Probability Theory Related Fields 137, 49–81 (2007)
Digital Object Identifier (DOI) 10.1007/s00440-006-0500-9

Sergey G. Bobkov† · Friedrich Götze

Concentration inequalities and limit theorems for randomized sums

Received: 25 June 2005 / Revised version: 18 December 2005 /
Published online: 24 April 2006 –

Abstract. Concentration properties and an asymptotic behaviour of distributions of normalized and self-normalized sums are studied in the randomized model where the observation times are selected from prescribed consecutive integer intervals.

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbf{P})$ denote a probability space. It is well known that any norm-bounded system in $L^2(\Omega, \mathcal{F}, \mathbf{P})$, say $\{X_n\}_{n=1}^{\infty}$, contains a lacunary-type subsystem $\{X_{n_k}\}$ whose elements behave like shifted independent random variables, possibly multiplied by an independent factor ρ (random or not). In particular, it was shown in 1960s by V. F. Gaposhkin [G2], that if X_n is convergent to zero weakly in L^2, a subsystem can be chosen such that the distributions of the normalized sums

$$S_N = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}}$$

will be weakly convergent to $N(0, \rho^2)$, that is, to the law of ρZ where Z is a standard normal random variable independent of the random variable $\rho \geq 0$. For short, this may be written as

$$S_N \Rightarrow N(0, \rho^2), \qquad \text{as } N \to \infty. \tag{1.1}$$

The first observation of this kind with a purely normal limit is apparently due to M. Kac in his 1938 paper [K1] and a somewhat later there is a similar result by R. Fortet [Fo], cf. also [K2]. On the unit interval $\Omega = (0, 1)$ equipped with Lebesgue measure $\mathbf{P}$, they considered special subsystems of the system $\{f(n\omega)\}$, where f is a fixed 1-periodic function on the real line.

Apart from questions concerning possible rates of increase of n_k, the general problem of the existence of subsequences of indices satisfying (1.1) was studied

S.G. Bobkov: School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA. e-mail: bobkov@math.umn.edu

F. Götze: Department of Mathematics, Bielefeld University, Bielefeld 33501, Germany. e-mail: goetze@math.uni-bielefeld.de

†Research supported in part by NSF Gr. No. 0405587

Key words or phrases: Concentration – typical distributions – central limit theorem – self-normalized statistics – orthogonal polynomials – pairwise independent random variables

by many authors with different methods and for more general schemes of sums involving weights. See for instance the work of G. W. Morgenthaler [Mo] for uniformly bounded orthonormal systems in L^2. Later on, various aspects of the above types of CLT have been intensively investigated by Gaposhkin, cf. [G1-4]. Some of his results were rediscovered by S. D. Chatterji [C] who introduced an informal statement known as principle of subsequences; it states that any limit theorem about independent, identically distributed random variables continues to hold under proper moment assumptions for a certain subsequence of a given sequence of random variables. This general observation was made precise and extended by D. J. Aldous [A], and later by I. Berkes and A. Péter [B-P].

However, not much is known about the speed of increase of a subsequence such that a central limit theorem holds. Historically, classical trigonometric systems like

$$X_n(\omega) = \cos(2\pi n\omega), \quad 0 < \omega < 1,$$

on the interval $\Omega = (0, 1)$ have been studied most intensively in this respect. The first result in this direction (after [K1-2]) was obtained by R. Salem and A. Zygmund [S-Z] showing in particular that $S_N \Rightarrow N(0, \frac{1}{2})$, whenever $\frac{n_{k+1}}{n_k} \geq q > 1$, for all k. This lacunary condition was weakened in terms of consecutive ratios of indices by P. Erdös [E] to the optimal condition $\frac{n_{k+1}}{n_k} \geq 1 + \frac{c_k}{\sqrt{k}}$, where $c_k \to +\infty$, cf. also [Ber1], [Mu], [Fu]. Note that here n_k must grow faster than $e^{\sqrt{k}}$. It was therefore an intriguing question whether a slower increasing sequence n_k can be chosen to satisfy the CLT and what the best possible rate is. To this aim, in 1978 I. Berkes [Ber2] proposed an implicit construction based on a random selection of indices.

Namely, assume the set of all natural numbers is partitioned into non-empty consecutive intervals Δ_k, $k \geq 1$, of respective lengths (cardinalities) $|\Delta_k| \to +\infty$. It was shown that, if we select each n_k from Δ_k independently and at random according to the discrete uniform distribution on Δ_k, then for almost all choices of indices, still $S_N \Rightarrow N(0, \frac{1}{2})$ holds. Hence, the gaps $n_{k+1} - n_k$ may grow as slow, as we wish. One of the purposes of the present note is to extend this result of Berkes in the case of self-normalized statistics

$$T_N = \frac{X_{n_1} + \dots + X_{n_N}}{\sqrt{X_{n_1}^2 + \dots + X_{n_N}^2}}$$

to general systems of random variables X_n in L^2 satisfying certain natural assumptions. Although it will not matter, one may define T_N to be zero in case the denominator is vanishing.

Throughout we assume that the "maximal spectral norm" λ of the associated correlation operators,

$$\lambda = \sup \mathbf{E}\, |a_1 X_1 + \dots + a_n X_n|^2,$$

is finite. Here the supremum is taken over all $n \geq 1$ and all collections $(a_1, \dots, a_n)$ of real (or, equivalently, complex) numbers such that $|a_1|^2 + \dots + |a_n|^2 \leq 1$. For example, any orthonormal system has spectral norm 1.

Introduce

$$V_k = \frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_n^2, \qquad \rho_N^2 = \frac{V_1 + \cdots + V_N}{N} \quad (\rho_N \geq 0).$$

Now we claim:

Theorem 1.1. *Let $|\Delta_k| \to +\infty$, and assume that*

a) $\max_{1 \leq k \leq N} \max_{n \in \Delta_k} |X_n| = o(\sqrt{N})$ *in probability, as $N \to \infty$;*
b) $\limsup_{N \to \infty} \mathbf{P}\{\rho_N \leq h\} \to 0$, *as $h \downarrow 0$.*

Then, for almost all indices $(n_k)_{k \geq 1}$, selected independently and uniformly from Δ_k,

$$T_N \Rightarrow N(0,1), \quad \text{as } N \to \infty. \tag{1.2}$$

If b) is replaced by the assumption that $\rho_N^2 \Rightarrow \rho^2$ weakly in distribution for some random variable ρ, then for almost all indices as selected above, we obtain the CLT (1.1).

We will comment on the assumption a) later on. The condition b) guarantees that the distributions of ρ_N should not have weak limit probability measures on the positive half-axis $[0, +\infty)$ with an atom at zero. This ensures that the random variables T_N are well-defined for large N with probability approaching 1.

Note that b) holds whenever $\rho_N^2 \Rightarrow \rho^2$ with $\rho > 0$ a.s., but the latter restriction is not important for the CLT in the form (1.1). For the cosine trigonometric system, we have $\rho_N^2 \to \rho^2 = \frac{1}{2}$ a.s., so it does not matter whether we consider normalized or self-normalized sums.

In case of orthogonal systems, a statement related to Theorem 1.1 has recently been derived in [B-G3] under similar hypotheses for a different randomized model, where every X_n is included in the partial sum $S_N = \varepsilon_1 X_1 + \cdots + \varepsilon_N X_N$ and the self-normalized statistic $T_N = \frac{\varepsilon_1 X_1 + \cdots + \varepsilon_N X_N}{\sqrt{\varepsilon_1 X_1^2 + \cdots + \varepsilon_N X_N^2}}$ with a prescribed probability p_n (thus, either $\varepsilon_n = 1$ or 0). As turns out, when p_n's approach zero in a certain stable way, (1.1)–(1.2) will hold with probability one in the sense of the infinite product Bernoulli measure on $\{0,1\}^\infty$. The latter was used to show that, whenever $\frac{m_k}{k} \to +\infty$, (1.1)–(1.2) hold for some fixed (non-random) sequence n_k satisfying $n_k \leq m_k$ for all k large enough. Although very close, this is however a weaker property in comparison with what one could potentially obtain in the Berkes model. Indeed, by the Erdös-Rényi "pure heads" theorem, in the scheme of Bernoulli trials we will be selecting indices $n_k = \min\{i > n_{k-1} : \varepsilon_i = 1\}$ with gaps $n_{k+1} - n_k$ of order at least $\log k$ for infinitely many k's, and therefore the rate of increase of gaps cannot be made as small as we wish.

Concerning the cosine trigonometric system, Berkes raised the natural question whether or not it is possible to find n_k with bounded gaps $n_{k+1} - n_k$ satisfying $S_N \Rightarrow N(0, \frac{1}{2})$, as in the randomized central limit theorem with growing gaps. We shall give a negative answer to this question. It turns out that a non-trivial limit $S_N \Rightarrow \xi$ is possible in case of bounded gaps, but then necessarily $\mathbf{E}\xi^2 < \frac{1}{2}$ (so, part of the second moment must be vanishing in the limit distribution). Nevertheless, it is of course a challenging problem to describe all possible weak limits of S_N. To see what may happen in the typical situation, let us look at the simplest case of a partition into intervals of cardinality 2.

Theorem 1.2. *For almost all indices n_k, selected independently and uniformly from the two point integer sets $\Delta_k = \{2k-1, 2k\}$, we have*

$$\sqrt{2}\, S_N \Rightarrow N(0, \rho^2), \quad \text{as well as} \quad T_N \Rightarrow N(0, \rho^2),$$

where ρ is distributed according to the arcsin law.

More precisely ρ has the distribution function $F(x) = \frac{2}{\pi}\,\arcsin(x), 0 < x < 1$. Hence, $\mathbf{E}\rho^2 = \frac{1}{2}$, while for the normalized partial sums $\mathbf{E}\,(\sqrt{2}\,S_N)^2 = 1$. Other typical distributions appear in the limit for intervals Δ_k of larger length. As we will see, the limit of S_N essentially reflects the density of the index sequence n_k in the set of all natural numbers.

It is of course a remarkable intrinsic feature that one may freely speak about typical distributions regardless of the behaviour of lengths $|\Delta_k|$ in randomized limit theorems. This is due to a strong concentration property of the family of the distributions F_τ of $S_N = S_N(\tau)$ with respect to finite selections $\tau = (n_1, \dots, n_N)$, which may be of independent interest. To be more precise, fix N and equip the space $M = \Delta_1 \times \dots \times \Delta_N$ of all τ's of length N with a (discrete) uniform probability measure μ. In terms of the Levy's distance between probability measures on the line we obtain, in particular:

Theorem 1.3. *For any $\delta > 0$,*

$$\mu\{\tau : L(F_\tau, F_N) \geq \delta\} \leq C\, e^{-cN}, \tag{1.3}$$

where $F_N = \int F_\tau \, d\mu(\tau)$ is the average distribution, and where C, c denote positive constants depending on λ and δ, only.

Note that the strength of concentration depends on λ only. Inequality (1.3) is obtained as an application of an abstract logarithmic Sobolev inequality in product probability spaces and a concentration phenomenon associated with it. Let us mention in this connection that the concentration property of $\{F_\tau\}$ around its μ-mean F_N is also known for some other distributions μ over the space of collections τ of indices, cf. [B-G3], [B2]. This line of applications of concentration methods originates in the work of V. N. Sudakov [S] and is now being actively discussed in the literature.

The paper is organized as follows. The basic concentration tools are discussed in section 2. In section 3 they are used to derive concentration inequalities for characteristic functions of randomized sums. A more detailed version of Theorem 1.3 is established in section 4. Section 5 is devoted to the asymptotic behaviour of the average distributions. Theorem 1.1 is proved in section 6. In section 7 we discuss typical distributions for the trigonometric system appearing in the model with bounded gaps and prove, in particular, Theorem 1.2. There it is shown as well that limit laws cannot be standard normal. Section 8 illustrates applications of Theorem 1.1 to sums of pairwise independent random variables.

2. Concentration in product spaces

In order to make the presentation self-contained, we recall in this section a rather abstract variant of the concentration phenomenon in product spaces, which will be used to study the concentration properties of normalized sums.

Given arbitrary probability spaces (M_k, μ_k), $1 \le k \le N$, consider the product space $(M, \mu) = (M_1, \mu_1) \otimes \cdots \otimes (M_N, \mu_N)$. With every complex-valued measurable function g on M, we associate the function of the "point" $x = (x_1, \dots, x_N) \in M$,

$$|\nabla g(x)|^2 = \frac{1}{2} \sum_{k=1}^{N} \int |g(x) - g(x_1, \dots, x_{k-1}, y_k, x_{k+1}, \dots, x_N)|^2 \, d\mu_k(y_k). \tag{2.1}$$

The quantity $|\nabla g(x)|$ may be viewed as the modulus of a gradient of g at the point x, and

$$\|g\|_{\mathrm{Lip}} = \operatorname{ess}\sup_{x \in M} |\nabla g(x)|$$

as the Lipschitz semi-norm of g. In particular, when the "dimension" $N = 1$, formula (2.1) simplifies to

$$|\nabla g(x)|^2 = \frac{1}{2} \int |g(x) - g(y)|^2 \, d\mu(y). \tag{2.2}$$

In the general case, the modulus of the gradient defined in (2.1) is of the additive type in the sense that

$$|\nabla g(x)|^2 = \sum_{k=1}^{N} |\nabla_{x_k} g(x)|^2, \qquad x = (x_1, \dots, x_N), \tag{2.3}$$

where we write $|\nabla_{x_k} g(x)|$ to emphasize that at this step the modulus of the gradient is evaluated as in dimension one with respect to the k-th variable (keeping the remaining variables fixed).

Proposition 2.1. *If $\|g\|_{\mathrm{Lip}} \le D$, then g is μ-integrable, and moreover, for any $h \ge 0$,*

$$\mu \left\{ \left| g - \int g \, d\mu \right| \ge h \right\} \le 4e^{-h^2/(8D^2)}. \tag{2.4}$$

A standard argument to derive (2.4) relies upon suitable Sobolev-type inequalities. As a first interesting example, let us note that according to (2.2), in dimension one we have $\mathrm{Var}_\mu(g) = \int |\nabla g|^2 \, d\mu$. In dimension N the variance functional possesses a subadditivity property

$$\mathrm{Var}_\mu(g) \le \int \sum_{k=1}^{N} \mathrm{Var}_{x_k}(g) \, d\mu$$

with the same understanding as in (2.3). This leads to a Poincaré-type inequality

$$\mathrm{Var}_\mu(g) \le \int |\nabla g|^2 \, d\mu \tag{2.5}$$

which can already be used to recover exponential tails of "Lipschitz" functions on the product space (M, μ). To reach the Gaussian decay as in (2.4), one has to work with a different class of analytic inequalities, and it appears that a modified logarithmic Sobolev inequality

$$\mathrm{Ent}_\mu(e^g) \le c \int |\nabla g|^2 e^g \, d\mu \tag{2.6}$$

represents the most convenient form responsible for the Gaussian-type concentration. Here g is an arbitrary real-valued function on M, and Ent_μ stands for the entropy functional defined for all non-negative integrable functions, say u, by

$$\mathrm{Ent}_\mu(u) = \int u \log(u) \, d\mu - \int u \, d\mu \ \log \int u \, d\mu.$$

In dimension one, by Jensen's inequality,

$$\begin{aligned} \mathrm{Ent}_\mu(e^g) \ \le \ \mathrm{cov}_\mu(g, e^g) &= \frac{1}{2} \int\!\!\int (g(x) - g(y))(e^{g(x)} - e^{g(y)}) \, d\mu(x) \, d\mu(y) \\ &\le \frac{1}{4} \int\!\!\int (g(x) - g(y))^2 (e^{g(x)} + e^{g(y)}) \, d\mu(x) \, d\mu(y) \\ &= \int |\nabla g|^2 e^g \, d\mu, \end{aligned}$$

where we used the elementary estimate $(a - b)(e^a - e^b) \le \frac{1}{2} (a - b)^2 (e^a + e^b)$, $a, b \in \mathbf{R}$. Therefore, (2.6) holds true with $c = 1$ when $N = 1$. Applying the subadditivity property of the entropy functional (similarly to the variance) and the additivity property for the gradient, one arrives at (2.6) for general product spaces.

Now, starting from (2.6) with $c = 1$ for functions tg with bounded g such that $|\nabla g| \le D$ on M, we get a distributional inequality $\mathrm{Ent}_\mu(e^{tg}) \le D^2 t^2 \int e^{tg} \, d\mu$ implying the bound on the Laplace transform

$$\int e^{t(g - \mathbf{E}_\mu g)} \, d\mu \le e^{D^2 t^2}, \quad t \in \mathbf{R},$$

where $\mathbf{E}_\mu g = \int g \, d\mu$. By Chebyshev's inequality, the latter yields

$$\mu\{g - \mathbf{E}_\mu g \ge h\} \le e^{-h^2/(4D^2)} \tag{2.7}$$

and therefore $\mu\{|g - \mathbf{E}_\mu g| \ge h\} \le 2 \exp\{-h^2/(4D^2)\}$. In the complex-valued case, this estimate applies separately to real and imaginary parts of g, and then we obtain (2.4).

This is how the standard argument leading to Proposition 4.1 works in the abstract framework. When every M_k consists of two points of equal μ_k-measure, (2.6) holds with optimal constant $c = \frac{1}{2}$, which follows from a logarithmic Sobolev inequality due to L. Gross [Gr]. Thus, (2.4) represents a natural generalization of the concentration phenomenon on the discrete cube. With the same constant, (2.6) can also be shown to hold on the discrete cube for non-symmetric product measures.

In the present form, the modified logarithmic Sobolev inequality (2.6) first appeared in M. Ledoux [L1] as a tool to recover a concentration inequality in abstract Hamming spaces, i.e., for functions on M that are Lipschitz with respect to the metric $d(x, y) = \mathrm{card}\{k \leq N : x_k \neq y_k\}$. M. Ledoux also used a similar argument as an alternative approach to an abstract scheme of penalties, developed in 1995 by M. Talagrand, cf. [T]. For other aspects related to isoperimetry and concentration in abstract product spaces, see also [L2], [B-G1, B-G2]. However, beyond the discrete cube, we are not aware of any applications of abstract deviation inequalities, such as (2.4), that are formulated in terms of the modulus of gradient defined via (2.2). So, we have found a little amazing that (2.4) can be effectively used to reach concentration for randomized distributions, as given in Theorem 3.1 below.

3. Concentration of characteristic functions

In the set of all natural numbers, let us fix non-empty consecutive integer intervals Δ_k, $1 \leq k \leq N$, of respective finite cardinalities $|\Delta_k|$, equipped with arbitrary probability measures μ_k.

Consider a random vector $X = (X_1, \dots, X_m)$ in $\mathbf{R}^m$, where $m = n_1 + \dots + n_N$.

With every collection of indices $\tau = (n_1, \dots, n_N)$ in $M = \Delta_1 \times \dots \times \Delta_N$, we associate the sums

$$S(\tau) = \frac{X_{n_1} + \dots + X_{n_N}}{\sqrt{N}}, \qquad R^2(\tau) = \frac{X_{n_1}^2 + \dots + X_{n_N}^2}{N},$$

and the two-dimensional random vector $W(\tau) = (S(\tau), R^2(\tau))$. Let G_τ denote the distribution of $W(\tau)$, and let

$$G = \int G_\tau \, d\mu(\tau)$$

be the μ-mean of these distributions on the plane with respect to the product measure $\mu = \mu_1 \otimes \dots \otimes \mu_N$ on M with marginals μ_k. In this section we study the concentration property of the family G_τ about G in terms of their characteristic functions

$$f_\tau(t, s) = \int e^{i(tx+sy)} \, dG_\tau(x, y),$$
$$f(t, s) = \int e^{i(tx+sy)} \, dG(x, y), \quad t, s \in \mathbf{R}.$$

A main step will be the following:

Theorem 3.1. *Assume* $\mathbf{E}X_n^2 \leq \sigma^2$, *for all* $1 \leq n \leq m$. *Then, for all* $t, s \in \mathbf{R}$, *and* $h > 0$,

$$\mu\{\tau : |f_\tau(t, s) - f(t, s)| \geq h\} \leq 4\, e^{-Nh^2/4D(t,s)^2}, \tag{3.1}$$

where $D(t,s) = \sqrt{2\lambda}\,|t| + 2\sigma^2(t^2 + |s|)$ with λ being the maximal eigenvalue of the correlation operator of X.

Although in further applications to randomized sums the measures μ will be taken to be discrete uniform on M, it should be emphasized that the bound on the rate of concentration can be chosen regardless of the product measure μ.

In the orthonormal case, $\lambda = \sigma = 1$, so $D(t,s) = \sqrt{2}\,|t| + 2t^2 + 2|s|$, and the right hand side of (3.1) has an exponential decay with respect to the "dimension" N. As for the general case, the given bound will be small provided that $\lambda = o(N)$ and $\sigma^2 = o(\sqrt{N})$.

Inequality (3.1) is immediately obtained from the deviation inequality (2.4) and the following bound on the Lipschitz semi-norm in the sense discussed in section 2:

Lemma 3.2. *For all $t, s \in \mathbf{R}$, the function $\tau \to f_\tau(t,s)$ has modulus of gradient satisfying, for all τ in M,*

$$|\nabla f_\tau(t,s)| \le \frac{\sqrt{2\lambda}\,|t| + 2\sigma^2(t^2 + |s|)}{\sqrt{2N}}. \tag{3.2}$$

Proof. To simplify notations, put $\mu_k(n) = \mu_k(\{n\})$ and

$$\tau_k(n) = (n_1, \dots, n_{k-1}, n, n_{k+1}, \dots, n_N), \quad n \in \Delta_k,$$

for a fixed collection $\tau = (n_1, \dots, n_N)$ in M. Then, for the function $g(\tau) = f_\tau(t,s)$ the definition (2.1) with $(M_k, \mu_k) = (\Delta_k, \mu_k)$ turns into

$$|\nabla g(\tau)|^2 = \frac{1}{2} \sum_{k=1}^N \sum_{n \in \Delta_k} |g(\tau) - g(\tau_k(n))|^2 \, \mu_k(n),$$

or equivalently,

$$|\nabla g(\tau)| = \frac{1}{\sqrt{2}} \max_a \left| \sum_{k=1}^N \sum_{n \in \Delta_k} a_k(n) \sqrt{\mu_k(n)}\, (g(\tau) - g(\tau_k(n))) \right|, \tag{3.3}$$

where summation is taken along all collections $a = \{a_k(n)\}$ of complex numbers such that

$$\sum_{k=1}^N \sum_{n \in \Delta_k} |a_k(n)|^2 = 1. \tag{3.4}$$

Now, since $S(\tau) - S(\tau_k(n)) = \frac{X_{n_k} - X_n}{\sqrt{N}}$ and $R^2(\tau) - R^2(\tau_k(n)) = \frac{X_{n_k}^2 - X_n^2}{N}$, we can split

$$e^{i\,(tS(\tau)+sR^2(\tau))} - e^{i\,(tS(\tau_k(n))+sR^2(\tau_k(n)))} = e^{i\,(tS(\tau)+sR^2(\tau))}\Big(1 - e^{-it(X_{n_k}-X_n)/\sqrt{N}}\Big)$$
$$+ e^{i\,(tS(\tau_k(n))+sR^2(\tau))}\Big(1 - e^{-is(X_{n_k}^2-X_n^2)/N}\Big).$$

Hence, the quantity after the max sign in (3.3) may be bounded from above by

$$\left|\mathbf{E}\, e^{i\,(tS(\tau)+sR^2(\tau))} \sum_{k=1}^{N} \sum_{n\in\Delta_k} a_k(n)\,\sqrt{\mu_k(n)}\left(1-e^{-it(X_{n_k}-X_n)/\sqrt{N}}\right)\right|$$
$$+\left|\mathbf{E}\sum_{k=1}^{N} \sum_{n\in\Delta_k} a_k(n)\,\sqrt{\mu_k(n)}\, e^{i\,(tS(\tau_k(n))+sR^2(\tau))}\left(1-e^{-is\,(X_{n_k}^2-X_n^2)/N}\right)\right|,$$

which in turn, using $|1-e^{i\alpha}|\le|\alpha|$, $\alpha\in\mathbf{R}$, inside the last sum, may be bounded by

$$\mathbf{E}\left|\sum_{k=1}^{N} \sum_{n\in\Delta_k} a_k(n)\,\sqrt{\mu_k(n)}\left(1-e^{-it(X_{n_k}-X_n)/\sqrt{N}}\right)\right| \tag{3.5}$$
$$+|s|\,\mathbf{E}\sum_{k=1}^{N} \sum_{n\in\Delta_k} |a_k(n)|\,\sqrt{\mu_k(n)}\;\frac{|X_{n_k}^2-X_n^2|}{N}. \tag{3.6}$$

By the assumption, $\mathbf{E}|X_{n_k}^2-X_n^2|\le 2\sigma^2$. Applying Cauchy's inequality, we also have $\sum_{n\in\Delta_k}|a_k(n)|\,\sqrt{\mu_k(n)}\le(\sum_{n\in\Delta_k}|a_k(n)|^2)^{1/2}$. Applying once more Cauchy's inequality and using (3.4), we get

$$\sum_{k=1}^{N} \sum_{n\in\Delta_k} |a_k(n)|\,\sqrt{\mu_k(n)}\le\sqrt{N}. \tag{3.7}$$

Hence, (3.6) is bounded by $\frac{2|s|\,\sigma^2}{\sqrt{N}}$. Similarly, we use $e^{i\alpha}=1+i\alpha+\theta\,\frac{\alpha^2}{2}$, $\alpha\in\mathbf{R}$, $|\theta|\le 1$, to bound (3.5) by

$$\frac{|t|}{\sqrt{N}}\mathbf{E}\left|\sum_{k=1}^{N} \sum_{n\in\Delta_k} a_k(n)\,\sqrt{\mu_k(n)}\,(X_{n_k}-X_n)\right|$$
$$+\frac{t^2}{2N}\mathbf{E}\sum_{k=1}^{N} \sum_{n\in\Delta_k} |a_k(n)|\,\sqrt{\mu_k(n)}\,(X_{n_k}-X_n)^2. \tag{3.8}$$

Again, $\mathbf{E}\,(X_{n_k}-X_n)^2\le 4\sigma^2$, so by (3.7), the second expectation in (3.8) is bounded by $4\sigma^2\sqrt{N}$. To estimate the first expectation, we shall use the maximal spectral eigenvalue (or the spectral radius) of the correlation operator of X, which may be described as the optimal constant λ satisfying

$$\mathbf{E}\left|\sum_{j=1}^{m}\alpha_j X_j\right|^2\le\lambda\sum_{j=1}^{m}\alpha_j^2,$$

for all $\alpha_j\in\mathbf{C}$. Hence, by the Cauchy inequality, the square of the first expectation in (3.8) is bounded by

$$\lambda \sum_{k=1}^{N} \Big| \sum_{n \in \Delta_k, n \neq n_k} a_k(n) \sqrt{\mu_k(n)} \Big|^2 + \lambda \sum_{k=1}^{N} \sum_{n \in \Delta_k, n \neq n_k} |a_k(n)|^2 \mu_k(n),$$

which does not exceed 2λ, by (3.4) and another application of the Cauchy inequality. Hence, (3.8) and thus (3.5) are bounded by

$$\frac{\sqrt{2\lambda}\,|t| + 2\sigma^2 t^2}{\sqrt{N}}.$$

Collecting the two estimates obtained for (3.5)-(3.6) and recalling (3.3), we arrive at the desired inequality (3.2). The proof is now complete. □

4. Concentration of distributions

Starting from Lemma 3.2 or Theorem 3.1, one may explore the rate of concentration of G_τ around G for various metrics metrizing weak convergence in the space of probability distributions on the plane. As a simple example, which can already be used in randomized limit theorems, consider the metric

$$\mathrm{dist}(H_1, H_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |h_1(t,s) - h_2(t,s)|\, e^{-2|t|-2|s|}\, dt\, ds,$$

where h_1 and h_2 are the characteristic functions of the probability measures H_1 and H_2, respectively, on the plane. Thus, with the same notations as in Theorem 3.1, we have:

Theorem 4.1. *For all $\delta > 0$,*

$$\mu\{\tau : \mathrm{dist}(G_\tau, G) \geq \delta\} \leq 2\, e^{-N\delta^2/(8B^2)}, \tag{4.1}$$

where $B = \sqrt{\lambda/2} + 2\sigma^2$.

Proof. At any fixed point τ in M, the modulus of the gradient $g \to |\nabla g(\tau)|$ represents a convex, homogeneous, and translation invariant functional. Hence, by Lemma 3.2,

$$\begin{aligned} |\nabla\, \mathrm{dist}(G_\tau, G)| &\leq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\nabla\,(f_\tau(t,s) - f(t,s))|\, e^{-2|t|-2|s|}\, dt\, ds \\ &\leq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{D(t,s)}{\sqrt{2N}}\, e^{-2|t|-2|s|} dt\, ds = \frac{\sqrt{\lambda/2} + 2\sigma^2}{\sqrt{2N}} \equiv D, \end{aligned}$$

where $D(t,s) = \sqrt{2\lambda}\,|t| + 2\sigma^2 t^2 + 2\sigma^2 |s|$, as before. In addition, by the Poincaré-type inequality (2.5), for all $t, s \in \mathbf{R}$,

$$\int |f_\tau(t,s) - f(t,s)|\, d\mu(\tau) \leq \left(\int |\nabla f_\tau(t,s)|^2\, d\mu \right)^{1/2} \leq \frac{D(t,s)}{\sqrt{2N}}.$$

Thus, for the distance-function $g(\tau) = \mathrm{dist}(G_\tau, G)$, we also have

$$\int g\,d\mu \ \le \ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{D(t,s)}{\sqrt{2N}}\,e^{-2|t|-2|s|}\,dt\,ds \ = \ D.$$

Therefore, applying to g the one-sided deviation inequality (2.7), we obtain that, for all $h \ge 0$,

$$\mu\{\tau : g(\tau) - D \ge Dh\} \le e^{-h^2/4}.$$

If $h \ge 1$, this implies $\mu\{g(\tau) \ge 2Dh\} \le e^{-h^2/4}$. Substituting $\delta = 2Dh$, we get

$$\mu\{g(\tau) \ge \delta\} \le e^{-\delta^2/16D^2}, \quad \delta \ge 2D.$$

Since $D = B/\sqrt{2N}$, we have arrived at the desired inequality (4.1) without factor 2. As for the range $\delta < 2D$, it remains to note that $\mu\{g \ge \delta\} \le 1 \le 2e^{-\delta^2/16D^2}$.

Theorem 4.1 follows.

More efforts are needed, if we want to control closeness of G_τ to G in terms of canonical metrics responsible for the weak convergence such as the Levy metric on the line. Here we consider the fiollowing most interesting particular case of this problem, restricting ourselves to randomized partial sums

$$S(\tau) = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}}, \qquad \tau = (n_1, \dots, n_N).$$

Denote by F_τ the distribution function of $S(\tau)$, and let $F(x) = \int F_\tau(x)\,d\mu(\tau)$ be the average distribution function with respect to μ, which is still supposed to be an arbitrary product measure on M. The Levy distance $L(F_\tau, F)$ is defined as the infimum over all $h \ge 0$ such that $F(x-h) - h \le F_\tau(x) \le F(x+h) + h$, for all $x \in \mathbf{R}$. □

Theorem 4.2. *For any $\delta > 0$,*

$$\mu\{\tau : L(F_\tau, F) \ge \delta\} \ \le \ C\,\frac{1+\lambda^{3/2}}{\delta^6}\ \exp\left\{-\frac{cN\,\delta^8}{\lambda + \sigma^4}\right\}, \tag{4.2}$$

where C, c are positive universal constants.

Theorem 1.3 follows from (4.2) by using $\sigma^2 \le \lambda$.

Proof. Without loss of generality, assume $\lambda > 0$. For the characteristic functions $f_\tau(t) = \int e^{itx}\,dF_\tau(x)$, $f(t) = \int e^{itx}\,dF(x)$, Theorem 3.1 with $s = 0$ gives, for any $h > 0$,

$$\mu\left\{\tau : \frac{|f_\tau(t) - f(t)|}{t} \ge h\right\} \le 4\,e^{-Nh^2/4D(t)^2}, \qquad t > 0, \tag{4.3}$$

where $D(t) = \sqrt{2\lambda} + 2\sigma^2 t$. If $t \le \frac{2}{h}$, we may use $D(t) \le D(\frac{2}{h})$ to derive from (4.3)

$$\mu\left\{\tau : \frac{|f_\tau(t) - f(t)|}{t} \ge h\right\} \le 4\,e^{-Nh^4/4V^2} \tag{4.4}$$

with $V = \sqrt{2\lambda}\, h + 4\sigma^2$. In the range $t > \frac{2}{h}$, the above inequality is fulfilled automatically, since $|f_\tau(t) - f(t)| \leq 2$, so that the left hand side is vanishing in that case. As $t \downarrow 0$, (4.4) becomes in the limit

$$\mu\{\tau : |\mathbf{E}S(\tau) - \mathbf{E}S| \geq h\} \leq 4\, e^{-Nh^4/4V^2}, \tag{4.5}$$

where S denotes a random variable with distribution function F. Thus, (4.5) may be viewed as a particular case of (4.4) when $t = 0$.

Now, in order to make (4.4) uniform with respect to t, we shall apply this inequality to points of the form $t = t_r = r \cdot ch^2$, $r = 0, \dots, \ell - 1$, with a positive integer ℓ and a real $c > 0$ to be specified later on. Then, for the set

$$M(h) = \left\{\tau \in M : \frac{|f_\tau(t_r) - f(t_r)|}{t_r} < h, \ \text{for all } r = 0, \dots, \ell - 1\right\}$$

we obtain from (4.4) that

$$1 - \mu(M(h)) \leq \sum_{r=0}^{\ell-1} \mu\left\{\frac{|f_\tau(t_r) - f(t_r)|}{t_r} \geq h\right\} \leq 4\ell\, e^{-Nh^4/4V^2}. \tag{4.6}$$

Take an arbitrary $t \in (0, \frac{1}{h}]$. By the definition of the spectral radius, for each $\tau \in M$, we have $\mathbf{E}S(\tau)^2 \leq \lambda$, so $|f_\tau'(t)| \leq \sqrt{\lambda}$, $|f_\tau''(t)| \leq \lambda$, and similarly for f.

Case $t \leq \frac{h}{\lambda}$. By Taylor's expansion in the integral form,

$$\frac{f_\tau(t) - f(t)}{t} = i\,(\mathbf{E}S_\tau - \mathbf{E}S) + t\int_0^1 (1 - v)(f_\tau''(tv) - f''(tv))\, dv.$$

Hence, if $\tau \in M(h)$ and, in particular, $|\mathbf{E}S_\tau - \mathbf{E}S| < h$, we obtain that $|\frac{f_\tau(t) - f(t)}{t}| < h + \lambda t \leq 2h$.

Case $\frac{h}{\lambda} < t \leq \frac{1}{h}$. Pick $r = 0, \dots, \ell - 1$ such that $t_r < t \leq t_{r+1}$. This is possible as long as $t_\ell \geq \frac{1}{h}$, i.e., we may take for ℓ the smallest integer which is greater than or equal to $\frac{1}{ch^3}$. Recalling that $t_{r+1} - t_r = ch^2$ and using the bound for the first derivative of f_τ and f, we may write, for any $\tau \in M(h)$,

$$\begin{aligned}
|f_\tau(t) - f(t)| &\leq |f_\tau(t) - f_\tau(t_r)| + |f_\tau(t_r) - f(t_r)| + |f(t_r) - f(t)| \\
&< \sqrt{\lambda}\,|t - t_r| + t_r h + \sqrt{\lambda}\,|t - t_r| \\
&\leq 2\sqrt{\lambda}\, ch^2 + t_r h \leq \ (2\lambda^{3/2}c + 1)\, th.
\end{aligned}$$

The assumption $h \leq \lambda t$ was used on the last step. Hence, taking $c = \frac{1}{2\lambda^{3/2}}$, we may conclude as in the first case that $|\frac{f_\tau(t) - f(t)}{t}| < 2h$.

Thus, the latter inequality holds in both cases, whenever $\tau \in M(h)$ and $0 < t \leq \frac{1}{h}$. It will also be fulfilled for $t > \frac{1}{h}$, just because $|f_\tau(t) - f(t)| \leq 2$. Hence, by (4.6),

$$\mu\left\{\sup_{t>0} \frac{|f_\tau(t) - f(t)|}{t} \geq 2h\right\} \leq 4\ell\, e^{-Nh^4/4V^2}. \tag{4.7}$$

Now, by Bohman's inequality [Bo], there is a general relationship $\frac{1}{2}\,L^2(F_\tau, F) \leq \sup_{t>0} \frac{|f_\tau(t)-f(t)|}{t}$. Hence, after the substitution $\delta = 2\sqrt{h}$, (4.7) implies

$$\mu\,\{L(F_\tau, F) \geq \delta\} \leq 4\ell\, e^{-N\delta^8/16V^2}. \tag{4.8}$$

Since $L(F_\tau, F) \leq 1$ always holds, only values $\delta \in [0, 1]$ or, equivalently, $h \in [0, \frac{1}{2}]$ are relevant in (4.8). In this case,

$$V^2 \leq \left(\sqrt{\lambda/2} + 4\sigma^2\right)^2 \leq \lambda + 32\,\sigma^4,$$

so the right hand side of (4.8) is bounded by $4\ell\ \exp\{-cN\delta^8/(\lambda+\sigma^4)\}$ with a universal constant $c > 0$. Finally, note that in both cases $\ell \leq \frac{1}{ch^3} + 1 = \frac{2\cdot 4^3\lambda^{3/2}}{\delta^6} + 1$. This finishes the proof. □

Remark 4.3. Similar concentration inequalities for the family F_τ around F remain to hold in a more general situation when $X_1, \dots, X_m$ represent random vectors, say, in $\mathbf{R}^d$. An appropriate definition of the involved parameters would be

1) $\max_{1\leq n\leq m} \mathbf{E}\,|X_n|^2 \ \leq\ \sigma^2$;
2) $\mathbf{E}\left|\sum_{n=1}^m \alpha_n X_n\right|^2 \ \leq\ \lambda\ \sum_{n=1}^m |\alpha_n|^2$, for all $\alpha_n \in \mathbf{R}$.

Here, $|\cdot|$ denotes the usual Euclidean norm in $\mathbf{R}^d$. Then, applying inequality (3.2) of Lemma 3.2 to the characteristic functions

$$f_\tau(t) = \mathbf{E}\,e^{i\langle t, S_\tau\rangle}, \quad f(t) = \int f_\tau(t)\,d\mu(\tau), \quad t \in \mathbf{R}^d,$$

we obtain

$$|\nabla f_\tau(t)| \leq \frac{\sqrt{2\lambda}\,|t| + 2\sigma^2|t|^2}{\sqrt{2N}}, \quad t \in \mathbf{R}^d.$$

Moreover, by a similar analysis as that used in the proof of Theorem 4.1, inequality (4.1) for the distance $\mathrm{dist}(F_\tau, F) = \int_{\mathbf{R}^d} |f_\tau(t) - f(t)|\, e^{-(2|t_1|+\dots+2|t_d|)}\,dt$, $t = (t_1, \dots, t_d)$, may be replaced by

$$\mu\{\tau : \mathrm{dist}(F_\tau, F) \geq \delta\} \leq 2\,e^{-N\delta^2/(8B^2)}, \tag{4.9}$$

where now $B = \sqrt{d\lambda} + d\sigma^2$ depends upon d as well.

For example, any orthonormal system $X_1, \dots, X_m$ of complex-valued random variables in the complex space $L^2(\Omega, \mathcal{F}, \mathbf{P})$ may be viewed as a system of two-dimensional random vectors, such that the conditions 1)-2) are fulfilled with $\sigma = \lambda = 1$. In this case, (4.9) turns into

$$\mu\{\tau : \mathrm{dist}(F_\tau, F) \geq \delta\} \leq 2\,e^{-cN\delta^2}$$

with an absolute positive constant c.

5. Average distributions

From now on, we will be dealing with an infinite system $(X_n)_{n\geq 1}$ of random variables in $L^2(\Omega, \mathcal{F}, \mathbf{P})$.

We assume that the set of all natural numbers is partitioned into non-empty consecutive intervals Δ_k, $k \geq 1$, of finite cardinalities $|\Delta_k|$, each equipped with a uniform discrete probability measure μ_k. As before, given a sequence of indices $\tau = (n_1, \dots, n_N)$ of length N in $M = \Delta_1 \times \cdots \times \Delta_N$, we consider the sums

$$S_N(\tau) = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}}, \qquad R_N^2(\tau) = \frac{X_{n_1}^2 + \cdots + X_{n_N}^2}{N},$$

and the two-dimensional random vectors $W_N(\tau) = (S_N(\tau), R_N^2(\tau))$. Recall that G_τ denotes the distribution of $W_N(\tau)$. Now, let

$$G_N = \int G_\tau \, d\mu(\tau) = \frac{1}{|\Delta_1| \dots |\Delta_N|} \sum_{\tau \in M} G_\tau$$

denote the corresponding average distribution with respect to the uniform discrete measure $\mu = \mu_1 \otimes \cdots \otimes \mu_N$ on M.

By Theorem 4.1, when N is a large number, under appropriate spectral and variance assumptions on X_n, most of G_τ are close to G_N in the sense of weak convergence of probability measures on the plane. Thus, it is natural to study the asymptotic behaviour of the average distributions G_N for growing N.

Note that G_N can be characterized as the joint distribution of the random variables

$$S_N = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}}, \qquad R_N^2 = \frac{X_{n_1}^2 + \cdots + X_{n_N}^2}{N},$$

where $n_1, \dots, n_N$ are now regarded as independent random indices, independent of all X_n, such that every n_k takes values in Δ_k with equal probabilities.

First we consider the sums S_N. Their asymptotic behaviour is mainly determined by the behaviour of the random variables

$$U_k = \frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_n, \qquad V_k = \frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_n^2,$$

provided that

$a)$ $\max_{1\leq k\leq N} \max_{n\in\Delta_k} |X_n| = o(\sqrt{N})$ in probability, as $N \to \infty$,
$b)$ $\sigma^2 = \sup_n \mathbf{E}X_n^2 < +\infty$,
$c)$ $\lambda = \sup_N \lambda_N < +\infty$,

where we use λ_N to denote the spectral radius of the correlation operator of the random vector $(X_1, \dots, X_m)$, $m = n_1 + \cdots + n_N$.

This may already be seen from:

Proposition 5.1. *Let Z denote a standard normal random variable independent of the sequence X_n. Under the conditions $a) - b)$, the distributions of S_N and*

$$\bar{S}_N = \frac{U_1 + \cdots + U_N}{\sqrt{N}} + \left(\frac{V_1 + \cdots + V_N}{N} - \frac{U_1^2 + \cdots + U_N^2}{N} \right)^{1/2} Z$$

are weakly convergent to each other in the sense that $\mathbf{E}\, e^{itS_N} - \mathbf{E}\, e^{it\bar{S}_N} \to 0$, as $N \to \infty$, uniformly on every bounded interval of the real line.

This statement about weak convergence may also be rewritten as the property that

$$\mathrm{dist}(F_N, \bar{F}_N) \to 0, \quad \text{as } N \to \infty, \tag{5.1}$$

where F_N and $\bar{F}_N$, respectively, are the distributions of S_N and $\bar{S}_N$, and where dist is a metric in the space of all probability measures on the line metrizing the weak convergence. For example, one may take $\mathrm{dist}(F, H) = \int_{-\infty}^{+\infty} |f(t) - h(t)|\, e^{-2|t|}\, dt$, where f, h are characteristic functions of probability measures F, H.

However, if we assume the hypothesis $c)$, which is stronger than $b)$, the property (5.1) will become independent of the choice of the metric. Indeed, by Prokhorov's compactness criterion, it is sufficient to make sure that the distributions involved have, say, uniformly bounded absolute moments. In our particular case, $\mathbf{E}S_N^2 \le \lambda$. For the other sequence, introduce

$$\bar{U}_N = \frac{U_1 + \cdots + U_N}{\sqrt{N}}.$$

By the spectral assumption, $\mathbf{E}\bar{U}_N^2 \le \frac{\lambda}{N} \sum_{k=1}^N \frac{1}{|\Delta_k|} \le \lambda$. By $b)$, $\mathbf{E}V_k \le \sigma^2 \le \lambda$, so we obtain that $\mathbf{E}\,|\bar{S}_N|^2 \le \mathbf{E}\,|\bar{U}_N|^2 + \mathbf{E}\,\frac{V_1 + \cdots + V_N}{N} \le 2\lambda$. This shows that no matter what metric is used in the assertion (5.1).

For the proof of Proposition 5.1, we need the following elementary statement of calculus:

Lemma 5.2. *Let ξ be a complex-valued random variable on some probability space (M, μ) such that $|\xi| \le \frac{1}{2}$. Then, for some complex θ, $|\theta| \le 1$,*

$$\mathbf{E}_\mu e^{\xi} = \exp\left\{ \mathbf{E}_\mu \xi + \frac{1}{2} \left[\mathbf{E}_\mu \xi^2 - (\mathbf{E}_\mu \xi)^2 \right] + C\theta\, \|\xi\|_\infty\, \mathbf{E}_\mu\, |\xi|^2 \right\},$$

where C is a universal positive constant, and $\|\xi\|_\infty = \operatorname{ess\,sup}_{x \in M} |\xi(x)|$.

Recall that we have introduced the notation

$$\rho_N^2 = \frac{V_1 + \cdots + V_N}{N}.$$

Proof of Proposition 5.1. By the very definition, the characteristic function $\varphi_N(t) = \mathbf{E}\, e^{itS_N}$, $t \in \mathbf{R}$, has the representation

$$\varphi_N(t) = \mathbf{E}\,\Pi_N(t), \qquad \text{where} \qquad \Pi_N(t) = \prod_{k=1}^N \left[\frac{1}{|\Delta_k|}\sum_{n\in\Delta_k} e^{itX_n/\sqrt{N}}\right].$$

Fix $t \neq 0$. The assumption $a)$ may formally be strengthened as the property that, for some sequence $\varepsilon_N \downarrow 0$, the events

$$\Omega_N = \left\{\omega \in \Omega : \max_{1\leq k\leq N}\max_{n\in\Delta_k} |X_n| \leq \varepsilon_N \sqrt{N}\right\}$$

have probabilities $\mathbf{P}(\Omega_N) \to 1$, as $N \to \infty$. Moreover, we may assume $\varepsilon_N \leq \frac{1}{2|t|}$, for all $N \geq 1$.

By Lemma 5.2, applied to $M = \Delta_k$ with uniform measure μ and to the random variable $\xi(n) = \frac{itX_n(\omega)}{\sqrt{N}}$ on Δ_k with a fixed $\omega \in \Omega_N$, we can write

$$\frac{1}{|\Delta_k|}\sum_{n\in\Delta_k} e^{itX_n/\sqrt{N}} = \exp\left\{\frac{it}{\sqrt{N}}\,U_k - \frac{t^2}{2N}\,(V_k - U_k^2) + C\theta_k t^3\,\varepsilon_N\,\frac{V_k}{N}\right\}$$

with some complex $\theta_k = \theta_k(\omega)$ such that $|\theta_k| \leq 1$. Therefore, putting $\eta_N = \frac{U_1^2+\cdots+U_N^2}{N}$, we arrive at

$$\Pi_N(t) = \exp\left\{it\bar{U}_N - \frac{t^2}{2}\,(\rho_N^2 - \eta_N) + Ct^3\theta_N'\varepsilon_N\rho_N^2\right\}$$

for another random complex θ_N' on Ω_N such that $|\theta_N'| \leq 1$.

Moreover, since $\mathbf{E}X_n^2 \leq \sigma^2$, we have $\mathbf{E}\,\rho_N^2 \leq \sigma^2$, and by Chebyshev's inequality, $\mathbf{P}\{\rho_N^2 \leq \frac{1}{\sqrt{\varepsilon_N}}\} \to 1$. Let $A_N = \Omega_N \cap \{\rho_N^2 \leq \frac{1}{\sqrt{\varepsilon_N}}\}$, so that $\mathbf{P}(A_N) \to 1$. On this set

$$\Pi_N(t) = \exp\left\{it\,\bar{U}_N - \frac{t^2}{2}\left(\rho_N^2 - \eta_N\right)\right\} e^{Ct^3\theta_N'\sqrt{\varepsilon_N}}.$$

Together with Cauchy's inequalities $V_k \geq U_k^2$, $\rho_N^2 \geq \eta_N$, the above representation implies that

$$\varphi_N(t) = \mathbf{E}\,e^{it\bar{S}_N} + o(1), \quad \text{as } N \to \infty,$$

where $o(1)$ is uniform with respect to t on bounded intervals of the real line.

As a corollary, we obtain:

Proposition 5.3. *Let $|\Delta_k| \to \infty$, and assume conditions $a) - c)$ are fulfilled. If $\rho_N^2 \Rightarrow \rho^2$ weakly in distribution for some random variable $\rho \geq 0$, then $S_N \Rightarrow N(0,\rho^2)$, as $N \to \infty$.*

The conclusion follows from Proposition 5.1 and the spectral assumption implying

$$\mathbf{E}\bar{U}_N^2 \leq \frac{\lambda}{N} \sum_{k=1}^{N} \frac{1}{|\Delta_k|} \to 0, \quad \text{as } N \to \infty. \tag{5.2}$$

If $C = \sup_n \mathbf{E}X_n^4 < +\infty$ and $\mathbf{E}X_k^2 X_j^2 \leq \mathbf{E}X_k^2\, \mathbf{E}X_j^2$, whenever $k < j$ (this holds, for example, for the cosine trigonometric system), then, $\mathrm{Var}(\frac{V_1+\cdots+V_N}{N}) \leq \frac{C}{N}$, so the assumption about ρ_N^2 is fulfilled with $\rho = 1$. That is, we arrive at a version of the central limit theorem, $S_N \Rightarrow N(0, 1)$, for sums with random indices.

Remark 5.4. In order to study applications to complex systems, let us state a multidimensional generalization of Proposition 5.3. Let $X_n = (X_{nj})_{1 \leq j \leq d}$ be a sequence of $\mathbf{R}^d$-valued random vectors on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ satisfying the properties $a)$ and $c)$ in the vector sense. The spectral norm λ is now defined by

$$\lambda = \sup \mathbf{E}\, |a_1 X_1 + \cdots + a_n X_n|^2,$$

where the sup is running over all $n \geq 1$ and over all collections $(a_1, \ldots, a_n)$ of real numbers such that $|a_1|^2 + \cdots + |a_n|^2 = 1$. Introduce the averages

$$V_k(j, l) = \frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_{nj} X_{nl}, \quad 1 \leq j, l \leq d,$$

and the associated sequence of (non-negatively definite) $d \times d$ random matrices ρ_N^2 with entries

$$\rho_N^2(j, l) = \frac{V_1(j, l) + \cdots + V_N(j, l)}{N}, \quad 1 \leq j, l \leq d.$$

With a similar argument, we have:

Proposition 5.5. *Assume $|\Delta_k| \to \infty$ and that, for some non-negatively definite $d \times d$ random matrix ρ, we have $\rho_N^2 \Rightarrow \rho^2$ in the sense of the weak convergence of distributions on $\mathbf{R}^{d \times d}$. Then, $S_N \Rightarrow \rho Z$ weakly in distribution, where Z is a standard normal random vector in $\mathbf{R}^d$ independent of ρ.*

Let us return to the real-valued case. Now we consider the asymptotic behaviour of the joint distribution of S_N and R_N^2.

Proposition 5.6. *Let $|\Delta_k| \to \infty$, and assume conditions $a) - c)$ are fulfilled. Then, the distributions of*

$$W_N = (S_N, R_N^2) \quad \textit{and} \quad \bar{W}_N = (\rho_N Z, \rho_N^2),$$

where Z is a standard normal random variable independent of ρ_n, are weakly convergent to each other, as $N \to \infty$.

Proof. is similar to that of Proposition 5.1. The statement about the weak convergence may be understood as

$$\mathrm{dist}(G_N, \bar{G}_N) \to 0, \quad \text{as } N \to \infty, \tag{5.3}$$

where $\bar{G}_N$ is the distribution of the random vector $\bar{W}_N$, and where dist is a metric in the space of all probability measures on the plane responsible for the weak convergence. Since the components of both W_N and $\bar{W}_N$ have bounded first absolute moments, this property does not depend on the choice of the metric, so it is sufficient to verify that the corresponding characteristic functions approach each other at every point.

Write the characteristic function of W_N

$$\varphi_N(t,s) = \mathbf{E}\, e^{itS_N + isR_N^2} = \mathbf{E}\, \Pi_N(t,s), \quad t, s \in \mathbf{R},$$

in terms of

$$\Pi_N(t,s) = \prod_{k=1}^N \left[\frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} e^{itX_n/\sqrt{N} + isX_N^2/N} \right].$$

Fix t, s not both zero. Choose a sequence $\varepsilon_N \downarrow 0$, such that the events

$$\Omega_N = \left\{ \omega \in \Omega : \max_{1 \le k \le N} \max_{n \in \Delta_k} |X_n| \le \varepsilon_N \sqrt{N} \right\}$$

have probabilities $\mathbf{P}(\Omega_N) \to 1$, as $N \to \infty$. Moreover, we may assume that all $\varepsilon_N \le \frac{1}{1+2(|t|+|s|)}$, so that $|\xi(n)| \le \frac{1}{2}$ on Ω_N, where

$$\xi(n) = \frac{itX_n(\omega)}{\sqrt{N}} + \frac{isX_n(\omega)^2}{N}.$$

By Lemma 5.2, applied to $M = \Delta_k$ with uniform measure μ_k and to the random variable ξ with a fixed $\omega \in \Omega_N$, there is a representation

$$\frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} e^{itX_n/\sqrt{N} + isX_N^2/N} = e^{z_k},$$

where

$$z_k = \int \xi\, d\mu_k + \frac{1}{2} \left[\int \xi^2\, d\mu_k - \left(\int \xi\, d\mu_k \right)^2 \right] + C\theta_k\, \|\xi\|_\infty \int |\xi|^2\, d\mu_k$$

with some complex $\theta_k = \theta_k(\omega)$ such that $|\theta_k| \le 1$. Now, $\|\xi\|_\infty \le (|t| + |s|)\, \varepsilon_N$ and

$$\int \xi\, d\mu_k = \frac{it}{\sqrt{N}}\, U_k + \frac{is}{N}\, V_k.$$

Squaring $\xi(n)^2 = -t^2 \, \frac{X_n^2}{N} - s^2 (\frac{X_n^2}{N})^2 - 2\, ts \, \frac{X_n^3}{N^{3/2}}$ and using $\varepsilon_N \le 1$, we get

$$\left| \xi(n)^2 + t^2 \, \frac{X_n^2}{N} \right| \le (s^2 + 2\, |t||s|) \, \varepsilon_N \, \frac{X_n^2}{N}.$$

Summing over all $n \in \Delta_k$ gives

$$\left| \int \xi^2 \, d\mu_k + t^2 \, \frac{V_k}{N} \right| \le (s^2 + 2\, |t||s|) \, \varepsilon_N \, \frac{V_k}{N}.$$

Similarly,

$$\int |\xi|^2 \, d\mu_k \le (|t| + |s|)^2 \, \frac{V_k}{N}.$$

All bounds together yield, for all $\omega \in \Omega_N$,

$$z_k = \frac{it}{\sqrt{N}} \, U_k + \frac{is}{N} \, V_k - \frac{t^2}{2N} \, V_k + \frac{1}{2} \left(\frac{t}{\sqrt{N}} \, U_k + \frac{s}{N} \, V_k \right)^2 + \theta_k(t,s) \, \varepsilon_N \frac{V_k}{N}$$

with a bounded random $\theta_k(t,s)$. Now, since $V_k \le \max_{n \in \Delta_k} |X_n|^2 \le \varepsilon_N^2 N$, we get

$$\frac{1}{2} \left| \frac{t}{\sqrt{N}} \, U_k + \frac{s}{N} \, V_k \right|^2 \le \frac{t^2}{N} \, U_k^2 + \frac{s^2}{N} \, V_k \, \varepsilon_N^2.$$

Therefore, after summation over all $k \le N$ and putting $\eta_N = \frac{U_1^2 + \dots + U_N^2}{N}$, we obtain that

$$\left| (z_1 + \dots + z_N) - \left(is - \frac{t^2}{2} \right) \rho_N^2 \right| \le |t| \, |\bar{U}_N| + t^2 \eta_N + \frac{s^2}{N} \, \rho_N^2 \, \varepsilon_N^2 + \theta_N'(t,s) \, \varepsilon_N \rho_N^2$$

with another bounded random $\theta_N'(t,s)$. Here the right hand side is getting small on sets of large probability. Indeed, since $\mathbf{E} X_n^2 \le \lambda$, we have $\mathbf{E} \, \rho_N^2 \le \lambda$, and by Chebyshev's inequality, $\mathbf{P}\{\rho_N^2 \le \frac{1}{\sqrt{\varepsilon_N}}\} \to 1$. Also recall that, by the spectral assumption c), both $\bar{U}_N$ and η_N tend to zero in $L^1(\Omega, \mathbf{P})$, cf. (5.2). So, $\mathbf{P}\{|\bar{U}_N| \le \varepsilon_N', \ \eta_N \le \varepsilon_N'\} \to 1$, for some $\varepsilon_N' \downarrow 0$. Let

$$A_N = \Omega_N \cap \left\{ \rho_N^2 \le \frac{1}{\sqrt{\varepsilon_N}}, \, |\bar{U}_N| \le \varepsilon_N', \ \eta_N \le \varepsilon_N' \right\},$$

so that $\mathbf{P}(A_N) \to 1$. On this set

$$\left| (z_1 + \dots + z_N) - \left(is - \frac{t^2}{2} \right) \rho_N^2 \right| \ \le \ \theta_N''(t,s) \, \varepsilon_N''$$

with some random bounded $\theta_N''(t,s)$ and non-random $\varepsilon_N'' \to 0$. Hence, $\Pi_N(t,s) = \exp\{(is - \frac{t^2}{2}) \, \rho_N^2\} + o(1)$ and, by the Lebesgue dominated convergence theorem,

$$\varphi_N(t,s) = \mathbf{E} \exp\left\{ \left(is - \frac{t^2}{2} \right) \rho_N^2 \right\} + o(1) = \mathbf{E} \exp\{it \, \rho_N Z + is\rho_N^2\} + o(1).$$

Thus, the proof is complete. □

6. Proof of Theorem 1.1

As before, we assume the set of all natural numbers is partitioned into non-empty consecutive intervals Δ_k, $k \geq 1$, of finite cardinalities $|\Delta_k|$, each equipped with a uniform discrete probability measure μ_k. Let μ_∞ denote the infinite product measure $\mu_1 \otimes \mu_2 \otimes \dots$ on $M_\infty = \Delta_1 \times \Delta_2 \times \cdots$. With every sequence of indices $\tau = (n_k)_{k \geq 1}$ in M_∞, we associate the sums

$$S_N(\tau) = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}}, \qquad R_N^2(\tau) = \frac{X_{n_1}^2 + \cdots + X_{n_N}^2}{N},$$

and the two-dimensional random vectors $W_N(\tau) = (S_N(\tau), R_N^2(\tau))$. Let $F_{\tau(N)}$ and $G_{\tau(N)}$ denote the distributions of $S_N(\tau)$ and $W_N(\tau)$, respectively, and let

$$F_N = \int F_{\tau(N)}\, d\mu_\infty(\tau), \qquad G_N = \int G_{\tau(N)}\, d\mu_\infty(\tau).$$

By Theorem 4.1, when N is a fixed and sufficiently large, most of the $G_{\tau(N)}$ are close to the average distribution G_N in the sense of the weak convergence of probability measures on the plane. This statement can be made precise, for example, by virtue of the finite dimensional concentration inequality (4.1), which now reads as

$$\mu_\infty\left\{\tau \in M_\infty : \mathrm{dist}(G_{\tau(N)}, G_N) \geq \delta\right\} \leq 2\, e^{-N\delta^2/(8B_N^2)}.$$

It holds true for any integer $N \geq 1$ and real $\delta > 0$ with $B_N = \sqrt{\lambda_N/2} + 2\sigma_N^2$, where $\sigma_N^2 = \max_{k \leq N} \max_{n \in \Delta_k} \mathbf{E}X_n^2$, and where λ_N is the spectral radius of the correlation operator of the random vector $(X_1, \dots, X_m)$, $m = n_1 + \cdots + n_N$. If $\lambda = \sup_N \lambda_N$ is finite, we get

$$\mu_\infty\left\{\tau \in M_\infty : \mathrm{dist}(F_{\tau(N)}, F_N) \geq \delta\right\} \leq 2\, e^{-cN\delta^2},$$
$$\mu_\infty\left\{\tau \in M_\infty : \mathrm{dist}(G_{\tau(N)}, G_N) \geq \delta\right\} \leq 2\, e^{-cN\delta^2}$$

with a constant c depending on λ, only. In particular, the series

$$\sum_{N=1}^{\infty} \mu_\infty\left\{\tau : \mathrm{dist}(F_{\tau(N)}, F_N) \geq \delta\right\}$$

is convergent. Therefore, by the Borel-Cantelli lemma, for any $\delta > 0$, for μ_∞-almost all τ, we have $\mathrm{dist}(F_{\tau(N)}, F_N) < \delta$, for all N large enough. If $\rho_N^2 \Rightarrow \rho^2$ weakly in distribution, then according to Proposition 5.3, $\mathrm{dist}(F_N, N(0, \rho^2)) \to 0$, so that

$$\limsup_{N \to \infty} \mathrm{dist}(F_{\tau(N)}, N(0, \rho^2)) \leq \delta$$

with μ_∞-probability one. This proves the second part of Theorem 1.1.

The first part is a little more delicate. Again applying the Borel-Cantelli lemma with $\delta = \delta_N$ of order, say, $N^{-1/4}$, we obtain that

$$\lim_{N\to\infty} \mathrm{dist}(G_{\tau(N)}, G_N) = 0$$

with μ_∞-probability one. Combining this with Proposition 5.6, we also get that with μ_∞-probability one

$$\lim_{N\to\infty} \mathrm{dist}(G_{\tau(N)}, \mathcal{L}(\rho_N Z, \rho_N^2)) = 0, \tag{6.1}$$

where $\mathcal{L}(\rho_N Z, \rho_N^2)$ is the joint distribution of $\rho_N Z$ and ρ_N^2, and where Z is a standard normal random variable independent of ρ_N.

Now, by the spectral assumption $\lambda < +\infty$, both families $\{G_{\tau(N)} : N \geq 1, \tau \in M_\infty\}$ and $\{G_N : N \geq 1\}$ are pre-compact in the space of all probability measures on the plane with respect to the dist-metric. Hence, (6.1) may be formulated as the property that, for any bounded, continuous function $u = u(x, y)$ in the upper half-plane $y \geq 0$, with μ_∞-probability one

$$\lim_{N\to\infty} \left| \mathbf{E}\, u(S_{\tau(N)}, R^2_{\tau(N)}) - \mathbf{E}\, u(\rho_N Z, \rho_N^2) \right| = 0. \tag{6.2}$$

By the assumption b) of Theorem 1.1, for large N the distributions of ρ_N^2 stay away from zero, so this also holds with μ_∞-probability one for the distributions of $R^2_{\tau(N)}$ (since, they approach each other, by (6.1)). That is,

$$\alpha(h) = \limsup_{N\to\infty} \mathbf{P}\{R^2_{\tau(N)} \leq h\} \to 0, \qquad \text{as } h \downarrow 0.$$

As a result, we can extend (6.2) to the class of all functions $u(x, y)$ that are bounded in the upper half-plane $y \geq 0$ and continuous at every point (x, y) with $y > 0$. Indeed, let $C = \sup_{x,y} |u(x, y)|$. Given $h > 0$, there exists a continuous function $u_h(x, y)$ in $y \geq 0$ that coincides with $u(x, y)$ whenever $y \geq h$ and satisfies $\sup_{x,y} |u_h(x, y)| \leq C$. Applying (6.2) to u_h, we then easily derive

$$\limsup_{N\to\infty} \left| \mathbf{E}\, u(S_{\tau(N)}, R^2_{\tau(N)}) - \mathbf{E}\, u(\rho_N Z, \rho_N^2) \right| \leq 2C\, (\alpha(h) + \beta(h)),$$

where $\beta(h) = \limsup_{N\to\infty} \mathbf{P}\{\rho_N^2 \leq h\}$. It remains to let $h \downarrow 0$.

Finally, starting from a bounded, continuous function v on the real line, apply (6.2) to $u(x, y) = v(x/\sqrt{y})$, $y > 0$, defining u to be zero in case $y = 0$. As a result, we obatin that $\lim_{N\to\infty} \mathbf{E}\, v(T_N) = \mathbf{E} v(Z)$.

This proves Theorem 1.1.

Remark 6.1. The condition a) of Theorem 1.1 is fulfilled automatically for uniformly bounded systems X_n, and more generally, for those satisfying

$$\sup_n |X_n| < +\infty \text{ a.s.}$$

If these random variables are unbounded, but have uniformly bounded moments, the condition a) will be fulfilled under an appropriate assumption on the growth of $|\Delta_k|$. Namely, assume

$$C_\Psi = \sup_n \mathbf{E}\,\Psi(|X_n|^2) < +\infty,$$

for some non-negative increasing function Ψ on $[0, +\infty)$ such that $\frac{\Psi(t)}{t} \to +\infty$, as $t \to +\infty$. Then, by Chebyshev's inequality,

$$\begin{aligned}
\mathbf{P}\Big\{\max_{1\le k\le N}\max_{n\in\Delta_k}|X_n| \ge c\sqrt{N}\Big\} &= \mathbf{P}\Big\{\Psi\Big(\max_{1\le k\le N}\max_{n\in\Delta_k}|X_n|^2\Big) \ge \Psi(c^2N)\Big\} \\
&\le \frac{\mathbf{E}\,\Psi\big(\max_{1\le k\le N}\max_{n\in\Delta_k}|X_n|^2\big)}{\Psi(c^2N)} \\
&\le \frac{\mathbf{E}\sum_{k=1}^N\sum_{n\in\Delta_k}\Psi(|X_n|^2)}{\Psi(c^2N)} \\
&\le C_\Psi\,\frac{|\Delta_1|+\dots+|\Delta_N|}{\Psi(c^2N)}.
\end{aligned}$$

Therefore, the condition a) will hold as long as, for any $c > 0$,

$$|\Delta_1| + \dots + |\Delta_N| = o(\Psi(cN)), \quad \text{as } N \to \infty, \tag{6.3}$$

Remark 6.2. With a similar argument and taking into account Remark 4.3 (in particular, inequality (4.9)) and Remark 5.4 (cf. Proposition 5.5), the second part of Theorem 1.1 may naturally be extended to $\mathbf{R}^d$-valued random vectors:

Theorem 6.3. *Under the assumptions of Proposition* 5.5, *for almost all indices* $(n_k)_{k\ge1}$, *selected independently and uniformly from* Δ_k, *weakly in distribution*

$$S_N \Rightarrow \rho Z, \quad \text{as } N \to \infty,$$

where Z *is a standard normal random vector in* $\mathbf{R}^d$ *independent of* ρ.

As a first illustratation of Theorem 6.3, we give a complex variant of Berkes' theorem about the cosine trigonometric system:

Corollary 6.4. *If* $|\Delta_k| \to \infty$, *then for almost all indices* $(n_k)_{k\ge1}$, *selected independently and uniformly from* Δ_k,

$$\frac{z^{n_1} + \dots + z^{n_N}}{\sqrt{N/2}} \Rightarrow N(0, \mathrm{I}_2)$$

weakly in distribution with respect to the normalized Lebesgue measure on the unit circle $S^1 = \{z \in \mathbf{C} : |z| = 1\}$.

Here $N(0, \mathrm{I}_2)$ denotes the standard normal distribution on the plane $\mathbf{R}^2$.

Proof. Consider

$$X_n(\omega) = (\cos(2\pi n\omega), \sin(2\pi n\omega)), \quad 0 < \omega < 1,$$

as a sequence of $\mathbf{R}^2$-valued random vectors on the probability space $\Omega = (0, 1)$ equipped with Borel σ-algebra $\mathcal{F}$ and a uniform probability measure $\mathbf{P}$. Then, in probability (and actually with probability 1)

$$\rho_N^2(j, j) = \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_{nj}^2 \right] \longrightarrow \frac{1}{2}, \qquad j = 1, 2,$$

which is due to the property that $\mathrm{cov}(X_{nj}^2, X_{nl}^2) = 0$, whenever $j < l$. In addition,

$$\rho_N^2(1, 2) = \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{|\Delta_k|} \sum_{n \in \Delta_k} X_{n1} X_{n2} \right] \longrightarrow 0,$$

since the random variables $X_{n1} X_{n2} = \frac{1}{2} X_{2n,1}$ are orthogonal in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ and have equal norms. Thus, the limit matrix

$$\rho^2 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

is non-random, and the conclusion follows. □

7. Bounded gaps

Here we continue the discussion of trigonometric systems, focusing on the case where the gaps $n_{k+1} - n_k$ remain bounded. To start with, let us first describe a general situation where a weak limit of partial sums S_N, if it exists, will not be standard normal.

Proposition 7.1. *Let $\{X_n\}_{n=1}^{\infty}$ be an orthonormal system in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ such that in probability*

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \to 0, \quad \text{as } n \to \infty. \tag{7.1}$$

Given an increasing sequence of indices $\tau = \{n_k\}_{k=1}^{\infty}$, assume that $S_N \Rightarrow \xi$ weakly in distribution, for some random variable ξ. Then,

$$\mathbf{E}\,\xi^2 \le 1 - \mathrm{den}(\tau). \tag{7.2}$$

Here, we use notation

$$\mathrm{den}(\tau) = \limsup_{N \to \infty} \frac{N}{n_N}$$

for the upper density of the sequence τ in the row of all natural numbers. In particular, if $\sup_k \big[n_{k+1} - n_k\big] < +\infty$, this quantity is positive, so ξ cannot be standard normal: part of the second moment is lost, while $\mathbf{E}S_N^2 = 1$, by orthonormality.

Note that the condition (7.1) is fulfilled for the (normalized) cosine trigonometric systems. Actually, in this case it does not matter whether we consider the sums S_N or self-normalized statistics T_N, since there holds true a strong law of large numbers $\frac{X_{n_1}^2+\cdots+X_{n_N}^2}{N} \to 1$.

Proof of Proposition 7.1. Without loss of generality, assume $\frac{N}{n_N} \to p = \mathrm{den}(\tau)$, as $N \to \infty$. Introduce the sums

$$Y_N = \frac{1}{\sqrt{n_N}} \sum_{1 \leq k \leq N} X_{n_k}, \qquad Z_N = \frac{1}{\sqrt{n_N}} \sum_{1 \leq j \leq n_N,\, j \neq n_k} X_j,$$

so that

1) $Y_N + Z_N = \frac{1}{\sqrt{n_N}}\,(X_1 + X_2 + \cdots + X_{n_N})$,
2) $\mathbf{E}\,Y_N^2 = \frac{N}{n_N} \to p, \mathbf{E}\,Z_N^2 = 1 - \frac{N}{n_N} \to 1 - p$,
3) $\mathbf{E}\,Y_N Z_N = 0$.

By the assumption $S_N \Rightarrow \xi$,

$$Y_N = \sqrt{\frac{N}{n_N}}\, S_N \Rightarrow \sqrt{p}\,\xi. \tag{7.3}$$

Note that $\varepsilon_N \equiv Y_N + Z_N \to 0$ in probability, by (7.1). Hence, given a positive parameter b, by (7.3),

$$\xi_N \equiv Y_N - bZ_N = (1+b)Y_N - b\varepsilon_N \Rightarrow (1+b)\sqrt{p}\,\xi.$$

Therefore, by Fatou's lemma, and using the properties 2)-3),

$$\begin{aligned} \mathbf{E}\big[(1+b)\sqrt{p}\,\xi\,\big]^2 &\leq \liminf_{N\to\infty}\ \mathbf{E}\,\xi_N^2 \\ &= \liminf_{N\to\infty}\ \Big[\mathbf{E}\,Y_N^2 + b^2\mathbf{E}\,Z_N^2\Big] \ = \ p + b^2(1-p). \end{aligned}$$

This yields

$$\mathbf{E}\,\xi^2 \leq \frac{p + b^2(1-p)}{p(1+b)^2}.$$

If $p < 1$, the right hand side is minimized for $b = \frac{p}{1-p}$ and is equal to $1-p$ at this value. In the case $p = 1$, let $b \to +\infty$.

Proposition 7.1 follows.

Now let us turn to the problem on the typical distributions in the randomized model of bounded gaps for the cosine trigonometric system

$$X_n(\omega) = \cos(2\pi n\omega)$$

on $\Omega = (0, 1)$ with uniform measure $\mathbf{P}$. For a fixed natural number d, we consider the canonical partition

$$\Delta_k = \{n : d\,(k-1) + 1 \leq n \leq dk\}, \quad k \geq 1.$$

Thus, $|\Delta_k| = d$, for all k. To study an asymptotic behaviour of sums

$$S_N = \frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}},$$

when the indices n_k are selected independently and uniformly from Δ_k, one may apply Proposition 5.1 together with the concentration argument used in the proof of Theorem 1.1. The latter reduces our task to the study of the average distributions, i.e., we may assume that (n_k) is a sequence of independent random indices, independent of the sequence X_n and with values in Δ_k according to the uniform distribution. In this case, the situation is considerably simplified by noting that

$$\frac{U_1 + \cdots + U_N}{\sqrt{N}} = \frac{X_1 + \cdots + X_{dN}}{d\sqrt{N}} \to 0,$$

for all $\omega \in (0, 1)$, since the sums $X_1 + \cdots + X_{dN}$ remain bounded, while N is increasing. In addition,

$$\frac{V_1 + \cdots + V_N}{N} \longrightarrow \frac{1}{2}$$

in probability, as was explained before. Therefore, by Proposition 5.1, the distributions of S_N and of the random variables

$$\left(\frac{1}{2} - \frac{U_1^2 + \cdots + U_N^2}{N}\right)^{1/2} Z$$

approach each other in the sense of the weak convergence. Now, for each $k \leq N$,

$$U_k^2 = \frac{1}{d^2} \sum_{n,m \in \Delta_k} \frac{X_{n+m} + X_{|n-m|}}{2},$$

where by convention, $X_0 = 1$. By direct evaluation, it should be clear that the sums $\sum_{k=1}^N \sum_{n,m\in\Delta_k} X_{n+m}$ remain bounded when N is increasing. On the other hand, $\sum_{n,m\in\Delta_k} X_{|n-m|}$ does not depend on k. Hence, S_N have in the limit the distribution of

$$S(d) = \frac{1}{\sqrt{2}} \left(1 - \frac{1}{d^2} \sum_{n,m=1}^{d} X_{|n-m|}\right)^{1/2} Z = \frac{1}{\sqrt{2}} \left(1 - \frac{1}{d} - \frac{2}{d^2} \sum_{n=1}^{d-1} (d-n)\, X_n\right)^{1/2} Z.$$

We may summarize:

Theorem 7.2. *For almost all indices n_k selected independently and uniformly from Δ_k,*

$$\frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N/2}} \Rightarrow N(0, \rho^2)$$

weakly in distribution with respect to the Lebesgue measure on $(0, 1)$*, where*

$$\rho = \left(1 - \frac{1}{d} - \frac{2}{d^2} \sum_{n=1}^{d-1} (d-n)\, X_n\right)^{1/2}. \tag{7.4}$$

Note that, although $\mathbf{E}\,(\sqrt{2}\,S_N)^2 = 1$, for the limit distribution the second moment is $\mathbf{E}\,(\sqrt{2}\,S(d))^2 = 1 - \frac{1}{d}$. This is consistent with Proposition 7.1 and actually shows optimality of (7.2), since $\mathrm{den}(\tau) = \frac{1}{d}$ with μ_∞-probability one.

If $d = 2$, formula (7.4) is simplified, and we get $\sqrt{2}\,S(2) = \sqrt{\frac{1-X_1}{2}}\,Z = N(0, \rho^2)$, where ρ has the arcsine distribution. Thus, we arrive at the conclusion made in Theorem 1.2. In this particular case, there is an alternative argument leading to Theorem 7.2, which we describe below.

The process of choosing indices n_k from $\Delta_k = \{2k-1, 2k\}$ may be connected with the Bernoullian scheme by noting that

$$X_{n_k} = \frac{1-\varepsilon_k}{2}\, X_{2k-1} + \frac{1+\varepsilon_k}{2}\, X_{2k},$$

where $\varepsilon_k = \pm 1$ are independent random variables taking the two values with probability $\frac{1}{2}$ (independently of the sequence X_n). Thus, the value $\varepsilon_k = -1$ corresponds to the choice $n_k = 2k-1$, while $\varepsilon_k = 1$ corresponds to $n_k = 2k$. Hence,

$$\sqrt{2}\,S_N = \frac{X_1 + \cdots + X_{2N}}{\sqrt{2N}} + \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \varepsilon_k\, \frac{X_{2k} - X_{2k-1}}{\sqrt{2}}. \tag{7.5}$$

As N is increasing to infinity, the first summand on the right of (7.5) will not contribute in the limit, by property (7.1). As for the second summand, one may apply the following theorem: Given an orthogonal sequence Y_n in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ such that

1) $\mathbf{E}\,Y_N^2 = \sigma^2$;
1) $\max\{|Y_1|, \dots, |Y_n|\} = o(\sqrt{n})$ in probability;
2) $\frac{Y_1^2 + \cdots + Y_n^2}{n} \to \rho^2$ weakly in distribution for some random $\rho \geq 0$,

we have that, for almost all choices of ε_k's, weakly in distribution

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_k Y_k \Rightarrow N(0, \rho^2). \tag{7.6}$$

This result has been proved in [B1] when $\rho = 1$, but the proof easily extends to the general case. Related results about convergence in probability with respect to ε_k's were obtained by H. von Weiszäcker [W] (see also the original work by V. N.

Sudakov [S]).The particular case $Y_n(\omega) = \cos(2\pi n\omega)$ in (7.6) corresponds to a result of R. Salem and A. Zygmund [S-Z2], who considered a more general scheme allowing weights. In our case

$$\begin{aligned} Y_k(\omega) &= \frac{X_{2k}(\omega) - X_{2k-1}(\omega)}{\sqrt{2}} \\ &= \frac{\cos(2\pi\, 2k\,\omega) - \cos(2\pi\,(2k-1)\,\omega)}{\sqrt{2}} \\ &= -\sqrt{2}\,\sin(\pi\omega)\,\sin((4k-1)\,\pi\omega), \end{aligned}$$

so these random variables are orthogonal, satisfy 1) with $\sigma^2 = \frac{1}{2}$, satisfy 2) since they are bounded, and also a.s.

$$\frac{Y_1^2 + \cdots + Y_N^2}{N} = 2\,\sin^2(\pi\omega)\,\frac{\sin^2(\pi\omega) + \cdots + \sin^2((4N-1)\,\pi\omega)}{N} \longrightarrow \sin^2(\pi\omega).$$

Hence, 3) is fulfilled with $\rho(\omega) = |\sin(\pi\omega)|$. This random variable takes values in [0,1] with respective distribution function and density

$$F(x) = \frac{2}{\pi}\,\arcsin(x), \quad p(x) = \frac{2}{\pi}\,\frac{1}{\sqrt{1-x^2}} \quad (0 < x < 1).$$

That is, ρ has the arcsin distribution.

8. Pairwise independent random variables

In this section we restrict ourselves to an important family of systems of random variables that are pairwise independent. It is well known, that the usual central limit theorem fails to hold in this situation. See, e. g., S. Janson [Ja], R. C. Bradley [Br], A. R. Pruss [P]; another counter-example is discussed below. It is therefore interesting to know whether weaker forms of the CLT are fulfilled. Theorem 1.1 immediately yields:

Theorem 8.1. *Let $(X_n)_{n=1}^{\infty}$ be pairwise independent random variables such that $\mathbf{E}X_n = 0$, $\mathbf{E}X_n^2 = 1$, and $\sup_n \operatorname{ess\,sup} |X_n| < +\infty$. If $|\Delta_k| \to \infty$, then for almost all indices n_k selected independently and uniformly from Δ_k,*

$$\frac{X_{n_1} + \cdots + X_{n_N}}{\sqrt{N}} \Rightarrow N(0,1). \tag{8.1}$$

Indeed, here we are dealing with an orthonormal system in $L^2(\Omega, \mathcal{F}, \mathbf{P})$, satisfying the condition $a)$ of Theorem 1.1. By pairwise independence and uniform boundedness of X_n, the random variables

$$\rho_N^2 = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{|\Delta_k|}\sum_{n\in\Delta_k} X_n^2$$

have variances of order at most $O(\frac{1}{N})$, so $\rho_N^2 \to 1$ in probability. Thus, Theorem 1.1 applies, and we arrive at the desired conclusion (8.1).

For illustration, let us start with a real-valued, 1-periodic measurable function f on the real line such that

$$\int_0^1 f(x)\,dx = 0, \qquad \int_0^1 f(x)^2\,dx = 1, \tag{8.2}$$

and consider a special system

$$X_n(t,s) = f(nt+s), \quad 0 < t, s < 1, \tag{8.3}$$

defined on the square $\Omega = (0,1) \times (0,1)$, which we equip with Borel σ-algebra $\mathcal{F}$ and Lebesgue measure $\mathbf{P}$. This system is related to the well studied case $\{f(nt)\}$, but the additional "mixing" argument s adds a number of remarkable properties. This will allow us to get rid of heavy restrictions such as "smoothness" usually required for the function f. In particular, we have:

Corollary 8.2. *If* $|\Delta_k| \to \infty$ *and* f *is essentially bounded, then the CLT* (8.1) *holds true for the system* (8.3) *with respect to Lebesgue measure* $\mathbf{P}$.

The crucial observation is contained in the following elementary statement.

Lemma 8.3. *Let* ζ *and* z *be independent, complex-valued random variables uniformly distributed on the unit circle* S^1 *of the complex plane. Then*

$$\xi_n = \zeta z^n, \qquad n = 1, 2 \dots,$$

represents a strictly stationary sequence of pairwise independent random variables.

At the same time, it is almost a deterministic sequence in the sense that $\xi_n = g_n(\xi_1, \xi_2)$, for certain measurable functions g_n on $S^1 \times S^1$. Moreover, there is relation $\xi_n = g_{i,j,n}(\xi_i, \xi_j)$, $i \neq j$, as long as $\frac{n-j}{i-j}$ is an integer. Systems of pairwise independent random variables with similar deterministic properties were first constructed by A. Joffe, cf. [Jo].

Let us also note that the sums $S_n = \xi_1 + \dots + \xi_n$ remain bounded for growing n, so the normalized sums $S_n/\sqrt{n}$, as well as their real parts $\mathrm{Re}\, S_n/\sqrt{n}$, fail to satisfy the central limit theorem. One may obtain other counter-examples in the form $f(\xi_n)$.

Proof of Lemma 8.3. By independence of ζ and z, we have for all integers $m_1, \dots, m_N$ and $h \geq 0$, $N \geq 1$,

$$\mathbf{E}\, \xi_{1+h}^{m_1} \cdots \xi_{N+h}^{m_N} = \mathbf{E}\, \zeta^{m_1 + \dots + m_N}\, \mathbf{E}\, z^{(1+h)m_1 + \dots + (N+h)m_N}. \tag{8.4}$$

Since

$$\mathbf{E}\, \zeta^m = \begin{cases} 1, & \text{if } m = 0, \\ 0, & \text{if } m \neq 0, \end{cases} \qquad \text{for all } m \in \mathbf{Z},$$

the right hand side of (8.4) is either equal to 0 or 1, and is equal to 1 if and only if $m_1 + \dots + m_N = 0$ and $(1+h)m_1 + \dots + (N+h)m_N = 0$, that is, if and only if

$m_1 + \cdots + m_N = 0$ and $1 \cdot m_1 + \cdots + N \cdot m_N = 0$. The latter description does not depend on h, which proves strict stationarity by applying Weierstrass' density theorem. Similarly, given $n > m \geq 0$ integers, and $a, b \in \mathbf{Z}$, one easily verifies that $\mathbf{E}\, \xi_n^a \xi_m^b = \mathbf{E}\, \xi_n^a \, \mathbf{E}\, \xi_m^b$, which is an equivalent form of independence.

One may equivalently formulate Lemma 8.3 as the property that

$$\eta_n(t, s) = nt + s \pmod 1$$

is a strictly stationary sequence of pairwise independent random variables on Ω. So is the sequence of the form $X_n = f(\eta_n)$. Since all η_n are uniformly distributed in the interval (0,1), the assumptions of Theorem 8.1 are fulfilled by (8.2).

The uniform boundedness assumption in Theorem 8.1 may be weakened to the uniform integrability at the expense of a certain condition posed on the growth of $|\Delta_k|$, with a rate depending upon integrability properties of X_n.

Theorem 8.4. *Let $(X_n)_{n=1}^{\infty}$ be pairwise independent random variables satisfying $\mathbf{E}X_n = 0$, $\mathbf{E}X_n^2 = 1$, and $\sup_n \mathbf{E}\, \Psi(|X_n|^2) < +\infty$ for some Young function Ψ. Then the CLT* (8.1) *holds provided that $|\Delta_k| \to \infty$ and, for any $c > 0$,*

$$|\Delta_1| + \cdots + |\Delta_N| = o(\Psi(cN)), \quad \text{as } N \to \infty, \tag{8.5}$$

Recall that a Young function represents a non-negative, increasing and convex function defined on the positive half-axis $t \geq 0$ such that $\Psi(0) = 0$ and $\frac{\Psi(t)}{t} \to +\infty$, as $t \to +\infty$. When X_n are equidistributed, the existence of Ψ follows from the finiteness of the second moment.

According to Remark 6.1, the condition *a*) of Theorem 1.1 is fulfilled, provided that (6.3) holds which is exactly the assumption (8.5). The condition *b*) is fulfilled in a stronger form, as a weak law of large numbers may be applied to the sequence $Y_n = X_n^2 - 1$.

Proposition 8.5. *Let $Y_1, \dots, Y_n$ be pairwise independent random variables satisfying $\mathbf{E}Y_i = 0$ and $\mathbf{E}\, \Psi(|Y_i|) \leq 1$, for all $i \leq n$. Then*

$$\mathbf{E} \left| \frac{Y_1 + \cdots + Y_n}{n} \right| \leq \varepsilon_\Psi(n) \tag{8.6}$$

with $\varepsilon_\Psi(n) \to 0$, as $n \to \infty$, depending on the Young function Ψ, only.

In particular, if we start with an infinite sequence Y_n under the same hypotheses, then in probability

$$\frac{Y_1 + \cdots + Y_n}{n} \longrightarrow 0, \qquad \text{as } n \to \infty.$$

This weak law of large numbers is due to D. Landers and L. Rogge [L-R]. Here the requirement that $\frac{\Psi(t)}{t} \to +\infty$ cannot be omitted even for totally independent random variables. For example, if $\mathbf{P}\{Y_n = \pm n\} = \frac{1}{2n}$, $\mathbf{P}\{Y_n = 0\} = 1 - \frac{1}{n}$, we have $\mathbf{E}Y_n = 0$, $\mathbf{E}|Y_n| = 1$, but the weak law does not hold. It may also be verified that $\mathbf{E} \left| \frac{Y_1 + \cdots + Y_n}{n} \right| \geq c$ for some absolute constant $c > 0$.

The refining inequality (8.6) is needed in Theorem 8.4 in view of the structure of the random variables ρ_N^2. Letting $C = \sup_n \mathbf{E}\Psi(X_n^2)$, $C_0 = 1 + \frac{1}{\Psi^{-1}(C)}$, $\Psi_0(t) = \frac{1}{C}\Psi(t)$, we get $\mathbf{E}\Psi_0(\frac{X_n^2-1}{C_0}) \leq 1$ and thus may apply Proposition 8.5 to the finite collection $Y_n = \frac{X_n^2-1}{C_0}$, $n \in \Delta_k$, with fixed $k = 1, \dots, N$. By (8.6), every $V_k = \frac{1}{|\Delta_k|}\sum_{n\in\Delta_k} X_n^2$ satisfies

$$\mathbf{E}\,|V_k - 1| \,\leq\, C_0\,\varepsilon_{\Psi_0}(|\Delta_k|).$$

Therefore, by the triangle inequality,

$$\mathbf{E}\,|\rho_N^2 - 1| \,\leq\, \frac{C_0}{N}\sum_{k=1}^N \varepsilon_{\Psi_0}(|\Delta_k|) \longrightarrow 0, \quad \text{as } N \to \infty,$$

since $|\Delta_k| \to \infty$. Hence $\rho_N^2 \to 1$ in probability, and thus the proof of Theorem 8.4 would be completed. To prove Proposition 8.5, we need:

Lemma 8.6. *Let Y be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ satisfying $\mathbf{E}\Psi(|Y|) \leq 1$, where Ψ is a Young function such that $\Psi(1) = 1$. Then*

a) $\mathbf{E}\,|Y|\,1_A \,\leq\, p\,\Psi^{-1}(\frac{1}{p})$, *for any $A \in \mathcal{F}$ with $\mathbf{P}(A) \leq p$;*
b) $\mathbf{E}\,|Y|\,1_{\{|Y|\geq c\}} \,\leq\, \frac{c}{\Psi(c)}$, *for any $c \geq 1$.*

As usual, Ψ^{-1} denotes the inverse function and 1_A the indicator function of a set A. Note that $p\,\Psi^{-1}(\frac{1}{p})$ is increasing in p.

Proof. The inequality in *b*) is obtained from *a*) by using Chebyshev's inequality $\mathbf{P}(A) \leq \frac{1}{\Psi(c)}$, where $A = \{|Y| \geq c\}$. For the first assertion, we may assume $0 < p \leq 1$, that Ω is finite and that Ψ has a continuous increasing derivative in $t > 0$. So, $R = \Psi^{-1}$ has a continuous decreasing derivative R' on the positive half-axis.

Fix $A \in \mathcal{F}$. Making the substitution $u = \Psi(|Y|)$, we are reduced to maximizing the functional $L(u) = \mathbf{E}\,R(u)\,1_A$ under the constraints $u \geq 0$, $\mathbf{E}u = 1$. Clearly, a maximizer u_0 exists and is vanishing outside A. Let $B = \{u_0 > 0\}$, so that $B \subset A$, and put $u_\varepsilon = u_0 + \varepsilon v$ with a real ε and an arbitrary function v on Ω vanishing outside B and such that $\mathbf{E}v = 0$. Then $\mathbf{E}u_\varepsilon = 1$ and $u_\varepsilon > 0$ on B, for all ε small enough. In addition, by Taylor's expansion,

$$L(u_\varepsilon) = L(u_0) + \varepsilon\,\mathbf{E}\,R'(u_0)v + o(\varepsilon), \quad \text{as } \varepsilon \to 0,$$

implying that $\mathbf{E}\,R'(u_0)v = 0$ for all v as above. Therefore, $R'(u_0)$ and thus u_0 itself must be constant on B. But when $u_0 = b1_B$ with $b\,\mathbf{P}(B) = \mathbf{E}u_0 = 1$, we have that $L(u_0) = R(b)\mathbf{P}(B) = \mathbf{P}(B)R(1/\mathbf{P}(B))$. Since the latter quantity is increasing in $\mathbf{P}(B)$ and $\mathbf{P}(B) \leq p$. it does not exceed $p\,R(1/p)$.

Proof of Proposition 8.5. We follow a standard argument as in [L-R], where a suitable double truncation procedure is applied. Consider $Y_i' = Y_i\,1_{\{|Y_i|<n\}}$ and put $s_n = \frac{Y_1+\dots+Y_n}{n}$, $s_n' = \frac{Y_1'+\dots+Y_n'}{n}$.

For a random variable ξ, we use the notation $\|\xi\|_\Psi = \inf\{\lambda > 0 : \Psi(|\xi|/\lambda) \le 1\}$ for the Orlicz norm of ξ generated by the Young function Ψ. First assume $\Psi(1) = 1$. Since, by the assumption, $\|Y_i\|_\Psi \le 1$, we also have $\|Y_i'\|_\Psi \le 1$, so $\|s_n\|_\Psi \le 1$ and $\|s_n'\|_\Psi \le 1$. Splitting $s_n = (s_n - s_n') + (s_n' - \mathbf{E}s_n') + \mathbf{E}s_n'$, we get

$$\mathbf{E}|s_n| \le \mathbf{E}|s_n - s_n'| + \mathbf{E}|s_n' - \mathbf{E}s_n'| + |\mathbf{E}s_n'|. \tag{8.7}$$

Note $\mathbf{E}s_n' = -\frac{1}{n}\sum_{i=1}^n Y_i\, 1_{\{|Y_i|\ge n\}}$, so by Lemma 8.4 b) applied to each Y_i,

$$|\mathbf{E}s_n'| \le \frac{n}{\Psi(n)}. \tag{8.8}$$

Now, by Chebyshev's inequality, for the set $A = \cup_{i=1}^n \{|Y_i| \ge n\}$, we have $\mathbf{P}(A) \le \frac{n}{\Psi(n)}$. Note that $s_n = s_n'$ on the complement of A. Hence, by Lemma 8.4 a) applied to $Y = \frac{s_n - s_n'}{2}$ with $p = \frac{n}{\Psi(n)}$, we get

$$\mathbf{E}\,|s_n - s_n'| \,=\, 2\,\mathbf{E}\,\frac{|s_n - s_n'|}{2}\,1_A \,\le\, \frac{2n}{\Psi(n)}\,\Psi^{-1}\Big(\frac{\Psi(n)}{n}\Big). \tag{8.9}$$

To bound the middle term on the right of (8.7), use pairwise independence of Y_i' to write

$$\begin{aligned}\mathbf{E}|s_n' - \mathbf{E}s_n'|^2 &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(Y_i') \,\le\, \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\,Y_i^2\,1_{\{|Y_i|<n\}} \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\,Y_i^2\,1_{\{|Y_i|<c\}} + \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\,Y_i^2\,1_{\{c\le|Y_i|<n\}},\end{aligned}$$

where $1 \le c \le n$. Next, $\mathbf{E}\,Y_i^2\,1_{\{|Y_i|<c\}} \le c\,\mathbf{E}|Y_i| \le c$, since, by Jensen's inequality, $\mathbf{E}|Y_i| \le \Psi^{-1}(\mathbf{E}\Psi(|Y_i|)) \le \Psi^{-1}(1) = 1$. Similarly,

$$\mathbf{E}\,Y_i^2\,1_{\{c\le|Y_i|<n\}} \le n\,\mathbf{E}\,|Y_i|\,1_{\{|Y_i|\ge c\}} \le n\,\frac{c}{\Psi(c)},$$

by Lemma 8.6 b). Hence, $\mathbf{E}|s_n' - \mathbf{E}s_n'|^2 \le \frac{c}{n} + \frac{c}{\Psi(c)}$. Together with (8.7)–(8.9), and using $\Psi(n) \ge n$, this yields

$$\mathbf{E}\,|s_n| \,\le\, \frac{3n}{\Psi(n)}\,\Psi^{-1}\Big(\frac{\Psi(n)}{n}\Big) + \inf_{1\le c\le n}\sqrt{\frac{c}{n} + \frac{c}{\Psi(c)}}. \tag{8.10}$$

The quantity on the right of (8.10) may be taken as $\varepsilon_\Psi(n)$, since the infimum is convergent to zero, as $n \to \infty$ (for $c = \sqrt{n}$, for example).

Finally, if $\Psi(1) \ne 1$, introduce a new Young function $\Psi_0(t) = \frac{1}{\Psi(1)}\Psi(t)$. Then, with $C = \max\{1, \frac{1}{\Psi(1)}\}$, we have $\mathbf{E}\Psi_0(\frac{|Y_i|}{C}) \le 1$ and, by the previous step applied to the sequence $\frac{Y_i}{C}$, we obtain that $\mathbf{E}|s_n| \le C\varepsilon_{\Psi_0}(n)$. Proposition 8.5 is proved.

Returning to the systems associated with periodic functions, we obtain from Theorem 8.4 the following analogue of Corollary 8.2.

Corollary 8.7. *For any* 1*-periodic measurable function* f*, normalized in accordance with* (8.2)*, the CLT* (8.1) *holds for* $X_n = f(nt + s)$ *with respect to the Lebesgue measure on the square* $0 < t, s < 1$*, as long as* $|\Delta_1| + \cdots + |\Delta_N|$ *grow to infinity sufficiently slowly.*

For example, as we know from (8.5), if

$$\int_0^1 |f(x)|^2 \log |f(x)| \, dx < +\infty,$$

the condition on the growth rate is given by $|\Delta_1| + \cdots + |\Delta_N| = o(N \log N)$. It is, however, not clear whether this condition is necessary.

References

[A] Aldous, D.J.: Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. Z. Wahrscheinlichkeitstheory Verw. Geb. **40**, 59–82 (1977)

[Ber1] Berkes, I.: On the central limit theorem for lacunary trigonometric series. Analysis Math. **4**, 159–180 (1978)

[Ber2] Berkes, I.: A central limit theorem for trigonometric series with small gaps. Z. Wahrscheinlichkeitstheory Verw. Geb. **47** (2), 151–161 (1979)

[B-P] Berkes, I., Péter, E.: Exchangeable random variables and the subsequence principle. Probab. Theory Rel. Fields, **73**, 395–414 (1986)

[B1] Bobkov, S.G.: Concentration of distributions of the weighted sums with Bernoullian coefficients. Geom. Aspects of Funct. Anal., Lecture Notes in Math. **1807**, 27–36 (2003)

[B2] Bobkov, S.G.: Concentration of normalized sums and a central limit theorem for non-correlated random variables. Ann. Probab. **32** (4), 2884–2907 (2004)

[B-G1] Bobkov, S.G., Götze, F.: Discrete isoperimetric and Poincaré-type inequalities. Probab. Theory Rel. Fields **114** (2), 245–277 (1999)

[B-G2] Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163** (1), 1–28 (1999)

[B-G3] Bobkov, S.G., Götze, F.: On the central limit theorem along subsequences of non-correlated observations. Probab. Theory Appl. **48** (4), 604–621 (2003)

[Bo] Bohman, H.: Approximate Fourier analysis of distribution functions. Arkiv Mat. **4** (2–3), 99–157 (1961)

[Br] Bradley, R.C.: A stationary, pairwise independent, absolutely regular sequence for which the central limit theorem fails. Probab. Theory Appl. **81** (1), 1–10 (1989)

[C] Chatterji, S.D.: A principle of subsequences in probability theory: the central limit theorem. Adv. in Math. **13**, 31–54 (1974)

[E] Erdös, P.: On trigonometric sums with gaps. Magyar Tud. Acad. Mat. Kutató Int. Kózl. **7**, 37–42 (1962)

[Fo] Fortet, R.: Sur une suite egalement répartie. Studia. Math. **9**, 54–69 (1940)

[Fu] Fukuyama, K.: Almost sure invariance principles for lacunary trigonometric series. Compt. Ren. de l'Acad. Sci. Ser. 1, Math. **332** (8), 685–690 (2001)

[G1] Gaposhkin, V.F.: Sequences of functions and the central limit theorem. Matem. Sbornik **70**, 145–171 (1966)

[G2] Gaposhkin, V.F.: Lacunary series and independent functions. Russian Math. Surveys **21** (6), 1–82 (1966)

[G3] Gaposhkin, V.F.: On the central limit theorem for subsequences of random variables. Probab. Theory Appl. **12** (4), 768–770 (1967)

[G4] Gaposhkin, V.F.: Convergence and limit theorems for subsequences of random variables. Probab. Theory Appl. **17** (3), 401–423 (1972)

[Gr] Gross, L.: Logarithmic Sobolev inequalities. Amer. J. Math. **97**, 1061–1083 (1975)

[Ja] Janson, S.: Some pairwise independent sequences for which the central limit theorem fails. Stochastics **23**, 439–448 (1988)

[Jo] Joffe, A. On a sequence of almost deterministic pairwise independent random variables. Proc. Amer. Math. Soc. **29** (2), 381–382 (1971)

[K1] Kac, M.: Sur les fonctions indépendantes V. Studia Math. **7**, 96–100 (1938)

[K2] Kac, M.: On the distribution of values of sums of the type $\sum f(2^k t)$. Ann. Math. 2nd Ser. **47** (1), 33–49 (1946)

[L-R] Landers, D., Rogge, L.: Laws of large numbers for pairwise independent uniformly integrable random variables. Math. Nachr. **130**, 189–192 (1987)

[L1] Ledoux, M.: On Talagrand's deviation inequalities. ESAIM: Probability and Statistics, **1**, 62–87 (1996)

[L2] Ledoux, M.: Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités, XXXIII. Lecture Notes in Math. **1799**, 120–216 (1999)

[Mo] Morgenthaler, G.W.: A central limit theorem for uniformly bounded orthonormal systems. Trans. Amer. Math. Soc. **79** (2), 281–311 (1955)

[Mu] Murai, T.: The central limit theorem for trigonometric series. Nagoya Math. J. **87**, 79–94 (1982)

[P] Pruss, A.R.: A bounded N-tuplewise independent and identically distributed counterexample to the CLT. Probab. Theory Rel. Fields. **111**, 323–332 (1998)

[S-Z] Salem, R., Zygmund, A.: On lacunary trigonometric systems. Proc. Nat. Acad. Sci. USA **33**, 333–338 (1947)

[S-Z2] Salem, R., Zygmund, A.: Trigonometric series whose terms have random sign. Acta Math. **91**, 245–301 (1954)

[Su] Sudakov, V.N.: Typical distributions of linear functionals in finite-dimensional spaces of higher dimension. Soviet Math. Dokl. **19** (6), 1578–1582 (1978); Translated from: Dokl. Akad. Nauk SSSR **243** (6) (1978)

[T] Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'I.H.E.S. **81** (9), 73–205 (1995)

[W] von Weizsäcker, H.: Sudakov's typical marginals, random linear functionals and a conditional central limit theorem. Probab. Theory Rel. Fields **107**, 313–324 (1997)