The Annals of Probability
2025, Vol. 53, No. 2, 453–477
https://doi.org/10.1214/24-AOP1711

CENTRAL LIMIT THEOREM FOR RÉNYI DIVERGENCE OF INFINITE ORDER

BY SERGEY G. BOBKOV[1,a] AND FRIEDRICH GÖTZE[2,b]

[1]*School of Mathematics, University of Minnesota,* [a]*bobkov@math.umn.edu*
[2]*Faculty of Mathematics, Bielefeld University,* [b]*goetze@math.uni-bielefeld.de*

For normalized sums Z_n of i.i.d. random variables, we explore necessary and sufficient conditions, which guarantee the normal approximation with respect to the Rényi divergence of infinite order. In terms of densities p_n of Z_n, this is a strengthened variant of the local limit theorem taking the form $\sup_x (p_n(x) - \varphi(x))/\varphi(x) \to 0$ as $n \to \infty$.

1. Introduction. Strict sub-Gaussianity. Let X be a random variable with density p. The Rényi divergence of order $\alpha > 0$ or the relative α-entropy of its distribution with respect to the standard normal law with density $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ is given by

$$D_\alpha(p\|\varphi) = \frac{1}{\alpha - 1} \log \int_{-\infty}^{\infty} \Big(\frac{p}{\varphi}\Big)^\alpha \varphi\, dx. \tag{1.1}$$

A closely related functional is the Tsallis distance

$$T_\alpha(p\|\varphi) = \frac{1}{\alpha - 1} \Big[\int_{-\infty}^{\infty} \Big(\frac{p}{\varphi}\Big)^\alpha \varphi\, dx - 1\Big]. \tag{1.2}$$

Since $T_\alpha = \frac{1}{\alpha-1}[e^{(\alpha-1)D_\alpha} - 1]$, both distances are of a similar order, when they are small. Hence, approximation problems in D_α and T_α are equivalent. Moreover, as the function $\alpha \to D_\alpha$ is nondecreasing, the convergence in D_α is getting stronger for growing indexes α.

Let us recall that, for the region $0 < \alpha < 1$, D_α is topologically equivalent to the total variation distance between the distribution of X and the standard normal law. For $\alpha = 1$, we obtain the Kullback–Leibler distance

$$D(p\|\varphi) = \int_{-\infty}^{\infty} p \log \frac{p}{\varphi}\, dx,$$

also called the informational divergence or the relative entropy. It is finite if and only if X has a finite second moment and finite Shannon's entropy. But the range $\alpha > 1$ leads to much stronger Rényi/Tsallis distances. For example, the finiteness of $D_\alpha(p\|\varphi)$ requires that X is sub-Gaussian, that is, the moments $\mathbb{E}\, e^{cX^2}$ should be finite for small $c > 0$. One important particular case $\alpha = 2$ in this hierarchy corresponds to the Pearson χ^2-distance $T_2 = \chi^2$. For various properties and applications of these distances, we refer an interested reader to [8, 15, 19, 21, 31, 32].

The study of the convergence in the central limit theorem (CLT) with respect to D_α and the associated problem of Berry–Esseen bounds have a long and rich history. Let us remind several results in this direction about the classical model of normalized sums

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

Received August 2023; revised May 2024.
MSC2020 subject classifications. 60E, 60F.
Key words and phrases. Central limit theorem, Rényi divergence.

of i.i.d. random variables $(X_k)_{k\geq 1}$. We will treat them as independent copies of a random variable X, assuming that it has mean zero and variance one.

The convergence $D_\alpha(p_n\|\varphi) \to 0$ as $n \to \infty$ holds true for $0 < \alpha < 1$, as long as Z_n have densities p_n for large n. This is due to the corresponding result by Prokhorov [29] about the total variation distance. The stronger property $D(p_n\|\varphi) \to 0$ in terms of relative entropy was studied by Barron [4] who showed that the condition $D(p_n\|\varphi) < \infty$ for some n is necessary and sufficient for the entropic CLT. The asymptotic behavior of such distances under higher-order moment assumptions, including Edgeworth-type expansions in powers of $1/n$, has been studied in [6]. It is worthwhile mentioning that this convergence is monotone with respect to n; cf. Artstein, Ball, Barthe and Naor [2] and Madiman and Barron [22]. See also [3] and [7] for various entropic bounds in the non-i.i.d. case.

The range $\alpha > 1$ was treated in detail in [8]. It was shown there that $D_\alpha(p_n\|\varphi) \to 0$ as $n \to \infty$, if and only if $D_\alpha(p_n\|\varphi)$ is finite for some n, and if X admits the following sub-Gaussian bound on the Laplace transform:

$$\mathbb{E}\, e^{tX} < e^{\alpha^* t^2/2}, \quad t \in \mathbb{R} \ (t \neq 0), \tag{1.3}$$

where $\alpha^* = \frac{\alpha}{\alpha-1}$. In that case, we have an equivalence $D_\alpha \sim T_\alpha \sim \frac{\alpha}{2}\chi^2$. These results have been extended to the multidimensional setting as well.

For indexes $\alpha \to \infty$ in (1.3), we arrive at the following characterization.

THEOREM 1.1. *Assume that $D_\alpha(p_n\|\varphi) < \infty$ for every $\alpha > 1$ with some $n = n_\alpha$. For the convergence $D_\alpha(p_n\|\varphi) \to 0$ for all α, it is necessary and sufficient that $\mathbb{E}\exp\{tX\} \leq \exp\{t^2/2\}$ for all $t \in \mathbb{R}$.*

The last inequality describes an interesting class of probability distributions, which appear naturally in many mathematical problems. More generally, one says that a random variable X with mean zero is strictly sub-Gaussian, or its distribution is strictly sub-Gaussian (regardless of whether or not it has a density), if the inequality

$$\mathbb{E}\, e^{tX} \leq e^{\sigma^2 t^2/2}, \quad t \in \mathbb{R}, \tag{1.4}$$

holds with constant $\sigma^2 = \mathrm{Var}(X)$, which is then best possible. Note that, when saying that X is sub-Gaussian (with mean zero), one means that (1.4) holds with some σ^2.

This class was apparently first introduced in an explicit form by Buldygin and Kozachenko in [12] under the name "strongly sub-Gaussian" and then analyzed in more details in their book [13]. Recent investigations include the work by Arbel, Marchal and Nguyen [1] providing some examples and properties and by Guionnet and Husson [17]. In the latter paper, (1.4) appears as a condition for the validity of large deviation principles for the largest eigenvalue of Wigner matrices with the same rate function as in the case of Gaussian entries.

A simple sufficient condition for the strict sub-Gaussianity was given by Newman in terms of location of zeros of the characteristic function $f(z) = \mathbb{E}\, e^{izX}$, $z \in \mathbb{C}$ (which is extended, by the sub-Gaussian property, from the real line to the complex plane as an entire function of order at most 2). As was stated in [23], X is strictly sub-Gaussian, as long as $f(z)$ has only real zeros in $\mathbb{C}$ (a detailed proof was later given in [13]). Such probability distributions form an important class denoted by $\mathfrak{L}$, introduced and studied by Newman in the mid 1970s in connection with the Lee–Yang property, which naturally arises in the context of ferromagnetic Ising models; cf. [23–26]. This class is rather rich; it is closed under infinite convergent convolutions and under weak limits. For example, it includes Bernoulli convolutions, and hence convolutions of uniform distributions on bounded symmetric intervals.

Some classes of strictly sub-Gaussian distributions outside $\mathfrak{L}$ have been recently discussed in [10]. It was shown that (1.4) continues to hold for symmetric distributions under the weaker requirement that all zeros of $f(z)$ with $\mathrm{Re}(z) > 0$ lie in the cone $|\mathrm{Arg}(z)| \le \frac{\pi}{8}$ (which is sharp when f has only one zero in the positive octant). In that case, if X is not normal, the inequality (1.4) may be sharpened as follows: For any $t_0 > 0$, there is $c = c(t_0)$, $0 < c < \sigma^2 = \mathrm{Var}(X)$, such that

$$\mathbb{E}\, e^{tX} \le e^{ct^2/2}, \quad |t| \ge t_0. \tag{1.5}$$

In general, this separation-type property is however not necessary for the strict sub-Gaussianity. It turns out ([10]) that there exists a large class of strictly sub-Gaussian distributions with mean zero and variance one, for which the Laplace transform has the form

$$\mathbb{E}\, e^{tX} = \Psi(t)\, e^{t^2/2}, \quad t \in \mathbb{R},$$

where $\Psi(t)$ is a *periodic* function with some period $h > 0$ and such that $\Psi(t) \le 1$ for all $t \in \mathbb{R}$. Hence, $\Psi(kh) = 1$ for all $k \in \mathbb{Z}$, so that (1.4) becomes an equality for infinitely many points t.

2. Main results for the convergence in D_∞. Thus, the strict sub-Gaussianity appears as a necessary condition for the convergence in all D_α and, therefore, in D_∞, which according to (1.1) is given by the limit

$$D_\infty(p\|\varphi) = \lim_{\alpha\to\infty} D_\alpha(p\|\varphi) = \operatorname{ess\,sup}_x \log\bigl(p(x)/\varphi(x)\bigr).$$

Although the Tsallis distance of infinite order may not be defined similarly as a limit of (1.2), we make the convention that

$$T_\infty(p\|\varphi) = \operatorname{ess\,sup}_x \frac{p(x) - \varphi(x)}{\varphi(x)}.$$

Then $T_\infty = e^{D_\infty} - 1$ like for the Tsallis distance of finite order, so that convergence in D_∞ and T_∞ are equivalent. In particular, in the setting of the normalized sums Z_n, the CLT $D_\infty(p_n\|\varphi) \to 0$ is equivalent to the assertion that Z_n have densities p_n such that

$$\sup_x \frac{p_n(x) - \varphi(x)}{\varphi(x)} \to 0 \quad \text{as } n \to \infty. \tag{2.1}$$

The purpose of this paper is to give necessary and sufficient conditions for this variant of the CLT in terms of the Laplace transform $L(t) = \mathbb{E}\, e^{tX}$. Consider the log-Laplace transform $K(t) = \log L(t)$ (which is a convex, smooth function) and the associated function

$$A(t) = \frac{1}{2} t^2 - K(t), \quad t \in \mathbb{R}.$$

As before, suppose that $(X_k)_{k\ge1}$ are independent copies of the random variable X with mean $\mathbb{E}X = 0$ and variance $\mathrm{Var}(X) = 1$. We assume that:

(1) Z_n has density p_n with $T_\infty(p_n\|\varphi) < \infty$ for some $n = n_0$;
(2) X is strictly sub-Gaussian, that is, $A(t) \ge 0$ for all $t \in \mathbb{R}$.

THEOREM 2.1. *For the convergence $T_\infty(p_n\|\varphi) \to 0$, it is necessary and sufficient that the following two conditions are fulfilled*:

(a) $A''(t) = 0$ *for every point* $t \in \mathbb{R}$ *such that* $A(t) = 0$;
(b) $\limsup_{k\to\infty} A''(t_k) \le 0$ *for every sequence* $t_k \to \pm\infty$ *such that* $A(t_k) \to 0$ *as* $k \to \infty$.

The conditions $a) - b)$ may be combined as $\lim_{A(t)\to 0} \max(A''(t), 0) = 0$, which is kind of continuity of A'' with respect to A.

Note added in proof: This combined condition can be strengthened to $\lim_{A(t)\to 0} A''(t) = 0$. It will be proved when extending Theorem 2.1 to the multivariate case in a forthcoming paper [11].

Under the separation property (1.5), the condition $b)$ is fulfilled automatically, while the equation $A(t) = 0$ has only one solution $t = 0$. But near zero, due to the strict sub-Gaussianity, $A(t) = O(t^4)$ and $A''(t) = O(t^2)$. Hence, the condition $a)$ is fulfilled as well, and we obtain the CLT with respect to D_∞. In particular, it is applicable to the class $\mathfrak{L}$ of Newman described above. In fact, for this conclusion, (1.5) may further be weakened to

$$\sup_{|t|\geq t_0} \big[e^{-t^2/2}\, \mathbb{E}\, e^{tX}\big] < 1 \quad \text{for all } t_0 > 0. \tag{2.2}$$

In this case, one can additionally explore the rate of convergence.

THEOREM 2.2. *Let X be a nonnormal random variable with* $\mathrm{Var}(X) = 1$ *satisfying* (2.2). *If $T_\infty(p_n \| \varphi) < \infty$ for some n, then*

$$T_\infty(p_n \| \varphi) = O\bigg(\frac{1}{n} (\log n)^3\bigg) \quad \textit{as } n \to \infty. \tag{2.3}$$

Furthermore, specializing Theorem 2.1 to the case where the Laplace transform contains a periodic component, we have the following.

THEOREM 2.3. *Suppose that the function $\Psi(t) = L(t)\, e^{-t^2/2}$ is h-periodic for a smallest value $h > 0$. For the convergence $T_\infty(p_n \| \varphi) \to 0$ as $n \to \infty$, it is necessary and sufficient that, for every* $0 < t < h$,

$$\Psi(t) = 1 \ \Rightarrow\ \Psi''(t) = 0. \tag{2.4}$$

Moreover, if the equation $\Psi(t) = 1$ *has no solution in* $0 < t < h$, *then the relation* (2.3) *about the rate of convergence continues to hold.*

For an illustration (cf. Section 9 for more details), consider random variables X with $\Psi(t) = 1 - c \sin^4 t$, where the parameter $c > 0$ is small enough. In this case, $\Psi(t)$ is π-periodic and all conditions in Theorem 2.3 are fulfilled. Hence, the CLT for T_∞ does hold with rate as in (2.3). On the other hand, in a similar π-periodic example

$$\Psi(t) = 1 - c\big(1 - 4\sin^2 t\big)^2 \sin^4 t,$$

the condition (2.4) is violated at the point $t = \pi/6$, so there is no CLT. Thus, the continuity condition of A'' with respect A in Theorem 2.1 may or may not be fulfilled in general in the class of strictly sub-Gaussian distributions.

Returning to the convergence property (2.1), it should be emphasized that it is not possible to put the absolute value sign in the numerator (this will be clarified in Section 4). The situation is of course different, when one considers the supremum over bounded increasing intervals. For example, under suitable moment assumptions (cf. [27, 28]), it follows from Edgeworth expansions for densities that

$$\sup_{|x|\leq c\sqrt{\log n}} \frac{|p_n(x) - \varphi(x)|}{\varphi(x)} \to 0 \quad \text{as } n \to \infty.$$

The proof of Theorem 2.1 is given in Section 8, with preliminary developments in Sections 3–7. Its application to the periodic case is discussed in Section 9. What is unusual in our approach is that the proof does not use in essence the tools from Complex Analysis (as one ingredient, we establish a uniform local limit theorem for bounded densities with a quantitative error term). However, in the study of rates of convergence with respect to T_∞, we employ an old result by Richter [30] in a certain refined form on the asymptotic behavior of ratios $p_n(x)/\varphi(x)$. This result is discussed in Section 10, where we also include the proof of Theorems 2.2–2.3 (for the rate of convergence). In the last section, we describe several examples of probability distributions satisfying the condition 1), needed for applicability of Theorems 2.1–2.2.

3. Semigroup of shifted distributions (Esscher transform). Let X be a sub-Gaussian random variable with density p. Here and in the sequel, the sub-Gaussianity is understood as the property that $\mathbb{E}\, e^{cX^2} < \infty$ for some $c > 0$ (which is equivalent to (1.4) with some σ^2 when X has mean zero).

Then the Laplace transform, or the moment generating function

$$(Lp)(t) = L(t) = \mathbb{E}\, e^{tX} = \int_{-\infty}^{\infty} e^{tx}\, p(x)\, dx$$

is finite for all complex numbers t and represents an entire function in the complex plane. Hence, the log-Laplace transform

$$(Kp)(t) = K(t) = \log L(t) = \log \mathbb{E}\, e^{tX}, \quad t \in \mathbb{R},$$

represents a convex, C^∞-smooth function on the real line.

DEFINITION 3.1. Introduce the family of probability densities

$$Q_h p(x) = \frac{1}{L(h)}\, e^{hx}\, p(x), \quad x \in \mathbb{R}, \tag{3.1}$$

with parameter $h \in \mathbb{R}$. We call the distribution with this density the shifted distribution of X at step h.

The early history of this well-known and popular transform goes back to 1930s. In actuarial science, following Esscher [16], the density $Q_h p$ is commonly called the Esscher transform of p. Other names "conjugate distribution laws," "the family of distribution laws conjugate to a system" were used by Khinchin [20] in the framework of statistical mechanics. See also Daniels [14] who applied this transform to develop asymptotic expansions for densities. In this paper, we prefer to use a different terminology as in Definition 3.1 in order to emphasize the following important fact: For the standard normal density $\varphi(x)$, the shifted normal law has density $Q_h\varphi(x) = \varphi(x+h)$.

A remarkable property of the transform (2.1) is the semigroup property

$$Q_{h_1}(Q_{h_2} p) = Q_{h_1+h_2} p, \quad h_1, h_2 \in \mathbb{R}.$$

Let us also mention how this transform acts under rescaling. Given $\lambda > 0$, the random variable λX has density $p_\lambda(x) = \frac{1}{\lambda} p(\frac{x}{\lambda})$ with Laplace transform $(Lp_\lambda)(t) = L(\lambda t)$. Hence,

$$Q_h p_\lambda(x) = \frac{1}{(Lp_\lambda)(h)}\, e^{hx}\, p_\lambda(x) = \frac{1}{\lambda}\, (Q_{\lambda h} p)\Big(\frac{x}{\lambda}\Big).$$

This identity implies that the maximum-of-density functional $M(X) = M(p) = \operatorname{ess\,sup}_x p(x)$ satisfies

$$M(Q_h p_\lambda) = \frac{1}{\lambda} M(Q_{\lambda h} p). \tag{3.2}$$

The transform Q_h is also multiplicative with respect to convolutions.

PROPOSITION 3.2. *If independent sub-Gaussian random variables have densities* $p_1, \dots, p_n$, *then for the convolution* $p = p_1 * \cdots * p_n$, *we have*

$$Q_h p = Q_h p_1 * \cdots * Q_h p_n. \tag{3.3}$$

PROOF. It is sufficient to compare the Laplace transforms of both sides in (3.3). The Laplace transform of p is given by $Lp(t) = (Lp_1)(t) \dots (Lp_n)(t)$. Hence, the Laplace transform of $Q_h p$ is given by

$$\begin{aligned}(LQ_h p)(t) &= \int_{-\infty}^{\infty} e^{tx} Q_h p(x)\,dx = \frac{1}{(Lp)(t)} \int_{-\infty}^{\infty} e^{(t+h)x} p(x)\,dx \\ &= \frac{(Lp)(t+h)}{(Lp)(t)} = \prod_{k=1}^{n} \frac{(Lp_k)(t+h)}{(Lp_k)(t)} = \prod_{k=1}^{n} (LQ_h p_k)(t).\end{aligned}$$

□

The formula (3.1) in Definition 3.1 may be written equivalently as

$$p(x) = L(h)\, e^{-xh} Q_h p(x) = e^{-xh+K(h)} Q_h p(x),$$

or

$$\frac{p(x)}{\varphi(x)} = \sqrt{2\pi}\, e^{\frac{1}{2}(x-h)^2 - \frac{1}{2}h^2 + K(h)} Q_h p(x).$$

Introduce the function

$$(Ap)(h) = A(h) = \frac{1}{2}h^2 - K(h), \tag{3.4}$$

which allows to reformulate the strict sub-Gaussianity via the inequality $A(h) \geq 0$ for all h (under the assumptions $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$). Thus,

$$\frac{p(x)}{\varphi(x)} = \sqrt{2\pi}\, e^{\frac{1}{2}(x-h)^2 - A(h)} Q_h p(x). \tag{3.5}$$

We will use this representation to bound the ratio on the left-hand side for the densities p_n of the normalized sums

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}} \tag{3.6}$$

of independent copies of the random variable X with density p. In order to apply (3.5) to p_n instead of p, put $x_n = x\sqrt{n}$, $h_n = h\sqrt{n}$. Note that in terms of $L = Lp$, $K = Kp$ and $A = Ap$, we may write

$$\begin{aligned}(Lp_n)(t) &= L(t/\sqrt{n})^n = e^{nK(t/\sqrt{n})}, \\ (Kp_n)(t) &= nK(t/\sqrt{n}), \\ (Ap_n)(h_n) &= \frac{1}{2}h_n^2 - (Kp_n)(h_n) = \frac{n}{2}h^2 - nK(h) = nA(h).\end{aligned}$$

Therefore, the ratio (3.5) being applied with (x_n, h_n) becomes the following.

PROPOSITION 3.3. *Putting* $x_n = x\sqrt{n}$, $h_n = h\sqrt{n}$ $(x, h \in \mathbb{R})$, *we have*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \sqrt{2\pi}\, e^{\frac{n}{2}(x-h)^2 - nA(h)} Q_{h_n} p_n(x_n). \tag{3.7}$$

This equality is useful, if we are able to bound the factor $Q_{h_n} p_n(x_n)$ uniformly over all x for a fixed value of h as stated in the following corollary.

COROLLARY 3.4. *For all* $x, h \in \mathbb{R}$,

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le \sqrt{2\pi}\, e^{\frac{n}{2}(x-h)^2 - nA(h)} M(Q_{h\sqrt{n}} p_n). \tag{3.8}$$

REMARK 3.5. Since the function K is convex, it follows from the definition (3.4) that $A''(h) \le 1$ for all $h \in \mathbb{R}$. As a consequence, this function satisfies a differential inequality

$$A'(h)^2 \le 2A(h), \quad h \in \mathbb{R}, \tag{3.9}$$

if $A(h) \ge 0$ for all $h \in \mathbb{R}$. For a short proof (proposed by the referee), one may apply the Taylor formula

$$\begin{aligned} 0 \le A(h+x) &= A(h) + A'(h)x + \frac{1}{2}A''(h_1)x^2 \\ &\le A(h) + A'(h)x + \frac{1}{2}x^2, \quad x \in \mathbb{R}, \end{aligned}$$

holding for some point h_1 in the segment with endpoints h and $h + x$. Minimizing the right-hand side over all x leads to (3.9).

4. Maximum of shifted densities. In order to bound the last term in (3.8), suppose that the distribution of X has a finite Rényi distance of infinite order to the standard normal law. This means that the density of X admits a pointwise upper bound

$$p(x) \le c\varphi(x), \quad x \in \mathbb{R} \text{ (a.e.)} \tag{4.1}$$

for some constant c. Note that its optimal value is $c = 1 + T_\infty(p\|\varphi)$. In that case, one may control the maximum of densities of shifted distributions

$$M(Q_h p) = \operatorname{ess\,sup}_x Q_h p(x).$$

Indeed, (4.1) implies that, for any $x \in \mathbb{R}$,

$$Q_h p(x) = \frac{1}{L(h)} e^{xh} p(x) \le \frac{ce^{xh - x^2/2}}{L(h)\sqrt{2\pi}} \le \frac{ce^{h^2/2}}{L(h)\sqrt{2\pi}} = \frac{c}{\sqrt{2\pi}} e^{A(h)},$$

where $L = Lp$ and $A = Ap$. Thus,

$$M(Q_h p) \le \frac{c}{\sqrt{2\pi}} e^{A(h)}. \tag{4.2}$$

However, it is useless to apply this bound directly to the densities p_n of the normalized sums Z_n as in (3.6), since then the right-hand side of (4.2) will contain the parameter $c_n = 1 + T_\infty(p_n\|\varphi)$. Instead, we use a semiadditive property of the maximum-of-density functional, which indicates that

$$M(X_1 + \cdots + X_n)^{-2} \ge \frac{1}{2} \sum_{k=1}^n M(X_k)^{-2}$$

for all independent random variables X_k having bounded densities; cf. [5] or [9]. If all X_k are identically distributed and have density p, this relation yields

$$M(p^{*n}) \le \sqrt{2/n}\, M(p)$$

for the convolution nth power of p. Applying Proposition 3.2 together with (4.2), we then have

$$M(Q_h p^{*n}) \le \sqrt{2/n}\, M(Q_h p) \le \sqrt{2/n}\, \frac{c}{\sqrt{2\pi}} e^{A(h)}.$$

On the other hand, since $p^{*n}(x) = \frac{1}{\lambda} p_n(\frac{x}{\lambda})$ with $\lambda = \sqrt{n}$, one may apply the identity (3.2):

$$M\big(Q_h p^{*n}\big) = \frac{1}{\sqrt{n}}\, M(Q_{h\sqrt{n}}\, p_n).$$

Hence,

$$M(Q_{h\sqrt{n}}\, p_n) \le \frac{c}{\sqrt{\pi}}\, e^{A(h)}.$$

Let us now return to Corollary 3.4 and apply this bound to get that

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le c\sqrt{2}\, e^{\frac{n}{2}(x-h)^2-(n-1)A(h)},$$

recalling that $c = 1 + T_\infty(p\|\varphi)$. In particular, with $h = x$ this yields the following.

PROPOSITION 4.1. *Let p_n denote the density of Z_n constructed for n independent copies of a sub-Gaussian random variable X whose density p has finite Rényi distance of infinite order to the standard normal law. Then, for almost all $x \in \mathbb{R}$,*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le c\sqrt{2}\, e^{-(n-1)A(x)}. \tag{4.3}$$

COROLLARY 4.2. *If additionally $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$ and X is strictly sub-Gaussian, then*

$$T_\infty(p_n\|\varphi) \le \sqrt{2}\big(1 + T_\infty(p\|\varphi)\big) - 1.$$

Thus, the finiteness of the Tsallis distance $T_\infty(p\|\varphi)$ for a strictly sub-Gaussian random variable X with density p ensures the boundedness of $T_\infty(p_n\|\varphi)$ for all normalized sums Z_n.

If $A(x)$ is bounded away from zero, the inequality (4.3) shows that $p_n(x\sqrt{n})/\varphi(x\sqrt{n})$ is exponentially small for growing n. In particular, this holds for any nonnormal random variable X satisfying the separation property (2.2). Then we immediately obtain the following.

COROLLARY 4.3. *Suppose that X has a density p with finite $T_\infty(p\|\varphi)$. Under the condition (2.2), for any $\tau_0 > 0$, there exist $A > 0$ and $\delta \in (0,1)$ such that the densities p_n of Z_n satisfy*

$$p_n(x) \le A\delta^n \varphi(x), \quad |x| \ge \tau_0\sqrt{n}. \tag{4.4}$$

In particular,

$$\liminf_{n\to\infty} \sup_{x\in\mathbb{R}} \frac{|p_n(x) - \varphi(x)|}{\varphi(x)} \ge 1.$$

Therefore, one cannot hope to strengthen the Tsallis distance by introducing a modulus sign in the definition of the distance.

Since (2.2) does not need be true in general, Proposition 4.1 will be applied outside the set of points where $A(x)$ is bounded away from zero. More precisely, for a parameter $a > 0$ and $n \ge 2$, define the critical zone

$$A_n(a) = \big\{h > 0 : A(h) \le a/(n-1)\big\}. \tag{4.5}$$

From (4.3), it follows that

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le c\sqrt{2}\, e^{-a}, \quad x \notin A_n(a). \tag{4.6}$$

If a is large, this bound may be used in the proof of the CLT with respect to the distance T_∞ restricted to the complement of the critical zone. As for this zone, the bound (4.3) is not appropriate, and we need to return to the basic representation from Proposition 3.3. To study the last term $Q_{h_n} p_n(x_n)$ in (3.7) for $x \in A_n(a)$, one may apply a variant of the local limit theorem, using the property that the density $Q_{h_n} p_n$ has a convolution structure. However, in order to justify this application, we should first explore the behavior of moments of densities participating in the convolution.

5. Moments of shifted distributions. For a sub-Gaussian random variable X with density p, denote by $X(h)$ a random variable with density $Q_h p$ ($h \in \mathbb{R}$). It is sub-Gaussian, and its Laplace and log-Laplace transforms are given by

$$\begin{aligned} L_h(t) &\equiv \mathbb{E}e^{tX(h)} = \frac{L(t+h)}{L(h)}, \\ K_h(t) &\equiv \log L_h(t) = K(t+h) - K(h). \end{aligned} \tag{5.1}$$

Furthermore, it has mean and variance

$$m_h \equiv \mathbb{E}X(h) = \frac{L'(h)}{L(h)} = K'(h),$$

$$\sigma_h^2 \equiv \operatorname{Var}(X(h)) = \frac{L''(h)L(h) - L'(h)^2}{L(h)^2} = K''(h).$$

The last equality shows that necessarily $K''(h) > 0$ for all $h \in \mathbb{R}$. Indeed, otherwise the random variable $X(h)$ would be a constant a.s.

The question of how to bound the standard deviation σ_h from below relies upon certain fine properties of the density p and the behavior of the function $A(h) = \frac{1}{2}h^2 - K(h)$, introduced in (3.4). As before, suppose that the distribution of X has finite Rényi distance of infinite order to the standard normal law, so that

$$p(x) \le c\varphi(x), \quad x \in \mathbb{R}, \tag{5.2}$$

with $c = 1 + T_\infty(p\|\varphi)$. Then one may control the maximum $M(X(h)) = \operatorname{ess\,sup}_x p_h(x)$ of densities of shifted distributions, using (4.2):

$$Q_h p(x) \le \frac{c}{\sqrt{2\pi}} e^{A(h)}.$$

For a lower bound, we employ a well-known general relation

$$M(\xi)^2 \operatorname{Var}(\xi) \ge \frac{1}{12}$$

(where the equality is attained for the uniform distribution on a bounded interval). Let us provide the following simple argument, assuming without loss of generality that a random variable ξ has finite variance and a density with $M(\xi) = 1$. Then the function

$$H(x) = \mathbb{P}\{|\xi - \mathbb{E}\xi| \ge x\}$$

is absolutely continuous, and its Radon–Nikodym derivative satisfies $H'(x) \ge -2$ a.e. in $x > 0$. Since $H(0) = 1$, we get $H(x) \ge 1 - 2x$ for all $x \ge 0$ and, therefore,

$$\operatorname{Var}(\xi) = 2\int_0^\infty xH(x)\,dx \ge 2\int_0^{1/2} x(1-2x)\,dx = \frac{1}{12}.$$

Applying this to $\xi = X(h)$ and combining the two bounds, we obtain that

$$\frac{1}{\sqrt{12}} \le M(X(h))\sigma_h \le \frac{c\sigma_h}{\sqrt{2\pi}} e^{A(h)}.$$

Thus we arrive at the following.

LEMMA 5.1. *Under the condition* (5.2), *for all* $h \in \mathbb{R}$,

$$\sigma_h \geq \sqrt{\frac{\pi}{6c^2}}\, e^{-A(h)}. \tag{5.3}$$

Since $\sigma_h > 0$, one may consider the normalized random variables

$$\widehat{X}(h) = \frac{X(h) - \mathbb{E}X(h)}{\sqrt{\mathrm{Var}(X(h))}} = \frac{X(h) - m_h}{\sigma_h}. \tag{5.4}$$

By (5.1), they have the moment generating function

$$\begin{aligned} \mathbb{E}\, e^{t\widehat{X}(h)} &= \mathbb{E}\exp\left\{\frac{t}{\sigma_h}\big(X(h) - m_h\big)\right\} \\ &= \exp\left\{-\frac{t}{\sigma_h}K'(h)\right\}\frac{L(h + \frac{t}{\sigma_h})}{L(h)} \end{aligned}$$

and the log-Laplace transform

$$\widehat{K}_h(t) = K\left(h + \frac{t}{\sigma_h}\right) - K(h) - \frac{t}{\sigma_h}K'(h). \tag{5.5}$$

In order to estimate (5.5) from above, assume that $K(h) \leq \frac{1}{2}h^2$, that is, $A(h) \geq 0$ for all h. For $h \in A_n(a)$, the definition (4.5) implies that

$$K(h) \geq \frac{1}{2}h^2 - \frac{a}{n-1},$$

and hence

$$\begin{aligned} \widehat{K}_h(t) &\leq \frac{1}{2}\left(h + \frac{t}{\sigma_h}\right)^2 - \frac{1}{2}h^2 + \frac{a}{n-1} - \frac{t}{\sigma_h}K'(h) \\ &= \frac{1}{2}\left(\frac{t}{\sigma_h}\right)^2 + \frac{a}{n-1} + \frac{t}{\sigma_h}\big(h - K'(h)\big). \end{aligned} \tag{5.6}$$

Here, the term $h - K'(h) = A'(h)$ can be estimated by virtue of the inequality (3.9), which gives

$$\big|h - K'(h)\big|^2 \leq 2A(h) \leq \frac{2a}{n-1}$$

and

$$\begin{aligned} \frac{|t|}{\sigma_h}\big|h - K'(h)\big| &\leq \frac{1}{2}\left(\frac{t}{\sigma_h}\right)^2 + \frac{1}{2}\big|h - K'(h)\big|^2 \\ &\leq \frac{1}{2}\left(\frac{t}{\sigma_h}\right)^2 + \frac{a}{n-1}, \end{aligned}$$

where we used $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$ $(a, b \in \mathbb{R})$. It follows from (5.6) that

$$\widehat{K}_h(t) \leq \left(\frac{t}{\sigma_h}\right)^2 + \frac{2a}{n-1}.$$

Here, the right-hand side is bounded for sufficiently small $|t|$ and sufficiently large n. One may require, for example, that $n \geq 4a + 1$ and $|t| \leq \frac{1}{\sqrt{2}}\sigma_h$, in which case $\widehat{K}_h(t) \leq 1$, so that

$$\mathbb{E}\, e^{|t|\widehat{X}(h)} \leq \mathbb{E}\, e^{t\widehat{X}(h)} + \mathbb{E}\, e^{-t\widehat{X}(h)} \leq 2e.$$

Using $x^3 e^{-|t|x} \leq (\frac{3}{e})^3\, |t|^{-3}$ $(x \geq 0)$, this gives $\mathbb{E}\,|\widehat{X}(h)|^3 \leq 2e(\frac{3}{e})^3\, |t|^{-3}$. One can summarize in the following statement.

LEMMA 5.2. *If the Laplace transform of a sub-Gaussian random variable X is such that $A(h) \geq 0$ for all $h \in \mathbb{R}$, then for all $h \in A_n(a)$ with $n \geq 4a+1$, we have*

$$\mathbb{E}\, e^{\sigma_h |\widehat{X}(h)|/\sqrt{2}} < 2e.$$

As a consequence,

$$\mathbb{E}\,|\widehat{X}(h)|^3 \leq C\sigma_h^{-3}$$

up to some absolute constant $C > 0$.

6. Local limit theorem for bounded densities. Before we can apply the representation (3.7), in the next step we need to establish a uniform local limit theorem with a quantitative error term. Let $(X_k)_{k\geq 1}$ be independent copies of a random variable X with $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$, $\beta_3 = \mathbb{E}|X|^3 < \infty$, which has a bounded density. Then the normalized sums Z_n have bounded continuous densities p_n for all $n \geq 2$ satisfying

$$\sup_x |p_n(x) - \varphi(x)| = O\Big(\frac{1}{\sqrt{n}}\Big) \quad (n \to \infty).$$

See, for example, [27, 28]. Let us quantify the error O-term in terms of β_3 and the maximum of density $M = M(X)$.

LEMMA 6.1. *With some positive absolute constant C, we have*

$$\sup_x |p_n(x) - \varphi(x)| \leq C\,\frac{M^2\beta_3}{\sqrt{n}}. \tag{6.1}$$

PROOF. Since $M \geq 1/\sqrt{12}$ and $\beta_3 \geq 1$, while $M(p_n) \leq \sqrt{2}M$ for all n (cf. Section 4), we may assume that $n \geq 4$.

Denote by $f(t)$ the characteristic function of X. By the boundedness assumption, the characteristic functions

$$f_n(t) = \mathbb{E}\, e^{itZ_n} = f(t/\sqrt{n})^n, \quad t \in \mathbb{R},$$

are integrable for all $n \geq 2$. Indeed, by the Plancherel theorem,

$$\int_{-\infty}^{\infty} |f(t)|^n\, dt \leq \int_{-\infty}^{\infty} |f(t)|^2\, dt = 2\pi \int_{-\infty}^{\infty} p(x)^2\, dx \leq 2\pi M.$$

Hence, one may apply the Fourier inversion formula to represent the densities of Z_n as

$$p_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f_n(t)\, dt, \quad x \in \mathbb{R}.$$

Using a similar representation for the normal density, we get

$$|p_n(x) - \varphi(x)| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} |f_n(t) - e^{t^2/2}|\, dt.$$

As is well known (cf., e.g., [28], p. 109),

$$|f_n(t) - e^{t^2/2}| \leq 16\,\frac{\beta_3}{\sqrt{n}}|t|^3 e^{-t^2/3}, \quad |t| \leq \frac{\sqrt{n}}{4\beta_3},$$

which yields

$$\int_{|t|\leq \frac{\sqrt{n}}{4\beta_3}} |f_n(t) - e^{t^2/2}|\, dt \leq \frac{C\beta_3}{\sqrt{n}}$$

with some absolute constant C. As for large values of $|t|$, it was shown in [9], page 145, that for any $\varepsilon \in (0, 1]$ and $n \geq 4$,

$$\int_{|t|\geq\varepsilon} |f(t)|^n\,dt \leq \frac{4\pi M}{\sqrt{2n}} \exp\{-n\varepsilon^2/(5200M^2)\}.$$

This gives

$$\int_{|t|\geq\frac{\sqrt{n}}{4\beta_3}} |f_n(t)|\,dt = \sqrt{n}\int_{|t|\geq\frac{1}{4\beta_3}} |f(t)|^n\,dt \leq \frac{4\pi M}{\sqrt{2}} \exp\{-c_0 n/(\beta_3^2 M^2)\}.$$

Since M is bounded away from zero, a similar estimate holds true for the normal characteristic function as well. As a result, we arrive at

$$|p_n(x) - \varphi(x)| \leq C_0\left(\frac{\beta_3}{\sqrt{n}} + M\exp\{-c_0 n/(\beta_3^2 M^2)\}\right)$$

with some positive absolute constants C_0 and c_0, Using $e^{-x} < x^{-1/2}$ ($x > 0$), the second term in the brackets is dominated by the first one up to the multiple of M^2. Hence, the above estimate may be simplified to (6.1). □

7. Local limit theorem for shifted densities. An application of Lemma 6.1 to the normalized sums of independent copies of random variables $\widehat{X}(h)$ defined in (5.4) leads to the following refinement of the representation (3.7) from Proposition 3.3, when the point x belongs to the critical zone $A(x) \leq \frac{a}{n-1}$. Define

$$v_x = \frac{x - m_x}{\sigma_x} = \frac{x - K'(x)}{\sigma_x} = \frac{A'(x)}{\sigma_x},$$

where we recall that $m_x = K'(x)$ and $\sigma_x^2 = K''(x)$.

LEMMA 7.1. *If the Laplace transform of a sub-Gaussian random variable X with finite constant $c = 1 + T_\infty(p\|\varphi)$ is such that $A(h) \geq 0$ for all $h \in \mathbb{R}$, then for all $x \in A_n(a)$ with $n \geq 4(a+1)$, we have*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \frac{1}{\sigma_x} e^{-nA(x) - nv_x^2/2} + \frac{Bc^4}{\sqrt{n}}, \tag{7.1}$$

where $B = B_n(x)$ is bounded by an absolute constant.

PROOF. Let us return to the term $Q_{h_n} p_n$ in (3.7) with $h_n = h\sqrt{n}$. By Proposition 3.2, this density has a convolution structure. Recall that, for any random variable X with density $p = p_X$,

$$Q_h p_{\lambda X}(x) = \frac{1}{\lambda}(Q_{\lambda h} p)\left(\frac{x}{\lambda}\right).$$

Using this notation, $p_n = p_{S_n/\sqrt{n}}$ in terms of the sum $S_n = X_1 + \cdots + X_n$. Hence, with $\lambda = 1/\sqrt{n}$,

$$Q_{h_n} p_n(x) = \sqrt{n}\,(Q_h p_{S_n})(x\sqrt{n}) = \sqrt{n}\,(Q_h p) * \cdots * (Q_h p)(x\sqrt{n}),$$

where we applied Proposition 3.2 in the last step. By the definition, $Q_h p$ is the density of the random variable $X(h)$. Hence, $Q_{h_n} p_n(x)$ represents the density for the normalized sum

$$Z_{n,h} \equiv (X_1(h) + \cdots + X_n(h))/\sqrt{n},$$

assuming that $X_k(h)$ are independent. Introduce the normalized sums

$$\widehat{Z}_{n,h} \equiv \big(\widehat{X}_1(h) + \cdots + \widehat{X}_n(h)\big)/\sqrt{n} \tag{7.2}$$

for the shifted distributions (5.4), that is, with $X_k(h) = m_h + \sigma_h \widehat{X}_k(h)$. Thus,

$$Z_{n,h} = m_h\sqrt{n} + \sigma_h \widehat{Z}_{n,h}.$$

Denote by $\widehat{p}_{n,h}$ the density of $\widehat{Z}_{n,h}$. Then the density of $Z_{n,h}$ is given by

$$p_{n,h}(x) = \frac{1}{\sigma_h}\widehat{p}_{n,h}\bigg(\frac{x - m_h\sqrt{n}}{\sigma_h}\bigg), \quad x \in \mathbb{R}.$$

At the points $x_n = x\sqrt{n}$ as in (3.7), we therefore obtain that

$$Q_{h_n} p_n(x_n) = p_{n,h}(x_n) = \frac{1}{\sigma_h}\widehat{p}_{n,h}\bigg(\frac{x - m_h}{\sigma_h}\sqrt{n}\bigg).$$

Consequently, the equality (3.7) may be equivalently stated as

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \sqrt{2\pi}\, e^{\frac{n}{2}(x-h)^2 - nA(h)}\frac{1}{\sigma_h}\widehat{p}_{n,h}\bigg(\frac{x - m_h}{\sigma_h}\sqrt{n}\bigg).$$

In particular, for $h = x$, we get

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \sqrt{2\pi} e^{-nA(x)}\frac{1}{\sigma_x}\widehat{p}_{n,x}(v_x\sqrt{n}). \tag{7.3}$$

We are now in a position to apply Lemma 6.1 to the sequence $\widehat{X}_k(x)$ and write

$$\widehat{p}_{n,x}(z) = \varphi(z) + B\frac{\beta_3(x)M(x)^2}{\sqrt{n}}, \quad z \in \mathbb{R}, \tag{7.4}$$

where the quantity $B = B_n(z)$ is bounded by an absolute constant, $\beta_3(x) = \mathbb{E}|\widehat{X}(x)|^3$ and $M(x) = M(\widehat{X}(x))$. The latter maximum can be bounded by virtue of the upper bound (4.2):

$$M\big(\widehat{X}(x)\big) = \sigma_x M\big(X(x)\big) = \sigma_x M(Q_x p) \le \frac{c\sigma_x}{\sqrt{2\pi}}\, e^{A(x)}.$$

In this case, (7.4) may be simplified with a new B to

$$\widehat{p}_{n,x}(z) = \varphi(z) + Bc^2\frac{\beta_3(x)\sigma_x^2}{\sqrt{n}} e^{2A(x)}.$$

Inserting this in (7.3) with $z = v_x\sqrt{n}$, again with a new B we arrive at

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \frac{1}{\sigma_x}e^{-nA(x) - nv_x^2/2} + Bc^2\frac{\beta_3(x)\sigma_x}{\sqrt{n}}e^{-(n-2)A(x)}.$$

To further simplify, assume that $x \in A_n(a)$ with $n \ge 4(a+1)$. Then, by Lemmas 5.1–5.2, $\beta_3(x) \le C\sigma_x^{-3}$, while $\sigma_x^{-1} \le 2ce^{A(x)}$. Hence,

$$\beta_3(x)\sigma_x\, e^{-(n-2)A(x)} \le 4Cc^2 e^{-(n-4)A(x)} \le 4Cc^2. \qquad \square$$

8. Proof of Theorem 2.1. Recall that the assumptions (1)–(2) stated before Theorem 2.1 are necessary for the convergence $T_\infty(p_n\|\varphi) \to 0$ as $n \to \infty$. For simplicity, we assume that $n_0 = 1$, that is, X is a strictly sub-Gaussian random variable with mean zero, variance one and with finite constant $c = 1 + T_\infty(p\|\varphi)$. In particular, the function

$$A(x) = \frac{1}{2}x^2 - K(x)$$

is nonnegative on the whole real line.

Sufficiency part. The critical zones $A_n(a) = \{x \in \mathbb{R} : A(x) \leq \frac{a}{n-1}\}$ were defined for a parameter $a > 0$ and $n \geq 2$. Choosing $a = \log(1/\varepsilon)$ for a given $\varepsilon \in (0, 1)$, we have by (4.6),

$$\sup_{x \notin A_n(a)} \frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \leq c\sqrt{2}\varepsilon. \tag{8.1}$$

In the case $x \in A_n(a)$ with $n \geq 4(a + 1)$, the equality (7.1) is applicable and implies

$$\sup_{x \in A_n(a)} \frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \leq \sup_{x \in A_n(a)} \frac{1}{\sigma_x} + O\Big(\frac{1}{\sqrt{n}}\Big),$$

where we recall that $\sigma_x^2 = K''(x)$. Using (8.1), we conclude that

$$1 + T_\infty(p_n\|\varphi) \leq \sup_{x \in A_n(a)} \frac{1}{\sigma_x} + c\sqrt{2}\varepsilon + O\Big(\frac{1}{\sqrt{n}}\Big).$$

Thus, a sufficient condition for the convergence $T_\infty(p_n\|\varphi) \to 0$ as $n \to \infty$ is that, for any $\varepsilon \in (0, 1)$,

$$\limsup_{n\to\infty} \sup_{x \in A_n(\log(1/\varepsilon))} \sigma_x^{-2} \leq 1.$$

Equivalently, we need to require that $\liminf_{n\to\infty} \inf_{x \in A_n(a)} K''(x) \geq 1$ for any $a > 0$, that is,

$$\limsup_{n\to\infty} \sup_{x \in A_n(a)} A''(x) \leq 0.$$

Since $A(x) = O(1/n)$ on every set $A_n(a)$, the above may be written as the following continuity condition:

$$\lim_{A(x)\to 0} \max\big(A''(x), 0\big) = 0. \tag{8.2}$$

Necessity part. To see that the condition (8.2) is also necessary for the convergence in T_∞, let us return to the representation (7.1). Assuming that $T_\infty(p_n\|\varphi) \to 0$, it implies that for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{x \in A_n(a)} \frac{1}{\sigma_x} \exp\Big\{-n\Big(A(x) + \frac{1}{2}v_x^2\Big)\Big\} \leq 1, \tag{8.3}$$

where $v_x = A'(x)/\sigma_x$. Recall that

$$A'(x)^2 \leq 2A(x), \quad \sigma_x^{-2} \leq \frac{6}{\pi}c^2 e^{A(x)}$$

(cf. Remark 3.5 and Lemma 5.1). Hence,

$$v_x^2 \leq \frac{2A(x)}{\sigma_x^2} \leq \frac{12}{\pi}c^2 e^{2A(x)} A(x) \leq 12c^2 A(x),$$

assuming that $x \in A_n(a)$ with $a \le 1/2$ and $n \ge 2$ in the last step. Since $nA(x) \le 2a$ on the set $A_n(a)$ and $c \ge 1$, it follows that

$$A(x) + \frac{1}{2}v_x^2 \le 7\,c^2 A(x) \le \frac{14c^2}{n}a.$$

Thus, (8.3) implies that

$$\limsup_{n\to\infty} \sup_{x\in A_n(a)} \frac{1}{\sigma_x} \le e^{14\,c^2 a}, \quad 0 < a \le 1/2.$$

Therefore, for all $n \ge n(a)$,

$$\inf_{x\in A_n(a)} K''(x) \ge e^{-30\,c^2 a}.$$

Since a may be as small as we wish, we conclude that, for any $\varepsilon > 0$, there is $\delta > 0$ such that $A(x) \le \delta \Rightarrow K''(x) \ge 1-\varepsilon$, or $A(x) \le \delta \Rightarrow A''(x) \le \varepsilon$. But this is the same as (8.2). □

One wide class of strictly sub-Gaussian distributions with mean zero and variance one is described in terms of the Laplace transform $L(t) = \mathbb{E}\,e^{tX}$ via the potential requirement (2.2), that is,

$$L(t) \le (1-\delta)\,e^{t^2/2} \tag{8.4}$$

for all $t_0 > 0$ and $|t| \ge t_0$ with some $\delta = \delta(t_0)$, $\delta \in (0,1)$. In this case, the log-Laplace transform and the A-function satisfy

$$K(t) \le \frac{1}{2}t^2 + \log(1-\delta), \quad A(t) \ge -\log(1-\delta).$$

Hence, the approach $A(t) \to 0$ is only possible when $t \to 0$. But, for strictly sub-Gaussian distributions, we necessarily have $A(t) = O(t^4)$ and $A''(t) = O(t^2)$ near zero. Therefore, the condition (8.2) is fulfilled.

COROLLARY 8.1. *If a random variable X with mean zero, variance one and finite distance $T_\infty(p\|\varphi)$ satisfies the separation property* (8.4), *then $T_\infty(p_n\|\varphi) \to 0$ as $n \to \infty$.*

9. Characterization in the periodic case. Examples. Let us apply Theorem 2.1 to the Laplace transforms $L(t)$ with

$$\Psi(t) = L(t)\,e^{-t^2/2} = \mathbb{E}\,e^{tX}e^{-t^2/2}, \quad t \in \mathbb{R}, \tag{9.1}$$

being periodic, with some period $h > 0$. Suppose that $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$, and assume that:

(1) Z_n has density p_n for some $n = n_0$ such that $T_\infty(p_n\|\varphi) < \infty$;
(2) X is strictly sub-Gaussian, that is, $L(t) \le e^{t^2/2}$, or equivalently $\Psi(t) \le 1$ for all $t \in \mathbb{R}$.

PROOF OF THEOREM 2.3 (FIRST PART). We need to show that the convergence $T_\infty(p_n\|\varphi) \to 0$ is equivalent to the assertion that, for every $0 < t < h$,

$$\Psi(t) = 1 \;\Rightarrow\; \Psi''(t) = 0. \tag{9.2}$$

First, note that due to $\Psi(t)$ being analytic, the equation $\Psi(t) = 1$ has finitely many solutions in the interval $[0,h]$ only, including the points $t = 0$ and $t = h$ (by the periodicity). Hence, the condition (b) in Theorem 2.1 may be ignored, and we obtain that $T_\infty(p_n\|\varphi) \to 0$ as $n \to \infty$, if and only if

$$A''(t) = 0 \text{ for every point } t \in [0,h] \text{ such that } A(t) = 0. \tag{9.3}$$

Here, one may exclude the endpoints, since $A''(0) = A''(h) = 0$, by the strict sub-Gaussianity and periodicity. As for the interior points $t \in (0, h)$, note that $A(t) = -\log \Psi(t)$ has the second derivative

$$A''(t) = \frac{\Psi'(t)^2 - \Psi''(t)\Psi(t)}{\Psi(t)^2} = -\Psi''(t)$$

at every point t such that $\Psi(t) = 1$ (in which case $\Psi'(t) = 0$ due to the property $\Psi \leq 1$). This shows that (9.3) is reduced to the condition (9.2). □

In order to describe examples illustrating Theorem 2.3, let us start with the following.

DEFINITION. We say that the distribution μ of a random variable X is periodic with respect to the standard normal law, with period $h > 0$, if it has a density $p(x)$ such that the function

$$q(x) = \frac{p(x)}{\varphi(x)} = \frac{d\mu(x)}{d\gamma(x)}, \quad x \in \mathbb{R},$$

is periodic with period h, that is, $q(x+h) = q(x)$ for all $x \in \mathbb{R}$.

Here, q represents the density of μ with respect to the standard Gaussian measure γ. We denote the class of all such distributions by $\mathfrak{F}_h$, and say that X belongs to $\mathfrak{F}_h$. Let us briefly collect and recall without proof several observations from [10] on this interesting class of probability distributions (cf. Sections 10–13).

PROPOSITION 9.1. *If X belongs to $\mathfrak{F}_h$, then X is sub-Gaussian, and the function $\Psi(t)$ in (9.1) is h-periodic. It may be extended to the complex plane as an entire function. Conversely, if $\Psi(t)$ for a sub-Gaussian random variable X is h-periodic, then X belongs to $\mathfrak{F}_h$, as long as the characteristic function $f(t)$ of X is integrable.*

Since

$$f(t) = L(it) = \Psi(it)\, e^{-t^2/2},$$

the integrability assumption in the reverse statement is fulfilled, as long as $\Psi(z)$ has order smaller than 2, that is, when $|\Psi(z)| = O(\exp\{|z|^\rho\})$ as $|z| \to \infty$ for some $\rho < 2$.

The periodicity property is stable under convolutions: The normalized sums Z_n belong to $\mathfrak{F}_{h\sqrt{n}}$, as long as X belongs to $\mathfrak{F}_h$.

This class contains distributions whose Laplace transform has the form $L(t) = \Psi(t)\, e^{t^2/2}$, where Ψ is a trigonometric polynomial. More precisely, consider functions of the form

$$\Psi(t) = 1 - cP(t), \qquad P(t) = a_0 + \sum_{k=1}^{N} \big(a_k \cos(kt) + b_k \sin(kt)\big),$$

where a_k, b_k are given real coefficients, and $c \in \mathbb{R}$ is a nonzero parameter.

PROPOSITION 9.2. *If $P(0) = 0$ and $|c|$ is small enough, then $L(t)$ represents the Laplace transform of a sub-Gaussian random variable X with density $p(x) = q(x)\varphi(x)$, where $q(x)$ is a nonnegative trigonometric polynomial of degree at most N.*

Necessarily q is bounded, so that $T_\infty(p\|\varphi) < \infty$.

As for the requirement that $P(0) = a_0 + a_1 + \cdots + a_N = 0$, it guarantees that $\int_{-\infty}^{\infty} p(x)\, dx = 1$. In order to apply Theorem 2.3, there are two more constraints coming from the assumption that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$.

COROLLARY 9.3. *Suppose that the polynomial $P(t)$ satisfies*:

(1) $P(0) = P'(0) = P''(0) = 0$;
(2) $P(t) \ge 0$ *for* $0 < t < h$, *where h is the smallest period of P*.

If $c > 0$ is small enough, then $L(t)$ represents the Laplace transform of a strictly sub-Gaussian random variable X. Moreover, if $P(t) > 0$ for $0 < t < h$, then $T_\infty(p_n\|\varphi) \to 0$ as $n \to \infty$.

In terms of the coefficients of the polynomial, the moment assumptions $P'(0) = P''(0) = 0$ are equivalent to $\sum_{k=1}^N kb_k = \sum_{k=1}^N k^2 a_k = 0$. The assumption (2) implies that $0 < \Psi(t) \le 1$, and if $P(t) > 0$ for $0 < t < h$, then the equation $\Psi(t) = 1$ has no solution in this interval.

EXAMPLE 9.4. Consider the transforms of the form

$$L(t) = \big(1 - c \sin^m(t)\big) e^{t^2/2} \tag{9.4}$$

with an arbitrary integer $m \ge 3$, where $|c|$ is small enough. Then $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, and the cumulants of X satisfy $\gamma_k(X) = 0$ for all $3 \le k \le m - 1$.

Moreover, if $m \ge 4$ is even, and $c > 0$ is small enough, the random variable X with the Laplace transform (9.4) is strictly sub-Gaussian. Hence, the conditions in Corollary 9.3 are met, and we obtain the statement about the Rényi divergence of infinite order. In the case $m = 4$, (9.4) corresponds to

$$P(t) = \sin^4 t = \frac{1}{8}\big(3 - 4\cos(2t) + \cos(4t)\big).$$

EXAMPLE 9.5. Put

$$P(t) = \big(1 - 4\sin^2 t\big)^2 \sin^4 t. \tag{9.5}$$

Then $P(t) = O(t^4)$, implying that $P(0) = P'(0) = P''(0) = 0$. Note that $\Psi(t) = 1 - cP(t)$ is π-periodic, and $h = \pi$ is the smallest period, although

$$\Psi(0) = \Psi(t_0) = \Psi(\pi) = 1, \quad t_0 = \pi/6.$$

As we know, if $c > 0$ is small enough, then $L(t) = 1 - c\Psi(t)$ represents the Laplace transform of a strictly sub-Gaussian random variable X. In this case, the last assertion in Corollary 9.3 is not applicable. Thus, the property that h is the smallest period for a periodic function $\Psi(t)$ such that $0 \le \Psi(t) \le 1$ and $\Psi(h) = 1$ does not guarantee that $0 < \Psi(t) < 1$ for $0 < t < h$.

Nevertheless, all assumptions of Theorem 2.3 are fulfilled for sufficiently small $c > 0$ with $h = \pi$, and we may check the condition (9.2). In this case,

$$\Psi(t) = 1 - cQ(t)^2, \qquad Q(t) = \big(1 - 4\sin^2 t\big)\sin^2 t = \sin^2 t - 4\sin^4 t,$$

so that $\Psi''(t) = -2cQ'(t)^2$ at the points t such that $Q(t) = 0$, that is, for $t = t_0$. Hence, $\Psi''(t) = 0 \Leftrightarrow Q'(t) = 0$. In our case,

$$Q'(t) = 2\sin t\cos t - 16\sin^3 t\cos t = \sin(2t)\big(1 - 8\sin^2 t\big),$$

and $Q'(t_0) = -\frac{1}{2}\sqrt{3} \ne 0$. Hence, $\Psi''(t_0) \ne 0$, showing that the condition (9.2) is *not* fulfilled. Thus, the CLT with respect to T_∞ does not hold in this example.

The examples based on trigonometric polynomials may be generalized to the setting of 2π-periodic functions represented by Fourier series

$$P(t) = a_0 + \sum_{k=1}^{\infty} \big(a_k\cos(kt) + b_k\sin(kt)\big).$$

Then the assertions in Proposition 9.2 and Corollary 9.3 will continue to hold, as long as the coefficients satisfy $\sum_{k=1}^{\infty} e^{k^2/2}(|a_k| + |b_k|) < \infty$.

10. Richter's local limit theorem and its refinement. We now turn to the problem of convergence rates with respect to T_∞, which can be explored, for example, under the separation-type condition (2.2). In this case, it was shown in Corollary 4.3 that $p_n(x)$ is much smaller than $\varphi(x)$ outside the interval $|x| = O(\sqrt{n})$. In the region $|x| = o(\sqrt{n})$, an asymptotic behavior of the densities p_n of the normalized sums

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

is governed by the following theorem due to Richter [30]. Assume that $(X_n)_{n\geq 1}$ are independent copies of a random variable X with mean $\mathbb{E}X = 0$ and variance $\mathrm{Var}(X) = 1$.

THEOREM 10.1. *Let $\mathbb{E}\,e^{c|X|} < \infty$ for some $c > 0$, and let Z_n have a bounded density for some n. Then Z_n with large n have bounded continuous densities p_n satisfying*

$$\frac{p_n(x)}{\varphi(x)} = \exp\Big\{\frac{x^3}{\sqrt{n}}\,\lambda\Big(\frac{x}{\sqrt{n}}\Big)\Big\}\Big(1 + O\Big(\frac{1+|x|}{\sqrt{n}}\Big)\Big) \tag{10.1}$$

uniformly for $|x| = o(\sqrt{n})$. The function $\lambda(z)$ is represented by an infinite power series which is absolutely convergent in a neighborhood of $z = 0$.

The corresponding representation

$$\lambda(z) = \sum_{k=0}^{\infty} \lambda_k z^k \tag{10.2}$$

is called Cramer's series; it is analytic in some disc $|z| \leq \tau_0$ of the complex plane. The proof of this theorem may also be found in the book by Ibragimov and Linnik [18] (cf. Theorem 7.1.1) where it was assumed that X has a continuous bounded density. The representation (10.1) was further investigated there for zones of normal attraction $|x| = o(n^\alpha)$, $\alpha < \frac{1}{2}$.

One immediate consequence of (10.1) is that

$$\frac{p_n(x)}{\varphi(x)} \to 1 \quad \text{as } n \to \infty \tag{10.3}$$

uniformly in the region $|x| = o(n^{1/6})$. However, in general this is no longer true outside this region. To better understand the possible behavior of densities, one needs to involve the information about the coefficients in the power series (10.2). As was already mentioned in [18], $\lambda_0 = \frac{1}{6}\gamma_3$, $\lambda_1 = \frac{1}{24}(\gamma_4 - 3\gamma_3^2)$. However, in order to judge the behavior of $\lambda(z)$ for small z, one should describe the leading term in this series. The analysis of the saddle point associated to the log-Laplace transform of the distribution of X shows that

$$\lambda(z) = \frac{\gamma_m}{m!}\,z^{m-3} + O\big(|z|^{m-2}\big) \quad \text{as } z \to 0, \tag{10.4}$$

where γ_m denotes the first nonzero cumulant of X (when X is not normal). Equivalently, m is the smallest integer such that $m \geq 3$ and $\mathbb{E}X^m \neq \mathbb{E}Z^m$, where $Z \sim N(0,1)$. In this case, $\gamma_m = \mathbb{E}X^m - \mathbb{E}Z^m$.

Using (10.4) in (10.1), we obtain a more informative representation

$$\frac{p_n(x)}{\varphi(x)} = \exp\Big\{\frac{\gamma_m}{m!}\,\frac{x^m}{n^{\frac{m}{2}-1}} + O\Big(\frac{x^{m+1}}{n^{\frac{m}{2}}}\Big)\Big\}\Big(1 + O\Big(\frac{1+|x|}{\sqrt{n}}\Big)\Big), \tag{10.5}$$

which holds uniformly for $|x| = o(\sqrt{n})$. With this refinement, the convergence in (10.3) holds true uniformly over all x in the potentially larger region

$$|x| \leq \varepsilon_n n^{\frac{1}{2}-\frac{1}{m}} \quad (\varepsilon_n \to 0).$$

For example, if the distribution of X is symmetric about the origin, then $\gamma_3 = 0$, so that necessarily $m \geq 4$.

Nevertheless, for an application to the T_∞-distance, it is desirable to get some information for larger intervals such as $|x| \leq \tau_0\sqrt{n}$ and to replace the term $O(\frac{|x|}{\sqrt{n}})$ in (10.5) with an explicit n-dependent quantity. For this aim, the following refinement of Theorem 10.1 was recently proved in [9].

THEOREM 10.2. *Let the conditions of Theorem* 10.1 *be fulfilled. There is* $\tau_0 > 0$ *with the following property. Putting* $\tau = x/\sqrt{n}$, *we have for* $|\tau| \leq \tau_0$,

$$\frac{p_n(x)}{\varphi(x)} = e^{n\tau^3 \lambda(\tau) - \mu(\tau)}\Big(1 + O\big(n^{-1}(\log n)^3\big)\Big), \tag{10.6}$$

where $\mu(\tau)$ *is an analytic function in* $|\tau| \leq \tau_0$ *such that* $\mu(0) = 0$.

Here, similar to (10.4),

$$\mu(\tau) = \frac{1}{2(m-2)!}\,\gamma_m \tau^{m-2} + O\big(|\tau|^{m-1}\big).$$

As a consequence of (10.6), we have the following assertion, which was also derived in [9] (note that it cannot be obtained on the basis of (10.1) or (10.5)).

COROLLARY 10.3. *Under the same conditions, suppose that first nonzero cumulant* γ_m *of* X *is negative and* $m \geq 4$ *is even. There exist constants* $\tau_0 > 0$ *and* $c > 0$ *with the following property. If* $|\tau| \leq \tau_0$, $\tau = x/\sqrt{n}$, *then*

$$\frac{p_n(x)}{\varphi(x)} \leq 1 + \frac{c(\log n)^3}{n}. \tag{10.7}$$

PROOF OF THEOREM 2.2. It remains to combine Corollary 4.3 with Corollary 10.3 and note that, for any strictly sub-Gaussian random variable X with variance one, m is even and $\gamma_m < 0$. Indeed, the log-Laplace transform of the distribution of X admits the following Taylor expansion near zero:

$$K(t) = \log \mathbb{E}\, e^{tX} = \frac{1}{2}t^2 + \sum_{k=3}^{\infty} \frac{\gamma_k}{k!} t^k$$
$$= \frac{1}{2}t^2 + \frac{\gamma_m}{m!} t^m + O\big(t^{m+1}\big),$$

which is a definition of cumulants. Hence, the strict sub-Gaussianity, that is, the property $K(t) \leq \frac{1}{2}t^2$ for all $t \in \mathbb{R}$ implies the claim. □

PROOF OF THEOREM 2.3 (CONVERGENCE PART). For simplicity, let $n_0 = 1$, so that the random variable X has density p with $T_\infty(p\|\varphi) < \infty$. By the assumption, $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$ and

$$L(t) = \mathbb{E}\, e^{tX} = \Psi(t)\, e^{t^2/2}, \quad t \in \mathbb{R},$$

for some periodic function $\Psi(t)$ with period $h > 0$ such that $0 < \Psi(t) < 1$ for all $0 < t < h$. Hence,

$$L(t/\sqrt{n})^n = \mathbb{E}\, e^{tZ_n} = \Psi_n(t)\, e^{t^2/2}, \quad \Psi_n(t) = \Psi(t/\sqrt{n})^n,$$

where the function $\Psi_n(t)$ has period $h\sqrt{n}$. Since the density p is bounded, the characteristic function of X is square integrable. Hence, the characteristic function of Z_n is integrable whenever $n \geq 2$. In this case, we are in position to apply Proposition 9.1 to the random variable Z_n and conclude that it has a continuous density p_n, which is periodic with respect to the standard normal law with period $h\sqrt{n}$. That is, $p_n(x) = q_n(x)\varphi(x)$ for some continuous, periodic function q_n with period $h\sqrt{n}$. We need to show that

$$\sup_x \big(q_n(x) - 1\big) = O\bigg(\frac{(\log n)^3}{n}\bigg) \quad \text{as } n \to \infty. \tag{10.8}$$

In view of periodicity, one may restrict this supremum to the interval $0 \leq x \leq h\sqrt{n}$. But, if $0 \leq x \leq \tau_0\sqrt{n}$, where τ_0 is taken as in Corollary 10.3, we obtain the desired rate due to (10.7). Here, without loss of generality one may assume that $\tau_0 < h$. Since $q_n(x) = q_n(x - h\sqrt{n})$, the same conclusion is also true, if we restrict the supremum to $(h - \tau_0)\sqrt{n} \leq x \leq h\sqrt{n}$. Finally, if $\tau_0\sqrt{n} \leq x \leq (h - \tau_0)\sqrt{n}$, we apply the bound (4.3), which gives

$$q_n(x) \leq c\sqrt{2}\,\Psi\bigg(\frac{x}{\sqrt{n}}\bigg)^{n-1}, \quad c = 1 + T_\infty(p\|\varphi).$$

Since $\Psi(t)$ is continuous, $\sup_{\tau_0 \leq t \leq h-\tau_0} \Psi(t) < 1$. Hence, the expression on the right-hand side is exponentially small for growing n. Collecting these estimates, we arrive at (10.8). □

11. Examples based on weighted sums. Here, we describe some examples illustrating Theorem 2.2. It involves the separation condition (2.2) on the Laplace transform,

$$\sup_{|t| \geq t_0} \big[e^{-t^2/2}\,\mathbb{E}\,e^{tX}\big] < 1 \quad \text{for all } t_0 > 0, \tag{11.1}$$

and states the following speed of convergence in the CLT:

$$D_\infty(p_n\|\varphi) = O\bigg(\frac{(\log n)^3}{n}\bigg) \quad \text{as } n \to \infty, \tag{11.2}$$

provided that the necessary condition $D_\infty(p_n\|\varphi) < \infty$ for some $n = n_0$ holds, where p_n denote the densities of the normalized sums Z_n constructed for independent copies of a random variable X with $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$.

While in general this condition is rather delicate, in the simplest case $n_0 = 1$, it reduces to the pointwise sub-Gaussian bound

$$p(x) \leq M\varphi(x), \quad x \in \mathbb{R}, \tag{11.3}$$

which should hold with some constant M for a density p of the random variable X. This property is obviously fulfilled, when the density p is bounded and compactly supported; the rate (11.2) holds as well for a family of probability distributions whose Laplace transform contains a periodic component (see remarks after Proposition 9.2). We now consider further examples where the density p is representable as a "weighted" convolution of at least two densities satisfying (11.3). More precisely, we have the following.

COROLLARY 11.1. *Assume that X satisfies* (11.1) *and is represented as*

$$X = c_0\eta_0 + c_1\eta_1 + c_2\eta_2, \quad c_0^2 + c_1^2 + c_2^2 = 1, c_1, c_2 > 0, \tag{11.4}$$

where the independent random variables η_k, $k = 0, 1, 2$, are strictly sub-Gaussian with variance one and satisfy $D_\infty(\eta_k\|\varphi) < \infty$ for $k = 1, 2$. Then the CLT holds with rate (11.2).

As an interesting subclass, one may consider infinite weighted convolutions, that is, random variables of the form

$$X = \sum_{k=1}^{\infty} a_k \xi_k, \quad \sum_{k=1}^{\infty} a_k^2 = 1. \tag{11.5}$$

COROLLARY 11.2. *Assume that the i.i.d. random variables ξ_k are strictly sub-Gaussian and have a bounded, compactly supported density with variance* $\mathrm{Var}(\xi_1) = 1$. *If ξ_1 satisfies* (11.1), *then the CLT holds with rate* (11.2).

This statement includes, for example, infinite weighted convolutions of the uniform distribution on a bounded symmetric interval.

By Theorem 2.2, Corollary 11.1 follows from the next general assertion.

LEMMA 11.3. *Suppose that the random variable X is represented in the form* (11.4), *where the random variables η_0, η_1, η_2 are independent and possess the properties*:

(a) *η_0 is strictly sub-Gaussian with* $\mathrm{Var}(\eta_0) = 1$;
(b) *η_1, η_2 have densities q_1, q_2 such that $q_k(x) \le M_k \varphi(x)$ for all $x \in \mathbb{R}$ with some constants M_k $(k = 1, 2)$.*

Then X has a density p satisfying (11.3) *with constant $M = \frac{1}{\sqrt{2c_1c_2}} M_1 M_2$.*

PROOF. The case $c_0 = 0$ is simple. Then X has density

$$p(x) = \frac{1}{c_1c_2} \int_{-\infty}^{\infty} q_1\left(\frac{x-y}{c_1}\right) q_2\left(\frac{y}{c_1}\right) dy, \quad x \in \mathbb{R},$$

which, by the assumption, does not exceed

$$\frac{M_1M_2}{c_1c_2} \int_{-\infty}^{\infty} \varphi\left(\frac{x-y}{c_1}\right) \varphi\left(\frac{y}{c_1}\right) dy = M_1 M_2 \, \varphi(x).$$

Hence, (11.3) is fulfilled with constant $M = M_1 M_2$ (which is better than what is claimed in the lemma, since $2c_1c_2 \le 1$).

In the basic case $c_0 > 0$, introduce the characteristic functions $f_k(t)$ of η_k and put $g_k(t) = f_k(c_k t)$, $k = 0, 1, 2$. Since the densities q_1, q_2 are bounded, they belong to $L^2(\mathbb{R})$ together with their characteristic functions f_1, f_2, according to the Plancherel theorem. The same is true for g_1, g_2, so that the characteristic function of X,

$$f(t) = g_0(t) g_1(t) g_2(t), \tag{11.6}$$

is integrable on the real line (using $|g_0(t)| \le 1$ for all $t \in \mathbb{R}$). As a consequence, the random variable X has a continuous density described by the inversion formula

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) \, dt, \quad x \in \mathbb{R}. \tag{11.7}$$

Moreover, the pointwise sub-Gaussian bounds on the densities q_k in b) for $k = 1, 2$ ensure that $\mathbb{E}\, e^{\lambda \eta_k^2} < \infty$ for $\lambda < \frac{1}{2}$, implying that the random variables η_k are sub-Gaussian. Since η_0 is also sub-Gaussian, we conclude that X is sub-Gaussian as well. Hence, all $g_k(t)$ and $f(t)$ may be extended from the real line to the complex plane as entire functions of order at most 2, and thus, (11.6) holds true for all $t \in \mathbb{C}$.

For definiteness, let $x < 0$ in (11.7). We use a contour integration to obtain a different representation for $p(x)$. Fix $T > 0$, $y > 0$, and apply Cauchy's formula for the oriented contour consisting of the segments $[-T, T]$, $[T, T+iy]$, $[T+iy, -T+iy]$, $[-T+iy, -T]$,

$$\begin{aligned} &\int_{-T}^{T} e^{-itx} f(t)\,dt + \int_0^y e^{-i(T+ih)x} f(T+ih)\,dh \\ &\quad = \int_{-T}^{T} e^{-i(t+iy)x} f(t+iy)\,dt + \int_0^y e^{-i(-T+ih)x} f(-T+ih)\,dh. \end{aligned} \tag{11.8}$$

Here, the two integrals taken over the interval $[0, y]$ are vanishing as $T \to \infty$. To prove this, first let us note that the functions

$$q_{k,h}(x) = e^{-hx} q_k(x), \quad x \in \mathbb{R}\ (k = 1, 2),$$

are integrable for every $h \in \mathbb{R}$ and have the Fourier transform

$$\widehat{q}_{k,h}(t) = \int_{-\infty}^{\infty} e^{itx} q_{k,h}(x)\,dx = \mathbb{E}e^{i(t+ih)\eta_k} = f_k(t+ih).$$

We may therefore conclude by applying the Riemann– Lebesgue lemma that $f_k(t+ih) \to 0$ as $|t| \to \infty$. Moreover, this convergence is uniform over all $0 \le h \le y$, which is due to the assumption b). Indeed, since the mapping $h \to q_{k,h}$ from $[0, h]$ to $L^1(\mathbb{R})$ is continuous, for any $\varepsilon > 0$, one can choose the points $0 = h_0 < h_1 < \cdots < h_N = y$ such that

$$\|q_{k,h} - q_{k,h_j}\|_{L^1} < \varepsilon \quad \text{for all } h \in [h_j, h_{j+1}],\ 0 \le j \le N-1.$$

In particular, $\sup_t |\widehat{q}_{k,h}(t) - \widehat{q}_{k,h_j}(t)| < \varepsilon$. By the Riemann–Lebesgue lemma, for every j, there is $t_j > 0$ such that $\sup_{|t| \ge t_j} |\widehat{q}_{k,h_j}(t)| < \varepsilon$. As a consequence,

$$\sup_{h \in [0,y]} \sup_{|t| \ge T} \bigl| f_k(t+ih)\bigr| < 2\varepsilon,$$

by choosing $T = \max\{t_1, \dots, t_N\}$. A similar property holds true for g_k, $k = 1, 2$ and, therefore, for the characteristic function f in (11.6), we get

$$\sup_{h \in [0,y]} \sup_{|t| \ge T} \bigl| f(t+ih)\bigr| \to 0 \quad \text{as } T \to \infty.$$

As a result, in the limit as $T \to \infty$ the identity (11.8) leads to the equivalent variant of (11.7),

$$p(x) = \frac{e^{yx}}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t+iy)\,dt,$$

which yields

$$p(x) \le \frac{e^{yx}}{2\pi} \int_{-\infty}^{\infty} \bigl| f(t+iy)\bigr|\,dt. \tag{11.9}$$

In the next step, we need to estimate the integrand in (11.9). In view of the bound,

$$\bigl| g_0(t+iy)\bigr| = \bigl|\mathbb{E}\, e^{ic_0(t+iy)\eta_0}\bigr| \le \mathbb{E}\, e^{-c_0 y \eta_0} = g_0(iy),$$

equation (11.6) gives

$$\bigl| f(t+iy)\bigr| \le g_0(iy)\, \bigl| g_1(t+iy)\bigr| \bigl| g_2(t+iy)\bigr|.$$

Applying this in (11.9) and using Cauchy's inequality, we get

$$\begin{aligned}
p(x) &\le e^{yx} g_0(iy)\,\frac{1}{2\pi}\int_{-\infty}^{\infty}\big|g_1(t+iy)\big|\big|g_2(t+iy)\big|\,dt \\
&\le e^{yx} g_0(iy)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\big|g_1(t+iy)\big|^2\,dt\right)^{1/2}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\big|g_2(t+iy)\big|^2\,dt\right)^{1/2} \\
&= \frac{e^{yx} f_0(ic_0y)}{2\pi\sqrt{c_1c_2}}\left(\int_{-\infty}^{\infty}\big|f_1(t+ic_1y)\big|^2\,dt\right)^{1/2}\left(\int_{-\infty}^{\infty}\big|f_2(t+ic_2y)\big|^2\,dt\right)^{1/2}.
\end{aligned}$$

Applying the Plancherel theorem and using the pointwise sub-Gaussian bound in (b), we get

$$\begin{aligned}
\frac{1}{2\pi}\int_{\infty}^{\infty}\big|f_k(t+ic_ky)\big|^2\,dt &= \int_{\infty}^{\infty} e^{-2c_kyx} q_k^2(x)\,dx \\
&\le M_k^2\int_{\infty}^{\infty} e^{-2c_kyx}\varphi^2(x)\,dx = \frac{M_k^2}{2\sqrt{\pi}}\,e^{c_k^2y^2}.
\end{aligned}$$

In addition, by the assumption $a)$, $f_0(ic_0y) = \mathbb{E}\,e^{-c_0y\eta_0} \le e^{c_0^2y^2/2}$. Combining these estimates, we arrive at

$$p(x) \le \frac{e^{yx}}{\sqrt{2c_1c_2}}\,\frac{M_1M_2}{\sqrt{2\pi}}\,e^{(c_0^2+c_1^2+c_2^2)\,y^2/2}.$$

It remains to choose $y = -x$ and recall the assumption $c_0^2 + c_1^2 + c_2^2 = 1$. □

We conclude this section with the following.

PROOF OF COROLLARY 11.2. To apply Theorem 2.2, we only need to check that X has a density $p(x)$ satisfying (11.3). Let $q(x)$ denote the common density of ξ_k, which is supposed to be bounded and compactly supported. Without loss of generality, let $a_1 \ge a_2 \ge \cdots \ge 0$.

Case 1: $a_1 = 1$ and $a_k = 0$ for all $k \ge 2$. Then $p = q$, so that $p(x) \le M_1\varphi(x)$ a.e. for some constant $M_1 \ge 1$.

Case 2: $a_2 > 0$. Then $X = c_0\eta_0 + c_1\eta_1 + c_2\eta_2$, where

$$c_0\eta_0 = \sum_{k=3}^{\infty} a_k\xi_k, \qquad \eta_1 = \xi_1, \eta_2 = \xi_2, \quad c_1 = a_1, c_2 = a_2.$$

If $a_3 > 0$, then $c_0 = \sqrt{1 - a_1^2 - a_2^2}$, so, η_0 is well-defined, strictly sub-Gaussian, and has variance one. Otherwise, we may put $c_0\eta_0 = 0$. By Lemma 11.3, the relation $p(x) \le M\varphi(x)$ a.e. holds true with constant $M = \frac{1}{\sqrt{2a_1a_2}}M_1^2$, thus proving (11.3). □

Acknowledgment. We would like to thank the referee for a careful reading and useful comments.

Funding. This research has been supported by the NSF Grant DMS-2154001 and the GRF–SFB 1283/2 2021—317210226.

REFERENCES

[1] ARBEL, J., MARCHAL, O. and NGUYEN, H. D. (2020). On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables. *ESAIM Probab. Stat.* **24** 39–55. MR4053001 https://doi.org/10.1051/ps/2019018

[2] 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 https://doi.org/10.1090/S0894-0347-04-00459-X

[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 https://doi.org/10.1007/s00440-003-0329-4

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

[5] BOBKOV, S. G. and CHISTYAKOV, G. P. (2012). Bounds for the maximum of the density of the sum of independent random variables. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **408** 62–73, 324. MR3032208 https://doi.org/10.1007/s10958-014-1836-9

[6] 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 https://doi.org/10.1214/12-AOP780

[7] 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 https://doi.org/10.1007/s00440-013-0510-3

[8] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2019). Rényi divergence and the central limit theorem. *Ann. Probab.* **47** 270–323. MR3909970 https://doi.org/10.1214/18-AOP1261

[9] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2023). Richter's local limit theorem, its refinement, and related results. *Lith. Math. J.* **63** 138–160. MR4621090 https://doi.org/10.1007/s10986-023-09598-9

[10] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2024). Strictly subgaussian probability distributions. *Electron. J. Probab.* **29** Paper No. 62, 28. MR4736269 https://doi.org/10.1214/24-ejp1120

[11] BOBKOV, S. G. and GÖTZE, F. (2024). Esscher transform and the central limit theorem. Preprint.

[12] BULDYGIN, V. V. and KOZAČENKO, YU. V. (1980). Sub-Gaussian random variables. *Ukr. Mat. Zh.* **32** 723–730. MR0598605

[13] BULDYGIN, V. V. and KOZACHENKO, YU. V. (2000). *Metric Characterization of Random Variables and Random Processes. Translations of Mathematical Monographs* **188**. Amer. Math. Soc., Providence, RI. Translated from the 1998 Russian original by V. Zaiats. MR1743716 https://doi.org/10.1090/mmono/188

[14] DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

[15] DEMBO, A., COVER, T. M. and THOMAS, J. A. (1991). Information-theoretic inequalities. *IEEE Trans. Inf. Theory* **37** 1501–1518. MR1134291 https://doi.org/10.1109/18.104312

[16] ESSCHER, F. (1932). On the probability function in the collective theory of risk. *Skand. Aktuarietidskr.* **15** 175–195.

[17] GUIONNET, A. and HUSSON, J. (2020). Large deviations for the largest eigenvalue of Rademacher matrices. *Ann. Probab.* **48** 1436–1465. MR4112720 https://doi.org/10.1214/19-AOP1398

[18] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov. Translation from the Russian edited by J. F. C. Kingman. MR0322926

[19] JOHNSON, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press, London. MR2109042 https://doi.org/10.1142/9781860945373

[20] KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. Dover, New York. Translated by G. Gamow. MR0029808

[21] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[22] MADIMAN, M. and BARRON, A. (2007). Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inf. Theory* **53** 2317–2329. MR2319376 https://doi.org/10.1109/TIT.2007.899484

[23] NEWMAN, C. M. (1975). Inequalities for Ising models and field theories which obey the Lee–Yang theorem. *Comm. Math. Phys.* **41** 1–9. MR0376061

[24] NEWMAN, C. M. (1975). Moment inequalities for ferromagnetic Gibbs distributions. *J. Math. Phys.* **16** 1956–1959. MR0391853 https://doi.org/10.1063/1.522748

[25] NEWMAN, C. M. (1975). An extension of Khintchine's inequality. *Bull. Amer. Math. Soc.* **81** 913–915. MR0375458 https://doi.org/10.1090/S0002-9904-1975-13884-5

[26] NEWMAN, C. M. and WU, W. (2019). Lee–Yang property and Gaussian multiplicative chaos. *Comm. Math. Phys.* **369** 153–170. MR3959556 https://doi.org/10.1007/s00220-019-03453-0

[27] PETROV, V. V. (1964). Local limit theorems for sums of independent random variables. *Teor. Veroyatn. Primen.* **9** 343–352. MR0163341

[28] PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* [*Results in Mathematics and Related Areas*], *Band* 82. Springer, New York. Translated from the Russian by A. A. Brown. MR0388499

[29] PROKHOROV, YU. V. (1952). A local theorem for densities. *Dokl*. *Akad*. *Nauk SSSR* **83** 797–800. MR0049501

[30] RIHTER, W. (1957). Lokale Grenzwertsätze für grosse Abweichungen. *Teor. Veroyatn*. *Primen*. **2** 214–229. MR0092260

[31] 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 https://doi.org/10.1007/978-1-4757-2539-1

[32] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans*. *Inf*. *Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500