The Annals of Probability
2013, Vol. 41, No. 4, 2479–2512
DOI: 10.1214/12-AOP780

RATE OF CONVERGENCE AND EDGEWORTH-TYPE EXPANSION IN THE ENTROPIC CENTRAL LIMIT THEOREM[1]

BY SERGEY G. BOBKOV, GENNADIY P. CHISTYAKOV AND FRIEDRICH GÖTZE

University of Minnesota, University of Bielefeld and University of Bielefeld

An Edgeworth-type expansion is established for the entropy distance to the class of normal distributions of sums of i.i.d. random variables or vectors, satisfying minimal moment conditions.

1. Introduction. Let $(X_n)_{n\geq 1}$ be independent, identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\mathrm{Var}(X_1) = 1$. According to the central limit theorem, the normalized sums

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

are weakly convergent in distribution to the standard normal law $Z_n \Rightarrow Z$, where $Z \sim N(0,1)$ with density $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. A much stronger statement (when applicable)—the entropic central limit theorem—states that, if for some n_0, or equivalently, for all $n \geq n_0$, the random variables Z_n have absolutely continuous distributions with finite entropies $h(Z_n)$, then these entropies converge,

$$h(Z_n) \to h(Z) \qquad \text{as } n \to \infty. \tag{1.1}$$

This theorem is due to Barron [3]. Some weaker variants of the theorem in case of regularized distributions were known before; they go back to the work of Linnik [16], initiating an information-theoretic approach to the central limit theorem.

To clarify in which sense (1.1) is strong, recall that, if a random variable X with finite second moment has a density $p(x)$, its entropy

$$h(X) = -\int_{-\infty}^{+\infty} p(x) \log p(x)\, dx$$

is well defined and is bounded from above by the entropy of the normal random variable Z, having the same mean a and the same variance σ^2 as X. Note that the value $h(X) = -\infty$ is possible. The relative entropy

$$D(X) = D(X\|Z) = h(Z) - h(X) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{\varphi_{a,\sigma}(x)}\, dx,$$

Received April 2011; revised May 2012.

[1]Supported in part by NSF Grant DMS-11-06530 and SFB 701.

MSC2010 subject classifications. 60E.

Key words and phrases. Entropy, entropic distance, central limit theorem, Edgeworth-type expansions.

where $\varphi_{a,\sigma}$ stands for the density of Z, is nonnegative and serves as kind of a distance to the class of normal laws, or to Gaussianity. This quantity does not depend on the mean or the variance of X, and can be related to the total variation distance between the distributions of X and Z by virtue of the Pinsker-type inequality $D(X) \geq \frac{1}{2}\|F_X - F_Z\|_{\mathrm{TV}}^2$. This already shows that the entropic convergence (1.1) is stronger than convergence in the total variation norm.

Thus, the entropic central limit theorem may be reformulated as $D(Z_n) \to 0$, as long as $D(Z_{n_0}) < +\infty$ for some n_0. This property itself gives rise to a number of intriguing questions, such as to the type and the rate of convergence. In particular, it has been proved only recently that the sequence $h(Z_n)$ is nondecreasing, so that $D(Z_n) \downarrow 0$; cf. [1, 17]. This leads to the question as to the precise rate of $D(Z_n)$ tending to zero; however, not much seems to be known about this problem. The best results in this direction are due to Artstein et al. [2] and to Barron and Johnson [15]. In the i.i.d. case as above, these authors have obtained an expected asymptotic bound $D(Z_n) = O(1/n)$ under the hypothesis that the distribution of X_1 admits an analytic inequality of Poincaré-type (in [15], a restricted Poincaré inequality is used). These inequalities involve a large variety of "nice" probability distributions which necessarily have a finite exponential moment.

The aim of this paper is to study the rate of $D(Z_n)$, using moment conditions $\mathbf{E}|X_1|^s < +\infty$ with fixed values $s \geq 2$, which are comparable to those required for classical Edgeworth-type approximations in the Kolmogorov distance. The cumulants

$$\gamma_r = i^{-r} \frac{d^r}{dt^r} \log \mathbf{E} e^{itX_1}\Big|_{t=0}$$

are then well defined for all $r \leq [s]$ (the integer part of s), and one may introduce the functions

$$q_k(x) = \varphi(x) \sum H_{k+2j}(x) \frac{1}{r_1! \cdots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \cdots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k} \tag{1.2}$$

involving the Chebyshev–Hermite polynomials H_k. The summation in (1.2) runs over all nonnegative integer solutions $(r_1, \dots, r_k)$ to the equation $r_1 + 2r_2 + \cdots + kr_k = k$, and one uses the notation $j = r_1 + \cdots + r_k$.

The functions q_k are defined for $k = 1, \dots, [s] - 2$. They appear in Edgeworth-type expansions including the local limit theorem, where q_k are used to construct the approximation of the densities of Z_n. These results can be applied to obtain an expansion in powers of $1/n$ for the distance $D(Z_n)$. For a multidimensional version of the following Theorem 1.1 for moments of integer order $s \geq 2$, see Theorem 6.1 below.

THEOREM 1.1. *Let $\mathbf{E}|X_1|^s < +\infty$ $(s \geq 2)$, and assume $D(Z_{n_0}) < +\infty$, for some n_0. Then*

$$D(Z_n) = \frac{c_1}{n} + \frac{c_2}{n^2} + \cdots + \frac{c_{[(s-2)/2]}}{n^{[(s-2)/2]}} + o\big((n \log n)^{-(s-2)/2}\big). \tag{1.3}$$

Here

$$c_j = \sum_{k=2}^{2j} \frac{(-1)^k}{k(k-1)} \sum \int_{-\infty}^{+\infty} q_{r_1}(x) \cdots q_{r_k}(x) \, \frac{dx}{\varphi(x)^{k-1}}, \tag{1.4}$$

where the summation runs over all positive integers $(r_1, \dots, r_k)$ *such that* $r_1 + \cdots + r_k = 2j$.

Each coefficient c_j in (1.3) represents a certain polynomial in the cumulants $\gamma_3, \dots, \gamma_{2j+1}$. For example, $c_1 = \frac{1}{12}\gamma_3^2$, and in the case $s = 4$, (1.3) gives

$$D(Z_n) = \frac{1}{12n}(\mathbf{E}X_1^3)^2 + o\left(\frac{1}{n \log n}\right) \qquad (\mathbf{E}X_1^4 < +\infty). \tag{1.5}$$

Thus, under the 4th moment condition, we have $D(Z_n) \le \frac{C}{n}$, where the constant depends on the underlying distribution. This has been conjectured by Johnson [14], page 49. Actually, the constant C may be expressed in terms of $\mathbf{E}X_1^4$ and $D(X_1)$, only.

When s varies in the range $4 \le s \le 6$, the leading linear term in (1.5) will be unchanged, while the remainder term improves and satisfies $O(\frac{1}{n^2})$ in case $\mathbf{E}X_1^6 < +\infty$. But for $s = 6$, the result involves the subsequent coefficient c_2 which depends on γ_3, γ_4 and γ_5. In particular, if $\gamma_3 = 0$, we have $c_2 = \frac{1}{48}\gamma_4^2$, thus

$$D(Z_n) = \frac{1}{48n^2}(\mathbf{E}X_1^4 - 3)^2 + o\left(\frac{1}{(n \log n)^2}\right) \qquad (\mathbf{E}X_1^3 = 0, \mathbf{E}X_1^6 < +\infty).$$

More generally, representation (1.3) simplifies if the first $k-1$ moments of X_1 coincide with the corresponding moments of $Z \sim N(0, 1)$.

COROLLARY 1.2. *Let* $\mathbf{E}|X_1|^s < +\infty$ $(s \ge 4)$, *and assume that* $D(Z_{n_0}) < +\infty$, *for some* n_0. *Given* $k = 3, 4, \dots, [s]$, *assume that* $\gamma_j = 0$ *for all* $3 \le j < k$. *Then*

$$D(Z_n) = \frac{\gamma_k^2}{2k!} \cdot \frac{1}{n^{k-2}} + O\left(\frac{1}{n^{k-1}}\right) + o\left(\frac{1}{(n \log n)^{(s-2)/2}}\right). \tag{1.6}$$

Johnson had noticed (though in terms of the standardized Fisher information, see [14], Lemma 2.12) that if $\gamma_k \ne 0$, $D(Z_n)$ cannot be of smaller order than $n^{-(k-2)}$.

Note that when $\mathbf{E}X_1^{2k} < +\infty$, the o-term may be removed in the representation (1.6). On the other hand, when $k > \frac{s+2}{2}$, the o-term will dominate the $n^{-(k-2)}$-term, and we can only conclude that $D(Z_n) = o((n \log n)^{-(s-2)/2})$.

As for the missing range $2 \le s < 4$, here there are no coefficients c_j appearing in the sum (1.3), and Theorem 1.1 just tells us that

$$D(Z_n) = o\left(\frac{1}{(n \log n)^{(s-2)/2}}\right). \tag{1.7}$$

This bound is worse than the rate $1/n$. In particular, it only gives $D(Z_n) = o(1)$ for $s = 2$, which is the statement of Barron's theorem. In fact, in this case the entropic distance to normality may decay to zero at an arbitrarily slow rate. In case of a finite 3rd absolute moment, $D(Z_n) = o(\frac{1}{\sqrt{n \log n}})$. To see that this and that the more general relation (1.7) cannot be improved with respect to the powers of $1/n$, we prove:

THEOREM 1.3. *Let $\eta > 1$. Given $2 < s < 4$, there exists a sequence of independent, identically distributed random variables $(X_n)_{n \geq 1}$ with $\mathbf{E}|X_1|^s < +\infty$, such that $D(X_1) < +\infty$ and*

$$D(Z_n) \geq \frac{c}{(n \log n)^{(s-2)/2} (\log n)^{\eta}}, \qquad n \geq n_1(X_1),$$

with a constant $c = c(\eta, s) > 0$, depending on η and s, only.

Known bounds on the entropy are commonly based on Bruijn's identity which may be used to represent the entropic distance to normality as an integral of the Fisher information for regularized distributions; cf. [3]. However, it is not clear how to reach exact asymptotics with this approach. The proofs of Theorems 1.1 and 1.3 stated above rely upon classical tools and results in the theory of sums of independent summands including Edgeworth-type expansions for convolution of densities formulated as local limit theorems with nonuniform remainder bounds. For noninteger values of s, the authors had to complete the otherwise extensive literature by recent, technically rather involved results based on fractional differential calculus; see [6, 7]. Our approach applies to random variables in higher dimension as well and to nonidentical distributions for summands with uniformly bounded sth moments.

We start with the description of a truncation-of-density argument, which allows us to reduce many questions about bounding the entropic distance to the case of bounded densities (Section 2). In Section 3 we discuss known results about Edgeworth-type expansions that will be used in the proof of Theorem 1.1. Main steps of the proofs are based on it in Sections 4 and 5. All auxiliary results cover the scheme of i.i.d. random vectors in $\mathbf{R}^d$ as well (however, with integer values of s) and are finalized in Section 6 to obtain multidimensional variants of Theorem 1.1 and Corollary 1.2. Sections 7 and 8 are devoted to lower bounds on the entropic distance to normality for a special class of probability distributions on the real line that are used in the proof Theorem 1.3.

2. Binomial decomposition of convolutions. First let us comment on the assumptions in Theorem 1.1. It may happen that X_1 has a singular distribution, but the distribution of $X_1 + X_2$ and of all next sums $S_n = X_1 + \cdots + X_n$ $(n \geq 2)$ are absolutely continuous; cf. [25].

If it exists, the density p of X_1 may or may not be bounded. In the first case, all the entropies $h(S_n)$ are finite. If p is unbounded, it may happen that all $h(S_n)$ are infinite, even if p is compactly supported. But if $h(S_n)$ is finite for some $n = n_0$ then, for all $n \geq n_0$, entropies are finite; see [3] for specific examples.

Denote by $p_n(x)$ the density of $Z_n = S_n/\sqrt{n}$ (when it exists). Since it is desirable to work with bounded densities, we will slightly modify p_n at the expense of a small change in the entropy. Variants of the next construction are well known; see, for example, [13, 23], where the central limit theorem was studied with respect to the total variation distance. Without any extra efforts, we may assume that X_n take values in $\mathbf{R}^d$ which we equip with the usual inner product $\langle \cdot, \cdot \rangle$ and the Euclidean norm $|\cdot|$. For simplicity, we describe the construction in the situation, where X_1 has a density $p(x)$; cf. Remark 2.5 on appropriate modifications in the general case.

Let $m_0 \geq 0$ be a fixed integer. (For the purposes of Theorem 1.1, one may take $m_0 = [s] + 1$.)

If p is bounded, we put $\widetilde{p}_n(x) = p_n(x)$ for all $n \geq 1$. Otherwise, the integral

$$b = \int_{p(x)>M} p(x)\,dx \tag{2.1}$$

is positive for all $M > 0$. Choose M to be sufficiently large to satisfy, for example, $0 < b < \frac{1}{2}$; cf. Remark 2.4. In this case (when p is unbounded), consider the decomposition

$$p(x) = (1-b)\rho_1(x) + b\rho_2(x), \tag{2.2}$$

where ρ_1, ρ_2 are the normalized restrictions of p to the sets $\{p(x) \leq M\}$ and $\{p(x) > M\}$, respectively. Hence, for the convolutions we have a binomial decomposition

$$p^{*n} = \sum_{k=0}^{n} C_n^k (1-b)^k b^{n-k} \rho_1^{*k} * \rho_2^{*(n-k)}.$$

For $n \geq m_0 + 1$, we split the above sum into the two parts, so that $p^{*n} = \rho_{n1} + \rho_{n2}$ with

$$\rho_{n1} = \sum_{k=m_0+1}^{n} C_n^k (1-b)^k b^{n-k} \rho_1^{*k} * \rho_2^{*(n-k)},$$

$$\rho_{n2} = \sum_{k=0}^{m_0} C_n^k (1-b)^k b^{n-k} \rho_1^{*k} * \rho_2^{*(n-k)}.$$

Note that, whenever $b < b_1 < \frac{1}{2}$,

$$\begin{aligned} \varepsilon_n \equiv \int \rho_{n2}(x)\,dx &= \sum_{k=0}^{m_0} C_n^k (1-b)^k b^{n-k} \\ &\leq n^{m_0} b^{n-m_0} = o(b_1^n) \qquad \text{as } n \to \infty. \end{aligned} \tag{2.3}$$

Finally define

$$\widetilde{p}_n(x) = p_{n1}(x) = \frac{1}{1-\varepsilon_n} n^{d/2} \rho_{n1}(x\sqrt{n}) \tag{2.4}$$

and similarly $p_{n2}(x) = \frac{1}{\varepsilon_n} n^{d/2} \rho_{n2}(x\sqrt{n})$. Thus, we have the desired decomposition

$$p_n(x) = (1-\varepsilon_n) p_{n1}(x) + \varepsilon_n p_{n2}(x). \tag{2.5}$$

The probability densities $p_{n1}(x)$ are bounded and provide an approximation for $p_n(x) = n^{d/2} p^{*n}(x\sqrt{n})$ in total variation. In particular, from (2.3)–(2.5) it follows that

$$\int \big|p_{n1}(x) - p_n(x)\big|\,dx < 2^{-n}$$

for all n large enough. One of the immediate consequences of this estimate is the bound

$$\big|v_{n1}(t) - v_n(t)\big| < 2^{-n} \qquad \big(t \in \mathbf{R}^d\big) \tag{2.6}$$

for the characteristic functions $v_n(t) = \int e^{i\langle t,x\rangle} p_n(x)\,dx$ and $v_{n1}(t) = \int e^{i\langle t,x\rangle} \times p_{n1}(x)\,dx$, corresponding to the densities p_n and p_{n1}.

This property may be sharpened in case of finite moments.

LEMMA 2.1. *If* $\mathbf{E}|X_1|^s < +\infty$ $(s \geq 0)$, *then for all n large enough,*

$$\int \big(1+|x|^s\big)\big|\widetilde{p}_n(x) - p_n(x)\big|\,dx < 2^{-n}.$$

In particular, (2.6) *also holds for all partial derivatives of v_{n1} and v_n up to order $m = [s]$.*

PROOF. By definition (2.5), $|p_{n1}(x) - p_n(x)| \leq \varepsilon_n (p_{n1}(x) + p_{n2}(x))$, hence

$$\begin{aligned}\int |x|^s \big|p_{n1}(x) - p_n(x)\big|\,dx &\leq \frac{\varepsilon_n}{1-\varepsilon_n} n^{-s/2} \int |x|^s \rho_{n1}(x)\,dx \\ &\quad + n^{-s/2} \int |x|^s \rho_{n2}(x)\,dx.\end{aligned}$$

Let $U_1, U_2, \dots$ be independent copies of U and $V_1, V_2, \dots$ be independent copies of V (that are also independent of U_n's), where U and V are random vectors with densities ρ_1 and ρ_2, respectively. From (2.2)

$$\beta_s \equiv \mathbf{E}|X_1|^s = (1-b)\mathbf{E}|U|^s + b\mathbf{E}|V|^s,$$

so $\mathbf{E}|U|^s \leq \beta_s/b$ and $\mathbf{E}|V|^s \leq \beta_s/b$ (using $b < \frac{1}{2}$). Therefore, for the normalized sums

$$R_{k,n} = \frac{1}{\sqrt{n}}(U_1 + \cdots + U_k + V_1 + \cdots + V_{n-k}), \qquad 0 \leq k \leq n,$$

we have $\mathbf{E}|R_{k,n}|^s \leq \frac{\beta_s}{b} n^{s/2}$, if $s \geq 1$, and $\mathbf{E}|R_{k,n}|^s \leq \frac{\beta_s}{b} n^{1-(s/2)}$, if $0 \leq s \leq 1$. Hence, by the definition of ρ_{n1} and ρ_{n2},

$$\int |x|^s \rho_{n1}(x)\,dx = n^{s/2} \sum_{k=m_0+1}^{n} C_n^k (1-b)^k b^{n-k} \mathbf{E}|R_{k,n}|^s \leq \frac{\beta_s}{b} n^{s+1},$$

$$\int |x|^s \rho_{n2}(x)\,dx = n^{s/2} \sum_{k=0}^{m_0} C_n^k (1-b)^k b^{n-k} \mathbf{E}|R_{k,n}|^s \leq \frac{\beta_s}{b} n^{s+1} \varepsilon_n.$$

It remains to apply estimate (2.3) on ε_n, and Lemma 2.1 follows. $\square$

We need to extend the assertion of Lemma 2.1 to the relative entropies with respect to the standard normal distribution on $\mathbf{R}^d$ with density $\varphi(x) = (2\pi)^{-d/2} e^{-|x|^2/2}$. Thus put

$$D_n = \int p_n(x) \log \frac{p_n(x)}{\varphi(x)}\,dx, \qquad \widetilde{D}_n = \int \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx.$$

LEMMA 2.2. *If X_1 has a finite second moment and finite entropy, then $|\widetilde{D}_n - D_n| < 2^{-n}$, for all n large enough.*

First, we collect a few elementary properties of the convex function $L(u) = u \log u$ ($u \geq 0$).

LEMMA 2.3. *For all $u, v \geq 0$ and $0 \leq \varepsilon \leq 1$:*

(a) $L((1-\varepsilon)u + \varepsilon v) \leq (1-\varepsilon)L(u) + \varepsilon L(v)$;
(b) $L((1-\varepsilon)u + \varepsilon v) \geq (1-\varepsilon)L(u) + \varepsilon L(v) + uL(1-\varepsilon) + vL(\varepsilon)$;
(c) $L((1-\varepsilon)u + \varepsilon v) \geq (1-\varepsilon)L(u) - \frac{1}{e}u - \frac{1}{e}$.

The first assertion is just Jensen's inequality applied to L. By the convexity of L, for each $y \geq 0$, the function $L(x+y) - L(x)$ is increasing in $x \geq 0$. Hence, $L(x+y) - L(x) \geq L(y)$, which is (b) for $x = (1-\varepsilon)u$ and $y = \varepsilon v$. Similarly, using $L \geq -\frac{1}{e}$, we obtain (c).

PROOF OF LEMMA 2.2. Assuming that p is (essentially) unbounded, define

$$D_{nj} = \int p_{nj}(x) \log \frac{p_{nj}(x)}{\varphi(x)}\,dx \qquad (j = 1, 2),$$

so that $\widetilde{D}_n = D_{n,1}$. By Lemma 2.3(a), $D_n \leq (1-\varepsilon_n)D_{n1} + \varepsilon_n D_{n2}$. On the other hand, by (b),

$$D_n \geq \big((1-\varepsilon_n)D_{n1} + \varepsilon_n D_{n2}\big) + \varepsilon_n \log \varepsilon_n + (1-\varepsilon_n)\log(1-\varepsilon_n).$$

In view of (2.3), the two estimates give

$$|D_{n1} - D_n| < C(n + D_{n1} + D_{n2})b_1^n, \tag{2.7}$$

which holds for all $n \geq 1$ with some constant C. In addition, by the inequality in (c) with $\varepsilon = b$, from (2.2) it follows that

$$D(X_1 \| Z) = \int L\biggl(\frac{p(x)}{\varphi(x)}\biggr)\varphi(x)\,dx \geq (1-b)\int \rho_1(x)\log\frac{\rho_1(x)}{\varphi(x)}\,dx - \frac{2}{e}, \tag{2.8}$$

where Z denotes a standard normal random vector in $\mathbf{R}^d$. By the same reasoning,

$$D(X_1 \| Z) \geq b\int \rho_2(x)\log\frac{\rho_2(x)}{\varphi(x)}\,dx - \frac{2}{e}. \tag{2.9}$$

Now, by the convexity of the function $L(u) = u\log u$,

$$D_{n1} \leq \frac{1}{1-\varepsilon_n}\sum_{k=m_0+1}^{n} C_n^k(1-b)^k b^{n-k}\int r_{k,n}(x)\log\frac{r_{k,n}(x)}{\varphi(x)}\,dx,$$

$$D_{n2} \leq \frac{1}{\varepsilon_n}\sum_{k=0}^{m_0} C_n^k(1-b)^k b^{n-k}\int r_{k,n}(x)\log\frac{r_{k,n}(x)}{\varphi(x)}\,dx,$$

where $r_{k,n}$ are densities of the normalized sums $R_{k,n}$ from the proof of Lemma 2.1. Here each integral may also be written as

$$\int r_{k,n}(x)\log\frac{r_{k,n}(x)}{\varphi(x)}\,dx = \int L\bigl(r_{k,n}(x)\bigr)\,dx + \frac{d}{2}\log(2\pi) + \frac{1}{2}\mathbf{E}|R_{k,n}|^2. \tag{2.10}$$

We have $\mathbf{E}|R_{k,n}|^2 \leq \frac{\beta^2}{b}n$, as noticed in the proof of Lemma 2.1. In addition, by the convexity of L, there is a general inequality

$$\int L\bigl((f*g)(x)\bigr)\,dx \leq \int L\bigl(f(x)\bigr)\,dx$$

valid for the convolution of any two probability densities f and g on $\mathbf{R}^d$ (if the integrals exist). In particular,

$$\int L\bigl(r_{k,n}(x)\bigr)\,dx \leq \frac{d}{2}\log n + \max\biggl\{\int L\bigl(\rho_1(x)\bigr)\,dx, \int L\bigl(\rho_2(x)\bigr)\,dx\biggr\},$$

which may actually be sharpened in case $1 < k < n$ by replacing max with min. By (2.8) and (2.9), the integrals on the right-hand side are finite, thus the integrals on the left-hand side of (2.10) are bounded by Cn with some constant C. Hence, a similar bound also holds for D_{nj}, and it remains to apply (2.7). Lemma 2.2 is proved. $\square$

REMARK 2.4. If X_1 has a finite second moment and $D(X_1) < +\infty$, the truncation level M in (2.1) can be chosen explicitly in terms of b using the entropic distance $D(X_1)$ and $\sigma^2 = \det(\Sigma)$, where Σ is the covariance matrix of X_1.

Indeed, putting $a = \mathbf{E}X_1$ and using an elementary inequality $t\log(1+t) \le t\log t + 1$ ($t \ge 0$), we have an upper estimate

$$\int p\log\Big(1+\frac{p}{\varphi_{a,\Sigma}}\Big)\,dx = \int \frac{p}{\varphi_{a,\Sigma}}\log\Big(1+\frac{p}{\varphi_{a,\Sigma}}\Big)\varphi_{a,\Sigma}\,dx$$
$$\le \int p\log\frac{p}{\varphi_{a,\Sigma}}\,dx + 1 = D(X_1)+1.$$

On the other hand, the original expression majorizes

$$\int_{\{p(x)>M\}} p(x)\log\frac{M}{\varphi_{a,\Sigma}(x)}\,dx \ge b\log\big(M\sigma(2\pi)^{d/2}\big),$$

hence

$$M \le \frac{1}{\sigma(2\pi)^{d/2}}e^{(D(X_1)+1)/b}.$$

REMARK 2.5. If Z_n have absolutely continuous distributions with finite entropies for $n \ge n_0 > 1$, the above construction should be properly modified.

Namely, one may put $\widetilde{p}_n = p_n$, if p_n are bounded, and otherwise apply the same decomposition (2.2) to p_{n_0} in place of p. As a result, for any $n = An_0 + B$ ($A \ge 1$, $0 \le B \le n_0 - 1$), the partial sum S_n will have the density

$$r_n(x) = \sum_{k=0}^{A} C_A^k(1-b)^k b^{A-k}\int\big(\rho_1^{*k} * \rho_2^{*(A-k)}\big)(x-y)\,dF_B(y),$$

where F_B is the distribution of S_B. For $A \ge m_0 + 1$, split the above sum into the two parts with summation over $m_0 + 1 \le k \le A$ and $0 \le k \le m_0$, respectively, so that $r_n = \rho_{n1} + \rho_{n2}$. Then, like in (2.4) and for the same sequence ε_n described in (2.3), define

$$\widetilde{p}_n(x) = \frac{1}{1-\varepsilon_n}n^{d/2}\rho_{n1}(x\sqrt{n}).$$

Clearly, these densities are bounded and approximate $p_n(x)$ in total variation. In particular, for all sufficiently large n, they satisfy the estimates that are similar to the estimates in Lemmas 2.1 and 2.2.

3. Edgeworth-type expansions. Let $(X_n)_{n\ge1}$ be independent, identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\mathrm{Var}(X_1) = 1$. In this section we collect some auxiliary results about Edgeworth-type expansions both for the distribution functions $F_n(x) = \mathbf{P}\{Z_n \le x\}$ and the densities $p_n(x)$ of the normalized sums $Z_n = S_n/\sqrt{n}$, where $S_n = X_1 + \cdots + X_n$.

If the absolute moment $\mathbf{E}|X_1|^s$ is finite for a given $s \geq 2$ and $m = [s]$, define

$$\varphi_m(x) = \varphi(x) + \sum_{k=1}^{m-2} q_k(x) n^{-k/2} \tag{3.1}$$

with the functions q_k described in (1.2). Introduce as well

$$\Phi_m(x) = \int_{-\infty}^{x} \varphi_m(y)\, dy = \Phi(x) + \sum_{k=1}^{m-2} Q_k(x) n^{-k/2}. \tag{3.2}$$

Similar to (1.2), the functions Q_k have an explicit description involving the cumulants $\gamma_3, \dots, \gamma_{k+2}$ of X_1. Namely,

$$Q_k(x) = -\varphi(x) \sum H_{k+2j-1}(x) \frac{1}{r_1! \cdots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \cdots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k},$$

where the summation is carried out over all nonnegative integer solutions $(r_1, \dots, r_k)$ to the equation $r_1 + 2r_2 + \cdots + kr_k = k$ with $j = r_1 + \cdots + r_k$; cf., for example, [4] or [21] for details.

THEOREM 3.1. *Assume that* $\limsup_{|t|\to+\infty} |\mathbf{E}e^{itX_1}| < 1$. *If* $\mathbf{E}|X_1|^s < +\infty$ $(s \geq 2)$, *then as* $n \to \infty$, *uniformly for all* x,

$$(1 + |x|^s)(F_n(x) - \Phi_m(x)) = o(n^{-(s-2)/2}). \tag{3.3}$$

For $2 \leq s < 3$ and $m = 2$, there are no expansion terms in the sum (3.2), and hence $\Phi_2(x) = \Phi(x)$ is the distribution function of the standard normal law. In this case, (3.3) becomes

$$(1 + |x|^s)(F_n(x) - \Phi(x)) = o(n^{-(s-2)/2}). \tag{3.4}$$

In fact, in this case Cramer's condition on the characteristic function of X_1 is not used. The result was obtained by Osipov and Petrov [19]; cf. also [5] where (3.4) is established with O.

In the case $s \geq 3$ Theorem 3.1 can be found in [21] (Theorem 2, Chapter VI, page 168). Note that when $s = m$ is integer, relation (3.3) without the factor $1 + |x|^m$ represents the classical Edgeworth expansion. It is essentially due to Cramér and is described in many papers and textbooks; cf. [9, 10]. However, the case of fractional values of s is more delicate, especially in the following local limit theorem.

THEOREM 3.2. *Let* $\mathbf{E}|X_1|^s < +\infty$ $(s \geq 2)$. *Suppose* Z_{n_0} *has a bounded density for some* n_0. *Then for all sufficiently large* n, *the random variables* Z_n *have continuous bounded densities* p_n *satisfying, as* $n \to \infty$,

$$(1 + |x|^m)(p_n(x) - \varphi_m(x)) = o(n^{-(s-2)/2}) \tag{3.5}$$

uniformly for all x. Moreover,

$$(1+|x|^s)\big(p_n(x)-\varphi_m(x)\big) = o\big(n^{-(s-2)/2}\big) + \big(1+|x|^{s-m}\big)\big(O\big(n^{-(m-1)/2}\big)+o\big(n^{-(s-2)}\big)\big). \tag{3.6}$$

If $s=m$ is integer and $m \geq 3$, Theorem 3.2 is well known; then (3.5) and (3.6) simplify to

$$\big(1+|x|^m\big)\big(p_n(x)-\varphi_m(x)\big) = o\big(n^{-(m-2)/2}\big). \tag{3.7}$$

In this formulation the result is due to Petrov [20]; cf. [21], page 211, or [4], page 192. Without the term $1+|x|^m$, relation (3.7) goes back to the results of Cramér and Gnedenko (cf. [11]).

In the general (fractional) case, Theorem 3.2 has recently been obtained in [6, 7] by using the technique of Liouville fractional integrals and derivatives. Assertion (3.6) gives an improvement over (3.5) on relatively large intervals of the real axis, and this is essential in the case of noninteger s.

An obvious weak point in Theorem 3.2 is that it requires the boundedness of the densities p_n, which is, however, necessary for conclusions, such as (3.5) or (3.7). Nevertheless, this condition may be removed, if we replace p_n by slightly modified densities $\widetilde{p}_n$.

THEOREM 3.3. *Let $\mathbf{E}|X_1|^s < +\infty$ ($s \geq 2$). Suppose that, for all for all sufficiently large n, Z_n have absolutely continuous distributions with densities p_n. Then there exist some bounded continuous densities $\widetilde{p}_n$ such that*:

(a) *the relations (3.5) and (3.6) hold true for $\widetilde{p}_n$ instead of p_n*;
(b) $\int_{-\infty}^{+\infty}(1+|x|^s)|\widetilde{p}_n(x)-p_n(x)|\,dx < 2^{-n}$, *for all sufficiently large n*;
(c) $\widetilde{p}_n(x) = p_n(x)$ *almost everywhere, if p_n is bounded (a.e.)*.

Here, property (c) is added to include Theorem 3.2 in Theorem 3.3 as a particular case. Moreover, one can use the densities $\widetilde{p}_n$ constructed in the previous section with $m_0 = [s]+1$. We refer to [6, 7] for detailed proofs.

This extended result allows us to immediately recover, for example, the central limit theorem with respect to the total variation distance (without the assumption of boundedness of p_n). Namely, we have

$$\|F_n - \Phi_m\|_{\mathrm{TV}} = \int_{-\infty}^{+\infty} \big|p_n(x)-\varphi_m(x)\big|\,dx = o\big(n^{-(s-2)/2}\big). \tag{3.8}$$

For $s=2$ and $\varphi_2(x)=\varphi(x)$, this statement corresponds to a theorem of Prokhorov [22], while for $s=3$ and $\varphi_3(x)=\varphi(x)(1+\gamma_3\frac{x^3-3x}{6\sqrt{n}})$—to the result of Sirazhdinov and Mamatov [23].

The multidimensional case. Similar results are also available in the multidimensional case for integer values $s = m$. In the remaining part of this section, let $(X_n)_{n\geq 1}$ denote independent identically distributed random vectors in the Euclidean space $\mathbf{R}^d$ with mean zero and identity covariance matrix.

Assuming $\mathbf{E}|X_1|^m < +\infty$ for some integer $m \geq 2$ (where now $|\cdot|$ denotes the Euclidean norm), introduce the cumulants γ_ν of X_1 and the associated cumulant polynomials $\gamma_k(it)$ up to order m by using the equality

$$\frac{1}{k!}\frac{d^k}{du^k}\log \mathbf{E}e^{iu\langle t,X_1\rangle}\bigg|_{u=0} = \frac{1}{k!}\gamma_k(it) = \sum_{|\nu|=k}\gamma_\nu \frac{(it)^\nu}{\nu!} \qquad (k = 1,\dots,m,\ t \in \mathbf{R}^d).$$

Here the summation runs over all d-tuples $\nu = (\nu_1,\dots,\nu_d)$ with integer components $\nu_j \geq 0$ such that $|\nu| = \nu_1 + \cdots + \nu_d = k$. We also write $\nu! = \nu_1!\cdots\nu_d!$ and use a standard notation for the generalized powers $z^\nu = z_1^{\nu_1}\cdots z_d^{\nu_d}$ of real or complex vectors $z = (z_1,\dots,z_d)$, which are treated as polynomials in z of degree $|\nu|$.

For $1 \leq k \leq m-2$, define the polynomials

$$P_k(it) = \sum_{r_1+2r_2+\cdots+kr_k=k} \frac{1}{r_1!\cdots r_k!}\left(\frac{\gamma_3(it)}{3!}\right)^{r_1}\cdots\left(\frac{\gamma_{k+2}(it)}{(k+2)!}\right)^{r_k}, \tag{3.9}$$

where the summation is performed over all nonnegative integer solutions $(r_1,\dots,r_k)$ to the equation $r_1 + 2r_2 + \cdots + kr_k = k$.

Furthermore, like in dimension one, define the approximating functions $\varphi_m(x)$ on $\mathbf{R}^d$ by virtue of the equality (3.1), where every q_k is determined by its Fourier transform

$$\int e^{i\langle t,x\rangle} q_k(x)\,dx = P_k(it)e^{-|t|^2/2}. \tag{3.10}$$

If Z_{n_0} has a bounded density for some n_0, then for all sufficiently large n, Z_n have continuous bounded densities p_n satisfying (3.7); see [4], Theorem 19.2. We need an extension of this theorem to the case of unbounded densities, as well as integral variants such as (3.8). The first assertion (3.11) in the next theorem is similar to the one-dimensional Theorem 3.3 in the case where $s = m$ is integer; cf. (3.5). For the proof (which we omit), one may apply Lemma 2.1 and follow the standard arguments from [4], Chapter 4.

THEOREM 3.4. *Suppose that $\mathbf{E}|X_1|^m < +\infty$ with some integer $m \geq 2$. If, for all sufficiently large n, Z_n have densities p_n, then the densities $\widetilde{p}_n$ introduced in Section 2 with $m_0 = m+1$ satisfy*

$$(1+|x|^m)\big(\widetilde{p}_n(x) - \varphi_m(x)\big) = o\big(n^{-(m-2)/2}\big) \tag{3.11}$$

uniformly for all x. In addition,

$$\int (1+|x|^m)\big|\widetilde{p}_n(x) - \varphi_m(x)\big|\,dx = o\big(n^{-(m-2)/2}\big). \tag{3.12}$$

The second assertion is Theorem 19.5 in [4], where it is stated for $m \geq 3$ under a slightly weaker hypothesis that X_1 has a nonzero absolutely continuous component. Note that, by Lemma 2.1, it does not matter whether $\widetilde{p}_n$ or p_n are used in (3.12).

4. Entropic distance to normality and moderate deviations. Let $X_1, X_2, \dots$ be independent, identically distributed random vectors in $\mathbf{R}^d$ with mean zero, identity covariance matrix and such that $D(Z_n) < +\infty$, for all n large enough.

According to Lemma 2.2 and Remark 2.5, up to an error at most 2^{-n} for sufficiently large n, the entropic distance to normality, $D_n = D(Z_n)$, is equal to the relative entropy

$$\widetilde{D}_n = \int \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx,$$

where φ is the density of a standard normal random vector Z in $\mathbf{R}^d$.

Given $T \geq 1$, split the integral into two parts by writing

$$\widetilde{D}_n = \int_{|x| \leq T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx + \int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx. \tag{4.1}$$

By Theorems 3.3 and 3.4, $\widetilde{p}_n$ are uniformly bounded, that is, $\widetilde{p}_n(x) \leq M$, for all $x \in \mathbf{R}^d$ and $n \geq 1$ with some constant M. Hence, the second integral in (4.1) may be treated by virtue of moderate deviations results (when T is not too large). Indeed, since $T \geq 1$,

$$\int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx \leq \int_{|x|>T} \widetilde{p}_n(x) \log \frac{M}{\varphi(x)}\, dx \leq C \int_{|x|>T} |x|^2 \widetilde{p}_n(x)\, dx,$$

where $C = \frac{1}{2} + \log(1 + M(2\pi)^{d/2})$. One the other hand, using $u \log u \geq u - 1$, we have a lower bound

$$\int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx \geq \int_{|x|>T} \big(\widetilde{p}_n(x) - \varphi(x)\big)\, dx \geq -\mathbf{P}\{|Z| > T\}.$$

The two estimates give

$$\left| \int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx \right| \leq \mathbf{P}\{|Z| > T\} + C \int_{|x|>T} |x|^2 \widetilde{p}_n(x)\, dx. \tag{4.2}$$

This is a very general upper bound, valid for any probability density $\widetilde{p}_n$ on $\mathbf{R}^d$, bounded by a constant M (with C as above).

Following (4.1), we are faced with two analytic problems. The first one is to give a sharp estimate of $\widetilde{p}_n(x) - \varphi(x)$ on a relatively large Euclidean ball $|x| \leq T$. Clearly, T has to be small enough, so that results like local limit theorems, such as Theorems 3.2–3.4 may be applied. The second problem is to give a sharp upper bound of the last integral in (4.2). To this aim, we need moderate deviations inequalities, so that Theorems 3.1 and 3.4 are applicable. Anyway, in order to use both types of results we are forced to choose T from a very narrow window only. This value turns out to be approximately

$$T_n = \sqrt{(s-2)\log n + s\log\log n + \rho_n} \qquad (s>2), \tag{4.3}$$

where $\rho_n \to +\infty$ is a sufficiently slowly growing sequence (whose growth will be restricted by the decay of the n-dependent constants in o-expressions of Theorems 3.2–3.4). In the case $s=2$, one may put $T_n = \sqrt{\rho_n}$ such that $T_n \to +\infty$ is a sufficiently slowly growing sequence.

LEMMA 4.1 (The case $d=1$ and s real). *If* $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 = 1, \mathbf{E}|X_1|^s < +\infty$ $(s \geq 2)$, *then*

$$\int_{|x|>T_n} x^2 \tilde{p}_n(x)\,dx = o\big((n\log n)^{-(s-2)/2}\big). \tag{4.4}$$

LEMMA 4.2 (The case $d \geq 2$ and s integer). *If X_1 has mean zero and identity covariance matrix, and* $\mathbf{E}|X_1|^m < +\infty$, *then*

$$\int_{|x|>T_n} x^2 \tilde{p}_n(x)\,dx = o\big(n^{-(m-2)/2}(\log n)^{-(m-d)/2}\big) \qquad (m \geq 3) \tag{4.5}$$

and $\int_{|x|>T_n} x^2 \tilde{p}_n(x)\,dx = o(1)$ *in the case* $m=2$.

Note that plenty of results and techniques concerning moderate deviations have been developed by now. Useful estimates can be found, for example, in [12]. Restricting ourselves to integer values of $s=m$, one may argue as follows.

PROOF OF LEMMA 4.2. Given $T \geq 1$, write

$$\begin{aligned}\int_{|x|>T} |x|^2 \tilde{p}_n(x)\,dx &\leq \frac{1}{T^{m-2}}\int |x|^m \tilde{p}_n(x)\,dx \\ &\leq \frac{1}{T^{m-2}}\int |x|^m \big|\tilde{p}_n(x) - \varphi_m(x)\big|\,dx \\ &\quad + \frac{1}{T^{m-2}}\int_{|x|>T} |x|^m \varphi_m(x)\,dx.\end{aligned} \tag{4.6}$$

By Theorem 3.4 [cf. (3.12)] the first integral in (4.6) is bounded by $o(n^{-(m-2)/2})$.

From the definition of q_k it follows that $q_k(x) = N(x)\varphi(x)$ with some polynomial N of degree at most $3(m-2)$; cf. Section 6 for details. Hence, from (3.1), $\varphi_m(x) \leq 2\varphi(x)$ on the balls of large radii $|x| < n^\delta$ with sufficiently large n (where $0<\delta<\frac{1}{2}$). On the other hand, with some constants C_d, C_d' depending on the dimension only,

$$\int_{|x|>T} |x|^m \varphi(x)\,dx = C_d \int_T^{+\infty} r^{m+d-1} e^{-r^2/2}\,dr \leq C_d' T^{m+d-2} e^{-T^2/2}. \tag{4.7}$$

But for $T = T_n$ and $s = m \geq 3$, we have $e^{-T^2/2} = T^{-m} o(n^{-(m-2)/2})$, so by (4.6) and (4.7),

$$\int_{|x|>T_n} |x|^2 \widetilde{p}_n(x)\,dx \leq C\Big(\frac{1}{T^{m-2}} + \frac{1}{T^{m-d}}\Big) o\big(n^{-(m-2)/2}\big).$$

Since T_n is of order $\sqrt{\log n}$, (4.5) follows. Furthermore, in the case $m = 2$, (4.6) gives the desired relation

$$\int_{|x|>T_n} |x|^2 \widetilde{p}_n(x)\,dx \leq o(1) + \int_{|x|>T_n} |x|^2 \varphi(x)\,dx \to 0 \qquad (n \to \infty). \qquad \square$$

PROOF OF LEMMA 4.1. The above argument also works for $d = 1$, but it can be refined applying Theorem 3.1 for real s. The case $s = 2$ is already covered, so let $s > 2$.

In view of decomposition (2.5), integrating by parts, we have, for any $T \geq 0$,

$$\begin{aligned}(1-\varepsilon_n)&\int_{|x|>T} x^2 \widetilde{p}_n(x)\,dx \\ &\leq \int_{|x|>T} x^2 p_n(x)\,dx = \int_{|x|>T} x^2\,dF_n(x)\end{aligned} \tag{4.8}$$

$$= T^2\big(1 - F_n(T) + F_n(-T)\big) + 2\int_T^{+\infty} x\big(1 - F_n(x) + F_n(-x)\big)\,dx, \tag{4.9}$$

where F_n denotes the distribution function of Z_n. [Note that the first inequality in (4.8) should be just ignored in the case, where p is bounded.]

By (3.3),

$$F_n(x) = \Phi_m(x) + \frac{r_n(x)}{n^{(s-2)/2}} \frac{1}{1+|x|^s}, \qquad r_n = \sup_x |r_n(x)| \to 0 \qquad (n \to \infty).$$

Hence, the first term in (4.9) can be replaced with

$$T^2\big(1 - \Phi_m(T) + \Phi_m(-T)\big) \tag{4.10}$$

at the expense of an error not exceeding (for the values $T \sim \sqrt{\log n}$)

$$\frac{2r_n}{n^{(s-2)/2}} \frac{T^2}{1+T^s} = o\big((n \log n)^{-(s-2)/2}\big). \tag{4.11}$$

Similarly, the integral in (4.9) can be replaced with

$$\int_T^{+\infty} x\big(1 - \Phi_m(x) + \Phi_m(-x)\big)\,dx \tag{4.12}$$

at the expense of an error not exceeding

$$\frac{2r_n}{n^{(s-2)/2}} \int_T^{+\infty} \frac{x\,dx}{1+x^s} = o\big((n \log n)^{-(s-2)/2}\big). \tag{4.13}$$

To explore the behavior of expressions (4.10) and (4.12) for $T = T_n$ using precise asymptotics as in (4.3), recall that, by (3.2),

$$1 - \Phi_m(x) = 1 - \Phi(x) - \sum_{k=1}^{m-2} Q_k(x) n^{-k/2}.$$

Moreover, we note that $Q_k(x) = N_{3k-1}(x)\varphi(x)$, where N_{3k-1} is a polynomial of degree at most $3k - 1$. Thus these functions admit a bound $|Q_k(x)| \leq C_m(1 + |x|^{3m})\varphi(x)$ with some constants C_m (depending on m and the cumulants $\gamma_3, \dots, \gamma_m$ of X_1), which implies with some other constants

$$|1 - \Phi_m(x)| \leq (1 - \Phi(x)) + \frac{C_m(1 + |x|^{3m})}{\sqrt{n}} \varphi(x). \tag{4.14}$$

Hence, using $1 - \Phi(x) < \frac{\varphi(x)}{x}$ ($x > 0$), we get

$$\begin{aligned} T_n^2 |1 - \Phi_m(T_n)| &\leq C T_n^2 (1 - \Phi(T_n)) \leq C T_n e^{-T_n^2/2} \\ &= o((n \log n)^{-(s-2)/2}). \end{aligned} \tag{4.15}$$

A similar bound also holds for $T_n^2 |\Phi_m(-T_n)|$.

Now, we use (4.14) to estimate (4.12) with $T = T_n$ up to a constant by

$$\int_T^\infty x(1 - \Phi(x))\, dx < 1 - \Phi(T) = o((n \log n)^{-(s-2)/2}).$$

It remains to combine the last relation with (4.11), (4.13) and (4.15). Since $\varepsilon_n \to 0$ in (4.8), Lemma 4.1 follows. $\square$

REMARK 4.3. Note that the probabilities $\mathbf{P}\{|Z| > T\}$ appearing in (4.2) yield a smaller contribution for $T = T_n$ in comparison with the right-hand sides of (4.4) and (4.5). Indeed, we have $\mathbf{P}\{|Z| > T\} \leq C_d T^{d-2} e^{-T^2/2}$ ($T \geq 1$). Hence, relations (4.4) and (4.5) may be extended to the integrals

$$\int_{|x| > T_n} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx.$$

5. Taylor-type expansion for the entropic distance. In this section we provide the last auxiliary step toward the proof of Theorem 1.1. In order to describe the multidimensional case, let $X_1, X_2, \dots$ be independent identically distributed random vectors in $\mathbf{R}^d$ with mean zero, identity covariance matrix, and such that $D(Z_{n_0}) < +\infty$ for some n_0.

If p_{n_0} is bounded, then the densities p_n of Z_n ($n \geq n_0$) are uniformly bounded, and we put $\widetilde{p}_n = p_n$. Otherwise, we use the modified densities $\widetilde{p}_n$ according to the construction of Section 2. In particular, if $\widetilde{Z}_n$ has density $\widetilde{p}_n$, then $|D(\widetilde{Z}_n \| Z) -$

$D(Z_n)| < 2^{-n}$ for all n large enough (where Z is a standard normal random vector; cf. Lemma 2.2 and Remark 2.5). Moreover, by Lemmas 4.1, 4.2 and Remark 4.3,

$$\left| D(Z_n) - \int_{|x| \le T_n} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx \right| = o(\Delta_n), \tag{5.1}$$

where T_n are defined in (4.3) and

$$\Delta_n = n^{-(s-2)/2} (\log n)^{-(s-\max(d,2))/2} \tag{5.2}$$

(with the convention that $\Delta_n = 1$ for the critical case $s = 2$).

Thus, all information about the asymptotics of $D(Z_n)$ is contained in the integral in (5.1). More precisely, writing a Taylor expansion for $\widetilde{p}_n$ using the approximating functions φ_m in Theorems 3.2–3.4 leads to the following representation (which is more convenient in applications such as Corollary 1.2).

THEOREM 5.1. *Let* $\mathbf{E}|X_1|^s < +\infty$ $(s \ge 2)$, *assuming that s is integer in case* $d \ge 2$. *Then*

$$\begin{aligned} D(Z_n) &= \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int \big(\varphi_m(x) - \varphi(x)\big)^k \frac{dx}{\varphi(x)^{k-1}} \\ &\quad + o(\Delta_n) \qquad \big(m = [s]\big). \end{aligned} \tag{5.3}$$

Note that in the case $2 \le s < 4$ there are no expansion terms in the sum of (5.3) which then simplifies to $D(Z_n) = o(\Delta_n)$.

PROOF OF THEOREM 5.1. In terms of $L(u) = u \log u$, rewrite the integral in (5.1) as

$$\begin{aligned} \widetilde{D}_{n,1} &= \int_{|x| \le T_n} L\left(\frac{\widetilde{p}_n(x)}{\varphi(x)}\right) \varphi(x)\, dx \\ &= \int_{|x| \le T_n} L\big(1 + u_m(x) + v_n(x)\big) \varphi(x)\, dx, \end{aligned} \tag{5.4}$$

where

$$u_m(x) = \frac{\varphi_m(x) - \varphi(x)}{\varphi(x)}, \qquad v_n(x) = \frac{\widetilde{p}_n(x) - \varphi_m(x)}{\varphi(x)}.$$

By Theorems 3.3 and 3.4, more precisely, by (3.6) for $d = 1$, and by (3.11) for $d \ge 2$ and $s = m$ integer, in the region $|x| = O(n^\delta)$ with an appropriate $\delta > 0$, we have

$$\big|\widetilde{p}_n(x) - \varphi_m(x)\big| \le \frac{r_n}{n^{(s-2)/2}} \frac{1}{1 + |x|^s}, \qquad r_n \to 0. \tag{5.5}$$

Since $\varphi(x)(1+|x|^s)$ is decreasing as a function of $|x|$ for large $|x|$, we obtain, for all $|x| \le T_n$,

$$|v_n(x)| \le C \frac{r_n}{n^{(s-2)/2}} \frac{e^{T_n^2/2}}{T_n^s} \le C' r_n e^{\rho_n/2}.$$

The last expression tends to zero by a suitable choice of $\rho_n \to \infty$ which we will assume from now on. In particular, for n large enough, $|v_n(x)| < \frac{1}{4}$ in $|x| \le T_n$.

From the definitions of q_k and φ_m [cf. (1.2), (3.1) and (3.10)], it follows that

$$|u_m(x)| \le C_m \frac{1+|x|^{3(m-2)}}{\sqrt{n}} \tag{5.6}$$

with some constants depending on m and the cumulants, only. Thus, we also have $|u_m(x)| < \frac{1}{4}$ for $|x| \le T_n$ with sufficiently large n.

Now, by Taylor's formula, for $|u| \le \frac{1}{4}$, $|v| \le \frac{1}{4}$,

$$L(1+u+v) = L(1+u) + v + 2\theta_1 uv + \theta_2 v^2$$

with some $|\theta_j| \le 1$ depending on (u,v). Applying this approximation with $u = u_m(x)$ and $v = v_n(x)$, we see that $v_n(x)$ can be removed from the right-hand side of (5.4) at the expense of an error not exceeding $|J_1| + J_2 + J_3$, where

$$J_1 = \int_{|x|\le T_n} (\tilde{p}_n(x) - \varphi_m(x))\,dx, \qquad J_2 = \int_{|x|\le T_n} |u_m(x)||\tilde{p}_n(x) - \varphi_m(x)|\,dx$$

and

$$J_3 = \int_{|x|\le T_n} \frac{(\tilde{p}_n(x) - \varphi_m(x))^2}{\varphi(x)}\,dx.$$

But

$$\begin{aligned} |J_1| &= \left| \int_{|x|>T_n} (\tilde{p}_n(x) - \varphi_m(x))\,dx \right| \\ &\le \int_{|x|>T_n} \tilde{p}_n(x)\,dx + \int_{|x|>T_n} |\varphi_m(x)|\,dx. \end{aligned} \tag{5.7}$$

By Lemmas 4.1 and 4.2, the first integral on the right-hand side is T_n^2-times smaller than $o(\Delta_n)$. Also, by (5.6), the last integral in (5.7) is bounded by

$$\begin{aligned} &\int_{|x|>T_n} |\varphi_m(x) - \varphi(x)|\,dx + \int_{|x|>T_n} \varphi(x)\,dx \\ &\quad \le \frac{C_m}{\sqrt{n}} \int_{|x|>T_n} (1+|x|^{3(m-2)})\varphi(x)\,dx + \mathbf{P}\{|Z| > T_n\} = o(\Delta_n). \end{aligned}$$

As a result, $J_1 = o(\Delta_n)$.

Applying (5.6) once more and then relation (3.12), we may also conclude that

$$J_2 \le C_m \frac{1+T_n^{3(m-2)}}{\sqrt{n}} \int_{|x|\le T_n} \left|\widetilde{p}_n(x) - \varphi_m(x)\right| dx = o(\Delta_n).$$

Finally, using (5.5) with $s > 2$, we get, up to some constants,

$$\begin{aligned} J_3 &\le C \frac{r_n^2}{n^{s-2}} \int_{|x|\le T_n} \frac{e^{|x|^2/2}}{1+|x|^{2s}}\, dx \le C_d \frac{r_n^2}{n^{s-2}} \int_1^{T_n} r^{d-2s-1} e^{r^2/2}\, dr \\ &\le C_d' \frac{r_n^2}{n^{s-2}} \frac{1}{T_n^{2s-d+2}} e^{T_n^2/2} = o\left(\frac{1}{n^{(s-2)/2}(\log n)^{(s-d+2)/2}}\right) = o(\Delta_n). \end{aligned}$$

If $s = 2$, all these steps are valid as well and give

$$J_3 \le C_d' \frac{r_n^2}{n^{s-2}} \frac{1}{T_n^{2s-d+2}} e^{T_n^2/2} \to 0$$

for a suitably chosen $T_n \to +\infty$.

Thus, at the expense of an error not exceeding $o(\Delta_n)$ one may remove $v_n(x)$ from (5.4), and we obtain the relation

$$\widetilde{D}_{n,1} = \int_{|x|\le T_n} L\big(1+u_m(x)\big)\varphi(x)\, dx + o(\Delta_n), \tag{5.8}$$

which contains specified expansion terms, only.

Moreover, $u_m(x) = u_2(x) = 0$ for $2 \le s < 3$, and then the theorem is proved.

Next, we consider the case $s \ge 3$. By Taylor's expansion around zero, we get, whenever $|u| < \frac{1}{4}$, for some positive constants θ_m,

$$L(1+u) = u + \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} u^k + \theta u^{m-1}, \qquad |\theta| \le \theta_m,$$

assuming that the sum has no terms in the case $m = 3$. Hence, with some $|\theta| \le \theta_m$,

$$\begin{aligned} &\int_{|x|\le T_n} L\big(1+u_m(x)\big)\varphi(x)\, dx \\ &\qquad = \int_{|x|\le T_n} \big(\varphi_m(x) - \varphi(x)\big)\, dx && (5.9) \\ &\qquad\quad + \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int_{|x|\le T_n} u_m(x)^k \varphi(x)\, dx \\ &\qquad\quad + \theta \int_{\mathbf{R}^d} \left|u_m(x)\right|^{m-1} \varphi(x)\, dx. && (5.10) \end{aligned}$$

For n large enough, by (5.6), the second integral in (5.9) has an absolute value

$$\left|\int_{|x|>T_n} \big(\varphi_m(x) - \varphi(x)\big)\, dx\right| \le \frac{C}{\sqrt{n}} \int_{|x|>T_n} \big(1+|x|^{3(m-2)}\big)\varphi(x)\, dx = o(\Delta_n).$$

This proves the theorem in the case $3 \le s < 4$ (when $m = 3$).

Now, let $s \ge 4$. The last integral in (5.10) can be estimated again by virtue of (5.6) by

$$\frac{C}{n^{(m-1)/2}} \int_{\mathbf{R}^d} (1 + |x|^{3(m-1)(m-2)}) \varphi(x)\, dx = o(\Delta_n).$$

In addition, the first integral in (5.10) can be extended to the whole space at the expense of an error not exceeding (for all n large enough)

$$\begin{aligned} \int_{|x|>T_n} |u_m(x)|^k \varphi(x)\, dx &\le \frac{C}{n^{k/2}} \int_{|x|>T_n} (1 + |x|^{3k(m-2)}) \varphi(x)\, dx \\ &\le \frac{C' T_n^{3k(m-2)}}{\sqrt{n}} e^{-T_n^2/2} = o(\Delta_n). \end{aligned}$$

Collecting these estimates in (5.9) and (5.10) and applying them in (5.8), we arrive at

$$\widetilde{D}_{n,1} = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int u_m(x)^k \varphi(x)\, dx + o(\Delta_n).$$

It remains to apply (5.1). Thus, Theorem 5.1 is proved. $\square$

6. Theorem 1.1 and its multidimensional extension. The desired representation (1.3) of Theorem 1.1 can be deduced from Theorem 5.1. Note that the latter covers the multidimensional case as well, although under somewhat stronger moment assumptions.

Thus, let $(X_n)_{n\ge 1}$ be independent identically distributed random vectors in $\mathbf{R}^d$ with finite second moment. If the normalized sum $Z_n = (X_1 + \cdots + X_n)/\sqrt{n}$ has density $p_n(x)$, the entropic distance to Gaussianity is defined as in dimension one to be the relative entropy

$$D(Z_n) = \int p_n(x) \log \frac{p(x)}{\varphi_{a,\Sigma}(x)}\, dx$$

with respect to the normal law on $\mathbf{R}^d$ with the same mean $a = \mathbf{E}X_1$ and covariance matrix $\Sigma = \mathrm{Var}(X_1)$. This quantity is affine invariant, and in this sense it does not depend on (a, Σ).

THEOREM 6.1. *If $D(Z_{n_0}) < +\infty$ for some n_0, then $D(Z_n) \to 0$, as $n \to \infty$. Moreover, given that $\mathbf{E}|X_1|^s < +\infty$ $(s \ge 2)$, and that X_1 has mean zero and identity covariance matrix, we have*

$$D(Z_n) = \frac{c_1}{n} + \frac{c_2}{n^2} + \cdots + \frac{c_{[(m-2)/2]}}{n^{[(m-2)/2]}} + o(\Delta_n) \qquad \big(m = [s]\big), \tag{6.1}$$

where Δ_n are defined in (5.2), and where we assume that s is integer in case $d \ge 2$.

Here, as in Theorem 1.1, each coefficient c_j is defined according to (1.4) again. It may be represented as a certain polynomial in the cumulants γ_ν, $3 \le |\nu| \le 2j+1$.

PROOF OF THEOREM 6.1. We shall start from the representation (5.3) of Theorem 5.1, so let us return to definition (3.1),

$$\varphi_m(x) - \varphi(x) = \sum_{r=1}^{m-2} q_r(x) n^{-r/2}.$$

In the case $2 \le s < 3$ (i.e., for $m = 2$), the right-hand side contains no terms and is therefore vanishing. Anyhow, raising this sum to the power $k \ge 2$ leads to

$$\big(\varphi_m(x) - \varphi(x)\big)^k = \sum_j n^{-j/2} \sum q_{r_1}(x) \cdots q_{r_k}(x),$$

where the inner sum is carried out over all positive integers $r_1, \dots, r_k \le m-2$ such that $r_1 + \cdots + r_k = j$. Respectively, the kth integral in (5.3) is equal to

$$\sum_j n^{-j/2} \sum \int q_{r_1}(x) \cdots q_{r_k}(x) \frac{dx}{\varphi(x)^{k-1}}. \tag{6.2}$$

Here the integrals are vanishing for odd j. In dimension one, this follows directly from definition (1.2) of q_r and the following property of the Chebyshev–Hermite polynomials [24]

$$\int_{-\infty}^{+\infty} H_{r_1}(x) \cdots H_{r_k}(x) \varphi(x)\, dx = 0 \qquad (r_1 + \cdots + r_k \text{ is odd}). \tag{6.3}$$

As for the general case, let us look at the structure of the functions q_r. Given a multi-index $\nu = (\nu_1, \dots, \nu_d)$ with integers $\nu_1, \dots, \nu_d \ge 1$, define $H_\nu(x_1, \dots, x_d) = H_{\nu_1}(x_1) \cdots H_{\nu_d}(x_d)$, so that

$$\int e^{i\langle t,x\rangle} H_\nu(x) \varphi(x)\, dx = (it)^\nu e^{-|t|^2/2}, \qquad t \in \mathbf{R}^d.$$

Hence, by definition (3.10),

$$q_r(x) = \varphi(x) \sum_\nu a_\nu H_\nu(x), \tag{6.4}$$

where the coefficients a_ν emerge from the expansion $P_r(it) = \sum_\nu a_\nu (it)^\nu$. Using (3.9), write these polynomials as

$$P_r(it) = \sum \frac{1}{l_1! \cdots l_r!} \bigg(\sum_{|\nu|=3} \gamma_\nu \frac{(it)^\nu}{\nu!} \bigg)^{l_1} \cdots \bigg(\sum_{|\nu|=r+2} \gamma_\nu \frac{(it)^\nu}{\nu!} \bigg)^{l_r}, \tag{6.5}$$

where the outer summation is performed over all nonnegative integer solutions $(l_1, \dots, l_r)$ to the equation $l_1 + 2l_2 + \cdots + rl_r = r$. Removing the brackets of the inner sums, we obtain a linear combination of the power polynomials $(it)^\nu$ with exponents of order

$$|\nu| = 3l_1 + \cdots + (r+2)l_r = r + 2b_l, \qquad b_l = l_1 + \cdots + l_r. \tag{6.6}$$

In particular, $r+2 \leq |\nu| \leq 3r$, so that $P_r(it)$ is a polynomial of degree at most $3r$, and thus $\varphi_m(x) = N(x)\varphi(x)$, where $N(x)$ is a polynomial of degree at most $3(m-2)$.

Moreover, from (6.4) and (6.6) it follows that

$$\frac{q_{r_1}(x)\cdots q_{r_k}(x)}{\varphi(x)^{k-1}} = \varphi(x)\sum a_{\nu^{(1)}}\cdots a_{\nu^{(k)}} H_{\nu^{(1)}}(x)\cdots H_{\nu^{(k)}}(x), \tag{6.7}$$

where $|\nu^{(1)}| + \cdots + |\nu^{(k)}| = r_1 + \cdots + r_k \pmod 2$. Hence, if $r_1 + \cdots + r_k$ is odd, the sum

$$|\nu^{(1)}| + \cdots + |\nu^{(k)}| = \sum_{i=1}^d \bigl(|\nu_i^{(1)}| + \cdots + |\nu_i^{(k)}|\bigr)$$

is odd as well. But then at least one of the inner sums, say with coordinate i, must be odd as well. Hence in this case, the integral of (6.7) over x_i will vanish by property (6.3).

Thus, in expression (6.2), only even values of j should be taken into account.

Moreover, since the terms containing $n^{-j/2}$ with $j > s-2$ will be absorbed into the remainder Δ_n in relation (6.1), we get from (5.3) and (6.2),

$$D(Z_n) = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \sum_{\text{even}\, j=2}^{m-2} n^{-j/2} \sum \int q_{r_1}(x)\cdots q_{r_k}(x)\,\frac{dx}{\varphi(x)^{k-1}} + o(\Delta_n).$$

Replace now j with $2j$ and rearrange the summation. Then

$$D(Z_n) = \sum_{2j \leq m-2} \frac{c_j}{n^j} + o(\Delta_n)$$

with

$$c_j = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \sum \int q_{r_1}(x)\cdots q_{r_k}(x)\,\frac{dx}{\varphi(x)^{k-1}}.$$

Here the inner summation is carried out over all positive integers $r_1, \dots, r_k \leq m-2$ such that $r_1 + \cdots + r_k = 2j$. This implies $k \leq 2j$. Furthermore, $2j \leq m-2$ is equivalent to $j \leq [\frac{s-2}{2}]$. As a result, we arrive at the required relation (6.1) with

$$c_j = \sum_{k=2}^{2j} \frac{(-1)^k}{k(k-1)} \sum_{r_1+\cdots+r_k=2j} \int q_{r_1}(x)\cdots q_{r_k}(x)\,\frac{dx}{\varphi(x)^{k-1}}. \tag{6.8}$$

Thus, Theorem 6.1 and therefore Theorem 1.1 are proved. □

REMARK. In order to show that c_j is a polynomial in the cumulants γ_ν, $3 \le |\nu| \le 2j+1$, first note that $r_1 + \cdots + r_k = 2j$, $r_1, \dots, r_k \ge 1$ imply $2j \ge \max_i r_i + (k-1)$, so $\max_i r_i \le 2j-1$. Thus, the maximal index for the functions q_{r_i} in (6.8) does not exceed $2j-1$. On the other hand, it follows from (6.4) and (6.5) that P_r and q_r are polynomials in the same set of the cumulants; more precisely, P_r is a polynomial in γ_ν with $3 \le |\nu| \le r+2$.

PROOF OF COROLLARY 1.2. By Theorem 5.1 [cf. (5.3)],

$$D(Z_n) = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int (\varphi_m(x) - \varphi(x))^k \frac{dx}{\varphi(x)^{k-1}} + o(\Delta_n). \tag{6.9}$$

Assume that $m \ge 4$ and $\gamma_3 = \cdots = \gamma_{k-1} = 0$ for a given integer $3 \le k \le m$. (This is no restriction, when $k = 3$.) Then, by (1.2), $q_1 = \cdots = q_{k-3} = 0$, while $q_{k-2}(x) = \frac{\gamma_k}{k!} H_k(x)\varphi(x)$. Hence, according to definition (3.1),

$$\varphi_m(x) - \varphi(x) = \frac{\gamma_k}{k!} H_k(x)\varphi(x) \frac{1}{n^{(k-2)/2}} + \sum_{j=k-1}^{m-2} \frac{q_j(x)}{n^{j/2}},$$

where the sum is empty in the case $m = 3$. Therefore, the sum in (1.3) will contain powers of $1/n$ starting from $1/n^{k-2}$, and the leading coefficient is due to the quadratic term in (6.9) when $k = 2$. More precisely, if $k - 2 \le \frac{m-2}{2}$, we get that $c_1 = \cdots = c_{k-3} = 0$, and

$$c_{k-2} = \frac{\gamma_k^2}{2k!^2} \int_{-\infty}^{+\infty} H_k(x)^2 \varphi(x)\, dx = \frac{\gamma_k^2}{2k!}. \tag{6.10}$$

Hence, if $k \le \frac{m}{2}$, (6.9) yields $D(Z_n) = \frac{\gamma_k^2}{2k!} \frac{1}{n^{k-2}} + O(n^{-(k-1)})$. Otherwise, the O-term should be replaced by $o((n \log n)^{-(s-2)/2})$. Thus Corollary 1.2 is proved. □

By a similar argument, the conclusion may be extended to the multidimensional case. Indeed, if $\gamma_\nu = 0$, for all $3 \le |\nu| < k$, then by (6.5), $P_1 = \cdots = P_{k-3} = 0$, while

$$P_{k-2}(it) = \sum_{|\nu|=k} \gamma_\nu \frac{(it)^\nu}{\nu!}.$$

Correspondingly, in (6.4) we have $q_1 = \cdots = q_{k-3} = 0$ and $q_{k-2}(x) = \varphi(x) \times \sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} H_\nu(x)$. Therefore,

$$\varphi_m(x) - \varphi(x) = \varphi(x) \sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} H_\nu(x) \frac{1}{n^{(k-2)/2}} + \sum_{j=k-1}^{m-2} \frac{q_j(x)}{n^{j/2}}.$$

Applying this relation in (6.9), we arrive at (6.1) with $c_1 = \cdots = c_{k-3} = 0$ and, by orthogonality of the polynomials H_ν,

$$c_{k-2} = \frac{1}{2} \int \bigg(\sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} H_\nu(x) \bigg)^2 \varphi(x)\, dx = \frac{1}{2} \sum_{|\nu|=k} \frac{\gamma_\nu^2}{\nu!}.$$

We may summarize our findings as follows.

COROLLARY 6.2. *Let $(X_n)_{n \geq 1}$ be i.i.d. random vectors in $\mathbf{R}^d$ $(d \geq 2)$ with mean zero and identity covariance matrix. Suppose that $\mathbf{E}|X_1|^m < +\infty$, for some integer $m \geq 4$, and $D(Z_{n_0}) < +\infty$, for some n_0. Given $k = 3, 4, \dots, m$, if $\gamma_\nu = 0$ for all $3 \leq |\nu| < k$, we have*

$$D(Z_n) = \frac{1}{2n^{k-2}} \sum_{|\nu|=k} \frac{\gamma_\nu^2}{\nu!} + O\bigg(\frac{1}{n^{k-1}}\bigg) + o\bigg(\frac{1}{n^{(m-2)/2}(\log n)^{(m-d)/2}}\bigg). \tag{6.11}$$

The conclusion corresponds to Corollary 1.2, if we replace d with 2 in the remainder on the right-hand side.

As in dimension one, when $\mathbf{E}X_1^{2k} < +\infty$, the o-term may be removed from this representation, while for $k > \frac{m}{2}$, the o-term dominates. Moreover, if $\frac{m+2}{2} < k \leq m$, we are left with this term, only, that is,

$$D(Z_n) = o\bigg(\frac{1}{n^{(m-2)/2}(\log n)^{(m-d)/2}}\bigg).$$

When $k = 3$, there is no restriction on the cumulants in Corollary 6.2, and (6.11) becomes

$$D(Z_n) = \frac{1}{2n} \sum_{|\nu|=3} \frac{\gamma_\nu^2}{\nu!} + O\bigg(\frac{1}{n^2}\bigg) + o\bigg(\frac{1}{n^{(m-2)/2}(\log n)^{(m-d)/2}}\bigg).$$

If $\mathbf{E}|X_1|^4 < +\infty$, we get $D(Z_n) = O(1/n)$ for $d \leq 4$, and the weaker bound $D(Z_n) = o((\log n)^{(d-4)/2}/n)$ for $d \geq 5$. However, if $\mathbf{E}|X_1|^5 < +\infty$, we always have $D(Z_n) = O(1/n)$ regardless of the dimension d.

Technically, this slight difference between conclusions for different dimensions is due to the dimension-dependent asymptotic $\int_{|x|>T} |x|^2 \varphi(x)\, dx \sim C_d T^d e^{-T^2/2}$.

REMARK. In case of discrete distributions when X_1 takes integer values, asymptotics for $D(S_n)$ were studied by Vilenkin and D'yachkov [26], who used an Edgeworth-type expansion for probabilities $\mathbf{P}\{S_n = k\}$ in the corresponding local limit theorem.

7. Convolutions of mixtures of normal laws. Is the asymptotic description of $D(Z_n)$ in Theorem 1.1 still optimal, if no expansion terms of order n^{-j} are present? This is exactly the case for $2 \le s < 4$.

In order to answer the question, we examine a special class of probability distributions that can be described as mixtures of normal laws on the real line with mean zero. They have densities of the form

$$p(x) = \int_0^{+\infty} \varphi_\sigma(x)\, dP(\sigma) \qquad (x \in \mathbf{R}), \tag{7.1}$$

where P is a (mixing) probability measure on the positive half-axis $(0, +\infty)$, and where

$$\varphi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}$$

is the density of the normal law with mean zero and variance σ^2 [as usual, we write $\varphi(x)$ in the standard normal case with $\sigma = 1$].

Equivalently, let $p(x)$ denote the density of the random variable $X_1 = \rho Z$, where the factors $Z \sim N(0, 1)$ and $\rho > 0$ (with the distribution P) are independent. Such distributions appear naturally, for example, as limit laws of sums with randomized length; cf., for example, [8].

For densities such as (7.1), we need a refinement of the local limit theorem for convolutions, described in the expansions (3.5) and (3.6). More precisely, our aim is to find a representation with an essentially smaller remainder term compared to $o(n^{-(s-2)/2})$.

Thus, let $X_1, X_2, \dots$ be independent random variables, having a common density $p(x)$ as in (7.1), and let $p_n(x)$ denote the density of the normalized sum $Z_n = (X_1 + \cdots + X_n)/\sqrt{n}$. If $X_1 = \rho Z$, where $Z \sim N(0, 1)$ and $\rho > 0$ are independent, then $\mathbf{E}X_1^2 = \mathbf{E}\rho^2$ and more generally,

$$\mathbf{E}|X_1|^s = \beta_s \mathbf{E}\rho^s = \beta_s \int_0^{+\infty} \sigma^s\, dP(\sigma),$$

where β_s denotes the sth absolute moment of Z.

Note that $p(x)$ is unimodal with mode at the origin, and $p(0) = \mathbf{E}\frac{1}{\rho\sqrt{2\pi}}$. If $\rho \ge \sigma_0 > 0$, the density is bounded, and therefore the entropy $h(X_1)$ is finite.

PROPOSITION 7.1. *Assume that* $\mathbf{E}\rho^2 = 1$, $\mathbf{E}\rho^s < +\infty$ $(2 < s \le 4)$. *If* $\mathbf{P}\{\rho \ge \sigma_0\} = 1$ *with some constant* $\sigma_0 > 0$, *then uniformly over all* x,

$$p_n(x) = \varphi(x) + n \int_0^{+\infty} (\varphi_{\sigma_n}(x) - \varphi(x))\, dP(\sigma) + O\left(\frac{1}{n^{s-2}}\right), \tag{7.2}$$

where $\sigma_n = \sqrt{1 + \frac{\sigma^2 - 1}{n}}$.

Of course, when $\mathbf{E}\rho^s < +\infty$ for $s > 4$, the proposition may be still applied, but with $s = 4$. In this case (7.2) has a remainder term of order $O(\frac{1}{n^2})$. Note that necessarily $\sigma_0 \le 1$ under the condition $\mathbf{E}\rho^2 = 1$.

The function p_n may also be described as the density of $Z_n = \sqrt{\frac{\rho_1^2+\cdots+\rho_n^2}{n}}\,Z$, where ρ_k are independent copies of ρ (independent of Z as well). This represention already indicates the closeness of p_n and φ and suggests to appeal to the law of large numbers. However, we shall choose a different approach based on the characteristic functions of Z_n.

Obviously, the characteristic function of X_1 is given by

$$v(t) = \mathbf{E}e^{itX_1} = \mathbf{E}e^{-\rho^2 t^2/2} \qquad (t \in \mathbf{R}).$$

Using Jensen's inequality and the assumption $\rho \ge \sigma_0 > 0$, we get a two-sided estimate

$$e^{-t^2/2} \le v(t) \le e^{-\sigma_0^2 t^2/2}. \tag{7.3}$$

In particular, the function $\psi(t) = e^{t^2/2} v(t) - 1$ is nonnegative for all t real.

LEMMA 7.2. *If* $\mathbf{E}\rho^2 = 1$, $M_s = \mathbf{E}\rho^s < +\infty$ $(2 \le s \le 4)$, *then for all* $|t| \le 1$,

$$0 \le \psi(t) \le M_s |t|^s.$$

PROOF. We may assume $0 < t \le 1$. Write $\psi(t) = \mathbf{E}(e^{-(\rho^2-1)t^2/2} - 1)$. The expression under the expectation sign is nonpositive for $\rho t > 1$, hence

$$\psi(t) \le \mathbf{E}\big(e^{-(\rho^2-1)t^2/2} - 1\big) 1_{\{\rho \le 1/t\}}.$$

Let $x = -(\rho^2 - 1)t^2$. Clearly, $|x| \le 1$ for $\rho \le 1/t$. Using $e^x \le 1 + x + x^2$ ($|x| \le 1$) and $\mathbf{E}\rho^2 = 1$, we get

$$\begin{aligned} \psi(t) &\le -\frac{t^2}{2}\mathbf{E}(\rho^2 - 1) 1_{\{\rho \le 1/t\}} + \frac{t^4}{4}\mathbf{E}(\rho^2 - 1)^2 1_{\{\rho \le 1/t\}} \\ &= \frac{t^2}{2}\mathbf{E}(\rho^2 - 1) 1_{\{\rho > 1/t\}} + \frac{t^4}{4}\mathbf{E}(\rho^2 - 1)^2 1_{\{\rho \le 1/t\}}. \end{aligned} \tag{7.4}$$

The last expectation is equal to

$$\begin{aligned} &\mathbf{E}\rho^4 1_{\{\rho \le 1/t\}} + 2\mathbf{E}(\rho^2 - 1) 1_{\{\rho > 1/t\}} - \mathbf{P}\{\rho \le 1/t\} \\ &\quad \le \mathbf{E}\rho^4 1_{\{\rho \le 1/t\}} + 2\mathbf{E}\rho^2 1_{\{\rho > 1/t\}} - 1 \\ &\quad \le \mathbf{E}\rho^4 1_{\{\rho \le 1/t\}} + \mathbf{E}\rho^2 1_{\{\rho > 1/t\}}. \end{aligned}$$

Together with (7.4), this gives

$$\psi(t) \le \frac{3t^2}{4}\mathbf{E}\rho^2 1_{\{\rho > 1/t\}} + \frac{t^4}{4}\mathbf{E}\rho^4 1_{\{\rho \le 1/t\}}. \tag{7.5}$$

Finally, $\mathbf{E}\rho^2 1_{\{\rho>1/t\}} \le \mathbf{E}\rho^s t^{s-2} 1_{\{\rho>1/t\}} \le M_s t^{s-2}$ and $\mathbf{E}\rho^4 1_{\{\rho\le 1/t\}} \le \mathbf{E}\rho^s t^{s-4} \times 1_{\{\rho\le 1/t\}} \le M_s t^{s-4}$. It remains to use these estimates in (7.5), and Lemma 7.2 is proved. $\square$

PROOF OF PROPOSITION 7.1. The characteristic functions $v_n(t) = v(\frac{t}{\sqrt{n}})^n$ of Z_n are real-valued and admit, by (7.3), similar bounds

$$e^{-t^2/2} \le v_n(t) \le e^{-\sigma_0^2 t^2/2}. \tag{7.6}$$

In particular, one may apply the inverse Fourier transform to represent the density of Z_n as

$$p_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx} v_n(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx-t^2/2}\big(1+\psi(t/\sqrt{n})\big)^n\,dt.$$

Letting $T_n = \frac{4}{\sigma_0}\log n$, we split the integral into the two regions, defined by

$$I_1 = \int_{|t|\le T_n} e^{-itx} v_n(t)\,dt, \qquad I_2 = \int_{|t|>T_n} e^{-itx} v_n(t)\,dt.$$

By the upper bound in (7.6),

$$|I_2| \le \int_{|t|>T_n} e^{-\sigma_0^2 t^2/2}\,dt \le \frac{\sqrt{2\pi}}{\sigma_0} e^{-\sigma_0^2 T_n^2/2} = \frac{\sqrt{2\pi}}{\sigma_0 n^8}. \tag{7.7}$$

In the interval $|t| \le T_n$, by Lemma 7.2, $\psi(\frac{t}{\sqrt{n}}) \le \frac{M_s|t|^s}{n^{s/2}} \le \frac{1}{n}$, for all $n \ge n_0$. But for $0 \le \varepsilon \le \frac{1}{n}$, there is the simple estimate $0 \le (1+\varepsilon)^n - 1 - n\varepsilon \le 2(n\varepsilon)^2$. Hence, once more by Lemma 7.2,

$$\begin{aligned} 0 &\le \big(1+\psi(t/\sqrt{n})\big)^n - 1 - n\psi(t/\sqrt{n}) \\ &\le 2\big(n\psi(t/\sqrt{n})\big)^2 \le 2M_s^2\frac{|t|^{2s}}{n^{s-2}} \qquad (n \ge n_0). \end{aligned}$$

This gives

$$\left| I_1 - \int_{|t|\le T_n} e^{-itx-t^2/2}\big(1+n\psi(t/\sqrt{n})\big)\,dt \right| \le \frac{2M_s^2}{n^{s-2}}\int_{-\infty}^{+\infty} |t|^{2s} e^{-t^2/2}\,dt. \tag{7.8}$$

In addition,

$$\begin{aligned} &\left| \int_{|t|>T_n} e^{-itx-t^2/2}\big(1+n\psi(t/\sqrt{n})\big)\,dt \right| \\ &\qquad \le \int_{|t|>T_n} e^{-t^2/2}\,dt + n\int_{|t|>T_n} e^{-t^2/2}\psi(t/\sqrt{n})\,dt. \end{aligned}$$

Here, the first integral on the right-hand side is of order $O(n^{-8})$. To estimate the second one, recall that, by (7.3), $\psi(t) = e^{t^2/2}v(t) - 1 \le e^{(1-\sigma_0^2)t^2/2}$. Hence,

$\psi(t/\sqrt{n}) \leq e^{(1-\sigma_0^2)t^2/2}$ and

$$\int_{|t|>T_n} e^{-t^2/2}\psi(t/\sqrt{n})\,dt \leq \int_{|t|>T_n} e^{-\sigma_0^2 t^2/2}\,dt \leq \frac{\sqrt{2\pi}}{\sigma_0 n^8}.$$

Together with (7.7) and (7.8) these bounds imply that

$$p_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx-t^2/2}\big(1 + n\psi(t/\sqrt{n})\big)\,dt + O\bigg(\frac{1}{n^{s-2}}\bigg)$$

uniformly over all x. It remains to note that

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx-t^2/2}\psi(t/\sqrt{n})\,dt &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx-t^2/2}\big(e^{t^2/2n}v(t/\sqrt{n}) - 1\big)\,dt \\ &= \int_0^{+\infty}\big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\,dP(\sigma).\end{aligned}$$

Proposition 7.1 is proved. □

REMARK 7.3. An inspection of (7.5) shows that, in the case $2 < s < 4$, Lemma 7.2 may slightly be sharpened to $\psi(t) = o(|t|^s)$. Correspondingly, the O-relation in Proposition 7.1 can be replaced with an o-relation. This improvement is convenient, but not crucial for the proof of Theorem 1.3.

8. Lower bounds. Proof of Theorem 1.3. Let $X_1, X_2, \dots$ be independent random variables with a common density of the form

$$p(x) = \int_0^{+\infty} \varphi_\sigma(x)\,dP(\sigma), \qquad x \in \mathbf{R}.$$

Equivalently, let $X_1 = \rho Z$ with independent random variables $Z \sim N(0,1)$ and $\rho > 0$ having distribution P.

A basic tool for proving Theorem 1.3 will be the following lower bound on the entropic distance to Gaussianity for the partial sums $S_n = X_1 + \cdots + X_n$.

PROPOSITION 8.1. *Let* $\mathbf{E}\rho^2 = 1$, $\mathbf{E}\rho^s < +\infty$ $(2 < s < 4)$ *and* $\mathbf{P}\{\rho \geq \sigma_0\} = 1$ *with* $\sigma_0 > 0$. *Assume that, for some* $\gamma > 0$,

$$\liminf_{n\to\infty} n^{s-1/2}\int_{n^{1/2+\gamma}}^{+\infty} \frac{1}{\sigma}\,dP(\sigma) > 0. \tag{8.1}$$

Then with some absolute constant $c > 0$ *and some constant* $\delta > 0$,

$$D(S_n) \geq cn\log n\,\mathbf{P}\{\rho \geq \sqrt{n\log n}\} + o\bigg(\frac{1}{n^{(s-2)/2+\delta}}\bigg). \tag{8.2}$$

In fact, in (8.2) one may take any positive number $\delta < \min\{\gamma s, \frac{s-2}{2}\}$.

PROOF OF PROPOSITION 8.1. By Proposition 7.1 and Remark 7.3, uniformly over all x,

$$p_n(x) = \varphi(x) + n \int_0^{+\infty} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dP(\sigma) + o\bigg(\frac{1}{n^{s-2}}\bigg), \tag{8.3}$$

where p_n is the density of $S_n/\sqrt{n}$ and $\sigma_n = \sqrt{1 + \frac{\sigma^2-1}{n}}$.

Define the sequence

$$N_n = \frac{n^{1/2+\gamma}}{5\sqrt{\log n}}$$

for n large enough (so that $N_n \geq 1$). By Chebyshev's inequality,

$$\mathbf{P}\{\rho \geq N_n\} \leq 5^s M_s \frac{\log^2 n}{n^{(1/2+\gamma)s}} = o\bigg(\frac{1}{n^{s/2+\delta}}\bigg), \qquad 0 < \delta < \gamma s. \tag{8.4}$$

Using $u \log u \geq u - 1$ ($u \geq 0$) and applying (8.3), we may write

$$\begin{aligned} I_n &\equiv \int_{|x| \leq 4\sqrt{\log n}} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \\ &\geq \int_{|x| \leq 4\sqrt{\log n}} \big(p_n(x) - \varphi(x)\big)\, dx \\ &\geq n \int_0^{+\infty} \int_{|x| \leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dx\, dP(\sigma) - C \frac{\sqrt{\log n}}{n^{s-2}} \end{aligned} \tag{8.5}$$

with some constant C.

Note that $\sigma_n < 1$ for $\sigma < 1$, and thus, for any $T > 0$,

$$\int_{|x| \leq T} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dx = 2\big(\Phi(T/\sigma_n) - \Phi(T)\big) > 0,$$

where Φ denotes the distribution function of the standard normal law. Hence, the outer integral in (8.5) may be restricted to the range $\sigma \geq 1$. Moreover, by (8.4), one may also restrict this integral, even to the range $\sigma \geq N_n$. More precisely, (8.4) gives

$$n \bigg| \int_{N_n}^{+\infty} \int_{|x| \leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dx\, dP(\sigma) \bigg| \leq n\mathbf{P}\{\rho \geq N_n\} = o\bigg(\frac{1}{n^{(s-2)/2+\delta}}\bigg).$$

Comparing this relation with (8.5) and imposing the additional requirement $\delta < \frac{s-2}{2}$, we get

$$\begin{aligned} I_n &\geq n \int_1^{N_n} \int_{|x| \leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dx\, dP(\sigma) + o\bigg(\frac{1}{n^{(s-2)/2+\delta}}\bigg) \\ &= -2n \int_1^{N_n} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma) + o\bigg(\frac{1}{n^{(s-2)/2+\delta}}\bigg). \end{aligned} \tag{8.6}$$

Now, let us estimate $p_n(x)$ from below in the region $4\sqrt{\log n} \le |x| \le n^\gamma$. If $|x| \ge 4\sqrt{\log n}$, it follows from (8.3) that

$$p_n(x) = n \int_0^{+\infty} \varphi_{\sigma_n}(x)\, dP(\sigma) + o\left(\frac{1}{n^{s-2}}\right). \tag{8.7}$$

Consider the function

$$g_n(x) = \int_0^{+\infty} \frac{\varphi_{\sigma_n}(x)}{\varphi(x)}\, dP(\sigma).$$

Note that $1 \le \sigma_n \le \sigma$ for $\sigma \ge 1$. In this case, the ratio $\frac{\varphi_{\sigma_n}(x)}{\varphi(x)}$ is nonincreasing in $x \ge 0$. Moreover, for $\sigma \ge \sqrt{3n+1}$, we have $\sigma_n^2 = 1 + \frac{\sigma^2-1}{n} \ge 4$, so $1 - \frac{1}{\sigma_n^2} \ge \frac{3}{4}$. Hence, for $|x| \ge 4\sqrt{\log n}$,

$$\frac{\varphi_{\sigma_n}(x)}{\varphi(x)} = \frac{1}{\sigma_n} e^{x^2(1-1/\sigma_n^2)/2} \ge \frac{n^6}{\sigma}.$$

Therefore,

$$g_n(x) \ge n^6 \int_{\sqrt{3n+1}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma).$$

But by assumption (8.1), the last expression tends to infinity with n, so for all n large enough, $g_n(x) \ge 2$ in the interval $|x| \ge 4\sqrt{\log n}$.

Furthermore, if $\sigma \ge |x|\sqrt{n}$, then $\sigma_n^2 = 1 + \frac{\sigma^2-1}{n} \ge x^2$, so $\frac{x^2}{2\sigma_n^2} \le \frac{1}{2}$. On the other hand,

$$\sigma_n^2 < 1 + \frac{\sigma^2}{n} = \frac{n+\sigma^2}{n} \le \frac{\sigma^2/x^2 + \sigma^2}{n} \le \frac{2\sigma^2}{n},$$

since $|x| \ge 4\sqrt{\log n} > 1$ for $n \ge 2$. The two estimates give

$$\varphi_{\sigma_n}(x) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-x^2/2\sigma_n^2} \ge \frac{\sqrt{n}}{6\sigma}.$$

Therefore, whenever $4\sqrt{\log n} \le |x| \le n^\gamma$,

$$n \int_0^{+\infty} \varphi_{\sigma_n}(x)\, dP(\sigma) \ge \frac{n^{3/2}}{6} \int_{|x|\sqrt{n}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma) \ge \frac{n^{3/2}}{6} \int_{n^{1/2+\gamma}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma).$$

By assumption (8.1), the last expression and therefore the left integral are larger than $\frac{c}{n^{s-2}}$ with some constant $c > 0$. Consequently, the remainder term in (8.7) is indeed smaller, so that for all n large enough, we may write, for example,

$$p_n(x) \ge 0.8n \int_0^{+\infty} \varphi_{\sigma_n}(x)\, dP(\sigma) = 0.8n g_n(x)\varphi(x) \qquad \left(4\sqrt{\log n} \le |x| \le n^\gamma\right).$$

Since $g_n(x) \geq 2$ for $|x| \geq 4\sqrt{\log n}$ with large n, we have in this region $\frac{p_n(x)}{\varphi(x)} \geq 1.6n > n$, thus

$$p_n(x) \log \frac{p_n(x)}{\varphi(x)} \geq p_n(x) \log n \geq 0.8n \log n \int_0^{+\infty} \varphi_{\sigma_n}(x)\, dx\, dP(\sigma).$$

Hence,

$$\begin{aligned} &\int_{4\sqrt{\log n} \leq |x| \leq n^\gamma} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \\ &\qquad \geq 0.8n \log n \int_0^{+\infty} \int_{4\sqrt{\log n} \leq |x| \leq n^\gamma} \varphi_{\sigma_n}(x)\, dx\, dP(\sigma) \\ &\qquad = 1.6n \log n \int_0^{+\infty} \int_{4\sqrt{\log n}/\sigma_n}^{n^\gamma/\sigma_n} \varphi(x)\, dx\, dP(\sigma). \end{aligned} \tag{8.8}$$

At this point, it is useful to note that $\frac{n^\gamma}{\sigma_n} \geq 4\sqrt{\log n}$, as long as $\sigma \leq N_n$ with n large enough. Indeed, in this case $\sigma_n^2 \leq (1 - \frac{1}{n}) + \frac{N_n^2}{n} < 1 + \frac{n^{2\gamma}}{25 \log n}$, so

$$(4\sigma_n \sqrt{\log n})^2 \leq 16 \log n \left(1 + \frac{n^{2\gamma}}{25 \log n}\right) < n^{2\gamma}$$

for all n large enough. Hence, from (8.8),

$$\int_{4\sqrt{\log n} \leq |x| \leq n^\gamma} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq 1.6n \log n \int_0^{N_n} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma).$$

But the last expression dominates the double integral in (8.6) with a factor of $2n$. Therefore, combining the above estimate with (8.6), we get

$$\begin{aligned} \int_{|x| \leq n^\gamma} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq 1.4n \log n \int_0^{N_n} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma) \\ + o\left(\frac{1}{n^{(s-2)/2+\delta}}\right). \end{aligned}$$

Finally, we may extend the outer integral on the right-hand side to all values $\sigma > 0$ by noting that, by (8.4),

$$n \log n \int_{N_n}^{+\infty} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma) \leq n \log n \mathbf{P}\{\rho > N_n\} = o\left(\frac{1}{n^{(s-2)/2+\delta}}\right).$$

Hence,

$$\begin{aligned} \int_{|x| \leq n^\gamma} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq 1.4n \log n \int_0^{+\infty} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma) \\ + o\left(\frac{1}{n^{(s-2)/2+\delta}}\right). \end{aligned} \tag{8.9}$$

For the remaining values $|x| \geq n^\gamma$, one can just use the property $u \log u \geq -\frac{1}{e}$ to get a simple lower bound

$$\int_{|x|>n^\gamma} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq \int_{|x|>n^\gamma,\, p_n(x)\leq\varphi(x)} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx$$
$$\geq -\frac{1}{e} \int_{|x|>n^\gamma,\, p_n(x)\leq\varphi(x)} \varphi(x)\, dx \geq -e^{-n^{2\gamma}/2}.$$

Together with (8.9) this yields

$$\int_{-\infty}^{+\infty} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq 1.4 n \log n \int_0^{+\infty} \int_{4\sqrt{\log n}/\sigma_n}^{4\sqrt{\log n}} \varphi(x)\, dx\, dP(\sigma)$$
$$+ o\left(\frac{1}{n^{(s-2)/2+\delta}}\right).$$

To simplify, finally note that $\frac{4}{\sigma_n}\sqrt{\log n} \leq 4$ for $\sigma \geq \sqrt{n \log n}$. In this case the last integral is separated from zero (for large n), hence with some absolute constant $c > 0$

$$\int_{-\infty}^{+\infty} p_n(x) \log \frac{p_n(x)}{\varphi(x)}\, dx \geq cn \log n \mathbf{P}\{\rho \geq \sqrt{n \log n}\} + o\left(\frac{1}{n^{(s-2)/2+\delta}}\right).$$

This is exactly the required inequality (8.2) and Proposition 8.1 is proved. □

PROOF OF THEOREM 1.3. Given $\eta > 0$, one may apply Proposition 8.1 to the probability measure P with density

$$\frac{dP(\sigma)}{d\sigma} = \frac{c}{\sigma^{s+1}(\log \sigma)^\eta}, \qquad \sigma > 2,$$

and extending it to an interval $[\sigma_0, 2]$ to meet the requirement $\int_{\sigma_0}^{+\infty} \sigma^2\, dP(\sigma) = 1$ (with some $0 < \sigma_0 < 1$ and a positive normalizing constant $c = c_{\eta,s}$). It is easy to see that in this case condition (8.1) is fulfilled for $0 < \gamma < \frac{s-2}{2(s+1)}$. In addition, if ρ has the distribution P, we have

$$\mathbf{P}\{\rho \geq \sigma\} \geq \mathrm{const}\, \frac{1}{\sigma^s (\log \sigma)^\eta}$$

for all σ large enough. Hence, by taking $\sigma = \sqrt{n \log n}$, (8.2) provides the desired lower bound. □

REMARK. In case $s = 2$ (i.e., with minimal moment assumptions), the mixtures of the normal laws with discrete mixing measures P were used by Matskyavichyus [18] in the central limit theorem in terms of the Kolmogorov distance. Namely, it is shown that, for any prescribed sequence $\varepsilon_n \to 0$, one may choose P such that $\Delta_n = \sup_x |F_n(x) - \Phi(x)| \geq \varepsilon_n$ for all n large enough (where

F_n is the distribution function of Z_n). In view of the Pinsker-type inequality, one may conclude that

$$D(Z_n) \geq \frac{1}{2}\Delta_n^2 \geq \frac{1}{2}\varepsilon_n^2.$$

Therefore, $D(Z_n)$ may decay at an arbitrarily slow rate.

REFERENCES

[1] ARTSTEIN, S., BALL, K. M., BARTHE, F. and NAOR, A. (2004). Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.* **17** 975–982 (electronic). MR2083473

[2] ARTSTEIN, S., BALL, K. M., BARTHE, F. and NAOR, A. (2004). On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields* **129** 381–390. MR2128238

[3] BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. MR0815975

[4] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. MR0436272

[5] BIKJALIS, A. (1964). An estimate for the remainder term in the central limit theorem. *Litovsk. Mat. Sb.* **4** 303–308. MR0174078

[6] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2011). Non-uniform bounds in local limit theorems in case of fractional moments. I. *Math. Methods Statist.* **20** 171–191. MR2908756

[7] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2011). Non-uniform bounds in local limit theorems in case of fractional moments. II. *Math. Methods Statist.* **20** 269–287. MR2886638

[8] BOBKOV, S. G. and GÖTZE, F. (2007). Concentration inequalities and limit theorems for randomized sums. *Probab. Theory Related Fields* **137** 49–81. MR2278452

[9] ESSEEN, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. MR0014626

[10] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. MR0270403

[11] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. MR0062975

[12] GÖTZE, F. and HIPP, C. (1978). Asymptotic expansions in the central limit theorem under moment conditions. *Z. Wahrsch. Verw. Gebiete* **42** 67–87. MR0467882

[13] IBRAGIMOV, I. A. and LINNIK, J. V. (1965). *Independent and Stationarily Connected Variables*. Nauka, Moscow.

[14] JOHNSON, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press, London. MR2109042

[15] JOHNSON, O. and BARRON, A. (2004). Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields* **129** 391–409. MR2128239

[16] LINNIK, J. V. (1959). An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Theory Probab. Appl.* **4** 288–299. MR0124081

[17] MADIMAN, M. and BARRON, A. (2007). Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory* **53** 2317–2329. MR2319376

[18] MATSKYAVICHYUS, V. K. (1983). A lower bound for the rate of convergence in the central limit theorem. *Teor. Veroyatn. Primen.* **28** 565–569. MR0716313

[19] OSIPOV, L. V. and PETROV, V. V. (1967). On the estimation of the remainder term in the central limit theorem. *Teor. Veroyatn. Primen.* **12** 322–329. MR0216552

[20] PETROV, V. V. (1964). On local limit theorems for sums of independent random variables. *Theory Probab. Appl.* **9** 312–320.

[21] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. MR0388499

[22] PROHOROV, Y. V. (1952). A local theorem for densities. *Doklady Akad. Nauk SSSR* (*N.S.*) **83** 797–800. MR0049501

[23] SIRAŽDINOV, S. H. and MAMATOV, M. (1962). On mean convergence for densities. *Teor. Veroyatn. Primen.* **7** 433–437. MR0143237

[24] SZEGŐ, G. (1967). *Orthogonal Polynomials*, 3rd ed. *American Mathematical Society Colloquium Publications* **23**. Amer. Math. Soc., Providence, RI. MR0310533

[25] TUCKER, H. G. (1965). On a necessary and sufficient condition that an infinitely divisible distribution be absolutely continuous. *Trans. Amer. Math. Soc.* **118** 316–330. MR0182061

[26] VILENKIN, P. A. and D'YACHKOV, A. G. (1998). Asymptotics of Shannon and Rényi entropies for sums of independent random variables. *Problemy Peredachi Informatsii* **34** 17–31. Translation in *Probl. Inf. Transm.* **34** (1999) 219–232. MR1663910

S. G. BOBKOV
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
127 VINCENT HALL, 206 CHURCH ST. S.E.
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: bobkov@math.umn.edu

G. P. CHISTYAKOV
F. GÖTZE
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
POSTFACH 100131
33501 BIELEFELD
GERMANY
E-MAIL: chistyak@math.uni-bielefeld.de
goetze@mathematik.uni-bielefeld.de