

RÉNYI DIVERGENCE AND THE CENTRAL LIMIT THEOREM

BY S. G. BOBKOV^{*,1}, G. P. CHISTYAKOV^{†,2} AND F. GÖTZE^{†,2}

University of Minnesota^{} and Bielefeld University[†]*

We explore properties of the χ^2 and Rényi distances to the normal law and in particular propose necessary and sufficient conditions under which these distances tend to zero in the central limit theorem (with exact rates with respect to the increasing number of summands).

CONTENTS

1. Introduction	270
2. Background on Rényi divergence	274
3. Pearson–Vajda distances	277
4. Basic exponential inequalities	279
5. Laplace and Weierstrass transforms	281
6. Connections with Fourier transform	285
7. Exponential series	287
8. Normal moments	289
9. Behavior of Rényi divergence under convolutions	292
10. Superadditivity of χ^2 with respect to marginals	296
11. Asymptotic expansions and lower bounds	298
12. Necessity part in Theorem 1.2 ($d = 1$)	303
13. Pointwise upper bounds for convolutions of densities	305
14. Sufficiency part in Theorem 1.2 ($d = 1$)	308
15. Nonuniform local limit theorem	311
16. The multidimensional case	313
17. Some examples and counterexamples	316
18. Convolution of Bernoulli with Gaussian	319
References	321

1. Introduction. Given random elements X and Z in a measurable space (Ω, μ) with densities p and q with respect to μ , the χ^2 -distance of Pearson

$$\chi^2(X, Z) = \int \frac{(p - q)^2}{q} d\mu$$

represents an important measure of deviation of the distribution P of X from the distribution Q of Z , which has been frequently used especially in statistics and

Received October 2016; revised January 2018.

¹Supported by the Humboldt Foundation, NSF Grant DMS-1612961 and CRC 1283.

²Supported by CRC 1283 at Bielefeld University.

MSC2010 subject classifications. 60E, 60F15.

Key words and phrases. χ^2 -divergence, Rényi and Tsallis entropies, central limit theorem.

information theory (cf., e.g., [29, 31, 48]). This strong distance-like quantity may be included in the hierarchy of Rényi divergences (relative α -entropies)

$$D_\alpha(X||Z) = \frac{1}{\alpha - 1} \log \int \left(\frac{p}{q}\right)^\alpha q d\mu \quad (\alpha > 0)$$

or equivalently, the Rényi divergence powers/the relative Tsallis entropies $T_\alpha = \frac{1}{\alpha-1}[e^{(\alpha-1)D_\alpha} - 1]$. The most important indexes are $\alpha = 0$, $\alpha = \frac{1}{2}$ (Hellinger distance), $\alpha = 1$ (Kullback–Leibler distance) and $\alpha = 2$ (quadratic Rényi/Tsallis divergence), in which case $T_2 = \chi^2$ and $D_2 = \log(1 + \chi^2)$.

The functionals D_α and T_α are nondecreasing in α , so, for growing indexes the distances are strengthening. In the range $0 < \alpha < 1$, all D_α are comparable to each other and are metrically equivalent to the total variation $\|P - Q\|_{TV}$. However, the informational divergence $D = D_1 = T_1$ (called also the relative entropy),

$$D(X||Z) = \int p \log(p/q) d\mu,$$

is much stronger, and this applies even more so to D_α with $\alpha > 1$. The difference between the different D_α 's appears in applications like the central limit theorem (CLT for short), which is studied in this paper.

For i.i.d. random variables X, X_1, X_2, \dots such that $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$, introduce the normalized sums

$$Z_n = (X_1 + \dots + X_n)/\sqrt{n} \quad (n = 1, 2, \dots)$$

together with their distributions F_n , which hence approach the standard normal law Φ in the weak sense. For convergence in the CLT using strong distances, recall that convergence in total variation was addressed in the 1950s by Prokhorov [39]. He showed that $\|F_n - \Phi\|_{TV}$ tends to zero as $n \rightarrow \infty$, if and only if F_n has a nontrivial absolutely continuous component for some $n = n_0$, that is, $\|F_{n_0} - \Phi\|_{TV} < 2$ (in particular, this is true, if X has density). A similar description is due to Barron [6] in the 1980s for the Kullback–Leibler distance: $D(Z_n||Z) \rightarrow \infty$ for $Z \sim N(0, 1)$, if and only if $D(Z_n||Z) < \infty$ for some $n = n_0$. The latter condition is fulfilled for a large family of underlying distributions, in particular, when X has density p with

$$\int_{-\infty}^{\infty} p(x) \log p(x) dx < \infty.$$

Different aspects of such strong CLTs, including the non-i.i.d. situation and the problem of rates or Berry–Esseen bounds, were studied by many authors, and we refer to [3, 5, 9, 10, 25, 26, 32, 42]. There is also an increasing interest to other limit theorems in other strong distances such as the relative Fisher information; cf. [11, 12, 44, 45].

As for convergence in the CLT with respect to D_α with $\alpha > 1$, not much is known so far. This case seems to be quite different in nature, and here the distance

restricts the range of applicability of the CLT quite substantially. When focusing on the particular value $\alpha = 2$, we are concerned with the behavior of the quantity

$$\chi^2(Z_n, Z) = \int_{-\infty}^{\infty} \frac{(p_n(x) - \varphi(x))^2}{\varphi(x)} dx,$$

where p_n denotes the density of Z_n and φ is the standard normal density. The finiteness of this integral already requires the existence of all moments of X and actually the existence of a ‘‘Gaussian moment.’’ This condition is to be expected, but the convergence to zero, and even the verification of the boundedness of $\chi^2(Z_n, Z)$ in n is rather delicate. This problem was studied in the early 1980s by Fomin [20] in terms of the exponential series (using Cramér’s terminology) for the density of X ,

$$p(x) = \varphi(x) \sum_{k=1}^{\infty} \frac{\sigma_k}{2^k k!} H_{2k}(x),$$

where H_r is the r th Chebyshev–Hermite polynomial. As a main result, he proved that $\chi^2(Z_n, Z) = O(\frac{1}{n})$ as $n \rightarrow \infty$, assuming that p is compactly supported, symmetric, piecewise differentiable, such that the series coefficients satisfy $\sup_{k \geq 2} \sigma_k < 1$. This sufficient condition was verified for the uniform distribution on the interval $(-\sqrt{3}, \sqrt{3})$ (this length is caused by the assumption $\mathbb{E}X^2 = 1$). However, for many other examples, Fomin’s result does not seem to provide applicable answers.

Fortunately, more or less simple necessary and sufficient conditions can be stated for the convergence in χ^2 by using the Laplace transform of the distribution of X . One of the purposes of this paper is to provide the following characterization of a class which may be called the ‘‘domain of χ^2 -attraction to the normal law.’’

THEOREM 1.1. *We have $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$, if and only if $\chi^2(Z_n, Z)$ is finite for some $n = n_0$, and*

$$(1.1) \quad \mathbb{E}e^{tX} < e^{t^2} \quad \text{for all real } t \neq 0.$$

In this case, the χ^2 -divergence admits an Edgeworth-type expansion

$$(1.2) \quad \chi^2(Z_n, Z) = \sum_{j=1}^{s-2} \frac{c_j}{n^j} + O\left(\frac{1}{n^{s-1}}\right) \quad \text{as } n \rightarrow \infty,$$

which is valid for every $s = 3, 4, \dots$ with coefficients c_j representing certain polynomials in the moments $\alpha_k = \mathbb{E}X^k$, $k = 3, \dots, j + 2$.

For $s = 3$, (1.2) becomes

$$\chi^2(Z_n, Z) = \frac{\alpha_3^2}{6n} + O\left(\frac{1}{n^2}\right),$$

and if $\alpha_3 = 0$ (as in the case of symmetric distributions), one may turn to the next moment of order $s = 4$, for which (1.2) yields

$$(1.3) \quad \chi^2(Z_n, Z) = \frac{(\alpha_4 - 3)^2}{24n^2} + O\left(\frac{1}{n^3}\right).$$

The property $\chi^2(Z_n, Z) < \infty$ is rather close to the sub-Gaussian condition (1.1). In particular, it implies that (1.1) is fulfilled for all t large enough, as well as near zero due to the variance assumption. It may happen, however, that (1.1) is fulfilled for all $t \neq 0$ except just one value $t = t_0$, and then there is no CLT for the χ^2 -distance. Various examples illustrating these conditions together with the convergence in χ^2 will be given in the end of the paper.

A similar characterization continues to hold in the *multidimensional* case for mean zero i.i.d. random vectors X, X_1, X_2, \dots in \mathbb{R}^d normalized to have identity covariance. Here, we endow the Euclidean space with the canonical norm and scalar product. Moreover, one may extend these results to the range of indexes $\alpha > 1$, arriving at the following statement. Let us denote by $\alpha^* = \frac{\alpha}{\alpha-1}$ the conjugate index, and by Z a random vector in \mathbb{R}^d having a standard normal distribution.

THEOREM 1.2. *$D_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$, if and only if $D_\alpha(Z_n||Z)$ is finite for some $n = n_0$, and*

$$(1.4) \quad \mathbb{E}e^{(t, X)} < e^{\alpha^*|t|^2/2} \quad \text{for all } t \in \mathbb{R}^d, t \neq 0.$$

In this case, $D_\alpha(Z_n||Z) = O(1/n)$, and even $D_\alpha(Z_n||Z) = O(1/n^2)$, provided that the distribution of X is symmetric about the origin.

Thus, in addition to the strength of normal approximation, the convergence in the Rényi distance says a lot about the character of the underlying distributions. Thanks to the existence of all moments of X , an Edgeworth-type expansion for D_α and T_α also holds similar to (1.2), involving the mixed cumulants of the components of X . Such expansion shows in particular an equivalence

$$D_\alpha(Z_n||Z) \sim T_\alpha(Z_n||Z) \sim \frac{\alpha}{2}\chi^2(Z_n, Z),$$

once these distances tend to zero. Moreover, an Edgeworth-type expansion allows to establish the monotonicity property of $D_\alpha(Z_n||Z)$ with respect to (large) n , in analogy with the known property of the relative entropy.

Note also that the restriction imposed by (1.4) is asymptotically vanishing as α approaches 1. This means that we may expect to arrive at Barron’s theorem in the limit, though this is not rigorously shown here.

As a closely related issue, and in fact, as an effective application, the Rényi divergence appears naturally in the study of normal approximation for densities p_n of Z_n in the form of nonuniform local limit theorems. Like in dimension one, denote by φ the standard normal density in \mathbb{R}^d .

THEOREM 1.3. *Suppose that $D_\alpha(Z_n||Z)$ is finite for some n , and let the property (1.4) be fulfilled. Then, for all n large enough and for all $x \in \mathbb{R}^d$,*

$$(1.5) \quad |p_n(x) - \varphi(x)| \leq \frac{c}{\sqrt{n}} e^{-|x|^2/(2\alpha^*)},$$

where the constant c does not depend on n . Moreover, the rate $1/\sqrt{n}$ may be improved to $1/n$, if the distribution of X is symmetric about the origin.

Thus, (1.5) is implied by the convergence $D_\alpha(Z_n||Z) \rightarrow 0$. Nonuniform bounds in the normal approximation have been intensively studied in the literature; cf. [1, 2, 24, 36, 37]. However, existing results start with weaker hypotheses (e.g., moment assumptions) and either provide a polynomial error of approximation with respect to x (such as $\frac{1}{1+|x|^3}$), or deal with narrow zones contained in regions $|x| = o(\sqrt{n})$.

The paper consists of two parts. In the first part results about the functional D_α are collected, including moment (exponential) inequalities and special properties of characteristic functions. A number of algebraic properties of the χ^2 -distance will also be derived. They are related to the associated exponential series, the behavior under convolutions and heat semi-group transformations, and in higher dimensions—to the super additivity of χ^2 with respect to its marginals. As a by-product, we establish the existence of densities in terms of the so-called normal moments. The second part is entirely devoted to the proof of Theorems 1.1–1.3. Employing an Edgeworth expansion for densities (together with the results from the first part), this proof heavily relies on the tools of complex analysis. To simplify the presentation, almost all proofs will be stated for the one-dimensional case, deferring the modifications needed to extend Theorems 1.1–1.3 to higher dimensions to separate sections.

2. Background on Rényi divergence. First, let us briefly review some general properties of the Rényi divergences. More details can be found in van Erven and Harremoës [49]; cf. also [21, 29, 41].

Let (Ω, μ) be a measure space (with a σ -finite measure), and let X and Z be random elements in Ω , having distributions P and Q with densities $p = \frac{dP}{d\mu}$, $q = \frac{dQ}{d\mu}$, respectively. The following basic definition goes back to [40].

DEFINITION 2.1. For $\alpha \in (0, \infty)$, $\alpha \neq 1$, the Rényi divergence of P from Q and the divergence power (relative Tsallis entropy) of index α are defined by

$$D_\alpha(X||Z) = D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \int \left(\frac{p}{q}\right)^\alpha q \, d\mu,$$

$$T_\alpha(X||Z) = T_\alpha(P||Q) = \frac{1}{\alpha - 1} \left[\int \left(\frac{p}{q}\right)^\alpha q \, d\mu - 1 \right].$$

These quantities do not depend on the choice of the dominating measure μ . The divergence D_α admits an axiomatic characterization via certain postulates. As a natural generalization of the Kullback–Leibler distance, the definition of T_α was introduced by Tsallis [46] (within the so-called “nonextensive thermostistical formalism”); cf. also [15]. Both quantities are related by monotone transformations:

$$D_\alpha = \frac{1}{\alpha - 1} \log(1 + (\alpha - 1)T_\alpha), \quad T_\alpha = \frac{1}{\alpha - 1} [e^{(\alpha-1)D_\alpha} - 1].$$

Thus, when they are small, D_α and T_α are equivalent. Both represent directional distances: $D_\alpha(P||Q) \geq 0$, and $D_\alpha(P||Q) = 0$ if and only if $P = Q$.

The Rényi divergence with $0 < \alpha < 1$ possesses some unique features, like the skew symmetry $D_\alpha(P||Q) = \frac{\alpha}{1-\alpha} D_{1-\alpha}(Q||P)$, where the coefficient is equal to 1 when $\alpha = \frac{1}{2}$. In this case, D_α represents a function of the square of the Hellinger metric:

$$D_{1/2}(P||Q) = -2 \log \left(1 - \frac{1}{2} \text{Hel}^2(P, Q) \right).$$

Another remarkable property is the equivalence of all D_α in this range:

$$\frac{\alpha}{1-\alpha} \frac{1-\beta}{1-\alpha} D_\beta(P||Q) \leq D_\alpha(P||Q) \leq D_\beta(P||Q), \quad 0 < \alpha < \beta < 1.$$

When $\alpha \in (0, 1)$ is fixed, $D_\alpha(P||Q)$ is a continuous function of the tuple (P, Q) with respect to the total variation distance in both coordinates. Conversely, it majorizes the total variation distance between P and Q . Gilardoni [22] has shown that

$$D_\alpha(P||Q) \geq \frac{\alpha}{2} \|P - Q\|_{\text{TV}}^2.$$

This extends the classical Pinsker inequality for the Kulback–Leibler distance (when $\alpha = 1$), with best constant due to Csiszár; cf. [18, 27, 38].

The following property is important for comparing the Rényi divergences.

PROPOSITION 2.2. *For all probability measures P and Q on Ω , the functions $\alpha \rightarrow D_\alpha(P||Q)$ and $\alpha \rightarrow T_\alpha(P||Q)$ are nondecreasing.*

The monotonicity of D_α is discussed in [49]. Let $0 < \alpha < \beta$, $\alpha, \beta \neq 1$. The functions $c \rightarrow e^{ct_0} - 1$ ($t_0 \geq 0$) and $t \rightarrow \frac{e^{ct}-1}{t}$ are nondecreasing in $c \geq 0$ and $t > 1$. Hence, in case $\alpha > 1$, we get, using monotonicity of D_α ,

$$\begin{aligned} T_\alpha(P||Q) &= \frac{1}{\alpha - 1} [e^{(\alpha-1)D_\alpha(P||Q)} - 1] \leq \frac{1}{\alpha - 1} [e^{(\alpha-1)D_\beta(P||Q)} - 1] \\ &\leq \frac{1}{\beta - 1} [e^{(\beta-1)D_\beta(P||Q)} - 1] = T_\beta(P||Q). \end{aligned}$$

In case $\alpha < 1$, we use the property that the function $c \rightarrow 1 - e^{-ct_0}$ is nondecreasing in $c \geq 0$, while $t \rightarrow \frac{1-e^{-ct}}{t}$ is nonincreasing on the half-axis $-\infty < t < 1$. This yields

$$\begin{aligned} T_\alpha(P||Q) &= \frac{1}{1-\alpha} [1 - e^{-(1-\alpha)D_\alpha(P||Q)}] \leq \frac{1}{1-\alpha} [1 - e^{-(1-\alpha)D_\beta(P||Q)}] \\ &\leq \frac{1}{1-\beta} [1 - e^{-(1-\beta)D_\beta(P||Q)}] = T_\beta(P||Q). \end{aligned}$$

The values $0 < \alpha < 1$ and $1 < \alpha < \infty$, for which the Rényi divergence was defined explicitly, are called simple. The monotonicity of $D_\alpha(P||Q)$ with respect to α allows to extend this function to the missing (extended) values $\alpha = 0$, $\alpha = 1$ and $\alpha = \infty$:

$$D_0(P||Q) = \lim_{\alpha \downarrow 0} D_\alpha(P||Q) = -\log Q\{p(x) > 0\},$$

$$D_\infty(P||Q) = \lim_{\alpha \rightarrow \infty} D_\alpha(P||Q) = \log \operatorname{ess\,sup}_P \frac{p(x)}{q(x)},$$

and $D_1(P||Q) = \lim_{\alpha \uparrow 1} D_\alpha(P||Q)$. The extended index $\alpha = 0$ may be used to characterize an absolute continuity or singularity of two given probability distributions: $D_0(P||Q) = 0$ if and only if Q is absolutely continuous with respect to P , and $D_0(P||Q) = \infty$ if and only if P and Q are orthogonal to each other. This can be illustrated by the Gaussian dichotomy—the property saying that any two Gaussian measures are either absolutely continuous to each other or orthogonal; cf. [41], page 366.

The extended index $\alpha = 1$ leads to the Kullback–Leibler distance

$$D(X||Z) = D(P||Q) = \int p \log \frac{p}{q} d\mu,$$

also known as the relative entropy or an informational divergence. It may also be defined as $\lim_{\alpha \downarrow 1} D_\alpha(P||Q)$, as long as $D_\alpha(P||Q)$ is finite for some $\alpha > 1$. Motivated by works of Shannon and Wiener on communication engineering, this quantity was introduced by Kullback and Leibler [28] under the name “the information of P relative to Q .” Note that $D_1 = T_1 = D$, and $D(P||Q) = \infty$, once P is not absolutely continuous with respect to Q .

In the case $\alpha = 2$, we arrive at the quadratic Rényi divergence and the quadratic Rényi divergence power also known as the χ^2 -distance:

$$D_2(X||Z) = \log \int \frac{p^2}{q} d\mu, \quad \chi^2(X, Z) = T_2(X||Z) = \int \frac{p^2}{q} d\mu - 1.$$

In all cases, by the Csiszár–Pinsker inequality for $\alpha = 1$, we have the relations

$$\frac{1}{2} \|P - Q\|_{\text{TV}}^2 \leq D(X||Z) \leq D_2(X||Z) \leq \chi^2(X, Z).$$

Another useful property of these distances is the contractivity under mappings. Let Ω' be an arbitrary measurable space.

PROPOSITION 2.3. For any measurable map S from Ω to Ω' ,

$$(2.1) \quad D_\alpha(S(X)||S(Z)) \leq D_\alpha(X||Z) \quad (\alpha \geq 1).$$

PROOF. Suppose that $D_\alpha(X||Z)$ is finite, so that the distribution P is absolutely continuous with respect to Q . Introducing $\xi = p/q$, $\beta = \alpha/(\alpha - 1)$ with $\alpha > 1$, one may write

$$\begin{aligned} (1 + (\alpha - 1)T_\alpha(X||Z))^{1/\alpha} &= (\mathbb{E}_Q \xi^\alpha)^{1/\alpha} = \sup_{\mathbb{E}_Q \eta^\beta \leq 1} \mathbb{E}_Q \xi \eta \\ &= \sup_{\mathbb{E}_Q \eta^\beta \leq 1} \mathbb{E}_P \eta = \sup_{\mathbb{E}_\eta(Z)^\beta \leq 1} \mathbb{E}_\eta(X), \end{aligned}$$

that is,

$$(2.2) \quad 1 + (\alpha - 1)T_\alpha(X||Z) = \sup(\mathbb{E}_\eta(X))^\alpha,$$

where the supremum is taken over all measurable $\eta : \Omega \rightarrow \mathbb{R}_+$ such that $\mathbb{E}_\eta(Z)^\beta \leq 1$. Similarly, $1 + (\alpha - 1)T_\alpha(S(X), S(Z))$ may be described as

$$\sup_{\mathbb{E}_\eta(S(Z))^\beta \leq 1} (\mathbb{E}_\eta(S(X)))^\alpha = \sup_{\mathbb{E}_{\eta'}(Z)^\beta \leq 1} (\mathbb{E}_{\eta'}(X))^\alpha,$$

where the second supremum has been restricted to the class of functions $\eta' = \eta(S)$. Hence, this supremum does not exceed the right-hand side of (2.2), thus proving (2.1) for T_α and, therefore, for D_α . \square

The property (2.1) is closely related to the so-called data processing inequality in information theory, namely $D_\alpha(P_{\mathfrak{A}}||Q_{\mathfrak{A}}) \leq D_\alpha(P||Q)$, where $P_{\mathfrak{A}}$ and $Q_{\mathfrak{A}}$ denote restrictions of the measures P and Q to an arbitrary σ -subalgebra \mathfrak{A} in Ω (cf. [49], Theorem 1).

3. Pearson–Vajda distances. Writing $\chi^2(X, Z) = \int \frac{|p-q|^2}{q} d\mu$, the χ^2 -distance may be regarded as a particular member in the family of Pearson–Vajda distances [34], described below.

DEFINITION 3.1. For $\alpha \geq 1$, the χ_α -distance of P from Q is defined by

$$\chi_\alpha(X, Z) = \chi_\alpha(P, Q) = \int \left| \frac{p-q}{q} \right|^\alpha q d\mu = \|p - q\|_{L^\alpha(q^{1-\alpha} d\mu)}^\alpha.$$

Here, X and Z denote random elements in (Ω, μ) , having distributions P and Q with densities $p = \frac{dP}{d\mu}$, $q = \frac{dQ}{d\mu}$. The quantity $\chi_\alpha(P, Q)$ (which is often denoted χ^α) does not depend on the choice of the dominating measure μ .

Clearly, the function $\chi_\alpha^{1/\alpha}$ is nondecreasing in α , and when $\alpha = 1$, we arrive at the total variation distance between P and Q .

For our further purpose, it will be useful to relate the Rényi divergence power $T_\alpha = T_\alpha(P||Q)$ to $\chi_\alpha = \chi_\alpha(P||Q)$. Both quantities are metrically equivalent, as seen by the following elementary observation.

PROPOSITION 3.2. *If $\alpha > 1$, we always have*

$$(3.1) \quad T_\alpha \leq \frac{1}{\alpha - 1} [(1 + \chi_\alpha^{1/\alpha})^\alpha - 1],$$

and conversely,

$$(3.2) \quad T_\alpha \geq \frac{3}{16} \min\{\chi_\alpha, \chi_\alpha^{2/\alpha}\} \quad (1 < \alpha \leq 2), \quad T_\alpha \geq \alpha 3^{-\alpha} \chi_\alpha \quad (\alpha \geq 2).$$

PROOF. Toward (3.1), the triangle inequality in $L^\alpha(q^{1-\alpha} d\mu)$ yields

$$\begin{aligned} \chi_\alpha^{1/\alpha} &= \|p - q\|_{L^\alpha(q^{1-\alpha} d\mu)} \geq \left| \|p\|_{L^\alpha(q^{1-\alpha} d\mu)} - \|q\|_{L^\alpha(q^{1-\alpha} d\mu)} \right| \\ &= \left(\int \left(\frac{p}{q} \right)^\alpha q d\mu \right)^{1/\alpha} - 1 = (1 + (\alpha - 1)T_\alpha)^{1/\alpha} - 1. \end{aligned}$$

For the opposite direction, put $\xi = p/q$. Since $dQ = q d\mu$, we may write

$$T_\alpha = \frac{1}{\alpha - 1} [\mathbb{E}\xi^\alpha - 1], \quad \chi_\alpha = \mathbb{E}|\xi - 1|^\alpha$$

with expectations taken on the probability space (Ω, Q) . We have $\xi \geq 0$ and $\mathbb{E}\xi = 1$. Consider the random variable $\eta = \xi - 1 \geq -1$ and the function

$$\psi(t) = \mathbb{E}(1 + t\eta)^\alpha - 1, \quad 0 \leq t \leq 1,$$

so that $\psi(1) = \mathbb{E}\xi^\alpha - 1$. This function is differentiable in $t > 0$, with

$$\psi'(t) = \alpha \mathbb{E}\eta(1 + t\eta)^{\alpha-1}, \quad \psi''(t) = \alpha(\alpha - 1)\mathbb{E}\eta^2(1 + t\eta)^{\alpha-2}.$$

Since $\psi(0) = \psi'(0) = 0$, by the Taylor integral formula,

$$(3.3) \quad \psi(1) = \int_0^1 (1 - t)\psi''(t) dt = \alpha(\alpha - 1)\mathbb{E}\eta^2 \int_0^1 (1 - t)(1 + t\eta)^{\alpha-2} dt.$$

Case $1 < \alpha \leq 2$. Since $t \rightarrow (1 + t\eta)^{\alpha-2}$ is convex on $(0, 1)$, Jensen's inequality with respect to the measure $d\nu(t) = 2(1 - t) dt$ on $(0, 1)$ yields

$$\begin{aligned} \int_0^1 (1 - t)(1 + t\eta)^{\alpha-2} dt &= \frac{1}{2} \int (1 + t\eta)^{\alpha-2} d\nu(t) \\ &\geq \frac{1}{2} \left(1 + \eta \int t d\nu(t) \right)^{\alpha-2} = \frac{1}{2} \left(1 + \frac{1}{3}\eta \right)^{\alpha-2}. \end{aligned}$$

Therefore,

$$\psi(1) \geq \frac{1}{2} \alpha(\alpha - 1) \mathbb{E}\eta^2 \left(1 + \frac{1}{3}\eta \right)^{\alpha-2}.$$

On the set $A = \{|\eta| \leq 1\}$, the expression $\eta^2(1 + \frac{1}{3}\eta)^{\alpha-2}$ is bounded from below by $(\frac{3}{4})^{2-\alpha}\eta^2$, and on $B = \{\eta > 1\}$ by $\eta^2 \cdot (\frac{4}{3}\eta)^{\alpha-2} = (\frac{3}{4})^{2-\alpha}\eta^\alpha$. Hence

$$\psi(1) \geq \frac{1}{2} \alpha(\alpha - 1) \left(\frac{3}{4} \right)^{2-\alpha} \mathbb{E}(\eta^2 1_A + \eta^\alpha 1_B).$$

For our range of α 's we may simply use $\alpha(\frac{3}{4})^{2-\alpha} \geq \frac{3}{4}$, so that

$$\psi(1) \geq \frac{3}{8}(\alpha - 1)\mathbb{E}(\eta^2 1_A + \eta^\alpha 1_B).$$

By Hölder's inequality,

$$\frac{1}{\mathbb{P}(A)}\mathbb{E}\eta^2 1_A \geq \left(\frac{1}{\mathbb{P}(A)}\mathbb{E}|\eta|^\alpha 1_A\right)^{2/\alpha},$$

so $\mathbb{E}\eta^2 1_A \geq (\mathbb{E}|\eta|^\alpha 1_A)^{2/\alpha}$, and thus $\mathbb{E}(\eta^2 1_A + \eta^\alpha 1_B) \geq U = u_0^{2/\alpha} + u_1$, where $u_0 = \mathbb{E}|\eta|^\alpha 1_A$ and $u_1 = \mathbb{E}|\eta|^\alpha 1_B$. If $u = u_0 + u_1 \leq 1$, then $u^{2/\alpha} \leq 2^{\frac{2}{\alpha}-1} \times (u_0^{2/\alpha} + u_1^{2/\alpha}) \leq 2U$, by Jensen's inequality and since $u_1 \leq 1$. In the case $u \geq 1$, we have $2U - u = 2u_0^{2/\alpha} - 2u_0 + u \geq 2u_0^{2/\alpha} - 2u_0 + 1$ which is positive for all $u_0 \geq 0$. So, $2U \geq \min(u, u^{2/\alpha})$ in both cases, that is,

$$\mathbb{E}(\eta^2 1_A + \eta^\alpha 1_B) \geq \frac{1}{2} \min\{\mathbb{E}|\eta|^\alpha, (\mathbb{E}|\eta|^\alpha)^{2/\alpha}\}.$$

As a result,

$$T_\alpha = \frac{1}{\alpha - 1}\psi(1) \geq \frac{3}{16} \min\{\mathbb{E}|\eta|^\alpha, (\mathbb{E}|\eta|^\alpha)^{2/\alpha}\} = \frac{3}{16} \min\{\chi_\alpha, \chi_\alpha^{2/\alpha}\},$$

which yields the first inequality in (3.2).

Case $\alpha > 2$. Let us return to the Taylor integral formula (3.3). Restricting integration to the interval $(\frac{1}{3}, \frac{2}{3})$, we get

$$\psi(1) \geq \frac{\alpha(\alpha - 1)}{3}\mathbb{E}\eta^2 \int_{1/3}^{2/3} (1 + t\eta)^{\alpha-2} dt.$$

Since $\eta \geq -1$, in case $\eta \leq 0$, we have $1 + t\eta \geq 1 + \frac{2}{3}\eta \geq -\frac{1}{3}\eta$. In case $\eta \geq 0$, we similarly have $1 + t\eta \geq t\eta \geq \frac{1}{3}\eta$. In both cases, $1 + t\eta \geq \frac{1}{3}|\eta|$, hence

$$\psi(1) \geq \frac{\alpha(\alpha - 1)}{3}\mathbb{E}\eta^2 \int_{1/3}^{2/3} \left(\frac{1}{3}|\eta|\right)^{\alpha-2} dt = \frac{\alpha(\alpha - 1)}{3^\alpha}\mathbb{E}|\eta|^\alpha,$$

and, therefore, $T_\alpha = \frac{1}{\alpha-1}\psi(1) \geq \alpha 3^{-\alpha}\mathbb{E}|\eta|^\alpha = \alpha 3^{-\alpha}\chi_\alpha$. \square

4. Basic exponential inequalities. We now focus on the case, where $\Omega = \mathbb{R}$ is the real line with Lebesgue measure μ , and where $Z \sim N(0, 1)$ is a standard normal random variable, that is, with density $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x \in \mathbb{R}$. Given a random variable X with density p , the Rényi divergence and the Tsallis distance of index $\alpha > 1$ with respect to Z are then given by

$$(\alpha - 1)D_\alpha(X||Z) = \log \int_{-\infty}^{\infty} \frac{p(x)^\alpha}{\varphi(x)^{\alpha-1}} dx,$$

$$(\alpha - 1)T_\alpha(X||Z) = \int_{-\infty}^{\infty} \frac{p(x)^\alpha}{\varphi(x)^{\alpha-1}} dx - 1.$$

If the distribution of X is not absolutely continuous with respect to μ , then $D_\alpha(X||Z) = T_\alpha(X||Z) = \infty$. These quantities are finite, if, for example, p is bounded and $\mathbb{E}e^{(\alpha-1)X^2/2} < \infty$. In fact, the finiteness of $D_\alpha(X||Z)$ or $T_\alpha(X||Z)$ implies a similar property. In the sequel, we put $\beta = \alpha^* = \frac{\alpha}{\alpha-1}$.

PROPOSITION 4.1. *If $T_\alpha = T_\alpha(X||Z) < \infty$, then X has an absolutely continuous distribution and finite moments of any order. Moreover,*

$$\mathbb{E}e^{cX^2} \leq C(1 - 2\beta c)^{-\frac{1}{2\beta}} \quad \text{for all } c < 1/(2\beta)$$

with $C = (1 + (\alpha - 1)T_\alpha)^{1/\alpha}$. It is possible that $T_\alpha < \infty$, while $\mathbb{E}e^{\frac{1}{2\beta}X^2} = \infty$.

PROOF. Let X have density p , and the integral $C = (\int p(x)^\alpha \varphi(x)^{1-\alpha} dx)^{1/\alpha}$ be finite. By the Hölder inequality with dual exponents (β, α) ,

$$\begin{aligned} \mathbb{E}e^{cX^2} &= \int_{-\infty}^{\infty} \frac{p(x)}{\varphi(x)^{1/\beta}} \cdot e^{cx^2} \varphi(x)^{1/\beta} dx \\ &\leq C \left(\int_{-\infty}^{\infty} e^{\beta cx^2} \varphi(x) dx \right)^{1/\beta} = C(1 - 2\beta c)^{-\frac{1}{2\beta}}. \end{aligned}$$

This proves the first assertion. For the second one, consider a density of the form $p(x) = \frac{a}{1+|x|} e^{-\frac{1}{2\beta}x^2}$. Then $T_\alpha < \infty$ and $\mathbb{E}e^{\frac{1}{2\beta}X^2} = \infty$. \square

An alternative (although almost equivalent) variant of Proposition 4.1 is the following.

PROPOSITION 4.2. *If $T_\alpha = T_\alpha(X||Z) < \infty$, then for all $t \in \mathbb{R}$,*

$$(4.1) \quad \mathbb{E}e^{tX} \leq C e^{\beta t^2/2}$$

with $C = (1 + (\alpha - 1)T_\alpha)^{1/\alpha}$. In particular, $\mathbb{P}\{X \geq r\} \leq C e^{-\frac{1}{2\beta}r^2}$ ($r \geq 0$).

Indeed, arguing as before, if p is density of X ,

$$\begin{aligned} \mathbb{E}e^{tX} &= \int_{-\infty}^{\infty} p(x)e^{tx} dx = \int_{-\infty}^{\infty} \frac{p(x)}{\varphi(x)^{1/\beta}} \cdot e^{tx} \varphi(x)^{1/\beta} dx \\ &\leq C \left(\int_{-\infty}^{\infty} e^{\beta t x} \varphi(x) dx \right)^{1/\beta} = C e^{\beta t^2/2}. \end{aligned}$$

This bound cannot be deduced from the bound of Proposition 4.1. In fact, the coefficient C in (4.1) may be chosen to be smaller than 1 for large values of $|t|$. The next assertion will be one of the steps needed in the proof of the sufficiency part of Theorems 1.1–1.2.

PROPOSITION 4.3. *If $T_\alpha(X||Z) < \infty$, then $\lim_{|t| \rightarrow \infty} \mathbb{E}e^{tX} e^{-\beta t^2/2} = 0$.*

PROOF. Write $\mathbb{E}e^{tX} = \int_{-\infty}^\infty e^{tx} p(x) dx$ in terms of density p of X . We split integration over $(0, \infty)$ into the two regions. By the Hölder inequality,

$$\begin{aligned} \int_0^{\beta t/2} e^{tx} p(x) dx &= \int_0^{\beta t/2} p(x) e^{\frac{x^2}{2\beta}} \cdot e^{tx - \frac{x^2}{2\beta}} dx \\ &\leq \left(\int_{-\infty}^\infty p(x)^\alpha e^{\frac{\alpha x^2}{2\beta}} dx \right)^{1/\alpha} \left(\int_0^{\beta t/2} e^{\beta t x - \frac{x^2}{2}} dx \right)^{1/\beta} \\ &\leq \frac{1}{(2\pi)^{1/(2\beta)}} (1 + (\alpha - 1)T_\alpha)^{1/\alpha} \left(\frac{\beta t}{2} \right)^{1/\beta} e^{3\beta t^2/8} \end{aligned}$$

for any $t > 0$, where $T_\alpha = T_\alpha(X||Z)$ and where we used the monotonicity of $\beta t x - \frac{1}{2}x^2$ in the interval $0 \leq x \leq \beta t$. Similarly,

$$\int_{\beta t/2}^\infty p(x) e^{\frac{x^2}{2\beta}} \cdot e^{tx - \frac{x^2}{2\beta}} dx \leq \left(\int_{\beta t/2}^\infty p(x)^\alpha e^{\frac{\alpha x^2}{2\beta}} dx \right)^{1/\alpha} \left(\int_{-\infty}^\infty e^{\beta t x - \frac{x^2}{2}} dx \right)^{1/\beta},$$

which is bounded by $\delta(t)e^{\frac{\beta t^2}{2}}$ with $\delta(t) \rightarrow 0$ as $t \rightarrow \infty$. Collecting the bounds, we get

$$\mathbb{E}e^{tX} 1_{\{X>0\}} e^{-\frac{\beta t^2}{2}} \leq (2\pi)^{-1/(2\beta)} (1 + (\alpha - 1)T_\alpha)^{1/\alpha} \left(\frac{\beta t}{2} \right)^{1/\beta} e^{-\frac{\beta t^2}{8}} + \delta(t) \rightarrow 0.$$

Since also $\mathbb{E}e^{tX} 1_{\{X<0\}} \rightarrow 0$ as $t \rightarrow \infty$, the conclusion follows. \square

5. Laplace and Weierstrass transforms. Although in general the critical constant $c = 1/(2\beta)$ cannot be included in the statement of Proposition 4.1, this turns out possible for sufficiently many normalized convolutions of the distribution of X with itself. Given independent copies X_1, \dots, X_n of X , here we consider ‘‘Gaussian’’ moments for the normalized sums

$$Z_n = (X_1 + \dots + X_n)/\sqrt{n}.$$

The following statement is crucial both in the necessity and sufficiency parts of the proof of Theorems 1.1–1.2. We always assume that $Z \sim N(0, 1)$.

PROPOSITION 5.1. *If $T_\alpha = T_\alpha(X||Z) < \infty$, then $\mathbb{E}e^{\frac{1}{2\beta}Z_n^2} < \infty$ for all $n \geq \alpha$, and*

$$(5.1) \quad \mathbb{E}e^{\frac{1}{2\beta}Z_n^2} \leq 2^n (1 + (\alpha - 1)T_\alpha)^{\frac{n}{\alpha}}.$$

Moreover, putting $\chi_\alpha = \chi_\alpha(X, Z)$, we have

$$(5.2) \quad |\mathbb{E}e^{\frac{1}{2\beta}Z_n^2} - \mathbb{E}e^{\frac{1}{2\beta}Z^2}| \leq 2^n ((1 + \chi_\alpha^{1/\alpha})^n - 1).$$

Thus, when X is close to Z in the sense of the Pearson–Vajda distance, we also obtain closeness of the Gaussian moments of Z_n and Z with fixed $n \geq \alpha$. Recall that χ_α in (5.2) can be estimated from above in terms of T_α according to Proposition 3.2 (while these distances coincide in case $\alpha = 2$).

One may equivalently rephrase inequality (5.1) in terms of the Laplace transform of the distribution of Z_n . Let us state one immediate corollary.

COROLLARY 5.2. *Let $T_\alpha = T_\alpha(Z_{n_0}||Z)$ be finite for some n_0 . Then the function $\psi(t) = \mathbb{E}e^{tX}e^{-\beta t^2/2}$ is integrable with power kn_0 for any integer $k \geq \alpha$, and moreover*

$$(5.3) \quad \int_{-\infty}^{\infty} \psi(t)^{kn_0} dt \leq 2^{k+1}(1 + (\alpha - 1)T_\alpha)^{\frac{k}{\alpha}}.$$

Indeed, representing $(\mathbb{E}e^{tX})^n = \mathbb{E}e^{t(X_1+\dots+X_n)} = \mathbb{E}e^{tZ_n\sqrt{n}}$, we find that

$$\int_{-\infty}^{\infty} \psi(t)^n dt = \int_{-\infty}^{\infty} e^{-\beta n t^2/2} \mathbb{E}e^{tZ_n\sqrt{n}} dt = \frac{\sqrt{2\pi}}{\sqrt{\beta n}} \mathbb{E}e^{\frac{1}{2\beta}Z_n^2}.$$

If $n = kn_0$, then Z_n represents the normalized sum of k independent copies of Z_{n_0} . Hence, one may apply (5.1) to Z_{n_0} in place of X , which yields (5.3) with constant $2^k\sqrt{2\pi}/\sqrt{\alpha\beta} < 2^{k+1}$.

The argument leading to (5.1)–(5.2) uses the contractivity properties of the Weierstrass transforms (a well-known heat semigroup of operators)

$$W_t u(x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{2t}} u(y) dy, \quad x \in \mathbb{R}, t > 0.$$

In the sequel, we denote by L^α the Lebesgue space $L^\alpha(\mathbb{R}, dx)$ of all measurable functions u on the real line with finite norm

$$\|u\|_\alpha = \left(\int_{-\infty}^{\infty} |u(x)|^\alpha dx \right)^{1/\alpha}, \quad \alpha \geq 1,$$

with usual convention $\|u\|_\infty = \text{ess sup}_x |u(x)|$. We refer an interested reader to [23] for a detail account on the Weierstrass transform, and here only mention one property. Since $W_t u$ represents the convolution of u —with the Gaussian density $\varphi_t(x) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/(2t)}$, we have $\|W_t u\|_\alpha \leq \|u\|_\alpha$ for all $\alpha \geq 1$ and $t > 0$. That is, W_t acts as a contraction from L^α to L^α .

This implies that W_t is a bounded operator from L^α to L^γ , $\gamma > \alpha$. Indeed, by Hölder’s inequality, $|W_t u(x)| \leq \|\varphi_t\|_\beta \|u\|_\alpha$ for all x (where $\beta = \alpha^*$), that is,

$$\|W_t u\|_\infty \leq (2\pi t)^{-\frac{1}{2\alpha}} \beta^{-\frac{1}{2\beta}} \|u\|_\alpha.$$

More generally, given $\alpha < \gamma < \infty$, we have

$$\int_{-\infty}^{\infty} |W_t u(x)|^\gamma dx = \int_{-\infty}^{\infty} |W_t u(x)|^{\gamma-\alpha} |W_t u(x)|^\alpha dx \leq (2\pi t)^{\frac{\alpha-\gamma}{2\alpha}} \beta^{\frac{\alpha-\gamma}{2\beta}} \|u\|_\alpha^\gamma.$$

Hence

$$(5.4) \quad \|W_t u\|_\gamma \leq (2\pi t)^{\frac{\alpha-\gamma}{2\gamma\alpha}} \beta^{\frac{\alpha-\gamma}{2\gamma\beta}} \|u\|_\alpha, \quad \alpha \leq \gamma \leq \infty.$$

In fact, since $\frac{\alpha-\gamma}{\gamma\alpha} = \frac{1}{\gamma} - \frac{1}{\alpha}$ may vary from zero to $-\frac{1}{\alpha}$, the latter bound can be made independent of γ . Namely, in the indicated range,

$$\|W_t u\|_\gamma \leq \max\{1, (2\pi t)^{-\frac{1}{2\alpha}} \beta^{\frac{1}{2\beta}}\} \|u\|_\alpha.$$

PROOF OF PROPOSITION 5.1. Let p be the density of X . The Weierstrass transform can be applied to the function $u(x) = \varphi(x)^{-1/\beta} p(x)$, which has finite norm $\|u\|_\alpha = (1 + (\alpha - 1)T_\alpha)^{1/\alpha}$. Putting $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ and using dx for $dx_1 \cdots dx_n$, the expectation we have to estimate is

$$\begin{aligned} \mathbb{E}e^{\frac{1}{2\beta}Z_n^2} &= \int_{\mathbb{R}^n} e^{\frac{n}{2\beta}\bar{x}^2} p(x_1) \cdots p(x_n) dx \\ &= (2\pi)^{-\frac{n}{2\beta}} \int_{\mathbb{R}^n} \exp\left\{\frac{n}{2\beta}\bar{x}^2 - \frac{1}{2\beta}(x_1^2 + \dots + x_n^2)\right\} u(x_1) \cdots u(x_n) dx \\ &= (2\pi)^{-\frac{n}{2\beta}} \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{4\beta n} \sum_{i=1}^n Q_i\right\} u(x_1) \cdots u(x_n) dx, \end{aligned}$$

where $Q_i = \sum_{j=1}^n (x_i - x_j)^2$. Putting $t = 2\beta$, by Hölder’s inequality, we get

$$\begin{aligned} \mathbb{E}e^{\frac{1}{2\beta}Z_n^2} &\leq (2\pi)^{-\frac{n}{2\beta}} \prod_{i=1}^n \left(\int_{\mathbb{R}^n} \exp\left\{-\frac{1}{4\beta} Q_i\right\} u(x_1) \cdots u(x_n) dx \right)^{1/n} \\ &= (2\pi)^{-\frac{n}{2\beta}} (2\pi t)^{\frac{n-1}{2}} \int_{-\infty}^{\infty} (W_t u(x_1))^{n-1} u(x_1) dx_1, \end{aligned}$$

where on the second step, inside the i th integral in the product we performed the integration over the variables $x_j, j \neq i$, which yielded the value $(2\pi t)^{\frac{n-1}{2}} \times (W_t u(x_i))^{n-1}$. By Hölder’s inequality once more, and applying (5.4) with $\gamma = \beta(n - 1)$, which satisfies $\gamma \geq \alpha$ due to the assumption $n \geq \alpha$, we see that the last one-dimensional integral does not exceed

$$\begin{aligned} \left(\int_{-\infty}^{\infty} (W_t u(x_1))^\gamma dx_1 \right)^{\frac{1}{\beta}} \|u\|_\alpha &= \|W_t u\|_\gamma^{n-1} \|u\|_\alpha \\ &\leq ((2\pi t)^{\frac{\alpha-\gamma}{2\gamma\alpha}} \beta^{\frac{\alpha-\gamma}{2\gamma\beta}} \|u\|_\alpha)^{n-1} \|u\|_\alpha. \end{aligned}$$

Hence $\mathbb{E}e^{\frac{1}{2\beta}Z_n^2} \leq c_{n,\alpha} \|u\|_\alpha^n$ with constant

$$\begin{aligned} c_{n,\alpha} &= (2\pi)^{-\frac{n}{2\beta}} (2\pi t)^{\frac{n-1}{2}} (2\pi t)^{\frac{n-1}{2} \frac{\alpha-\gamma}{\gamma\alpha}} \beta^{\frac{n-1}{2} \frac{\alpha-\gamma}{\gamma\beta}} \\ &= t^{\frac{n}{2\beta}} \beta^{\frac{\alpha-n}{2\beta}} = 2^{\frac{n}{2\beta}} \beta^{\frac{1}{2(\beta-1)}} < 2^{\frac{n}{2}} \sqrt{e} < 2^n \quad (n \geq \alpha > 1). \end{aligned}$$

This proves (5.1). It is also interesting to note that $c_{n,\alpha} \rightarrow 1$ as $\alpha \rightarrow 1$.

This argument can easily be extended to not necessarily equal positive functions. Namely, for the integral

$$I = I(p_1, \dots, p_n) = \int_{\mathbb{R}^n} e^{\frac{n}{2\beta} \bar{x}^2} p_1(x_1) \cdots p_n(x_n) dx$$

we similarly obtain

$$\begin{aligned} |I| &\leq (2\pi)^{-\frac{n}{2\beta}} \prod_{i=1}^n \left(\int_{\mathbb{R}^n} \exp\left\{-\frac{1}{4\beta} \sum_{i=1}^n Q_i\right\} |u_1(x_1)| \cdots |u_n(x_n)| dx \right)^{1/n} \\ &= (2\pi)^{-\frac{n}{2\beta}} (2\pi t)^{\frac{n-1}{2}} \prod_{i=1}^n \left(\int_{-\infty}^{\infty} |u_i(x_i)| \prod_{j \neq i} (W_t |u_j|)(x_i) dx_i \right)^{1/n}, \end{aligned}$$

where $u_j = \varphi^{-1/\beta} p_j$. An application of Hölder’s inequality together with (5.4) allows one to estimate the last integral by

$$\begin{aligned} \|u_i\|_\alpha \prod_{j \neq i} \|W_t |u_j|\|_\gamma &\leq \|u_i\|_\alpha \prod_{j \neq i} (2\pi t)^{\frac{\alpha-\gamma}{2\gamma\alpha}} \beta^{\frac{\alpha-\gamma}{2\gamma\beta}} \|u_j\|_\alpha \\ &= (2\pi t)^{\frac{n-1}{2} \frac{\alpha-\gamma}{\gamma\alpha}} \beta^{\frac{n-1}{2} \frac{\alpha-\gamma}{\gamma\beta}} \|u_1\|_\alpha \cdots \|u_n\|_\alpha. \end{aligned}$$

This leads to

$$(5.5) \quad |I(p_1, \dots, p_n)| \leq c_{n,\alpha} \|u_1\|_\alpha \cdots \|u_n\|_\alpha$$

with the same constant as before (so that $c_{n,\alpha} < 2^n$).

We use the latter bound to derive (5.2). Splitting the density of X as $p = \varphi + \varphi^{1/\beta} v$, such that $\|v\|_\alpha^\alpha = \chi_\alpha(X, Z)$, we get a decomposition

$$\begin{aligned} \mathbb{E} e^{\frac{1}{2\beta} Z_n^2} &= \int_{\mathbb{R}^n} e^{\frac{n}{2\beta} \bar{x}^2} p(x_1) \cdots p(x_n) dx \\ &= \sum_{k=0}^n \frac{n!}{k!(n-k)!} \int e^{\frac{n}{2\beta} \bar{x}^2} \varphi(x_1) \cdots \varphi(x_k) \\ &\quad \times \varphi^{1/\beta}(x_{k+1}) v(x_{k+1}) \cdots \varphi^{1/\beta}(x_n) v(x_n) dx. \end{aligned}$$

We apply (5.5) with p_1 to p_k replaced by φ , and with p_{k+1} to p_n replaced with $\varphi^{1/\beta} v$, that is, $u_j = \varphi^{1/\alpha}$ for $j \leq k$ and $u_j = v$ for $j > k$. Moving the last term with $k = n$ of this decomposition to the left, we then get the bound

$$\begin{aligned} \left| \mathbb{E} e^{\frac{1}{2\beta} Z_n^2} - \mathbb{E} e^{\frac{1}{2\beta} Z^2} \right| &\leq c_{n,\alpha} \sum_{k=0}^{n-1} \frac{n!}{k!(n-k)!} \|\varphi^{1/\alpha}\|_\alpha^k \|v\|_\alpha^{n-k} \\ &= c_{n,\alpha} ((1 + \|v\|_\alpha)^n - 1). \end{aligned} \quad \square$$

6. Connections with Fourier transform. In the next sections, we restrict ourselves to the particular interesting index $\alpha = 2$, that is, to the χ^2 -distance from the standard normal law,

$$\chi^2(X, Z) = \int_{-\infty}^{\infty} \frac{p(x)^2}{\varphi(x)} dx - 1, \quad Z \sim N(0, 1).$$

In this case, necessary and sufficient conditions for finiteness of this distance may be given in terms of the characteristic function $f(t) = \mathbb{E}e^{itX}$, $t \in \mathbb{R}$.

PROPOSITION 6.1. *The condition $\chi^2(X, Z) < \infty$ ensures that $f(t)$ has square integrable derivatives of any order. Moreover, in that case*

$$1 + \chi^2(X, Z) = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} |f^{(n)}(t)|^2 dt.$$

PROOF. By the very definition,

$$1 + \chi^2(X, Z) = \sqrt{2\pi} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} x^{2n} p(x)^2 dx.$$

We know that f has finite derivatives of any order given by

$$f^{(n)}(t) = \mathbb{E}(iX)^n e^{itX} = \int_{-\infty}^{\infty} (ix)^n e^{itx} p(x) dx.$$

It remains to apply Plancherel’s theorem. \square

In view of Proposition 4.1, existence of $\chi^2(X, Z)$ does not guarantee existence of the “Gaussian” moment $\mathbb{E}e^{X^2/4}$. Nevertheless, it is true for the normalized convolution of the distribution of X with itself, as indicated in Proposition 5.1. In this case, inequality (5.1) can be stated more precisely as

$$\mathbb{E}e^{\frac{1}{4}(\frac{X+\tilde{X}}{\sqrt{2}})^2} \leq 2(1 + \chi^2(X, Z)),$$

where \tilde{X} is an independent copy of X . Equivalently, there is a corresponding refinement of inequality (5.3) in Corollary 5.2 (without any convolution).

PROPOSITION 6.2. *For any random variable X ,*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(iy)^2 e^{-2y^2} dy \leq 1 + \chi^2(X, Z).$$

The argument is based on the following observation of independent interest.

LEMMA 6.3. *Given a function p on the real line, suppose that the function $g(x) = p(x)e^{x^2/4}$ belongs to L^2 . Then the Fourier transforms*

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx, \quad \rho(t) = \int_{-\infty}^{\infty} e^{itx} g(x) dx$$

are connected by the identity

$$(6.1) \quad f(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(t-u)^2} \rho(u) du \quad (t \in \mathbb{R}),$$

which may analytically be extended to the complex plane. Moreover,

$$(6.2) \quad \int_{-\infty}^{\infty} |f(iy)|^2 e^{-2y^2} dy = \int_{-\infty}^{\infty} |\rho(t)|^2 e^{-2t^2} dt.$$

Thus, the characteristic function f appears as the Weierstrass transform of the Fourier transform of g . While Proposition 5.1 and its Corollary 5.2 are key ingredients of the proof of Theorem 1.2, Proposition 6.2 can be used as an alternative approach to Theorem 1.1 for the particular case $\alpha = 2$. Lemma 6.3 and Proposition 6.2 can be adapted to cover the range $1 < \alpha \leq 2$ by considering the Fourier transform on the Lebesgue space L^α . However, these results do not extend to indexes $\alpha > 2$.

Note that g does not need to be integrable, so one should understand ρ as the limit $\rho(t) = \lim_{N \rightarrow \infty} \int_{-N}^N e^{itx} g(x) dx$ in the norm of the space L^2 .

Let us also note that the second integral in (6.2) can be bounded by the squared L^2 -norm of ρ , which is, by the Plancherel theorem, equal to

$$2\pi \|g\|_2^2 = 2\pi \int_{-\infty}^{\infty} |p(x)|^2 e^{x^2/2} dx.$$

If p is density of X , the last expression is $\sqrt{2\pi}(1 + \chi^2(X, Z))$, thus proving Proposition 6.2.

PROOF OF LEMMA 6.3. First, assume that p is compactly supported; in particular, both p and g are integrable and have analytic Fourier transforms. By Fubini's theorem,

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} e^{itx} g(x) \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-ixu-u^2} du \right] dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} \left[\int_{-\infty}^{\infty} e^{i(t-u)x} g(x) dx \right] du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(u-t)^2} \rho(u) du, \end{aligned}$$

and we obtain (6.1). Rewriting the last integral as $e^{-t^2} \int_{-\infty}^{\infty} e^{2ut-u^2} \rho(u) du$ and changing the variable, we have another representation

$$\sqrt{\pi} f\left(\frac{iz}{2}\right) e^{-z^2/4} = \int_{-\infty}^{\infty} e^{izu-u^2} \rho(u) du \quad (z \in \mathbb{R}).$$

Hence, the left-hand side represents the Fourier transform of the function $e^{-u^2} \rho(u)$, and by Plancherel’s theorem,

$$(6.3) \quad \|e^{-u^2} \rho(u)\|_2^2 = \frac{1}{2} \int_{-\infty}^{\infty} \left| f\left(\frac{iz}{2}\right) \right|^2 e^{-z^2/2} dz = \int_{-\infty}^{\infty} |f(iy)|^2 e^{-2y^2} dy,$$

thus proving (6.2). In the general case, we have $p \in L^1 \cap L^2$, and arguing as in the proof of Proposition 4.1 (for the case $\alpha = 2$), we also get

$$\int_{-\infty}^{\infty} e^{cx^2} |p(x)| dx \leq C(1 - 4c)^{-1/4} < \infty \quad \text{for all } c < \frac{1}{4},$$

where $C^2 = \int_{-\infty}^{\infty} \frac{p(x)^2}{\varphi(x)} dx$. In particular, f is an entire function. Let p_N be the restriction of p to $[-N, N]$, $g_N(x) = p_N(x)e^{x^2/4}$, and put

$$f_N(t) = \int_{-\infty}^{\infty} e^{itx} p_N(x) dx, \quad \rho_N(t) = \int_{-\infty}^{\infty} e^{itx} g_N(x) dx.$$

According to the previous step, for all $t \in \mathbb{R}$,

$$(6.4) \quad f_N(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(t-u)^2} \rho_N(u) du.$$

By the Lebesgue dominated convergence theorem, we have $f_N(t) \rightarrow f(t)$ for all real t and $\|g_N - g\|_2 \rightarrow 0$ as $N \rightarrow \infty$. By the continuity of the Fourier transform on L^2 , we obtain $\|\rho_N - \rho\|_2 \rightarrow 0$, which in turn implies

$$\int_{-\infty}^{\infty} e^{-(t-u)^2} \rho_N(u) du \rightarrow \int_{-\infty}^{\infty} e^{-(t-u)^2} \rho(u) du.$$

Hence, in the limit (6.4) yields the desired identity (6.1). Its right-hand side is well defined and finite for all complex t , and clearly represents an entire function. Moreover, as before, one may apply Plancherel’s theorem, leading to (6.3) and, therefore, to (6.2). \square

7. Exponential series. The χ^2 -distance from the standard normal law on the real line admits a nice description in terms of a so-called exponential series as well. Let us introduce basic notation and recall several well-known facts. By H_k , we denote the k th Chebyshev–Hermite polynomial

$$H_k(x) = (-1)^k (e^{-x^2/2})^{(k)} e^{x^2/2}, \quad k = 0, 1, 2, \dots \quad (x \in \mathbb{R}).$$

In particular, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$. Each H_k is a polynomial of degree k with integer coefficients. Depending on k being even or odd, H_k contains even respectively odd powers only. These polynomials may be defined explicitly via $H_k(x) = \mathbb{E}(x + iZ)^k$, $Z \sim N(0, 1)$. Being orthogonal to

each other with weight function $\varphi(x)$, they form a complete orthogonal system in the Hilbert space $L^2(\mathbb{R}, \varphi(x) dx)$, with

$$\mathbb{E}H_k(Z)^2 = \int_{-\infty}^{\infty} H_k(x)^2 \varphi(x) dx = k!.$$

Equivalently, the Hermite functions $\varphi_k = H_k \varphi$ form a complete orthogonal system in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$ with $\int_{-\infty}^{\infty} \frac{\varphi_k(x)^2}{\varphi(x)} dx = k!$ Summarizing we have the following.

PROPOSITION 7.1. *Any complex valued function u with $\int_{-\infty}^{\infty} |u(x)|^2 e^{\frac{x^2}{2}} dx < \infty$ admits a unique representation in the form of the orthogonal series*

$$(7.1) \quad u(x) = \varphi(x) \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x),$$

which converges in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$. Here, the coefficients are given by

$$c_k = \int_{-\infty}^{\infty} u(x) H_k(x) dx,$$

and we have Parseval's identity

$$\sum_{k=0}^{\infty} \frac{|c_k|^2}{k!} = \int_{-\infty}^{\infty} \frac{|u(x)|^2}{\varphi(x)} dx.$$

The functional series (7.1) representing u is called an exponential series. The question of its pointwise convergence is rather delicate similar to the pointwise convergence of ordinary Fourier series based on trigonometric functions. In Cramér's paper [17], the following two propositions are stated, together with an explanation of the basic ingredients of the proof.

PROPOSITION 7.2. *If $u(x)$ is vanishing at infinity and has a continuous derivative such that the integral $\int_{-\infty}^{\infty} |u'(x)|^2 e^{x^2/2} dx$ is finite, then it may be developed in an exponential series, which is absolutely and uniformly convergent for $-\infty < x < \infty$.*

PROPOSITION 7.3. *If $u(x)$ has bounded variation in every finite interval, and if the integral $\int_{-\infty}^{\infty} |u(x)| e^{x^2/4} dx$ is finite, then the exponential series for $u(x)$ converges to $\frac{u(x+) + u(x-)}{2}$ uniformly in every finite interval of continuity.*

Proposition 7.3 is illustrated in [17] on the example of the Gaussian functions $u(x) = e^{-\lambda x^2}$ ($\lambda > 0$). In this case, the corresponding exponential series can be explicitly computed. At $x = 0$, it is absolutely convergent for $\lambda > \frac{1}{4}$, simply convergent for $\lambda = \frac{1}{4}$ and divergent for $\lambda < \frac{1}{4}$.

8. Normal moments. Let X be a random variable with density p , and let Z be a standard normal random variable independent of X . Applying Proposition 7.1 to p , we obtain the following: If

$$(8.1) \quad \int_{-\infty}^{\infty} p(x)^2 e^{x^2/2} dx < \infty,$$

then p admits a unique representation in the form of the exponential series

$$(8.2) \quad p(x) = \varphi(x) \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x),$$

which converges in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$. Here, the coefficients are given by

$$c_k = \int_{-\infty}^{\infty} H_k(x) p(x) dx = \mathbb{E}H_k(X) = \mathbb{E}(X + iZ)^k,$$

which we call the normal moments of X . In particular, $c_0 = 1, c_1 = \mathbb{E}X$.

In general, c_k exists, as long as the k th absolute moment of X is finite. These moments are needed to develop the characteristic function of X in a Taylor series around zero as follows:

$$(8.3) \quad f(t) = \mathbb{E}e^{itX} = e^{-t^2/2} \sum_{k=0}^N \frac{c_k}{k!} (it)^k + o(|t|^N), \quad t \rightarrow 0.$$

In particular, $c_k = 0$ for $k \geq 1$ in case X is standard normal, similar to the property of the cumulants

$$\gamma_k(X) = \frac{d^k}{i^k dt^k} \log f(t) \Big|_{t=0}$$

with $k \geq 3$ [using the branch of the logarithm determined by $\log f(0) = 0$].

Let us emphasize one simple algebraic property of normal moments.

PROPOSITION 8.1. *Let X be a random variable with $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$ and $\mathbb{E}|X|^k < \infty$ for some integer $k \geq 3$, and let $Z \sim N(0, 1)$. The following three properties are equivalent:*

- (a) $\gamma_j(X) = 0$ for all $j = 3, \dots, k - 1$;
- (b) $\mathbb{E}H_j(X) = 0$ for all $j = 3, \dots, k - 1$;
- (c) $\mathbb{E}X^j = \mathbb{E}Z^j$ for all $j = 3, \dots, k - 1$.

In this case,

$$(8.4) \quad \gamma_k(X) = \mathbb{E}H_k(X) = \mathbb{E}X^k - \mathbb{E}Z^k.$$

PROOF. The repeated differentiation of $f(t)e^{t^2/2} = \mathbb{E}e^{it(X+iZ)}$ leads to $\frac{d^j}{i^j dt^j} [f(t)e^{t^2/2}]|_{t=0} = \mathbb{E}(X + iZ)^j$. Hence, $c_j = \mathbb{E}H_j(X)$, $j \leq N$, in (8.3).

Now, assuming that (b) holds, the expansion (8.3) simplifies to

$$(8.5) \quad f(t) = e^{-t^2/2} \left(1 + \frac{c_k}{k!} (it)^k \right) + o(|t|^k),$$

so that $\log f(t) = -\frac{1}{2}t^2 + \frac{c_k}{k!}(it)^k + o(|t|^k)$. The latter expansion immediately yields (a). The argument may easily be reversed in order to show that (a) \Rightarrow (b) as well. Next, differentiating (8.5) j times at zero, $j \leq k - 1$, we get that $\mathbb{E}X^j = (-i)^j H_j(0) = \mathbb{E}Z^j$. Hence, (c) follows from (b). Moreover, differentiating (8.5) k times at zero, we arrive at $\mathbb{E}X^k = \mathbb{E}Z^k + c_k$, which is the second equality in (8.4). Again, the argument may be reversed in the sense that, starting from (c), we obtain (8.5) and, therefore, (b). Thus, all the three properties are equivalent.

Finally, the first equality in (8.4) is obtained when differentiating the expression $\log f(t) = -\frac{1}{2}t^2 + \frac{c_k}{k!}(it)^k + o(|t|^k)$ k times. \square

The moments of X may be expressed in terms of the normal moments. Indeed, the Chebyshev–Hermite polynomials have the generating function

$$\sum_{k=0}^{\infty} H_k(x) \frac{z^k}{k!} = e^{xz - z^2/2}, \quad x, z \in \mathbb{C},$$

which follows from the identity $H_k(x) = \mathbb{E}(x + iZ)^k$. Equivalently,

$$e^{xz} = e^{z^2/2} \sum_{i=0}^{\infty} H_i(x) \frac{z^i}{i!} = \sum_{i,j=0}^{\infty} H_i(x) \frac{z^{i+2j}}{i!j!2^j}.$$

Expanding e^{xz} into the power series and comparing the coefficients, we get $x^k = k! \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{1}{(k-2j)!j!2^j} H_{k-2j}(x)$. Hence, if $\mathbb{E}|X|^k < \infty$, then

$$(8.6) \quad \mathbb{E}X^k = k! \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{1}{(k-2j)!j!2^j} \mathbb{E}H_{k-2j}(X).$$

Now, let us describe the connection between the normal moments and the χ^2 -distance. The series in (8.3) is absolutely convergent as $N \rightarrow \infty$, when f is analytic in \mathbb{C} . Hence, assuming condition (8.1), we have the expansion

$$(8.7) \quad f(t) = e^{-t^2/2} \sum_{k=0}^{\infty} \frac{c_k}{k!} (it)^k, \quad t \in \mathbb{C}.$$

Moreover, using the Parseval identity in Proposition 7.1, we have

$$(8.8) \quad \sum_{k=0}^{\infty} \frac{c_k^2}{k!} = \int_{-\infty}^{\infty} \frac{p(x)^2}{\varphi(x)} dx = 1 + \chi^2(X, Z),$$

and hence arrive at the following relation.

PROPOSITION 8.2. *If $\chi^2(X, Z) < \infty$, then*

$$(8.9) \quad \chi^2(X, Z) = \sum_{k=1}^{\infty} \frac{1}{k!} (\mathbb{E}H_k(X))^2.$$

Recall that, if $\chi^2(X, Z) < \infty$, then X has finite moments of any order, and moreover, $\mathbb{E}e^{cX^2} < \infty$ for any $c < \frac{1}{4}$. Hence, the normal moments $\mathbb{E}H_k(X)$ are well defined and finite, so that the representation (8.9) makes sense. We now show a converse to Proposition 8.2.

PROPOSITION 8.3. *Let X be a random variable with finite moments of any order. If the series in (8.9) is convergent, then X has an absolutely continuous distribution with finite distance $\chi^2(X, Z)$.*

It looks surprising that a simple sufficient condition for the existence of a density p of X can be formulated in terms of moments of X , only. Note that if X is bounded, then it has finite moments of any order, and the property $\chi^2(X, Z) < \infty$ just means that p is in L^2 .

COROLLARY 8.4. *A bounded random variable X has an absolutely continuous distribution with a square integrable density, if and only if the series in (8.9) is convergent.*

PROOF OF PROPOSITION 8.3. Let $C^2 = \sum_{k=0}^{\infty} \frac{1}{k!} (\mathbb{E}H_k(X))^2$ be finite ($C \geq 1$). Then $|\mathbb{E}H_{2k}(X)| \leq C\sqrt{(2k)!}$ for all $k \geq 0$, and from (8.6) we get

$$\mathbb{E}X^{2k} \leq (2k)! \sum_{j=0}^k \frac{|\mathbb{E}H_{2k-2j}(X)|}{(2k-2j)!j!2^j} \leq C(2k)! \sum_{j=0}^k \frac{1}{\sqrt{(2k-2j)!j!2^j}}.$$

Using $\frac{(2k)!}{(2k-2j)!} \leq (2k)^{2j}$, we obtain that

$$\begin{aligned} \mathbb{E}X^{2k} &\leq C\sqrt{(2k)!} \sum_{j=0}^k \frac{\sqrt{(2k)!}}{\sqrt{(2k-2j)!j!2^j}} \\ &\leq C\sqrt{(2k)!} \sum_{j=0}^k \frac{(2k)^j}{j!2^j} < C\sqrt{(2k)!} \sum_{j=0}^{\infty} \frac{k^j}{j!} = Ce^k\sqrt{(2k)!}. \end{aligned}$$

Thus, $\mathbb{E}X^{2k} < Ce^k\sqrt{(2k)!}$ for all k , which means that $\mathbb{E}e^{cX^2} < \infty$ for some $c > 0$. In particular, X has an entire characteristic function $f(t) = \mathbb{E}e^{itX}$ which admits a power series representation (8.7), where necessarily $c_k = \mathbb{E}H_k(X)$. Consider the N th partial sum of that series

$$f_N(t) = e^{-t^2/2} \sum_{k=0}^N c_k \frac{(it)^k}{k!}.$$

It represents the Fourier transform of $p_N(x) = \varphi(x) \sum_{k=0}^N c_k \frac{H_k(x)}{k!}$ which is the N th partial sum of the series in (8.2). Since $\sum_{k=0}^{\infty} \frac{c_k^2}{k!} < \infty$ (by the assumption), the functions p_N converge to some function p in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$, by Proposition 7.1. In particular, p_N converge in $L^2(\mathbb{R}, dx)$, and by Plancherel's theorem, f_N also converge in $L^2(\mathbb{R}, dx)$ to the Fourier transform \hat{p} of p . But $f_N(t) \rightarrow f(t)$ for all t , so $f(t) = \hat{p}(t)$ almost everywhere. Thus we conclude that f belongs to $L^2(\mathbb{R}, dx)$ and is equal to the Fourier transform of p . Hence, X has an absolutely continuous distribution, and p is density of X .

It remains to use once more the series (8.2). By Proposition 7.1, we have Parseval's equality (8.8), which means that $\chi^2(X, Z) = \sum_{k=0}^{\infty} \frac{c_k^2}{k!} < \infty$. \square

The identity (8.9) admits a natural generalization for the random variables $X_t = \sqrt{t}X + \sqrt{1-t}Z$, where $Z \sim N(0, 1)$ is independent of X .

PROPOSITION 8.5. *If $\chi^2(X, Z) < \infty$, then for all $t \in [0, 1]$,*

$$\chi^2(X_t, Z) = \sum_{k=1}^{\infty} \frac{t^k}{k!} (\mathbb{E}H_k(X))^2.$$

This yields another description of the normal moments via the derivatives of the χ^2 -distance:

$$(\mathbb{E}H_k(X))^2 = \left. \frac{d^k t}{dt^k} \chi^2(X_t, Z) \right|_{t=0}, \quad k = 1, 2, \dots$$

PROOF OF PROPOSITION 8.5. The Hermite polynomials satisfy the binomial formula

$$(8.10) \quad H_k(ax + by) = \sum_{i=0}^k C_k^i a^i b^{k-i} H_i(x) H_{k-i}(y), \quad x, y \in \mathbb{R},$$

whenever $a^2 + b^2 = 1$. In particular, $\mathbb{E}H_k(aX + bZ) = a^k \mathbb{E}H_k(X)$, which may be used in the formula (8.9) with $a = \sqrt{t}$ and $b = \sqrt{1-t}$. \square

9. Behavior of Rényi divergence under convolutions. The obvious question, when describing convergence in the CLT in the D_α -distance is, does it remain finite for sums of independent summands with finite D_α -distances? The answer is affirmative and is made precise in the following.

PROPOSITION 9.1. *Let X and Y be independent random variables. Given $\alpha > 1$, for all $a, b \in \mathbb{R}$ such that $a^2 + b^2 = 1$, we have*

$$D_\alpha(aX + bY||Z) \leq D_\alpha(X||Z) + D_\alpha(Y||Z),$$

where $Z \sim N(0, 1)$. Equivalently,

$$(9.1) \quad \begin{aligned} 1 + (\alpha - 1)T_\alpha(aX + bY||Z) \\ \leq (1 + (\alpha - 1)T_\alpha(X||Z))(1 + (\alpha - 1)T_\alpha(Y||Z)). \end{aligned}$$

The statement may be extended by induction to finitely many independent summands X_1, \dots, X_n by the relation

$$D_\alpha(a_1X_1 + \dots + a_nX_n||Z) \leq D_\alpha(X_1||Z) + \dots + D_\alpha(X_n||Z),$$

where $a_1^2 + \dots + a_n^2 = 1$. For the relative entropy, there is a stronger property,

$$D(a_1X_1 + \dots + a_nX_n||Z) \leq \max\{D(X_1||Z), \dots, D(X_n||Z)\},$$

which follows from the entropy power inequality (cf. [19]). However, this is no longer true for D_α . Nevertheless, for the normalized sums $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$ with i.i.d. summands, Proposition 9.1 guarantees a sublinear growth of the Rényi divergence with respect to n , that is,

$$(9.2) \quad D_\alpha(Z_n||Z) \leq nD_\alpha(X_1||Z).$$

PROOF OF PROPOSITION 9.1. Let Z' be an independent copy of Z , so that the random vector $\tilde{Z} = (Z, Z')$ is standard normal in \mathbb{R}^2 . By Definition 2.1, the Rényi distance of the random vector $\tilde{X} = (X, Y)$ to \tilde{Z} is given by

$$D_\alpha(\tilde{X}||\tilde{Z}) = D_\alpha(X||Z) + D_\alpha(Y||Z').$$

Hence, by the contractivity property (2.1) (cf. Proposition 2.3), we have

$$D_\alpha(S(\tilde{X})||S(\tilde{Z})) \leq D_\alpha(X||Z) + D_\alpha(Y||Z')$$

for any Borel measurable function $S : \mathbb{R}^2 \rightarrow \mathbb{R}$. It remains to apply this inequality with the linear function $S(x, y) = ax + by$. \square

Let us describe a simple alternative argument in the case $\alpha = 2$, which relies upon normal moments only. Assume that $D_2(X||Z)$ and $D_2(Y||Z)$ are finite, so that X and Y have finite moments of any order. Without loss of generality, let $a, b > 0$. From the binomial formula (8.10), it follows that

$$\mathbb{E}H_k(aX + bY) = \sum_{i=0}^k C_k^i a^i b^{k-i} \mathbb{E}H_i(X)\mathbb{E}H_{k-i}(Y).$$

By Cauchy's inequality,

$$\begin{aligned} (\mathbb{E}H_k(aX + bY))^2 &\leq \sum_{i=0}^k C_k^i (a^i b^{k-i})^2 \sum_{i=0}^k C_k^i (\mathbb{E}H_i(X))^2 (\mathbb{E}H_{k-i}(Y))^2 \\ &= \sum_{i=0}^k C_k^i (\mathbb{E}H_i(X))^2 (\mathbb{E}H_{k-i}(Y))^2. \end{aligned}$$

This gives

$$\frac{(\mathbb{E}H_k(aX + bY))^2}{k!} \leq \sum_{i=0}^k \frac{(\mathbb{E}H_i(X))^2}{i!} \frac{(\mathbb{E}H_{k-i}(Y))^2}{(k-i)!},$$

and summation over all integers $k \geq 0$ leads to

$$\sum_{k=0}^{\infty} \frac{(\mathbb{E}H_k(aX + bY))^2}{k!} \leq \sum_{i=0}^{\infty} \frac{(\mathbb{E}H_i(X))^2}{i!} \sum_{j=0}^{\infty} \frac{(\mathbb{E}H_j(Y))^2}{j!}.$$

But, by Proposition 8.2, this inequality is the same as

$$1 + \chi^2(aX + bY, Z) \leq (1 + \chi^2(X, Z))(1 + \chi^2(Y, Z)),$$

which is exactly (9.1) for $\alpha = 2$.

One may also ask whether or not $\chi^2(aX + bY, Z)$ remains finite, when $\chi^2(X, Z)$ is finite, and Y is “small” enough. If p is density of X , the density of $aX + bY$ represents a convex mixture of densities on the line, given by

$$q(x) = \frac{1}{|a|} \mathbb{E} p\left(\frac{x - bY}{a}\right), \quad x \in \mathbb{R}.$$

By Cauchy’s inequality, $\frac{1}{\varphi(x)} q(x)^2 \leq \frac{1}{a^2} \mathbb{E} \frac{1}{\varphi(x)} p\left(\frac{x - bY}{a}\right)^2$. Hence, applying $(ax + by)^2 \leq x^2 + y^2$, we get an elementary bound

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{q(x)^2}{\varphi(x)} dx &\leq \frac{1}{|a|} \mathbb{E} \int_{-\infty}^{\infty} \frac{p(x)^2}{\varphi(ax + bY)} dx \\ &\leq \frac{1}{|a|} \mathbb{E} \int_{-\infty}^{\infty} \sqrt{2\pi} p(x)^2 e^{\frac{1}{2}(x^2 + Y^2)} dx = \frac{1}{|a|} (1 + \chi^2(X, Z)) \mathbb{E} e^{Y^2/2}. \end{aligned}$$

That is, we arrive at the following.

PROPOSITION 9.2. *Let X and Y be independent random variables. For all $a, b \in \mathbb{R}$ such that $a^2 + b^2 = 1$, we have*

$$1 + \chi^2(aX + bY, Z) \leq \frac{1}{|a|} (1 + \chi^2(X, Z)) \mathbb{E} e^{Y^2/2}, \quad Z \sim N(0, 1).$$

Let us now describe two examples of i.i.d. random variables X, X_1, \dots, X_n such that for the normalized sums $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$, and any prescribed integer $n_0 > 1$, we have

$$(9.3) \quad \chi^2(Z_1, Z) = \dots = \chi^2(Z_{n_0-1}, Z) = \infty, \quad \chi^2(Z_{n_0}, Z) < \infty.$$

EXAMPLE 9.3. Suppose that X has density of the form

$$(9.4) \quad p(x) = \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} d\pi(\sigma^2), \quad x \in \mathbb{R},$$

where π is a probability measure on the positive half-axis. The existence of $\chi^2(X, Z)$ implies that $\sigma^2 < 2$ for π -almost all σ^2 , that is, π should be supported on the interval $(0, 2)$. Squaring (9.4) and integrating over x , we find that

$$1 + \chi^2(X, Z) = \int_{-\infty}^{\infty} \frac{p(x)^2}{\varphi(x)} dx = \int_0^2 \int_0^2 \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2 - \sigma_1^2 \sigma_2^2}} d\pi(\sigma_1^2) d\pi(\sigma_2^2).$$

It is easy to see that the last double integral is convergent, if and only if

$$\int_0^1 \int_0^1 \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} d\pi(\sigma_1^2) d\pi(\sigma_2^2)$$

and

$$\int_1^2 \int_1^2 \frac{1}{\sqrt{4 - (\sigma_1^2 + \sigma_2^2)}} d\pi(\sigma_1^2) d\pi(\sigma_2^2)$$

are finite. These conditions may be simplified in terms of the distribution function $F(\varepsilon) = \pi\{\sigma^2 \leq \varepsilon\}$, $0 \leq \varepsilon \leq 2$, by noting that

$$F(\varepsilon/2)^2 \leq (\pi \otimes \pi)\{\sigma_1^2 + \sigma_2^2 \leq \varepsilon\} \leq F(\varepsilon)^2.$$

Hence, the first integral is convergent, if and only if

$$\int_0^1 \frac{1}{\sqrt{\varepsilon}} dF(\varepsilon)^2 = F(1-) ^2 + \frac{1}{2} \int_0^1 \frac{F(\varepsilon)^2}{\varepsilon^{3/2}} d\varepsilon$$

is finite. A similar description applies to the second double integral.

Thus, $\chi^2(X, Z) < \infty$ for the random variable X with density (9.4), iff π is supported on the interval $(0, 2)$, and its distribution function satisfies

$$(9.5) \quad \int_0^1 \frac{F(\varepsilon)^2}{\varepsilon^{3/2}} d\varepsilon < \infty, \quad \int_1^2 \frac{(1 - F(\varepsilon))^2}{(2 - \varepsilon)^{3/2}} d\varepsilon < \infty.$$

Based on this description, we now investigate convolutions. Note that Z_n has density of a similar type as before

$$p_n(x) = \int_0^{\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2} d\pi_n(\sigma^2).$$

More precisely, if ξ_1, \dots, ξ_n are independent copies of a random variable ξ distributed according to π , then the mixing measure π_n can be recognized as the distribution of the normalized sum $S_n = \frac{1}{n}(\xi_1 + \dots + \xi_n)$. Therefore, by (9.5), $\chi^2(Z_n, Z) < \infty$, if and only if $\mathbb{P}\{S_n < 2\} = 1$ and

$$\int_0^1 \frac{F_n(\varepsilon)^2}{\varepsilon^{3/2}} d\varepsilon < \infty, \quad \int_1^2 \frac{(1 - F_n(\varepsilon))^2}{(2 - \varepsilon)^{3/2}} d\varepsilon < \infty,$$

where F_n is the distribution function of S_n . Since $F(\varepsilon/n)^n \leq F_n(\varepsilon) \leq F(\varepsilon)^n$, which is needed near zero, and using similar relations near the point 2, these conditions may be simplified to

$$(9.6) \quad \int_0^1 \frac{F(\varepsilon)^{2n}}{\varepsilon^{3/2}} d\varepsilon < \infty, \quad \int_1^2 \frac{(1 - F(\varepsilon))^{2n}}{(2 - \varepsilon)^{3/2}} d\varepsilon < \infty.$$

Now, suppose that π is supported on $(0, 1)$, so that the second integral in (9.6) is vanishing, and let $F(\varepsilon) \sim \varepsilon^\kappa$ for $\varepsilon \rightarrow 0$ with parameter $\kappa > 0$ (where the equivalence is understood up to a positive factor). Then the first integral in (9.6) will be finite, if and only if $n > 1/(4\kappa)$. Choosing $\kappa = \frac{1}{4(n_0-1)}$, we obtain the required property (9.3).

EXAMPLE 9.4. Consider a density of the form

$$p(x) = \frac{a_k}{1 + |x|^{1/2k}} e^{-x^2/4}, \quad x \in \mathbb{R},$$

where a_k is a normalizing constant, $k = n_0 - 1$, and let f_1 denote its Fourier transform (i.e., the characteristic function). Define the distribution of X via its characteristic function $f(t) = \alpha f_1(t) + (1 - \alpha) \frac{\sin(\gamma t)}{\gamma t}$ with a sufficiently small $\alpha > 0$ and $\gamma = (3(1 + \alpha f_1''(0))/(1 - \alpha))^{1/2}$. It is easy to check that $f''(0) = -1$, which guarantees that $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$. Furthermore, it is not difficult to show that the densities p_n of Z_n admit the two-sided bounds

$$\frac{b'_n}{1 + |x|^{n/2k}} e^{-x^2/4} \leq p_n(x) \leq \frac{b''_n}{1 + |x|^{n/2k}} e^{-x^2/4} \quad (x \in \mathbb{R}),$$

up to some n -dependent factors. Hence, again we arrive at the property (9.3).

10. Superadditivity of χ^2 with respect to marginals. A multidimensional version of Theorem 1.1 requires to involve some other properties of the χ^2 -distance in higher dimensions. The contractivity under mappings,

$$\chi^2(S(X), S(Z)) \leq \chi^2(X, Z),$$

has already been shown in Proposition 2.3 in a general setting. This inequality may be considerably sharpened, when distance is measured to the standard normal law in $\Omega = \mathbb{R}^d$. In order to compare the behavior of χ^2 -divergence with often used information-theoretic quantities, recall the definition of Shannon entropy and Fisher information,

$$h(X) = - \int_{\mathbb{R}^d} p(x) \log p(x) dx, \quad I(X) = \int_{\mathbb{R}^d} \frac{|\nabla p(x)|^2}{p(x)} dx,$$

where X is a random vector in \mathbb{R}^d with density p (assuming that the above integrals exists). These functionals are known to be subadditive and superadditive

with respect to the components: Writing $X = (X', X'')$ with $X' \in \mathbb{R}^{d_1}$, $X'' \in \mathbb{R}^{d_2}$ ($d_1 + d_2 = d$), one always has

$$(10.1) \quad h(X) \leq h(X') + h(X''), \quad I(X) \geq I(X') + I(X'');$$

cf. [16, 30]. Both $h(X)$ and $I(X)$ themselves are not yet distances, so one also considers the relative entropy and the relative Fisher information with respect to other distributions. In particular, in case of the standard normal random vector $Z \sim N(0, I_d)$ and random vectors X with mean zero and identity covariance matrix I_d , they are given by

$$D(X||Z) = h(Z) - h(X), \quad I(X||Z) = I(X) - I(Z).$$

Hence, by (10.1), these distances are both superadditive, that is,

$$D(X||Z) \geq D(X'||Z') + D(X''||Z''), \quad I(X||Z) \geq I(X'||Z') + I(X''||Z''),$$

where Z' and Z'' are standard normal in \mathbb{R}^{d_1} and \mathbb{R}^{d_2} , respectively (both inequalities become equalities, when X' and X'' are independent).

We now establish a similar property for the χ^2 -distance, which can be more conveniently stated in the setting of a Euclidean space H , say of dimension d , with norm $|\cdot|$ and inner product $\langle \cdot, \cdot \rangle$. If X is a random vector in H with density p , and Z is a normal random vector with mean zero and an identity covariance operator I_d , then (according to the abstract definition),

$$\chi^2(X, Z) = \int_H \frac{p(x)^2}{\varphi(x)} dx - 1 = \int_H \frac{(p(x) - \varphi(x))^2}{\varphi(x)} dx,$$

where $\varphi(x) = (2\pi)^{-d/2} e^{-|x|^2/2}$ ($x \in H$) is the density of Z .

PROPOSITION 10.1. *Given a random vector X in H and an orthogonal decomposition $H = H' \oplus H''$ into two linear subspaces $H', H'' \subset H$ of dimensions $d_1, d_2 \geq 1$, for orthogonal projections $X' = \text{Proj}_{H'}(X)$, $X'' = \text{Proj}_{H''}(X)$, we have*

$$(10.2) \quad \chi^2(X, Z) \geq \chi^2(X', Z') + \chi^2(X'', Z''),$$

where Z, Z', Z'' are standard normal random vectors in H, H', H'' , respectively.

Note, however, that (10.2) will not become an equality for independent components X', X'' .

PROOF OF PROPOSITION 10.1. Let $H = \mathbb{R}^d$ and $X = (\xi_1, \dots, \xi_d)$. Note that $\chi^2(X, Z)$ is invariant under orthogonal transformations U of the space, that is, $\chi^2(U(X), Z) = \chi^2(X, Z)$. Hence, without loss of generality, one may assume that $X' = (\xi_1, \dots, \xi_{d_1})$ and $X'' = (\xi_{d_1+1}, \dots, \xi_d)$. Moreover, to simplify the argument

(notationally), let $d_1 = d_2 = 1$. The finiteness of $\chi^2(X, Z)$ then means that the random vector $X = (\xi_1, \xi_2)$ has density $p = p(x_1, x_2)$ such that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2)^2 e^{(x_1^2+x_2^2)/2} dx_1 dx_2 < \infty.$$

The Hermite functions $\varphi_{k_1, k_2}(x_1, x_2) = \varphi(x_1)\varphi(x_2)H_{k_1}(x_1)H_{k_2}(x_2)$ form a complete orthogonal system in $L^2(\mathbb{R}^2)$ (where now φ denotes the one-dimensional standard normal density). Hence, the density p admits a unique representation in the form of the exponential series

$$(10.3) \quad p(x_1, x_2) = \varphi(x_1)\varphi(x_2) \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \frac{c_{k_1, k_2}}{k_1!k_2!} H_{k_1}(x_1)H_{k_2}(x_2),$$

converging in $L^2(\mathbb{R}, \frac{dx_1 dx_2}{\varphi(x_1)\varphi(x_2)})$, with coefficients (mutual normal moments)

$$c_{k_1, k_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_{k_1}(x_1)H_{k_2}(x_2)p(x_1, x_2) dx_1 dx_2 = \mathbb{E}H_{k_1}(\xi_1)H_{k_2}(\xi_2).$$

Moreover, we have Parseval’s equality

$$(10.4) \quad 1 + \chi^2(X, Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{p(x_1, x_2)^2}{\varphi(x_1)\varphi(x_2)} dx_1 dx_2 = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \frac{c_{k_1, k_2}^2}{k_1!k_2!}.$$

Now, integrating (10.3) over x_2 and separately over x_1 , we obtain similar representations for the marginal densities

$$p_1(x_1) = \varphi(x_1) \sum_{k_1=0}^{\infty} \frac{c_{k_1, 0}}{k_1!} H_{k_1}(x_1), \quad p_2(x_2) = \varphi(x_2) \sum_{k_2=0}^{\infty} \frac{c_{0, k_2}}{k_2!} H_{k_2}(x_2).$$

Hence, by Proposition 8.1,

$$\chi^2(\xi_1, \xi) = \sum_{k_1=1}^{\infty} \frac{c_{k_1, 0}^2}{k_1!}, \quad \chi^2(\xi_2, \xi) = \sum_{k_2=1}^{\infty} \frac{c_{0, k_2}^2}{k_2!} \quad [\xi \sim N(0, 1)].$$

But, the quantities $\chi^2(\xi_1, \xi)$ and $\chi^2(\xi_2, \xi)$ appear as summands in (10.4). \square

11. Asymptotic expansions and lower bounds. Let X, X_1, X_2, \dots be independent identically distributed random variables such that $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$, with characteristic function $f(t) = \mathbb{E}e^{itX}$. The normalized sums

$$Z_n = (X_1 + \dots + X_n)/\sqrt{n}$$

weakly converge in distribution to the standard normal law: $Z_n \Rightarrow Z$ for $Z \sim N(0, 1)$. In this connection, the following question arises: When is it true that $D_\alpha(Z_n||Z) \rightarrow 0$ or equivalently $T_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$? And if so, what is the rate of convergence?

We shall give a complete solution of this problem in the next sections. First, we shall describe here asymptotic expansions for “truncated” T_α -distances, which yield reasonable lower bounds for $T_\alpha(Z_n||Z)$. More precisely, given $M > 0$, we have an obvious estimate

$$(11.1) \quad T_\alpha(Z_n||Z) \geq \frac{1}{\alpha - 1} (I(M) - 1)$$

with

$$(11.2) \quad I(M) = \int_{|x| \leq M} \left(\frac{p_n(x)}{\varphi(x)} \right)^\alpha \varphi(x) dx,$$

where p_n denotes the density of Z_n . We will see that, under suitable conditions, while choosing

$$M = M_n(s) = \sqrt{2(s - 1) \log n}$$

with a fixed integer $s \geq 2$, inequality (11.1) can be reversed up to an error term of order $o(n^{-(s-1)})$. This reduces our task to the study of the asymptotic behavior of the integrals $I(M_n(s))$, using the following result due to Petrov (cf. [8, 36, 37]).

PROPOSITION 11.1. *Let X have a finite absolute moment of order $k \geq 3$, and Z_n admit a bounded density for some n . Then, for all n large enough, Z_n have continuous bounded densities p_n satisfying uniformly in $x \in \mathbb{R}$*

$$(11.3) \quad p_n(x) = \varphi(x) + \varphi(x) \sum_{\nu=1}^{k-2} \frac{q_\nu(x)}{n^{\nu/2}} + o\left(\frac{1}{n^{(k-2)/2}}\right) \frac{1}{1 + |x|^k}.$$

In this formula,

$$(11.4) \quad q_\nu(x) = \sum H_{\nu+2l}(x) \prod_{m=1}^{\nu} \frac{1}{k_m!} \left(\frac{\gamma_{m+2}}{(m+2)!} \right)^{k_m},$$

where the sum extends over all nonnegative integer solutions (k_1, k_2, \dots, k_ν) to the equation $k_1 + 2k_2 + \dots + \nu k_\nu = \nu$. Here, γ_r denotes the r th cumulant of X , and we put $l = k_1 + k_2 + \dots + k_\nu$. The sum in (11.3) defines a polynomial in x of degree at most $3(k - 2)$. For example, for $k = 3$ (11.3) yields

$$p_n(x) = \varphi(x) + \frac{\gamma_3}{3! \sqrt{n}} H_3(x) \varphi(x) + o\left(\frac{1}{\sqrt{n}}\right) \frac{1}{1 + |x|^3}, \quad \gamma_3 = \mathbb{E}X^3,$$

where $H_3(x) = x^3 - 3x$. More generally, if $\gamma_3 = \dots = \gamma_{k-1} = 0$, that is, the first $k - 1$ moments of X are the same as for $Z \sim N(0, 1)$, then (11.3) simplifies

$$(11.5) \quad p_n(x) = \varphi(x) + \frac{\gamma_k}{k!} H_k(x) \varphi(x) n^{-\frac{k-2}{2}} + o(n^{-\frac{k-2}{2}}) \frac{1}{1 + |x|^k}.$$

Proposition 11.1 with $k = 2s$ ($s \geq 2$) may be used to derive the following expansion. We use the standard notation $(\alpha)_m = \alpha(\alpha - 1) \dots (\alpha - m + 1)$.

LEMMA 11.2. *Under the assumptions of Proposition 11.1,*

$$(11.6) \quad I(M_n(s)) = 1 + \sum_{j=1}^{s-1} b_j n^{-j} + o(n^{-(s-1)})$$

with

$$(11.7) \quad b_j = \sum \frac{(\alpha)_{m_1+\dots+m_{2j-1}}}{m_1! \dots m_{2j-1}!} \int_{-\infty}^{\infty} q_1(x)^{m_1} \dots q_{2j-1}(x)^{m_{2j-1}} \varphi(x) dx.$$

Here, the sum extends over all integers $m_1, \dots, m_{2j-1} \geq 0$ such that $m_1 + 2m_2 + \dots + (2j - 1)m_{2j-1} = 2j$. In particular, when $\gamma_j = 0$ for $j = 3, \dots, s - 1$ ($s \geq 3$), we have

$$(11.8) \quad I(M_n(s)) = 1 + \alpha(\alpha - 1) \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-2}} + O(n^{-(s-1)}).$$

Using (11.4), one can evaluate the integrals in (11.7) and rewrite them as polynomials in the cumulants $\gamma_3, \dots, \gamma_{2j+1}$, which in turn may be expressed polynomially in terms of the moments $\alpha_r = \mathbb{E}X^r$, $r \leq 2j + 1$.

PROOF OF LEMMA 11.2. The representation (11.3) with $k = 2s$ may be written as

$$\frac{p_n(x)}{\varphi(x)} = 1 + R_n(x) + \frac{\varepsilon_n(x)}{n^{s-1}} \frac{1}{\varphi(x)(1 + |x|^{2s})}, \quad R_n(x) = \sum_{v=1}^{2s-2} \frac{q_v(x)}{n^{v/2}},$$

where $\sup_x |\varepsilon_n(x)| = o(1)$ as $n \rightarrow \infty$. Since every polynomial q_v has degree at most $r = 3(2s - 2)$, we have $|R_n(x)| \leq \frac{C}{\sqrt{n}}(1 + |x|^r)$ up to some constant C . It follows that

$$\left| \frac{p_n(x)}{\varphi(x)} - 1 \right| \leq \frac{C}{\sqrt{n}}(1 + |x|^r) + \frac{1}{\varphi(x)(1 + |x|^{2s})} o(n^{-(s-1)}) \leq \delta_n \rightarrow 0$$

as $n \rightarrow \infty$ uniformly in $|x| \leq M_n(s)$. Using the Lipschitz property of the power function near the point 1, we thus obtain that

$$\left(\frac{p_n(x)}{\varphi(x)} \right)^\alpha = (1 + R_n(x))^\alpha + \frac{1}{\varphi(x)(1 + |x|^{2s})} o(n^{-(s-1)}),$$

so that

$$(11.9) \quad I(M_n(s)) = \int_{|x| \leq M_n(s)} (1 + R_n(x))^\alpha \varphi(x) dx + o(n^{-(s-1)}).$$

Using a Taylor expansion for $(1 + x)^\alpha$ around zero yields

$$(1 + R_n(x))^\alpha = 1 + \sum_{m=1}^{2s-2} \frac{(\alpha)_m}{m!} R_n(x)^m + \frac{C_n(x)}{n^{s-1/2}} (1 + |x|^{r(2s-1)})$$

with $\sup_x |C_n(x)| \leq C$ (where C is a constant). Thus, integration in (11.9) leads to the representation of $I(M_n(s))$ in the form

$$\int_{|x| \leq M_n(s)} \varphi(x) dx + \sum_{m=1}^{2s-2} \frac{(\alpha)_m}{m!} \int_{|x| \leq M_n(s)} R_n(x)^m \varphi(x) dx + O\left(\frac{1}{n^{s-1/2}}\right).$$

Here, the integrals may be extended to the whole real line at the expense of an error $= o(n^{-(s-1)})$. Indeed, with some constant C_l depending on $l \geq 0$,

$$\int_{|x| > M_n(s)} |x|^l \varphi(x) dx \leq C_l M_n(s)^{l-1} e^{-M_n(s)^2/2} = O(n^{-(s-1)} \log^{\frac{l-1}{2}} n),$$

which we use in the polynomial bound on R_n (with factor $1/\sqrt{n}$). Thus,

$$I(M_n(s)) = 1 + \sum_{m=1}^{2s-2} \frac{(\alpha)_m}{m!} \int_{-\infty}^{\infty} R_n(x)^m \varphi(x) dx + o(n^{-(s-1)}).$$

Now, using a multinomial formula, write

$$R_n(x)^m = \sum_{m_1 + \dots + m_{2s-2} = m} \frac{m!}{m_1! \dots m_{2s-2}!} n^{-N/2} q_1(x)^{m_1} \dots q_{2s-2}(x)^{m_{2s-2}},$$

where $N = m_1 + 2m_2 + \dots + (2s - 2)m_{2s-2}$. Hence, up to a $o(n^{-(s-1)})$ -term, one can describe $I(M_n(s)) - 1$ as the sum

$$(11.10) \quad \sum \frac{(\alpha)_m}{m_1! \dots m_{2s-2}!} n^{-N/2} \int_{-\infty}^{\infty} q_1(x)^{m_1} \dots q_{2s-2}(x)^{m_{2s-2}} \varphi(x) dx$$

over all $m_1, \dots, m_{2s-2} \geq 0$ such that $1 \leq m = m_1 + \dots + m_{2s-2} \leq 2s - 2$. This representation simplifies due to the next property of Hermite polynomials [43]:

$$\int_{-\infty}^{\infty} H_{v_1}(x) \dots H_{v_k}(x) \varphi(x) dx = 0 \quad (v_1 + \dots + v_k \text{ is odd}).$$

It follows from (11.4) that a similar property holds for q_j as well: the integral in (11.10) is vanishing, as long as N is odd. Restricting ourselves to the values $N = 2j$, we necessarily have $m_l = 0$ for $l > 2j$, and (11.10) becomes

$$(11.11) \quad \sum \frac{(\alpha)_m}{m_1! \dots m_{2j}!} n^{-j} \int_{-\infty}^{\infty} q_1(x)^{m_1} \dots q_{2j}(x)^{m_{2j}} \varphi(x) dx,$$

where the summation extends over all $m_1, \dots, m_{2j} \geq 0$ with $m_1 + 2m_2 + \dots + 2jm_{2j} = 2j$ and $m = m_1 + \dots + m_{2j}$. Finally, we may exclude the case $m_{2j} = 1$, where again the above integral is vanishing. As a result, we arrive at the required expansion (11.6) with coefficients (11.7). As for the last assertion, we necessarily have $b_j = 0$ for $j = 1, \dots, s - 3$, while $b_{s-2} = \alpha(\alpha - 1)\gamma_s^2/(2s!)$ and then we arrive at (11.8). \square

The integral in (11.11) is zero as well for $m = 1$ (when only one $m_l = 1$). Also, for $\alpha = 2$, the factor $(\alpha)_m$ is vanishing unless $m \leq 2$. Hence, we are reduced to tuples m_1, \dots, m_{2j-1} such that $m_l = 1$ for two different indexes, say, $l = \nu_1$ and $l = \nu_2$, and also for tuples where $m_l = 2$ holds for one l only. Hence, the description of the coefficients is simplified to

$$b_j = \sum_{\substack{\nu_1, \nu_2 > 0 \\ \nu_1 + \nu_2 = 2j}} \int_{-\infty}^{\infty} q_{\nu_1}(x)q_{\nu_2}(x)\varphi(x) dx \quad (\alpha = 2).$$

Recall that if $T_\alpha(Z_n||Z)$ is finite, then $\mathbb{E}e^{cZ_n^2} < \infty$, and hence $\mathbb{E}e^{cX^2} < \infty$ for some $c > 0$ (so that X has finite moments of all orders). In addition, Z_n must have a density in L^2 , and then Z_{n+1} has a bounded density. Therefore, all conditions of Proposition 11.1 are fulfilled, and in view of the lower bound (11.1), Lemma 11.2 yields the following.

PROPOSITION 11.3. *For every fixed $s = 3, 4, \dots$, we have, as $n \rightarrow \infty$,*

$$T_\alpha(Z_n||Z) \geq \frac{1}{\alpha - 1} \sum_{j=1}^{s-2} \frac{b_j}{n^j} + O\left(\frac{1}{n^{s-1}}\right)$$

with coefficients as in (11.7). In particular, if $\gamma_j = 0$ for $j = 3, \dots, s - 1$, then

$$(11.12) \quad T_\alpha(Z_n||Z) \geq \alpha \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-2}} + O\left(\frac{1}{n^{s-1}}\right).$$

The last lower bound extends to D_α as well (which is equivalent to T_α when these two distances are small). Hence we get the following.

COROLLARY 11.4. *If $\liminf_{n \rightarrow \infty} \frac{\log D_\alpha(Z_n||Z)}{\log n} < -K$ for some integer $K > 1$, then $\gamma_j = 0$ for all $j = 3, \dots, K$. In particular, the random variable X is standard normal, if and only if*

$$\liminf_{n \rightarrow \infty} \frac{\log D_\alpha(Z_n||Z)}{\log n} = -\infty.$$

Combining the lower bound (11.2) with the upper bound (9.2) yields the following.

COROLLARY 11.5. *Let $D_\alpha(X||Z) < \infty$ with $\gamma_j = 0$ for $j = 3, \dots, s - 1$ and $\gamma_s \neq 0$ ($s \geq 3$). Then as $n \rightarrow \infty$,*

$$\left(1 + O\left(\frac{1}{n}\right)\right) \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-2}} \leq D_\alpha(Z_n||Z) \leq n D_\alpha(X||Z).$$

12. Necessity part in Theorem 1.2 ($d = 1$). Again, let X, X_1, X_2, \dots be i.i.d. random variables with characteristic function $f(t) = \mathbb{E}e^{itX}$, and let $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$. The necessity part in Theorem 1.2 does not require any moment assumptions on the mean and variance. As a preliminary step, the next lemma provides a sub-Gaussian bound on the Laplace transform $f(iy) = \mathbb{E}e^{-yX}$ subject to the sublinear growth of $D_\alpha(Z_n||Z)$. Recall that $\alpha > 1$, and we denote its conjugate value by $\beta = \alpha/(\alpha - 1)$.

LEMMA 12.1. *If $\liminf_{n \rightarrow \infty} [\frac{1}{n} D_\alpha(Z_n||Z)] = 0$, then*

$$(12.1) \quad f(iy) \leq e^{\beta y^2/2}, \quad y \in \mathbb{R}.$$

PROOF. By Proposition 4.2, applied to Z_n in place of X , for all $y \in \mathbb{R}$,

$$f(iy/\sqrt{n})^n \leq (1 + (\alpha - 1)T_\alpha)^{1/\alpha} e^{\beta y^2/2},$$

where $T_\alpha = T_\alpha(Z_n||Z)$. After a change of the variable, we then get

$$f(iy) \leq \exp\left\{ \frac{1}{\alpha n} \log(1 + (\alpha - 1)T_\alpha) \right\} e^{\beta y^2/2}.$$

But $\liminf_{n \rightarrow \infty} [\frac{1}{n} D_\alpha(Z_n||Z)] = 0$, iff $\liminf_{n \rightarrow \infty} [\frac{1}{n} \log(1 + (\alpha - 1)T_\alpha)] = 0$. Hence, we arrive at the required conclusion by letting $n \rightarrow \infty$ along a suitable subsequence. \square

In other words, if $f(iy_0) > e^{\beta y_0^2/2}$ for some $y_0 \in \mathbb{R}$, then $D_\alpha(Z_n||Z) \geq cn$ with some constant $c > 0$. In this case, $D_\alpha(Z_n||Z)$ has a maximal growth rate, in view of the sublinear upper bound (9.2).

The assumption of Lemma 12.1 is fulfilled, when $D_\alpha(Z_n||Z) \rightarrow 0$, which provides a slightly weakened variant of the necessary condition (1.4) in Theorem 1.2 for dimension $d = 1$ (replacing the strict inequality with a nonstrict inequality). To arrive at a more precise condition, we have to add another preliminary step.

LEMMA 12.2. *If $\lim_{n \rightarrow \infty} D_\alpha(Z_n||Z) = 0$, then for any integer $k \geq \alpha/2$,*

$$(12.2) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(iy/\sqrt{kn})^{2kn} e^{-\beta y^2} dy = \sqrt{\pi(\alpha - 1)}.$$

PROOF. The integral in (12.2) is

$$\begin{aligned} \int_{-\infty}^{\infty} (\mathbb{E}e^{-yZ_{nk}})^2 e^{-\beta y^2} dy &= \int_{-\infty}^{\infty} \mathbb{E}e^{-y(Z_{nk} + Z'_{nk})} e^{-\beta y^2} dy \\ &= \int_{-\infty}^{\infty} \mathbb{E}e^{-\sqrt{2}yZ_{2nk}} e^{-\beta y^2} dy \\ &= \sqrt{\frac{\pi}{\beta}} \mathbb{E}e^{\frac{1}{2\beta} Z_{2nk}^2}, \end{aligned}$$

where Z'_{nk} is an independent copy of Z_{nk} . But Z_{nk} is a normalized sum of k independent copies of Z_n , so we may apply Proposition 5.1 with X replaced by Z_n and with n replaced by $2k$. In this case, by (5.2), whenever $2k \geq \alpha$,

$$|\mathbb{E}e^{\frac{1}{2\beta}Z'^2_{2nk}} - \mathbb{E}e^{\frac{1}{2\beta}Z^2}| \leq 4^k((1 + \chi_\alpha(Z_n||Z)^{1/\alpha})^{2k} - 1), \quad Z \sim N(0, 1).$$

Since, by the assumption, $\chi_\alpha(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$, the limit in (12.2) is equal to $\sqrt{\frac{\pi}{\beta}}\mathbb{E}e^{\frac{1}{2\beta}Z^2}$ which is the same as the right-hand side of (12.2). \square

PROOF OF THE NECESSITY PART IN THEOREM 1.2 FOR $d = 1$. Assume that $D_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$ and fix an integer $k \geq \alpha/2$. Given a fixed number $\delta > 0$, let us decompose

$$\begin{aligned} & \int_{-\infty}^{\infty} f(iy/\sqrt{nk})^{2nk} e^{-\beta y^2} dy \\ (12.3) \quad & = I_1 + I_2 \\ & = \left(\int_{|y| \leq \delta\sqrt{nk}} + \int_{|y| > \delta\sqrt{nk}} \right) f(iy/\sqrt{nk})^{2nk} e^{-\beta y^2} dy. \end{aligned}$$

The characteristic function f is entire, with $f(0) = 1$, hence it is nonvanishing in some disc $|t| < R$ on the complex plane. Define $g(t) = \log f(t)$ for $|t| < R$, choosing the standard branch of the logarithm. The function g is analytic in the same disc and admits a power series representation

$$g(t) = -\frac{1}{2}t^2 + \sum_{m=3}^{\infty} a_m t^m.$$

For a suitable $r \in (0, R)$ and a constant C , we have $\sum_{m=3}^{\infty} |a_m t^m| \leq C|t|^3$ in the disc $|t| \leq r$, so

$$f(iy/\sqrt{nk})^{2nk} = \exp\{y^2 + \theta y^3/\sqrt{n}\} \quad \text{for } y \in [-r\sqrt{nk}, r\sqrt{nk}]$$

with $\theta = \theta(y)$, $|\theta| \leq C$. Assuming that $\delta \leq \min\{r, (\beta - 1)/(2C\sqrt{k})\}$, this relation allows us to rewrite the integral I_1 as

$$I_1 = \int_{|y| \leq \delta\sqrt{nk}} e^{-(\beta-1)y^2 + \theta y^3/\sqrt{n}} dy.$$

Here, the term $\theta y^3/\sqrt{n}$ may be removed at the expense of an error of order $O(\frac{1}{\sqrt{n}})$. This is justified by the bounds

$$\begin{aligned} \int_{|y| \leq \delta\sqrt{nk}} |e^{-(\beta-1)y^2 + \frac{\theta y^3}{\sqrt{n}}} - e^{-(\beta-1)y^2}| dy & \leq \int_{|y| \leq \delta\sqrt{nk}} \frac{C|y|^3}{\sqrt{n}} e^{-(\beta-1)y^2 + \frac{C|y|^3}{\sqrt{n}}} dy \\ & \leq \frac{C}{\sqrt{n}} \int_{|y| \leq \delta\sqrt{nk}} |y|^3 e^{-(\beta-1)y^2/2} dy \\ & = o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Hence

$$\begin{aligned}
 I_1 &= \int_{|y| \leq \delta \sqrt{nk}} e^{-(\beta-1)y^2} dy + O\left(\frac{1}{\sqrt{n}}\right) \\
 &= \sqrt{\pi(\alpha-1)} + O\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.
 \end{aligned}$$

In particular, $I_1 = \sqrt{\pi(\alpha-1)} + o(1)$. Applying this result in (12.3), the equality (12.2) implies that $I_2 \rightarrow 0$, or equivalently

$$(12.4) \quad \int_{|u| > \delta} (f(iu)e^{-\beta u^2/2})^{2nk} du = o\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \rightarrow \infty,$$

which holds for any sufficiently small $\delta > 0$, and hence for any $\delta > 0$.

Now, the function $\psi(u) = f(iu)e^{-\beta u^2/2}$ is analytic, and $0 < \psi(u) \leq 1$ on the real line; cf. (12.1). In order to show that $\psi(u) < 1$ for all $u \neq 0$, suppose for a moment that $\psi(u_0) = 1$ for some $u_0 > 0$. Obviously, u_0 has to be local maximum point, which implies $\psi'(u_0) = 0$. Hence the power series representation at this point, that is,

$$\psi(u) - 1 = c_l(u - u_0)^l + \sum_{j=l+1}^{\infty} c_j(u - u_0)^j$$

starts with a nonzero term $c_l \neq 0$ for some $l \geq 2$. Since $\psi(u) - 1 \leq 0$ for all $u \in \mathbb{R}$, necessarily $l = 2m$ is even ($m \geq 1$) and $c_l < 0$. Hence, in some neighborhood $|u - u_0| \leq r_0 < u_0$ and for some constants $c_1, c_0 > 0$, we have $\psi(u) \geq 1 - c_1(u - u_0)^{2m} \geq e^{-c_0(u - u_0)^{2m}}$. Choosing $\delta = u_0 - r_0$, this neighborhood is contained in (δ, ∞) , and with some constant $c > 0$ we get

$$\begin{aligned}
 \int_{|u| > \delta} (f(iu)e^{-\beta u^2/2})^{2nk} du &\geq \int_{|u - u_0| < \delta} \psi(u)^{2nk} du \\
 &\geq \int_{|u - u_0| < \delta} \exp\{-2nk \cdot c_0(u - u_0)^{2m}\} du \\
 &= 2 \int_0^\delta \exp\{-2nk \cdot c_0 x^{2m}\} dx \geq \frac{c}{n^{1/(2m)}},
 \end{aligned}$$

which contradicts to the asymptotic relation (12.4). The case $u_0 < 0$ is similar, and thus we necessarily arrive at $\psi(u) < 1$ for all real $u \neq 0$. \square

13. Pointwise upper bounds for convolutions of densities. Before turning to the sufficiency part in Theorem 1.2, we shall derive several upper bounds for the densities p_n of the normalized sums Z_n . In general, bounds for the density $p(x)$ of X at individual points x cannot be deduced using the condition $D_\alpha(X||Z) < \infty$. However, this is possible after several convolutions of p with itself, and even under a weaker condition $D_\alpha(Z_{n_0}||Z) < \infty$. The following observation holds without assuming that X has mean zero and variance one. Let f be the characteristic function

of X , and define

$$\psi(u) = f(iu)e^{-\beta u^2/2} = \mathbb{E}e^{-uX}e^{-\beta u^2/2}, \quad u \in \mathbb{R},$$

where $\beta = \frac{\alpha}{\alpha-1}$. As usual, Z denotes a standard normal random variable.

PROPOSITION 13.1. *If $T_\alpha = T_\alpha(Z_{n_0}||Z) < \infty$ for some $n_0 \geq 1$, then for any $n \geq n_\beta = \max(\beta, 2)n_0$, Z_n has a continuous bounded density p_n satisfying*

$$(13.1) \quad p_n(x) \leq \frac{A_\alpha \sqrt{n}}{\sqrt{2\pi n_0}} e^{-x^2/(2\beta)} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n-n_\beta}, \quad x \in \mathbb{R}.$$

Here, $A_\alpha = (1 + (\alpha - 1)T_\alpha)^{k_\alpha}$ with $k_\alpha = \frac{1}{\alpha-1}$ for $\alpha \in (1, 2]$ and $k_\alpha = \frac{2}{\alpha}$ for $\alpha > 2$.

In particular, under the condition (1.1), that is, when $\psi \leq 1$, we arrive at the sub-Gaussian pointwise bound

$$p_n(x) \leq \frac{A_\alpha \sqrt{n}}{\sqrt{2\pi n_0}} e^{-x^2/(2\beta)},$$

which is effective for $|x| \gg \sqrt{\log n}$. It can be sharpened for larger values of $|x|$ by virtue of Proposition 4.3. Being applied to Z_{n_0} in place of X , it gives

$$\lim_{|t| \rightarrow \infty} (\mathbb{E}e^{tX/\sqrt{n_0}})^{n_0} e^{-\beta t^2/2} = 0,$$

that is, $\psi(u) \rightarrow 0$ as $|u| \rightarrow \infty$. Combined with (13.1), it immediately provides an exponential pointwise bound (with respect to n).

COROLLARY 13.2. *If $T_\alpha(Z_{n_0}||Z) < \infty$ for some n_0 , then there exist $x_0 > 0$ and $\delta \in (0, 1)$ such that, for all n large enough,*

$$(13.2) \quad p_n(x) \leq \delta^n e^{-x^2/(2\beta)} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n/2} \quad \text{whenever } |x| \geq x_0\sqrt{n}.$$

Here, the last ψ -term in (13.2) will become crucial for bounding $T_\alpha(Z_n||Z)$.

PROOF OF PROPOSITION 13.1. By the assumption, Z_{n_0} has a density p_{n_0} in $L^\alpha(\mathbb{R}, dx)$, and hence in all $L^\gamma(\mathbb{R}, dx)$, $1 \leq \gamma \leq \alpha$. In particular, if $\alpha \geq 2$, both p_{n_0} and f_{n_0} are in $L^2(\mathbb{R}, dx)$ (by Plancherel’s theorem), so f_n is integrable whenever $n \geq 2n_0$. In the case $1 < \alpha \leq 2$, according to the Hausdorff–Young inequality, f_{n_0} must belong to $L^\beta(\mathbb{R}, dx)$, and then f_n is integrable whenever $n \geq \beta n_0$. Thus, f_n is integrable for every $n \geq n_\beta$, in which case Z_n has a bounded, continuous density given by the Fourier inversion formula

$$p_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t/\sqrt{n})^n dt = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T e^{-itx} f(t/\sqrt{n})^n dt.$$

By Proposition 4.3, $\mathbb{E}e^{cZ_{n_0}^2} < \infty$, and hence $\mathbb{E}e^{cX^2} < \infty$ for some $c > 0$. In particular, $f(t)$ and all characteristic functions $f_n(t) = \mathbb{E}e^{itZ_n} = f(t/\sqrt{n})^n$

are extended as entire functions to the complex plane. Moreover, since the map $h \rightarrow e^{hx} p_n(x)$ for a fixed $n \geq \beta n_0$ is continuous from \mathbb{R} to $L^1(\mathbb{R})$, the family $\{e^{hx} p_n(x)\}_{0 \leq h \leq y}$ is compact in $L^1(\mathbb{R})$. Hence, $f_n(t)$ tends to zero at infinity uniformly in every strip $|\operatorname{Im} t| \leq y < \infty$ (by the Riemann–Lebesgue lemma). Applying Cauchy’s theorem to the rectangle contour $[-T, T] \cup [T, T + iy] \cup [T + iy, -T + iy] \cup [-T + iy, -T]$, the inversion formula may therefore be written as

$$(13.3) \quad p_n(x) = e^{yx} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f((t + iy)/\sqrt{n})^{2n_0} dt$$

for any fixed $y > 0$. Without loss of generality, let $x < 0$.

Case $\alpha > 2, n \geq 2n_0$. Using $|f(t + iy)| \leq f(y)$ ($t, y \in \mathbb{R}$) and changing variable in (13.3), we get

$$(13.4) \quad p_n(x) \leq e^{yx} f\left(\frac{iy}{\sqrt{n}}\right)^{2n_0} \frac{1}{2\pi} \sqrt{\frac{n}{n_0}} \int_{-\infty}^{\infty} \left|f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right)\right|^{2n_0} dt.$$

The function $t \rightarrow f_{n_0}(t + iy/\sqrt{n}) = \mathbb{E}e^{itZ_{n_0} - yZ_{n_0}/\sqrt{n}}$ represents the Fourier transform of $g_y(u) = e^{-yu/\sqrt{n}} p_{n_0}(u)$. Hence, by Plancherel’s theorem,

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right)\right|^{2n_0} dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |f_{n_0}(t + iy\sqrt{n_0}/\sqrt{n})|^2 dt \\ &= \int_{-\infty}^{\infty} e^{-2yu\sqrt{n_0}/\sqrt{n}} p_{n_0}(u)^2 du. \end{aligned}$$

We factorize the latter integrand as $(e^{-2yu\sqrt{n_0}/\sqrt{n}} \varphi(u)^{2/\beta}) \frac{p_{n_0}(u)^2}{\varphi(u)^{2/\beta}}$ and apply Hölder’s inequality with exponents $r = \frac{\alpha}{\alpha-2}, r^* = \frac{\alpha}{2}$. It gives that, up to the factor $(1 + (\alpha - 1)T_\alpha)^{2/\alpha}$, the last integral can be estimated from above by

$$\left(\int_{-\infty}^{\infty} e^{-2ryu\sqrt{\frac{n_0}{n}}} \varphi(u)^{\frac{2r}{\beta}} du\right)^{\frac{1}{r}} = \frac{1}{\sqrt{2\pi}} \left(\frac{\alpha - 2}{2\alpha - 2}\right)^{\frac{1}{2r}} e^{\frac{\beta y^2 n_0}{n}} \leq \frac{1}{\sqrt{2\pi}} e^{\frac{\beta y^2 n_0}{n}}.$$

Therefore,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right)\right|^{2n_0} dt \leq \frac{1}{\sqrt{2\pi}} (1 + (\alpha - 1)T_\alpha)^{2/\alpha} e^{\beta y^2 n_0/n},$$

and (13.4) results in the upper bound

$$\begin{aligned} p_n(x) &\leq \sqrt{\frac{n}{2\pi n_0}} (1 + (\alpha - 1)T_\alpha)^{2/\alpha} e^{yx + \beta y^2 n_0/n} f(iy/\sqrt{n})^{2n_0} \\ &= \sqrt{\frac{n}{2\pi n_0}} (1 + (\alpha - 1)T_\alpha)^{2/\alpha} e^{yx + \beta y^2/2} \psi(y/\sqrt{n})^{2n_0}. \end{aligned}$$

Choosing here $y = -x/\beta$, we arrive at (13.1).

Case $1 < \alpha \leq 2, n \geq \beta n_0$. Again using $|f(t + iy)| \leq f(iy)$ and changing variable, we obtain from (13.3) that

$$(13.5) \quad p_n(x) \leq e^{yx} f(iy/\sqrt{n})^{n-\beta n_0} \frac{1}{2\pi} \sqrt{\frac{n}{n_0}} \int_{-\infty}^{\infty} \left| f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right) \right|^{\beta n_0} dt.$$

Since $\beta \geq 2$, we are allowed to apply the Hausdorff–Young inequality to get

$$\begin{aligned} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right) \right|^{\beta n_0} dt \right)^{\frac{1}{\beta}} &= \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| f_{n_0}\left(t + \frac{iy\sqrt{n_0}}{\sqrt{n}}\right) \right|^{\beta} dt \right)^{\frac{1}{\beta}} \\ &\leq \|g_{y\sqrt{n_0}}\|_{\alpha} \\ &= \left(\int_{-\infty}^{\infty} e^{-\alpha y u \sqrt{n_0}/\sqrt{n}} p_{n_0}(u)^{\alpha} du \right)^{\frac{1}{\alpha}}. \end{aligned}$$

We factorize the latter integrand as $(e^{-\alpha y u \sqrt{n_0}/\sqrt{n}} \varphi(u)^{\alpha-1}) \frac{p_{n_0}(u)^{\alpha}}{\varphi(u)^{\alpha-1}}$ and bound the last integral by $(2\pi)^{-\frac{\alpha-1}{2}} e^{\frac{\alpha\beta y^2 n_0}{2n}}$. This gives

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| f\left(\frac{t}{\sqrt{n_0}} + \frac{iy}{\sqrt{n}}\right) \right|^{\beta n_0} dt &\leq ((2\pi)^{-\frac{\alpha-1}{2}} e^{\frac{\alpha\beta y^2 n_0}{2n}} (1 + (\alpha - 1)T_{\alpha}))^{\frac{\beta}{\alpha}} \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{\beta^2 y^2 n_0}{2n}} (1 + (\alpha - 1)T_{\alpha})^{\frac{1}{\alpha-1}}. \end{aligned}$$

Hence, (13.5) results in the upper bound

$$\begin{aligned} p_n(x) &\leq \sqrt{\frac{n}{2\pi n_0}} (1 + (\alpha - 1)T_{\alpha})^{\frac{1}{\alpha-1}} e^{yx + \beta^2 y^2 n_0/2n} f(iy/\sqrt{n})^{n-\beta n_0} \\ &= \sqrt{\frac{n}{2\pi n_0}} (1 + (\alpha - 1)T_{\alpha})^{\frac{1}{\alpha-1}} e^{yx + \beta y^2/2} \psi(y/\sqrt{n})^{n-\beta n_0}. \end{aligned}$$

Again choosing $y = -x/\beta$, we arrive at (13.1). \square

14. Sufficiency part in Theorem 1.2 ($d = 1$). Let X, X_1, X_2, \dots be i.i.d. random variables such that $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$, with characteristic function $f(t) = \mathbb{E}e^{itX}$. As before, put $\psi(u) = f(iu)e^{-\beta u^2/2}, \beta = \frac{\alpha}{\alpha-1}$, and let $Z \sim N(0, 1)$. Assuming that $\psi(u) < 1$ for all real $u \neq 0$, here it will be shown that the normalized sums

$$Z_n = (X_1 + \dots + X_n)/\sqrt{n}$$

satisfy $T_{\alpha}(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$, as long as $T_{\alpha}(Z_{n_0}||Z) < \infty$ for some n_0 . We also derive an asymptotic expansion for this distance generalizing (1.2) in case $\alpha = 2$. Recall that, by Proposition 13.1, Z_n have densities p_n which are continuous and bounded whenever $n \geq n_{\beta}$.

According to Lemma 11.2, the integrals of the form

$$I_0 = \int_{|x| \leq M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} dx \quad \text{with } M_n = \sqrt{2(s-1) \log n} \quad (s = 3, 4, \dots)$$

admit an asymptotic expansion in powers of $1/n$ up to $1/n^{s-1}$. Hence, for the proof of Theorem 1.2 (in dimension one), it remains to bound the integral of $p_n^\alpha/\varphi^{\alpha-1}$ over the complementary region $|x| > M_n$ by a polynomially small quantity with respect to n . More precisely, it will be sufficient to show that, for any large enough $s \geq 3$ and some constant $\kappa > 0$,

$$(14.1) \quad \int_{|x| > M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} dx = O\left(\frac{1}{n^{\kappa s}}\right), \quad n \rightarrow \infty.$$

To this aim, we need to properly estimate $p_n(x)$, which can be done using the pointwise bounds of the previous section. For definiteness, let us consider the half-axis $x < -M_n$, which we split into three intervals reflecting the possible different behavior of these densities. Namely, define

$$I_1 = \int_{-\infty}^{-x_0\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} dx, \quad I_2 = \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} dx,$$

$$I_3 = \int_{-x_1\sqrt{n}}^{-M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} dx$$

with parameters $0 < x_1 < x_0$ and assuming $M_n < x_1\sqrt{n}$ (otherwise, $I_3 = 0$).

Using (13.2), we get that, for all large n , with some $\delta \in (0, 1)$, $x_0 > 0$,

$$I_1 \leq (2\pi)^{\frac{\alpha-1}{2}} \delta^{\alpha n} \int_{-\infty}^{-x_0\sqrt{n}} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{\alpha n/2} dx$$

$$\leq (2\pi)^{\frac{\alpha-1}{2}} \delta^{\alpha n} \beta\sqrt{n} \int_{-\infty}^{\infty} \psi(u)^m du, \quad m \leq \frac{\alpha n}{2},$$

where on the last step we used $\psi \leq 1$. By Corollary 5.2, the last integral is convergent whenever $m = kn_0$, $k \geq \alpha$. One may take, for example, $k = [\alpha] + 1$, which ensures the condition $m \leq \frac{\alpha n}{2}$ for all sufficiently large n . Hence $I_1 \leq C\delta_1^n$ ($n \geq n_1$) with some constants $C > 0$, $x_0 > 0$ and $\delta < \delta_1 < 1$, depending on the density p only.

To bound I_2 (with any fixed $0 < x_1 < x_0$), we employ Proposition 13.1. By the condition (1.4), the function ψ is bounded away from 1 on any compact interval in $(-\infty, 0)$, so, $\delta_2 = \max_{-x_0 \leq u \leq -x_1} \psi(u) < 1$. Hence, by (13.1),

$$I_2 \leq A_\alpha n^{\alpha/2} \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n-n\beta} dx$$

$$= A_\alpha^\alpha \beta n^{\frac{\alpha+1}{2}} \int_{-x_0/2}^{-x_1/2} \psi(u)^{n-n\beta} du \leq A_\alpha^\alpha \beta n^{\frac{\alpha+1}{2}} (x_0 - x_1) \delta_2^{n-n\beta},$$

which again decays exponentially fast like I_1 .

It remains to estimate I_3 with some $x_1 > 0$. In order to estimate $p_n(x)$ in $[-x_1\sqrt{n}, -M_n]$, we use the bound (13.1) once more. As discussed in Section 12, the function $h(u) = \log f(iu)$ is analytic in some disc $|u| \leq r$, and since $h(0) = h'(0) = 0, h''(0) = 1$, we have $h(u) \sim \frac{1}{2}u^2$ near zero. Hence $|h(u)| \leq \frac{1+\beta}{4}|u|^2$ throughout this disc, when r is sufficiently small, implying $|f(iu)| \leq e^{(1+\beta)|u|^2/4}$. Hence $\psi(u) \leq e^{-(\beta-1)|u|^2/4}$ for u real, $|u| \leq r$, so,

$$\psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n-n\beta} \leq \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n/2} \leq \exp\left\{-\frac{\beta-1}{4}\frac{x^2}{2\beta^2}\right\} = e^{-x^2/(8\alpha\beta)}$$

for all $n \geq 2n_\beta$ and $-\beta r\sqrt{n} < x < 0$. Therefore, by (13.1), in this interval

$$\frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} \leq A_\alpha^\alpha n^{\alpha/2} e^{-x^2/(8\beta)},$$

which results with $x_1 = \beta r$ in the bound

$$\begin{aligned} I_3 &\leq A_\alpha^\alpha n^{\alpha/2} \int_{-x_1\sqrt{n}}^{-M_n} e^{-x^2/(8\beta)} dx \\ &\leq \sqrt{2\pi\beta} A_\alpha^\alpha n^{\alpha/2} e^{-M_n^2/(8\beta)} = \sqrt{2\pi\beta} A_\alpha^\alpha n^{-(\frac{s-1}{4\beta} - \frac{\alpha}{2})}, \end{aligned}$$

where we used a well-known inequality $\int_M^\infty \varphi(x) dx \leq \frac{1}{2}e^{-M^2/2}$ ($M > 0$).

Collecting these bounds, we get $I_1 + I_2 + I_3 = o(n^{-s/8\beta})$ for a sufficiently large s . A similar relation holds for $x > M_n$, which proves (14.1).

Since $T_\alpha(Z_n||Z) = \frac{1}{\alpha-1}(I_0 + I_1 + I_2 + I_3 - 1)$, and using the expansion (11.6), we conclude that, for any sufficiently large s and hence for any $s \geq 3$,

$$(14.2) \quad T_\alpha(Z_n||Z) = \frac{1}{\alpha-1} \sum_{j=1}^{s-2} \frac{b_j}{n^j} + O(n^{-(s-1)})$$

with coefficients b_j as in (11.7). In particular, by (11.8), (14.2) simplifies to

$$(14.3) \quad T_\alpha(Z_n||Z) = \alpha \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-2}} + O(n^{-(s-1)}) \quad \text{if } \gamma_j = 0 \text{ for } j = 3, \dots, s-1.$$

Since D_α and T_α are equivalent (when these quantities are small), the last relation holds true for the Rényi distance $D_\alpha(Z_n||Z)$ as well. Thus, Theorem 1.2 is proved in dimension one. \square

A. Marsiglietti pointed us that expansions like (14.2) have to do with the monotonicity properties of the functionals under consideration.

COROLLARY 14.1. *Given a random variable X ($\mathbb{E}X = 0, \mathbb{E}X^2 = 1$), there exists $n_1 \geq 1$ such that the sequence $D_\alpha(Z_n||Z)$ is nonincreasing for $n \geq n_1$.*

Indeed, assuming that X is non-Gaussian, let $D_\alpha(Z_{n_0}||Z)$ be finite for some n_0 . In particular, X has finite moments of any order. Let γ_s be the first nonzero cumulant of X ($s \geq 3$). Then, according to (14.2)–(14.3),

$$T_\alpha(Z_n||Z) = \alpha \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-2}} + \frac{b_{s-1}}{n^{s-1}} + O(n^{-s}).$$

Hence, the increments

$$T_\alpha(Z_n||Z) - T_\alpha(Z_{n+1}||Z) = \alpha(s-2) \frac{\gamma_s^2}{2s!} \frac{1}{n^{s-1}} + O(n^{-s})$$

are positive for all sufficiently large n (and we know also the rate).

It is interesting to know, whether Corollary 14.1 is true with $n_1 = 1$, which would generalize the monotonicity of the relative entropy along normalized convolutions; cf. [4, 33, 47].

15. Nonuniform local limit theorem. We prove Theorem 1.3 in dimension one in a more precise form, using the cumulants γ_k of X . We keep the basic assumptions $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$ and remind that $\beta = \frac{\alpha}{\alpha-1}$ ($\alpha > 1$).

THEOREM 15.1. *Suppose that $D_\alpha(Z_n||Z)$ is finite for some $n = n_0$, and let the condition (1.4) hold. If $\gamma_3 = \dots = \gamma_{s-1} = 0$ for some $s \geq 3$, then*

$$(15.1) \quad \sup_{x \in \mathbb{R}} \frac{|p_n(x) - \varphi(x)|}{\varphi(x)^{1/\beta}} = \frac{a_s |\gamma_s|}{s!} n^{-\frac{s-2}{2}} + O(n^{-\frac{s-1}{2}}),$$

where $a_s = \sup_{x \in \mathbb{R}} [\varphi(x)^{1/\alpha} |H_s(x)|]$.

In case $s = 3$, we thus obtain the inequality (1.5), and if $\mathbb{E}X^3 = 0$ (and hence $\gamma_3 = 0$), one may turn to the next moment of order $s = 4$, which yields the rate $1/n$ in (15.1). As for the cumulant coefficient, let us recall that $\gamma_s = \mathbb{E}H_s(X) = \mathbb{E}X^s - \mathbb{E}Z^s$.

To compare this result with Proposition 11.1, note that, assuming the existence of moments of order s , and that Z_n has a bounded continuous density p_n for large n , the Edgeworth expansion (11.5) with $k = s$ allows to derive a weaker statement

$$\sup_{x \in \mathbb{R}} (1 + |x|^s) |p_n(x) - \varphi(x)| = \frac{a'_s |\gamma_s|}{s!} n^{-\frac{s-2}{2}} + o(n^{-\frac{s-2}{2}})$$

with $a'_s = \sup_{x \in \mathbb{R}} (1 + |x|^s) |H_s(x)| \varphi(x)$.

Note in addition that the condition (1.4) is almost necessary for the conclusion such as (15.1) and even for a weaker one. Indeed, suppose that

$$(15.2) \quad \liminf_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} (p_n(x) - \varphi(x)) \varphi(x)^{-1/\beta} < \infty,$$

so that $p_n(x) \leq \varphi(x) + C\varphi(x)^{1/\beta}$ for infinitely many n with some constant C . Multiplying this inequality by e^{tx} and integrating, we get

$$(\mathbb{E}e^{tX/\sqrt{n}})^n = \mathbb{E}e^{tZ_n} \leq e^{t^2/2} + BCe^{\beta t^2/2}, \quad B = (2\pi)^{1/(2\alpha)}\sqrt{\beta}.$$

Now substitute t with $t\sqrt{n}$ and raise the above inequality to the power $1/n$. Letting $n \rightarrow \infty$ along a suitable subsequence, we arrive in the limit at $\mathbb{E}e^{tX} \leq e^{\beta t^2/2}$ for all $t \in \mathbb{R}$. Thus, this sub-Gaussian property is indeed implied by the local limit theorem (15.2).

PROOF OF THEOREM 15.1. In contrast with the proof of Theorem 1.2, we need to consider a decomposition into a smaller number of zones, $\Delta_0 : |x| \leq M_n$, $\Delta_1 : |x| \geq x_1\sqrt{n}$, and $\Delta_2 : M_n < |x| < x_1\sqrt{n}$, with $x_1 > 0$ and

$$M_n = \sqrt{2(l-1)\log n}, \quad l = 2\alpha\beta(s-1) + 1.$$

It will be sufficient to restrict the supremum in (15.1) to Δ_0 and to bound

$$J_1 = \sup_{x \in \Delta_1} [p_n(x)\varphi(x)^{-1/\beta}], \quad J_2 = \sup_{x \in \Delta_2} [p_n(x)\varphi(x)^{-1/\beta}]$$

by $O(n^{-\frac{s-1}{2}})$, assuming that $M_n < x_1\sqrt{n}$ (otherwise, $J_2 = 0$). Recall that, as shown in the proof of the sufficiency part of Theorem 1.2,

$$p_n(x)\varphi(x)^{-1/\beta} \leq A_\alpha\sqrt{n}e^{-x^2/(8\alpha\beta)}, \quad |x| \leq x_1\sqrt{n},$$

for some $x_1 > 0$, which we take for the definition of the zones. This gives

$$J_2 \leq A_\alpha\sqrt{n}e^{-M_n^2/(8\alpha\beta)} = A_\alpha n^{-\left(\frac{l-1}{4\alpha\beta} - \frac{1}{2}\right)} = A_\alpha n^{-\frac{s-1}{2}}.$$

Next, we again invoke the bounds of Proposition 13.1 and Corollary 13.2. By the assumption (1.4), $\psi(u) = \mathbb{E}e^{-uX}e^{-\beta u^2/2}$ satisfies $\psi(u) < 1$ for all $u \neq 0$. Hence, the bound (13.2) yields, for all n large enough,

$$p_n(x)\varphi(x)^{-1/\beta} \leq \delta^n, \quad |x| \geq x_0\sqrt{n},$$

with some $\delta \in (0, 1)$, $x_0 > x_1$. Moreover, since $\delta_2 = \max_{x_1 \leq |u| \leq x_0} \psi(u/\beta) < 1$, the bound (13.1) yields for $x_1\sqrt{n} \leq |x| \leq x_0\sqrt{n}$

$$p_n(x)\varphi(x)^{-1/\beta} \leq A_\alpha\sqrt{n}\delta_2^{n-n\beta} = O(\delta_1^n) \quad [n \geq n_\beta = n_0 \max(\beta, 2)]$$

if $\max(\delta, \delta_2) < \delta_1 < 1$. Both estimates imply $J_1 = O(\delta_1^n)$ as $n \rightarrow \infty$.

Finally, in order to study the asymptotic behavior of

$$J_0 = \sup_{x \in \Delta_0} |p_n(x) - \varphi(x)|\varphi(x)^{-1/\beta},$$

we invoke the expansion (11.3) of Proposition 11.1. After division by $\varphi(x)^{1/\beta}$, the remainder term there will be $O(n^{-\frac{s-1}{2}})$ uniformly on Δ_0 , as soon as $\frac{k-2}{2} -$

$\frac{l-1}{\beta} \geq s - 1$. Pick up such an integer k (necessarily $k \geq 3$). As a result, we may replace $p_n(x)$ in the definition of J_0 by the Edgeworth approximation described on the right-hand side of (11.3). The first (potentially) nonzero term in it has the form $\frac{1}{s!} \gamma_s H_s(x) \varphi(x) n^{-\frac{s-2}{2}}$, while all remaining terms are $\varphi(x) q_\nu(x) n^{-\nu/2}$ with $\nu \geq s - 1$. After division by $\varphi(x)^{1/\beta}$, such terms may therefore contribute in the supremum a quantity which is $O(n^{-\frac{s-1}{2}})$. This means that we are left with the leading term, resulting in (15.1). \square

16. The multidimensional case. Let us now turn to the multidimensional variant of Theorems 1.1–1.3. We will denote by Z a standard normal random vector in \mathbb{R}^d , that is, having mean zero and an identity covariance matrix. Given i.i.d. random vectors X, X_1, X_2, \dots in \mathbb{R}^d with mean zero and identity covariance, consider the normalized sums

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n}} \quad (n = 1, 2, \dots).$$

We need to show that $D_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$, if and only if $D_\alpha(Z_n||Z)$ is finite for some $n = n_0$, and

$$(16.1) \quad \mathbb{E}e^{\langle X, t \rangle} < e^{\beta|t|^2/2} \quad \text{for all } t \in \mathbb{R}^d, t \neq 0.$$

Moreover, in this case $D_\alpha(Z_n||Z) = O(1/n)$, and $D_\alpha(Z_n||Z) = O(1/n^2)$ when the distribution of X is symmetric about the origin. In fact, a more precise Edgeworth-type expansion holds for $T_\alpha(Z_n||Z)$ in powers of $1/n$ similar to (14.2)–(14.3), with the coefficients being polynomials of mixed cumulants of the components of X .

As for the proof of the theorems, much of the analysis developed before about the convergence in T_α (or D_α), as well pointwise upper bounds on the densities p_n of Z_n , may easily be extended from dimension one to an arbitrary dimension d . Actually, the contractivity property of the functional D_α (Proposition 2.3) allows one to reduce the necessity part in Theorem 1.2 to the one-dimensional case using a standard Wold-type device. Indeed, consider the i.i.d. sequence $\langle X_j, \theta \rangle$ with unit vectors θ . Then, assuming that $D_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$, we get

$$D_\alpha(\langle Z_n, \theta \rangle || \langle Z, \theta \rangle) \leq D_\alpha(Z_n||Z) \rightarrow 0.$$

Since $\mathbb{E}\langle X_i, \theta \rangle = 0$, $\mathbb{E}\langle X_i, \theta \rangle^2 = 1$, and $\langle Z, \theta \rangle \sim N(0, 1)$, we may apply the one-dimensional variant of this theorem which gives

$$\mathbb{E}e^{r\langle X, \theta \rangle} < e^{\beta r^2/2} \quad \text{for all } r \neq 0.$$

This is the condition (16.1), thus proving the necessity part in Theorem 1.2.

Like in dimension one, the finiteness of $D_\alpha(Z_{n_0}||Z)$ guarantees the existence of bounded continuous densities p_n for Z_n for all $n \geq n_\beta = \max(\beta, 2)n_0$; cf. Proposition 13.1. In addition, $\mathbb{E}e^{c|X|^2} < \infty$ for some $c > 0$. In particular, the characteristic

function $f(t) = \mathbb{E}e^{i\langle X,t \rangle}$ extends as an entire function to the d -dimensional complex space \mathbb{C}^d . Most important properties of the densities p_n rely upon the function

$$\psi(u) = f(iu)e^{-\beta|u|^2/2} = \mathbb{E}e^{-\langle X,u \rangle} e^{-\beta|u|^2/2} \quad (u \in \mathbb{R}^d).$$

LEMMA 16.1. *If $T_\alpha = T_\alpha(Z_{n_0}||Z) < \infty$ for some n_0 , then ψ vanishes at infinity and lies in $L^{kn_0}(\mathbb{R}^d)$ for any integer $k \geq \alpha$. Moreover, up to some (k, d) -dependent constants $c_{k,d}$,*

$$(16.2) \quad \int_{\mathbb{R}^d} \psi(u)^{kn_0} du \leq c_{k,d}(1 + (\alpha - 1)T_\alpha)^{\frac{k}{\alpha}}.$$

The first assertion is a multidimensional analog of Proposition 4.3; it can be proved with very similar arguments as in dimension one. The second assertion generalizing Corollary 5.2 can be proved by using the contractivity properties of the d -dimensional Weierstrass transform

$$W_t u(x) = (2\pi t)^{-d/2} \int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2t}} u(y) dy, \quad x \in \mathbb{R}^d, t > 0.$$

In particular, (5.1) takes the form $\mathbb{E}e^{\frac{1}{2\beta}|Z_k|^2} \leq c_{k,d}(1 + (\alpha - 1)T_\alpha)^{\frac{k}{\alpha}}$ in \mathbb{R}^d , from which (16.2) easily follows. In case $\alpha = 2$, one may adapt Lemma 6.3 as well to the multidimensional situation with its Parseval identity in \mathbb{R}^d . Furthermore, Proposition 6.2 is extended as

$$(2\pi)^{-d/2} \int_{\mathbb{R}^d} \psi(u)^2 du \leq 1 + \chi^2(X, Z),$$

thus refining (16.2) for $k = 2$ and $n_0 = 1$. Repeating the arguments as in Section 13, one may also extend the upper pointwise bounds on the densities.

LEMMA 16.2. *If $T_\alpha(Z_{n_0}||Z) < \infty$ for some n_0 , then for all $x \in \mathbb{R}^d$,*

$$(16.3) \quad p_n(x) \leq A_{\alpha,d} n^{d/2} e^{-|x|^2/(2\beta)} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n-n\beta}, \quad n \geq n\beta,$$

where $A_{\alpha,d}$ depends on (α, d) only. In particular, there exist constants $x_0 > 0$ and $\delta \in (0, 1)$ depending on the density p such that for all n large enough

$$(16.4) \quad p_n(x) \leq \delta^n e^{-|x|^2/(2\beta)} \psi\left(-\frac{x}{\beta\sqrt{n}}\right)^{n/2} \quad \text{whenever } |x| \geq x_0\sqrt{n}.$$

PROOF OF THEOREM 1.2 (SUFFICIENCY PART) AND THEOREM 1.3. We need to explore the asymptotic behavior of

$$(\alpha - 1)T_\alpha(Z_n||Z) = \int_{\mathbb{R}^d} w_n^\alpha(x) dx - 1, \quad w_n(x) = p_n(x)\varphi(x)^{-1/\beta},$$

where φ is the standard normal density on \mathbb{R}^d . To this aim, we split the integration into the four shell-type regions. The behavior of the integrals

$$I_0 = \int_{|x| < M_n} w_n^\alpha(x) dx, \quad M_n = \sqrt{2(l-1) \log n},$$

may be studied as in dimension one (Lemma 11.2) by virtue of the Edgeworth expansion for $p_n(x)$ on the balls $|x| < M_n$ with a nonuniform error term. To this aim, a multidimensional variant of Proposition 11.1 is used as stated in [7], Theorem 19.2: Uniformly in \mathbb{R}^d ,

$$(16.5) \quad p_n(x) = \varphi_s(x) + \frac{o(n^{-(s-2)/2})}{1 + |x|^s}, \quad \varphi_s(x) = \varphi(x) + \varphi(x) \sum_{k=1}^{s-2} \frac{q_k(x)}{n^{k/2}},$$

where each q_k represents a polynomial whose coefficients involve mixed cumulant of the components of X of order up to $k + 2$. In particular, if the distribution of X is symmetric about the origin, then $q_1(x) = 0$ and there is no $1/\sqrt{n}$ term in (16.5). In this way, we will arrive at the Edgeworth-type expansion for I_0 similar to dimension one, which implies that $I_0 - 1 = O(1/n)$ in general, and $I_0 - 1 = O(1/n^2)$ when the distribution of X is symmetric.

As a result, it remains to establish a polynomial smallness of the integrals

$$I_1 = \int_{|x| > x_0 \sqrt{n}} w_n^\alpha(x) dx, \quad I_2 = \int_{x_1 \sqrt{n} < |x| < x_0 \sqrt{n}} w_n^\alpha(x) dx,$$

$$I_3 = \int_{M_n < |x| < x_1 \sqrt{n}} w_n^\alpha(x) dx$$

with $x_1 > 0$ being any fixed small number, and $x_0 > x_1$ depending on the density p . The bounds (16.2)–(16.4) allow us to properly estimate these integrals as functions of n , by modifying the arguments from the previous section. Using (16.4) and (16.2) with $k = [\alpha] + 1$ and assuming that $\psi \leq 1$, we get for all n large enough

$$I_1 \leq C_1 \delta^{\alpha n} \int_{|x| > x_0 \sqrt{n}} \psi \left(-\frac{x}{\beta \sqrt{n}} \right)^{\alpha n/2} dx \leq C_1 \delta^{\alpha n} \int_{\mathbb{R}^d} \psi(u)^{kn_0} du \leq C_2 \delta_1^n$$

with some constants C_j , $x_0 > 0$ and $0 < \delta < \delta_1 < 1$ not depending on n .

For the region of I_2 , thanks to (1.4), $\delta_2 = \max_{x_0 \leq |u| \leq x_1} \psi \left(\frac{u}{\beta} \right) < 1$. Hence, by (16.3), putting $n_1 = n - n\beta$, we get that with some constants $C_j > 0$

$$I_2 \leq C_1 n^{d\alpha/2} \int_{x_1 \sqrt{n} < |x| < x_0 \sqrt{n}} \psi \left(-\frac{x}{\beta \sqrt{n}} \right)^{n_1} dx$$

$$= C_2 n^{d(\alpha+1)/2} \int_{x_1 \sqrt{n} < |x| < x_0 \sqrt{n}} \psi(u)^{n_1} du \leq C_3 n^{d(\alpha+2)/2} x_0^d \delta_2^{n_1},$$

which is decaying exponentially fast like I_1 . Finally, using the analyticity of f , we have $\psi(u) \leq e^{-(\beta-1)|u|^2/4}$ in a sufficiently small ball $|u| < r$, so that

$$\psi \left(-\frac{x}{\beta \sqrt{n}} \right)^{n_1} \leq \psi \left(-\frac{x}{\beta \sqrt{n}} \right)^{n/2} \leq e^{-x^2/(8\alpha\beta)}, \quad |x| < \beta r \sqrt{n}, n \geq 2n\beta.$$

Therefore, by (16.2), in this ball $w_n(x) \leq A_{\alpha,d}^\alpha n^{d\alpha/2} e^{-|x|^2/(8\alpha\beta)}$, which gives

$$I_3 \leq C_1 n^{d\alpha/2} \int_{M_n < |x| < x_1 \sqrt{n}} e^{-|x|^2/(8\alpha\beta)} dx$$

$$< C_2 n^{d\alpha/2} \mathbb{P} \left\{ |Z|^2 > \frac{M_n^2}{8d\alpha\beta} \right\} \leq C_3 n^{d\alpha/2} e^{-\frac{M_n^2}{8d\alpha\beta}} = C_3 n^{-\left(\frac{l-1}{4d\alpha\beta} - \frac{\alpha}{2}\right)}$$

with $x_1 = \beta r$. Collecting these bounds, we get that $I_1 + I_2 + I_3 = o(n^{-l/8d\alpha\beta})$ for all sufficiently large l , thus proving Theorem 1.2.

For the proof of Theorem 1.3 in \mathbb{R}^d , we need to investigate the suprema

$$J_0 = \sup_{|x| \leq M_n} |p_n(x) - \varphi_s(x)| \varphi(x)^{-1/\beta},$$

$$J_1 = \sup_{|x| \geq x_1 \sqrt{n}} p_n(x) \varphi(x)^{-1/\beta}, \quad J_2 = \sup_{M_n \leq |x| \leq x_1 \sqrt{n}} p_n(x) \varphi(x)^{-1/\beta}$$

with some $x_1 > 0$ and assuming that $M_n < x_1 \sqrt{n}$. An application of (16.5) implies that $J_0 = O(1/\sqrt{n})$ in general and $J_0 = O(1/n)$ when the distribution of X is symmetric. The polynomial smallness of J_1 and J_2 (for sufficiently large values of l in the definition of M_n) follows from Lemma 16.2, by repeating the arguments of the proof of Theorem 15.1. \square

17. Some examples and counterexamples. Given a random variable X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, consider the function $\psi(t) = e^{-t^2} \mathbb{E}e^{tX}$ ($t \in \mathbb{R}$). As before, put $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$, where X_j are independent copies of X . One immediate consequence of Theorem 1.1 (with $n_0 = 1$) is the following characterization. As usual, Z denotes a standard normal r.v.

THEOREM 17.1. *Let the random variable X have a density p such that*

$$(17.1) \quad \int_{-\infty}^{\infty} p(x)^2 e^{x^2/2} dx < \infty.$$

Then $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$, if and only if $\psi(t) = e^{-t^2} \mathbb{E}e^{tX}$ satisfies

$$(17.2) \quad \psi(t) < 1 \quad \text{for all } t \neq 0.$$

The assumption (17.1) is fulfilled, for example, when X is bounded and has a square integrable density. We now illustrate Theorem 17.1 and the more general Theorem 1.2 with a few examples (mostly in dimension one).

Uniform distribution. Let X be uniformly distributed on the segment $[-\sqrt{3}, \sqrt{3}]$. The characteristic function of X is given by $f(t) = \sin(t\sqrt{3})/(t\sqrt{3})$, and for imaginary values $t = iy$, we have the simple estimate

$$(17.3) \quad f(iy) = \frac{\sinh(y\sqrt{3})}{y\sqrt{3}} < e^{y^2/2}, \quad y \in \mathbb{R} \ (y \neq 0),$$

so that (17.2) holds. In this case, the first moments are given by $\alpha_2 = 1$, $\alpha_3 = 0$, $\alpha_4 = \frac{9}{5}$. Therefore, by Theorem 17.1, $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$. Moreover, Theorem 1.1 provides an asymptotic expansion (1.3) which becomes

$$\chi^2(Z_n, Z) = \frac{3}{50n^2} + O\left(\frac{1}{n^3}\right).$$

In fact, the property (17.3) means that the condition (1.4) of a more general Theorem 1.2 is fulfilled in the whole range of indexes $\alpha > 1$. Using the formula (14.3), we therefore obtain a stronger assertion $T_\alpha(Z_n || Z) = \frac{\alpha}{2} \chi^2(Z_n, Z) + O(\frac{1}{n^3})$ and a similar one for D_α .

Convex mixtures of centered Gaussian measures. Consider the densities of the form

$$p(x) = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} d\pi(\sigma^2), \quad x \in \mathbb{R},$$

where π is a probability measure on the positive half-axis with $\int_0^\infty \sigma^2 d\pi(\sigma^2) = 1$. The random variable X with this density has mean zero and variance one, and its distribution is equal to that of $\sqrt{\xi}Z$, where ξ is independent of Z and is distributed according to π . As in Example 9.3, $\chi^2(Z_n, Z) < \infty$ for some $n = n_0$, if and only if π is supported on the interval $(0, 2)$, and its distribution function $F(\varepsilon) = \pi((0, \varepsilon])$ satisfies the condition (9.6).

On the other hand, the distribution of X has the Laplace transform

$$\mathbb{E}e^{tX} = \int_0^\infty e^{\sigma^2 t^2/2} d\pi(\sigma^2) = \mathbb{E}e^{\xi t^2/2}, \quad t \in \mathbb{R}.$$

Hence, the condition $\chi^2(Z_n, Z) < \infty$ guarantees that (17.2) is fulfilled. Without that condition, $\mathbb{E}e^{tX} < e^{t^2}$ for all $t \neq 0$, if and only if $\mathbb{P}\{\xi \leq 2\} = 1$ and $\mathbb{P}\{\xi = 2\} < 1$. Here, $\mathbb{P}\{\xi = 2\} = 1$ is not possible in view of $\mathbb{E}X^2 = \mathbb{E}\xi = 1$.

Hence, $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$, if and only if the measure π is supported on the interval $(0, 2)$ and satisfies the condition (9.6). In this case, we obtain the expansion (1.3) which reads

$$\chi^2(Z_n, Z) = \frac{3(m-1)^2}{8n^2} + O\left(\frac{1}{n^3}\right), \quad m = \int_0^\infty \sigma^4 d\pi(\sigma^2).$$

Distributions with Gaussian component. Consider random variables

$$X = a\xi + bZ \quad (a^2 + b^2 = 1, a, b > 0),$$

assuming that $\mathbb{E}\xi = 0$, $\mathbb{E}\xi^2 = 1$, and where $Z \sim N(0, 1)$ is independent of ξ . The distribution of X is a convex mixture of shifted Gaussian measures with variance b^2 . It admits a density

$$p(x) = \frac{1}{b} \mathbb{E}\varphi\left(\frac{x - a\xi}{b}\right), \quad x \in \mathbb{R}.$$

To ensure finiteness of $\chi^2(Z_n, Z)$ with some n , the random variable ξ should have a finite Gaussian moment, or equivalently, the Laplace transform of the distribution of ξ should admit a sub-Gaussian bound

$$(17.4) \quad \mathbb{E}e^{t\xi} \leq e^{\sigma^2 t^2/2}, \quad t \in \mathbb{R},$$

with some finite $\sigma > 0$. Let σ be an optimal value in this inequality (necessarily $\sigma \geq 1$). It then follows that $\mathbb{E}e^{c\xi^2} < \infty$ whenever $c < 1/(2\sigma^2)$.

Squaring $p(x)$, we easily find an expression for the χ^2 -distance,

$$1 + \chi^2(X, Z) = \frac{1}{\sqrt{1-a^4}} \mathbb{E} \exp \left\{ \frac{a^2}{2(1-a^2)} \left(\frac{2}{1+a^2} (\xi + \eta)^2 - (\xi^2 + \eta^2) \right) \right\}$$

with η an independent copy of ξ . Using $(\xi + \eta)^2 \leq 2\xi^2 + 2\eta^2$, we get

$$1 + \chi^2(X, Z) \leq \frac{1}{\sqrt{1-a^4}} \left(\mathbb{E}e^{\frac{a^2}{2(1+a^2)} \xi^2} \right)^2.$$

Hence, $\chi^2(X, Z) < \infty$ whenever $a < a_\sigma = \frac{1}{\sqrt{\sigma^2-1}}$, which is automatically fulfilled in case $\sigma^2 \leq 2$. Moreover, for all $t \neq 0$, from (17.4),

$$\mathbb{E}e^{tX} = \mathbb{E}e^{at\xi} e^{b^2 t^2/2} \leq e^{(\sigma^2 a^2 + b^2) t^2/2} = e^{((\sigma^2-1)a^2-1)t^2/2} < e^{t^2}$$

under the same constraint $a < a_\sigma$. Thus, $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$, if $a < \frac{1}{\sqrt{\sigma^2-1}}$.

In case $\sigma^2 \leq 2$, this convergence holds for all admissible (a, b) .

Distributions with finite Gaussian moment. If a random variable X with mean zero and variance, one has finite $M = \mathbb{E}e^{cX^2}$ ($c > 0$), then (17.4) is fulfilled for some $\sigma \geq 1$, in fact with $\sigma^2 \leq \frac{4 \log M}{c \log 2}$. This means that (1.4) is fulfilled for any $\alpha > 1$ such that $\beta < \sigma^2$. Therefore, if $D_\alpha(X||Z) < \infty$, then $D_\alpha(Z_n||Z) \rightarrow 0$ with any $\alpha < \frac{\sigma^2}{\sigma^2-1}$.

Conditions in terms of exponential series. Consider a symmetric density of the form

$$p(x) = \varphi(x) \sum_{k=0}^{\infty} \frac{\sigma_k}{2^k k!} H_{2k}(x), \quad x \in \mathbb{R},$$

with $\sigma_0 = 1$ and $\sigma_1 = 0$ (i.e., $\mathbb{E}X^2 = 1$ for the random variable with density p). In view of Section 6, condition (17.1) is fulfilled, if and only if the series

$$\chi^2(X, Z) = \sum_{k=2}^{\infty} \frac{(2k)!}{4^k k!^2} \sigma_k^2 \sim \sum_{k=2}^{\infty} \frac{1}{\sqrt{k}} \sigma_k^2$$

is convergent (which is fulfilled automatically, when p is compactly supported and bounded). Assuming additionally that $\sup_{k \geq 2} \sigma_k \leq 1$, we also have

$$\mathbb{E}e^{tX} = e^{t^2/2} \left[1 + \sum_{k=2}^{\infty} \frac{\sigma_k}{k!} \left(\frac{t^2}{2} \right)^k \right] \leq e^{t^2/2} \left(e^{t^2/2} - \frac{t^2}{2} \right) < e^{t^2}, \quad t \neq 0.$$

Hence, in this case, by Theorem 17.1, $\chi^2(Z_n, Z) \rightarrow 0$ as $n \rightarrow \infty$. Moreover, according to the expansion (1.3), we have $\chi^2(Z_n, Z) = O(1/n^2)$. This assertion strengthens the result of [20] (under weaker assumptions).

Log-concave probability distributions. More examples including those in higher dimensions illustrate the multidimensional Theorem 1.2 within the class of densities $p(x) = e^{-V(x)}$ supported on some open convex region $\Omega \subset \mathbb{R}^d$. Let V be a C^2 -convex function with Hessian satisfying $V''(x) \geq cI_d$ in the sense of positive definite matrices ($c > 0$). The probability measures with such densities are known to admit logarithmic Sobolev inequalities (via the Bakry–Emery criterion). In particular, they satisfy transport-entropy inequalities which in turn can be used to get a sub-Gaussian bound

$$\mathbb{E}e^{rg(X)} \leq e^{r^2/(2c)}, \quad r \in \mathbb{R}.$$

Here, g may be an arbitrary function on \mathbb{R}^d with Lipschitz semi-norm $\|g\|_{\text{Lip}} \leq 1$, such that $\mathbb{E}g(X) = 0$ (cf. [13, 35]). In particular, if $\mathbb{E}X = 0$, one may choose an arbitrary linear function $g(x) = \langle x, \theta \rangle$ with $|\theta| = 1$. Hence, the condition (1.4) will be fulfilled, as long as $c > \frac{1}{\beta}$. Moreover, the property $D_\alpha(X||Z) < \infty$ will also hold in this case, since necessarily

$$V(x) \geq V(x_0) + \langle V'(x_0), x - x_0 \rangle + \frac{c}{2}|x - x_0|^2$$

for all $x, x_0 \in \Omega$. Applying Theorem 1.2, we get the following.

COROLLARY 17.2. *If a random vector X in \mathbb{R}^d with mean zero and identity covariance matrix has density $p = e^{-V}$ such that $V'' \geq cI_d$ ($0 < c \leq 1$) on the supporting open convex region, then $D_\alpha(Z_n||Z) \rightarrow 0$ as $n \rightarrow \infty$, whenever $\alpha < \frac{1}{1-c}$.*

18. Convolution of Bernoulli with Gaussian. One might wonder whether or not it is possible to replace the condition (1.1) in Theorem 1.1 with a slightly weaker requirement $\mathbb{E}e^{tX} \leq e^{t^2}$ (hoping, e.g., that the strict inequality would automatically hold, in view of the assumption $\mathbb{E}X^2 = 1$). The answer is negative, including the D_α -case as in Theorem 1.2 with its condition (1.4). Put $\beta = \frac{\alpha}{\alpha-1}$ for a fixed $\alpha > 1$.

PROPOSITION 18.1. *There exists a random variable X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, and $D_\alpha(X||Z) < \infty$ for $Z \sim N(0, 1)$, such that the inequality*

$$(18.1) \quad \mathbb{E}e^{tX} < e^{\beta t^2/2}$$

is fulfilled for all $t \neq 0$ except for exactly one point $t_0 \neq 0$.

Since (18.1) is violated (although at one point only), Theorem 1.2 implies that convergence $D_\alpha(Z_n||Z) \rightarrow 0$ does not hold.

Let us describe explicitly one family of distributions satisfying the assertion of this proposition. Returning to one of the previous examples, consider random variables of the form $X_p = a\xi + bZ$ ($a, b > 0$), assuming that ξ takes two values q and $-p$ with probabilities p and q , respectively ($p, q > 0$, $p + q = 1$), while $Z \sim N(0, 1)$ is independent of ξ . Then $\mathbb{E}X_p = 0$, and we have the constraint

$$(18.2) \quad \mathbb{E}X_p^2 = pqa^2 + b^2 = 1.$$

The density w of X_p is a convex mixture of two shifted Gaussian densities,

$$w(x) = \frac{p}{b}\varphi\left(\frac{x-aq}{b}\right) + \frac{q}{b}\varphi\left(\frac{x+ap}{b}\right),$$

and the condition $D_\alpha(X||Z) < \infty$ obviously holds (since necessarily $b < 1$).

Let $\sigma^2 = \sigma^2(p, q)$ denote the smallest positive constant in the inequality

$$(18.3) \quad \mathbb{E}e^{t\xi} = pe^{qt} + qe^{-pt} \leq e^{\sigma^2 t^2/2}, \quad t \in \mathbb{R}.$$

This is the so-called sub-Gaussian constant for the Bernoulli distribution. Since $\mathbb{E}e^{tX_p} = \mathbb{E}e^{at\xi} e^{b^2 t^2/2}$, (18.3) yields

$$\mathbb{E}e^{tX_p} \leq e^{(\sigma^2 a^2 + b^2)t^2/2}, \quad t \in \mathbb{R},$$

with an optimal constant $\sigma^2 a^2 + b^2$ in the exponent on the right-hand side. Thus, according to the requirement (18.1), we get another constraint $\sigma^2 a^2 + b^2 = \beta$. Combining it with (18.2), we find that necessarily

$$a^2 = \frac{\beta - 1}{\sigma^2 - pq}, \quad b^2 = \frac{\sigma^2 - \beta pq}{\sigma^2 - pq},$$

which makes sense if $\sigma^2 > \beta pq$. According to [14], Proposition 2.3, the sub-Gaussian constant for the Bernoulli distribution is known to be

$$\sigma^2 = \frac{p - q}{2(\log p - \log q)}.$$

It is easy to see that (18.3) becomes equality for $t_0 = -2(\log p - \log q)$, which is a unique nonzero point with such property, as long as $p \neq q$. Hence, the random variable X_p satisfies the assertion of Proposition 18.1, if and only if

$$(18.4) \quad \frac{p - q}{2(\log p - \log q)} > \beta pq.$$

This inequality does hold, provided that p is sufficiently close to 0 or 1, although it is not true for a neighborhood of $1/2$ (since at this point the inequality becomes $1 > \beta$). More precisely, for some constant $p_\alpha \in (0, \frac{1}{2})$, (18.4) holds for all p from the set $(0, p_\alpha) \cup (1 - p_\alpha, 1)$, while for p from $(p_\alpha, 1 - p_\alpha)$ it holds with an opposite inequality sign.

REFERENCES

- [1] AMOSOVA, N. N. (1990). Narrow zones of local normal attraction. *Teor. Veroyatn. Primen.* **35** 138–143. Translation in: *Theory Probab. Appl.* **35** (1990) 140–145 (1991). [MR1050062](#)
- [2] AMOSOVA, N. N. (1990). A remark on a local limit theorem for large deviations. *Teor. Veroyatn. Primen.* **35** 754–756. Translation in: *Theory Probab. Appl.* **35** (1990), 758–760 (1991). [MR1090501](#)
- [3] 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](#)
- [4] 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. [MR2083473](#)
- [5] BALLY, V. and CAMELLINO, L. (2016). Asymptotic development for the CLT in total variation distance. *Bernoulli* **22** 2442–2485. [MR3498034](#)
- [6] BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. [MR0815975](#)
- [7] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. [MR0436272](#)
- [8] 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](#)
- [9] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2013). Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **41** 2479–2512. [MR3112923](#)
- [10] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2014). Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159** 435–478. [MR3230000](#)
- [11] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2014). Fisher information and the central limit theorem. *Probab. Theory Related Fields* **159** 1–59. [MR3201916](#)
- [12] BOBKOV, S. G., CHISTYAKOV, G. P. and KÖSTERS, H. (2015). The entropic Erdős–Kac limit theorem. *J. Theoret. Probab.* **28** 1520–1555. [MR3422941](#)
- [13] BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. [MR1682772](#)
- [14] BOBKOV, S. G., HOUDRÉ, C. and TETALI, P. (2006). The subgaussian constant and concentration inequalities. *Israel J. Math.* **156** 255–283. [MR2282379](#)
- [15] BORLAND, L., PLASTINO, A. R. and TSALLIS, C. (1998). Information gain within nonextensive thermostatics. *J. Math. Phys.* **39** 6490–6501. [MR1656933](#)
- [16] CARLEN, E. A. (1991). Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.* **101** 194–211. [MR1132315](#)
- [17] CRAMÉR, H. (1925). On some classes of series used in mathematical statistics. In *Proc. 6th Scand. Math. Congr. Copenhagen* 399–425. Also: HARALD CRAMÉR (1994) *Collected Works, Vol. I* (A. Martin-Löf, ed.) 438–464. Springer, Berlin. [MR1280267](#)
- [18] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318. [MR0219345](#)
- [19] DEMBO, A., COVER, T. M. and THOMAS, J. A. (1991). Information-theoretic inequalities. *IEEE Trans. Inform. Theory* **37** 1501–1518. [MR1134291](#)
- [20] FOMIN, S. V. (1982). The central limit theorem: Convergence in the norm $\|u\| = (\int_{-\infty}^{\infty} u^2(x)e^{x^2/2} dx)^{1/2}$. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **119** 218–229, 242, 245. Problems of the theory of probability distribution, VII. [MR0666101](#)
- [21] GIBBS, A. L. and SU, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** 419–435.

- [22] GILARDONI, G. L. (2010). On Pinsker's and Vajda's type inequalities for Csiszár's f -divergences. *IEEE Trans. Inform. Theory* **56** 5377–5386. [MR2808583](#)
- [23] HIRSCHMAN, I. I. and WIDDER, D. V. (1955). *The Convolution Transform*. Princeton Univ. Press, Princeton, NJ. [MR0073746](#)
- [24] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing, Groningen. [MR0322926](#)
- [25] JOHNSON, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press, London. [MR2109042](#)
- [26] JOHNSON, O. and BARRON, A. (2004). Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields* **129** 391–409. [MR2128239](#)
- [27] KULLBACK, S. (1967). A lower bound for discrimination in terms of variation. *IEEE Trans. Inform. Theory* **13** 126–127.
- [28] KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. [MR0039968](#)
- [29] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. [MR0856411](#)
- [30] LIEB, E. H. (1975). Some convexity and subadditivity properties of entropy. *Bull. Amer. Math. Soc.* **81** 1–13. [MR0356797](#)
- [31] LIESE, F. and VAJDA, I. (1987). *Convex Statistical Distances*. Teubner-Texte zur Mathematik [Teubner Texts in Mathematics] **95**. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.
- [32] LINNIK, JU. V. (1959). An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Theory Probab. Appl.* **4** 288–299. [MR0124081](#)
- [33] MADIMAN, M. and BARRON, A. (2007). Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory* **53** 2317–2329. [MR2319376](#)
- [34] NIELSEN, F. (2014). On the Chi square and higher-order Chi distances for approximating f -divergences. *IEEE Signal Process. Lett.* **21** 10–13.
- [35] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. [MR1760620](#)
- [36] PETROV, V. V. (1964). Local limit theorems for sums of independent random variables. *Teor. Veroyatn. Primen.* **9** 343–352. [MR0163341](#)
- [37] PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **82**. Springer, New York. [MR0388499](#)
- [38] PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. Holden-Day, Inc., San Francisco, CA. Translated and edited by Amiel Feinstein. [MR0213190](#)
- [39] PROKHOROV, YU. V. (1952). A local theorem for densities. *Dokl. Akad. Nauk SSSR (N.S.)* **83** 797–800. [MR0049501](#)
- [40] RÉNYI, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. 1* 547–561. Univ. California Press, Berkeley, CA. [MR0132570](#)
- [41] SHIRYAEV, A. N. (1996). *Probability*, 2nd ed. *Graduate Texts in Mathematics* **95**. Springer, New York. Translated from the first (1980) Russian edition by R. P. Boas. [MR1368405](#)
- [42] SIRAŽDINOV, S. H. and MAMATOV, M. (1962). On mean convergence for densities. *Theory Probab. Appl.* **7** 424–428. [MR0143237](#)
- [43] SZEGÖ, G. *Orthogonal Polynomials*, 3rd ed. *American Math. Soc. Publications* **23**. Amer. Math. Soc., Providence, RI.
- [44] TOSCANI, G. (2016). Entropy inequalities for stable densities and strengthened central limit theorems. *J. Stat. Phys.* **165** 371–389. [MR3554386](#)
- [45] TOSCANI, G. (2016). The fractional Fisher information and the central limit theorem for stable laws. *Ric. Mat.* **65** 71–91. [MR3513902](#)

- [46] TSALLIS, C. (1998). Generalized entropy-biased criterion for consistent testing. *Phys. Rev. E* **58** 1442–1445.
- [47] TULINO, A. M. and VERDÚ, S. (2006). Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof. *IEEE Trans. Inform. Theory* **52** 4295–4297. [MR2298559](#)
- [48] VAJDA, I. (1989). *Theory of Statistical Inference and Information*. Kluwer Academic, Dordrecht.
- [49] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inform. Theory* **60** 3797–3820. [MR3225930](#)

S. G. BOBKOV
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
VINCENT HALL 228
206 CHURCH ST. S.E.
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: bobkov@math.umn.edu

G. P. CHISTYAKOV
F. GÖTZE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BIELEFELD
POSTBOX 100131
33501 BIELEFELD
GERMANY
E-MAIL: chistyak@math.uni-bielefeld.de
goetze@math.uni-bielefeld.de