for some fixed $\alpha \in]0,1[$. However, in that case, we would need to adjust the value of α depending on k in order to obtain (2.5), and then passing from the truncated variables to the original ones would require some minor additional argument. In fact, the function $(\log\log N)^{1/3}$ used to defined the truncation could be replaced by any function going to infinity slower than $(\log\log N)^{1/2}$.

Step 2. (Moment computation) We now begin the proof of (2.5). We use the definition of $\tilde{\omega}_0(n)$ and expand the k-th power in $\mathbf{E}_N(\tilde{\mathsf{X}}_N^k)$ to derive

$$\mathbf{E}_N(\tilde{\mathsf{X}}_N^k) = \frac{1}{\sigma_N^{k/2}} \sum_{p_1 \leqslant Q} \cdots \sum_{p_k \leqslant Q} \mathbf{E}_N\Big(\Big(\mathsf{B}_{p_1} - \frac{1}{p_1}\Big) \cdots \Big(\mathsf{B}_{p_k} - \frac{1}{p_k}\Big)\Big).$$

The crucial point is that the random variable

$$\Big(\mathsf{B}_{p_1} - \frac{1}{p_1}\Big) \cdots \Big(\mathsf{B}_{p_k} - \frac{1}{p_k}\Big) \tag{2.6}$$

can be expressed as $f(\pi_q)$ for some modulus $q \geqslant 1$ and some function $f : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C}$, so that the basic result of Theorem 1.2.1 may be applied to each summand.

To be precise, the value at $n \in \Omega_N$ of the random variable (2.6) only depends on the residue class x of n in $\mathbf{Z}/q\mathbf{Z}$, where q is the least common multiple of $p_1, \ldots, p_k$. In fact, this value is equal to $f(x)$ where

$$f(x) = \Big(\delta_{p_1}(x) - \frac{1}{p_1}\Big) \cdots \Big(\delta_{p_k}(x) - \frac{1}{p_k}\Big)$$

with δ_{p_i} denoting the characteristic function of the residues classes modulo q which are 0 modulo p_i. It is clear that $|f(x)| \leqslant 1$, as product of terms which are all $\leqslant 1$, and hence we have the bound

$$\|f\|_1 \leqslant q$$

(this is extremely imprecise, as we will see later). From this we get

$$\Big|\mathbf{E}_N\Big(\Big(\mathsf{B}_{p_1} - \frac{1}{p_1}\Big) \cdots \Big(\mathsf{B}_{p_k} - \frac{1}{p_k}\Big)\Big) - \mathbf{E}(f)\Big| \leqslant \frac{2q}{N} \leqslant \frac{2Q^k}{N}$$

by Theorem 1.2.1.

But by the definition of f, we also see that

$$\mathbf{E}(f) = \mathbf{E}\Big(\Big(B_{p_1} - \frac{1}{p_1}\Big) \cdots \Big(B_{p_k} - \frac{1}{p_k}\Big)\Big)$$

where the random variables (B_p) form a sequence of *independent* Bernoulli random variables with $\mathbf{P}(B_p = 1) = 1/p$ (the (B_p) for p dividing q are realized concretely as the characteristic functions δ_p on $\mathbf{Z}/q\mathbf{Z}$ with uniform probability measure).

Therefore we derive

$$\begin{aligned}
\mathbf{E}_N(\tilde{\mathsf{X}}_N^k) &= \frac{1}{\sigma_N^{k/2}} \sum_{p_1 \leqslant Q} \cdots \sum_{p_k \leqslant Q} \mathbf{E}\Big(\Big(B_{p_1} - \frac{1}{p_1}\Big) \cdots \Big(B_{p_k} - \frac{1}{p_k}\Big)\Big) + O(Q^k N^{-1})\Big) \\
&= \Big(\frac{\tau_N}{\sigma_N}\Big)^{k/2} \mathbf{E}(X_N^k) + O(Q^{2k} N^{-1}) \\
&= \Big(\frac{\tau_N}{\sigma_N}\Big)^{k/2} \mathbf{E}(X_N^k) + o(1)
\end{aligned}$$

by our choice (2.3) of Q, where

$$\tau_N = \sum_{p \leqslant Q} \frac{1}{p}\Big(1 - \frac{1}{p}\Big) = \sum_{p \leqslant Q} \mathbf{V}(B_p)$$

and

$$X_N = \frac{1}{\sqrt{\tau_N}} \sum_{p \leqslant Q} \Big(B_p - \frac{1}{p}\Big).$$

Step 3. (Conclusion) We now note that the version of the Central Limit Theorem recalled in Theorem B.5.1 applies to the random variables (B_p), and implies precisely that X_N converges in law to $\mathcal{N}$. But moreover, the sequence (X_N) satisfies the uniform integrability assumption

in the converse of the method of moments (see Theorem B.3.6 (2), applied to the variables $B_p - 1/p$, which are bounded by 1 and independent), and hence we have in particular

$$\mathbf{E}(X_N^k) \longrightarrow \mathbf{E}(\mathcal{N}^k).$$

Since $\tau_N \sim \sigma_N$ by the Mertens formula, we deduce that $\mathbf{E}_N(\tilde{\mathrm{X}}_N^k)$ converges also to $\mathbf{E}(\mathcal{N}^k)$, which was our desired goal (2.5).

Exercise 2.1.3. One can avoid appealing to the converse of the method of moments by directly using proofs of the Central Limit Theorem based on the moments that give directly the convergence of moments for (X_N). Find such a proof in this special case. (See for instance [3, p. 391]; note that this requires knowledge of the moments of gaussian random variables, which we recall in Proposition B.5.2).

Exercise 2.1.4. For an integer $N \geqslant 1$, let $m(N)$ denote the set of integers that occur in the multiplication table for integers $1 \leqslant n \leqslant N$:

$$m(N) = \{k = ab \mid 1 \leqslant a \leqslant N, \quad 1 \leqslant b \leqslant N\} \subset \Omega_{N^2}.$$

Prove that $\mathbf{P}(m(N)) = 0$, i.e., that

$$\lim_{N \to +\infty} \frac{|m(N)|}{N^2} = 0.$$

This result is the basic statement concerning the "multiplication table" problem of Erdős; the precise asymptotic behavior of $|m(N)|$ has been determined by K. Ford [9] (improving results of Tenenbaum): we have

$$\frac{|m(N)|}{N^2} \asymp (\log N)^{-\alpha} (\log\log N)^{-3/2}$$

where

$$\alpha = 1 - \frac{1 + \log\log 2}{\log 2}.$$

See also the work of Koukoulopoulos [17] for generalizations.

Exercise 2.1.5. Let $\Omega(n)$ be the number of prime divisors of an integer $n \geqslant 1$, counted *with* multiplicity (so $\Omega(12) = 3$).[1] Prove that

$$\mathbf{P}_N\Big(\Omega(n) - \omega(n) \geqslant (\log\log N)^{1/4}\Big) \leqslant (\log\log N)^{-1/4},$$

and deduce that the random variables

$$n \mapsto \frac{\Omega(n) - \log\log N}{\sqrt{\log\log N}}$$

also converge in law to $\mathcal{N}$.

2.2. Generalizations

Reviewing the proof of Theorem 2.1.1, we see that, as promised when explaining the intuitive idea behind the result, very little arithmetic information was used. More precisely:

- In Step 1, the truncation is guided by the fact that Ω_N (the support of $\mathbf{P}_N$) consists of integers $\leqslant N$;
- In Step 2, we appealed to Theorem 1.2.1, which gives the asymptotic distribution of integers $n \in \Omega_N$ modulo q for q relatively large compared with N – again, N appears as related to the size of integers in the support of $\mathbf{P}_N$;
- Finally, in Step 3, we use the fact that the quantities σ_N tend to infinity as N grows, and to be precise that $\sigma_N \sim \log\log N$ (but this precise order of growth is, to some extent, just an additional precision that is nice to have but not completely essential to obtain an interesting statement).

[1] We only use this function in this section and hope that confusion with Ω_N will be avoided.

From this, it is not difficult to obtain a wide-ranging generalization of Theorem 2.1.1 to other sequences of integer-supported random variables. This is enough to deal for instance with the measures associated to irreducible polynomials $P \in \mathbf{Z}[X]$ (as in Exercise 1.2.4), or to the random walks on groups discussed in Section 1.3.2. However, we might want to consider the also the distribution of $\omega(p-1)$, for p prime, which is related instead to Section 1.3.1, and in this situation we lack the analogue of Theorem 1.2.1, as we observed (to be more precise, it would only follow from a close relative of the Generalized Riemann Hypothesis). But the proof of Step 2 shows that we are dealing there with how close the reduction of n modulo q, for $n \in \Omega_N$, to the limiting distribution μ_q, but rather *to an average over* q (namely, over q arising as the lcm of primes $p_1, \dots, p_k \leqslant Q$). And, for primes, we *do* have some control over such an average by means of the Bombieri-Vinogradov estimates (1.6)!

Building on these observations, we can prove a very general abstract form of the Erdős-Kac Theorem. (In a first reading, readers should probably just glance at the statement, and skip to the examples, or to the next section or chapter, without looking at the proof, which is something of an *exercice de style*).

We begin by defining suitable sequences of random variables.

Definition 2.2.1. A *balanced random integer* is a sequence $(X_N)_{N\geqslant 1}$ of random variables with values in $\mathbf{Z}$ that satisfy the following properties:

(1) The expectation $m_N = \mathbf{E}(|X_N|)$ is finite, $m_N \geqslant 1$, and m_N tends to infinity as $N \to +\infty$;

(2) There exist a real number $\theta > 0$ and, for each integer $q \geqslant 1$, there exists a probability measure μ_q on $\mathbf{Z}/q\mathbf{Z}$ such that for any $A > 0$ and $N \geqslant 1$, we have

$$\sum_{q\leqslant m_N^\theta} \max_{\|f_q\|_1\leqslant 1} \Big|\mathbf{E}(f_q(\pi_q(X_N))) - \mathbf{E}_{\mu_q}(f)\Big| \ll \frac{1}{(\log 2m_N)^A}$$

where f_q ranges over all functions $f_q : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C}$ with $\|f_q\| \leqslant 1$, and the implied constant depends on A.

Our main result essentially states that under suitable conditions, the number of prime factors of a balanced random integer tends to a standard normal random variable. Here it will be useful to use the convention $\omega(0) = 0$.

Theorem 2.2.2 (General Erdős-Kac theorem). *Let (X_N) be a balanced random integer. Assume furthermore that*

(1) ...

Then we have convergence in law

$$\frac{\omega(X_N) - \sigma_N}{\sqrt{\sigma_N}} \longrightarrow \mathcal{N}$$

as $N \to +\infty$.

Proof. We first assume that X_N takes values in $\mathbf{Z} - \{0\}$. For a given N, we define

$$\tau_N = \sum_{p\leqslant m_N} \mu_p(0)$$

and

$$Q = m_N^{1/\tau_N^{1/3}}.$$

The assumption (**??**) ensures that $Q \to +\infty$ as $N \to +\infty$. We then define the truncation

$$\omega_{0,N} = \sum_{p\leqslant Q} (\mathsf{B}_p - \mathbf{E}(\mathsf{B}_p))$$

where B_p is the Bernoulli random variable equal to 1 if and only if $p \mid X_N$. We claim next that if we define

$$\sigma_N = \sum_{p\leqslant Q} \mu_p(0),$$

then $\omega_{0,N}/\sqrt{\sigma_N}$ converges in law to $\mathcal{N}$.

To see this, we apply the method of moments. For a fixed integer $k \geqslant 0$, we have

$$\mathbf{E}\Big(\Big(\frac{\omega_{0,N}}{\sqrt{\sigma_N}}\Big)^k\Big) = \frac{1}{\sigma_N^{k/2}} \sum_{p_1 \leqslant Q} \cdots \sum_{p_k \leqslant Q} \mathbf{E}\Big((\mathsf{B}_{p_1} - \mathbf{E}(\mathsf{B}_{p_1})) \cdots (\mathsf{B}_{p_k} - \mathbf{E}(\mathsf{B}_{p_k}))\Big).$$

We rearrange the sum according to the value q of the lcm $[p_1, \ldots, p_k]$. For each $\boldsymbol{p} = (p_1, \ldots, p_k)$, there is a function $f_{\boldsymbol{p}} : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{R}$ such that

$$(\mathsf{B}_{p_1} - \mathbf{E}(\mathsf{B}_{p_1})) \cdots (\mathsf{B}_{p_k} - \mathbf{E}(\mathsf{B}_{p_k})) = f_{\boldsymbol{p}}(\pi_q(X_N)).$$

Indeed, if $\nu(p)$ is the multiplicity of a divisor p of q in the tuple $\boldsymbol{p}$, then we have

$$f_{\boldsymbol{p}}(x) = \prod_{p|q} (\delta_p(x) - \mathbf{E}(\mathsf{B}_p))^{\nu(p)}.$$

Finally, we must deal with the case of random integers (X_N) that may take the value 0. We do this by replacing the sequence (X_N) by $(\tilde{X}_N)$ where $\tilde{X}_N$ is the restriction of X_N to the subset $\tilde{\Omega}_N = \{X_N \neq 0\}$ of the original probability space, which is equiped with the conditional probability measure

$$\tilde{\mathbf{P}}_N(A) = \frac{\mathbf{P}(A)}{\mathbf{P}(\tilde{\Omega}_N)}.$$

We then have

$$\frac{\omega(\tilde{X}_N) - \tilde{\sigma}_N}{\sqrt{\tilde{\sigma}_N}} \longrightarrow \mathcal{N}$$

by the first case, and the result follows for (X_N) because

$$\mathbf{P}(X_N \neq 0) \longrightarrow 1.$$

□

2.3. Convergence without renormalization

One important point that is made clear by the proof of the Erdős-Kac Theorem is that, although one might think that a statement about the behavior of the number of primes factors of integers tells us something about the distribution of primes (which are those integers n with $\omega(n) = 1$), the Erdős-Kac Theorem *gives no information about these*. This can be seen mechanically from the proof (where the truncation step means in particular that primes are disregarded unless they are smaller than the truncation level Q), or intuitively from the fact that the statement itself implies that "most" integers of size about N have $\log\log N$ prime factors. For instance, we have

$$\mathbf{P}_N\Big(|\omega(n) - \log\log N| > a\sqrt{\log\log N}\Big) \longrightarrow \mathbf{P}(|\mathcal{N}| > a) \leqslant \sqrt{\frac{2}{\pi}} \int_a^{+\infty} e^{-x^2/2} dx \leqslant e^{-a^2/4},$$

as $N \to +\infty$.

The problem lies in the normalization used to obtain a definite theorem of convergence in law: this "crushes" to some extent the more subtle aspect of the distribution of values of $\omega(n)$, especially with respect to extreme values. One can however still study this function probabilistically, but one must use less generic methods, to go beyond the "universal" behavior given by the Central Limit Theorem. There are at least two possible approaches in this direction, and we now briefly survey some of the results.

Both methods have in common a switch in probabilistic focus: instead of looking for a normal approximation of a normalized version of $\omega(n)$, one looks for a *Poisson approximation* of the unnormalized function.

Recall (see also Section B.7 in the Appendix) that a Poisson distribution with real parameter $\lambda \geqslant 0$ satisfies

$$\mathbf{P}(\lambda = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

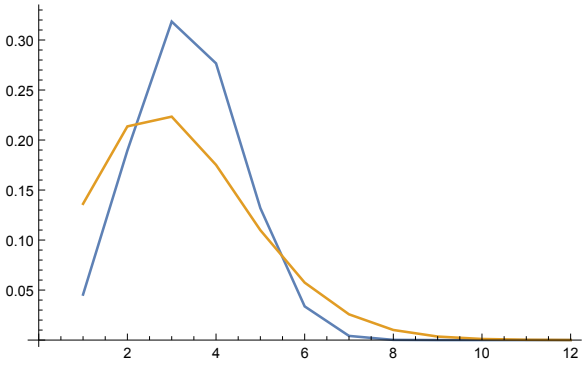

FIGURE 2.2. The number of prime divisors for $n \leqslant 10^{10}$ (blue) compared with a Poisson distribution.

for any integer $k \geqslant 0$. It turns out that an inductive computation using the Prime Number Theorem leads to the asymptotic formula

$$\frac{1}{N}|\{n \leqslant N \mid \omega(n) = k\}| \sim \frac{1}{(k-1)!}\frac{(\log\log N)^{k-1}}{\log N} = e^{-\log\log N}\frac{(\log\log N)^{k-1}}{(k-1)!},$$

for any fixed integer $k \geqslant 1$. This suggests that a better probabilistic approximation to the arithmetic function $\omega(n)$ on Ω_N is a Poisson distribution with parameter $\log\log N$. The Erdős-Kac Theorem would then be, in essence, a consequence of the simple fact that a sequence (X_n) of Poisson random variables with parameters $\lambda_n \to +\infty$ has the property that

$$\frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \to \mathcal{N}, \tag{2.7}$$

as explained in Proposition B.7.1. Figure 2.2 shows the density of the values of $\omega(n)$ for $n \leqslant 10^{10}$ and the corresponding Poisson density. (The values of the probabilities for consecutive integers are joined by line segments for readability).

The trick to make this precise is to give a meaning to this statement, which can not be a straightforward convergence statement since the parameter is varying with N.

Harper [14] (to the author's knowledge) was the first to implement explicitly such an idea. He derived an explicit upper-bound for the *total variation distance* between a truncated version of $\omega(n)$ on Ω_N and a suitable Poisson random variable, namely between

$$\sum_{\substack{p|n\\ p\leqslant Q}} 1, \quad \text{where } Q = N^{1/(3\log\log N)^2}$$

and a Poisson random variable Po_N with parameter

$$\lambda_N = \sum_{p\leqslant Q}\frac{1}{N}\left\lfloor\frac{N}{p}\right\rfloor$$

(so that the Mertens formula implies that $\lambda_N \sim \log\log N$).

Precisely, Harper proves that for *any* subset A of the non-negative integers, we have

$$\left|\mathbf{P}_N\Big(\sum_{\substack{p|n\\ p\leqslant Q}} 1 \in A\Big) - \mathbf{P}(\mathrm{Po}_N \in A)\right| \ll \frac{1}{\log\log N},$$

and moreover that the decay rate $(\log\log N)^{-1}$ is best possible. This requires some additional arithmetic information than the proof of Theorem 2.1.1 (essentially some form of sieve), but the arithmetic ingredients remain to a large extent elementary. On the other hand, new ingredients from probability theory are involved, especially cases of Stein's Method for Poisson approximation.

A second approach starts from a proof of the Erdős-Kac Theorem due to Rényi and Turán [29], which is the implementation of the Lévy Criterion for convergence in law. Precisely, they prove that

$$\mathbf{E}_N(e^{it\omega(n)}) = (\log N)^{e^{it}-1}(\Phi(t)+o(1)) \tag{2.8}$$

for any $t \in \mathbf{R}$ as $N \to +\infty$ (in fact, uniformly for $t \in \mathbf{R}$ – note that the function here is 2π-periodic), with a factor $\Phi(t)$ given by

$$\Phi(t) = \frac{1}{\Gamma(e^{it})}\prod_p \Big(1-\frac{1}{p}\Big)^{e^{it}}\Big(1+\frac{e^{it}}{p-1}\Big),$$

where the Euler product is absolutely convergent. Recognizing that the term $(\log N)^{e^{it}-1}$ is the characteristic function of a Poisson random variable Po_N with parameter $\log\log N$, one can then obtain the Erdős-Kac Theorem by the same computation that leads to (2.7), combined with the continuity of Φ that shows that

$$\Phi\Big(\frac{t}{\sqrt{\log\log N}}\Big) \longrightarrow \Phi(0) = 1$$

as $N \to +\infty$.

The computation that leads to (2.8) is now interpreted as an instance of the Selberg-Delange method (see [33, II.5, Th. 3] for the general statement, and [33, II.6, Th. 1] for the special case of interest here).

It should be noted that the proof of (2.8) is quite a bit deeper than the proof of Theorem 2.1.1, and this is at it should, because this formula contains precise information about the extreme values of $\omega(n)$, which we saw are not relevant to the Erdős-Kac Theorem. Indeed, taking $t = \pi$ and observing that $\Phi(\pi) = 0$ (because of the pole of the Gamma function), we obtain

$$\frac{1}{N}\sum_{n\leqslant N}(-1)^{\omega(n)} = \mathbf{E}(e^{-i\pi\omega(n)}) = o\Big(\frac{1}{(\log N)^2}\Big)$$

This is well-known to imply the Prime Number Theorem

$$\sum_{p\leqslant N} 1 \sim \frac{N}{\log N}$$

(see, for instance [15, §2.1], where the Möbius function is used instead of $n \mapsto (-1)^{\omega(n)}$, noting that these functions coincide on squarefree integers).

The link between the formula (2.8) and Poisson distribution was noticed in joint work with Nikeghbali [23]. Among other things, we remarked that it implies easily a bound for the Kolmogorov-Smirnov distance between $n \mapsto \omega(n)$ on Ω and a Poisson random variable Po_N. Additional work with A. Barbour [2] leads to bounds in total variation distance, and to better (non-Poisson) approximations of even better quality. Another suggestive remark is that if we consider the independent random variables that appear in the proof of the Erdős-Kac theorem, namely

$$X_N = \sum_{p\leqslant N}\Big(B_p - \frac{1}{p}\Big),$$

where (B_p) is a sequence of independent Bernoulli random variables with $\mathbf{P}(B_p = 1) = 1/p$, then we have (by a direct computation) the following analogue of (2.8):

$$\mathbf{E}(e^{itX_N}) = (\log N)^{e^{it}-1}\Big(\prod_p\Big(1-\frac{1}{p}\Big)^{e^{it}}\Big(1+\frac{e^{it}}{p-1}\Big)+o(1)\Big).$$

It is natural to ask then if there is a similar meaning to the factor $1/\Gamma(e^{it})$ that also appears. And there is: for $N \geqslant 1$, define ℓ_N as the random variable on the symmetric group $\mathfrak{S}_N$ that

maps a permutation σ to the number of cycles in its canonical cyclic representation. Then, giving $\mathfrak{S}_N$ the uniform probability measure, we have

$$\mathbf{E}(e^{it\ell_N}) = N^{e^{it}-1}\Big(\frac{1}{\Gamma(e^{it})} + o(1)\Big),$$

corresponding to a Poisson distribution with parameter $\log N$ this time. This is not an isolated property: see the survey paper of Granville [12] for many significant analogies between (multiplicative) properties of integers and random permutations.

REMARK 2.3.1. Observe that (2.8) would be true *if* we had a decomposition

$$\omega(n) = \mathrm{Po}_N(n) + Y_N(n)$$

as random variables on Ω_N, where Y_N is independent of Po_N and converges in law to a random variable with characteristic function Φ. However, this is not in fact the case, because Φ is not a characteristic function of a probability measure! (It is unbounded on $\mathbf{R}$).

2.4. Further reading

CHAPTER 3

The distribution of values of the Riemann zeta function

3.1. Introduction

The Riemann zeta function is defined first for complex numbers s such that $\mathrm{Re}(s) > 1$, by means of the series

$$\zeta(s) = \sum_{n \geqslant 1} \frac{1}{n^s}.$$

It plays an important role in prime number theory, arising because of the famous Euler product formula, which expresses $\zeta(s)$ as a product over primes, in this region: we have

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} \tag{3.1}$$

if $\mathrm{Re}(s) > 1$. By standard properties of series of holomorphic functions (note that $s \mapsto n^s = e^{s \log n}$ is entire for any $n \geqslant 1$), the Riemann zeta function is holomorphic for $\mathrm{Re}(s) > 1$. It is of crucial importance however that it admits an analytic continuation to $\mathbf{C} - \{1\}$, with furthermore a simple pole at $s = 1$ with residue 1.

This analytic continuation can be performed simultaneously with the proof of the *functional equation*: the function defined by

$$\Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$$

satisfies

$$\Lambda(1 - s) = \Lambda(s).$$

Since $\zeta(s)$ is quite well-behaved for $\mathrm{Re}(s) > 1$, and since the Gamma function is a very well-known function, this relation shows that one can understand the behavior of $\zeta(s)$ for s outside of the *critical strip*

$$S = \{s \in \mathbf{C} \mid 0 \leqslant \mathrm{Re}(s) \leqslant 1\}.$$

The Riemann Hypothesis is a crucial statement about $\zeta(s)$ when s is in the critical strip: it states that if $s \in S$ satisfies $\zeta(s) = 0$, then the real part of s must be $1/2$. Because holomorphic functions (with relatively slow growth, a property true for ζ, although this requires some argument to prove) are essentially characterized by their zeros (just like polynomials are!), the proof of this conjecture would enormously expand our understanding of the properties of the Riemann zeta function. Although it remains open, this should motivate our interest in the distribution of values of the zeta function.

We first focus our attention to a vertical line $\mathrm{Re}(s) = \tau$, wehre τ is a fixed real number such that $\tau \geqslant 1/2$ (the case $\tau \leqslant 1$ will be the most interesting, but some statements do not require this assumption). We consider real numbers $T \geqslant 1$ and we view

$$t \mapsto \zeta(\tau + it)$$

as random variables on the probability space $[-T, T]$ given with the uniform probability measure $dt/(2T)$. These are arithmetically defined random variables. Do they have some specific, interesting, asymptotic behavior?

The answer to this question turns out to depend on σ, as the following first result reveals:

THEOREM 3.1.1 (Bohr-Jessen). *Let $\tau > 1/2$ be a fixed real number. Define $\mathsf{Z}_{\tau,T}$ as the random variable $t \mapsto \zeta(\tau + it)$ on $[-T, T]$. There exists a probability measure μ_τ on $\mathbf{C}$ such that*

$\mathsf{Z}_{\tau,T}$ *converges in law to* μ_τ *as* $T \to +\infty$. *Moreover, the support of* μ_τ *is compact if* $\tau > 1$, *and is equal to* $\mathbf{C}$ *if* $1/2 < \tau \leqslant 1$.

We will describe precisely the measure μ_τ in Section 3.2: it is a highly non-generic probability distribution, whose definition (and hence properties) retains a significant amount of arithmetic, in contrast with the Erdős-Kac Theorem, where the limit is a very generic distribution.

The analogue of Theorem 3.1.1 fails for $\tau = 1/2$. This shows that the Riemann zeta function is significantly more complicated on the critical line. However, there is a limit theorem after normalization, due to Selberg, which we will prove in Section 3.4, at least for the modulus of $\zeta(1/2 + it)$.

THEOREM 3.1.2 (Selberg). *Define* L_T *as the random variable* $t \mapsto \log \zeta(1/2+it)$ *on* $[-T, T]$, *defined by continuity along the line* $\mathrm{Im}(s) = t$ *for* t *such that* $\zeta(1/2 + it) = 0$, *and extended to be* 0 *for* t *among the ordinates of zeros of* ζ. *Then the sequence of random variables*

$$\frac{\mathsf{L}_T}{\sqrt{\frac{1}{2}\log\log T}}$$

converges in law as $T \to +\infty$ *to a standard complex gaussian random variable.*

There is another generalization of Theorem 3.1.1, due to Voronin and Bagchi, that we will discuss, and that extends it in a very surprising direction. Instead of fixing $\tau \in]1/2, 1[$ and looking at the distribution of the single values $\zeta(\tau + it)$ as t varies, we consider for such τ some radius r such that the disc

$$D = \{s \in \mathbf{C} \mid |s - \tau| \leqslant r\}$$

is contained in the interior of the critical strip, and we look for $t \in \mathbf{R}$ at the *functions*

$$\zeta_{D,t} : \begin{cases} D & \longrightarrow & \mathbf{C} \\ s & \mapsto & \zeta(s + it) \end{cases}$$

which are "vertical translates" of the Riemann zeta function restricted to D. For each $T \geqslant 0$, we view $t \mapsto \zeta_{D,t}$ as a random variable (say $\mathsf{Z}_{D,T}$) on $([-T, T], dt/(2T))$ with values in the space $\mathcal{H}(D)$ of functions which are holomorphic in the interior of D and continuous on its boundary. Bagchi's remarkable result is a convergence in law in this space: there exists a probability measure ν on $\mathcal{H}(D)$ such that the random variables $\mathsf{Z}_{D,T}$ converge in law to ν as $T \to +\infty$. Computing the support of ν (which is a non-trivial task) leads to a proof of Voronin's universality theorem: for any function $f \in \mathcal{H}(D)$ which does not vanish on D, and for any $\varepsilon > 0$, there exist $t \in \mathbf{R}$ such that

$$\|\zeta(\cdot + it) - f\|_\infty < \varepsilon,$$

where the norm is the supremum norm on D. In other words, up to arbitrarily small error, all functions f (that do not vanish) can be seen by looking at some vertical translate of the Riemann zeta function!

We illustrate this fact in Figure 3.1, which presents density plots of $|\zeta(s + it)|$ for various values of $t \in \mathbf{R}$, as functions of s in the square $[3/4 - 1/8, 3/4 + 1/8] \times [-1/8, 1/8]$. Voronin's Theorem implies that, for suitable t, such a picture will be arbitrarily "close" to that for any holomorphic function on this square that never vanishes there; one such function displayed in Figure 3.2 is $f(s) = 2 + \cos(10s)^3$.

We will prove the Bohr-Jessin-Bagchi theorems in the next section, and use in particular the computation of the support of Bagchi's limiting distribution for translates of the Riemann zeta function to prove Voronin's universality theorem in Section 3.3.

3.2. The Bohr-Jessen-Bagchi theorems

We begin by stating a precise version of Bagchi's Theorem. In the remainder of this chapter, we denote by Ω_T the probability space $([-T, T], dt/(2T))$ for $T \geqslant 1$. We will often write $\mathbf{E}_T(\cdot)$ and $\mathbf{P}_T(\cdot)$ for the corresponding expectation and probability.

FIGURE 3.1. The modulus of $\zeta(s+it)$ for s in the square $[3/4-1/8, 3/4+1/8] \times [-1/8, 1/8]$, for $t = 0$, 21000, 58000 and 75000.

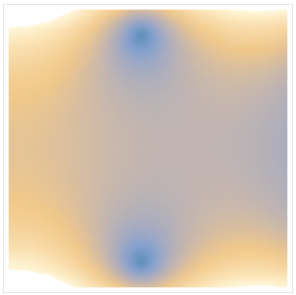

FIGURE 3.2. The modulus of $2+\cos(10s)^3$ for s in the square $[3/4-1/8, 3/4+1/8] \times [-1/8, 1/8]$.

THEOREM 3.2.1 (Bagchi [1]). *Let τ be such that $1/2 < \tau$. If $1/2\tau < 1$, let $r > 0$ be such that*

$$D = \{s \in \mathbf{C} \mid |s-\tau| \leqslant r\} \subset \{s \in \mathbf{C} \mid 1/2 < \mathrm{Re}(s) < 1\},$$

and if $\tau \geqslant 1$, let D be any compact subset of $\{s \in \mathbf{C} \mid \mathrm{Re}(s) \geqslant 1\}$ such that $\tau \in D$.

Consider the $\mathcal{H}(D)$-valued random variables $\mathsf{Z}_{D,T}$ defined by

$$t \mapsto (s \mapsto \zeta(s+it))$$

on Ω_T. Let $(X_p)_{p \text{ prime}}$ be a sequence of independent random variables which are identically distributed, with distribution uniform on the unit circle $\mathbf{S}^1 \subset \mathbf{C}^\times$.

Then we have convergence in law $\mathsf{Z}_{D,t} \longrightarrow Z_D$, where Z_D is the random Euler product

$$Z_D(s) = \prod_p (1 - p^{-s} X_p^{-it})^{-1}.$$

In this theorem, the space $\mathcal{H}(D)$ is viewed as a Banach space (hence a metric space, so that convergence in law makes sense) with the norm

$$\|f\|_\infty = \sup_{z \in D} |f(z)|.$$

We can already see that Theorem 3.2.1 is (much) stronger than the convergence in law component of Theorem 3.1.1 which we now prove assuming this result:

COROLLARY 3.2.2. *Fix τ such that $1/2 < \tau$. As $T \to +\infty$, the random variables $\mathsf{Z}_{\tau,T}$ of Theorem 3.1.1 converge in law to the random variable $Z_D(\tau)$, where D is either a disc*

$$D = \{s \in \mathbf{C} \mid |s-\tau| \leqslant r\}$$

contained in the interior of the critical strip, if $\tau < 1$, or any compact subset of $\{s \in \mathbf{C} \mid \mathrm{Re}(s) \geqslant 1\}$ such that $\tau \in D$.

PROOF. Fix D as in the statement. Tautologically, we have

$$\mathsf{Z}_{\tau,T} = \zeta_{D,T}(\tau)$$

or $\mathsf{Z}_{\tau,T} = e_\tau \circ \zeta_{D,T}$, where

$$e_\tau \left\{ \begin{array}{ccc} \mathcal{H}(D) & \longrightarrow & \mathbf{C} \\ f & \mapsto & f(\tau) \end{array} \right.$$

is the evaluation map. Since this map is continuous for the topology on $\mathcal{H}(D)$, it is tautological (see Proposition B.2.1 in Appendix B) that the convergence in law $\mathsf{Z}_{D,T} \longrightarrow Z_D$ of Bagchi's Theorem implies the convergence in law of $\mathsf{Z}_{\tau,T}$ to the random variable $e_\tau \circ Z_D$, which is simply $Z_D(\tau)$. □

In order to prove the final part of Theorem 3.1.1, and to derive Voronin's universality theorem, we need to understand the support of the limit Z_D in Bagchi's Theorem. We will prove in Section 3.3:

THEOREM 3.2.3 (Bagchi, Voronin). *Let τ be such that $1/2 < \tau < 1$, and r such that*

$$D = \{s \in \mathbf{C} \mid |s - \tau| \leqslant r\} \subset \{s \in \mathbf{C} \mid 1/2 < \mathrm{Re}(s) < 1\}.$$

The support of the distribution of the law of Z_D contains

$$\mathcal{H}(D)^\times = \{f \in \mathcal{H}(D) \mid f(z) \neq 0 \text{ for all } z \in D\}$$

and is equal to $\mathcal{H}(D)^\times \cup \{0\}$.

In particular, for any function $f \in \mathcal{H}(D)^\times$, and for any $\varepsilon > 0$, there exists $t \in \mathbf{R}$ such that

$$\sup_{s \in D} |\zeta(s + it) - f(s)| < \varepsilon. \tag{3.2}$$

It is then obvious that if $1/2 < \tau < 1$, the support of the Bohr-Jessen random variable $Z_D(\tau)$ is equal to $\mathbf{C}$.

We now begin the proof of Theorem 3.2.1 by giving some intuition for the result and in particular for the shape of the limiting distribution. Indeed, this very elementary argument will suffice to prove Bagchi's Theorem in the case $\tau > 1$. This turns out to be similar to the intuition behind the Erdős-Kac Theorem. We begin with the Euler product

$$\zeta(s + it) = \prod_p (1 - p^{-s-it})^{-1},$$

which is valid for $\mathrm{Re}(s) > 1$. We compute the logarithm

$$\log \zeta(s + it) = -\sum_p \log(1 - p^{-s-it}), \tag{3.3}$$

and see that this is a sum of random variables where the randomness lies in the behavior of the sequence of random variables $(\mathsf{X}_{p,T})_p$ on Ω_T given by $\mathsf{X}_p(t) = p^{-it}$, which take values in the unit circle $\mathbf{S}^1$. We view $(\mathsf{X}_{p,T})_p$, for each $T \geqslant 1$, as taking value in the infinite product $\hat{\mathbf{S}}^1 = \prod_p \mathbf{S}^1$ of circles parameterized by primes, which is still a compact metric space. It is therefore natural to study the behavior of these sequences as $T \to +\infty$, in order to guess how $\mathsf{Z}_{D,T}$ will behave. This has a very simple answer:

PROPOSITION 3.2.4. *For $T \geqslant 0$, let*

$$\mathsf{X}_T = (\mathsf{X}_{p,T})_p$$

be the $\hat{\mathbf{S}}^1$-valued random variable on Ω_T. given by

$$t \mapsto (p^{-it})_p.$$

Then X_T converges in law as $T \to +\infty$ to a random variable $X = (X_p)_p$, where the X_p are independent and uniformly distributed on $\mathbf{S}^1$.

Bagchi's Theorem is therefore to be understood as saying that we can "pass to the limit" in the formula (3.3) to obtain a convergence in law of $\log \zeta(s + it)$, for $s \in D$, to

$$-\sum_p \log(1 - p^{-s} X_p).$$

This sketch is of course incomplete in general, the foremost objection being that we are interested in particular in the zeta function outside of the region of absolute convergence, where the Euler product does not converge absolutely, so the meaning of (3.3) is unclear. But we will see that nevertheless enough connections remain to carry the argument through.

We isolate the crucial part of the proof of Proposition 3.2.4 as a lemma, since we will use it in Section 3.4 in the proof of Selberg's Theorem.

LEMMA 3.2.5. *Let $r > 0$ be a real numbers. We have*

$$|\mathbf{E}_T(r^{-it})| \leqslant \min\Big(1, \frac{1}{T|\log r|}\Big). \tag{3.4}$$

In particular, if $r = n_1/n_2$ for some positive integers $n_1 \neq n_2$, then we have

$$\mathbf{E}_T(r^{-it}) \ll \min\Big(1, \frac{\sqrt{n_1 n_2}}{T}\Big) \tag{3.5}$$

where the implied constant is absolute.

PROOF OF LEMMA 3.2.5. Since $|r^{-it}| = 1$, we see that the expectation is always $\leqslant 1$. If $r \neq 1$, then we get

$$\mathbf{E}(r^{-it}) = \frac{1}{2T}\Big[\frac{i}{\log r} r^{-it}\Big]_{-T}^{T} = \frac{i(r^{iT} - r^{-iT})}{2T(\log r)},$$

which has modulus at most $|\log r|^{-1}T^{-1}$, hence the first bound holds.

Assume now that $r = n_1/n_2$ with $n_1 \neq n_2$ positive integers. Assume that $n_2 > n_1 \geqslant 1$. Then $n_2 \geqslant n_1 + 1$, and hence

$$\Big|\log \frac{n_1}{n_2}\Big| = \Big|\log\Big(1 + \frac{1}{n_1}\Big)\Big| \gg \frac{1}{n_1} \geqslant \frac{1}{\sqrt{n_1 n_2}}.$$

If $n_2 < n_1$, we exchange the role of n_1 and n_2, and since both sides of the bound (3.5) are symmetric in terms of n_1 and n_2, the result follows. □

PROOF OF PROPOSITION 3.2.4. The impression of having an infinite-dimensional situation is illusory. Because the limiting measure (the law of (X_p)) is simply the Haar measure on the compact (abelian) group $\hat{\mathbf{S}}^1$, the well-known Weyl Criterion (see Section B.4 in Appendix B) shows that the statement is equivalent with the property that

$$\lim_{T \to +\infty} \mathbf{E}(\chi(\mathsf{X}_{p,T})) = 0 \tag{3.6}$$

for any non-trivial continuous character $\chi : \hat{\mathbf{S}}^1 \longrightarrow \mathbf{S}^1$. An elementary property of compact groups shows that for any such character there exists a finite non-empty subset S of primes, and for each $p \in S$ some integer $m_p \in \mathbf{Z} - \{0\}$, such that

$$\chi(z) = \prod_{p \in S} z_p^{m_p}$$

for any $z = (z_p)_p \in \hat{\mathbf{S}}^1$ (see Example B.4.2(2)). We then have by definition

$$\mathbf{E}(\chi(\mathsf{X}_{p,T})) = \frac{1}{2T}\int_{-T}^{T} \prod_{p \in S} p^{-itm_p} dt = \frac{1}{2T}\int_{-T}^{T} r^{-it} dt$$

where $r > 0$ is the rational number given by

$$r = \prod_{p \in S} p^{m_p}.$$

Since we have $r \neq 1$ (because S is not empty and $m_p \neq 0$), we obtain $\mathbf{E}(\chi(\mathsf{X}_{p,T})) \to 0$ as $T \to +\infty$ from (3.4). □

As a corollary, Bagchi's Theorem follows formally for $\tau > 1$ and D contained in the set of complex numbers with real part > 1. This is once more a very simple fact which is often not specifically discussed, but which gives an indication and a motivation for the more difficult study in the critical strip.

Special case of Theorem 3.2.1 for $\tau > 1$. Assume that $\tau > 1$ and that D is a compact subset containing τ contained in $\{s \in \mathbf{C} \mid \mathrm{Re}(s) > 1\}$. We view $\mathsf{X}_T = (\mathsf{X}_{p,T})$ as random variables with values in the topological space $\hat{\mathbf{S}}^1$, as before. This is also (as a countable product of metric spaces) a metric space. We claim that the map

$$\varphi \quad \left\{ \begin{array}{ccc} \hat{\mathbf{S}}^1 & \longrightarrow & \mathcal{H}(D) \\ (x_p) & \mapsto & \Big(s \mapsto -\displaystyle\sum_p \log(1 - x_p p^{-s})\Big) \end{array} \right.$$

is continuous. If this is so, then the composition principle (see Proposition B.2.1) and Proposition 3.2.4 imply that $\varphi(\mathsf{X}_T)$ converges in law to the $\mathcal{H}(D)$-valued random variable $\varphi(X)$, where $X = (X_p)$ with the X_p uniform and independent on $\mathbf{S}^1$. But this is exactly the statement of Bagchi's Theorem for D.

Now we check the claim. Fix $\varepsilon > 0$. Let $T > 0$ be some parameter to be chosen later in terms of ε. For any $x = (x_p)$ and $y = (y_p)$ in $\hat{\mathbf{S}}^1$, we have

$$\|\varphi(x) - \varphi(y)\|_\infty \leqslant \sum_{p \leqslant T} \|\log(1 - x_p p^{-s}) - \log(1 - y_p p^{-s})\|_\infty + \sum_{p>T} \|\log(1 - x_p p^{-s})\|_\infty + \sum_{p>T} \|\log(1 - y_p p^{-s})\|_\infty.$$

Because D is compact in the half-plane $\mathrm{Re}(s) > 1$, the minimum of the real part of $s \in D$ is some real numbers $\sigma_0 > 1$. Since $|x_p| = |y_p| = 1$ for all primes, and

$$|\log(1 - z)| \leqslant 2|z|$$

for $|z| \leqslant 1/2$, it follows that

$$\sum_{p>T} \|\log(1 - x_p p^{-s})\|_\infty + \sum_{p>T} \|\log(1 - y_p p^{-s})\|_\infty \leqslant 4 \sum_{p>T} p^{-\sigma_0} \ll T^{1-\sigma_0}.$$

We fix T so that $T^{1-\sigma_0} < \varepsilon/2$. Now the map

$$(x_p)_{p \leqslant T} \mapsto \sum_{p \leqslant T} \|\log(1 - x_p p^{-s}) - \log(1 - y_p p^{-s})\|_\infty$$

is obviously continuous, and therefore uniformly continuous since the definition set is compact. This function has value 0 when $x_p = y_p$ for $p \leqslant T$, so there exists $\delta > 0$ such that

$$\sum_{p \leqslant T} |\log(1 - x_p p^{-s}) - \log(1 - y_p p^{-s})| < \frac{\varepsilon}{2}$$

if $|x_p - y_p| \leqslant \delta$ for $p \leqslant T$. Therefore, provided that

$$\max_{p \leqslant T} |x_p - y_p| \leqslant \delta,$$

we have

$$\|\varphi(x) - \varphi(y)\|_\infty \leqslant \varepsilon.$$

This proves the (uniform) continuity of φ. □

We now begin the proof of Bagchi's Theorem in the critical strip. The argument follows closely his original proof [1], which is quite different from the Bohr-Jessen approach (as we will briefly discuss at the end). Here are the main steps of the proof:

- We prove convergence almost surely of the random Euler product, and of its formal Dirichlet series expansion; this also shows that they define random *holomorphic* functions;
- We prove that both the Riemann zeta function and the limiting Dirichlet series are, in suitable mean sense, limits of smoothed partial sums of their respective Dirichlet series;
- We then use an elementary argument to conclude using Proposition 3.2.4.

We fix from now on a sequence $(X_p)_p$ of independent random variables all uniformly distributed on $\mathbf{S}^1$. We often view the sequence (X_p) as an $\hat{\mathbf{S}}^1$-valued random variable. Furthermore, for any positive integer $n \geqslant 1$, we define

$$X_n = \prod_{p|n} X_p^{v_p(n)}$$

where $v_p(n)$ is the p-adic valuation of n. Thus (X_n) is a sequence of $\mathbf{S}^1$-valued random variables.

EXERCISE 3.2.6. Prove that the sequence $(X_n)_{n\geqslant 1}$ is neither independent nor symmetric.

We first show that the limiting random functions are indeed well-defined as $\mathcal{H}(D)$-valued random variables.

PROPOSITION 3.2.7. *Let $\tau \in]1/2, 1[$ and let $U_\tau = \{s \in \mathbf{C} \mid \Re(s) > \tau\}$.*
(1) *The random Euler product defined by*

$$Z(s) = \prod_p (1 - X_p p^{-s})^{-1}$$

converges almost surely for any $s \in U_\tau$; for any compact subset $K \subset U_\tau$, the random function

$$Z_K : \begin{cases} K & \longrightarrow \mathbf{C} \\ s & \mapsto Z(s) \end{cases}$$

is an $\mathcal{H}(K)$-valued random variable.
(2) *The random Dirichlet series defined by*

$$\tilde{Z} = \sum_{n\geqslant 1} X_n n^{-s}$$

converges almost surely for any $s \in U_\tau$; for any compact subset $K \subset U_\tau$, the random function $\tilde{Z}_K : s \mapsto \tilde{Z}(s)$ on K is an $\mathcal{H}(K)$-valued random variable, and moreover, we have $\tilde{Z}_K = Z_K$ almost surely.

PROOF. (1) For $N \geqslant 1$ and $s \in K$ we have

$$\sum_{p\leqslant N} \log(1 - X_p p^{-s})^{-1} = \sum_{p\leqslant N} \frac{X_p}{p^s} + \sum_{k\geqslant 2} \sum_{p\leqslant N} \frac{X_{p^k}}{p^{ks}}.$$

Since $\mathrm{Re}(s) > 1/2$ for $s \in K$, the series

$$\sum_{k\geqslant 2} \sum_p \frac{X_{p^k}}{p^{ks}}$$

converges absolutely for $s \in U_\tau$. By Lemma A.2.1, its sum is therefore a random holomorphic function in $\mathcal{H}(K)$-valued random variable for any compact subset K of U_τ.

Fix now $\tau_1 < \tau$ such that $\tau_1 > \tau$. We can apply Kolmogorov's Theorem B.8.1 to the independent random variables $(X_p p^{-\tau_1})$, since

$$\sum_p \mathbf{V}(p^{-\tau_1} X_p) = \sum_p \frac{1}{p^{2\tau_1}} < +\infty.$$

Thus the series

$$\sum_p \frac{X_p}{p^{\tau_1}}$$

converges almost surely. By Lemma A.2.1 again, it follows that

$$P(s) = \sum_p \frac{X_p}{p^s}$$

converges almost surely for all $s \in U_\tau$, and is holomorphic on U_τ. By restriction, its sum is an $\mathcal{H}(K)$-valued random variable for any K compact in U_τ.

These facts show that the sequence of partial sums

$$\sum_{p \leqslant N} \log(1 - X_p p^{-s})^{-1}$$

converges almost surely as $N \to +\infty$ to a random holomorphic function on K. Taking the exponential, we obtain the almost sure convergence of the random Euler product to a random holomorphic function Z_K on K.

(2) The argument is similar, except that the sequence $(X_n)_{n \geqslant 1}$ is not independent. However, it is orthonormal: if $n \neq m$, we have

$$\mathbf{E}(X_n \overline{X}_m) = 0, \qquad \mathbf{E}(|X_n|^2) = 1$$

(indeed X_n and X_m may be viewed as characters of $\hat{\mathbf{S}}^1$, and they are distinct if $n \neq m$, so that this is the orthogonality property of characters of compact groups). We can then apply the Menshov-Rademacher Theorem B.8.4 to (X_n) and $a_n = n^{-\tau_1}$: since

$$\sum_{n \geqslant 1} |a_n|^2 (\log n)^2 = \sum_{n \geqslant 1} \frac{(\log n)^2}{n^{2\tau_1}} < +\infty,$$

the series $\sum X_n n^{-\tau_1}$ converges almost surely, and Lemma A.2.1 shows that $\tilde{Z}$ converges almost surely on U_τ, and defines a holomorphic function there. Restricting to K leads to $\tilde{Z}_K$ as $\mathcal{H}(K)$-valued random variable.

Finally, to prove that $Z_K = \tilde{Z}_K$ almost surely, we may replace K by the compact subset

$$K_1 = \{s \in \mathbf{C} \mid \tau_1 \leqslant \sigma \leqslant A, \quad |t| \leqslant B\},$$

with $A \geqslant 2$ and B chosen large enough to ensure that $K \subset K_1$. The previous argument shows that the random Euler product and Dirichlet series converge almost surely on K_1. But K_1 contains the open set

$$V = \{s \in \mathbf{C} \mid 1 < \mathrm{Re}(s) < 2, \quad |t| < B\}$$

where the Euler product and Dirichlet series converge absolutely, so that Lemma C.1.2 proves that the random holomorphic functions Z_{K_1} and $\tilde{Z}_{K_1}$ are equal when restricted to V. By analytic continuation (and continuity), they are equal also on K_1, hence *a posteriori* on K. □

We will prove Bagchi's Theorem using the random Dirichlet series, which is easier to handle than the Euler product. However, we will still denote it $Z(s)$, which is justified by the last part of the proposition.

For the proof of Bagchi's Theorem, some additional properties of this random Dirichlet series are needed. Most importantly, we need to find a finite approximation that also applies to the Riemann zeta function. This will be done using *smooth partial sums*.

First we need to check that $Z(s)$ is of polynomial growth on average on vertical strips.

LEMMA 3.2.8. *Let $Z(s)$ be the random Dirichlet series $\sum X_n n^{-s}$ defined and holomorphic almost surely for $\mathrm{Re}(s) > 1/2$. For any $\sigma_1 > 1/2$, we have*

$$\mathbf{E}(|Z(s)|) \ll 1 + |s|$$

uniformly for all s such that $\mathrm{Re}(s) \geqslant \sigma_1$.

PROOF. The series

$$\sum_{n\geqslant 1}\frac{X_n}{n^{\sigma_1}}$$

converges almost surely. Therefore the partial sums

$$S_u=\sum_{n\leqslant u}\frac{X_n}{n^{\sigma_1}}$$

are bounded almost surely.

By summation by parts (see Appendix A), it follows that for any s with real part $\sigma>\sigma_1$, we have

$$Z(s)=(s-\sigma_1)\int_1^{+\infty}\frac{S_u}{u^{s-\sigma_1+1}}du,$$

where the integral converges almost surely. Hence

$$|Z(s)|\leqslant(1+|s|)\int_1^{+\infty}\frac{|S_u|}{u^{\sigma-\sigma_1+1}}du.$$

Fubini's Theorem (for non-negative functions) and the Cauchy-Schwarz inequality then imply

$$\begin{aligned}\mathbf{E}(|Z(s)|)&\leqslant(1+|s|)\int_1^{+\infty}\mathbf{E}(|S_u|)\frac{du}{u^{\sigma-\sigma_1+1}}\\&\leqslant(1+|s|)\int_1^{+\infty}\mathbf{E}(|S_u|^2)^{1/2}\frac{du}{u^{\sigma-\sigma_1+1}}\\&=(1+|s|)\int_1^{+\infty}\Big(\sum_{n\leqslant u}\frac{1}{n^{2\sigma_1}}\Big)^{1/2}\frac{du}{u^{\sigma-\sigma_1+1}}\ll 1+|s|,\end{aligned}$$

using the orthonormality of the variables X_n. □

We can then deduce a good result on average approximation by partial sums.

PROPOSITION 3.2.9. *Let $\varphi:[0,+\infty[\longrightarrow[0,1]$ be a smooth function with compact support such that $\varphi(0)=1$. Let $\hat{\varphi}$ denote its Mellin transform. For $N\geqslant 1$, define the $\mathcal{H}(D)$-valued random variable*

$$Z_{D,N}=\sum_{n\geqslant 1}X_n\varphi\Big(\frac{n}{N}\Big)n^{-s}.$$

There exists $\delta>0$ such that

$$\mathbf{E}(\|Z_D-Z_{D,N}\|_\infty)\ll N^{-\delta}$$

for $N\geqslant 1$.

We recall that the norm $\|\cdot\|_\infty$ refers to the sup norm on the compact set D.

PROOF. The first step is to apply the smoothing process of Proposition A.2.4 in Appendix A. Since the random Dirichlet series

$$Z(s)=\sum_{n\geqslant 1}X_n n^{-s}$$

converges almost surely for $\mathrm{Re}(s)>1/2$, we deduce that almost surely, we have

$$Z_D(s)-Z_{D,N}(s)=-\frac{1}{2i\pi}\int_{(-\delta)}Z(s+w)\hat{\varphi}(w)N^w dw$$

for $\sigma>1/2$ and any $\delta>0$ such that $-\delta+\sigma\geqslant 1/2$. (It is important that the "almost surely" is independent of s, which is simply because we work with random variables taking values in $\mathcal{H}(D)$, and not with particular evaluations of these random functions at a specific $s\in D$).

Using a rectangle R containing D in its interior, we deduce from Cauchy's Theorem that

$$\|Z_D-Z_{D,N}\|_\infty\ll N^{-\delta}\int_R\int_{\mathbf{R}}|Z(-\delta+\sigma+i(t+u))|\ |\hat{\varphi}(-\delta+iu)||ds|du.$$

Therefore, taking the expectation, we get

$$\mathbf{E}(\|Z_D - Z_{D,N}\|_\infty) \ll N^{-\delta} \int_R \int_{\mathbf{R}} \mathbf{E}\big(|Z(-\delta+\sigma+i(t+u))|\big)\ |\hat{\varphi}(-\delta+iu)||ds|du.$$

Applying Fubini's Theorem, we therefore need to bound

$$\int_{\mathbf{R}} \mathbf{E}\big(|Z(-\delta+\sigma+i(t+u))|\big)\ |\hat{\varphi}(-\delta+iu)|du.$$

for some fixed $\sigma+it$ in the compact rectangle R. Since $\hat{\varphi}$ decays faster than any polynomial at infinity in vertical strips, and

$$\mathbf{E}(|Z(s)|) \ll 1+|s|$$

by Lemma 3.2.8, we have

$$\int_{\mathbf{R}} \mathbf{E}\big(|Z(-\delta+\sigma+i(t+u))|\big)\ |\hat{\varphi}(-\delta+iu)|du \ll 1$$

uniformly for $\sigma+it$ in a compact set contained in $]1/2,1[$. □

The last preliminary result is a similar approximation result for the translates of the Riemann zeta function by smooth partial sums of its Dirichlet series.

Proposition 3.2.10. *Let $\varphi : [0,+\infty[\longrightarrow [0,1]$ be a smooth function with compact support such that $\varphi(0)=1$. Let $\hat{\varphi}$ denote its Mellin transform. For $N \geqslant 1$, define*

$$\zeta_N(s) = \sum_{n\geqslant 1} \varphi\Big(\frac{n}{N}\Big) n^{-s},$$

and define $\mathsf{Z}_{N,T}$ to be the $\mathcal{H}(D)$-valued random variable $t \mapsto (s \mapsto \zeta_N(s+it))$.

There exists $\delta > 0$ such that

$$\mathbf{E}_T(\|\mathsf{Z}_{D,T} - \mathsf{Z}_{N,T}\|_\infty) \ll N^{-\delta} + NT^{-1}$$

for $N \geqslant 1$ and $T \geqslant 1$.

Note that ζ_N is an entire function, since φ has compact support, so that the range of the sum is in fact finite. The meaning of the statement is that the smoothed partial sums ζ_N give very uniform and strong approximations to the vertical translates of the Riemann zeta function.

Proof. We will write Z_T for $\mathsf{Z}_{D,T}$ for simplicity. We begin by applying the smoothing process of Proposition A.2.4 in Appendix A in the case $a_n = 1$. For $\sigma > 1/2$ and any $\delta > 0$ such that

$$-\delta + \sigma \geqslant 1/2,$$

we have

$$\zeta(s) - \zeta_N(s) = -\frac{1}{2i\pi} \int_{(-\delta)} \zeta(s+w)\hat{\varphi}(w)N^w dw + N^{1-s}\hat{\varphi}(1-s) \tag{3.7}$$

since the Riemann zeta function has a pole at $s=1$ with residue 1 (Figure 3.3 may help understand the location of the regions involved in the proof).

We need in fact to have some control of the supremum norm on D, since this is the norm on the space $\mathcal{H}(D)$. For this purpose, we use Cauchy's inequality.

Let S be a compact segment in $]1/2,1[$ such that the fixed rectangle $R = S \times [-1/2,1/2] \subset \mathbf{C}$ contains D in its interior. Then for any v with $\mathrm{Re}(v) > 1/2$ and $t \in \mathbf{R}$, Cauchy's theorem gives

$$\zeta(v+it) - \zeta_N(v+it) = \frac{1}{2i\pi} \int_{\partial R} (\zeta(s+it) - \zeta_N(s+it)) \frac{ds}{s-v},$$

where the boundary of R is oriented counterclockwise. Since the definition of R ensures that $|s-v|^{-1} \gg 1$ for $v \in D$ and $s \in \partial R$, we deduce that the random variable $\|\mathsf{Z}_T - \mathsf{Z}_{N,T}\|_\infty$, which takes the value

$$\sup_{s\in D} |\zeta(s+it) - \zeta_N(s+it)|$$

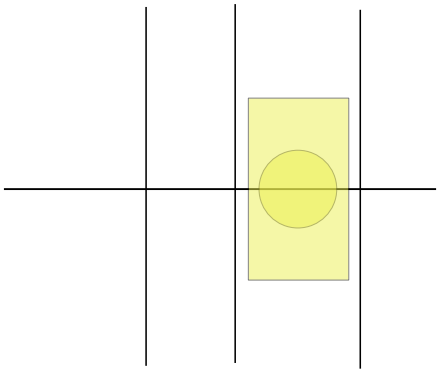

FIGURE 3.3. Regions and contours in the proof of Proposition 3.2.10.

at $t \in \Omega_T$, satisfies

$$\|\mathsf{Z}_T - \mathsf{Z}_{N,T}\|_\infty \ll \int_{\partial R} |\zeta(s+it) - \zeta_N(s+it)||ds|$$

for $t \in \Omega_T$. Integrating over $t \in [-T, T]$ leads to

$$\mathbf{E}_T\Big(\|\mathsf{Z}_T - \zeta_{N,T}\|_\infty\Big) \ll \frac{1}{2T}\int_{-T}^{T}\int_{\partial R} |\zeta(s+it) - \zeta_N(s+it)||ds|dt.$$

We split the integral over ∂R in the two horizontal and the two vertical segments. The contribution of the vertical segments are estimated by

$$\frac{1}{2T}\int_{-T}^{T}\int_{-1/2}^{1/2} |\zeta(a+iu+it) - \zeta_N(a+iu+it)|dudt \leqslant \frac{1}{2T}\int_{-2T}^{2T} |\zeta(a+iu) - \zeta_N(a+iu)|du$$

where a is the minimum or the maximum of the segment S. The contribution of the horizontal segments can be expressed as a two-dimensional integral over a rectangle contained in

$$S \times [-2T, 2T],$$

Therefore, applying Fubini's Theorem, and then estimating the integrals for a fixed real part, we get

$$\mathbf{E}_T\Big(\|\mathsf{Z}_T - \mathsf{Z}_{N,T}\|_\infty\Big) \ll \sup_{\sigma\in S} \frac{1}{T}\int_{-2T}^{2T} |\zeta_N(\sigma+it) - \zeta(\sigma+it)|dt,$$

since the vertical contributions are (by the above) also bounded by the right-hand side.

We now fix $\sigma \in S$. We use (3.7) to express $\zeta_N(\sigma+it) - \zeta(\sigma+it)$ for $t \in [-2T, 2T]$. We take

$$\delta = \frac{1}{2}(\min S - 1/2) > 0$$

(since S is compact in $]1/2, 1[$), so that

$$-\delta + \sigma > 1/2, \qquad 0 < \delta < 1.$$

Thus

$$\begin{aligned}\frac{1}{T}\int_{-2T}^{2T} |\zeta_N(\sigma+it) - \zeta(\sigma+it)|dt \ll \frac{N^{1-\sigma}}{T}\int_{-2T}^{2T} |\hat\varphi(1-\sigma-it)|dt \\ + \frac{N^{-\delta}}{T}\int_{-2T}^{2T}\int_{\mathbf{R}} |\zeta(1/2+\varepsilon+i(u+t))||\hat\varphi(-\delta+iu)|dudt.\end{aligned}$$

The first terms is bounded by $\ll NT^{-1}$ using the fast decay of $\hat\varphi$ on vertical strips, which shows that the integral is uniformly bounded.

For the second term, we apply Fubini's Theorem again. For a fixed u, the integral over t is then the integral of ζ along a vertical segment contained in

$$\{1/2+\varepsilon\} \times [-|u| - 2T, |u| + 2T].$$

Using the fact that

$$\frac{1}{X}\int_{-X}^{X}|\zeta(1/2+\varepsilon+it)|dt \ll 1 \tag{3.8}$$

for a fixed $\varepsilon > 0$ and $X \geqslant 1$ (see Proposition C.2.1 in Appendix C), we obtain

$$\frac{1}{T}\int_{-2T}^{2T}|\zeta_N(\sigma+it)-\zeta(\sigma+it)|dt \ll N^{-\delta}\int_{\mathbf{R}}|\hat{\varphi}(-\delta+iu)|\Big(1+\frac{|u|}{T}\Big)du.$$

Now the fast decay of $\hat{\varphi}(s)$ on the vertical line $\mathrm{Re}(s) = -\delta$ shows that

$$\frac{N^{-\delta}}{T}\int_{-2T}^{2T}\int_{\mathbf{R}}|\zeta(1/2+\varepsilon+i(u+t))||\hat{\varphi}(-\delta+iu)|dudt \ll N^{-\delta},$$

and this concludes the proof. □

Finally we can prove Theorem 3.2.1:

Proof of Bagchi's Theorem. By Proposition B.2.2, it is enough to prove that for any bounded and Lipschitz function $f : \mathcal{H}(D) \longrightarrow \mathbf{C}$, we have

$$\mathbf{E}_T(f(\mathsf{Z}_{D,T})) \longrightarrow \mathbf{E}(f(Z_D))$$

as $T \to +\infty$. We may use the Dirichlet series expansion of Z_D according to Proposition 3.2.7, (2).

Since D is fixed, we omit it from the notation for simplicity, denoting $\mathsf{Z}_T = \mathsf{Z}_{D,T}$ and $Z = Z_D$. Fix some integer $N \geqslant 1$ to be chosen later. We denote

$$\mathsf{Z}_{T,N} = \sum_{n\geqslant 1} n^{-s}\varphi\Big(\frac{n}{N}\Big)$$

(viewed as random variable on $[-T, T]$) and

$$Z_N = \sum_{n\geqslant 1} X_n n^{-s}\varphi\Big(\frac{n}{N}\Big)$$

the smoothed partial sums of the Dirichlet series as in Propositions 3.2.10 and 3.2.9.

We then write

$$\begin{aligned}|\,\mathbf{E}_T(f(\mathsf{Z}_T)) - \mathbf{E}(f(Z))| \leqslant |\,\mathbf{E}_T(f(\mathsf{Z}_T) - f(\mathsf{Z}_{T,N}))| + \\ |\,\mathbf{E}_T(f(\mathsf{Z}_{T,N})) - \mathbf{E}(f(Z_N))| + |\,\mathbf{E}(f(Z_N) - f(Z))|.\end{aligned}$$

Since f is a Lipschitz function on $\mathcal{H}(D)$, there exists a constant $C \geqslant 0$ such that

$$|f(x) - f(y)| \leqslant C\|x - y\|_\infty$$

for all x, $y \in \mathcal{H}(D)$. Hence we have

$$\begin{aligned}|\,\mathbf{E}_T(f(\mathsf{Z}_T)) - \mathbf{E}(f(Z))| \leqslant C\,\mathbf{E}_T(\|\mathsf{Z}_T - \mathsf{Z}_{T,N}\|_\infty) + \\ |\,\mathbf{E}_T(f(\mathsf{Z}_{T,N})) - \mathbf{E}(f(Z_N))| + C\,\mathbf{E}(\|Z_N - Z\|_\infty).\end{aligned}$$

Fix $\varepsilon > 0$. Propositions 3.2.10 and 3.2.9 together show that there exists some $N \geqslant 1$ such that

$$\mathbf{E}_T(\|\mathsf{Z}_T - \mathsf{Z}_{T,N}\|_\infty) < \varepsilon + \frac{N}{T}$$

for all $T \geqslant 1$ and

$$\mathbf{E}(\|Z_N - Z\|_\infty) < \varepsilon.$$

We fix such a value of N. By Proposition 3.2.4 and composition, the random variables $\mathsf{Z}_{T,N}$ (which are Dirichlet polynomials) converge in law to Z_N as $T \to +\infty$. Since $N/T \to 0$ also for $T \to +\infty$, we deduce that for all T large enough, we have

$$|\,\mathbf{E}_T(f(\mathsf{Z}_T)) - \mathbf{E}(f(Z))| < 4\varepsilon.$$

This finishes the proof. □

EXERCISE 3.2.11. Prove that if $\sigma > 1/2$ is fixed, then we have almost surely

$$\lim_{T\to+\infty} \frac{1}{2T}\int_{-T}^{T} |Z(\sigma+it)|^2 dt = \zeta(2\sigma).$$

[*Hint.* Use the Birkhoff-Khintchine pointwise ergodic theorem for flows, see e.g. [**?**, §8.6.1].]

3.3. The support of Bagchi's measure

Our goal in this section is to explain the proof of Theorem 3.2.3, which is due to Bagchi [1, Ch. 5]. Since it involves results of complex analysis that are quite far from the main interest of these notes, we will only treat in detail the part of the proof that involves arithmetic, giving references for the results that are used.

The support is easiest to compute using the random Euler interpretation of the random Dirichet series, essentially because it is essentially a sum of independent random variables. To be precise, define

$$P(s) = \sum_p \frac{X_p}{p^s}, \qquad \tilde{P}(s) = \sum_p \sum_{k\geqslant 1} \frac{X_p^k}{p^{ks}}$$

(see the proof of Proposition 3.2.7). The series converge almost surely for $\mathrm{Re}(s) > 1/2$. We claim that the support of the distribution of $\tilde{P}$, when viewed as an $\mathcal{H}(D)$-valued random variable, is equal to $\mathcal{H}(D)$. Let us first assume this.

Since $Z = \exp(\tilde{P})$, we deduce by composition (see Lemma B.1.1) that the support of Z is the closure of the set of functions of the form e^g, where $g \in \mathcal{H}(D)$. But this last set is precisely $\mathcal{H}(D)^\times$, and Lemma A.3.5 in Appendix A shows that its closure in $\mathcal{H}(D)$ is $\mathcal{H}(D)^\times \cup \{0\}$.

Finally, to prove the approximation property (3.2), which is the original version of Voronin's universality theorem, we simply apply Lemma B.2.3 to the family of random variables Z_T, which gives the much stronger statement that for any $\varepsilon > 0$, we have

$$\liminf_{T\to+\infty} \lambda(\{t \in [-T,T] \mid \sup_{s\in D} |\zeta(s+it) - f(s)| < \varepsilon\}) > 0,$$

where λ denotes Lebesgue measure.

From Proposition B.8.7 in Appendix B, the following proposition will imply that the support of the random Dirichlet series P is $\mathcal{H}(D)$. The statement is slightly more general to help with the last step afterwards.

PROPOSITION 3.3.1. *Let τ be such that $1/2 < \tau < 1$. Let $r > 0$ be such that*

$$D = \{s \in \mathbf{C} \mid |s-\tau| \leqslant r\} \subset \{s \in \mathbf{C} \mid 1/2 < \mathrm{Re}(s) < 1\}.$$

Let N be an arbitrary positive real number. The set of all series

$$\sum_{p>N} \frac{x_p}{p^s}, \qquad (x_p) \in \hat{\mathbf{S}}^1,$$

which converge in $\mathcal{H}(D)$ is dense in $\mathcal{H}(D)$.

We will deduce the proposition from the density criterion of Theorem A.3.1 in Appendix A, applied to the space $\mathcal{H}(D)$ and the sequence (f_p) with $f_p(s) = p^{-s}$ for p prime. Since $\|f_p\|_\infty = p^{-\sigma_1}$, where $\sigma_1 = \tau - r > 1/2$, the condition

$$\sum_p \|f_p\|_\infty^2 < +\infty$$

holds. Furthermore, Proposition 3.2.7 certainly shows that there exist *some* $(x_p) \in \hat{\mathbf{S}}^1$ such that the series $\sum_p x_p f_p$ converges in $\mathcal{H}(D)$. Hence the conclusion of Theorem A.3.1 is what we seek, and we only need to check the following lemma to establish the last hypothesis required to apply it:

LEMMA 3.3.2. *Let $\mu \in C(\bar{D})'$ be a continuous linear functional. Let*

$$g(z) = \mu(s \mapsto e^{sz})$$

be its Laplace transform. If

$$\sum_p |g(\log p)| < +\infty, \tag{3.9}$$

then we have $g = 0$.

Indeed, the point is that $\mu(f_p) = \mu(s \mapsto p^{-s}) = g(\log p)$, so that the assumption (3.9) concerning g is precisely (A.1).

This is a statement that has some arithmetic content, as we will see, and indeed the proof involves the Prime Number Theorem.

PROOF. Let

$$\varrho = \limsup_{r \to +\infty} \frac{\log |g(r)|}{r},$$

which is finite by Lemma A.3.2 (1). By Lemma A.3.2 (3), it suffices to prove that $\varrho \leqslant 1/2$ to conclude that $g = 0$. To do this, we will use Theorem A.3.3, that provides access to the value of ϱ by "sampling" g along certain sequences of real numbers tending to infinity.

The idea is that (3.9) implies that $|g(\log p)|$ cannot be often of size at least $1/p = e^{-\log p}$, since the series $\sum p^{-1}$ diverges. Since the $\log p$ increase slowly, this makes it possible to find real numbers $r_k \to +\infty$ growing linearly and such that $|g(r_k)| \leqslant e^{-r_k}$, and from this and Theorem A.3.3 we will get a contradiction.

To be precise, we first note that for $y \in \mathbf{R}$, we have

$$|g(iy)| \leqslant \|\mu\| \|s \mapsto e^{iys}\|_\infty \leqslant \|\mu\| e^{r|y|}$$

(since the maximum of the absolute value of the imaginary part of $s \in \bar{D}$ is r), and therefore

$$\limsup_{\substack{y \in \mathbf{R} \\ |y| \to +\infty}} \frac{\log |g(iy)|}{|y|} \leqslant r.$$

We put $\alpha = r \leqslant 1/4$. Then the first condition of Theorem A.3.3 holds for the function g. We also take $\beta = 1$ so that $\alpha\beta < \pi$.

For any $k \geqslant 0$, let I_k be the set of primes p such that $e^k \leqslant p < e^{k+1}$. By the Prime Number Theorem, we have

$$\sum_{p \in I_k} \frac{1}{p} \sim \frac{1}{k}$$

as $k \to +\infty$. Let further A be the set of those $k \geqslant 0$ for which the inequality

$$|g(\log p)| \geqslant \frac{1}{p}$$

holds for *all* primes $p \in I_k$, and let B be its complement among the non-negative integers. We then note that

$$\sum_{k \in A} \frac{1}{k} \ll \sum_{k \in A} \sum_{p \in I_k} \frac{1}{p} \ll \sum_{k \in A} \sum_{p \in I_k} |g(\log p)| < +\infty.$$

This shows that B is infinite. For $k \in B$, let p_k be a prime in I_k such that $|g(\log p_k)| < p_k^{-1}$. Let $r_k = \log p_k$. We then have

$$\limsup_{k \to +\infty} \frac{\log |g(r_k)|}{r_k} \leqslant -1.$$

Since $p_k \in I_k$, we have

$$r_k = \log p_k \sim k.$$

Furthermore, if we order B in increasing order, the fact that

$$\sum_{k\notin B}\frac{1}{k}<+\infty$$

implies that the k-th element n_k of B satisfies $n_k\sim k$.

Now we consider the sequence formed from the r_{2k}, arranged in increasing order. We have $r_{2k}/k\to 2$ from the above. Moreover, by construction, we have

$$r_{2k+2}-r_{2k}\geqslant 1,$$

hence $|r_{2k}-r_{2l}|\gg|k-l|$. Since $|g(r_{2k})|\leqslant e^{-r_{2k}}$ for all $k\in B$, we can apply Theorem A.3.3 to this increasing sequence and we get

$$\varrho=\limsup_{k\to+\infty}\frac{\log|g(r_{2k})|}{r_{2k}}\leqslant -1<1/2,$$

as desired. □

There remains a last lemma to prove, that allows us to go from the support of the series $P(s)$ of independent random variables to that of the full series $\tilde{P}(s)$.

Lemma 3.3.3. *The support of $\tilde{P}(s)$ is $\mathcal{H}(D)$.*

Proof. We can write

$$\tilde{P}=-\sum_p\log(1-X_pp^{-s})$$

where the random variables $(\log(1-X_pp-s))_p$ are independent, and the series converges almost surely in $\mathcal{H}(D)$. Therefore it is enough by Proposition B.8.7 to prove that the set of convergent series

$$-\sum_p\log(1-x_pp^{-s}),\qquad (x_p)\in\hat{\mathbf{S}}^1,$$

is dense in $\mathcal{H}(D)$.

Fix $f\in\mathcal{H}(D)$ and $\varepsilon>0$ be fixed. Let $N\geqslant 1$ be a parameter to be chosen later. For (x_p) such that the series converges, we write

$$-\sum_p\log(1-x_pp^{-s})=f_N(s)+g_N(s)+h_N(s),$$

where

$$f_N(s)=\sum_{p\leqslant N}\sum_{k\geqslant 1}\frac{x_p^k}{p^{ks}},\qquad g_N(s)=\sum_{p>N}\frac{x_p}{p^s},$$

and $h_N(s)$ is the remainder. The series defining f_N converges absolutely for any $(x_p)_{p\leqslant N}$ with $|x_p|=1$. We pick for instance $x_p=1$ for $p\leqslant N$, and denote $f_0=f-f_N$ for this particular choice.

We have furthermore

$$\|h_N\|_\infty\leqslant\sum_{k\geqslant 2}\sum_{p>N}\frac{1}{p^{k/2}}\to 0$$

as $N\to+\infty$. In particular, we can select N so that $\|h_N\|_\infty<\varepsilon/2$. We then see that

$$\|f_N+g_N+h_N-f\|_\infty<\varepsilon$$

is true provided $(x_p)_{p>N}$ is such that

$$\|g_N-f_0\|_\infty<\frac{\varepsilon}{2},$$

and such a choice exists by Proposition 3.3.1. This concludes the proof. □

3.4. Selberg's theorem

We present in this section a recent proof of "half" of Theorem 3.1.2 due to Radziwill and Soundararajan [28]. Namely, we will use their methods to prove:

THEOREM 3.4.1 (Selberg). *Define* L_T *as the random variable* $t \mapsto \log|\zeta(1/2+it)|$ *on* $\Omega_T = [-T,T]$, *defined by continuity along the line* $\mathrm{Im}(s) = t$ *for* t *such that* $\zeta(1/2+it) = 0$, *and extended to be* 0 *for* t *among the ordinates of zeros of* ζ. *Then the sequence of random variables*

$$\frac{\mathsf{L}_T}{\sqrt{\frac{1}{2}\log\log T}}$$

converges in law as $T \to +\infty$ *to a standard gaussian random variable.*

The difference with Theorem 3.1.2 is that we only consider the modulus of $\zeta(1/2+it)$, or in other words we consider only the real part of the random variables of Theorem 3.1.2.

The strategy of the proof is again to use approximations, and to obtain the desired probabilistic behavior from the independence of the vector $t \mapsto (p^{-it})_p$ (as in Proposition 3.2.4). However, one has to be much more careful than in the previous section. The approximation used by Radziwill and Soundararajan goes in three steps:

- An approximation of L_T with the random variable $\tilde{\mathsf{L}}_T$ given by $t \mapsto \log|\zeta(\sigma_0+it)|$ for σ_0 sufficiently close to $1/2$ (where σ_0 depends on T);
- For the random variable Z_T given by $t \mapsto \zeta(\sigma_0+it)$, an approximation of the inverse $1/\mathsf{Z}_T$ by a short Dirichlet polynomial D_T of the type
$$\mathsf{D}_T(s) = \sum_{n\geqslant 1} a_T(n)\mu(n)n^{-s}$$
where $a_T(n)$ is zero for n large enough (again, depending on T);
- An approximation of D_T by what is essentially a short Euler product, namely by $\exp(\mathsf{P}_T)$, where
$$\mathsf{P}_T(s) = \sum_{p^k\leqslant X} \frac{1}{k}\frac{1}{p^{ks}}$$
for suitable X (again depending on T).

Finally, the last probabilistic step is to prove that the random variables

$$t \mapsto \frac{\mathrm{Re}(\mathsf{P}_T(\sigma_0+it))}{\sqrt{\frac{1}{2}\log\log T}}$$

converge in law to a standard normal random variable as $T \to +\infty$, where σ_0 is the specific real part of the first step (that depends therefore on T).

None of these steps is especially easy (in comparison with the results discussed up to now), and the specific approximations that are used (namely, the choices of the coefficients $a_T(n)$ as well as of the length parameter X) are quite subtle and by no means obvious. Even the nature of the approximation will not be the same in the three steps!

In order to simplify the reading of the proof, we first specify the relevant parameters. For $T \geqslant 3$, we denote

$$\varrho = \varrho_T = \sqrt{\log\log T}$$

the normalizing factor in the theorem. We then define

$$W = (\log\log\log T)^4 \asymp (\log\varrho)^4, \qquad \sigma_0 = \frac{1}{2} + \frac{W}{\log T} = \frac{1}{2} + O\Big(\frac{(\log\varrho)^4}{\log T}\Big) \tag{3.10}$$

$$X = T^{1/(\log\log\log T)^2} = T^{1/\sqrt{W}}. \tag{3.11}$$

Note that we usually omit the dependency on T in these notation. We will also require a further parameter

$$Y = T^{1/(\log\log T)^2} = T^{16/\varrho^4} \leqslant X. \tag{3.12}$$

We begin by stating the precise approximation statements. We will then show how they combine to imply Theorem 3.4.1, and finally we will prove them in turn. We note that unless otherwise specified, all results in this section are due to Radziwill and Soundararajan.

PROPOSITION 3.4.2 (Moving outside of the critical line). *With notation as above, we have*

$$\mathbf{E}_T\big(|\mathsf{L}_T - \tilde{\mathsf{L}}_T|\big) = o(\varrho_T)$$

as $T \to +\infty$.

We now define properly the Dirichlet polynomial of the second step. We denote by $\mathcal{D}_T$ the set of squarefree integers $n \geqslant 1$ such that

(1) All prime factors of n are $\leqslant X$;

(2) There are at most $\omega_1 = 100\log\log T = 100\varrho_T$ prime factors p of n such that $p \leqslant Y$;

(3) There are at most $\omega_2 = 100\log\log\log T \sim 100\log\varrho_T$ prime factors p of n such that $Y < p \leqslant X$.

Let a_T be the characteristic function of the set $\mathcal{D}_T$. Then we define

$$D(s) = \sum_{n\geqslant 1} a_T(n)\mu(n)n^{-s}$$

for $s \in \mathbf{C}$. This is a finite sum and therefore an entire function (see the next remark).

REMARK 3.4.3. Although the definition of $D(s)$ may seem complicated, we will see its different components coming together in the proofs of this proposition and the next. A first comment is that $\mathcal{D}_T$ is a set of relatively small integers: if $n \in \mathcal{D}_T$, then we have

$$n \leqslant Y^{100\log\log T} X^{100\log\log\log T} = T^c$$

where

$$c = \frac{100}{\log\log T} + \frac{100}{\log\log\log T} \to 0.$$

Moreover, we can express a_T as the Dirichlet convolution of the characteristic functions a'_T and a''_T of squarefree integers satisfying the first and the second (resp. the third) condition in Proposition 3.4.4. Each is also supported on $n \leqslant T^{\varepsilon}$ for arbitrarily small ε. Among those integers, note that (by the Erdős-Kac Theorem, extended to squarefree integers) the typical number of prime factors is still about $\log\log T$. Therefore the integers satisfying the second condition are quite typical, and only extreme outliers (in terms of number of prime factors) are excluded. However, the integers satisfying the third condition have much fewer prime factors than is typical, and are therefore very rare. This indicates that a_T is a subtle *arithmetic* truncation of the characteristic function of integers $n \leqslant T^c$, and hence that

$$\sum_{n\geqslant 1} a_T(n)\mu(n)n^{-s}$$

is an arithmetic truncation of the Dirichlet series that formally gives the inverse of $\zeta(s)$. This should be contrasted with the more traditional *analytic* truncations used in Lemma 3.2.8 and Proposition 3.2.9. Selberg used, for this purpose and for many other results concerning the Riemann zeta function, some truncations that are roughly of the shape

$$\sum_{n\leqslant X} \frac{\mu(n)}{n^s}\Big(1 - \frac{\log n}{\log X}\Big).$$

PROPOSITION 3.4.4 (Dirichlet polynomial approximation). *With notation as above, Let* D_T *be the random variable on* Ω_T *defined by*

$$t \mapsto D(\sigma_0 + it).$$

We then have

$$\mathbf{E}_T(|1 - \mathsf{Z}_T\mathsf{D}_T|^2) \to 0$$

as $T \to +\infty$.

PROPOSITION 3.4.5 (Short Euler product approximation). *With notation as above, for any* $\varepsilon > 0$, *we have*

$$\mathbf{P}_T\big(|\mathsf{D}_T \exp(\mathsf{P}_T) - 1| > \varepsilon\big) \to 0$$

as $T \to +\infty$.

Note that one can also state this last result as stating that the random variables $\mathsf{D}_T \exp(\mathsf{P}_T)$ converge to 1 in probability. However, these three previous statement are really theorems of number theory, and could have been expressed without any probabilistic notation; for instance, the first one means that

$$\frac{1}{T}\int_{-T}^{T} |\log|\zeta(1/2+it)| - \log|\zeta(\sigma_0+it)||dt = o(\log\log T).$$

The last result finally introduces the probabilistic behavior,

PROPOSITION 3.4.6 (Normal Euler products). *With notation as above, the random variables* $\varrho_T^{-1}\mathsf{P}_T$ *converge in law to a standard complex gaussian as* $T \to +\infty$. *In particular, we have convergence in law*

$$\frac{\mathrm{Re}(\mathsf{P}_T)}{\sqrt{\frac{1}{2}\log\log T}} \to \mathcal{N}$$

as $T \to +\infty$.

We now show how to combine these ingredients.

PROOF OF THEOREM 3.4.1. Until Proposition 3.4.6 is used, this is essentially a variant of the fact that convergence in probability implies convergence in law, and that convergence in L^1 or L^2 implies convergence in probability.

For the details, fix some standard Gaussian random variable $\mathcal{N}$. Let f be a bounded Lipschitz function $\mathbf{R} \longrightarrow \mathbf{R}$, with

$$|f(x) - f(y)| \leqslant C|x-y|, \qquad |f(x)| \leqslant C,$$

for some $C \geqslant 0$. We consider the difference

$$\mathbf{E}_T\Big(f\Big(\frac{\mathsf{L}_T}{\varrho_T}\Big)\Big) - \mathbf{E}(f(\mathcal{N})),$$

and must show that this tends to 0 as $T \to +\infty$.

We estimate this quantity using the "chain" of approximations introduced above: we have

$$\begin{aligned}(3.13)\quad \Big|\mathbf{E}_T\Big(f\Big(\frac{\mathsf{L}_T}{\varrho_T}\Big)\Big) - \mathbf{E}(f(\mathcal{N}))\Big| \leqslant \\ \mathbf{E}_T\Big(\Big|f\Big(\frac{\mathsf{L}_T}{\varrho_T}\Big) - f\Big(\frac{\tilde{\mathsf{L}}_T}{\varrho_T}\Big)\Big|\Big) + \mathbf{E}_T\Big(\Big|f\Big(\frac{\tilde{\mathsf{L}}_T}{\varrho_T}\Big) - f\Big(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\Big)\Big|\Big) + \\ \mathbf{E}_T\Big(\Big|f\Big(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\Big) - f\Big(\frac{\mathrm{Re}\,\mathsf{P}_T}{\varrho_T}\Big)\Big|\Big) + \Big|\mathbf{E}_T\Big(f\Big(\frac{\mathrm{Re}\,\mathsf{P}_T}{\varrho_T}\Big)\Big) - \mathbf{E}(f(\mathcal{N}))\Big|,\end{aligned}$$

and we discuss each of the four terms on the right-hand side using the four previous propositions.

The first one is handled straightforwardly using Proposition 3.4.2: we have

$$\mathbf{E}_T\Big(\Big|f\Big(\frac{\mathsf{L}_T}{\varrho_T}\Big) - f\Big(\frac{\tilde{\mathsf{L}}_T}{\varrho_T}\Big)\Big|\Big) \leqslant \frac{C}{\varrho_T}\,\mathbf{E}_T(|\mathsf{L}_T - \tilde{\mathsf{L}}_T|) \longrightarrow 0$$

as $T \to +\infty$.

For the second term, let $A_T \subset \Omega_T$ be the event

$$|\tilde{\mathsf{L}}_T - \log|\mathsf{D}_T|^{-1}| > 1/2,$$

and A_T' its complement. Since $\log|\mathsf{Z}_T| = \tilde{\mathsf{L}}_T$, we then have

$$\mathbf{E}_T\Big(\Big|f\Big(\frac{\tilde{\mathsf{L}}_T}{\varrho_T}\Big) - f\Big(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\Big)\Big|\Big) \leqslant 2C\,\mathbf{P}_T(A_T) + \frac{C}{2\varrho_T}.$$

Proposition 3.4.5 implies that $\mathbf{P}_T(A_T) \to 0$ (convergence to 1 in L^2 implies convergence to 1 in probability for $\mathsf{Z}_T\mathsf{D}_T$, hence convergence to 0 in probability for the logarithm of the modulus) and therefore

$$\mathbf{E}_T\Big(\Big|f\Big(\frac{\tilde{\mathsf{L}}_T}{\varrho_T}\Big) - f\Big(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\Big)\Big|\Big) \to 0$$

as $T \to +\infty$.

We now come to the third term in (3.13). Distinguishing according to the events

$$B_T = \big\{\big|\log|\mathsf{D}_T\exp(\mathsf{P}_T)|\big| > 1/2\big\}$$

and its complement, we get as before

$$\mathbf{E}_T\Big(\Big|f\Big(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\Big) - f\Big(\frac{\mathrm{Re}\,\mathsf{P}_T}{\varrho_T}\Big)\Big|\Big) \leqslant 2C\,\mathbf{P}_T(B_T) + \frac{C}{2\varrho_T},$$

and this tends to 0 as $T \to +\infty$ by Proposition 3.4.5.

Finally, Proposition 3.4.6 implies that

$$\Big|\mathbf{E}_T\Big(f\Big(\frac{\mathrm{Re}\,\mathsf{P}_T}{\varrho_T}\Big)\Big) - \mathbf{E}(f(\mathcal{N}))\Big| \to 0$$

as $T \to +\infty$, and hence we conclude the proof of the theorem, assuming the approximation statements. □

We now explain the proofs of these four propositions. A key tool is the quantitative form of Proposition 3.2.4 contained in Lemma 3.2.5.

Proof of Proposition 3.4.2. □

Proof of Proposition 3.4.4. □

Proof of Proposition 3.4.5. □

Proof of Proposition 3.4.6. □

3.5. Further topics

If we look back at the proof of Bagchi's Theorem, and at the proof of Voronin's Theorem, to see precisely which arithmetic ingredients appear, we can see (as in Chapter 2) that relatively little is really needed. Indeed, arithmetic only appears in the following steps of the argument:

- In Proposition 3.2.4, which depends on the unique factorization of integers into primes, and which illustrates again the asymptotic independence of prime numbers, similarly to Proposition 1.2.5;
- In the crucial approximation step of Proposition 3.2.10, where the mean-value property 3.8 of the Riemann zeta function is required; this estimate has arithmetic meaning through the large sieve inequalities that enter into its proof.
- In some features of the Random Dirichlet Series that arises as the limit of translates of the Riemann zeta function, because of its Euler product expansion, or equivalently, because of the multiplicativity of its coefficients (this contrasts with the Erdős-Kac Theorem, where the limit is the universal gaussian distribution).
- In the proof of Voronin's Theorem, where the Prime Number Theorem is used to control the distribution of primes in (roughly) dyadic intervals.

Because of this, it is not very surprising that Bagchi's Theorem can be generalized to many other situations. The most interesting concerns perhaps the limiting behavior, in $\mathcal{H}(D)$, of families of L-functions of the type

$$\sum_{n\geqslant 1} \lambda_f(n) n^{-s}$$

where f runs over some sequence of arithmetic objects with associated L-functions, ordered in a sequence of probability spaces (which need not be continuous like Ω_T). We refer to [15, Ch. 5] or [**?**,] for surveys of the possible types of L-functions that might be used here. There are some rather obvious special cases, such as the vertical translates $L(\chi, s+it)$ of a fixed Dirichlet L-function $L(\chi, s)$, since all properties of the Riemann zeta function extend to this case.

However, extending Bagchi's Theorem to vertical translates of an L-function of higher rank requires restrictions, in the current state of knowledge, because the analogue of (3.8) is only known for the average over $\sigma + it$ for $\sigma \in]1/2, 1[$ large enough (depending on f). Thus only domains D contained in a region to the right of this limit may be handled.

One may also vary the character (and avoid the vertical shifts). This would mean considering the probability space of (say) all primitive Dirichlet characters modulo some integer q, and the statistic distribution of the restriction of $L(s, \chi)$ to D.

[**Further references:** Titchmarsh [35], especially Chapter 11, discusses the older work of Bohr and Jessen, which has some interesting geometric aspects that are not apparent in modern treatments. Bagchi's Thesis [1] contains some generalizations as well as more information concerning the limit theorem and Voronin's Theorem.]

CHAPTER 4

The shape of exponential sums

4.1. Introduction

We consider in this chapter a rather type of arithmetic objects: exponential sums and their partial sums. Although the principles apply to very general situations, we consider as usual only an important special case: the partial sums of *Kloosterman sums* modulo primes.

Thus let p be a prime number. For any pair (a, b) of invertible elements in the finite field $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$, the (normalized) Kloosterman sum $S(a, b; p)$ is defined by the formula

$$S(a,b;p) = \frac{1}{\sqrt{p}} \sum_{x \in \mathbf{F}_p^\times} e\Big(\frac{ax + b\bar{x}}{p}\Big),$$

where we recall that we denote by $e(z)$ the 1-periodic function defined by $e(z) = e^{2i\pi z}$, and that $\bar{x}$ is the inverse of x modulo p.

These are finite sums, and they are of great importance in many areas of number theory, especially in relation with automorphic forms and with analytic number theory (see [19] for a survey of the origin of these sums and of their applications, due to Poincaré, Kloosterman, Linnik, Iwaniec, and others). Among their remarkable properties is the following estimate for the modulus of $S(a, b; p)$, due to A. Weil: for any $(a, b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times$, we have

$$|S(a,b;p)| \leqslant 2.$$

This is a very strong result if one considers that $S(a, b; p)$ is, up to dividing by $\sqrt{p}$, the sum of $p - 1$ roots of unity, so that the only "trivial" estimate is that $|S(a, b; p)| \leqslant (p - 1)/\sqrt{p}$. What this reveals is that the arguments of the summands $e((ax + b\bar{x})/p)$ in $\mathbf{C}$ vary in a very complicated manner that leads to this remarkable cancellation property. This is due essentially to the very "random" behavior of the map $x \mapsto \bar{x}$ when seen at the level of representatives of x and $\bar{x}$ in the interval $\{0, \dots, p - 1\}$.

From a probabilistic point of view, the order of magnitude $\sqrt{p}$ of the sum (before normalization) is not unexpected. If we simply heuristically model an exponential sum as above by a random walk with independent summands uniformly distributed on the unit circle, say

$$X_N = S_1 + \cdots + S_N$$

(where the (S_n) are independent and uniform on the unit circle), then the Central Limit Theorem implies a convergence in law of $X_N/\sqrt{N}$ to a standard complex gaussian random variable, which shows that $\sqrt{N}$ is the "right" order of magnitude.

This probabilistic analogy and the study of random walks (or sheer curiosity) suggests to look at the partial sums of Kloosterman sums, and the way they move in the complex plane. This requires some ordering of the sum defining $S(a, b; p)$, which we simply arrange by summing over $1 \leqslant x \leqslant p - 1$ in increasing order. Thus we will consider the $p - 1$ points

$$\frac{1}{\sqrt{p}} \sum_{1 \leqslant x \leqslant n} e\Big(\frac{ax + b\bar{x}}{p}\Big)$$

for $1 \leqslant n \leqslant p - 1$. We illustrate this for the sum $S(1, 1; 139)$ in Figure 4.1.

Because this cloud of points is not particularly enlightening, we refine the construction by joining the successive points with line segments. This gives the result in Figure 4.2 for $S(1, 1; 139)$. If we change the values of a and b, we observe that the figures change in apparently

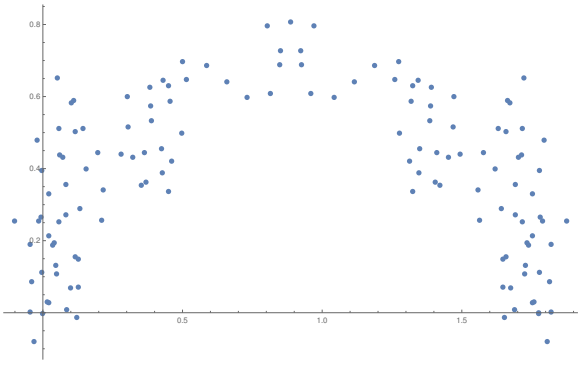

FIGURE 4.1. The partial sums of $S(1,1;139)$.

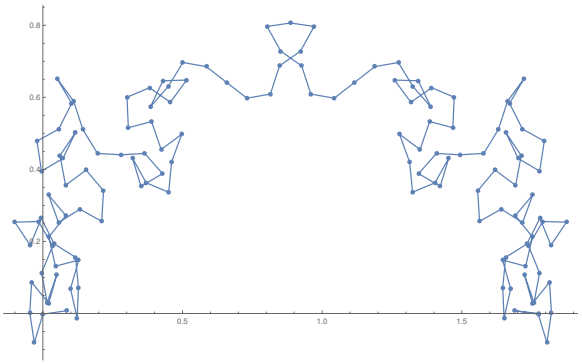

FIGURE 4.2. The partial sums of $S(1,1;139)$, joined by line segments.

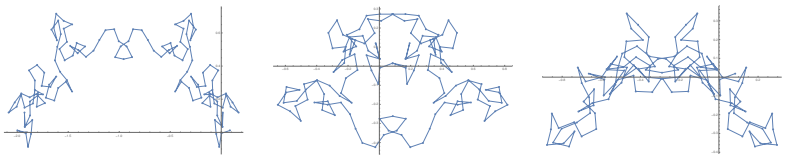

FIGURE 4.3. The partial sums of $S(a,1;139)$ for $a = 2,3,4$.

random way, although some basic features remain (the final point is on the real axis, which reflects the easily-proven fact that $S(a,b;p) \in \mathbf{R}$, and there is a reflection symmetry with respect to the line $x = \frac{1}{2}S(a,b;p)$). For instance, Figure 4.3 shows the curves corresponding to $S(2,1;139)$, $S(3,1;139)$ and $S(4,1;139)$.

If we vary a and b, for a fixed p, we see that the shapes of these polygonal paths changes in seemingly unpredictable manner (see [22] for many more pictures). We then ask whether there is a definite statistical behavior for these *Kloosterman paths* as $p \to +\infty$, when we pick $(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times$ uniformly at random. As we will see, this is indeed the case!

To state the precise result, we introduce some further notation. Thus, for p prime and $(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times$, we denote by $\mathsf{K}_p(a,b)$ the function

$$[0,1] \longrightarrow \mathbf{C}$$

such that, for $0 \leqslant j \leqslant p-2$, the value at t such that

$$\frac{j}{p-1} \leqslant t < \frac{j+1}{p-1}$$

is obtained by interpolating linearly between the consecutive partial sums

$$\frac{1}{\sqrt{p}}\sum_{1\leqslant x\leqslant j} e\Big(\frac{ax+b\bar{x}}{p}\Big) \text{ and } \frac{1}{\sqrt{p}}\sum_{1\leqslant x\leqslant j+1} e\Big(\frac{ax+b\bar{x}}{p}\Big).$$

The path $t\mapsto \mathsf{K}_p(a,b)(t)$ is the polygonal path described above; for $t=0$, we have $\mathsf{K}_p(a,b)(0)=0$, and for $t=1$, we obtain $\mathsf{K}_p(a,b)(1)=S(a,b;p)$.

Let $\Omega_p=\mathbf{F}_p^\times\times\mathbf{F}_p^\times$. We view K_p as a random variable

$$\Omega\longrightarrow C([0,1]),$$

where $C([0,1])$ is the Banach space of continuous functions $\varphi:[0,1]\to\mathbf{C}$ with the supremum norm $\|\varphi\|_\infty=\sup|\varphi(t)|$. Alternatively, we sometimes think of the *family* of random variables $(\mathsf{K}_p(t))_{t\in[0,1]}$ such that

$$(a,b)\mapsto \mathsf{K}_p(a,b)(t),$$

and view it as a stochastic process with t playing the role of "time".

Here is the theorem that gives the limiting behavior of these arithmetically-defined random variables (or processes), proved in [24].

THEOREM 4.1.1 (Kowalski–Sawin). *Let $(\mathrm{ST}_h)_{h\in\mathbf{Z}}$ be a sequence of independent random variables, all distributed according to the Sato-Tate measure*

$$\mu_{ST}=\frac{1}{\pi}\sqrt{1-\frac{x^2}{4}}dx$$

on $[-2,2]$.

(1) The random series

$$K(t)=t\mathrm{ST}_0+\sum_{\substack{h\in\mathbf{Z}\\ h\neq 0}}\frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h$$

defined for $t\in[0,1]$ converges uniformly almost surely, in the sense of symmetric partial sums

$$K(t)=t\mathrm{ST}_0+\lim_{H\to+\infty}\sum_{\substack{h\in\mathbf{Z}\\ 1\leqslant|h|<H}}\frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h.$$

This random series defines a $C([0,1])$-valued random variable K.

(2) As $p\to+\infty$, the random variables K_p converge in law to K, in the sense of $C([0,1])$-valued variables.

The Sato-Tate measure is better known in probability a a semi-circle law, but its appearance in Theorem 4.1.1 is really due to the group-theoretic interpretation that often arises in number theory, and reflects the choice of name. Namely, we recall (see Example B.4.1 (3)) that μ_{ST} is the direct image under the trace map of the probability Haar measure on the compact group $\mathrm{SU}_2(\mathbf{C})$.

Note in particular that the theorem implies, by taking $t=1$, that the Kloosterman sums $S(a,b;p)=\mathsf{K}_p(a,b)(1)$, viewed as random variables on Ω_p, become asymptotically distributed like $K(1)=\mathrm{ST}_0$, i.e., that Kloosterman sums are Sato-Tate distributed in the sense that for any real numbers $-2\leqslant\alpha<\beta\leqslant 2$, we have

$$\frac{1}{(p-1)^2}|\{(a,b)\in\mathbf{F}_p^\times\times\mathbf{F}_p^\times\ \mid\ \alpha<S(a,b;p)<\beta\}|\longrightarrow\int_\alpha^\beta d\mu_{ST}(t).$$

This result is a famous theorem of N. Katz [16]. In some sense, Theorem 4.1.1 is a "functional" extension of this equidistribution theorem. In fact, the key arithmetic ingredient in the proof is a relatively simple extension of the results and methods developed by Katz in the proof of such statements.

4.2. Proof of the distribution theorem

We will explain the proof of the theorem, following to some extent the original article but with some simplifications arising from the consideration of this single example (instead of more general settings, as found in [24]), and with a number of probabilistic steps explained in detail, instead of using references to standard properties of random Fourier series.

The proof will be complete from a probabilistic point of view, but it relies on an extremely deep arithmetic result that we will only be able to view as a black box in this book. The crucial underlying result is the very general form of the Riemann Hypothesis over finite fields, and the formalism that is attached to it, which is due to Deligne, with the particular applications we use relying extensively on the additional work of Katz. All of this builds on the algebraic-geometric foundations of Grothendieck and his school, and we give a few further references in Section 4.4.

In outline, the proof has three steps:

(1) Show that the random Fourier series K exists, as a $C([0,1])$-valued random variable, and establish properties such as existence of moments of all order for any $K(t)$;

(2) Prove that K_p converges to K in the sense of finite distributions (Definition B.9.2);

(3) Prove that the sequence $(\mathsf{K}_p)_p$ is tight (Definition B.2.4), using Kolmogorov's Criterion (Proposition B.9.5).

Once this is done, Prokhorov's Theorem (Theorem B.9.4) shows that the combination of (2) and (3) implies that K_p converges to K.

We denote by $\mathbf{P}_p$ and $\mathbf{E}_p$ the probability and expectation with respect to the uniform measure on $\Omega_p = \mathbf{F}_p^\times \times \mathbf{F}_p^\times$. Before we begin the proof in earnest, it is useful to see *why* the limit arises, and why it is precisely this random Fourier series. The idea is to use discrete Fourier analysis to represent the partial sums of Kloosterman sums.

LEMMA 4.2.1. *Let $p \geqslant 3$ be a prime and a, $b \in \mathbf{F}_p^\times$. Let $t \in [0,1]$. Then we have*

$$\frac{1}{\sqrt{p}} \sum_{1 \leqslant n \leqslant (p-1)t} e\Big(\frac{an + b\bar{n}}{p}\Big) = \sum_{|h|<p/2} \alpha_p(h,t) S(a-h,b;p), \tag{4.1}$$

where

$$\alpha_p(h,t) = \frac{1}{p} \sum_{1 \leqslant n \leqslant (p-1)t} e\Big(\frac{nh}{p}\Big).$$

PROOF. This is a case of the discrete Plancherel formula, applied to the characteristic function of the discrete interval of summation; to check it quickly, insert the definitions of $\alpha_p(h,t)$ and of $S(a-h,b;p)$ in the right hand-side of (4.1). This shows that it is equal to

$$\begin{aligned}
\sum_{|h|<p/2} \alpha_p(h,t) S(a-h,b;p) &= \frac{1}{p^{3/2}} \sum_{|h|<p/2} \sum_{1 \leqslant n \leqslant (p-1)t} \sum_{m \in \mathbf{F}_p} e\Big(\frac{nh}{p}\Big) e\Big(\frac{(a-h)m + b\bar{m}}{p}\Big) \\
&= \frac{1}{\sqrt{p}} \sum_{1 \leqslant n \leqslant (p-1)t} \sum_{m \in \mathbf{F}_p} e\Big(\frac{am + b\bar{m}}{p}\Big) \frac{1}{p} \sum_{h \in \mathbf{F}_p} e\Big(\frac{h(n-m)}{p}\Big) \\
&= \frac{1}{\sqrt{p}} \sum_{1 \leqslant n \leqslant (p-1)t} e\Big(\frac{an + b\bar{n}}{p}\Big),
\end{aligned}$$

as claimed, since by the orthogonality of characters we have

$$\frac{1}{p} \sum_{h \in \mathbf{F}_p} e\Big(\frac{h(n-m)}{p}\Big) = \delta(n,m)$$

for any n, $m \in \mathbf{F}_p$. □

If we observe that $\alpha_p(h,t)$ is essentially a Riemann sum for the integral

$$\int_0^t e(ht)dt = \frac{e(ht)-1}{2i\pi h}$$

for all $h \neq 0$, and that $\alpha_p(0,t) \to t$ as $p \to +\infty$, we see that the right-hand side of (4.1) looks like a Fourier series of the same type as $K(t)$, with coefficients given by shifted Kloosterman sums $S(a-h,b;p)$ instead of ST_h. Now the crucial arithmetic information is contained in the following very deep theorem:

THEOREM 4.2.2 (Katz; Deligne). *Fix an integer $b \neq 0$. For p prime not dividing b, consider the random variable*

$$\mathsf{S}_p \,:\, a \mapsto (S(a-h,b;p))_{h\in\mathbf{Z}}$$

on $\mathbf{F}_p^\times$ with uniform probability measure, taking values in the compact topological space

$$\hat{T} = \prod_{h\in\mathbf{Z}} [-2,2]$$

Then S_p converges in law to the product probability measure

$$\bigotimes_{h\in\mathbf{Z}} \mu_{ST}.$$

In other words, the sequence of random variables $a \mapsto S(a-h,b;p)$ converges in law to a sequence $(\mathrm{ST}_h)_{h\in\mathbf{Z}}$ of independent Sato-Tate distributed random variables.

Because of this theorem, the formula (4.1) suggests that $\mathsf{K}_p(t)$ converges in law to the random series

$$t\mathrm{ST}_0 + \sum_{\substack{h\in\mathbf{Z}\\ h\neq 0}} \frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h,$$

which is exactly $K(t)$. We now proceed to the implementation of the three steps above, which will use this deep arithmetic ingredient.

REMARK 4.2.3. There is a subtlely in the argument: although Theorem 4.2.2 holds for any fixed b, when averaging only over a, we cannot at the current time prove the analogue of Theorem 4.1.1 for fixed b, because the proof of tightness in the last step uses crucially both averages.

Step 1. (Existence and properties of the random Fourier series)
We can write the series $K(t)$ as

$$K(t) = t\mathrm{ST}_0 + \sum_{h\geqslant 1}\Big(\frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h - \frac{e(-ht)-1}{2i\pi h}\mathrm{ST}_{-h}\Big).$$

The summands here, namely

$$X_h = \frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h - \frac{e(-ht)-1}{2i\pi h}\mathrm{ST}_{-h}$$

for $h \geqslant 1$, are independent and have expectation 0 since $\mathbf{E}(\mathrm{ST}_h) = 0$ (see (B.5)). Furthermore, since ST_h is independent of ST_{-h}, and they have variance 1, we have

$$\sum_{h\geqslant 1}\mathbf{V}(X_h) = \sum_{h\geqslant 1}\Big(\Big|\frac{e(ht)-1}{2i\pi h}\Big|^2 + \Big|\frac{e(-ht)-1}{2i\pi h}\Big|^2\Big) \leqslant \sum_{h\geqslant 1}\frac{1}{h^2} < +\infty$$

for any $t \in [0,1]$. From Kolmogorov's criterion for almost sure convergence of random series with finite variance (Theorem B.8.1), it follows that for any $t \in [0,1]$, the series $K(t)$ converges almost surely to a complex-valued random variable.

To prove convergence in $C([0,1])$, we will use convergence of finite distributions combined with Kolmogorov's Tightness Criterion Consider the partial sums

$$K_H(t) = t\mathrm{ST}_0 + \sum_{1\leqslant|h|\leqslant H}\frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h$$

for $H \geqslant 1$. These are $C([0,1])$-valued random variables. The convergence of $K_H(t)$ to $K(t)$ in L^1, for any $t \in [0,1]$, implies (see Lemma B.9.3) that the sequence $(K_H)_{H\geqslant 1}$ converges to K in the sense of finite distributions. Therefore, by Proposition B.9.5, the sequence converges in the sense of $C([0,1])$-valued random variables if there exist constants $C \geqslant 0$, $\alpha > 0$ and $\delta > 0$ such that for any $H \geqslant 1$, and real numbers $0 \leqslant s < t \leqslant 1$, we have

$$\mathbf{E}(|K_H(t) - K_H(s)|^\alpha) \leqslant C|t-s|^{1+\delta}. \tag{4.2}$$

We will take $\alpha = 4$. We have

$$K_H(t) - K_H(s) = (t-s)\mathrm{ST}_0 + \sum_{1\leqslant|h|\leqslant H} \frac{e(ht) - e(hs)}{2i\pi h}\mathrm{ST}_h.$$

This is a sum of independent, centered and bounded random variables, so that by Proposition B.6.2 (1) and (2), it is σ_H^2-subgaussian with

$$\sigma_H^2 = |t-s|^2 + \sum_{1\leqslant|h|\leqslant H} \Big|\frac{e(ht) - e(hs)}{2i\pi h}\Big|^2.$$

By Parseval's formula for ordinary Fourier series, we see that σ_H^2 is equal to

$$\int_0^1 |f(x)|^2 dx,$$

where f is the characteristic function of the interval $[s,t]$. Therefore $\sigma_H^2 = |t-s|$. By the properties of subgaussian random variables (see Proposition B.6.3 in Section B.6), we deduce that there exists $C \geqslant 0$ such that

$$\mathbf{E}(|K_H(t) - K_H(s)|^4) \leqslant C\sigma_H^4 = C|t-s|^2,$$

which establishes (4.2).

Some additional remarks will be useful in the next step. First, since

$$\sigma_t^2 = |t|^2 + \sum_{h\neq 0} \Big|\frac{e(ht) - 1}{2i\pi h}\Big|^2 < +\infty$$

for all t, the series defining $K(t)$ converges in L^2 for any t. Applying Proposition B.6.2 again shows that $K(t)$ is σ_t^2-subgaussian.

Similarly, for any H, the remainder $K(t) - K_H(t)$ is $\sigma_{t,H}^2$-subgaussian with

$$\sigma_{t,H}^2 = \sum_{|h|>H} \Big|\frac{e(ht) - 1}{2i\pi h}\Big|^2.$$

So Proposition B.6.3 implies that, for any integer $k \geqslant 0$, there exists a constant c_k such that

$$\mathbf{E}(|K(t) - K_H(t)|^k) \leqslant c_k \sigma_{t,H}^k.$$

Since $\sigma_{t,H} \to 0$ as H tends to infinity (for any fixed t), it follows that K_H converges to K in L^k as H tends to infinity.

Step 2. (Convergence in the sense of finite distributions)

Fix an integer $k \geqslant 1$ and real numbers

$$0 \leqslant t_1 < \cdots < t_k \leqslant 1.$$

The goal is to prove that the vectors

$$(\mathsf{K}_p(t_1), \ldots, \mathsf{K}_p(t_k))$$

converge in law to $(K(t_1), \ldots, K(t_k))$ as p tends to infinity. Because the real numbers t_i are fixed, we may replace $\mathsf{K}_p(t)$ by the "discontinuous" version

$$\tilde{\mathsf{K}}_p(t) = \frac{1}{\sqrt{p}} \sum_{1\leqslant n\leqslant (p-1)t} e\Big(\frac{an + b\bar{n}}{p}\Big).$$

Indeed, we have

$$|\tilde{\mathsf{K}}_p(t) - \mathsf{K}_p(t)| \leqslant \frac{1}{\sqrt{p}}$$

for any t, and therefore Lemma B.3.3 shows that the convergence in law of

$$(\tilde{\mathsf{K}}_p(t_1), \dots, \tilde{\mathsf{K}}_p(t_k)) \tag{4.3}$$

to $(K(t_1), \dots, K(t_k))$ implies that of $(\mathsf{K}_p(t_1), \dots, \mathsf{K}_p(t_k))$ to the same limit.

To prove the convergence of (4.3), we may use the method of moments. Indeed, each $K(t)$ is subgaussian, as we saw in Step 1, and a subgaussian random variable has Laplace transform defined everywhere.

Let $n_1, \dots, n_k$ and $m_1, \dots, m_k$ be non-negative integers. We denote

$$A = \sum_i n_i + \sum_i m_i.$$

We will show that the moments

$$M_p = \mathbf{E}_p\Big(\tilde{\mathsf{K}}_p(t_1)^{n_1}\overline{\tilde{\mathsf{K}}_p(t_1)}^{m_1} \cdots \tilde{\mathsf{K}}_p(t_k)^{n_k}\overline{\tilde{\mathsf{K}}_p(t_k)}^{m_k}\Big)$$

converge to

$$M = \mathbf{E}\Big(K(t_1)^{n_1}\overline{K(t_1)}^{m_1} \cdots K(t_k)^{n_k}\overline{K(t_k)}^{m_k}\Big). \tag{4.4}$$

We insert in the definition of M_p the discrete Fourier expansion (4.1) for each $\tilde{AK}_p(t_i)$ and their powers. Exchanging the resulting sums and expectations, we obtain a sum over k tuples $\boldsymbol{h}_i$ of integers, with

$$\boldsymbol{h}_i = (h_{i,1}, \dots, h_{i,n_i}, h_{i,n_i+1}, \dots, h_{i,n_i+m_i})$$

itself an $(n_i + m_i)$-tuple of integers, each of which is (strictly) between $-p/2$ and $p/2$. If we denote

$$\alpha_p(\boldsymbol{h}_i, t_i) = \prod_{j=1}^{n_i} \alpha_p(h_{i,j}, t) \prod_{j=n_i+1}^{n_i+m_i} \overline{\alpha_p(h_{i,j}, t)},$$

the resulting formula is

$$M_p = \sum_{\boldsymbol{h}_1,\dots,\boldsymbol{h}_k} \cdots \sum \alpha_p(\boldsymbol{h}_1, t_1) \cdots \alpha_p(\boldsymbol{h}_k, t_k) \tag{4.5}$$
$$\mathbf{E}_p\Big(\prod_{j=1}^{n_1+m_1} S(a - h_{1,j}, b; p) \cdots \prod_{j=1}^{n_k+m_k} S(a - h_{k,j}, b; p)\Big)$$

(where we use the fact that Kloosterman sums are real numbers).

For each fixed tuples $(\boldsymbol{h}_1, \dots, \boldsymbol{h}_k)$, we can rearrange the expectation that occurs according to the multiplicities of the shifts $h_{j,i}$: for any integer h with $|h| < p/2$, we denote

$$\nu(h) = \sum_{i=1}^{k} |\{j \leqslant n_i + m_i \mid h_{i,j} = h\}|.$$

Then we have

$$\mathbf{E}_p\Big(\prod_{j=1}^{n_1+m_1} S(a - h_{1,j}, b; p) \cdots \prod_{j=1}^{n_k+m_k} S(a - h_{k,j}, b; p)\Big) = \mathbf{E}_p\Big(\prod_{|h|<p/2} S(a - h, b; p)^{\nu(h)}\Big).$$

According to Theorem 4.2.2, we have then

$$\mathbf{E}_p\Big(\prod_{|h|<p/2} S(a - h, b; p)^{\nu(h)}\Big) \longrightarrow \mathbf{E}\Big(\prod_{|h|<p/2} \mathrm{ST}_h^{\nu(h)}\Big)$$

as $p \to +\infty$. This implies, by reversing the computation, that M_p is close to

$$\sum_{\boldsymbol{h}_1,\dots,\boldsymbol{h}_k} \cdots \sum \alpha_p(\boldsymbol{h}_1,t_1)\cdots\alpha_p(\boldsymbol{h}_k,t_k)\,\mathbf{E}\Big(\prod_{|h|<p/2} \mathrm{ST}_h^{\nu(h)}\Big) = \mathbf{E}\Big(\tilde{K}_p(t_1)^{n_1}\overline{\tilde{K}_p(t_1)}^{m_1}\cdots\tilde{K}_p(t_k)^{n_k}\overline{\tilde{K}_p(t_k)}^{m_k}\Big)\Big)$$

with

$$\tilde{K}_p(t) = \sum_{|h|<p/2} \alpha_p(h,t)\mathrm{ST}_h.$$

However, to control the approximation in this step, we need a quantitative version of the convergence in law. It follows from the proof of Theorem 4.2.2, and especially from a strong form of the Riemann Hypothesis over finite fields of Deligne, that we have

$$\mathbf{E}_p\Big(\prod_{|h|<p/2} S(a-h,b;p)^{\nu(h)}\Big) = \mathbf{E}\Big(\prod_{|h|<p/2} \mathrm{ST}_h^{\nu(h)}\Big) + O(p^{-1/2}),$$

where the implied constant depends only on the sum A of the integers m_i and n_i.

In addition, a simple direct computation shows that for any $t \in [0,1]$, we have

$$\sum_{|h|<p/2} |\alpha_p(h,t)| \leqslant (\log 3p),$$

and therefore (4.5) leads to the approximation

$$\begin{aligned} M_p &= \sum_{\boldsymbol{h}_1,\dots,\boldsymbol{h}_k} \cdots \sum \alpha_p(\boldsymbol{h}_1,t_1)\cdots\alpha_p(\boldsymbol{h}_k,t_k)\,\mathbf{E}\Big(\prod_{|h|<p/2} \mathrm{ST}_h^{\nu(h)}\Big) + O(p^{-1/2}(\log 3p)^A) \\ &= \mathbf{E}\Big(\tilde{K}_p(t_1)^{n_1}\overline{\tilde{K}_p(t_1)}^{m_1}\cdots\tilde{K}_p(t_k)^{n_k}\overline{\tilde{K}_p(t_k)}^{m_k}\Big) + O(p^{-1/2}(\log 3p)^A). \end{aligned}$$

Then the following lemma finishes the proof of convergence of M_p to the corresponding moment M given by (4.4) for the limit random Fourier series $K(t)$.

LEMMA 4.2.4. *For $p \geqslant 3$ prime, consider $t \mapsto \tilde{K}_p(t)$ as a $C([0,1])$-valued random variable. Then $(\tilde{K}_p(t))$ converges to $(K(t))$ in the sense of finite distributions.*

PROOF. By Lemma B.9.3, it is enough to prove that for any fixed $t \in [0,1]$, the sequence $(\tilde{K}_p(t))_p$ converges in L^1 to $K(t)$. In fact, we will show convergence in L^2. Let

$$\beta_p(h,t) = \frac{e(ht)-1}{2i\pi h}$$

if $h \neq 0$, and $\beta_p(0,t) = t$, and

$$K_p(t) = \sum_{|h|<p/2} \beta_p(h,t)\mathrm{ST}_h$$

be the partial sum of the series $K(t)$. We have

$$\|\tilde{K}_p(t) - K(t)\|_{L^2} \leqslant \|\tilde{K}_p(t) - K_p(t)\|_{L^2} + \|K_p(t) - K(t)\|_{L^2}.$$

The second term tends to 0 as $p \to +\infty$, since we know (from the remarks made at the end of Step 1) that $K_p(t)$ converges to $K(t)$ in L^2. Then by independence of the random variables (ST_h), we obtain

$$\|\tilde{K}_p(t) - K_p(t)\|_{L^2}^2 = \sum_{|h|<p/2} |\alpha_p(h,t) - \beta_p(h,t)|^2.$$

Simple computations show that $|\alpha_p(h,t) - \beta_p(h,t)| \ll p^{-1}$ for all h with $|h| < p/2$ and all $t \in [0,1]$. Hence we get

$$\|\tilde{K}_p(t) - K_p(t)\|_{L^2} \ll p^{-1/2},$$

and therefore $\tilde{K}_p(t) \to K(t)$ in L^2, which implies convergence in L^1. □

REMARK 4.2.5. As explained in Remark 4.2.3, we could have fixed the value of b and averaged only over $a \in \mathbf{F}_p^\times$ in this step.

Step 3. (Tightness of the Kloosterman paths)

We now come to the last step of the proof of Theorem 4.1.1: the fact that the sequence $(\mathsf{K}_p)_p$ is tight. According to Kolmogorov's Criterion (Proposition B.9.5), it is enough to find constants $C \geqslant 0$, $\alpha > 0$ and $\delta > 0$ such that, for all primes $p \geqslant 3$ and all t and s with $0 \leqslant s < t \leqslant 1$, we have

$$\mathbf{E}_p(|\mathsf{K}_p(t) - \mathsf{K}_p(s)|^\alpha) \leqslant C|t-s|^{1+\delta}. \tag{4.6}$$

We denote by $\gamma \geqslant 0$ the real number such that

$$|t-s| = (p-1)^{-\gamma}.$$

So γ is larger when t and s are closer. The proof of (4.6) involves two different ranges.

First range. Assume that $\gamma > 1$ (that is, that $|t-s| < 1/(p-1)$). In that range, we use the polygonal nature of the paths $x \mapsto \mathsf{K}_p(x)$, which implies that

$$|\mathsf{K}_p(t) - \mathsf{K}_p(s)| \leqslant \sqrt{p-1}|t-s| \leqslant \sqrt{|t-s|}$$

(since the "velocity" of the path is $(p-1)/\sqrt{p} \leqslant \sqrt{p-1}$). Consequently, for any $\alpha > 0$, we have

$$\mathbf{E}_p(|\mathsf{K}_p(t) - \mathsf{K}_p(s)|^\alpha) \leqslant |t-s|^{\alpha/2}. \tag{4.7}$$

In the remaining ranges, we will use the discontinuous partial sums $\tilde{\mathsf{K}}_p(t)$ instead of $\mathsf{K}_p(t)$. To check that this is legitimate, note that

$$|\tilde{\mathsf{K}}_p(t) - \mathsf{K}_p(t)| \leqslant \frac{1}{\sqrt{p}}$$

for all primes $p \geqslant 3$ and all t. Hence, using Hölder's inequality, we derive for $\alpha \geqslant 1$ the relation

$$\begin{aligned}\mathbf{E}_p(|\mathsf{K}_p(t) - \mathsf{K}_p(s)|^\alpha) &= \mathbf{E}_p(|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)|^\alpha) + O(p^{-\alpha/2}) \\ &= \mathbf{E}_p(|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)|^\alpha) + O(|t-s|^{\alpha/2})\end{aligned} \tag{4.8}$$

where the implied constant depends only on α.

We take $\alpha = 4$. The following computation of the fourth moment is an idea that goes back to Kloosterman's very first non-trivial estimate for individual Kloosterman sums.

We have

$$\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s) = \frac{1}{\sqrt{p}} \sum_{n \in I} e\Big(\frac{an + b\bar{n}}{p}\Big),$$

where I is the discrete interval

$$(p-1)s < n \leqslant (p-1)t$$

of summation. The length of I is

$$\lfloor (p-1)t \rfloor - \lceil (p-1)s \rceil \leqslant 2(p-1)|t-s|$$

since $(p-1)|t-s| \geqslant 1$.

By expanding the fourth power, we get

$$\begin{aligned}\mathbf{E}_p(|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)|^4) &= \frac{1}{(p-1)^2} \sum_{(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times} \Big| \frac{1}{\sqrt{p}} \sum_{(p-1)s < n \leqslant (p-1)t} e\Big(\frac{an + b\bar{n}}{p}\Big)\Big|^4 \\ &= \frac{1}{p^2(p-1)^2} \sum_{a,b} \sum_{n_1,\dots,n_4 \in I} e\Big(\frac{a(n_1+n_2-n_3-n_4)}{p}\Big) e\Big(\frac{b(\bar{n}_1+\bar{n}_2-\bar{n}_3-\bar{n}_4)}{p}\Big).\end{aligned}$$

After exchanging the order of the sums, which "separates" the two variables a and b, we get

$$\frac{1}{p^2(p-1)^2} \sum_{n_1,\dots,n_4 \in I} \Big(\sum_{a \in \mathbf{F}_p^\times} e\Big(\frac{a(n_1+n_2-n_3-n_4)}{p}\Big)\Big)\Big(\sum_{b \in \mathbf{F}_p^\times} e\Big(\frac{b(\bar{n}_1+\bar{n}_2-\bar{n}_3-\bar{n}_4)}{p}\Big)\Big).$$

The orthogonality relations for additive character (namely the relation

$$\frac{1}{p}\sum_{a\in\mathbf{F}_p^\times} e\Big(\frac{ah}{p}\Big) = \delta(h,0) - \frac{1}{p}$$

for any $h \in \mathbf{F}_p$) imply that

$$\mathbf{E}_p(|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)|^4) = \frac{1}{(p-1)^2} \sum_{\substack{n_1,\dots,n_4\in I\\ n_1+n_2=n_3+n_4\\ \bar{n}_1+\bar{n}_2=\bar{n}_3+\bar{n}_4}} 1 + O(|I|^3(p-1)^{-3}). \tag{4.9}$$

Fix first n_1 and n_2 in I with $n_1 + n_2 \neq 0$. Then if (n_3, n_4) satisfy

$$n_1 + n_2 = n_3 + n_4, \qquad \bar{n}_1 + \bar{n}_2 = \bar{n}_3 + \bar{n}_4,$$

the value of $n_3 + n_4$ is fixed, and

$$n_3 n_4 = \frac{n_3 + n_4}{\bar{n}_1 + \bar{n}_2}$$

(in $\mathbf{F}_p^\times$) is also fixed. Hence there are at most two pairs (n_3, n_4) that satisfy the equations for these given (n_1, n_2). This means that the contribution of these n_1, n_2 to (4.9) is $\leqslant 2|I|^2(p-1)^{-2}$. Similarly, if $n_1 + n_2 = 0$, the equations imply that $n_3 + n_4 = 0$, and hence the solutions are determined uniquely by (n_1, n_3). Hence the contribution is then $\leqslant |I|^2(p-1)^2$, and we get

$$\mathbf{E}_p(|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)|^4) \ll |I|^2(p-1)^{-2} + |I|^3(p-1)^{-3} \ll |t-s|^2,$$

where the implied constants are absolute. Using (4.8), this gives

$$\mathbf{E}_p(|\mathsf{K}_p(t) - \mathsf{K}_p(s)|^4) \ll |t-s|^2 \tag{4.10}$$

with an absolute implied constant. Combined with (4.7) with $\alpha = 4$ in the former range, this completes the proof of tightness, and therefore of Theorem 4.1.1.

The proof of tightness uses crucially that we average over both a and b to reduce the problem to a count of solutions of equations over $\mathbf{F}_p$ (see (4.9)). Since $S(a, b; p) = S(ab; 1, p)$ for all a and b in $\mathbf{F}_p^\times$, it seems natural to try to prove an analogue of Theorem 4.1.1 when averaging only over a, with $b = 1$ fixed. The convergence of finite distributions extends, as we saw, to that setting, but the final tightness step is currently out of reach. Using the method involved in convergence of finite distributions, and the trivial bound

$$\Big|\tilde{\mathsf{K}}_p(t) - \tilde{\mathsf{K}}_p(s)\Big| \leqslant |I|p^{-1/2},$$

one can check that it is enough to prove a suitable estimate for the average over a in the restricted range where

$$\frac{1}{2} - \eta \leqslant \gamma \leqslant \frac{1}{2} + \eta$$

for some fixed but arbitrarily small value of $\eta > 0$ (see [24, §3]). The next exercise illustrates this point.

Exercise 4.2.6. Assume p is odd. Let $\Omega'_p = \mathbf{F}_p^\times \times (\mathbf{F}_p^\times)^2$, where $(\mathbf{F}_p^\times)^2$ is the set of non-zero squares in $\mathbf{F}_p^\times$. We denote by $\mathsf{K}'_p(t)$ the random variable $\mathsf{K}_p(t)$ restricted to Ω'_p, with the uniform probability measure, for which $\mathbf{P}'_p(\cdot)$ and $\mathbf{E}'_p(\cdot)$ denote probability and expectation.

(1) Prove that $(\mathsf{K}'_p(t))$ converges to $(K(t))$ in the sense of finite distributions.

(2) For $n \in \mathbf{F}_p$, prove that

$$\sum_{b\in(\mathbf{F}_p^\times)^2} e\Big(\frac{bn}{p}\Big) = \frac{p-1}{2}\delta(n,0) + O(\sqrt{p})$$

where the implied constant is absolute. [*Hint*: show that if $n \in \mathbf{F}_p^\times$, we have

$$\Big|\sum_{b\in\mathbf{F}_p^\times} e\Big(\frac{nb^2}{p}\Big)\Big| = \sqrt{p},$$

where the left-hand sum is known as a *quadratic Gauss sum*.]

(3) Deduce that if $|t-s| \geqslant 1/p$, then

$$\mathbf{E}'_p(|\mathsf{K}'_p(t) - \mathsf{K}'_p(s)|^4) \ll \sqrt{p}|t-s|^3 + |t-s|^2$$

where the implied constant is absolute.

(3) Using notation as in the proof of tightness for K_p, prove that if $\eta > 0$, $\alpha \geqslant 1$ and

$$\frac{1}{2} + \eta \leqslant \gamma \leqslant 1,$$

then

$$\mathbf{E}'_p(|\mathsf{K}'_p(t) - \mathsf{K}'_p(s)|^\alpha) \ll |t-s|^{\alpha\eta} + |t-s|^{\alpha/2},$$

where the implied constant depends only on α.

(4) Prove that if $\eta > 0$ and

$$0 \leqslant \gamma \leqslant \frac{1}{2} - \eta,$$

then there exists $\delta > 0$ such that

$$\mathbf{E}'_p(|\mathsf{K}'_p(t) - \mathsf{K}'_p(s)|^4) \ll |t-s|^{1+\delta},$$

where the implied constant depends only on η.

(5) Conclude that (K'_p) converges in law to K in $C([0,1])$.

4.3. Application: large values

We can use Theorem 4.1.1 to gain information on partial sums of Kloosterman sums. Since this gives the occasion to illustrate some rather interesting results of probability theory in Banach spaces, we present one application.

THEOREM 4.3.1 (Kowalski–Sawin). *For p prime and $A > 0$, let $M_p(A)$ and $N_p(A)$ be the events*

$$M_p(A) = \Big\{(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times \mid \max_{1\leqslant j\leqslant p-1} \frac{1}{\sqrt{p}}\Big|\sum_{1\leqslant n\leqslant j} e\Big(\frac{an+b\bar{n}}{p}\Big)\Big| > A\Big\},$$

$$N_p(A) = \Big\{(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times \mid \max_{1\leqslant j\leqslant p-1} \frac{1}{\sqrt{p}}\Big|\sum_{1\leqslant n\leqslant j} e\Big(\frac{an+b\bar{n}}{p}\Big)\Big| \geqslant A\Big\}.$$

There exists a positive constant $c > 0$ such that, for any $A > 0$, we have

$$c^{-1}\exp(-\exp(cA)) \leqslant \liminf_{p\to+\infty} \mathbf{P}_p(N_p(A)) \leqslant \limsup_{p\to+\infty} \mathbf{P}_p(M_p(A)) \leqslant c\exp(-\exp(c^{-1}A)).$$

In particular, partial sums of normalized Kloosterman sums are unbounded (whereas the full normalized Kloosterman sums are always of modulus at most 2), but large values of partial sums are extremely rare.

PROOF. The functions $t \mapsto \mathsf{K}_p(a,b)(t)$ describe polygonal paths in the complex plane. Since the maximum modulus of a point on such a path is achieved at one of the vertices, it follows that

$$\max_{1\leqslant j\leqslant p-1} \frac{1}{\sqrt{p}}\Big|\sum_{1\leqslant n\leqslant j} e\Big(\frac{an+b\bar{n}}{p}\Big)\Big| = \|\mathsf{K}_p(a,b)\|_\infty,$$

so that the events $M_p(A)$ are the same as $\{\|\mathsf{K}_p\|_\infty > A\}$, and $N_p(A)$ is the same as $\{\|\mathsf{K}_p\|_\infty \geqslant A\}$.

By Theorem 4.1.1 and composition with the norm map (Proposition B.2.1), the real-valued random variables $\|\mathsf{K}_p\|_\infty$ converge in law to the random variable $\|K\|_\infty$, the norm of the random Fourier series K. By elementary properties of convergence in law, we have therefore

$$\mathbf{P}(\|K\|_\infty > A) \leqslant \liminf_{p\to+\infty} \mathbf{P}_p(N_p(A)) \leqslant \limsup_{p\to+\infty} \mathbf{P}_p(M_p(A)) \leqslant \mathbf{P}(\|K\|_\infty \geqslant A).$$

So the problem is reduced to questions about the limiting random Fourier series.

We begin by proving that there exists a constant $c > 0$ such that

$$\mathbf{P}(|\operatorname{Im}(K(1/2))| > A) \geqslant c^{-1}\exp(-\exp(cA)), \tag{4.11}$$

which implies that

$$\mathbf{P}(\|K\|_\infty > A) \geqslant c^{-1}\exp(-\exp(cA)).$$

(The point $1/2$ could be replaced by any $t \in]0,1[$ for the imaginary part, and one could also use the real part and any t such that $t \notin \{0, 1/2, 1\}$; the symmetry of the Kloosterman paths with respect to the line $x = \frac{1}{2}S(a,b;p)$ shows that the real part of $\mathsf{K}_p(a,b)(1/2)$ is $\frac{1}{2}S(a,b;p)$, which is a real number between -1 and 1).

We have

$$\operatorname{Im}(K(1/2)) = -\frac{1}{2\pi}\sum_{h\neq 0}\frac{\cos(\pi h)-1}{h}\mathrm{ST}_h = \frac{1}{\pi}\sum_{h\geqslant 1}\frac{1}{h}\mathrm{ST}_h.$$

Recalling that

$$\sum_{1\leqslant h\leqslant H}\frac{1}{h} \geqslant \log(H),$$

for any real number $H \geqslant 1$, we therefore derive

$$\mathbf{P}(|\operatorname{Im}(K(1/2))| > A) \geqslant \mathbf{P}\Big(\mathrm{ST}_h > 1 \text{ for } 1 \leqslant h \leqslant e^{\pi A} \text{ and } \frac{1}{\pi}\sum_{h>e^{\pi A}}\frac{1}{h}\mathrm{ST}_h \geqslant 0\Big).$$

Since the random variables $(\mathrm{ST}_h)_{h\geqslant 1}$ are independent and identically distributed, this leads to

$$\mathbf{P}(|\operatorname{Im}(K(1/2))| > A) \geqslant \mathbf{P}(\mathrm{ST}_1 > 1)^{\exp(\pi A)}\,\mathbf{P}\Big(\sum_{h>e^{\pi A}}\frac{1}{h}\mathrm{ST}_h \geqslant 0\Big).$$

Furthermore, each ST_h is symmetric, and hence so is the sum

$$\sum_{h>e^{\pi A}}\frac{1}{h}\mathrm{ST}_h,$$

which means that it has probability $\geqslant 1/2$ to be $\geqslant 0$. Therefore, writing $\mathbf{P}(\mathrm{ST}_1 > 1) = \exp(-\kappa)$ for some $\kappa > 0$,

$$\mathbf{P}(|\operatorname{Im}(K(1/2))| > A) \geqslant \frac{1}{2}\exp(-\kappa\exp(\pi A)).$$

This is of the right form asymptotically, so that (4.11) is proved.

We now consider the upper-bound. Here it suffices to prove the existence of a constant $c > 0$ such that

$$\mathbf{P}(\|\operatorname{Im}(K)\|_\infty > A) \leqslant c\exp(-\exp(c^{-1}A)), \quad \mathbf{P}(\|\operatorname{Re}(K)\|_\infty > A) \leqslant c\exp(-\exp(c^{-1}A)).$$

We will do this for the real part; the imaginary part is very similar and left as an exercise. Let $R = \operatorname{Re}(K)$. This is a random variable with values in the real Banach space $C_{\mathbf{R}}([0,1])$ of real-valued continuous functions on $[0,1]$. We write R as the random Fourier series

$$R = \sum_{h\geqslant 0}\varphi_h Y_h,$$

where $\varphi_h \in C_{\mathbf{R}}([0,1])$ and the random variables Y_h are defined by

$$\varphi_0(t) = t, \qquad Y_0 = \mathrm{ST}_0,$$
$$\varphi_h(t) = \frac{\sin(2\pi h t)}{2\pi h}, \quad Y_h = \mathrm{ST}_h + \mathrm{ST}_{-h} \text{ for } h \geqslant 1.$$

We note that the random variables (Y_h) are independent and that for all $h \geqslant 0$, we have $|Y_h| \leqslant 4$ (almost surely).

The idea that will now be implemented is the following: if we were evaluating R at a fixed $t \in]0,1[$ (not $1/2$), the series would be a subgaussian random variable, and standard estimates would give a subgaussian bound for $\mathbf{P}(|R(t)| > A)$, of the type $\exp(-cA^2)$. Such a bound would be essentially sharp for a *gaussian* series. But although it is already quite strong, it is far from the truth here, intuitively because in the gaussian case, the lower-bound for the probability arises from the probability, which is very small but non-zero, that a single summand (distributed like a gaussian) might be very large. This cannot happen for the series $R(t)$, because each Y_h is absolutely bounded.

For the actual proof, we "interpolate" between the subgaussian behavior (given by Talagrand's inequality in this case) and the boundedness of the first few steps. This principle goes back (at least) to Montgomery-Smith.

Fix an auxiliary parameter $s \geqslant 1$. We write $R = R_1 + R_2$, where

$$R_1 = \sum_{0 \leqslant h \leqslant s^2} \varphi_h Y_h, \qquad R_2 = \sum_{h > s^2} \varphi_h Y_h.$$

Let m be a median of the real random variable $\|R_2\|_\infty$. Then for any $\alpha > 0$ and $\beta > 0$, we have

$$\mathbf{P}(\|R\|_\infty \geqslant \alpha + \beta + m) \leqslant \mathbf{P}(\|R_1\|_\infty \geqslant \alpha) + \mathbf{P}(\|R_2\|_\infty \geqslant m + \beta),$$

by the triangle inequality. We pick

$$\alpha = 8 \sum_{0 \leqslant h \leqslant s^2} \|\varphi_h\|_\infty = 8 + \frac{4}{\pi} \sum_{1 \leqslant h \leqslant s^2} \frac{1}{h},$$

so that by the estimate $|Y_h| \leqslant 4$, we have

$$\mathbf{P}(\|R_1\|_\infty \geqslant \alpha) = 0.$$

Then we take $\beta = s\sigma$, where $\sigma \geqslant 0$ is such that

$$\sigma^2 = \sup_{\|\lambda\| \leqslant 1} \sum_{h > s^2} |\lambda(\varphi_h)|^2,$$

where λ runs over continuous linear functions $C_{\mathbf{R}}([0,1]) \longrightarrow \mathbf{R}$ with norm at most 1. Then Talagrand's Inequality (Theorem B.9.6) shows that

$$\mathbf{P}(\|R_2\|_\infty \geqslant m + \beta) \leqslant 4\exp(-s^2/8).$$

Hence, for all $s \geqslant 1$, we have

$$\mathbf{P}(\|R\|_\infty \geqslant \alpha + \beta + m) \leqslant 4\exp(-s^2/8).$$

We now select s as large as possible so that $m + \alpha + \beta \leqslant A$. We have

$$m \leqslant 2\,\mathbf{E}(\|R_2\|_\infty) \leqslant 2 \sum_{0 \leqslant h \leqslant s^2} \|\varphi_h\|_\infty$$

(by Chebychev's inequality, see (B.10)), so that

$$m + \alpha \ll \sum_{1 \leqslant h \leqslant s^2} \frac{1}{h} \ll \log(2s)$$

for any $s \geqslant 1$. Also, for any λ with $\|\lambda\| \leqslant 1$, we have

$$\sum_{h > s^2} |\lambda(\varphi_h)|^2 \leqslant \sum_{h > s^2} \frac{1}{4\pi^2 h^2} \ll \frac{1}{s^2}$$

so that $\sigma \ll s^{-1}$ and $\beta = s\sigma \ll 1$. It follows that $m + \alpha + \beta \leqslant c \log(cs)$ for some constant $c \geqslant 1$ and all $s \geqslant 1$. We finally select s so that $c \log(cs) = A$, i.e.

$$s = \frac{1}{c} \exp(c^{-1} A)$$

(assuming, as we may, that A is large enough so that $s \geqslant 1$) and deduce that

$$\mathbf{P}(\|R\|_\infty \geqslant A) \leqslant 4 \exp(-s^2/8) = 4 \exp\Big(-\frac{1}{8c^2} \exp\Big(\frac{A}{c}\Big)\Big).$$

This gives the desired upper bound. □

4.4. Further topics

[**Further references:** Iwaniec and Kowalski [15, Ch. 11], Kowalski and Sawin [24].]

CHAPTER 5

Further topics

APPENDIX A

Complex analysis

In Chapter 3, we use a number of facts of complex analysis which are not necessarily included in most introductory graduate courses. We therefore review them here, and give either full proofs of detailed references.

A.1. Mellin transform

The Mellin transform is a multiplicative analogue of the Fourier transform, to which it can indeed in principle be reduced. We consider it only in simple cases. Let

$$\varphi : [0, +\infty[\longrightarrow \mathbf{C}$$

be a continuous function that decays faster than any polynomial at infinity (for instance, a function with compact support). Then the Mellin transform $\hat{\varphi}$ of φ is the holomorphic function defined by the integral

$$\hat{\varphi}(s) = \int_0^{+\infty} \varphi(x) x^s \frac{dx}{x},$$

for all those $s \in \mathbf{C}$ for which the integral makes sense, which under our assumption includes all complex numbers with $\mathrm{Re}(s) > 0$.

The basic properties of the Mellin transform that are relevant for us are summarized in the next proposition:

PROPOSITION A.1.1. *Let $\varphi : [0, +\infty[\longrightarrow \mathbf{C}$ be smooth and assume that it and all its derivatives decay faster than any polynomial at infinity.*

(1) *The Mellin transform $\hat{\varphi}$ extends to a meromorphic function on $\mathrm{Re}(s) > -1$, with at most a simple pole at $s = 0$ with residue $\varphi(0)$.*

(2) *For any real numbers $-1 < A < B$, the Mellin transform has rapid decay in the strip $A \leqslant \mathrm{Re}(s) \leqslant B$, in the sense that for any integer $k \geqslant 1$, there exists a constant $C_k \geqslant 0$ such that*

$$|\hat{\varphi}(s)| \leqslant C_k (1 + |t|)^{-k}$$

for all $s = \sigma + it$ with $A \leqslant \sigma \leqslant B$ and $|t| \geqslant 1$.

(3) *For any $\sigma > 0$ and any $x \geqslant 0$, we have the Mellin inversion formula*

$$\varphi(x) = \frac{1}{2i\pi} \int_{(\sigma)} \hat{\varphi}(s) x^{-s} ds.$$

In the last formula, the notation $\int_{(\sigma)} (\cdots) ds$ refers to an integral over the vertical line $\mathrm{Re}(s) = \sigma$, oriented upward.

PROOF. (1) We integrate by parts in the definition of $\hat{\varphi}(s)$ for $\mathrm{Re}(s) > 0$, and obtain

$$\hat{\varphi}(s) = \Big[\varphi(x)\frac{x^s}{s}\Big]_0^{+\infty} - \frac{1}{s}\int_0^{+\infty} \varphi'(x) x^{s+1} \frac{dx}{x} = -\frac{1}{s}\int_0^{+\infty} \varphi'(x) x^{s+1} \frac{dx}{x}$$

since φ and φ' decay faster than any polynomial at ∞. It follows that $\psi(s) = s\hat{\varphi}(s)$ is holomorphic for $\mathrm{Re}(s) > -1$, and hence that $\hat{\varphi}(s)$ is meromorphic in this region. Since

$$\psi(0) = -\int_0^{+\infty} \varphi'(x) dx = \varphi(0),$$

it follows that there is at most a simple pole with residue $\varphi(0)$ at $s = 0$.

(2) Iterating the integration by parts $k \geqslant 2$ times, we obtain for $\mathrm{Re}(s) > -1$ the relation

$$\hat{\varphi}(s) = \frac{(-1)^k}{s(s+1)\cdots(s+k)} \int_0^{+\infty} \varphi^{(k)}(x) x^{s+k} \frac{dx}{x}.$$

Hence for $A \leqslant \sigma \leqslant B$ and $|t| \geqslant 1$ we obtain the bound

$$|\hat{\varphi}(s)| \ll \frac{1}{(1+|t|)^k} \int_0^{+\infty} |\varphi^{(k)}(x)| x^{B+k} \frac{dx}{x} \ll \frac{1}{(1+|t|)^k}.$$

(3) We interpret $\hat{\varphi}(s)$, for $s = \sigma + it$ with $\sigma > 0$ fixed, as a Fourier transform: we have

$$\hat{\varphi}(s) = \int_0^{+\infty} \varphi(x) x^{\sigma} x^{it} \frac{dx}{x} = \int_{\mathbf{R}} \varphi(e^y) e^{\sigma y} e^{iyt} dy,$$

which shows that $t \mapsto \hat{\varphi}(\sigma + it)$ is the Fourier transform (with the above normalization) of the function $g(y) = \varphi(e^y)e^{\sigma y}$. Note that g is smooth, and tends to zero very rapidly at infinity (for $y \to -\infty$, this is because φ is bounded close to 0, but $e^{\sigma y}$ then tends exponentially fast to 0). Therefore the Fourier inversion formula holds, and for any $y \in \mathbf{R}$, we obtain

$$\varphi(e^y)e^{\sigma y} = \frac{1}{2\pi} \int_{\mathbf{R}} \hat{\varphi}(\sigma + it) e^{-ity} dt.$$

Putting $x = e^y$, this translates to

$$\varphi(x) = \frac{1}{2\pi} \int_{\mathbf{R}} \hat{\varphi}(\sigma + it) x^{-\sigma - it} dt = \frac{1}{2i\pi} \int_{(\sigma)} \hat{\varphi}(s) x^{-s} ds.$$

□

A.2. Dirichlet series

We present in this section some of the basic analytic properties of Dirichlet series of the type

$$\sum_{n \geqslant 1} a_n n^{-s},$$

where $a_n \in \mathbf{C}$ for $n \geqslant 1$. For more information, see for instance [34, Ch. 9].

LEMMA A.2.1. *Let $(a_n)_{n\geqslant 1}$ be a sequence of complex numbers. Let $s_0 \in \mathbf{C}$. If the series*

$$\sum_{n \geqslant 1} a_n n^{-s_0}$$

converges, then the series

$$\sum_{n \geqslant 1} a_n n^{-s}$$

converges uniformly on compact subsets of $U = \{s \in \mathbf{C} \mid \mathrm{Re}(s) > \mathrm{Re}(s_0)\}$. In particular the function

$$D(s) = \sum_{n \geqslant 1} a_n n^{-s}$$

is holomorphic on U.

PROOF. TODO □

REMARK A.2.2. (1) In general, the convergence is *not* absolute.

(2) We see in this lemma a first instance of a fairly general principle concerning Dirichlet series: if some particular property holds for some $s_0 \in \mathbf{C}$ (or for all s_0 with some fixed real part), then it holds – or even a stronger property holds – for any s with $\mathrm{Re}(s) > \mathrm{Re}(s_0)$.

This principle also applies in many cases to the possible analytic continuation of Dirichlet series beyond the region of convergence. One example is found below in Proposition A.2.3 (concerning the size of the Dirichlet series).

PROPOSITION A.2.3. *Let $\sigma \in \mathbf{R}$ be a real number and let $(a_n)_{n\geqslant 1}$ be a bounded sequence of complex numbers such that the Dirichlet series*

$$D(s) = \sum_{n\geqslant 1} a_n n^{-s}$$

converges for $\mathrm{Re}(s) > \sigma$. Then for any $\sigma_1 > \sigma$, we have

$$|D(s)| \ll 1 + |t|$$

uniformly for $\mathrm{Re}(s) \geqslant \sigma_1$.

PROOF. This is [34, 9.33]. □

In order to express in a practical manner a Dirichlet series outside of its region of convergence, one can use *smooth partial sums*, which exploit harmonic analysis.

PROPOSITION A.2.4. *Let $\varphi : [0, +\infty[\longrightarrow [0, 1]$ be a smooth function with compact support such that $\varphi(0) = 1$. Let $\hat{\varphi}$ denote its Mellin transform. Let $\sigma > 0$ be given with $0 < \sigma_0 < 1$, and let $(a_n)_{n\geqslant 1}$ be any sequence of complex numbers with $|a_n| \leqslant 1$ such that the Dirichlet series*

$$\sum_{n\geqslant 1} a_n n^{-s}$$

extends to a holomorphic function $f(s)$ in the region $\mathrm{Re}(s) > \sigma_0$ with at most a simple pole at $s = 1$ with residue $c \in \mathbf{C}$.

For $N \geqslant 1$, define

$$f_N(s) = \sum_{n\geqslant 1} a_n \varphi\Big(\frac{n}{N}\Big) n^{-s}.$$

Let σ be a real number such that $\sigma_0 < \sigma < 1$. Then we have

$$f(s) - f_N(s) = -\frac{1}{2i\pi} \int_{(-\delta)} f(s + w) N^w \hat{\varphi}(w) dw +$$

for any $s = \sigma + it$ and any $\delta > 0$ such that $-\delta + \sigma > \sigma_0$.

It is of course possible that $c = 0$ (corresponding to a Dirichlet series that is holomorphic for $\mathrm{Re}(s) > \sigma_0$).

This result gives a convergent approximation of $f(s)$, inside the strip $\mathrm{Re}(s) > \sigma_1$, using the *finite* sums $f_N(s)$: The point is that $|N^w| = N^{-\delta}$, so that the polynomial growth of f on vertical lines combined with the fast decay of the Mellin transform show that the integral on the right tends to 0 as $N \to +\infty$. Moreover, the shape of the formula makes it very accessible to further manipulations, as done in Chapter 3.

PROOF. Fix $\alpha > 1$ such that the Dirichlet series $f(s)$ converges absolutely for $\mathrm{Re}(s) = \alpha$. By the Mellin inversion formula, followed by exchanging the order of the sum and integral, we have

$$\begin{aligned} f_N(s) &= \sum_{n\geqslant 1} a_n \varphi\Big(\frac{n}{N}\Big) \frac{1}{2i\pi} \int_{(\alpha)} N^w n^{-w} \hat{\varphi}(w) dw \\ &= \frac{1}{2i\pi} \int_{(-\delta)} \Big(\sum_{n\geqslant 1} a_n n^{-s-w}\Big) N^w \hat{\varphi}(w) dw \\ &= \frac{1}{2i\pi} \int_{(-\delta)} f(s + w) N^w \hat{\varphi}(w) dw, \end{aligned}$$

where the absolute convergence justifies the exchange of sum and integral.

Now consider some $T \geqslant 1$, and some δ such that $0 < \delta < 1$. Let $\mathcal{R}_T$ be the rectangle in $\mathbf{C}$ with sides $[\alpha - iT, \alpha + iT]$, $[\alpha + iT, -\delta + iT]$, $[-\delta + iT, -\delta - T]$, $[-\delta - iT, \alpha - iT]$, oriented counterclockwise. Inside this rectangle, the function

$$w \mapsto f(s + w) N^w \hat{\varphi}(w)$$

is meromorphic. It has a simple pole at $w = 0$, by our choice of δ and the properties of the Mellin transform of φ given by Proposition A.1.1, where the residue at $w = 0$ is $\varphi(0)f(s) = f(s)$, again by Proposition A.1.1. It may (if $c \neq 0$) also have a simple pole at $w = 1 - s$, with residue equal to $cN^{1-s}\hat{\varphi}(1 - s)$.

Cauchy's Integral theorem therefore implies that

$$f_N(s) = f(s) + \frac{1}{2i\pi}\int_{\mathcal{R}_T} f(s + w)N^w\hat{\varphi}(w)dw + cN^{1-s}\hat{\varphi}(1 - s).$$

Now we let $T \to +\infty$. Our assumptions imply that $w \mapsto f(s + w)$ has polynomial growth on the strip $-\delta \leqslant \mathrm{Re}(w) \leqslant \alpha$, and therefore the fast decay of $\hat{\varphi}$ (Proposition A.1.1 again) shows that the contribution of the two horizontal segments to the integral along $\mathcal{R}_T$ tends to 0 as $T \to +\infty$. Taking into account orientation, we get

$$f(s) - f_N(s) = -\frac{1}{2i\pi}\int_{(-\delta)} f(s + w)N^w\hat{\varphi}(w)dw - cN^{1-s}\hat{\varphi}(1 - s),$$

as claimed. □

A.3. Density of certain sets of holomorphic functions

Let D be a non-empty open disc in $\mathbf{C}$ and $\bar{D}$ its closure. We denote by $\mathcal{H}(D)$ the Banach space of all continuous functions $f : \bar{D} \longrightarrow \mathbf{C}$ which are holomorphic in D, with the norm

$$\|f\|_\infty = \sup_{z\in\bar{D}} |f(z)|.$$

We also denote by $C(K)$ the Banach space of continuous functions on a compact space K, also with the norm

$$\|f\|_\infty = \sup_{x\in K} |f(x)|$$

(so that there is no risk of confusion if $K = D$ and we apply this to a function that also belongs to $\mathcal{H}(D)$). We denote by $C(K)'$ the dual of $C(K)$, namely the space of continuous linear functionals $C(K) \longrightarrow \mathbf{C}$. An element $\mu \in C(K)'$ can also be interpreted as a *complex measure* on K (by the Riesz-Markov Theorem, see e.g. [8, Th. 7.17]), and in this interpretation one would write

$$\mu(f) = \int_K f(x)d\mu(x).$$

THEOREM A.3.1. *Let D be as above. Let $(f_n)_{n\geqslant 1}$ be a sequence of elements of $\mathcal{H}(D)$ with*

$$\sum_{n\geqslant 1} \|f_n\|_\infty^2 < +\infty.$$

Let X be the set of sequences (α_n) of complex numbers with $|\alpha_n| = 1$ such that the series

$$\sum_{n\geqslant 1} \alpha_n f_n$$

converges in $\mathcal{H}(D)$.

Assume that X is not empty and that, for any continuous linear functional $\mu \in C(\bar{D})'$ such that

$$\sum_{n\geqslant 1} |\mu(f_n)| < +\infty, \tag{A.1}$$

the Laplace transform of μ is identically 0. Then for any $N \geqslant 1$, the set of series

$$\sum_{n\geqslant N} \alpha_n f_n$$

for (α_n) in X is dense *in $\mathcal{H}(D)$.*

Here, the Laplace transform of μ is defined by

$$g(z) = \mu(w \mapsto e^{wz})$$

for $z \in \mathbf{C}$. In the interpretation of μ as a complex measure, which can be viewed as a complex measure on $\mathbf{C}$ that is supported on $\bar{D}$, one would write

$$g(z) = \int_{\mathbf{C}} e^{wz} d\mu(w).$$

PROOF. This result is proved, for instance, in [1, Lemma 5.2.9], except that only the case $N = 1$ is considered. However, if the assumptions hold for $(f_n)_{n\geqslant 1}$, they hold equally for $(f_n)_{n>N}$, hence the general case follows. □

We will use the last part of the following lemma as a criterion to establish that the Laplace transform is zero in certain circumstances.

LEMMA A.3.2. *Let K be a complex subset of $\mathbf{C}$ and $\mu \in C(K)'$ a continuous linear functional. Let*

$$g(z) = \int e^{wz} d\mu(z) = \mu(w \mapsto e^{wz})$$

be its Laplace transform.

(1) *The function g is an entire function on $\mathbf{C}$, i.e., it is holomorphic on $\mathbf{C}$.*

(2) *We have*

$$\limsup_{|z|\to+\infty} \frac{\log|g(z)|}{|z|} < +\infty.$$

(3) *If $g \neq 0$, then*

$$\limsup_{r\to+\infty} \frac{\log|g(r)|}{r} \geqslant \inf_{z\in K} \mathrm{Re}(z).$$

PROOF. (1) Let $z \in \mathbf{C}$ be fixed. For $h \neq 0$, we have

$$\frac{g(z+h) - g(z)}{h} = \mu(f_h)$$

where $f_h(w) = (e^{w(z+h)} - e^{wz})/h$. We have

$$f_h(w) \to we^{wz}$$

as $h \to 0$, and the convergence is uniform on K. Hence we get

$$\frac{g(z+h) - g(z)}{h} \longrightarrow \mu(w \mapsto we^{wz}),$$

which shows that g is holomorphic at z with derivative $\mu(w \mapsto we^{wz})$. Since z is arbitrary, this means that g is entire.

(2) We have

$$|g(z)| \leqslant \|\mu\| \ \|w \mapsto e^{wz}\|_\infty \leqslant \|\mu\| e^{|z|M}$$

where $M = \sup_{w\in K} |w|$, and therefore

$$\limsup_{|z|\to+\infty} \frac{\log|g(z)|}{|z|} \leqslant M < +\infty.$$

(3) This is proved, for instance, in [1, Lemma 5.2.2], using relatively elementary properties of entire functions satisfying growth conditions such as those in (2). □

Finally, we will use the following theorem of Bernstein, extending a result of Pólya.

THEOREM A.3.3. *Let $g : \mathbf{C} \longrightarrow \mathbf{C}$ be an entire function such that*

$$\limsup_{|z|\to+\infty} \frac{\log|g(z)|}{|z|} < +\infty.$$

Let (r_k) be a sequence of positive real numbers, and let α, β be real numbers such that
(1) *We have $\alpha\beta < \pi$;*
(2) *We have*

$$\limsup_{\substack{y\in\mathbf{R}\\|y|\to+\infty}} \frac{\log|g(iy)|}{|y|} \leqslant \alpha.$$

(3) *We have $|r_k - r_l| \gg |k-l|$ for all k, $l \geqslant 1$, and $r_k/k \to \beta$.*
Then it follows that

$$\limsup_{k\to+\infty} \frac{\log|g(r_k)|}{r_k} = \limsup_{r\to+\infty} \frac{\log|g(r)|}{r}.$$

This is explained in Lemma [1, 5.2.3].

EXAMPLE A.3.4. Taking $g(z) = \sin(\pi z)$, with $\alpha = 1$, $r_n = n\pi$ so that $\beta = \pi$, we see that the first condition is best possible.

We also use a relatively elementary lemma due to Hurwitz on zeros of holomorphic functions

LEMMA A.3.5. *Let D be a non-empty open disc in $\mathbf{C}$. Let (f_n) be a sequence of holomorphic functions in $\mathcal{H}(D)$. Assume f_n converges to f in $\mathcal{H}(D)$. If $f_n(z) \neq 0$ all $n \geqslant 1$ and $z \in K$, then either $f = 0$ or f does not vanish on D.*

PROOF. We assume that f is not zero, and show that it has no zero in D. Let $z_0 \in D$ be fixed, and let C be a circle of radius $r > 0$ centered at z_0 and such that $C \subset D$. Since f is non-zero in $\mathcal{H}(D)$, it is non-zero in the disc with boundary C, and by the maximum modulus principle, it is non-zero on C. In particular, we have $\delta = \inf_{z\in C}|f(z)| > 0$. For n large enough, we get

$$\sup_{z\in C}|f(z) - f_n(z)| < \delta,$$

and then the relation $f = f - f_n + f_n$ combined with Rouché's Theorem (see, e.g., [34, 3.42]) shows that f has the same number of zeros as f_n in the disc bounded by C. This means that f has no zeros there, and in particular that $f(z_0) \neq 0$. □

APPENDIX B

Probability

This Appendix summarizes the probabilistic notions that are most important in the notes. Although many readers will not need to be reminded of the basic definitions, they might still refer to it to check some easy probabilistic statements whose proof we have included here to avoid disrupting the arguments in the main part of the book.

B.1. Support of a measure

Let M be a topological space. If M is either second countable (i.e., there is basis of open sets that is countable) or compact, then any Radon measure μ on M has a well-defined closed *support*, denoted $\mathrm{supp}(\mu)$, which is characterized by either of the following properties: (1) it is the complement of the largest open set U, with respect to inclusion, such that $\mu(U) = 0$; or (2) it is the set of those $x \in M$ such that, for any open neighborhood U of x, we have $\mu(U) > 0$.

If X is a random variable with values in M, we will say that the *support of* X is the support of the law of X, which is a probability measure on M.

We need the following elementary property of the support of a measure:

LEMMA B.1.1. *Let M and N be topological spaces that are each either second countable or compact. Let μ be a probability measure on M, and let $f : M \longrightarrow N$ be a continuous map. The support of $f_*(\mu)$ is the closure of $f(\mathrm{supp}(\mu))$.*

PROOF. First, if $y = f(x)$ for some $x \in \mathrm{supp}(\mu)$, and if U is an open neighborhood of y, then we can find an open neighborhood $V \subset M$ of x such that $f(V) \subset U$. Then $(f_*\mu)(V) \geqslant \mu(U) > 0$. This shows that y belongs to the support of $f_*\mu$. Since the support is closed, we deduce that $\overline{f(\mathrm{supp}(\mu))} \subset \mathrm{supp}(f_*\mu)$.

For the converse, let $y \in N$ be in the support of $f_*\mu$. For any open neighborhood U of y, we have $\mu(f^{-1}(U)) = (f_*\mu)(U) > 0$. This implies that $f^{-1}(U) \cap \mathrm{supp}(\mu)$ is not empty, and since U is arbitrary, that y belongs to the closure of $f(\mathrm{supp}(\mu))$. □

Recall that a family $(X_i)_{i\in I}$ of random variables, each taking possibly values in a different metric space M_i, is *independent* if, for any finite subset $J \subset I$, the joint distribution of $(X_j)_{j\in J}$ is the measure on $\prod M_j$ which is the product measure of the laws of the X_j's.

LEMMA B.1.2. *Let $X = (X_i)_{i\in I}$ be a finite family of random variables with values in a topological space M that is compact or second countable. Viewed as a random variable taking values in M^I, we have*

$$\mathrm{supp}(X) = \prod_{i\in I} \mathrm{supp}(X_i).$$

PROOF. If $x = (x_i) \in M^I$, then an open neighborhood U of x contains a product set $\prod U_i$, where U_i is an open neighborhood of x_i in M. Then we have

$$\mathbf{P}(X \in U) \geqslant \mathbf{P}(X \in \prod_i U_i) = \prod_i \mathbf{P}(X_i \in U_i)$$

by independence. If $x_i \in \mathrm{supp}(X_i)$ for each i, then this is > 0, and hence $x \in \mathrm{supp}(X)$.

Conversely, if $x \in \mathrm{supp}(X)$, then for any $j \in I$, and any open neighborhood U of x_j, the set

$$V = \{y = (y_i)_{i\in I} \in M^I \mid y_j \in U\} \subset M^I$$

is an open neighborhood of x. Hence we have $\mathbf{P}(X \in V) > 0$, and since $\mathbf{P}(X \in V) = \mathbf{P}(X_i \in U)$, it follows that x_j is in the support of X_j. □

B.2. Convergence in law

Let M be a metric space. We view it as given with the Borel σ-algebra generated by open sets, and we denote by $C_b(M)$ the Banach space of bounded complex-valued continuous functions on M, with the norm

$$\|f\|_\infty = \sup_{x \in M} |f(x)|.$$

Given a sequence (μ_n) of probability measures on M, and a probability measure μ on M, one says that μ_n converges weakly to μ if and only if, for any bounded and continuous function $f : M \longrightarrow \mathbf{R}$, we have

$$\int_M f(x) d\mu_n(x) \longrightarrow \int_M f(x) d\mu(x). \tag{B.1}$$

If $(\Omega, \Sigma, \mathbf{P})$ is a probability space and $(X_n)_{n \geqslant 1}$ is a sequence of M-valued random variables, and if X is an M-valued random variable, then one says that (X_n) converges in law to X if and only if the measures $X_n(\mathbf{P})$ converge weakly to $X(\mathbf{P})$. If μ is a probability measure on M, then we will also say that X_n converges to μ if $X_n(\mathbf{P})$ converge weakly to μ.

The probabilistic versions of (B.1) in those cases is that

$$\mathbf{E}(f(X_n)) \longrightarrow \mathbf{E}(f(X)), \qquad \mathbf{E}(f(X_n)) \longrightarrow \int_M f d\mu \tag{B.2}$$

for all functions $f \in C_b(M)$.

In both cases, the definition immediately implies the following very useful fact, which we state in probabilistic language:

Proposition B.2.1. *Let M be a metric space. Let (X_n) be a sequence of M-valued random variables such that X_n converges in law to a random variable M. For any metric space N and any continuous function $\varphi : M \to N$, the N-valued random variables $\varphi \circ X_n$ converge in law to $\varphi \circ X$.*

Proof. For any continuous and bounded function $f : N \longrightarrow \mathbf{C}$, the composite $f \circ \varphi$ is bounded and continuous on M, and therefore convergence in law implies that

$$\mathbf{E}(f(\varphi(X_n))) \longrightarrow \mathbf{E}(f(\varphi(X))).$$

By definition, this formula, valid for all f, means that $\varphi(X_n)$ converges in law to $\varphi(X)$. □

Another property that is useful in Chapter 3 is the following:

Proposition B.2.2. *Let M be a complete separable metric space. Let (X_n) be a sequence of M-valued random variables, and μ a probability measure on M. Then X_n converges in law to μ if and only if we have*

$$\mathbf{E}(f(X_n)) \to \int_M f(x) d\mu(x)$$

for all bounded Lipschitz functions $f : M \longrightarrow \mathbf{C}$.

In other words, it is enough to prove the convergence property (B.2) for Lipschitz test functions.

Proof. A classical argument shows that convergence in law of (X_n) to μ is equivalent to

$$\mu(F) \geqslant \limsup_{n \to +\infty} \mathbf{P}(X_n \in F) \tag{B.3}$$

for all closed subsets F of M (see, e.g., [4, Th. 2.1, (iii)]).

However, the proof that convergence in law *implies* this property uses only Lipschitz test functions f (see for instance [4, (ii)⇒(iii), p. 16, and (1.1), p. 8], where it is only stated that

the relevant functions f are uniformly continuous, but this is shown by checking that they are Lipschitz). Hence the assumption that (B.2) holds for Lipschitz functions implies (B.3) for all closed subsets F, and consequently it implies convergence in law. □

A last general property is the following:

LEMMA B.2.3. *Let M be a second countable or compact topological space. Let (X_n) be a sequence of M-valued random variables, defined on some probability spaces Ω_n. Assume that (X_n) converges in law to some random variable X, and let $N \subset M$ be the support of the law of X.*

Then, for any $x \in N$ and for any open neighborhood U of x, we have

$$\liminf_{n\to+\infty} \mathbf{P}(X_n \in U) > 0,$$

and in particular there exists some $n \geqslant 1$ and some $\omega \in \Omega_n$ such that $X_n(\omega) \in U$.

PROOF. Another standard equivalent form of convergence in law is that, for any open set $U \subset M$, we have

$$\liminf_{n\to+\infty} \mathbf{P}(X_n \in U) \geqslant \mathbf{P}(X \in U)$$

(see [4, Th. 2.1, (i) and (iv)]). If $x \in N$ and U is an open neighborhood of x, then by definition we have $\mathbf{P}(X \in U) > 0$, and therefore

$$\liminf_{n\to+\infty} \mathbf{P}(X_n \in U) > 0.$$

□

We also recall an important definition that is a property of weak-compactness for a family of probability measures (or random variables).

DEFINITION B.2.4 (Tightness). Let M be a complete separable metric space. Let $(\mu_i)_{i\in I}$ be a family of probability measures on M. One says that (μ_i) is *tight* if for any $\varepsilon > 0$, there exists a compact subset $K \subset M$ such that $\mu_i(K) \geqslant 1 - \varepsilon$ for all $i \in I$.

It is a non-obvious fact that a single probability measure on a complete separable metric space is tight (see [4, Th. 1.3]).

B.3. Convergence in law in a finite-dimensional vector space

We will use two important criteria for convergence in law for random variables with values in a finite-dimensional real vector space V, which both amount to testing (B.1) for a restricted set of functions. Another important criterion applies to variables with values in a compact topological group, and is reviewed below in Section B.4.

The first result is valid in all cases, and is based on the Fourier transform. Given an integer $m \geqslant 1$ and a probability measure μ on $\mathbf{R}^m$, recall that the *characteristic function* (or *Fourier transform*) of μ is the function

$$\varphi_\mu : \mathbf{R}^m \longrightarrow \mathbf{C}$$

defined by

$$\varphi_\mu(t) = \int_{\mathbf{R}^m} e^{it\cdot x} d\mu(x),$$

where $t \cdot x = t_1x_1 + \cdots + t_mx_m$ is the standard inner-product. This is a continuous bounded function on $\mathbf{R}^m$. For a random vector X with values in $\mathbf{R}^m$, we denote by φ_X the characteristic function of $X(\mathbf{P})$, namely

$$\varphi_X(t) = \mathbf{E}(e^{it\cdot X}).$$

We state two (obviously equivalent) versions of P. Lévy's theorem for convenience:

THEOREM B.3.1 (Lévy). *Let $m \geqslant 1$ be an integer.*

(1) *Let (μ_n) be a sequence of probability measures on $\mathbf{R}^m$, and let μ be a probability measure on $\mathbf{R}^m$. Then (μ_n) converges weakly to μ if and only if, for any $t \in \mathbf{R}^m$, we have*

$$\varphi_{\mu_n}(t) \longrightarrow \varphi_\mu(t)$$

as $n \to +\infty$.

(2) *Let $(\Omega, \Sigma, \mathbf{P})$ be a probability space. Let $(X_n)_{n\geqslant 1}$ be $\mathbf{R}^m$-valued random vectors on Ω, and let X be an $\mathbf{R}^m$-valued random vector. Then (X_n) converges in law to X if and only if, for all $t \in \mathbf{R}^m$, we have*

$$\mathbf{E}(e^{it\cdot X_n}) \longrightarrow \mathbf{E}(e^{it\cdot X}).$$

REMARK B.3.2. In fact, the precise version of Lévy's Theorem does not require to know in advance the limit of the sequence: if a sequence (μ_n) of probability measures is such that, for all $t \in \mathbf{R}^m$, we have

$$\varphi_{\mu_n}(t) \longrightarrow \varphi(t)$$

for *some* function φ, then one can show that φ is the characteristic function of a probability measure μ (and hence that μ_n converges weakly to μ). So, for instance, it is not necessary to know beforehand that $\varphi(t) = e^{-t^2/2}$ is the characteristic function of a probability measure in order to prove the Central Limit Theorem using Lévy's Criterion.

LEMMA B.3.3. *Let $m \geqslant 1$ be an integer. Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables with values in $\mathbf{R}^m$ on some probability space. Let (β_n) be sequences of positive real numbers such that $\beta_n \to 0$ as $n \to +\infty$. If (X_n) converges in law to an $\mathbf{R}^m$-valued random variable X, then for any sequence (Y_n) of $\mathbf{R}^m$-valued random variables such that $\|X_n - Y_n\|_\infty \leqslant \beta_n$ for all $n \geqslant 1$, the random variables Y_n converge to X.*

PROOF. We use Lévy's criterion. We fix $t \in \mathbf{R}^m$ and write

$$\mathbf{E}(e^{it\cdot Y_n}) - \mathbf{E}(e^{it\cdot X}) = \mathbf{E}(e^{it\cdot Y_n} - e^{it\cdot X_n}) + \mathbf{E}(e^{it\cdot X_n} - e^{it\cdot X}).$$

By Lévy's Theorem and our assumption on the convergence of the sequence(X_n), the second term on the right converges to 0 as $n \to +\infty$. For the first, we can simply apply the dominated convergence theorem to derive the same conclusion: we have

$$\|(X_n - Y_n)\|_\infty \leqslant \beta_n \to 0$$

hence

$$e^{it\cdot Y_n} - e^{it\cdot X_n} = e^{it\cdot Y_n}\Big(1 - e^{it\cdot(X_n - Y_n)}\Big) \to 0$$

(pointwise) as $n \to +\infty$. Moreover, we have

$$\Big|e^{it\cdot Y_n} - e^{it\cdot X_n}\Big| \leqslant 2$$

for all $n \geqslant 1$. Hence the dominated convergence theorem implies that the expectation $\mathbf{E}(e^{it\cdot Y_n} - e^{it\cdot X_n})$ converges to 0.

Lévy's Theorem applied once more allows us to conclude that (Y_n) converges in law to X, as claimed. □

The second convergence criterion is known as the *method of moments*. It is more restricted than Lévy's criterion, but is sometimes analytically more flexible.

DEFINITION B.3.4. Let μ be a probability measure on $\mathbf{R}^m$. We say that μ is *mild* if the moments

$$M_{\boldsymbol{k}}(\mu) = \int_{\mathbf{R}^m} |x_1|^{k_1} \cdots |x_m|^{k_m} d\mu(x_1, \ldots, x_m)$$

exist for all tuples of non-negative integers $\boldsymbol{k} = (k_1, \ldots, k_m)$, and if there exists $\delta > 0$ such that the power series

$$\sum_{k_i \geqslant 0} \sum M_{\boldsymbol{k}}(\mu) \frac{z_1^{k_1} \cdots z_m^{k_m}}{k_1! \cdots k_m!}$$

converges in the region

$$\{(z_1, \ldots, z_m) \in \mathbf{C}^m \mid |z_i| \leqslant \delta\}.$$

If X is a random variable, we will say as usual that a random vector $X = (X_1, \ldots, X_m)$ is mild if its law $X(\mathbf{P})$ is mild. The moments are then

$$M_{\boldsymbol{k}}(X) = \mathbf{E}(|X_1|^{k_1} \cdots |X_m|^{k_m}).$$

We again give two versions of the method of moments for weak convergence when the limit is mild:

THEOREM B.3.5 (Method of moments). *Let $m \geqslant 1$ be an integer.*

(1) *Let (μ_n) be a sequence of probability measures on $\mathbf{R}^m$ such that all moments $M_{\boldsymbol{k}}(\mu_n)$ exist, and let μ be a probability measure on $\mathbf{R}^m$.* Assume that μ is mild. *Then (μ_n) converges weakly to μ if for any m-tuple $\boldsymbol{k}$ of non-negative integers, we have*

$$M_{\boldsymbol{k}}(\mu_n) \longrightarrow M_{\boldsymbol{k}}(\mu)$$

as $n \to +\infty$.

(2) *Let $(\Omega, \Sigma, \mathbf{P})$ be a probability space. Let $(X_n)_{n\geqslant 1}$ be $\mathbf{R}^m$-valued random vectors on Ω such that all moments $M_{\boldsymbol{k}}(X)$ exist, and let Y be an $\mathbf{R}^m$-valued random vector.* Assume that Y is mild. *Then (X_n) converges in law to Y if for any m-tuple $\boldsymbol{k}$ of non-negative integers, we have*

$$\mathbf{E}(|X_{n,1}|^{k_1} \cdots |X_{n,m}|^{k_m}) \longrightarrow \mathbf{E}(|Y_1|^{k_1} \cdots |Y_n|^{k_m}).$$

For a proof (in the case $m = 1$), see for instance [3, Th. 30.1].

This only gives one implication in comparison with the Lévy Criterion. It is often useful to have a converse, which is however more restricted. We only state one version:

THEOREM B.3.6 (Converse to the method of moments). (1) *Let $m \geqslant 1$ be an integer. Let $(\Omega, \Sigma, \mathbf{P})$ be a probability space. Let $(X_n)_{n\geqslant 1}$ be $\mathbf{R}^m$-valued random vectors on Ω such that all moments $M_{\boldsymbol{k}}(X)$ exist, and such that there exist constants $c_{\boldsymbol{k}} \geqslant 0$ with*

$$\mathbf{E}(|X_{n,1}|^{k_1} \cdots |X_{n,m}|^{k_m}) \leqslant c_{\boldsymbol{k}}$$

for all $n \geqslant 1$. Assume that X_n converges in law to a random vector Y. Then Y is mild and for any m-tuple $\boldsymbol{k}$ of non-negative integers, we have

$$\mathbf{E}(|X_{n,1}|^{k_1} \cdots |X_{n,m}|^{k_m}) \longrightarrow \mathbf{E}(|Y_1|^{k_1} \cdots |Y_n|^{k_m}).$$

(2) *In particular, this applies for $m = 1$ if X_n is given by*

$$X_n = \frac{B_1 + \cdots + B_n}{\sigma_n}$$

where the variables (B_n) are independent and satisfy

$$\mathbf{E}(B_n) = 0, \qquad |B_n| \leqslant 1, \qquad \sigma_n^2 = \sum_{i=1}^{n} \mathbf{V}(B_n) \longrightarrow +\infty.$$

PROOF. See [3, Th 25.12 and Cor.] for a proof (again for $m = 1$). The fact that the final example satisfies the uniform integrability follows from [3, p. 391], or from the following argument: for any $k \geqslant 0$, there exists a constant $C_k \geqslant 0$ such that

$$|x|^k \leqslant C_k(e^x + e^{-x})$$

for all $x \in \mathbf{R}$. In particular, if we can show that there exists $D \geqslant 0$ such that

$$\mathbf{E}(e^{X_n}) \leqslant D, \qquad \mathbf{E}(e^{-X_n}) \leqslant D \tag{B.4}$$

for all $n \geqslant 1$, then we obtain $\mathbf{E}(|X_n|^k) \leqslant 2C_kD$ for all n, which gives the desired conclusion.

To prove (B.4), fix more generally $t \in [-1,1]$. For n large enough (which we may assume), we have $\sigma_n \geqslant 1 \geqslant |t|$. We then have, by independence, the formula

$$\mathbf{E}(e^{tX_n}) = \prod_{i=1}^{m} \mathbf{E}\Big(\exp\Big(\frac{tB_i}{\sigma_n}\Big)\Big).$$

Since $|tB_i/\sigma_n| \leqslant 1$ from our assumptions, we have

$$\exp\Big(\frac{tB_i}{\sigma_n}\Big) \leqslant 1 + \frac{tB_i}{\sigma_n} + \frac{t^2 B_i^2}{\sigma_n^2}$$

(because $e^x \leqslant 1 + x + x^2$ for $|x| \leqslant 1$), and hence

$$\mathbf{E}(e^{tX_n}) \leqslant \prod_{i=1}^{n} \Big(1 + \frac{t^2}{\sigma_n^2}\,\mathbf{E}(B_i^2)\Big)$$

since $\mathbf{E}(B_i) = 0$. Using $1 + x \leqslant e^x$, this leads to

$$\mathbf{E}(e^{tX_n}) \leqslant \exp\Big(\frac{t^2}{\sigma_n^2} \sum_{i=1}^{m} \mathbf{E}(B_i^2)\Big) = \exp(t^2).$$

Applying this with $t = 1$ and $t = -1$, we get (B.4) with $D = e$. □

Remark B.3.7. In the case $m = 2$, one often deals with random variables that are naturally seen as complex-valued, instead of $\mathbf{R}^2$-valued. In that case, it is sometimes quite useful to use the complex moments

$$\tilde{M}_{k_1,k_2}(X) = \mathbf{E}(X^{k_1}\bar{X}^{k_2})$$

of a $\mathbf{C}$-valued random variable instead of $M_{k_1,k_2}(X)$. The corresponding statements are that X is mild if and only if the power series

$$\sum_{k_1,k_2 \geqslant 0} \sum \tilde{M}_{k_1,k_2}(X) \frac{z_1^{k_1} z_2^{k_2}}{k_1! k_2!}$$

converges in a region

$$\{(z_1, z_2) \in \mathbf{C} \mid |z_1| \leqslant \delta, \quad |z_2| \leqslant \delta\}$$

for some $\delta > 0$, and that if X is mild, then (X_n) converges weakly to X if and only if

$$\tilde{M}_{k_1,k_2}(X_n) \longrightarrow \tilde{M}_{k_1,k_2}(X)$$

for all k_1, $k_2 \geqslant 0$.

Example B.3.8. (1) Any bounded random vector is mild. Indeed, if $\|X\|_\infty \leqslant B$, say, then we get

$$|M_{\boldsymbol{k}}(X)| \leqslant B^{k_1+\cdots+k_m},$$

and therefore

$$\sum_{k_i \geqslant 0} \sum |M_{\boldsymbol{k}}(\mu)| \frac{|z_1|^{k_1} \cdots |z_m|^{k_m}}{k_1! \cdots k_m!} \leqslant e^{B|z_1|+\cdots+B|z_m|},$$

so that the power series converges, in that case, for all $\boldsymbol{z} \in \mathbf{C}^m$.

(2) Any gaussian random vector is mild (see the next section).

(3) If X is mild, and Y is another random vector with $|Y_i| \leqslant |X_i|$ (almost surely) for all i, then Y is also mild.

We refer to Billingsley's book [4] for further results on convergence in law.

B.4. The Weyl criterion

One important special case of convergence in law is known as *equidistribution* in the context of topological groups in particular. We only consider compact groups here for simplicity. Let G be such a group. Then there exists on G a unique Borel probability measure μ_G which is invariant under left (and right) translations: for any integrable function $f : G \longrightarrow \mathbf{C}$ and for any fixed $g \in G$, we have

$$\int_G f(gx)d\mu_G(x) = \int_G f(xg)d\mu_G(x) = \int_G f(x)d\mu_G(x).$$

If a G-valued random variable X is distributed according to μ_G, one says that X is *uniformly distributed* on G.

EXAMPLE B.4.1. (1) If $G = \mathbf{S}^1$ is the multiplicative group of complex numbers of modulus 1, then the measure μ_G is the Lebesgue measure $d\theta/(2\pi)$ under the identification $\mathbf{R}/2\pi\mathbf{Z} \longrightarrow \mathbf{S}^1$ given by $t \mapsto e^{it}$.

(2) If $(G_i)_{i\in I}$ is any family of compact groups, each with a probability Haar measure μ_i, then the (possibly infinite) tensor product

$$\bigotimes_{i\in I} \mu_i$$

is the probability Haar measure μ on the product G of the groups G_i. Probabilistically, one would interpret this as saying that μ is the law of a family (X_i) of *independent* random variables, where each X_i is uniformly distributed on G_i.

(3) Let G be the non-abelian compact group $\mathrm{SU}_2(\mathbf{C})$, i.e.

$$G = \left\{ \begin{pmatrix} \alpha & \bar{\beta} \\ -\beta & \bar{\alpha} \end{pmatrix} \mid \alpha,\ \beta \in \mathbf{C},\ |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

Writing $\alpha = a + ib$, $\beta = c + id$, we can identify G, as a topological space, with the unit 3-sphere

$$\{(a, b, c, d) \in \mathbf{R}^4 \mid a^2 + b^2 + c^2 + d^2 = 1$$

in $\mathbf{R}^4$. Then the left-multiplication by some element on G is the restriction of a rotation of $\mathbf{R}^4$. Hence the surface (Lebesgue) measure μ_0 on the 3-sphere is a Borel invariant measure on G. By uniqueness, we see that the probability Haar measure on G is

$$\mu = \frac{1}{2\pi^2}\mu_0$$

(since the surface area of the 3-sphere is $2\pi^2$).

Consider now the trace $\mathrm{Tr} : G \longrightarrow \mathbf{R}$, which is given by $(a, b, c, d) \mapsto 2a$ in the sphere coordinates. One can show that the direct image $\mathrm{Tr}_*(\mu)$ is the so-called *Sato-Tate* measure

$$\mu_{ST} = \frac{1}{\pi}\sqrt{1 - \frac{x^2}{4}}dx,$$

supported on $[-2, 2]$. One obtains from either description of μ_{ST} the expectation and variance

$$\int_{\mathbf{R}} td\mu_{ST} = 0, \qquad \int_{\mathbf{R}} t^2 d\mu_{ST} = 1. \tag{B.5}$$

For a topological group G, a *unitary character* χ of G is a continuous homomorphism

$$\chi : G \longrightarrow \mathbf{S}^1.$$

The *trivial character* is the character $g \mapsto 1$ of G. The set of all characters of G is denoted $\hat{G}$.

It is a basic fact (known as *Pontryagin duality*, see e.g. [21, §7.3] for a survey) that there are many characters if G is a locally compact abelian group, and indeed that (for instance) the characters of a compact abelian group form an orthonormal basis of the space $L^2(G, \mu_G)$.

For an integrable function $f \in L^1(G,\mu)$, its *Fourier transform* is the function $\hat{f} : \hat{G} \longrightarrow \mathbf{C}$ defined by

$$\hat{f}(\chi) = \int_G f(x)\overline{\chi(x)}d\mu(x)$$

for all $\chi \in \hat{G}$. For a compact group G, and $f \in L^2(G,\mu)$, we have

$$f = \sum_{\chi\in\hat{G}} \hat{f}\chi,$$

as a series converging in $L^2(G,\mu)$. It follows easily that $f \in L^1(G)$ is almost everywhere constant if and only if $\hat{f}(\chi) = 0$ for all $\chi \neq 0$.

The following relation is immediate from the invariance of Haar measure: for f integrable and any fixed $y \in G$, if we let $g(x) = f(xy)$, then

$$\hat{g}(\chi) = \int_G f(xy)\overline{\chi(x)}d\mu(x) = \chi(y)\int_G f(x)\overline{\chi(x)}d\mu(x) = \chi(y)\hat{f}(y),$$

so that

$$\hat{g} = \chi\hat{f}. \tag{B.6}$$

Example B.4.2. (1) The characters of $\mathbf{S}^1$ are given by

$$z \mapsto z^m$$

for $m \in \mathbf{Z}$.

(2) If $(G_i)_{i\in I}$ is any family of compact groups, each with a probability Haar measure μ_i, then the characters of the product G of the G_i are given in a unique way as follows: take a finite subset S of I, and for any $i \in I$, pick a non-trivial character χ_i of G_i, then define

$$\chi(x) = \prod_{i\in S} \chi_i(x_i)$$

for any $x = (x_i)_{i\in I}$ in G. Here, the trivial character corresponds to $S = \emptyset$. See, e.g., [21, Example 5.6.10] for a proof.

Weyl's Criterion is a criterion for a sequence of G-valued random variables to converge in law to a uniformly distributed random variable. We state it for compact abelian groups only:

Theorem B.4.3 (Weyl's Criterion). *Let G be a compact topological group. A sequence (X_n) of G-valued random variables converges in law to a uniformly distributed random variable on G if and only if, for any non-trivial character χ of G, we have*

$$\lim_{n\to+\infty} \mathbf{E}(\chi(X_n)) \longrightarrow 0.$$

Remark B.4.4. Note that the orthogonality of characters implies that

$$\int_G \chi(x)d\mu_G(x) = \langle\chi, 1\rangle = 0$$

for any non-trivial character χ of G. Hence the Weyl criterion has the same flavor of Lévy's criterion (note that, for any $t \in \mathbf{R}^m$, the function $x \mapsto e^{ix\cdot t}$ is a character of $\mathbf{R}^m$).

Exercise B.4.5. The following is a classical example of application of the Weyl Criterion, known as Kronecker's Theorem.

Let $d \geqslant 1$ be an integer and let $\xi = (\xi_1, \ldots, \xi_d) \in (\mathbf{R}/\mathbf{Z})^d$ be given. Let T be the closure of the set $\{n\xi \mid n \in \mathbf{Z}\} \subset (\mathbf{R}/\mathbf{Z})^d$.

(1) Prove that T is a closed subgroup of $(\mathbf{R}/\mathbf{Z})^d$.

(2) Prove that the probability measures

$$\frac{1}{N}\sum_{1\leqslant n\leqslant N} \delta_{n\xi}$$

on $(\mathbf{R}/\mathbf{Z})^d$ converge to the (unique) probability Haar measure on T.

B.5. Gaussian random variables

By definition, a random vector X with values in $\mathbf{R}^m$ is called a *(centered) gaussian vector* if there exists a non-negative quadratic form Q on $\mathbf{R}^m$ such that the characteristic function φ_X of X is of the form

$$\varphi_X(t) = e^{-Q(t)/2}$$

for $t \in \mathbf{R}^m$. The quadratic form can be recovered from X by the relation

$$Q(t_1, \ldots, t_m) = \sum_{1 \leqslant i,j \leqslant m} a_{i,j} t_i t_j,$$

with

$$a_{i,j} = \mathbf{E}(X_i X_j).$$

More generally, if X is a gaussian random vector, then X is mild, and in fact

$$\sum_{\boldsymbol{k}} M_{\boldsymbol{m}}(X) \frac{t_1^{k_1} \cdots t_m^{k_m}}{k_1! \cdots k_m!} = \mathbf{E}(e^{t \cdot X}) = e^{Q(t)/2}$$

for $t \in \mathbf{R}^m$, so that the power series converges on all of $\mathbf{C}^m$.

For $m = 1$, this means that a random variable is a centered gaussian if and only if there exists $\sigma \geqslant 0$ such that

$$\varphi_X(t) = e^{-\sigma^2 t/2},$$

and in fact we have

$$\mathbf{E}(X^2) = \mathbf{V}(X) = \sigma^2.$$

If $\sigma = 1$, then we say that X is a *standard gaussian random variable*, or a *standard normal random variable*. We then have

$$\mathbf{P}(a < X < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

for all real numbers $a < b$.

We will use the following simple version of the Central Limit Theorem:

THEOREM B.5.1. *Let $B \geqslant 0$ be a fixed real number. Let (X_n) be a sequence of independent real-valued random variables with $|X_n| \leqslant B$ for all n. Let*

$$\alpha_n = \mathbf{E}(X_n), \qquad \beta_n = \mathbf{V}(X_n^2).$$

Let $\sigma_N \geqslant 0$ be defined by

$$\sigma_N^2 = \beta_1 + \cdots + \beta_N$$

for $N \geqslant 1$. If $\sigma_N \to +\infty$ as $n \to +\infty$, then the random variables

$$Y_N = \frac{(X_1 - \alpha_1) + \cdots + (X_N - \alpha_N)}{\sigma_N}$$

converge in law to a standard gaussian random variable.

PROOF. Although this is a very simple case of the general Central Limit Theorem for sums of independent random variables (indeed, even of Lyapunov's well-known version), we give a proof using Lévy's criterion for convenience. First of all, we may assume that $\alpha_n = 0$ for all n by replacing X_n by $X_n - \alpha_n$ (up to replacing B by $2B$, since $|\alpha_n| \leqslant B$).

By independence of the variables (X_n), the characteristic function φ_N of Y_N is given by

$$\varphi_N(t) = \mathbf{E}(e^{itY_N}) = \prod_{1 \leqslant n \leqslant N} \mathbf{E}(e^{itX_n/\sigma_N})$$

for $t \in \mathbf{R}$. We then have

$$\mathbf{E}(e^{itX_n/\sigma_N}) = \varphi_{X_n}\Big(\frac{t}{\sigma_N}\Big) = 1 - \frac{1}{2}\Big(\frac{t}{\sigma_N}\Big)^2 \mathbf{E}(X_n^2) + O\Big(\Big(\frac{|t|}{\sigma_N}\Big)^3 \mathbf{E}(|X_n|^3)\Big).$$

Observe that with our assumption, we have

$$\mathbf{E}(|X_n|^3) \leqslant B\,\mathbf{E}(X_n^2) = B\beta_n$$

and hence a simple computation shows that

$$\begin{aligned}\varphi_N(t) &= \exp\Big(\sum_{n=1}^{N} \log \mathbf{E}(e^{itX_n/\sigma_N})\Big)\\ &= \exp\Big(-\frac{t^2}{2\sigma_N}\sum_{n=1}^{N}\beta_n + O\Big(\frac{B|t|^3}{\sigma_N^3}\sum_{n=1}^{N}\beta_n\Big)\Big)\\ &= \exp\Big(-\frac{t^2}{2} + O\Big(\frac{B|t|^3}{\sigma_N}\Big)\Big) \longrightarrow \exp(-t^2/2)\end{aligned}$$

as $N \to +\infty$. □

If one uses directly the method of moments to get convergence in law to a gaussian random variable, it is useful to know the values of their moments. We only state the one-dimensional and the simplest complex case:

Proposition B.5.2. (1) *Let X be a real-valued gaussian random variable with expectation 0 and variance σ^2. For $k \geqslant 0$, we have*

$$\mathbf{E}(X^k) = \begin{cases} 0 & \text{if } k \text{ is odd,}\\ \sigma^k \frac{k!}{2^{k/2}(k/2)!} = 1\cdot 3\cdot\cdots\cdot(k-1) & \text{if } k \text{ is even.}\end{cases}$$

(1) *Let X be a complex-valued gaussian random variable with covariance matrix*

$$\begin{pmatrix}\sigma & 0\\ 0 & \sigma\end{pmatrix}$$

for some $\sigma > 0$. For $k \geqslant 0$ and $l \geqslant 0$, we have

$$\mathbf{E}(X^k \bar{X}^l) = \begin{cases} 0 & \text{if } k \neq l,\\ \sigma^k 2^k k! & \text{if } k = l.\end{cases}$$

Exercise B.5.3. Prove this proposition.

B.6. Subgaussian random variables

Gaussian random variables have many remarkable properties. It is a striking fact that a number of these, especially with respect to integrability properties, are shared by a much more general class of random variables.

Definition B.6.1 (Subgaussian random variable). Let $\sigma > 0$ be a real number. A real-valued random variable X is σ^2-subgaussian if we have

$$\mathbf{E}(e^{tX}) \leqslant e^{-\sigma^2 t^2/2}$$

for all $t \in \mathbf{R}$. A complex-valued random variable X is σ^2-subgaussian if $X = Y + iZ$ with Y and Z real-valued σ^2-subgaussian random variables.

By definition, a gaussian random variable is therefore subgaussian. But there are many more examples, in particular the random variables described in the next proposition.

Proposition B.6.2. (1) *Let X be a complex-valued random variable and $m > 0$ a real number such that $\mathbf{E}(X) = 0$ and $|X| \leqslant m$, for some real number $m \geqslant 0$. Then X is m^2-subgaussian.*

(2) *Let X_1 and X_2 be independent random variables such that X_i is σ_i^2-subgaussian. Then $X_1 + X_2$ is $(\sigma_1^2 + \sigma_2^2)$-subgaussian.*

PROOF. (1) We may assume that X is real-valued, and by considering $m^{-1}X$ instead of X, we may assume that $|X| \leqslant 1$, and of course that X is not almost surely 0. In particular, the function $\varphi(t) = \mathbf{E}(e^{tX})$ is well-defined, and $\varphi(t) > 0$ for all $t \in \mathbf{R}$. Moreover, it is smooth on $\mathbf{R}$ with

$$\varphi'(t) = \mathbf{E}(Xe^{tX}), \qquad \varphi''(t) = \mathbf{E}(X^2e^{tX}),$$

and in particular

$$\varphi(0) = 1, \qquad \varphi'(0) = \mathbf{E}(X) = 0.$$

We now define $f(t) = \log(\varphi(t)) - \frac{1}{2}t^2$. The function f is also smooth and satisfies $f(0) = f'(0) = 0$. Moreover we have

$$f''(t) = \frac{\varphi''(t)\varphi(t) - \varphi'(t)^2 - \varphi(t)^2}{\varphi(t)^2}.$$

The formula for φ'' and the condition $|X| \leqslant 1$ imply that $0 \leqslant \varphi''(t) \leqslant \varphi(t)$ for all $t \in \mathbf{R}$. Therefore

$$\varphi''(t)\varphi(t) - \varphi'(t)^2 - \varphi(t)^2 \leqslant -\varphi'(t)^2 \leqslant 0,$$

and $f''(t) \leqslant 0$ for all $t \in \mathbf{R}$. Hence the derivative of f is decreasing, which means that $f'(t)$ is $\leqslant 0$ for $t \geqslant 0$, and $\geqslant 0$ for $t \leqslant 0$. Therefore f is increasing for $t \leqslant 0$ and decreasing for $t \geqslant 0$. It follows that $f(t) \leqslant f(0) = 0$ for all $t \in \mathbf{R}$, which means exactly that $\mathbf{E}(e^{tX}) \leqslant e^{t^2/2}$.

(2) Since X_1 and X_2 are independent and subgaussian, we have

$$\mathbf{E}(e^{t(X_1+X_2)}) = \mathbf{E}(e^{tX_1})\,\mathbf{E}(e^{tX_2}) \leqslant \exp(\tfrac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2)$$

for any $t \in \mathbf{R}$. □

PROPOSITION B.6.3. *Let $\sigma > 0$ be a real number and let X be a σ^2-subgaussian random variable, either real or complex-valued. For any integer $k \geqslant 0$, there exists $c_k \geqslant 0$ such that*

$$\mathbf{E}(|X|^k) \leqslant c_k\sigma^k.$$

PROOF. The random variable $Y = \sigma^{-1}X$ is 1-subgaussian. As in the proof of Theorem B.3.6 (2), we observe that there exists $c_k \geqslant 0$ such that

$$|Y|^k \leqslant c_k(e^{X_k} + e^{-X_k}),$$

and therefore

$$\sigma^{-k}\,\mathbf{E}(|X|^k) = \mathbf{E}(|Y|^k) \leqslant c_k(e^{1/2} + e^{-1/2}),$$

which gives the result. □

REMARK B.6.4. A more precise argument leads to specific values of c_k. For instance, if X is real-valued, one can show that the inequality holds with $c_k = k2^{k/2}\Gamma(k/2)$.

B.7. Poisson random variables

Let $\lambda > 0$ be a real number. A random variable X is said to have a Poisson distribution with parameter $\lambda \in [0, +\infty[$ if and only if it is integral-valued, and if for any integer $k \geqslant 0$, we have

$$\mathbf{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

One checks immediately that

$$\mathbf{E}(X) = \lambda, \qquad \mathbf{V}(X) = \lambda,$$

and that the characteristic function of X is

$$\varphi_X(t) = e^{-\lambda}\sum_{k\geqslant 0} e^{ikt}\frac{\lambda^k}{k!} = \exp(\lambda(e^{it} - 1)).$$

PROPOSITION B.7.1. *Let (λ_n) be a sequence of real numbers such that $\lambda_n \to +\infty$ as $n \to +\infty$. Then*

$$\frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

converges in law to a standard normal random variable.

PROOF. Use the Lévy Criterion: the characteristic function φ_n of X_n is given by

$$\varphi_n(t) = \mathbf{E}(e^{it(X_n-\lambda_n)/\sqrt{\lambda_n}}) = \exp\Big(-it\sqrt{\lambda_n} + \lambda_n(e^{it/\sqrt{\lambda_n}} - 1)\Big)$$

for $t \in \mathbf{R}$, by the formula for the characteristic function of X_n. Since

$$-\frac{it}{\sqrt{\lambda_n}} + \lambda_n(e^{it/\sqrt{\lambda_n}} - 1) = it\sqrt{\lambda_n} + \lambda_n\Big(\frac{it}{\sqrt{\lambda_n}} - \frac{t^2}{2\lambda_n} + O\Big(\frac{|t|^3}{\lambda_n^{3/2}}\Big)\Big) = -\frac{t^2}{2} + O\Big(\frac{|t|^3}{\lambda_n^{1/2}}\Big),$$

we obtain $\varphi_n(t) \to \exp(-t^2/2)$, which is the characteristic function of a standard normal random variable. □

B.8. Random series

We will need some fairly elementary results on certain random series, especially concerning almost sure convergence. We first have a well-known criterion of Kolmogorov for convergence in the case of independent summands:

THEOREM B.8.1 (Kolmogorov). *Let (X_n) be a sequence of independent complex-valued random variables such that $\mathbf{E}(X_n) = 0$ and*

$$\sum_{n\geqslant 1} \mathbf{V}(X_n) < +\infty.$$

Then the series

$$\sum_{n\geqslant 1} X_n$$

converges almost surely, and hence also in law.

PROOF. We will show that the sequence of partial sums

$$S_N = \sum_{1\leqslant n\leqslant N} X_n$$

is almost surely a Cauchy sequence. For this purpose, denote

$$Y_{N,M} = \sup_{1\leqslant k\leqslant M} |S_{N+k} - S_N|$$

for N, $M \geqslant 1$. For fixed N, $Y_{N,M}$ is an increasing sequence of random variables; we denote by $Y_N = \sup_{k\geqslant 1} |S_{N+k} - S_N|$ its limit. Because of the estimate

$$|S_{N+k} - S_{N+l}| \leqslant |S_{N+k} - S_N| + |S_{N+l} - S_N| \leqslant 2Y_N$$

for $N \geqslant 1$ and k, $l \geqslant 1$, we have

$$\{(S_N)_{N\geqslant 1} \text{ is not Cauchy}\} = \bigcup_{k\geqslant 1}\bigcap_{N\geqslant 1}\bigcup_{k\geqslant 1}\bigcup_{l\geqslant 1}\{|S_{N+k} - S_{N+l}| > 2^{-k}\} \subset \bigcup_{k\geqslant 1}\bigcap_{N\geqslant 1}\{Y_N > 2^{-k-1}\}.$$

It is therefore sufficient to prove that

$$\mathbf{P}(\bigcap_{N\geqslant 1}\{Y_N > 2^{-k-1}\}) = 0$$

for each $k \geqslant 1$, or what amounts to the same thing, to prove that for any $\varepsilon > 0$, we have

$$\lim_{N\to+\infty} \mathbf{P}(Y_N > \varepsilon) = 0.$$

We begin by estimating $\mathbf{P}(Y_{N,M} > \varepsilon)$. *If* $Y_{N,M}$ was defined as $S_{N+M} - S_N$ (without the sup over $k \leqslant M$) this would be easy using the Markov inequality. To handle it, we use Kolmogorov's

Maximal Inequality (see Lemma B.8.3 below): since the $(X_n)_{N+1\leqslant n\leqslant N+M}$ are independent, it shows that for any $\varepsilon > 0$, we have

$$\mathbf{P}(Y_{N,M} > \varepsilon) = \mathbf{P}\Big(\sup_{k\leqslant M}\Big|\sum_{1\leqslant n\leqslant k} X_{N+n}\Big| > \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\sum_{n=N+1}^{N+M}\mathbf{V}(X_n).$$

Letting $M \to +\infty$, we obtain

$$\mathbf{P}(Y_N > \varepsilon) \leqslant \frac{1}{\varepsilon^2}\sum_{n\geqslant N+1}\mathbf{V}(X_n).$$

From the assumption on the convergence of the series of variance, this tends to 0 as $N \to +\infty$, which finishes the proof. □

Remark B.8.2. This result is one ingredient (and a special case) of Kolmogorov's Three Series Theorem which gives a necessary and sufficient condition for almost sure convergence of a series of independent complex random variables. It is worth mentioning two further results for context: (1) the event "the series converges" is an asymptotic event, in the sense that it doesn't depend on any finite number of the random variables; Kolmogorov's Zero-One Law then shows that this event can only have probability 0 or 1; (2) a theorem of P. Lévy shows that, again for independent summands, the almost sure convergence is equivalent to convergence in law, or to convergence in probability. For proofs and discussion of these facts, see for instance [25, §0.III].

Here is Kolmogorov's inequality:

Lemma B.8.3. *Let $M \geqslant 1$ be an integer, $Y_1, \ldots, Y_M$ independent complex random variables in L^2 with $\mathbf{E}(Y_n) = 0$ for all n. Then for any $\varepsilon > 0$, we have*

$$\mathbf{P}\Big(\sup_{1\leqslant k\leqslant M}|Y_1 + \cdots + Y_k| > \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\sum_{n=1}^{M}\mathbf{V}(Y_n).$$

Proof. Let

$$S_n = Y_1 + \cdots + Y_n$$

for $1 \leqslant n \leqslant M$. We define a random variable T with values in $[0, +\infty]$ by $T = \infty$ if $|S_n| \leqslant \varepsilon$ for all $n \leqslant M$, and otherwise

$$T = \inf\{n \leqslant M \mid |S_n| > \varepsilon\}.$$

We then have

$$\sup_{1\leqslant k\leqslant M}|Y_1 + \cdots + Y_k| > \varepsilon\} = \bigcup_{1\leqslant n\leqslant M}\{T = n\},$$

and the union is disjoint. In particular, we get

$$\mathbf{P}\Big(\sup_{1\leqslant k\leqslant M}|S_k| > \varepsilon\Big) = \sum_{n=1}^{M}\mathbf{P}(T = n).$$

We now note that $|S_n|^2 \geqslant \varepsilon^2$ on the event $\{T = n\}$, so that we can also write

$$\mathbf{P}\Big(\sup_{1\leqslant k\leqslant M}|S_k| > \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\sum_{n=1}^{M}\mathbf{E}(|S_n|^2\mathbf{1}_{\{T=n\}}). \tag{B.7}$$

We claim next that

$$\mathbf{E}(|S_n|^2\mathbf{1}_{\{T=n\}}) \leqslant \mathbf{E}(|S_M|^2\mathbf{1}_{\{T=n\}}) \tag{B.8}$$

for all $n \leqslant M$.

Indeed, if we write $S_M = S_n + R_n$, the independence assumption shows that R_n is independent of $(X_1, \ldots, X_n)$, and in particular is independent of the characteristic function of the event

$\{T = n\}$, which only depends on $X_1, \dots, X_n$. Moreover, we have $\mathbf{E}(R_n) = 0$. Now, taking the modulus square in the definition and multiplying by this characteristic function, we get

$$|S_M|^2 \mathbf{1}_{\{T=n\}} = |S_n|^2 \mathbf{1}_{\{T=n\}} + S_n \overline{R_n} \mathbf{1}_{\{T=n\}} + \overline{S_n} R_n \mathbf{1}_{\{T=n\}} + |R_n|^2 \mathbf{1}_{\{T=n\}}.$$

Taking then the expectation, and using the positivity of the last term, this gives

$$\mathbf{E}(S_n \overline{R_n} \mathbf{1}_{\{T=n\}}) \leqslant \mathbf{E}(S_M \overline{R_n} \mathbf{1}_{\{T=n\}}) + \mathbf{E}(S_n \overline{R_n} \mathbf{1}_{\{T=n\}}) + \mathbf{E}(\overline{S_n} R_n \mathbf{1}_{\{T=n\}}).$$

But, by independence, we have

$$\mathbf{E}(S_n \overline{R_n} \mathbf{1}_{\{T=n\}}) = \mathbf{E}(S_n \mathbf{1}_{\{T=n\}}) \, \mathbf{E}(\overline{R_n}) = 0,$$

and similarly $\mathbf{E}(\overline{S_n} R_n \mathbf{1}_{\{T=n\}}) = 0$. Thus we get the bound (B.8).

Using this in (B.7), this gives

$$\mathbf{P}\Big(\sup_{1 \leqslant k \leqslant M} |S_k| > \varepsilon \Big) \leqslant \frac{1}{\varepsilon^2} \sum_{n=1}^{M} \mathbf{E}(|S_N|^2 \mathbf{1}_{\{T=n\}}) \leqslant \frac{1}{\varepsilon^2} \sum_{n=1}^{M} \mathbf{E}(|S_N|^2)$$

by positivity once again. □

The second result we need is more subtle. It concerns similar series, but *without* the independence assumption, which is replaced by an orthogonality condition.

THEOREM B.8.4 (Menshov-Rademacher). *Let (X_n) be a sequence of complex-valued random variables such that $\mathbf{E}(X_n) = 0$ and*

$$\mathbf{E}(X_n \overline{X_m}) = \begin{cases} 0 & \text{if } n \neq m, \\ 1 & \text{if } n = m. \end{cases}$$

Let (a_n) be any sequence of complex numbers such that

$$\sum_{n \geqslant 1} |a_n|^2 (\log n)^2 < +\infty.$$

Then the series

$$\sum_{n \geqslant 1} a_n X_n$$

converges almost surely, and hence also in law.

REMARK B.8.5. Consider the probability space $\Omega = \mathbf{R}/\mathbf{Z}$ with the Lebesgue measure, and the random variables $X_n(t) = e(nt)$ for $n \in \mathbf{Z}$. One easily sees (adapting to double-sided sequences and symmetric partial sums) that Theorem B.8.4 implies that the series

$$\sum_{n \in \mathbf{Z}} a_n e(nt)$$

converges almost everywhere (with respect to Lebesgue measure), provided

$$\sum_{n \in \mathbf{Z}} |a_n|^2 (\log |n|)^2 < +\infty.$$

This may be proved more directly (see, e.g., [36, III, th. 4.4]), using properties of Fourier series, but it is not an obvious fact. Note that, in this case, the well-known theorem of Carleson shows that the condition may be replaced with $\sum |a_n|^2 < +\infty$. Menshov proved that Theorem B.8.4 can *not* be relaxed in this way (in fact, the term $(\log n)^2$ can not be replaced by any positive function $f(n)$ such that $f(n) = o((\log n)^2)$, even for $\mathbf{R}/\mathbf{Z}$.

We begin with a lemma which will play an auxiliary role similar to Kolmogorov's inequality.

LEMMA B.8.6. *Let $(X_1, \dots, X_N)$ be orthonormal random variables, $(a_1, \dots, a_N)$ be complex numbers and $S_k = a_1X_1 + \cdots + a_kX_k$ for $1 \leqslant k \leqslant N$. We have*

$$\mathbf{E}\big(\max_{1\leqslant k\leqslant N} |S_k|^2\big) \ll (\log N)^2 \sum_{n=1}^{N} |a_n|^2,$$

where the implied constant is absolute.

PROOF. The basic ingredient is a simple combinatorial property, which we present a bit abstractly. We claim that there exist a family $\mathcal{J}$ of discrete intervals

$$I = \{n_I, \dots, m_I - 1\}, \qquad m_I - n_I \geqslant 1,$$

for $I \in \mathcal{J}$, with the following two properties:

(1) Any interval $1 \leqslant n \leqslant M$ with $M \leqslant N$ is the disjoint union of $\ll \log N$ intervals $I \in \mathcal{J}$;

(2) An integer n with $1 \leqslant n \leqslant N$ belongs to $\ll \log N$ intervals in $\mathcal{J}$;

and in both cases the implied constant is independent of N.

To see this, let $n \geqslant 1$ be such that $2^{n-1} \leqslant N \leqslant 2^n$ (so that $n \ll \log N$), and consider for instance the family of dyadic intervals

$$I_{i,j} = \{n \mid 1 \leqslant n \leqslant N \text{ and } i2^j \leqslant n < (i+1)2^j\}$$

for $0 \leqslant j \leqslant N$ and $1 \leqslant i \leqslant 2^{n-j}$.

Now, having fixed such a collection of intervals, we denote by T the smallest integer between 1 and N such that

$$\max_{1\leqslant k\leqslant N} |S_k| = |S_T|.$$

By our first property of the intervals $\mathcal{J}$, we can write

$$S_T = \sum_I \tilde{S}_I$$

where I runs over a set of $\ll \log N$ disjoint intervals in $\mathcal{J}$, and

$$\tilde{S}_I = \sum_{n\in I} a_n X_n$$

is the corresponding partial sum. By the Cauchy-Schwarz inequality, and the first property again, we get

$$|S_T|^2 \ll (\log N) \sum_I |\tilde{S}_I|^2 \ll (\log N) \sum_{I\in\mathcal{J}} |\tilde{S}_I|^2.$$

Taking the expectation and using orthonormality, we derive

$$\begin{aligned}\mathbf{E}\big(\max_{1\leqslant k\leqslant N} |S_k|^2\big) = \mathbf{E}(|S_T|^2) &\ll (\log N) \sum_{I\in\mathcal{J}} \mathbf{E}(|\tilde{S}_I|^2) \\ &= (\log N) \sum_{I\in\mathcal{J}} \sum_{n\in I} |a_n|^2 \ll (\log N)^2 \sum_{1\leqslant n\leqslant N} |a_n|^2\end{aligned}$$

by the second property of the intervals $\mathcal{J}$. □

PROOF OF THE MENSHOV–RADEMACHER THEOREM. If the factor $(\log N)^2$ in Lemma B.8.6 was replaced with $(\log n)^2$ inside the sum, we would proceed just like the deduction of Theorem B.8.1 from Lemma B.8.3. Since this is not the case, a slightly different argument is needed.

We define

$$S_n = a_1X_1 + \cdots + a_nX_n$$

for $n \geqslant 1$. For $j \geqslant 0$, we also define the dyadic sum

$$\tilde{S}_j = \sum_{2^j\leqslant n<2^{j+1}} a_nX_n = S_{2^{j+1}-1} - S_{2^j}.$$

We first note that the series

$$T = \sum_{j\geqslant 0}(j+1)|\tilde{S}_j|^2$$

converges almost surely. Indeed, since it is a series of non-negative terms, it suffices to show that $\mathbf{E}(T) < +\infty$. But we have

$$\mathbf{E}(T) = \sum_{j\geqslant 0}(j+1)\,\mathbf{E}(|\tilde{S}_j|^2) = \sum_{j\geqslant 0}(j+1)\sum_{2^j\leqslant n<2^{j+1}}|a_n|^2 \ll \sum_{n\geqslant 1}|a_n|^2(\log 2n)^2 < +\infty$$

by orthonormality and by the assumption of the theorem.

Next, we observe that for $j \geqslant 0$ and $k \geqslant 0$, we have

$$|S_{2^{j+k}} - S_{2^j}| \leqslant \sum_{i=j}^{j+k-1}|\tilde{S}_i| \leqslant \Big(\sum_{j\leqslant i<j+k}\frac{1}{(i+1)^2}\Big)^{1/2}|T|^{1/2} \ll \Big(\frac{|T|}{j+1}\Big)^{1/2}$$

by the Cauchy-Schwarz inequality. Hence the sequence (S_{2^j}) is almost surely a Cauchy sequence, and hence converges almost surely to a random variable S.

Finally, to prove that (S_n) converges almost surely to S, we observe that for any $n \geqslant 1$, and $j \geqslant 0$ such that $2^j \leqslant n < 2^{j+1}$, we have

$$|S_n - S_{2^j}| \leqslant M_j = \max_{2^j<k\leqslant 2^{j+1}}\Big|\sum_{m=2^j}^{k}a_nX_n\Big|. \tag{B.9}$$

Lemma B.8.6 implies that

$$\mathbf{E}\Big(\sum_{j\geqslant 0}M_j^2\Big) = \sum_{j\geqslant 0}\mathbf{E}(M_j^2) \ll \sum_{n\geqslant 1}(\log 2n)^2|a_n|^2 < +\infty,$$

which means in particular that M_j tends to 0 as $j \to +\infty$ almost surely. From (B.9) and the convergence of $(S_{2^j})_j$ to S, we deduce that (S_n) converges almost surely to S. This finishes the proof. □

We will also use information on the support of the distribution of a random series with independent summands.

Proposition B.8.7. *Let B be a separable Banach space. Let $(X_n)_{n\geqslant 1}$ be a sequence of* independent *B-valued random variables such that the series $S = \sum X_n$ converges almost surely.*[1] *The support of the law of S contains the closure of the set of all convergent series of the form $\sum x_n$, where x_n belongs to the support of the law of X_n for all $n \geqslant 1$.*

Proof. For $N \geqslant 1$, we write

$$S_N = \sum_{n=1}^{N}X_n, \qquad R_N = X - S_N.$$

The variables S_N and R_N are independent.

First, we observe that Lemmas B.1.1 and B.1.2 imply that the support of S_N is the closure of the set of elements $x_1 + \cdots + x_N$ with $x_n \in \mathrm{supp}(X_n)$ for $1 \leqslant n \leqslant N$ (apply Lemma B.1.1 to the law of $(X_1, \ldots, X_N)$ on B^N, which has support the product of the $\mathrm{supp}(X_n)$ by Lemma B.1.2, and to the addition map $B^N \to B$).

We will prove that all convergent series $\sum x_n$ with $x_n \in \mathrm{supp}(X_n)$ belong to the support of X, hence the closure of this set is contained in the support of X, as claimed. Thus let $x = \sum x_n$ be of this type. Let $\varepsilon > 0$ be fixed.

For all N large enough, we have

$$\Big|\sum_{n>N}x_n\Big| < \varepsilon,$$

[1] Recall that by the result of P. Lévy mentioned in Remark B.8.2, this is equivalent in that case to convergence in law.

and it follows that $x_1 + \cdots + x_N$, which belongs to the support of S_N as first remarked, also belongs to the open ball U_ε of radius ε around x. Hence

$$\mathbf{P}(S_N \in U_\varepsilon) > 0$$

for all N large enough (U_ε is an open neighborhood of some element in the support of S_N).

Now the almost sure convergence implies (by the dominated convergence theorem, for instance) that $\mathbf{P}(\|R_N\| > \varepsilon) \to 0$ as $N \to +\infty$. Therefore, taking N suitably large, we get

$$\begin{aligned}\mathbf{P}(\|S - x\| < 2\varepsilon) &\geqslant \mathbf{P}(\|S_N - x\| < \varepsilon \text{ and } \|R_N\| < \varepsilon) \\ &= \mathbf{P}(\|S_N - x\| < \varepsilon)\,\mathbf{P}(\|R_N\| < \varepsilon) > 0\end{aligned}$$

(by independence). Since ε is arbitrary, this shows that $x \in \mathrm{supp}(S)$, as was to be proved. □

Exercise B.8.8. With assumptions as in Proposition B.8.7, show that the support of X is the set of all $x \in M$ such that

$$x = \lim_{N\to+\infty} \sum_{n=1}^{N} x_n^{(N)}$$

where $x_n^{(N)} \in \mathrm{supp}(X_n)$ for all n.

B.9. Some probability in Banach spaces

We consider in this section some simple facts about probability in a (complex) Banach space V. For simplicity, we will always assume that V is separable (so that, in particular, Radon measures on V have a well-defined support).

The first result concerns series

$$\sum_n X_n$$

where (X_n) is a sequence of *symmetric* random variables, which means that for any $N \geqslant 1$, and for any choice $(\varepsilon_1, \ldots, \varepsilon_N)$ of signs $\varepsilon_n \in \{-1, 1\}$ for $1 \leqslant n \leqslant N$, the random vectors

$$(X_1, \ldots, X_N) \text{ and } (\varepsilon_1 X_1, \ldots, \varepsilon_N X_N)$$

have the same distribution.

Symmetric random variables have remarkable properties. For instance, the next proposition can be compared with Kolmogorov's Theorem (Theorem B.8.1), but note that we make no assumption of integrability or independence on the summands!

Proposition B.9.1 (Lévy). *Let V be a separable Banach space with norm $\|\cdot\|$, and (X_n) a sequence of V-valued random variables. Assume that the sequence (X_n) is symmetric. Let*

$$S_N = X_1 + \cdots + X_N$$

for $N \geqslant 1$.

(1) For $N \geqslant 1$ and $\varepsilon > 0$, we have

$$\mathbf{P}(\max_{1\leqslant n\leqslant N} \|S_N\| > \varepsilon) \leqslant 2\,\mathbf{P}(\|S_N\| > \varepsilon).$$

Part (1) is known as *Lévy's reflection principle*, and can be compared with Kolmogovo's maximal inquality (Lemma B.8.3).

Proof. (1) Similarly to the proof of Lemma B.8.3, we define a random variable T by $T = \infty$ if $\|S_n\| \leqslant \varepsilon$ for all $n \leqslant N$, and otherwise

$$T = \inf\{n \leqslant N \mid \|S_n\| > \varepsilon\}.$$

Assume $T = k$ and consider the random variables

$$X'_n = X_n \text{ for } 1 \leqslant n \leqslant k, \quad X'_n = -X_n \text{ for } k+1 \leqslant n \leqslant N.$$

The sequence $(X'_n)_{1\leqslant n\leqslant N}$ has the same distribution as $(X_n)_{1\leqslant n\leqslant N}$. Let S'_n denote the partial sums of the sequence (X'_n), and T' the analogue of T for the sequence (X'_n). The event $\{T'=k\}$ is the same as $T=k$ since $X'_n=X_n$ for $n\leqslant k$. On the other hand, we have

$$S'_N = X_1+\cdots+X_k-X_{k+1}-\cdots-X_N = 2S_k-S_N.$$

Therefore

$$\mathbf{P}(\|S_N\|>\varepsilon \text{ and } T=k) = \mathbf{P}(\|S'_N\|>\varepsilon \text{ and } T'=k) = \mathbf{P}(\|2S_k-S_N\|>\varepsilon \text{ and } T=k).$$

By the triangle inequality we have

$$\{T=k\}\subset\{\|S_N\|>\varepsilon \text{ and } T=k\}\cup\{\|2S_n-S_K\|>\varepsilon \text{ and } T=k\}.$$

We deduce

$$\begin{aligned}\mathbf{P}(\max_{1\leqslant n\leqslant N}\|S_n\|>\varepsilon) &= \sum_{k=1}^{N}\mathbf{P}(T=k)\\ &\leqslant \sum_{k=1}^{N}\mathbf{P}(\|S_N\|>\varepsilon \text{ and } T=k)+\sum_{k=1}^{N}\mathbf{P}(\|2S_n-S_K\|>\varepsilon \text{ and } T=k)\\ &= 2\,\mathbf{P}(\|S_N\|>\varepsilon).\end{aligned}$$

□

We now consider the special case where the Banach space V is $C([0,1])$, the space of complex-valued continuous functions on $[0,1]$ with the norm

$$\|f\|_\infty = \sup_{t\in[0,1]}|f(t)|.$$

For a $C([0,1])$-valued random variable X and any fixed $t\in[0,1]$, we will denote by $X(t)$ the complex-valued random variable that is the value of the random function X at t, i.e., $X(t)=e_t\circ X$, where $e_t : C([0,1])\longrightarrow\mathbf{C}$ is the evaluation at t.

Definition B.9.2 (Convergence of finite distributions). Let (X_n) be a sequence of $C([0,1])$-valued random variables and let X be a $C([0,1])$-valued random variable. One says that (X_n) converges to X in the sense of finite distributions if and only if, for all integers $k\geqslant 1$, and for all

$$0\leqslant t_1<\cdots<t_k\leqslant 1,$$

the vectors $(X_n(t_1),\dots,X_n(t_k))$ converge in law to $(X(t_1),\dots,X(t_k))$, in the sense of convergence in law in $\mathbf{C}^k$.

One sufficient condition for convergence in finite distributions is the following:

Lemma B.9.3. *Let (X_n) be sequence of $C([0,1])$-valued random variables and let X be a $C([0,1])$-valued random variable, all defined on the same probability space. Assume that, for any $t\in[0,1]$, the random variables $(X_n(t))$ converge in L^1 to $X(t)$. Then (X_n) converges to X in the sense of finite distributions.*

Proof. Fix $k\geqslant 1$ and

$$0\leqslant t_1<\cdots<t_k\leqslant 1.$$

Let φ be a Lipschitz function on $\mathbf{C}^k$ (given the distance associated to the norm

$$\|(z_1,\dots,z_k)\| = \sum_i |z_i|,$$

for instance) with Lipschitz constant $C\geqslant 0$. Then we have

$$\Big|\mathbf{E}(\varphi(X_n(t_1),\dots,X_n(t_k)))-\mathbf{E}(\varphi(X_n(t_1),\dots,X_n(t_k)))\Big| \leqslant C\sum_{i=1}^{k}\mathbf{E}(|X_n(t_i)-X(t_i)|)$$

which tends to 0 as $n \to +\infty$ by our assumption. Hence Proposition B.2.2 shows that $(X_n(t_1), \dots, X_n(t_k))$ converges in law to $(X(t_1), \dots, X(t_k))$. This proves the lemma. □

Convergence in finite distributions is a necessary condition for convergence in law of (X_n) to X, but it is not sufficient (see, e.g., [4, Example 2.5] for a counterexample). However, it suffices under the additional condition of *tightness* (see Definition B.2.4):

THEOREM B.9.4 (Prokhorov). *Let (X_n) be a sequence of $C([0,1])$-valued random variables and let X be a $C([0,1])$-valued random variable. Suppose that that (X_n) converges to X in the sense of finite distributions. Then (X_n) converges in law to X in the sense of $C([0,1])$-valued random variables if and only if (X_n) is tight.*

For a proof, see, e.g., [4, Th. 7.1]. In applications, we need some criteria to detect tightness. One such criterion is due to Kolmogorov:

PROPOSITION B.9.5 (Kolmogorov's tightness criterion). *Let (X_n) be a sequence of $C([0,1])$-valued random variables. If there exists real numbers $\alpha > 0$, $\delta > 0$ and $C \geqslant 0$ such that, for any real numbers $0 \leqslant s < t \leqslant 1$ and any $n \geqslant 1$, we have*

$$\mathbf{E}(|X_n(t) - X_n(s)|^\alpha) \leqslant C|t-s|^{1+\delta},$$

then (X_n) is tight.

See for instance [30, Th. XIII.1.8] for a proof. The statement does not hold if the exponent $1+\delta$ is replaced by 1.

We will also use the following inequality of Talagrand, which gives a type of subgaussian behavior of sums of random variable in Banach spaces, extending standard properties of real or complex-valued random variables.

THEOREM B.9.6 (Talagrand). *Let V be a separable real Banach space and V' its dual. Let $(X_n)_{n\geqslant 1}$ be a sequence of independent real-valued random variables with $|X_n| \leqslant 1$ almost surely, and let $(v_n)_n$ be a sequence of elements of V. Assume that the series $\sum v_n X_n$ converges almost surely in V. Let $m \geqslant 0$ be a median of*

$$\Big\|\sum_n v_n X_n\Big\|.$$

Let $\sigma \geqslant 0$ be the real number such that

$$\sigma^2 = \sup_{\substack{\lambda \in V' \\ \|\lambda\| \leqslant 1}} \sum_n |\lambda(v_n)|^2.$$

For any real number $t > 0$, we have

$$\mathbf{P}\Big(\Big\|\sum_n v_n X_n\Big\| \geqslant t\sigma + m\Big) \leqslant 4\exp\Big(-\frac{t^2}{16}\Big).$$

This is a consequence of [32, Th. 13.2]. We recall that a median m of a real-valued random variable X is any real number such that

$$\mathbf{P}(X \geqslant m) \geqslant \frac{1}{2}, \qquad \mathbf{P}(X \leqslant m) \geqslant \frac{1}{2}.$$

A median always exists. If X is integrable, then Chebychev's inequality

$$\mathbf{P}(X \geqslant t) \leqslant \frac{\mathbf{E}(|X|)}{t} \tag{B.10}$$

shows that $m \leqslant 2\,\mathbf{E}(|X|)$.

APPENDIX C

Number theory

We review here the facts of number theory that we use, and give references for their proofs.

C.1. Primes and their distribution

One of the first non-trivial estimates about prime numbers is given by the Mertens formula:

PROPOSITION C.1.1. *There exists a constant $C \in \mathbf{R}$ such that, for any $x \geqslant 3$, we have*

$$\sum_{p \leqslant x} \frac{1}{p} = \log\log x + O(1),$$

and more precisely

$$\sum_{p \leqslant x} \frac{1}{p} = \log\log x + C + O((\log x)^{-2}). \tag{C.1}$$

Recall that an arithmetic function f is multiplicative if $f(nm) = f(n)f(m)$ whenever n and m are coprime. For such functions, the associated Dirichlet series has an Euler product expansion when it converges absolutely.

LEMMA C.1.2. *Let f be a multiplicative function. For all $s \in \mathbf{C}$ such that*

$$\sum_{n \geqslant 1} \frac{f(n)}{n^s}$$

converges absolutely*, we have*

$$\sum_{n \geqslant 1} \frac{f(n)}{n^s} = \prod_p (1 + f(p)p^{-s} + \cdots + f(p^k)p^{-ks} + \cdots),$$

where the right-hand side converges absolutely.

PROOF. Since

$$1 + f(p)p^{-s} + \cdots + f(p^k)p^{-ks} + \cdots$$

is, for any prime p, a subseries of $\sum f(n)n^{-s}$, the absolute convergence of the latter implies that all of these series are also absolutely convergent.

We first consider the case when $f(n) \geqslant 0$ for all n. Then for $N \geqslant 1$, we have

$$\prod_{p \leqslant N} \sum_{k \geqslant 0} f(p^k)p^{-ks} = \sum_{\substack{n \geqslant 1 \\ p \mid n \Rightarrow p \leqslant N}} f(n)n^{-s}$$

by expanding the product and using the absolute convergence and the uniqueness of factorization of integers. It follows that

$$\Big| \prod_{p \leqslant N} \sum_{k \geqslant 0} f(p^k)p^{-ks} - \sum_{n \leqslant N} f(n)n^{-s} \Big| \leqslant \sum_{n > N} f(n)n^{-\sigma}$$

(since we assume $f(n) \geqslant 0$). This converges to 0 as $N \to +\infty$, because the series $\sum f(n)n^{-s}$ is absoluvely convergent. Thus this case is done.

In the general case, replacing f by $|f|$, the previous argument shows that the product converges absolutely. Then we get in the same manner

$$\Big|\prod_{p\leqslant N}\sum_{k\geqslant 0} f(p^k)p^{-ks} - \sum_{n\leqslant N} f(n)n^{-s}\Big| \leqslant \sum_{n>N} |f(n)|n^{-\sigma} \longrightarrow 0$$

as $N \to +\infty$. □

C.2. The Riemann zeta function

The Riemann zeta function is the holomorphic function defined by the absolutely convergent Dirichlet series

$$\zeta(s) = \sum_{n\geqslant 1} n^{-s}$$

for $\mathrm{Re}(s) > 1$. By Lemma C.1.2, it has also the Euler product expansion

$$\zeta(s) = \prod_p (1-p^{-s})^{-1}$$

in this region.

It is known that the zeta function extends to a meromorphic function on all of $\mathbf{C}$, with a unique pole located at $s = 1$; this is simple pole with residue 1. Moreover, the analytic continuation of $\zeta(s)$ satisfies the functional equation

$$\pi^{-s/2}\Gamma\Big(\frac{s}{2}\Big)\zeta(s) = \pi^{-(1-s)/2}\Gamma\Big(\frac{1-s}{2}\Big)\zeta(1-s).$$

Because the Gamma function has poles at integers $-k$ for $k \geqslant 0$, it follows that $\zeta(-2k) = 0$ for $k \geqslant 1$ (the case $k = 0$ is special because of the pole at $s = 1$). The negative even integers are called the *trivial zeros* of $\zeta(s)$. Hadamard and de la Vallée Poussin proved (independently) that $\zeta(s) \neq 0$ for $\mathrm{Re}(s) = 1$, and it follows that the non-trivial zeros of $\zeta(s)$ are located in the *critical strip* $0 < \mathrm{Re}(s) < 1$.

Proposition C.2.1. (1) *For $1/2 < \sigma < 1$, we have*

$$\frac{1}{2T}\int_{-T}^{T} |\zeta(\sigma+it)|^2 dt \longrightarrow \zeta(2\sigma)$$

as $T \to +\infty$.

(2) *We have*

$$\frac{1}{2T}\int_{-T}^{T} |\zeta(\tfrac{1}{2}+it)|^2 dt \sim T(\log T)$$

for $T \geqslant 1$.

Proof. According to Proposition C.2.1, the first part will follow if we can establish that, for $1/ < \sigma < 1$, we have

$$\frac{1}{2T}\int_{-T}^{T} |\zeta(\sigma+it)|^2 dt \ll 1.$$

We will establish this, using a basic inequality that will imply that

$$\frac{1}{2T}\int_{-T}^{T} |\zeta(\tfrac{1}{2}+it)|^2 dt \ll T(\log T),$$

(which is an upper bound of the right order of magnitude for the second part).

□

For much more information concerning the analytic properties of the Riemann zeta function, see [35]. Note however that the deeper *arithmetic* aspects are best understood in the larger framework of L-functions, from Dirichlet L-functions to automorphic L-functions (see, e.g, [15, Ch. 5]).

C.3. Exponential sums

Bibliography

[1] B. Bagchi: *Statistical behaviour and universality properties of the Riemann zeta function and other allied Dirichlet series*, PhD thesis, Indian Statistical Institute, Kolkata, 1981; available at library.isical.ac.in/jspui/handle/10263/4256

[2] A. Barbour, E. Kowalski and A. Nikeghbali: *Mod-discrete expansions*, Probability Theory and Related Fields, 2013; doi:10.1007/s00440-013-0498-8.

[3] P. Billingsley: *Probability and measure*, 3rd edition, Wiley, 1995.

[4] P. Billingsley: *Convergence of probability measures*, 2nd edition, Wiley, 1999.

[5] A. Borel: *Linear algebraic groups*, 2nd edition, GTM 126, Springer 1991.

[6] E. Breuillard and H. Oh (eds): *Thin groups and super-strong approximation*, MSRI Publications Vol. 61, Cambridge Univ. Press, 2014.

[7] P. Erdős and M. Kac: *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math. 62 (1940), 738–742.

[8] G.B. Folland: *Real analysis*, Wiley, 1984.

[9] K. Ford: *The distribution of integers with a divisor in a given interval*, Annals of Math. 168 (2008), 367–433.

[10] J. Friedlander and H. Iwaniec: *Opera de cribro*, Colloquium Publ. 57, A.M.S, 2010.

[11] P.X. Gallagher: *The large sieve and probabilistic Galois theory*, in Proc. Sympos. Pure Math., Vol. XXIV, Amer. Math. Soc. (1973), 91–101.

[12] A. Granville: *The anatomy of integers and permutations*, preprint (2008), http://www.dms.umontreal.ca/~andrew/PDF/Anatomy.pdf

[13] A. Granville and K. Soundararajan: *Sieving and the Erdös-Kac theorem*, in "Equidistribution in number theory, an introduction", 15–27, Springer 2007.

[14] A. Harper: *Two new proofs of the Erdős–Kac Theorem, with bound on the rate of convergence, by Stein's method for distributional approximations*, Math. Proc. Camb. Phil. Soc. 147 (2009), 95–114.

[15] H. Iwaniec and E. Kowalski: *Analytic Number Theory*, Colloquium Publ. 53, A.M.S, 2004.

[16] N.M. Katz: *Gauss sums, Kloosterman sums and monodromy groups*, Annals of Math. Studies 116, Princeton Univ. Press (1988).

[17] D. Koukoulopoulos: *Localized factorizations of integers*, Proc. London Math. Soc. 101 (2010), 392–426.

[18] E. Kowalski: *The large sieve and its applications*, Cambridge Tracts in Math., vol 175, C.U.P (2008).

[19] E. Kowalski: *Poincaré and analytic number theory*, in "The scientific legacy of Poincaré", edited by É. Charpentier, É. Ghys and A. Lesne, A.M.S, 2010.

[20] E. Kowalski: *Sieve in expansion*, Séminaire Bourbaki, Exposé 1028 (November 2010).

[21] E. Kowalski: *An introduction to the representation theory of groups*, Grad. Studies in Math. 155, A.M.S, 2014.

[22] E. Kowalski: *The Kloostermania page*, blogs.ethz.ch/kowalski/the-kloostermania-page/

[23] E. Kowalski and A. Nikeghbali: *Mod-Poisson convergence in probability and number theory*, International Mathematics Research Notices 2010; doi:10.1093/imrn/rnq019

[24] E. Kowalski and W. Sawin: *Kloosterman paths and the shape of exponential sums*, Compositio Math., to appear.

[25] D. Li and H. Queffélec: *Introduction à l'étude des espaces de Banach; Analyse et probabilités*, Cours Spécialisés 12, S.M.F, 2004.

[26] W. Bosma, J. Cannon and C. Playoust: *The Magma algebra system, I. The user language*, J. Symbolic Comput. 24 (1997), 235–265; also http://magma.maths.usyd.edu.au/magma/

[27] PARI/GP, version 2.6.0, Bordeaux, 2011, http://pari.math.u-bordeaux.fr/.

[28] M. Radziwill and K. Soundararajan: *Selberg's central limit theorem for* $\log|\zeta(1/2+it)|$, preprint (2015).

[29] A. Rényi and P. Turán: *On a theorem of Erdös and Kac*, Acta Arith. 4 (1958), 71–84.

[30] D. Revuz and M. Yor: *Continuous Martingales and Brownian Motion*, 3rd ed., Springer-Verlag, Berlin, 1999.

[31] J.-P. Serre: *Linear representations of finite groups*, Grad. Texts in Math. 42, Springer (1977).

[32] M. Talagrand: *Concentration of measure and isoperimetric inequalities in product spaces*, Publ. Math. I.H.É.S 81 (1995), 73–205.

[33] G. Tenenbaum: *Introduction to analytic and probabilistic number theory*, Cambridge studies adv. math. 46, 1995.
[34] E.C. Titchmarsh: *The theory of functions*, 2nd edition, Oxford Univ. Press, 1939.
[35] E.C. Titchmarsh: *The theory of the Riemann zeta function*, 2nd edition, Oxford Univ. Press, 1986.
[36] A. Zygmund: *Trigonometric sums*, 3rd Edition, Cambridge Math Library, Cambridge, 2002.