

## ON THE CENTRAL LIMIT THEOREM ALONG SUBSEQUENCES OF NONCORRELATED OBSERVATIONS\*

S. G. BOBKOV<sup>†</sup> AND F. GÖTZE<sup>‡</sup>

**Abstract.** We study the asymptotic behavior of distributions of normalized and self-normalized sums along suitable subsequences of noncorrelated random variables.

**Key words.** concentration, typical distributions, central limit theorem, self-normalized statistics, orthogonal polynomials

**DOI.** 10.1137/S0040585X97980737

**1. Introduction.** Let  $\{X_n\}_{n=1}^\infty$  be a sequence of random variables on a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  with finite second moments and such that, for all  $k, j$ ,

$$(1.1) \quad \mathbf{E}X_k X_j = \delta_{kj},$$

where  $\delta_{kj}$  denotes the Kronecker symbol.

In this paper, we study the question of whether or not the sequence  $\{X_n\}_{n=1}^\infty$  contains a subsequence, say  $\{X_{i_n}\}_{n=1}^\infty$ , which behaves like a sequence of independent random variables with respect to some intensively studied functionals. As main examples, we consider normalized sums

$$S_n = \frac{X_{i_1} + \cdots + X_{i_n}}{\sqrt{n}}$$

and self-normalized statistics

$$T_n = \frac{X_{i_1} + \cdots + X_{i_n}}{\sqrt{X_{i_1}^2 + \cdots + X_{i_n}^2}}.$$

To avoid degenerate situations, we assume that in probability, as  $n \rightarrow \infty$ ,

$$(1.2) \quad \max_{1 \leq k \leq n} |X_k| = o(\sqrt{n}).$$

This condition is weaker than the uniform integrability of the sequence  $X_n^2$  and is implied by usual moment assumptions such as  $\sup_{n \geq 1} \mathbf{E}X_n^2 \log |X_n| < +\infty$ , for example.

**THEOREM 1.1.** *Assume that in probability as  $n \rightarrow \infty$ ,*

$$(1.3) \quad \frac{X_1^2 + \cdots + X_n^2}{n} \rightarrow 1.$$

*Then, under the hypotheses (1.1)–(1.2), there exists an increasing sequence of indices  $\{i_n\}_{n \geq 1}$  such that weakly in distribution*

$$S_n \Rightarrow N(0, 1).$$

---

\*Received by the editors July 10, 2003. This work was supported by the Alexander von Humboldt Foundation and NSF grant DMS-0103929.

<http://www.siam.org/journals/tvp/48-4/98073.html>

<sup>†</sup>School of Mathematics, University of Minnesota, 127 Vincent Hall, Church St. S. E., Minneapolis, MN 55455 (bobkov@math.umn.edu).

<sup>‡</sup>Department of Mathematics, Bielefeld University, Bielefeld 33501, Germany (goetze@mathematik.uni-bielefeld.de).

As we will see, the desired sequence  $\{i_n\}_{n \geq 1}$  can be chosen to be sufficiently tight in a sense that, for any prescribed sequence  $\{j_n\}_{n=1}^\infty$  such that  $j_n/n \rightarrow +\infty$ ,  $\{i_n\}_{n \geq 1}$  is dominated by  $\{j_n\}_{n \geq 1}$ , i.e.,

$$(1.4) \quad i_n \leq j_n \quad \text{for all } n \text{ large enough.}$$

Note that no assumption is made about the expectations  $\mathbf{E}X_n$ .

Theorem 1.1 may be viewed as a certain generalization of the central limit theorem for independent and identically distributed random variables with mean zero and a finite second moment. In that case, all the conditions are fulfilled, while the conclusion of Theorem 1.1 does not depend on the choice of subsequences (by exchangeability).

The meaning of the hypothesis (1.2) is better clarified in the triangular arrays scheme or in the scheme of finite sequences  $X_1, \dots, X_N$ . Consider, for example, normalized indicator functions

$$X_n = \sqrt{N} \mathbf{1}_{A_n}, \quad 1 \leq n \leq N,$$

where the  $A_n$ 's form a partition of  $\Omega$  into  $N$  measurable subsets of  $\mathbf{P}$ -measure  $1/N$ . Then, (1.1) and (1.3) are satisfied, but, for any choice of  $1 \leq i_1 < \dots < i_n \leq N$ , the random variable  $S_n$  takes only two values and therefore cannot be approximated by the standard normal law.

While condition (1.3) is important for asymptotic normality of  $S_n$ , it can be considerably weakened in the study of self-normalized statistics.

**THEOREM 1.2.** *Assume that*

$$(1.5) \quad \limsup_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{X_1^2 + \dots + X_n^2}{n} \leq h \right\} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

*Then, under hypotheses (1.1), (1.2), there exists an increasing sequence of indices  $\{i_n\}_{n \geq 1}$  such that weakly in distribution as  $n \rightarrow \infty$ ,*

$$T_n \Rightarrow N(0, 1).$$

Assumption (1.5) means that the distributions of the quadratic forms  $R_n = (X_1^2 + \dots + X_n^2)/n$  cannot approximate those distributions which have atoms at zero. This ensures, in particular, that the random variables  $T_n$  are well defined for large  $n$  with probability almost 1 (for proper subsequences).

Actually, a statement similar to (1.4) holds in this case, as well: in both statements the sequences of indices can be chosen almost arbitrarily. Moreover, they form a subset  $I$  in the space  $\mathcal{I}$  of all increasing sequences (of natural numbers) with  $\mu(I) = 1$  for certain probability measures  $\mu$  on  $\mathcal{I}$ .

To make this argument precise, recall that every increasing sequence of natural numbers may be canonically associated to a 0–1 sequence  $\varepsilon = \{\varepsilon_n\}_{n \geq 1}$ . Hence, we may equivalently consider the problem on the functionals

$$S_n(\varepsilon) = \frac{\varepsilon_1 X_1 + \dots + \varepsilon_n X_n}{\sqrt{\varepsilon_1 + \dots + \varepsilon_n}} \quad \text{and} \quad T_n(\varepsilon) = \frac{\varepsilon_1 X_1 + \dots + \varepsilon_n X_n}{\sqrt{\varepsilon_1 X_1^2 + \dots + \varepsilon_n X_n^2}}.$$

It turns out that assuming (1.1) only, the distributions of  $S_n(\varepsilon)$  and  $T_n(\varepsilon)$  are strongly concentrated around certain “typical distributions” with respect to products of some Bernoullian measures. In this sense they almost do not depend on the choice of  $\varepsilon$ . The first observation of this kind was made by V. N. Sudakov [17], who considered a

related problem about the concentration of distributions of linear forms  $\sum_{k=1}^n \theta_k X_k$  with coefficients representing coordinates of a random point on the unit sphere in  $\mathbf{R}^n$ . Some extensions and refinements in this randomized model, including the rate of concentration and different approaches to this concentration phenomenon, were later studied by a number of authors; cf., e.g., [7], [19], [1], [3]. In these papers, the additional assumption (1.3) appears as a natural condition which ensures asymptotic normality of typical distributions (regardless of whether (1.2) holds). The symmetric Bernoullian case, where  $\theta_k = \varepsilon_k/\sqrt{n}$  with  $\varepsilon_k = \pm 1$ , was considered in [4]. There it is shown that, under (1.1)–(1.3), for  $\mu$ -almost all sequences of signs,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \varepsilon_k X_k \Rightarrow N(0, 1) \quad \text{as } n \rightarrow \infty,$$

where  $\mu$  is the product symmetric Bernoullian measure on the infinite discrete cube  $\{-1, 1\}^\infty$ . In contrast, such a statement for “0–1”-coefficients  $\varepsilon_k$  (with probability weights  $\frac{1}{2}$ ) is no longer valid (even with other than  $\sqrt{n}$ -normalization). Therefore we need to look for different suitable measures  $\mu$ .

As for  $T_n$ , this functional is of a more complicated nature, but it is known to be more well behaved in comparison with linear forms. In the nonrandomized model, i.e., for  $i_n = n$ , the self-normalized statistics have been of great interest already in the classical case of independent, identically distributed random variables  $X_n$ . In this case, the question of limit distributions of  $T_n$  was raised in 1973 by B. Logan et al. [12] solving the problem for stable laws  $F$  of  $X_n$ 's. After contributions [13], [9], [8], and others, a complete answer in terms of  $F$  has recently been given by G. Chistyakov and F. Götze [6]. In particular, the central limit theorem for  $T_n$  holds under weaker assumptions than that for  $S_n$  (that is, a little less than finiteness of the second moment of  $F$  is required). For other important aspects of the asymptotic normality of  $T_n$ , such as the Berry–Esseen-type bounds, see, e.g., [2], [16], [18].

On the other hand, there is no hope of reaching the asymptotic normality of  $S_n$  and  $T_n$  if we replace the independence with the orthogonality condition (1.1). One surprising possible situation was recently noticed by A. R. Pruss [15]: For any natural number  $N$ , it is possible that  $X_{i_1}, \dots, X_{i_N}$  are independent for any choice of distinct indices, and that all  $X_n$  are equidistributed according to a prescribed symmetric non-degenerate law, but the central limit theorem fails. See also [10] for other examples. As we have seen, in such “bad” situations, one can still obtain some positive results at the expense of removing certain “bad” terms in the given sequence  $X_n$ . This scheme involves many standard examples such as the sequence of trigonometric functions and various systems of orthogonal polynomials such as Chebyshev's.

In the randomized situation as above, let us note that, in order to study a typical law of  $T_n(\varepsilon)$ , there is good reason to first look at the joint distribution  $F_\varepsilon$  of both linear and quadratic forms, namely, of random variables

$$\frac{\varepsilon_1 X_1 + \dots + \varepsilon_n X_n}{\sqrt{B_n}} \quad \text{and} \quad \frac{\varepsilon_1 X_1^2 + \dots + \varepsilon_n X_n^2}{B_n}$$

with a suitably chosen normalization  $B_n$ . As we will see, under (1.1)–(1.2), for “most”  $\varepsilon$ , the distributions  $F_\varepsilon$  are very close to the distribution of the random vector

$$W_n = (R_n Z, R_n^2),$$

where  $Z$  is a standard normal random variable independent of  $R_n$ . This explains why conditions (1.3) and (1.5) appear in the formulations of Theorems 1.1 and 1.2.

This paper is (almost) self-contained and is organized as follows. In section 2 we discuss a concentration property of product measures on the finite dimensional discrete cube in terms of deviations of “Lipschitz” functions from their means. In section 3 we consider characteristic functions  $f_\varepsilon$  of  $F_\varepsilon$  and investigate the Lipschitz seminorm of functions  $\varepsilon \rightarrow f_\varepsilon(t, s)$  with fixed  $(t, s)$ . A concentration property for this family, uniform with respect to the variables  $(t, s)$ , is studied in section 4. An asymptotic behavior of typical distributions is discussed in section 5, and finally in section 6 we complete the proof of Theorems 1.1 and 1.2, including the refining assertion (1.4).

**2. Concentration on the discrete cube.** Given  $p \in (0, 1)$ , denote by  $\mu_p$  the Bernoullian measure on  $\{0, 1\}$  assigning mass  $p$  to the point 1 and  $q = 1 - p$  to the point 0. In this section we recall a Gaussian deviation inequality for functions on the discrete cube  $\{0, 1\}^n$  with respect to arbitrary product probability measures  $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$ .

With every complex-valued function  $f$  on  $\{0, 1\}^n$ , we connect the length of the discrete gradient  $|\nabla f|$  defined by

$$|\nabla f(\varepsilon)|^2 = \sum_{k=1}^n |f(\varepsilon) - f(s_k(\varepsilon))|^2, \quad \varepsilon \in \{0, 1\}^n.$$

Here,  $s_k(\varepsilon)$  denotes the neighbor of  $\varepsilon$  along the  $k$ th coordinate, i.e.,  $(s_k(\varepsilon))_j = \varepsilon_j$  for  $j \neq k$  and  $(s_k(\varepsilon))_k = 1 - \varepsilon_k$ . Set  $\|\nabla f\|_\infty = \max_\varepsilon |\nabla f(\varepsilon)|$ . This quantity may be viewed as a Lipschitz seminorm of  $f$ .

PROPOSITION 2.1. *For every complex-valued  $f$  on  $\{0, 1\}^n$  such that  $\|\nabla f\|_\infty \leq \sigma$ ,*

$$(2.1) \quad \mu \left\{ \left| f - \int f d\mu \right| \geq h \right\} \leq 4e^{-h^2/\sigma^2}, \quad h > 0.$$

This Gaussian bound is standard, although in the literature it appeared with a worse constant in the exponent (except for the case of the uniform measure). So, we include a proof of the refined estimate.

LEMMA 2.1. *For every real-valued function  $f$  on  $\{0, 1\}$ , for any  $p \in (0, 1)$ ,*

$$(2.2) \quad \text{Ent}_{\mu_p}(e^f) \leq \frac{1}{8} \int |\nabla f|^2 e^f d\mu_p.$$

Here, the entropy functional  $\text{Ent}_\mu(g)$  with respect to a probability measure  $\mu$  is defined, for any  $g \geq 0$ , by

$$\text{Ent}_\mu(g) = \int g \log g d\mu - \int g d\mu \log \int g d\mu.$$

Thus, the worst case in the modified logarithmic Sobolev inequality (2.2) corresponds to the symmetric measure, i.e., when  $p = q = \frac{1}{2}$ .

*Proof.* Instead of (2.2), consider an equivalent inequality

$$(2.3) \quad \text{Ent}_{\mu_p}(f) \leq C(p) \int |\nabla \log f|^2 f d\mu_p$$

in the class of all positive  $f$  on  $\{0, 1\}$  with some optimal  $C(p)$ . An exact value of this constant is unknown, but a simple argument shows that it is maximized at  $p = \frac{1}{2}$ .

Since (2.3) is homogeneous in  $f$ , we may assume  $\int f d\mu_p = 1$ . In terms of  $x = f(1)$ ,  $y = f(0)$ , this condition reads as

$$(2.4) \quad px + qy = 1, \quad x, y > 0,$$

and (2.3) itself simplifies to

$$px \log x + qy \log y \leq C(p) (\log x - \log y)^2.$$

We may think of  $x$  as a free variable with values in  $(0, 1/p)$ , while  $y = (1 - px)/q$  is dependent.

Introduce the function

$$u(x) = \frac{px \log x + qy \log y}{(\log x - \log y)^2}, \quad 0 < x < \frac{1}{p}, \quad y = \frac{1 - px}{q},$$

so that  $C(p) = \sup_x u(x)$ . Since  $u(0+) = u(1/p-) = 0$ , the function  $u$  attains maximum at some point  $x_0 \in (0, 1/p)$ . We have

$$u'(x) = \frac{pqxy(\log x - \log y)^2 - 2(px \log x + qy \log y)}{(\log x - \log y)^3 \cdot qy}.$$

Hence  $u'(x) = 0$  if and only if  $px \log x + qy \log y = (pqxy/2)(\log x - \log y)^2$ . Therefore,  $C(p) = u(x_0) = px_0 qy_0/2 \leq \frac{1}{8}$  by (2.4). Thus, Lemma 2.1 follows.

*Proof of Proposition 2.1.* Inequality (2.2) is extended to the product space  $\{0, 1\}^n$  equipped with product measures  $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$ . Namely, for any real-valued function  $f$  on  $\{0, 1\}^n$ ,

$$\text{Ent}_\mu(e^f) \leq \frac{1}{8} \int |\nabla f|^2 e^f d\mu.$$

In turn, given that  $\|\nabla f\|_\infty \leq \sigma$ , the above inequality easily implies the following bound on probabilities of large deviations of  $f$  (cf. [11], [5]),

$$(2.5) \quad \mu \left\{ f - \int f d\mu \geq h \right\} \leq e^{-2h^2/\sigma^2}, \quad h > 0,$$

and the following two-sided estimate, as well:

$$(2.6) \quad \mu \left\{ \left| f - \int f d\mu \right| \geq h \right\} \leq 2e^{-2h^2/\sigma^2}, \quad h > 0.$$

In the complex-valued case, the latter bound can be applied separately to  $f_1 = \text{Re}(f)$  and  $f_2 = \text{Im}(f)$ . Assume  $\int f d\mu = 0$ . Since  $|f| \geq h$  implies  $|f_1| \geq h/\sqrt{2}$  or  $|f_2| \geq h/\sqrt{2}$ , and since  $|\nabla f_j| \leq |\nabla f| \leq \sigma$ ,  $j = 1, 2$ , (2.6) yields

$$\mu\{|f| \geq h\} \leq \mu\left\{|f_1| \geq \frac{h}{\sqrt{2}}\right\} + \mu\left\{|f_2| \geq \frac{h}{\sqrt{2}}\right\} \leq 2e^{-h^2/\sigma^2} + 2e^{-h^2/\sigma^2}.$$

Thus, Proposition 2.1 is proved.

**3. Bounding the Lipschitz constant.** Consider a finite sequence  $(X_1, \dots, X_n)$  of random variables satisfying the orthonormality condition (1.1).

Given a collection of probabilities  $p_1, \dots, p_n \in (0, 1)$ , we equip the discrete cube  $\{0, 1\}^n$  with the product of Bernoullian measures  $\mu = \mu_{p_1} \otimes \dots \otimes \mu_{p_n}$ . For  $\varepsilon \in \{0, 1\}^n$ , put

$$S_n(\varepsilon) = \frac{\varepsilon_1 X_1 + \dots + \varepsilon_n X_n}{\sqrt{B_n}},$$

$$V_n(\varepsilon) = \frac{\varepsilon_1 X_1^2 + \dots + \varepsilon_n X_n^2}{B_n},$$

where here and throughout we write  $B_n = p_1 + \dots + p_n$ . The normalization is chosen canonically so that

$$(3.1) \quad \int \mathbf{E} S_n(\varepsilon)^2 d\mu(\varepsilon) = \int \mathbf{E} V_n(\varepsilon) d\mu(\varepsilon) = 1.$$

Note that the self-normalized statistics  $T_n(\varepsilon)$  may be expressed as

$$T_n(\varepsilon) = \frac{S_n(\varepsilon)}{\sqrt{V_n(\varepsilon)}}.$$

Our next goal will be to show that, under certain conditions on the growth of  $B_n$ , the joint distribution  $F_\varepsilon$  of  $S_n(\varepsilon)$  and  $V_n(\varepsilon)$  is almost independent of the choice of  $\varepsilon$  with respect to the measure  $\mu$ . In other words,  $F_\varepsilon$ 's are strongly concentrated around a certain "typical" distribution  $F$  on the plane  $\mathbf{R}^2$  and, moreover, for this typical distribution one can take the  $\mu$ -average of  $F_\varepsilon$ . Identifying distributions with their corresponding distribution functions, the average  $F$  is given by

$$F(x, y) = \int \mathbf{P} \{ S_n(\varepsilon) \leq x, V_n(\varepsilon) \leq y \} d\mu(\varepsilon), \quad x, y \in \mathbf{R}.$$

Thus,  $F$  represents the distribution of  $(S_n(\varepsilon), V_n(\varepsilon))$ , where now  $\varepsilon$  is viewed as a random vector independent of all  $X_k$ 's, with values in the discrete cube and distributed according to  $\mu$ .

To measure the distance between distribution functions of two variables, say  $F_1$  and  $F_2$ , we use the metric

$$\rho(F_1, F_2) = \sup_{t, s} \frac{|f_1(t, s) - f_2(t, s)|}{1 + |t| + |s|},$$

where  $f$ 's denote the associated characteristic functions. Thus, introduce the family of characteristic functions of  $F_\varepsilon$  with parameter  $\varepsilon \in \{0, 1\}^n$ ,

$$f_\varepsilon(t, s) = \mathbf{E} \exp \{ it S_n(\varepsilon) + is V_n(\varepsilon) \}, \quad t, s \in \mathbf{R},$$

and their  $\mu$ -mean

$$f(t, s) = \int \mathbf{E} \exp \{ it S_n(\varepsilon) + is V_n(\varepsilon) \} d\mu(\varepsilon),$$

representing the characteristic function of the average distribution  $F$ .

As a first step, we fix  $t, s$  and derive a bound on the modulus of the gradient of  $f_\varepsilon(t, s)$  viewed as a function of  $\varepsilon \in \{0, 1\}^n$ .

LEMMA 3.1. For all  $t, s \in \mathbf{R}$  and  $\varepsilon \in \{0, 1\}^n$ ,

$$(3.2) \quad |\nabla f_\varepsilon(t, s)| \leq \frac{|t|}{\sqrt{B_n}} + \frac{(|t|^2 + 2|s|)\sqrt{n}}{2B_n}.$$

COROLLARY 3.1. Given  $t, s \in \mathbf{R}$  with  $|t| + |s| > 0$ , for any  $h \in (0, 2]$ ,

$$(3.3) \quad \mu \left\{ \varepsilon : \frac{|f_\varepsilon(t, s) - f(t, s)|}{|t| + |s|} \geq h \right\} \leq 4e^{-B_n^2 h^2 / (9n)}.$$

Note that the right-hand side is small once  $B_n$  becomes bigger than  $\sqrt{n}$ .

Inequality (3.3) remains true in the limit case  $t, s \rightarrow 0$ . In particular, since for small  $t$  Taylor’s representation reads as

$$f_\varepsilon(t, 0) - 1 = it \mathbf{E}S_n(\varepsilon) + o(t) = it \frac{\varepsilon_1 \mathbf{E}X_1 + \dots + \varepsilon_n \mathbf{E}X_n}{\sqrt{B_n}} + o(t),$$

we obtain the following corollary.

COROLLARY 3.2. For any  $h \in (0, 2]$ ,

$$(3.4) \quad \mu \left\{ \varepsilon : \left| \sum_{k=1}^n (\varepsilon_k - p_k) \mathbf{E}X_k \right| \geq h\sqrt{B_n} \right\} \leq 4e^{-B_n^2 h^2 / (9n)}.$$

To show how Corollary 3.1 can be derived from Lemma 3.1 and Proposition 2.1, first let us observe that, since  $|f_\varepsilon(t, s) - f(t, s)| \leq 2$ , inequality (3.3) is immediate in the case  $|t| + |s| > 2/h$ . In the other case, consider on the discrete cube the function

$$g(\varepsilon) = \frac{f_\varepsilon(t, s) - f(t, s)}{|t| + |s|}.$$

It has  $\mu$ -mean zero, and the modulus of its gradient satisfies, by (3.2),

$$|\nabla g(\varepsilon)| \leq \frac{1}{|t| + |s|} \left( \frac{|t|\sqrt{B_n} + |s|\sqrt{n}}{B_n} + \frac{|t||t|\sqrt{n}}{2B_n} \right) \leq \frac{\sqrt{n}}{B_n} + \frac{\sqrt{n}}{B_n h} \leq \frac{3\sqrt{n}}{B_n h},$$

where we have made use of the property  $B_n \leq n$  and the assumptions  $h \leq 2$  and  $|t| \leq 2/h$  in the two last steps. It remains to apply inequality (2.1) to the function  $g$ .

*Proof of Lemma 3.1.* We use the following equivalent definition of the modulus of the gradient of  $f$ ,

$$(3.5) \quad |\nabla f(\varepsilon)| = \max_a \left| \sum_{k=1}^n a_k (f(\varepsilon) - f(s_k(\varepsilon))) \right|, \quad \varepsilon \in \{0, 1\}^n;$$

here the maximum is computed over all  $n$ -tuples of complex numbers  $a = (a_1, \dots, a_n)$  such that  $|a_1|^2 + \dots + |a_n|^2 = 1$ .

Note that, for each  $k \leq n$ ,

$$S_n(s_k(\varepsilon)) = S_n(\varepsilon) - \frac{\delta_k X_k}{\sqrt{B_n}}, \quad V_n(s_k(\varepsilon)) = V_n(\varepsilon) - \frac{\delta_k X_k}{B_n},$$

where

$$\delta_k = 2\varepsilon_k - 1 = \begin{cases} 1 & \text{if } \varepsilon_k = 1, \\ -1 & \text{if } \varepsilon_k = 0. \end{cases}$$

Hence, we may write

$$\begin{aligned} & e^{itS_n(\varepsilon)} e^{isV_n(\varepsilon)} - e^{itS_n(s_k(\varepsilon))} e^{isV_n(s_k(\varepsilon))} \\ &= e^{itS_n(\varepsilon)+isV_n(\varepsilon)} \left(1 - e^{-it\delta_k X_k/\sqrt{B_n}}\right) + e^{itS_n(s_k(\varepsilon))+isV_n(\varepsilon)} \left(1 - e^{-is\delta_k X_k^2/B_n}\right). \end{aligned}$$

By (3.5),

$$\begin{aligned} |\nabla f_\varepsilon(t, s)| &\leq \max_a \left| \mathbf{E} e^{itS_n(\varepsilon)+isV_n(\varepsilon)} \sum_{k=1}^n a_k \left(1 - e^{-it\delta_k X_k/\sqrt{B_n}}\right) \right| \\ &\quad + \max_a \left| \mathbf{E} e^{itS_n(s_k(\varepsilon))+isV_n(\varepsilon)} \sum_{k=1}^n a_k \left(1 - e^{-is\delta_k X_k^2/B_n}\right) \right| \\ &\leq \max_a \mathbf{E} \left| \sum_{k=1}^n a_k \left(1 - e^{-it\delta_k X_k/\sqrt{B_n}}\right) \right| + \max_a \mathbf{E} \left| \sum_{k=1}^n a_k \left(1 - e^{-is\delta_k X_k^2/B_n}\right) \right|. \end{aligned}$$

To estimate the first expectation, we use the estimate  $|e^{i\alpha} - 1 - i\alpha| \leq \frac{1}{2} \alpha^2$  ( $\alpha \in \mathbf{R}$ ), which yields

$$\left| \sum_{k=1}^n a_k \left(1 - e^{-it\delta_k X_k/\sqrt{B_n}}\right) \right| \leq \frac{|t|}{\sqrt{B_n}} \left| \sum_{k=1}^n a_k \delta_k X_k \right| + \frac{t^2}{2B_n} \sum_{k=1}^n |a_k| X_k^2.$$

By Cauchy's inequality and the basic assumption (1.1),

$$\left( \mathbf{E} \left| \sum_{k=1}^n a_k \delta_k X_k \right| \right)^2 \leq \mathbf{E} \left| \sum_{k=1}^n a_k \delta_k X_k \right|^2 = 1.$$

On the other hand,  $\sum_{k=1}^n |a_k| \leq \sqrt{n}$ . Together with  $\mathbf{E} X_k^2 = 1$ , we get

$$(3.6) \quad \mathbf{E} \left| \sum_{k=1}^n a_k \left(1 - e^{-it\delta_k X_k/\sqrt{B_n}}\right) \right| \leq \frac{|t|}{\sqrt{B_n}} + \frac{t^2 \sqrt{n}}{2B_n}.$$

To estimate the second expectation, just use  $|e^{i\alpha} - 1| \leq |\alpha|$  ( $\alpha \in \mathbf{R}$ ), which leads to

$$(3.7) \quad \mathbf{E} \left| \sum_{k=1}^n a_k \left(1 - e^{-is\delta_k X_k^2/B_n}\right) \right| \leq \frac{|s|}{B_n} \mathbf{E} \sum_{k=1}^n |a_k| X_k^2 \leq \frac{|s| \sqrt{n}}{B_n}.$$

It remains to combine (3.6) with (3.7) in order to finish the proof of Lemma 3.1.

**4. Concentration of distributions in the metric  $\rho$ .** In the next step we need to show that the distance  $\rho(F_\varepsilon, F)$  is small for most of  $\varepsilon \in \{0, 1\}^n$  with respect to the measure  $\mu$ . Starting from (3.3), it is unlikely that one can bound the  $\mu$ -probability

$$\mu \left\{ \varepsilon : \sup_{t,s} \frac{|f_\varepsilon(t, s) - f(t, s)|}{|t| + |s|} \geq h \right\}$$

without stronger moment assumptions on  $X_k$ . We thus use a metric  $\rho$  which is less sensitive to the behavior of characteristic functions near zero. Note that always  $\rho(F, G) \leq 2$ .

PROPOSITION 4.1. *For any  $h \in (0, 2]$ ,*

$$(4.1) \quad \mu \{ \varepsilon : \rho(F_\varepsilon, F) \geq h \} \leq \frac{C}{h^2} e^{-B_n^2 h^2 / (9n)},$$

where  $C$  is a universal constant.



*Proof.* In the definition

$$\rho(F_\varepsilon, F) = \max_{t,s} \frac{|f_\varepsilon(t, s) - f(t, s)|}{1 + |t| + |s|}$$

one may restrict oneself to the half-plane  $s \geq 0$ . Consider points

$$\begin{aligned} t_r &= r \cdot ch, & r &= 0, \pm 1, \pm 2, \dots, \pm N, \\ s_l &= l \cdot ch, & l &= 0, 1, 2, \dots, N. \end{aligned}$$

where  $N = [2/(ch)]$  (the integer part) and where the constant  $c$  (such that  $0 < c < 1$ ) will be specified later on. Let

$$\Omega_0(h) = \left\{ \varepsilon \in \{0, 1\}^n : \frac{|f_\varepsilon(t_r, s_l) - f(t_r, s_l)|}{|t_r| + |s_l|} < h \text{ for all } 0 \leq |r|, l \leq N - 1 \right\}$$

with the convention that the ratio is defined to be zero for  $r = l = 0$ . By Corollary 3.1,

$$\begin{aligned} 1 - \mu(\Omega_0(h)) &\leq \sum_{|r| < N, l < N} \mu \left\{ \frac{|f_\varepsilon(t_r, s_l) - f(t_r, s_l)|}{|t_r| + |s_l|} \geq h \right\} \\ (4.2) \quad &\leq 4N(2N - 1) e^{-B_n^2 h^2 / (9n)} \leq \frac{16}{c^2 h^2} e^{-B_n^2 h^2 / (9n)}. \end{aligned}$$

Take  $\varepsilon \in \Omega_0(h)$  and a point  $(t, s)$  from the square  $[0, t_N] \times [0, s_N]$ . Then, for some  $r, l = 1, \dots, N$ ,

$$t_{r-1} \leq t \leq t_r, \quad s_{l-1} \leq s \leq s_l,$$

so  $t - t_{r-1} \leq ch, s - s_{l-1} \leq ch$ . Since

$$|f_\varepsilon(t, s) - f_\varepsilon(t_{r-1}, s_{l-1})| \leq \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right| (t - t_{r-1}) + \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| (s - s_{l-1})$$

and similarly for  $f$ , we obtain that

$$\begin{aligned} |f_\varepsilon(t, s) - f(t, s)| &\leq |f_\varepsilon(t_{r-1}, s_{l-1}) - f(t_{r-1}, s_{l-1})| \\ &\quad + |f_\varepsilon(t, s) - f_\varepsilon(t_{r-1}, s_{l-1})| + |f(t, s) - f(t_{r-1}, s_{l-1})| \\ &\leq (t_{r-1} + s_{l-1}) h + \left( \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right| + \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| + \sup_{t,s} \left| \frac{\partial f}{\partial t} \right| + \sup_{t,s} \left| \frac{\partial f}{\partial s} \right| \right) ch \\ (4.3) \quad &\leq (t + s) h + \left( \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right| + \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| + \sup_{t,s} \left| \frac{\partial f}{\partial t} \right| + \sup_{t,s} \left| \frac{\partial f}{\partial s} \right| \right) ch. \end{aligned}$$

Recall that  $f_\varepsilon(t, s) = \mathbf{E} e^{itS_n(\varepsilon) + isV_n(\varepsilon)}$ . To estimate the absolute values of the partial derivatives of  $f_\varepsilon(t, s)$ , we may use the trivial bounds

$$(4.4) \quad \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right| \leq \mathbf{E} |S_n(\varepsilon)| \leq (\mathbf{E} |S_n(\varepsilon)|^2)^{1/2} = \left( \frac{\varepsilon_1 + \dots + \varepsilon_n}{B_n} \right)^{1/2},$$

$$(4.5) \quad \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| \leq \mathbf{E} V_n(\varepsilon) = \frac{\varepsilon_1 + \dots + \varepsilon_n}{B_n}.$$

Indeed, consider on the discrete cube the function  $g(\varepsilon) = (\varepsilon_1 + \dots + \varepsilon_n)/B_n$ . It has  $\mu$ -mean 1 and a constant modulus of gradient  $|\nabla g(\varepsilon)| \equiv \sigma = \sqrt{n}/B_n$ . Hence, by the one-sided deviation inequality (2.5) applied to  $g$ ,

$$\mu\{g - 1 \geq R\} \leq e^{-2R^2/\sigma^2}, \quad R > 0.$$

Using it with  $R = \frac{\sqrt{2}}{3}$ , we arrive at

$$(4.6) \quad \mu \left\{ \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| \geq 1 + \frac{\sqrt{2}}{3} \right\} \leq e^{-4B_n^2/(9n)} \leq e^{-B_n^2 h^2/(9n)}$$

(since  $h \leq 2$ ), and similarly

$$(4.7) \quad \mu \left\{ \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right|^2 \geq 1 + \frac{\sqrt{2}}{3} \right\} \leq e^{-B_n^2 h^2/(9n)}.$$

Also note that by (4.4)–(4.5) and (3.1), since  $f$  represents the  $\mu$ -mean of  $f_\varepsilon$ 's,

$$(4.8) \quad \sup_{t,s} \left| \frac{\partial f}{\partial t} \right| \leq \int \mathbf{E}|S_n(\varepsilon)| d\mu(\varepsilon) \leq 1, \quad \sup_{t,s} \left| \frac{\partial f}{\partial s} \right| \leq \int \mathbf{E}V_n(\varepsilon) d\mu(\varepsilon) = 1.$$

Now, introduce

$$\Omega_1 = \left\{ \varepsilon \in \{0, 1\}^n : \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial t} \right|^2 < 1 + \frac{\sqrt{2}}{3}, \quad \sup_{t,s} \left| \frac{\partial f_\varepsilon}{\partial s} \right| < 1 + \frac{\sqrt{2}}{3} \right\}.$$

Then, by (4.8), if  $\varepsilon \in \Omega_0(h) \cap \Omega_1$ , (4.3) implies

$$|f_\varepsilon(t, s) - f(t, s)| \leq (t + s)h + \left( 3 + \frac{\sqrt{2}}{3} + \left( 1 + \frac{\sqrt{2}}{3} \right)^{1/2} \right) ch < (t + s)h + 5ch.$$

A similar argument works in the case  $-t_N \leq t \leq 0$ . So taking, for example,  $c = \frac{1}{5}$ , for all  $(t, s) \in [-t_N, t_N] \times [0, s_N]$ , we obtain that

$$(4.9) \quad |f_\varepsilon(t, s) - f(t, s)| \leq (1 + |t| + s)h \quad \text{for } \varepsilon \in \Omega_0(h) \cap \Omega_1.$$

On the other hand, if  $|t| > t_N$  or  $s > s_N$ , then  $(1 + |t| + s)h > (1 + [2/h])h > 2$ , so inequality (4.9) will be fulfilled automatically. Thus, applying (4.2) and (4.6)–(4.7), we conclude that

$$\begin{aligned} \mu \left\{ \sup_{t,s} \frac{|f_\varepsilon(t, s) - f(t, s)|}{1 + |t| + |s|} \geq h \right\} &\leq 1 - \mu(\Omega_0(h) \cap \Omega_1) \\ &\leq \frac{16}{c^2 h^2} e^{-B_n^2 h^2/(9n)} + 2e^{-B_n^2 h^2/(9n)}. \end{aligned}$$

Finally, note that  $16/(c^2 h^2) + 2 = 16 \cdot 25/h^2 + 2 \leq 408/h^2$ , for  $0 < h \leq 2$ . Thus, Proposition 4.1 follows with  $C = 408$ .

**5. Average distributions.** Let us start with an infinite sequence  $\{X_n\}_{n \geq 1}$  of random variables satisfying the orthonormality condition (1.1). Here we concentrate on the asymptotic behavior of the average distributions  $F = F_n$  for a growing “dimension” parameter  $n$ . Thus, fix a sequence  $p_n \in (0, 1)$  and, for each  $n$ , equip the discrete cube  $\{0, 1\}^n$  with the product of Bernoullian measures  $\mu_n = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$ .

Introduce the random variable  $R_n \geq 0$  by

$$(5.1) \quad R_n^2 = \frac{1}{B_n} \sum_{k=1}^n p_k X_k^2,$$

and denote by  $G_n$  the distribution of the two-dimensional random vector

$$W_n = (R_n Z, R_n^2),$$

where  $Z$  is an  $N(0, 1)$ -random variable independent of all  $X_n$ 's. Note that  $\mathbf{E}R_n^2 = 1$ .

PROPOSITION 5.1. *We have  $\rho(F_n, G_n) \rightarrow 0$  as  $n \rightarrow \infty$ , if*

- (a)  $p_n \rightarrow 0$ ;
- (b)  $B_n \rightarrow +\infty$ ;
- (c)  $\max_{1 \leq k \leq n} |X_k| = o(\sqrt{B_n})$  in probability.

It is obvious that assumption (b) is implied by (c). The latter is defined as

$$(5.2) \quad \mathbf{P} \left\{ \frac{\max_{1 \leq k \leq n} |X_k|}{\sqrt{B_n}} \leq \lambda \right\} \rightarrow 1 \quad \text{for any } \lambda > 0.$$

*Proof of Proposition 5.1.* Changing the order of integration in the definition of the characteristic function  $f_n$  of  $F_n$ , we may write it as  $f_n(t, s) = \mathbf{E}\Pi_n(t, s)$ , where

$$\Pi_n(t, s) = \prod_{k=1}^n \left( q_k + p_k e^{itX_k/\sqrt{B_n} + isX_k^2/B_n} \right), \quad t, s \in \mathbf{R}, \quad q_k = 1 - p_k.$$

By conditions (a) and (b), we may assume without loss of generality that  $p_n \leq \frac{1}{4}$  for all  $n$ . Fix a point  $(t, s)$ . We use  $\log$  to denote the principal value of the logarithm. By Taylor's expansion,  $|\log(1 + z) - z| < |z|^2$  for any complex  $z$  such that  $|z| \leq \frac{1}{2}$ . Applying this to  $z = p_k(\exp\{itX_k/\sqrt{B_n} + isX_k^2/B_n\} - 1)$ , we may write

$$(5.3) \quad \log \Pi_n(t, s) = \sum_{k=1}^n p_k \left( e^{itX_k/\sqrt{B_n} + isX_k^2/B_n} - 1 \right) + \sum_{k=1}^n \theta_k p_k^2 \left| e^{itX_k/\sqrt{B_n} + isX_k^2/B_n} - 1 \right|^2$$

with some random complex  $\theta_k$  such that  $|\theta_k| < 1$ .

First let us note that the second sum in (5.3) is becoming small for large  $n$ . Using  $|e^{i\alpha} - 1| \leq |\alpha|$  ( $\alpha \in \mathbf{R}$ ), we observe that on the set  $\Omega_n = \{\max_{1 \leq k \leq n} |X_k|/\sqrt{B_n} \leq 1\}$

$$\left| e^{itX_k/\sqrt{B_n} + isX_k^2/B_n} - 1 \right|^2 \leq (|t| + |s|)^2 \frac{X_k^2}{B_n}.$$

But the random variables  $\xi_n = B_n^{-1} \sum_{k=1}^n \theta_k p_k^2 X_k^2$  have first absolute moments satisfying

$$\mathbf{E}|\xi_n| \leq \frac{1}{B_n} \sum_{k=1}^n p_k^2 = \frac{p_1^2 + \cdots + p_n^2}{p_1 + \cdots + p_n} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

due to assumptions (a) and (b). Together with (5.2) we conclude that the second sum in (5.3) represents a random variable, say  $Y_n$ , such that  $Y_n \rightarrow 0$  in probability as  $n \rightarrow \infty$ .

Now, let us turn to the first sum in (5.3). We will need a strengthened (but still equivalent) version of (5.2), namely, the property that there is a sequence of reals  $\lambda_n \downarrow 0$  such that the events

$$\Omega_n(\lambda_n) = \left\{ \frac{\max_{1 \leq k \leq n} |X_k|}{\sqrt{B_n}} \leq \lambda_n \right\}$$

have probabilities  $\mathbf{P}(\Omega_n(\lambda_n)) \rightarrow 1$ . On the set  $\Omega_n(\lambda_n)$ , by Taylor's expansion, uniformly over all  $k \leq n$

$$e^{itX_k/\sqrt{B_n} + isX_k^2/B_n} = 1 + it \frac{X_k}{\sqrt{B_n}} - \left( \frac{t^2}{2} - is \right) \frac{X_k^2}{B_n} + o(1) \frac{X_k^2}{B_n},$$

so the first sum in (5.3) simplifies

$$\frac{it}{\sqrt{B_n}} \sum_{k=1}^n p_k X_k - \left( \frac{t^2}{2} - is \right) R_n^2 + o(1) R_n^2.$$

But by the basic assumption (1.1), the random variables  $\eta_n = B_n^{-1/2} \sum_{k=1}^n p_k X_k$  have second moments  $\mathbf{E}\eta_n^2 = B_n^{-1} \sum_{k=1}^n p_k^2 \rightarrow 0$ , as already mentioned. Hence, the first sum behaves asymptotically like  $(-t^2/2 + is) R_n^2$ .

Eventually, with probability converging to 1, an identity

$$\log \Pi_n(t, s) = \left( -\frac{t^2}{2} + is \right) R_n^2 + \alpha_n(t, s)$$

holds with some random  $\alpha_n$  such that  $\alpha_n \rightarrow 0$ . This implies that

$$f_n(t, s) = \mathbf{E}e^{(-t^2/2 + is) R_n^2 + \alpha_n} = g_n(t, s)(1 + o(1)),$$

where  $g_n(t, s) = \mathbf{E}e^{(-t^2/2 + is) R_n^2}$  is just the characteristic function of  $W_n$ .

This finishes the proof.

*Remark 5.1.* The statement of Proposition 5.1 holds if we replace in (5.1) the random variable  $R_n^2$  with a more canonical expression

$$\widehat{R}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2.$$

More precisely, the sequence  $p_n$  should possess an additional property

$$(5.4) \quad \sum_{k=1}^n \left| \frac{p_k}{B_n} - \frac{1}{n} \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

In this case,  $\mathbf{E}|\widehat{R}_n^2 - R_n^2| \rightarrow 0$  as  $n \rightarrow \infty$ , so  $\mathbf{E}|\widehat{R}_n - R_n| \rightarrow 0$  as well. As a result, the distributions of  $W_n$  and  $\widehat{W}_n = (\widehat{R}_n Z, \widehat{R}_n^2)$  will converge.

*Remark 5.2.* The convergence  $\rho(F_n, G_n) \rightarrow 0$  is equivalent to the (formally stronger) property that, for any bounded, continuous function  $u = u(x, y)$  on the plane,

$$\lim_{n \rightarrow \infty} \left| \int u dF_n - \int u dG_n \right| = 0.$$

This is due to the fact that both sequences  $\{F_n\}_{n \geq 1}$  and  $\{G_n\}_{n \geq 1}$  form precompact families in the space of probability measures on the half-plane  $y \geq 0$  equipped with the topology of weak convergence. Indeed, they have bounded first moments:

$$\iint (|x| + y) dF_n(x, y) = \int (\mathbf{E}|S_n(\varepsilon)| + \mathbf{E}V_n(\varepsilon)) \mu_n(\varepsilon) \leq 2$$

(cf. (4.8)), and similarly for  $G_n$ ,

$$\iint (|x| + y) dG_n(x, y) = \mathbf{E}R_n \mathbf{E}|Z| + \mathbf{E}R_n^2 < 2.$$

**6. Limit behavior along subsequences.** At last, we are prepared to prove Theorems 1.1 and 1.2. Keeping the notation of the previous section, introduce the product of Bernoullian measures  $\mu = \mu_{p_1} \otimes \mu_{p_2} \otimes \dots$  on the infinite dimensional discrete cube  $M = \{0, 1\}^\infty$  parametrized by  $p_n \in (0, 1)$ ,  $n \geq 1$ . For infinite sequences  $\varepsilon = (\varepsilon_n)_{n \geq 1} \in M$ , we use the notation as before, namely,

$$S_n(\varepsilon) = \frac{\varepsilon_1 X_1 + \dots + \varepsilon_n X_n}{\sqrt{B_n}}, \quad V_n(\varepsilon) = \frac{\varepsilon_1 X_1^2 + \dots + \varepsilon_n X_n^2}{B_n}, \quad T_n(\varepsilon) = \frac{S_n(\varepsilon)}{\sqrt{V_n(\varepsilon)}}.$$

Also recall that

$$R_n^2 = \frac{1}{B_n} \sum_{k=1}^n p_k X_k^2, \quad B_n = p_1 + \dots + p_n,$$

and that  $Z$  denotes a standard normal random variable independent of  $R_n$ . First we derive a more general theorem.

**THEOREM 6.1.** *Assume that*

- (a)  $p_n \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (b) for any  $h > 0$ , the series  $\sum_{n=1}^\infty e^{-B_n h^2/n}$  is convergent;
- (c)  $\max_{1 \leq k \leq n} |X_k| = o(\sqrt{B_n})$  in probability.

*Then, for  $\mu$ -almost all  $\varepsilon \in M$ , the two-dimensional distributions of random vectors  $(S_n(\varepsilon), V_n(\varepsilon))$  and  $(R_n Z, R_n^2)$  are convergent in the sense of the weak topology.*

Condition (b) strengthens the assumption  $B_n \rightarrow \infty$  of Proposition 5.1.

As an application, we have (taking into account Remark 5.2) the following corollary.

**COROLLARY 6.1.** *Assuming conditions (a)–(c) of Theorem 6.1, we have the following for  $\mu$ -almost all  $\varepsilon \in M$ :*

- (1) *If  $R_n \rightarrow 1$  in probability as  $n \rightarrow \infty$ , then  $S_n(\varepsilon) \Rightarrow N(0, 1)$ ;*
- (2) *if  $\limsup_{n \rightarrow \infty} \mathbf{P}\{R_n < h\} \rightarrow 0$  as  $h \downarrow 0$ , then  $T_n(\varepsilon) \Rightarrow N(0, 1)$ .*

Note that the assumption of statement (2) is much weaker than that of statement (1).

*Proof of Theorem 6.1.* Given  $\varepsilon \in M$ , denote by  $U_n(\varepsilon)$  its projection  $(\varepsilon_1, \dots, \varepsilon_n)$  onto  $\{0, 1\}^n$ . Clearly,  $U_n(\varepsilon)$  has distribution  $\mu_n = \mu_{p_1} \otimes \dots \otimes \mu_{p_n}$ , and the random vector  $(S_n(\varepsilon), V_n(\varepsilon))$  has distribution  $F_{U_n(\varepsilon)}$ , according to previous notation.

In view of Proposition 5.1, we need to show that  $\rho(F_{U_n(\varepsilon)}, F_n) \rightarrow 0$  for  $\mu$ -almost all  $\varepsilon$ . Fix a number  $h > 0$ . By Proposition 4.1 and assumption (b),

$$\begin{aligned} \sum_{n=1}^{\infty} \mu\{\rho(F_{U_n(\varepsilon)}, F_n) \geq h\} &= \sum_{n=1}^{\infty} \mu_n\{\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) : \rho(F_\varepsilon, F_n) \geq h\} \\ &\leq \sum_{n=1}^{\infty} \frac{C}{h^2} e^{-B_n^2 h^2/n} < +\infty. \end{aligned}$$

Therefore, by the Borel–Cantelli lemma, for  $\mu$ -almost all  $\varepsilon \in M$ , we obtain that  $\rho(F_{U_n(\varepsilon)}, F_n) < h$ , for all sufficiently large  $n$ . So,

$$\mu\left\{\limsup_{n \rightarrow \infty} \rho(F_{U_n(\varepsilon)}, F_n) \leq h\right\} = 1.$$

Letting  $h$  tend to zero finishes the proof.

*Proof of Corollary 6.1.* To deduce statement (2), we need to make sure, as is discussed in Remark 5.2, that for  $\mu$ -almost all  $\varepsilon \in M$ , the sequence of distributions  $\{F_{U_n(\varepsilon)}\}_{n \geq 1}$  forms a precompact family in the space of probability measures on the half-plane  $y \geq 0$ . Indeed, for the first moments we have

$$\begin{aligned} \int \int (|x| + y) dF_{U_n(\varepsilon)}(x, y) &= \int (\mathbf{E}|S_n(\varepsilon)| + \mathbf{E}V_n(\varepsilon)) d\mu_n(\varepsilon) \\ &\leq \left(\frac{\varepsilon_1 + \dots + \varepsilon_n}{B_n}\right)^{1/2} + \frac{\varepsilon_1 + \dots + \varepsilon_n}{B_n}, \end{aligned}$$

which is bounded in  $n$  with  $\mu$ -probability 1, by the strong law of large numbers. Hence, for any bounded, continuous function  $u = u(x, y)$ , we have for  $\mu$ -almost all  $\varepsilon$ ,

$$(6.1) \quad \lim_{n \rightarrow \infty} |\mathbf{E}u(S_n(\varepsilon), V_n(\varepsilon)) - \mathbf{E}u(R_n Z, R_n^2)| = 0.$$

The assumption in statement (2) implies that  $\limsup_{n \rightarrow \infty} \mathbf{P}\{V_n(\varepsilon) < h\} \rightarrow 0$  as  $h \downarrow 0$ . Hence, we may apply (6.1) to functions of the form  $u(x, y) = v(x/\sqrt{y})$ , which gives  $\mathbf{E}v(U_n(\varepsilon)) \rightarrow \mathbf{E}v(Z)$  as  $n \rightarrow \infty$ . Thus Corollary 6.1 follows.

Now, let us focus on specific choices of product measures  $\mu$  on  $M$ , that is, of sequences  $p_n$ , which make assumptions on  $X_n$  as weak as possible. For example, since  $B_n \leq n$ , one can try to weaken assumption (c) of Theorem 6.1 to

$$(6.2) \quad \max_{1 \leq k \leq n} |X_k| = o(\sqrt{n}) \quad \text{in probability as } n \rightarrow \infty.$$

Note that the basic hypothesis (1.1) or, more precisely, the assumption  $\mathbf{E}X_n^2 = 1$ , implies only  $\max_{1 \leq k \leq n} |X_k| = O(\sqrt{n})$  so that (6.2) represents a mild strengthening. On the other hand, (6.2) is already fulfilled under a mild moment assumption such as

$$(6.3) \quad \sup_n \mathbf{E}\Psi(X_n^2) < +\infty,$$

where  $\Psi(x)$  is a nonnegative increasing function in  $x \geq 0$  such that  $\Psi(x)/x \rightarrow +\infty$  as  $x \rightarrow +\infty$ . Indeed, by Chebyshev’s inequality, for any  $\lambda > 0$ , we would get

$$\begin{aligned} \mathbf{P}\left\{\frac{\max_{1 \leq k \leq n} |X_k|}{\sqrt{n}} \geq \lambda\right\} &\leq \frac{\mathbf{E}\Psi(\max_{1 \leq k \leq n} |X_k|^2)}{\Psi(\lambda^2 n)} \\ &\leq \frac{\mathbf{E}\Psi(X_1^2) + \dots + \mathbf{E}\Psi(X_n^2)}{\Psi(\lambda^2 n)} \rightarrow 0. \end{aligned}$$

The existence of such a function  $\Psi$  is equivalent to the property of the sequence  $\{X_n^2\}$  to be uniformly integrable in the Lebesgue space  $L^1(\Omega, \mathcal{F}, \mathbf{P})$ . In that case,  $\Psi$  can be chosen to be convex; cf., e.g., [14].

**THEOREM 6.2.** *If (6.2) is fulfilled, then for some product probability measure  $\mu$  on  $M = \{0, 1\}^\infty$ , the distributions of  $(S_n(\varepsilon), V_n(\varepsilon))$  and  $(R_n Z, R_n^2)$  converge for  $\mu$ -almost all  $\varepsilon$  in  $M$ .*

The proof is based on Theorem 6.1 and the following simple observation, which is needed for the proof of Theorems 1.1 and 1.2 as well.

**LEMMA 6.1.** *Given a sequence of positive reals  $(A_n)_{n \geq 1}$  such that  $A_n = o(n)$  as  $n \rightarrow \infty$ , one can find a sequence  $(p_n)_{n \geq 1}$  with the following properties:*

- (1)  $0 < p_n < 1$ ;
- (2)  $p_n$  is nonincreasing;
- (3)  $p_n \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (4)  $p_{[\sqrt{n}]} \sim p_n$  as  $n \rightarrow \infty$ ;
- (5)  $\sum_{n=1}^\infty e^{-B_n^2 h^2/n} < \infty$  for all  $h > 0$ , where  $B_n = p_1 + \dots + p_n$ ;
- (6)  $\sum_{k=1}^n |p_k/B_n - 1/n| \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (7)  $B_n \geq A_n$  for all  $n$  large enough.

Here  $[\sqrt{n}]$  denotes the integer part of  $\sqrt{n}$  and  $\sim$  denotes the equivalence of sequences. Thus, property (4) means that  $p_n$  is varying very slowly. Together with other properties, it actually implies property (6), appearing in Remark 5.1; cf. (5.4). Indeed, write

$$\begin{aligned} \sum_{k=1}^n \left| \frac{p_k}{B_n} - \frac{1}{n} \right| &= 2 \sum_{1 \leq k < \sqrt{n}} \left( \frac{p_k}{B_n} - \frac{1}{n} \right)^+ + 2 \sum_{\sqrt{n} \leq k \leq n} \left( \frac{p_k}{B_n} - \frac{1}{n} \right)^+ \\ &\leq 2 \sum_{k < \sqrt{n}} \frac{p_k}{B_n} + 2n \left( \frac{p_{[\sqrt{n}]}}{B_n} - \frac{1}{n} \right)^+ \leq 2 \frac{\sqrt{n}}{B_n} + 2 \left( \frac{np_{[\sqrt{n}]}}{B_n} - 1 \right)^+, \end{aligned}$$

where we used notation  $a^+ = \max\{a, 0\}$  and where we applied properties (1) and (2) to justify the last inequality. Now, by property (5),  $\sqrt{n}/B_n \rightarrow 0$  as  $n \rightarrow \infty$ . Finally, in terms of  $c_n = p_{[\sqrt{n}]} / p_n$ , since  $np_n \leq B_n$ , we obtain that

$$\left( \frac{np_{[\sqrt{n}]}}{B_n} - 1 \right)^+ = \left( \frac{c_n np_n}{B_n} - 1 \right)^+ \leq c_n - 1 \rightarrow 0.$$

*Proof of Lemma 6.4.* The last requirement, (7), may be weakened to  $A_n = O(B_n)$ , since one can always start with a sequence  $A'_n$  such that  $A'_n = o(n)$  and  $A_n = o(A'_n)$ .

Put  $\alpha_n = \max_{k \geq n} A_k/k$ . Since this sequence is not increasing, it majorizes  $A_n/n$  and satisfies  $\alpha_n \downarrow 0$  as  $n \rightarrow \infty$ . Hence, the sequence  $\beta_n = (\alpha_1 + \dots + \alpha_n)/n$  also majorizes  $A_n/n$  and converges to zero. Let  $A'_n = \beta_n n = \alpha_1 + \dots + \alpha_n$  so that  $A'_n \geq A_n$ ,  $A'_n = o(n)$ , and  $A'_{n+1} - A'_n \downarrow 0$  as  $n \rightarrow \infty$ . Given  $c > 0$ , define

$$B'_n = c(A'_n + n^{3/4}), \quad n \geq 1.$$

This sequence is strictly increasing, and again we have  $B'_n \geq A_n$ ,  $B'_n = o(n)$ ,  $B'_{n+1} - B'_n \downarrow 0$  as  $n \rightarrow \infty$  (since the increments  $(n+1)^{3/4} - n^{3/4}$  decrease as well). Moreover, the constant  $c$  can be chosen to satisfy  $p'_n \equiv B'_{n+1} - B'_n < 1$  for all  $n \geq 1$ .

The sequence  $p'_n$  meets all the requirements of the lemma except perhaps property (4) (and therefore (6)), and with property (7) replaced by  $B'_n \geq cA_n$ . Hence, it

remains to modify  $p'_n$  to make it vary slowly. For this purpose, introduce the sequence  $r'_n \geq 0$  through  $R'_n \equiv -\log p'_n = r'_1 + \dots + r'_n$  and define  $\Delta'_k = r'_{2^{k-1}} + \dots + r'_{2^k-1}$ . Then  $\sum_{k=1}^\infty \Delta'_k = \lim_{k \rightarrow \infty} R'_{2^k-1} = +\infty$ . Therefore, one can choose a sequence  $\Delta_k$  with the following properties:

- (a)  $0 \leq \Delta_k \leq \Delta'_k$ ;
- (b)  $\Delta_k \rightarrow 0$  as  $k \rightarrow \infty$ ;
- (c)  $\sum_{k=1}^\infty \Delta_k = +\infty$ .

Moreover, by the first property, for each  $k \geq 1$ , one can choose nonnegative real numbers  $r_{2^{k-1}} \leq r'_{2^{k-1}}, \dots, r_{2^k-1} \leq r'_{2^k-1}$  such that  $r_{2^{k-1}} + \dots + r_{2^k-1} = \Delta_k$ . Finally, put  $R_n = r_1 + \dots + r_n$  and  $p_n = e^{-R_n}$ . It follows that  $0 < p'_n \leq p_n < 1$ ,  $(p_n)_{n \geq 1}$  is nonincreasing, and  $B_n = p_1 + \dots + p_n \geq p'_1 + \dots + p'_n = B'_n \geq cA_n$ . Hence, properties (1), (2), (5) and the weakened version of property (7) are fulfilled automatically. Moreover, property (3) is also satisfied, since by monotonicity and property (c),  $p_{2^k-1} = e^{-(\Delta_1 + \dots + \Delta_k)} \rightarrow 0$  as  $k \rightarrow \infty$ . Finally,  $(p_{2^k-1})/p_{2^k-1} = e^{-\Delta_k} \rightarrow 1$  by property (b), which implies  $p_{[\log n]}/p_n \rightarrow 1$  as  $n \rightarrow \infty$ . This is stronger than property (4), so Lemma 6.1 is proved.

*Proof of Theorem 6.2.* Condition (6.2) can formally be strengthened as

$$(6.4) \quad \max_{1 \leq k \leq n} |X_k| = o(\lambda_n \sqrt{n}) \quad \text{in probability as } n \rightarrow \infty,$$

for some  $\lambda_n \rightarrow 0$ . Put

$$(6.5) \quad A_n = \lambda_n^2 n$$

so that  $A_n = o(n)$ . Let us recall how one can choose  $A_n$ , i.e.,  $\lambda_n$  in (6.4). For any sequence of random variables  $\xi_n \geq 0$ , the convergence  $\xi_n \rightarrow 0$  in probability as  $n \rightarrow \infty$  is equivalent to  $\mathbf{E}L(\xi_n) \rightarrow 0$ , where

$$L(x) = \frac{x}{1+x}, \quad x \geq 0.$$

This function is concave, positive for  $x > 0$ , and satisfies  $L(0) = 0, 0 \leq L < 1$ . In particular, the inequality  $L(x/\lambda) \leq \lambda^{-1} L(x)$  holds for all  $x \geq 0$  and  $\lambda \in (0, 1]$ . Moreover, if  $\mathbf{E}\xi_n > 0$ , the sequence  $\lambda_n = (\mathbf{E}L(\xi_n))^{1/2}$  satisfies the properties  $0 < \lambda_n < 1, \lambda_n \rightarrow 0$ , and  $\mathbf{E}L(\xi_n/\lambda_n) \leq \lambda_n$ . Thus, in the case  $\xi_n = M_n/n, M_n = \max_{1 \leq k \leq n} |X_k|^2$ , one can take  $A_n = \lambda_n^2 n$  with

$$(6.6) \quad \lambda_n^2 = \left( \mathbf{E}L\left(\frac{M_n}{n}\right) \right)^{1/2}.$$

Picking a sequence  $p_n$  from Lemma 6.1, all the conditions of Theorem 6.1 will be fulfilled, and thus the conclusion follows.

Taking into account Remark 5.1, from Lemma 6.1 and Theorem 6.2 we obtain immediately the assertions of Theorems 1.1 and 1.2 in a stronger “almost surely” form similarly to statements (1)–(2) of Corollary 6.1 (now with  $R_n$  replaced by  $\tilde{R}_n$ ). It remains to prove the refining property (1.4).

**THEOREM 6.3.** *Given a sequence  $\{j_n\}_{n \geq 1}$  such that  $\lim_{n \rightarrow \infty} j_n/n \rightarrow +\infty$ , a sequence  $\{i_n\}_{n \geq 1}$  in Theorems 1.1 and 1.2 can be chosen to satisfy  $i_n \leq j_n$  for all  $n$  large enough.*

*Proof.* Let  $\mu = \mu_{p_1} \otimes \mu_{p_2} \otimes \dots$  be product of Bernoulli measures on  $M = \{0, 1\}^\infty$  constructed for a sequence  $p_n$  from Lemma 6.1 with  $A_n$  defined in (6.5) and (6.6). In



the conditions of Theorems 1.1 and 1.2, we have proved the weak convergence

$$S_n(\varepsilon) \Rightarrow N(0, 1), \quad T_n(\varepsilon) \Rightarrow N(0, 1)$$

for  $\mu$ -almost all  $\varepsilon = (\varepsilon_n)_{n \geq 1}$  in  $M$ . Thus, such an  $\varepsilon$  exists, but we need to choose a 0–1-sequence satisfying (1.4). Note that since  $B_n = p_1 + \cdots + p_n \rightarrow +\infty$ , for  $\mu$ -almost  $\varepsilon$ , we have, by the strong law of large numbers,

$$(6.7) \quad \frac{\varepsilon_1 + \cdots + \varepsilon_n}{B_n} \longrightarrow 1.$$

Thus,  $S_n(\varepsilon)$  may be replaced with  $S_n$ , introduced in section 1, for the increasing sequence of indices

$$i_k = \min\{n \geq 1: \varepsilon_1 + \cdots + \varepsilon_n = k\}, \quad k \geq 1.$$

Hence, by (6.7), for all sufficiently large  $k$ , we have  $k = \varepsilon_1 + \cdots + \varepsilon_{i_k} \geq B_{i_k}/2$ , that is,  $B_{i_k} \leq 2k$ . Recall that in Lemma 6.1 the sequence  $B_n$  may be chosen to majorize any prescribed  $A_n$  such that  $A_n = o(n)$ . Therefore, with  $\mu$ -probability 1, we also have  $A_{i_k} \leq 2k$  for all sufficiently large  $k$ . Let us start with an increasing, continuous function  $A = A(x)$  of  $x \geq 0$  such that  $A(0) = 0$ ,  $A(x)/x \rightarrow 0$  as  $x \rightarrow +\infty$ , with inverse function  $A^{-1}$ . Put  $A_i = A(i)$  and solve  $A_{i_k} \leq 2k$  as  $i_k \leq A^{-1}(2k)$ . Finally, the inequality  $2k \leq A(j_k)$  holds by an appropriate choice of the function  $A$ .

#### REFERENCES

- [1] M. ANTILA, K. BALL, AND I. PERISSINAKI, *The central limit problem for convex bodies*, Trans. Amer. Math. Soc., 355 (2003), pp. 4723–4735.
- [2] V. BENTKUS AND F. GÖTZE, *The Berry–Esseen bound for Student’s statistic*, Ann. Probab., 24 (1996), pp. 491–503.
- [3] S. G. BOBKOV, *On concentration of distributions of random weighted sums*, Ann. Probab., 31 (2003), pp. 195–215.
- [4] S. G. BOBKOV, *Concentration of distributions of the weighted sums with Bernoullian coefficients*, in Geometric Aspects of Functional Analysis, V. D. Milman, ed., Lecture Notes in Math. 1807, Springer, Berlin, 2003, pp. 27–36.
- [5] S. G. BOBKOV AND F. GÖTZE, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal., 163 (1999), pp. 1–28.
- [6] G. CHISTYAKOV AND F. GÖTZE, *Limit distributions of Studentized means*, Ann. Probab., 32 (2004), pp. 28–77.
- [7] P. DIACONIS AND D. FREEDMAN, *Asymptotics of graphical projection pursuit*, Ann. Statist., 12 (1984), pp. 793–815.
- [8] E. GINÉ, F. GÖTZE, AND D. M. MASON, *When is the Student t-statistic asymptotically standard normal?* Ann. Probab., 25 (1997), pp. 1514–1531.
- [9] P. S. GRIFFIN AND D. M. MASON, *On the asymptotic normality of self-normalized sums*, Proc. Cambridge Philos. Soc., 109 (1991), pp. 597–610.
- [10] S. JANSON, *Some pairwise independent sequences for which the central limit theorem fails*, Stochastics, 23 (1988), pp. 439–448.
- [11] M. LEDOUX, *Concentration of measure and logarithmic Sobolev inequalities*, in Séminaire de Probabilités XXXIII, Lecture Notes in Math. 1709, Springer, Berlin, 1999, pp. 120–216.
- [12] B. LOGAN, C. MALLOWS, S. RICE, AND L. SHEPP, *Limit distributions of self-normalized sums*, Ann. Probab., 1 (1973), pp. 788–809.
- [13] R. A. MALLER, *A theorem on products of random variables, with application to regression*, Austral. J. Statist., 23 (1981), pp. 177–185.
- [14] P.-A. MEYER, *Probability and Potentials*, Blaisdell, Waltham, MA, 1966.
- [15] A. R. PRUSS, *A bounded N-tuplewise independent and identically distributed counterexample to the CLT*, Probab. Theory Related Fields, 111 (1998), pp. 323–332.
- [16] Q.-M. SHAO, *Self-normalized large deviations*, Ann. Probab., 25 (1997), pp. 285–328.

- [17] V. N. SUDAKOV, *Typical distributions of linear functionals in finite-dimensional spaces of higher dimension*, Sov. Math. Dokl., 19 (1978), pp. 1578–1582.
- [18] Q. WANG AND B.-Y. JING, *An exponential nonuniform Berry–Esseen bound for self-normalized sums*, Ann. Probab., 27 (1999), pp. 2068–2088.
- [19] H. VON WEIZSÄCKER, *Sudakov’s typical marginals, random linear functionals and a conditional central limit theorem*, Probab. Theory Related Fields, 107 (1997), pp. 313–324.