Journal of Functional Analysis 289 (2025) 110999

Contents lists available at ScienceDirect

Journal of Functional Analysis

journal homepage: www.elsevier.com/locate/jfa

Regular Article

Esscher transform and the central limit theorem

Sergey G. Bobkov [a], Friedrich Götze [b,*]

[a] School of Mathematics, University of Minnesota, Minneapolis, MN, USA
[b] Faculty of Mathematics, Bielefeld University, Germany

A R T I C L E I N F O

Article history:
Received 31 July 2024
Accepted 6 April 2025
Available online 11 April 2025
Communicated by Pietro Caputo

MSC:
primary 60E, 60F

Keywords:
Central limit theorem
Rényi divergence
Esscher transform

A B S T R A C T

The paper is devoted to the investigation of Esscher's transform on high dimensional Euclidean spaces in the light of its application to the central limit theorem. With this tool, we explore necessary and sufficient conditions of normal approximation for normalized sums of i.i.d. random vectors in terms of the Rényi divergence of infinite order, extending recent one dimensional results.

Contents

* Corresponding author.
E-mail addresses: bobkov@math.umn.edu (S.G. Bobkov), goetze@math.uni-bielefeld.de (F. Götze).

https://doi.org/10.1016/j.jfa.2025.110999

1. Introduction

Introduce the normalized sums

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

of independent copies of a random vector X in $\mathbb{R}^d$ with mean zero and identity covariance matrix I_d. It is well known that, if Z_n have densities p_n for large n, their distributions are convergent in total variation norm to the standard normal law on $\mathbb{R}^d$ with density $\varphi(x) = \frac{1}{(2\pi)^{d/2}} \exp(-|x|^2/2)$. That is, we have the convergence in L^1-norm for densities

$$\int_{\mathbb{R}^d} |p_n - \varphi| \, dx \to 0 \quad (n \to \infty),$$

which was first emphasized by Prokhorov [18]. A much stronger property, which may or may not hold in general, is described by means of the Rényi divergence

$$D_\alpha(p_n||\varphi) = \frac{1}{\alpha - 1} \log \int_{\mathbb{R}^d} \Big(\frac{p_n}{\varphi}\Big)^\alpha \varphi \, dx$$

of order $\alpha \geq 1$ (the relative α-entropy). These distance-like functionals are increasing for growing α, and in the limit they define the Rényi divergence of infinite order

$$D_\infty(p_n||\varphi) = \text{ess sup}_x \, \log \frac{p_n(x)}{\varphi(x)}.$$

The convergence in relative entropy

$$D(p_n||\varphi) = D_1(p_n||\varphi) = \int_{\mathbb{R}^d} p_n \log \frac{p_n}{\varphi} \, dx$$

(which is the Kullback-Leibler distance) was the subject of numerous investigations starting with Linnik [15], who initiated an information-theoretic approach to the central limit theorem. Let us only mention the work by Barron [2] for necessary and sufficient conditions and later [1], [4], [5] for the problems of rates and Berry-Esseen bounds. The case

$\alpha > 1$ and, in particular, the Pearson χ^2-distance was treated in detail in [6], where the following statement was obtained as a consequence of a certain characterization of the convergence with respect to D_α in terms of the Laplace transform.

Theorem 1.1. *Assume that $D_\alpha(p_n||\varphi) < \infty$ for every $\alpha \in (1, \infty)$ with some $n = n_\alpha$. For the convergence $D_\alpha(p_n||\varphi) \to 0$ for any finite α, it is necessary and sufficient that*

$$\mathbb{E}\, e^{\langle t,X\rangle} \le e^{|t|^2/2} \quad \text{for all } t \in \mathbb{R}^d. \tag{1.1}$$

The assumption of finiteness of $D_\alpha(p_n||\varphi)$ may be equivalently stated as the property that Z_n have bounded densities for large n.

The inequality (1.1) describes a remarkable class of probability distributions which appear naturally in many mathematical problems. In modern literature, (1.1) is often called strict sub-Gaussianity. We refer an interested reader to [7] for the history, references, and recent developments towards the problem of characterization of such distributions in dimension one.

The convergence in D_∞ is equivalent to the convergence with respect to

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

in view of the relation $T_\infty = e^{D_\infty} - 1$. This quantity looks more natural than D_∞ for local limit theorems, where $|p_n(x) - \varphi(x)|$ is commonly estimated on $\mathbb{R}^d$ with polynomially growing weights. However, if X is not normal, the statement

$$\operatorname{ess\,sup}_x \frac{p_n(x) - \varphi(x)}{\varphi(x)} \to 0 \quad \text{as } n \to \infty, \tag{1.2}$$

cannot be strengthened to absolute value under the supremum. Indeed, if for example, X is bounded, then $p_n(x)$ is compactly supported, and the above ratio is equal to -1 for large $|x|$ (see also Corollary 6.5 and the relation (6.5) after it).

One of the purposes of this paper is to give necessary and sufficient conditions for the multidimensional CLT such as (1.2) in terms of the Laplace transform $L(t) = \mathbb{E}\, e^{\langle t,X\rangle}$. Introduce $K(t) = \log L(t)$ (which is a convex smooth function) and define the function

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

As before, suppose that $(X_k)_{k\ge1}$ are independent copies of the random vector X in $\mathbb{R}^d$ with mean zero and identity covariance matrix. Below 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 $t \in \mathbb{R}^d$.

Theorem 1.2. *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* $t \in \mathbb{R}^d$ *such that* $A(t) = 0$*;*
b) $\lim_{k\to\infty} A''(t_k) = 0$ *for every sequence* $|t_k| \to \infty$ *such that* $A(t_k) \to 0$ *as* $k \to \infty$*.*

Here and elsewhere A'' denotes the Hessian, that is, the $d \times d$ matrix of second order partial derivatives of A. The conditions $a) - b)$ may be combined in the requirement

$$\lim_{A(t)\to 0} A''(t) = 0, \quad \text{or equivalently} \quad \lim_{A(t)\to 0} K''(t) = I_d$$

which is a kind of continuity of A'' with respect to A. As we will see, these conditions may also be stated in a formally weaker form as

$a')$ $\det K''(t) = 1$ for every $t \in \mathbb{R}^d$ such that $A(t) = 0$;
$b')$ $\lim_{k\to\infty} \det K''(t_k) = 1$ for any sequence $|t_k| \to \infty$ such that $A(t_k) \to 0$ as $k \to \infty$.

In dimension $d = 1$ Theorem 1.2 has been obtained in [8], where $a) - b)$ are stated as a weaker condition $\limsup_{A(t)\to 0} A''(t) \le 0$. The multidimensional situation turns out to be more complicated, since it requires a careful treatment of eigenvalues of the matrix $K''(t)$, when $A(t)$ approaches zero. Another ingredient in the proof is a quantitative version of the uniform local limit theorem, which was recently developed in [9].

Assuming the strict sub-Gaussianity (1.1), the conditions $a) - b)$ may or may not hold in general. This shows that the convergence in D_∞ is stronger than the convergence in D_α simultaneously for all finite α. Nevertheless, for a wide class of strictly sub-Gaussian distributions the Laplace transform possesses a separation-type property

$$\sup_{|t|\ge t_0} \Big[e^{-|t|^2/2}\, \mathbb{E}\, e^{\langle t,X\rangle}\Big] < 1 \quad \text{for all } t_0 > 0. \tag{1.3}$$

This is a strengthened form of condition 2), which entails properties $a) - b)$.

Corollary 1.3. *If a random vector X with mean zero and identity covariance matrix satisfies (1.3), then $T_\infty(p_n||\varphi) \to 0$ as $n \to \infty$.*

On the other hand, the case of equality in the sub-Gaussian bound (1.1) is quite possible, and one can observe new features in the multidimensional case. While in dimension one, an equality $L(t) = e^{|t|^2/2}$ is only possible for a discrete set of points t, in higher dimensions the set of points where this equality holds may have dimension $d-1$. In order to clarify this behavior, we will discuss the class of Laplace transforms which contain periodic components. Specializing Theorem 1.2 to this class, the general characterization may be simplified.

Corollary 1.4. *Suppose that the function* $\Psi(t) = L(t)\, e^{-|t|^2/2}$ *is* h*-periodic for some vector* $h \in \mathbb{R}^d_+$ $(h \neq 0)$*. For the convergence* $T_\infty(p_n||\varphi) \to 0$ *as* $n \to \infty$*, it is necessary and sufficient that, for every* $t \in [0, h]$*,*

$$\Psi(t) = 1 \Rightarrow \Psi''(t) = 0.$$

Our approach to Theorem 1.2 is based on the application of the Esscher transforms Q_h on $\mathbb{R}^d$ defined below. Recall that a random vector X in $\mathbb{R}^d$ is called sub-Gaussian, if $\mathbb{E}\, e^{c|X|^2} < \infty$ for some $c > 0$.

Definition 1.5. Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with distribution μ. Introduce the family of probability distributions $\mu_h = Q_h \mu$ on $\mathbb{R}^d$ with parameter $h \in \mathbb{R}^d$ which have densities with respect to μ

$$\frac{d\mu_h(x)}{d\mu(x)} = \frac{1}{L(h)}\, e^{\langle h,x\rangle}, \quad x \in \mathbb{R}^d. \tag{1.4}$$

The early history of this transform in dimension one goes back to Esscher [13] in actuarial science, Khinchin [14] in statistical mechanics, and Daniels [12] who applied it to develop asymptotic expansions for densities. It is used to study large deviations for instance in Central Limit theorems and sometimes called Cramér's transform as well. For examples, see [19].

As a result, in (1.4) we obtain a semi-group of probability measures $\{Q_h \mu\}_{h \in \mathbb{R}^d}$ on $\mathbb{R}^d$ with the following remarkable property: Every Q_h transforms the convolution of several measures to the convolutions of their Q_h-transforms. This is analogous to the property that the Fourier transform of convolutions represents the product of Fourier transforms. Somewhat surprisingly, for the study of convergence in T_∞ an application of the Q_h-transform effectively replaces Fourier calculus. Indeed, the proof of Theorem 1.2 makes use of the Fourier analysis only in a minor way.

We will discuss the action of the Q-transform on single distributions in Sections 2-5 and then turn to convolutions in Sections 6-7. The proof of Theorem 1.2 and Corollaries 1.3-1.4 are given in Section 8-9. The remaining Sections 10-14 illustrate these results for several classes of probability distributions and specific examples.

2. Semigroup of shifted distributions. A basic identity

Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with distribution μ. Then, the Laplace transform, or the moment generating function

$$(L\mu)(t) = L(t) = \mathbb{E}\, e^{\langle t,X\rangle} = \int_{\mathbb{R}^d} e^{\langle t,x\rangle}\, d\mu(x), \quad t \in \mathbb{R}^d,$$

is finite and represents a C^∞-smooth function on $\mathbb{R}^d$. Correspondingly, the log-Laplace transform

$$(K\mu)(t) = K(t) = \log L(t) = \log \mathbb{E}\, e^{\langle t, X\rangle}$$

is a convex, C^∞-smooth function on $\mathbb{R}^d$.

The measures $\mu_h = Q_h\mu$ are defined in (1.4) by means of the Esscher transform. In the absolutely continuous case, Definition 1.5 is reduced to the following:

Definition 2.1. If the random vector X has density p, the measure μ_h has density which we denote similarly as

$$Q_h p(x) = \frac{1}{L(h)}\, e^{\langle h,x\rangle} p(x). \tag{2.1}$$

In this case, let us also write $L\mu = Lp$ and $K\mu = Kp$.

We will call μ_h the shifted distribution of X at step h 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 (1.4) is the semi-group property

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

Similarly, in the space of probability densities as in (2.1), we have

$$Q_{h_1}(Q_{h_2}p) = Q_{h_1+h_2}p.$$

Let us also mention how this transform acts under rescaling, for simplicity in the absolutely continuous case. Given $\lambda > 0$, the random vector λX has density $p_\lambda(x) = \lambda^{-d}\, p(x/\lambda)$ with Laplace transform $(Lp_\lambda)(t) = L(\lambda t)$. Hence

$$Q_h p_\lambda(x) = \frac{1}{(Lp_\lambda)(h)}\, e^{\langle h,x\rangle} p_\lambda(x) = \frac{1}{\lambda^d}\, (Q_{\lambda h}p)\Big(\frac{x}{\lambda}\Big). \tag{2.2}$$

The transform Q_h is multiplicative with respect to convolutions.

Proposition 2.2. *If independent sub-Gaussian random vectors in $\mathbb{R}^d$ have distributions $\mu_1, \dots, \mu_n$, then for the convolution $\mu = \mu_1 * \dots * \mu_n$, we have*

$$Q_h\mu = Q_h\mu_1 * \dots * Q_h\mu_n. \tag{2.3}$$

In particular, if μ_k have densities p_k, then for the convolution $p = p_1 * \dots * p_n$, we have

$$Q_h p = Q_h p_1 * \dots * Q_h p_n.$$

Proof. It is sufficient to compare the Laplace transforms of both sides in (2.3). The Laplace transform of μ is given by

$$L\mu(t) = (L\mu_1)(t) \dots (L\mu_n)(t).$$

Hence, the Laplace transform of $Q_h\mu$ is given by

$$\begin{aligned}(LQ_h\mu)(t) &= \int_{\mathbb{R}^d} e^{\langle t,x\rangle}\, dQ_h\mu(x) \;=\; \frac{1}{(L\mu)(t)} \int_{\mathbb{R}^d} e^{\langle t+h,x\rangle} d\mu(x) \\ &= \frac{(L\mu)(t+h)}{(L\mu)(t)} \;=\; \prod_{k=1}^n \frac{(L\mu_k)(t+h)}{(L\mu_k)(t)} \;=\; \prod_{k=1}^n (LQ_h\mu_k)(t). \qquad \square\end{aligned}$$

If the random vector X has density p, the formula (2.1) may be rewritten as

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

or

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

As in the introductory section, we consider the smooth function on $\mathbb{R}^d$

$$(Ap)(h) = A(h) = \frac{1}{2}\,|h|^2 - K(h). \tag{2.4}$$

It allows one to reformulate the property $L(h) \le e^{\frac{1}{2}\,|h|^2}$ as $A(h) \ge 0$ for all $h \in \mathbb{R}^d$. Note that this is equivalent to the strict sub-Gaussianity of X, when this random vector has mean zero and identity covariance matrix. Thus, we have:

Proposition 2.3. *Given a sub-Gaussian random vector X in $\mathbb{R}^d$ with density p,*

$$\frac{p(x)}{\varphi(x)} = (2\pi)^{d/2}\, e^{\frac{1}{2}\,|x-h|^2 - A(h)}\, Q_h p(x), \quad x, h \in \mathbb{R}^d, \tag{2.5}$$

where A is the associated function to p.

This is a basic identity which will be used with $h = x$ to bound from above the ratio on the left for densities of normalized sums of independent copies of X.

Since the function $A(h)$ plays an essential role in the representation (2.5), a number of its properties will be important in the sequel. Some of them may be explored in the general situation where X does not need to have a density. We denote by $A'(h) = \nabla A(h)$ the gradient and by $A''(h)$ the Hessian of A at the point h.

Proposition 2.4. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with Laplace transform satisfying $A(h) \geq 0$ for all $h \in \mathbb{R}^d$. Then, for all $h \in \mathbb{R}^d$,*

$$|A'(h)|^2 \leq 2A(h), \tag{2.6}$$

$$A''(h) \leq I_d, \tag{2.7}$$

$$A(h) = 0 \;\Rightarrow\; A'(h) = 0 \quad \text{and} \quad A''(h) \geq 0. \tag{2.8}$$

Proof. The inequalities in (2.7)-(2.8) are understood in matrix sense. The implication in (2.8) is obvious. Also, since the function K is convex, the assertion (2.7) follows from the definition (2.4). To prove (2.6), let us write the Taylor integral formula for A at the point h up to the quadratic term

$$A(h+x) = A(h) + \langle A'(h), x\rangle + \int_0^1 (1-s)\,\langle A''(h+sx)x, x\rangle\, ds. \tag{2.9}$$

By (2.7), and using the assumption $A \geq 0$, we have, for all $x \in \mathbb{R}^d$,

$$0 \leq A(h) + \langle A'(h), x\rangle + \frac{1}{2}\,|x|^2.$$

Minimizing the right-hand side over all x, we arrive at (2.6). $\square$

Note that an application of (2.6)-(2.7) in the Taylor formula (2.9) also implies that

$$|A(h+x) - A(h)| \leq \sqrt{2A(h)}\,|x| + \frac{1}{2}\,|x|^2, \quad x, h \in \mathbb{R}^d. \tag{2.10}$$

To explore some other properties of the function A, we need to look at the moments of shifted distributions.

3. Moments of shifted distributions

For a sub-Gaussian random vector X in $\mathbb{R}^d$ with distribution μ, denote by $X(h)$ a random vector with distribution $\mu_h = Q_h\mu$ ($h \in \mathbb{R}^d$). It is sub-Gaussian, and its Laplace and log-Laplace transforms are given by

$$L_h(t) = \mathbb{E}\, e^{\langle t, X(h)\rangle} = \frac{L(t+h)}{L(h)},$$

$$K_h(t) = \log L_h(t) = K(t+h) - K(h). \tag{3.1}$$

This random vector has mean

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

where we recall that $L'(h) = \nabla L(h)$ and $K'(h) = \nabla K(h)$ denote the gradients of L and K respectively. In addition, since for all $t, h \in \mathbb{R}^d$,

$$\mathrm{Var}\big(\langle t, X(h)\rangle\big) \ = \ \frac{d^2}{d\lambda^2}\, K_h(\lambda t)\Big|_{\lambda=0} \ = \ \langle K''(h)t, t\rangle\,,$$

the covariance matrix of $X(h)$ represents the matrix of second order partial derivatives (Hessian)

$$R_h = K''(h), \quad h \in \mathbb{R}^d.$$

This shows that necessarily $K''(h)$ is positive semi-definite. Moreover, if X has a density, it is strictly positive definite, since otherwise the random vector $X(h)$ may not have a density.

The centered random vector

$$\bar{X}(h) = X(h) - m_h \tag{3.2}$$

has mean zero and covariance matrix R_h. Involving the function A and assuming that it is non-negative, the log-Laplace transform is given by and satisfies

$$\begin{aligned}
\bar{K}_h(t) \ = \ \log \mathbb{E}\, e^{\langle t, \bar{X}(h)\rangle} &= K(t+h) - K(h) - \langle t, m_h\rangle \\
&= \Big(\frac{1}{2}\,|t+h|^2 - A(t+h)\Big) - \Big(\frac{1}{2}\,|h|^2 - A(h)\Big) - \langle t, K'(h)\rangle \\
&\leq \frac{1}{2}\,|t+h|^2 - \Big(\frac{1}{2}\,|h|^2 - A(h)\Big) - \langle t, K'(h)\rangle \\
&= \frac{1}{2}\,|t|^2 + A(h) + \langle t, A'(h)\rangle\,.
\end{aligned}$$

Applying (2.6), we get

$$\begin{aligned}
\bar{K}_h(t) &\leq \frac{1}{2}\,|t|^2 + A(h) + |t|\,|A'(h)| \\
&\leq \frac{1}{2}\,|t|^2 + A(h) + |t|\sqrt{2A(h)} \ = \ \frac{1}{2}\,\Big(|t| + \sqrt{2A(h)}\Big)^2.
\end{aligned}$$

Thus, we obtain an upper bound for the Laplace transform.

Proposition 3.1. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with Laplace transform such that $A(h) \geq 0$ for all $h \in \mathbb{R}^d$. Then*

$$\mathbb{E}\, e^{\langle t, \bar{X}(h)\rangle} \leq \exp\Big\{\frac{1}{2}\Big(|t| + \sqrt{2A(h)}\Big)^2\Big\}, \quad t, h \in \mathbb{R}^d. \tag{3.3}$$

This relation shows in a quantitative form the sub-Gaussian character of $\bar{X}$. Indeed, to simplify, one may use $(a+b)^2 \leq 2a^2 + 2b^2$ $(a, b \in \mathbb{R})$, so that (3.3) yields

$$\mathbb{E}\, e^{\langle t, \bar{X}(h)\rangle} \leq \exp\big\{|t|^2 + 2A(h)\big\}.$$

Let us apply this with $t = \frac{1}{2}\,\xi\theta$, $|\theta| = 1$, where ξ is a standard normal random variable independent of X. Taking the expectation with respect to ξ and using $\mathbb{E}\, e^{\frac{1}{4}\,\xi^2} = \sqrt{2}$, we then get

$$\mathbb{E}\, e^{\frac{1}{4}\langle \theta, \bar{X}(h)\rangle^2} \leq \sqrt{2}\, e^{2A(h)}. \tag{3.4}$$

Moreover, by Jensen's inequality, the left expectation is greater than or equal to $\exp\big\{\frac{1}{4}\,\mathbb{E}\,\big\langle \theta, \bar{X}(h)\big\rangle^2\big\}$ which leads to

$$\frac{1}{4}\,\langle R_h\theta, \theta\rangle \leq 2A(h) + \log\sqrt{2}.$$

In particular, one may apply this inequality to orthonormal eigenvectors θ of R_h.

Corollary 3.2. *Under the conditions of Proposition 3.1, the eigenvalues $\lambda_j(h)$ of R_h satisfy, for any $h \in \mathbb{R}^d$,*

$$\max_{1 \leq j \leq d} \lambda_j(h) \ \leq\ 8A(h) + 2\log 2. \tag{3.5}$$

Starting from (3.4), one may similarly estimate higher order moments of linear functionals of $\bar{X}(h)$. The following bound will be needed with $q = 3$.

Corollary 3.3. *Under the conditions of Proposition 3.1, up to some absolute constant C,*

$$\Big(\mathbb{E}\,\big|\big\langle \theta, \bar{X}(h)\big\rangle\big|^q\Big)^{2/q} \leq C\big(q + A(h)\big), \quad q \geq 1. \tag{3.6}$$

Proof. One may assume that $q \geq 2$. By (3.4), for the random variable

$$\eta = \frac{1}{4}\,\big\langle \theta, \bar{X}(h)\big\rangle^2 - 2A(h) - \log\sqrt{2}$$

we have $\mathbb{E}\, e^{\eta} \leq 1$, implying $\mathbb{P}\{\eta \geq x\} \leq e^{-x}$ for all $x \geq 0$. Hence, for any $r \geq 1$,

$$\mathbb{E}\,(\eta^+)^r = r\int_0^\infty x^{r-1}\,\mathbb{P}\{\eta \geq x\}\,dx \leq \Gamma(r+1),$$

where $\eta^+ = \max(\eta, 0)$. Using $\big\langle \theta, \bar{X}(h)\big\rangle^2 \leq 4\eta^+ + 8A(h) + 2\log 2$, it follows that

$$\Big(\mathbb{E}\,\big|\,\langle\theta,\bar X(h)\rangle\,\big|^{2r}\Big)^{1/r} \le 4\,\Gamma(r+1)^{1/r} + 8A(h) + 2\log 2.$$

It remains to apply this inequality with $r = q/2$. $\square$

4. Behavior of eigenvalues near the critical zone

Another application of the sub-Gaussian bound (3.4), more precisely – of the inequality (3.6) with $q = 3$, concerns the behavior of the Hessian $A''(h)$ when $A(h)$ is small (which we shall call the critical zone in view of the inequality (6.3) from Proposition 6.3 below). The following statement complements the property (2.8) from Proposition 2.4 which asserts that $A''(h) \ge 0$ as long as $A(h) = 0$.

Proposition 4.1. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with Laplace transform such that $A(h) \ge 0$ for all $h \in \mathbb{R}^d$. If $0 \le A(h) \le 1$, then*

$$\inf_{|\theta|=1} \langle A''(h)\theta,\theta\rangle \ge -C_d A(h)^{1/4} \tag{4.1}$$

with some constant $C_d > 0$ depending on d only.

One consequence of (4.1), which will be needed for the characterization of the CLT with respect to the Rényi divergence of infinite order is that

$$\liminf_{A(h)\to 0}\ \inf_{|\theta|=1} \langle A''(h)\theta,\theta\rangle \ge 0,$$

or equivalently, since $A'' = I_d - K''$,

$$\limsup_{A(h)\to 0}\ \sup_{|\theta|=1} \langle K''(h)\theta,\theta\rangle \le 1.$$

In terms of the eigenvalues $\lambda_j(h)$ of the covariance matrix R_h of the random vector $X(h)$ this may also be restated as:

Corollary 4.2. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with Laplace transform such that $A(h) \ge 0$ for all $h \in \mathbb{R}^d$. Then*

$$\limsup_{A(h)\to 0}\ \max_{1\le j\le d} \lambda_j(h) \le 1. \tag{4.2}$$

In particular,

$$\limsup_{A(h)\to 0}\ \det K''(h) \le 1.$$

Proof. We keep notations from the previous sections. Recall that

$$\bar{X}(h) = X(h) - \mathbb{E}X(h) = (\xi_1, \dots, \xi_d), \quad h = (h_1, \dots, h_d) \in \mathbb{R}^d.$$

The following elementary identity is needed for the 3rd order partial derivatives of the log-Laplace transform with respect to the variables h_i, h_j, h_k:

$$\partial_{ijk} K(h) = \mathbb{E}\, \xi_i \xi_j \xi_k = -\partial_{ijk} A(h), \quad 1 \le i, j, k \le d.$$

By (3.6) with $q = 3$,

$$\left(\mathbb{E}\, |\xi_i|^3\right)^{1/3} \le C\, (1 + A(h))^{1/2},$$

where C is an absolute constant. Hence, by Hölder's inequality, for any $h \in \mathbb{R}^d$,

$$|\partial_{ijk} A(h)| \le C^3\, (1 + A(h))^{3/2}. \tag{4.3}$$

Let $0 < A(h) \le 1$. Applying the inequality (2.10), we get that, whenever $|x| \le 1$,

$$|A(h + x) - A(h)| \le \sqrt{2A(h)}\, |x| + \frac{1}{2}\, |x|^2 \,<\, 2.$$

Hence $A(h + x) < 3$, so that by (4.3),

$$|\partial_{ijk} A(h + x)| \le C \tag{4.4}$$

with some absolute constant $C > 0$.

We now apply the multidimensional integral Taylor formula up to cubic terms which indicates that

$$\begin{aligned} A(h + x) &= A(h) + \langle A'(h), x\rangle + \frac{1}{2}\, \langle A''(h)x, x\rangle \\ &\quad + \sum_{|\beta| = 3} \frac{3}{\beta!}\, x^\beta \int_0^1 (1 - s)^2\, D^\beta A(h + sx)\, ds. \end{aligned} \tag{4.5}$$

Here we use the standard notation for the partial derivative

$$D^\beta A = \frac{\partial^{|\beta|} A}{\partial h_1^{\beta_1} \dots \partial h_d^{\beta_d}}, \quad \beta = (\beta_1, \dots, \beta_d),$$

where β is a multi-index of length $|\beta| = \beta_1 + \dots + \beta_d$ (with integers $\beta_i \ge 0$), and where

$$x^\beta = x_1^{\beta_1} \dots x_d^{\beta_d} \quad \text{for } x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

If $|x| \le 1$, then, by (4.4), all these partial derivatives for $|\beta| = 3$ do not exceed C in absolute value. Since also $|x^\beta| \le |x|^{|\beta|}$, we obtain from (4.5) that

$$A(h+x) \le A(h) + \sqrt{2A(h)}\,|x| + \frac{1}{2}\,\langle A''(h)x, x\rangle + C_d\,|x|^3 \tag{4.6}$$

with $C_d = C\sum_{|\beta|=3}\frac{1}{\beta!}$. Here we also used (2.6) to bound the linear term in (4.5). Let us choose $x = r\theta$ with $0 < r \le 1$, $|\theta| = 1$, and use the assumption $A(h+x) \ge 0$. Then (4.6) yields

$$\langle A''(h)\theta, \theta\rangle \ge -\frac{2}{r^2}\Big[A(h) + \sqrt{2A(h)}\,r + C_d r^3\Big].$$

In particular, the choice $r = A(h)^{1/4}$ leads to (4.1). $\square$

5. Maximum of shifted densities

In order to bound the last term $Q_h p(x)$ in the basic identity (2.5), suppose that the distribution of X has a finite Rényi distance of infinite order to the standard normal law. This means that X has a density p which admits a point-wise upper bound

$$p(x) \le c\varphi(x), \quad x \in \mathbb{R}^d \text{ (a.e.)} \tag{5.1}$$

with optimal value $c = 1 + T_\infty(p||\varphi)$. In that case, one may control the maximum of the density $Q_h p$ as follows. In the sequel, we use the notation

$$M(\xi) = M(q) = \operatorname{ess\,sup}_x q(x)$$

for a random vector ξ in $\mathbb{R}^d$ with density q.

By (5.1), for almost all $x \in \mathbb{R}^d$, using $\langle x, h\rangle \le \frac{1}{2}\,|x|^2 + \frac{1}{2}\,|h|^2$, we have

$$\begin{aligned} Q_h p(x) &= \frac{1}{L(h)}\,e^{\langle x,h\rangle}p(x) \\ &\le \frac{c\,e^{\langle x,h\rangle - \frac{1}{2}\,|x|^2}}{L(h)\,(2\pi)^{d/2}} \le \frac{c\,e^{\frac{1}{2}\,|h|^2}}{L(h)\,(2\pi)^{d/2}} = \frac{c}{(2\pi)^{d/2}}\,e^{A(h)}, \end{aligned}$$

where L is the Laplace transform of the distribution of X and $A(h) = \frac{1}{2}\,|h|^2 - K(h)$, $K(h) = \log L(h)$. Thus, we arrive at the following elementary relation.

Proposition 5.1. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with density p such that $c = 1 + T_\infty(p||\varphi)$ is finite. Then, for all $h \in \mathbb{R}^d$,*

$$M(Q_h p) \le \frac{c}{(2\pi)^{d/2}}\,e^{A(h)}. \tag{5.2}$$

This inequality may be used to bound the eigenvalues of the covariance matrix R_h in terms of $A(h)$ from below. This complements the upper bound of Corollary 3.2.

In the sequel, put

$$\sigma_h = (\mathrm{det} R_h)^{\frac{1}{2d}} = (\mathrm{det} K''(h))^{\frac{1}{2d}}, \quad h \in \mathbb{R}^d,$$

which is everywhere positive. In dimension $d = 1$, this quantity represents the standard deviation of the random variable $X(h)$.

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

$$\sigma_h^d \geq \frac{1}{c}\, e^{-A(h)-d/2}. \tag{5.3}$$

Proof. The argument employs the following known relation between the covariance matrix and maximum of density whose proof we include for completeness at the end of this section: Given a random vector ξ in $\mathbb{R}^d$ with finite second moment and finite $M = M(\xi)$, we have

$$\big(M^2\, \mathrm{det} R\big)^{\frac{1}{d}} \geq \frac{1}{2\pi e}, \tag{5.4}$$

where R is the covariance matrix of ξ. Applying this inequality to the random vector $\xi = X(h)$ with its covariance matrix $R = R_h$ and using (5.2), we get

$$\frac{1}{(2\pi e)^{d/2}} \leq M(X(h))\, \sigma_h^d \leq \frac{c\, \sigma_h^d}{(2\pi)^{d/2}}\, e^{A(h)},$$

from which (5.3) follows immediately. □

Corollary 5.3. *Let X be a sub-Gaussian random vector in $\mathbb{R}^d$ with Laplace transform such that $A(h) \geq 0$ for all $h \in \mathbb{R}^d$ and with finite $c = 1 + T_\infty(p||\varphi)$. Then, the eigenvalues $\lambda_j(h)$ of R_h satisfy, for any $h \in \mathbb{R}^d$,*

$$\begin{aligned} \frac{1}{c^2}\, \frac{e^{-2A(h)-d}}{(8A(h)+2)^{d-1}} &\leq \min_{1 \leq j \leq d} \lambda_j(h) \\ &\leq \max_{1 \leq j \leq d} \lambda_j(h) \,\leq\, 8A(h) + 2. \end{aligned} \tag{5.5}$$

Proof. Put $\lambda_j = \lambda_j(h)$ and $\alpha = A(h)$. The upper bound in (5.5) is provided in Corollary 3.2. On the other hand, by (5.3),

$$\sigma_h^{2d} = \mathrm{det} R_h = \lambda_1 \dots \lambda_d \geq \frac{1}{c^2}\, e^{-2\alpha-d}. \tag{5.6}$$

Therefore, by the upper bound in (5.5),

$$\lambda_1 \dots \lambda_d \le \min_j \lambda_j \, (\max_j \lambda_j)^{d-1} \le \min_j \lambda_j \, (8\alpha + 2)^{d-1},$$

and (5.6) yields the lower bound in (5.5). $\square$

Proof of (5.4). We follow a simple information-theoretic approach proposed in [10]. Introduce the entropy functional

$$h(p) = -\int_{\mathbb{R}^d} p(x) \log p(x) \, dx. \tag{5.7}$$

It is well defined for absolutely continuous distributions with finite second moment and is maximized for the normal distribution when the covariance matrix is fixed. Indeed, without loss of generality, let ξ have mean zero. If ζ has a normal density q on $\mathbb{R}^d$ with mean zero and the same covariance matrix R, then $\mathbb{E} \left\langle R^{-1}\xi, \xi \right\rangle = \mathbb{E} \left\langle R^{-1}\zeta, \zeta \right\rangle = d$, implying

$$h(q) - h(p) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{q(x)} \, dx.$$

Here the right-hand side defines the relative entropy $D(p||q)$ of p with respect to q (the Kullback-Leibler distance), which is non-negative, by Jensen's inequality.

Now, from one hand,

$$h(q) = \frac{d}{2} \log(C\sigma^2), \quad C = 2\pi e, \quad \sigma = (\det R)^{\frac{1}{2d}}.$$

On the other hand, $h(p) \ge -\log M$, which follows from (5.7) using $\log p(x) \le \log M$ (a.e.) Hence

$$D(p||q) \le \frac{d}{2} \log(C\sigma^2) + \log M,$$

that is,

$$M^2 \det R \ge \frac{1}{(2\pi e)^d} \, e^{2D(q||p)}.$$

This is a sharpened form of (5.4). $\square$

6. Representation for convolutions

We are now prepared to apply these results to the normalized sums

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

of independent copies of a sub-Gaussian random vector X in $\mathbb{R}^d$ with density p. Note that in terms of $L = Lp$, $K = Kp$ and $A = Ap$, for the density p_n of Z_n we have

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

where $h_n = h\sqrt{n}$. Hence, the basic identity (2.5) yields a similar formula.

Proposition 6.1. *Putting $x_n = x\sqrt{n}$, $h_n = h\sqrt{n}$ ($x, h \in \mathbb{R}^d$), the density p_n of Z_n admits the representation*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = (2\pi)^{d/2}\, e^{\frac{n}{2}\,|x-h|^2 - nA(h)}\, Q_{h_n} p_n(x_n). \tag{6.1}$$

This equality becomes 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 6.2. *For all $x, h \in \mathbb{R}^d$,*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le (2\pi)^{d/2}\, e^{\frac{n}{2}\,|x-h|^2 - nA(h)}\, M(Q_{h\sqrt{n}}\, p_n). \tag{6.2}$$

A general upper bound on the M-functional was given in Proposition 5.1. However, it is useless to apply this bound directly in (6.2) to the densities p_n, since then the right-hand side of (5.2) will contain the parameter $c_n = 1 + T_\infty(p_n||\varphi)$. Instead, we use a semi-additive property of the maximum-of-density functional, which indicates that

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

for all independent random vectors X_k in $\mathbb{R}^d$ having bounded densities, cf. [3]. If all X_k are identically distributed and have a density p, this relation yields

$$M(p^{*n}) \le \Big(\frac{e}{n}\Big)^{d/2} M(p)$$

for the convolution n-th power of p. Applying the multiplicativity property of the Esscher transform (Proposition 2.2) together with (5.2), we then have

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

where we recall that $c = 1 + T_\infty(p||\varphi)$. Now, since $p^{*n}(x) = \lambda^{-d}\, p_n(x/\lambda)$ with $\lambda = \sqrt{n}$, one may apply the scaling identity (2.2):

$$Q_h p^{*n}(x) = \frac{1}{\lambda^d}\,(Q_{\lambda h} p_n)\Big(\frac{x}{\lambda}\Big).$$

Hence, a similar identity holds true for the M-functional,

$$M(Q_h p^{*n}) = \frac{1}{n^{d/2}}\,M(Q_{h\sqrt{n}}\,p_n),$$

and we get

$$M(Q_{h\sqrt{n}}\,p_n) \le c\,\Big(\frac{e}{2\pi}\Big)^{d/2}\,e^{A(h)}.$$

One can now return to Corollary 6.2 and apply the above bound to obtain that

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

In particular, with $h = x$ this yields:

Proposition 6.3. *If the density p has finite Rényi distance of infinite order to the standard normal law, then, for almost all $x \in \mathbb{R}^d$,*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le c\,e^{d/2}\,e^{-(n-1)A(x)}, \tag{6.3}$$

where $c = 1 + T_\infty(p||\varphi)$.

Corollary 6.4. *If additionally X has identity covariance matrix and is strictly sub-Gaussian, then*

$$T_\infty(p_n||\varphi) \le e^{d/2}\,(1 + T_\infty(p||\varphi)) - 1.$$

Thus, the finiteness of the Tsallis distance $T_\infty(p||\varphi)$ for a strictly sub-Gaussian random vector 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 (6.3) shows that its left-hand side is exponentially small for growing n. In particular, this holds for any non-normal random vector X satisfying the separation property (1.3). Then we immediately obtain:

Corollary 6.5. *Suppose that X has a density p with finite $T_\infty(p||\varphi)$. Under the condition (1.3), for any $\tau_0 > 0$, there exist $A > 0$ and $\delta \in (0,1)$ such that the densities p_n of Z_n satisfy*

$$p_n(x) \le A\delta^n \varphi(x), \quad |x| \ge \tau_0\sqrt{n}. \tag{6.4}$$

In particular,

$$\liminf_{n\to\infty} \sup_{x\in\mathbb{R}^d} \frac{|p_n(x)-\varphi(x)|}{\varphi(x)} \geq 1. \tag{6.5}$$

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

Although (1.3) does not hold in general for strictly sub-Gaussian distributions, Proposition 6.3 may be applied outside the set of points where $A(x)$ is bounded away from zero. If $A(x)$ is close to zero, we say that the point x belongs to the critical zone (a precise definition will be given in Section 8). In this case, we need to return to the basic representation of Proposition 6.1 and study the last term $Q_{h_n}p_n(x_n)$. This requires to apply a variant of the local limit theorem, using the property that the density $Q_{h_n}p_n$ has a convolution structure.

7. Local limit theorem for shifted distributions

Keeping the notations and assumptions from the previous section, recall that $R_h = K''(h)$ represents the covariance matrix of $X(h)$. Since it is symmetric and strictly positive definite, one may consider the normalized random vectors

$$\begin{aligned}\widehat{X}(h) &= R_h^{-1/2}(X(h) - \mathbb{E}X(x)) \\ &= R_h^{-1/2}(X(h) - m_h), \qquad h\in\mathbb{R}^d,\end{aligned}$$

which have mean zero and identity covariance matrix. We will have to consider convolution powers of distributions of $\widehat{X}(h)$ by means of a multidimensional local limit theorem for the points h where the value

$$A(h) = \frac{1}{2}\,|h|^2 - K(h) = \frac{1}{2}\,|h|^2 - \log\mathbb{E}\,e^{\langle X,h\rangle}$$

is small. More precisely, here we prove the following refinement of the representation (6.1). Consider the vector-function

$$\begin{aligned}v_x &= R_x^{-1/2}(x - m_x) \\ &= R_x^{-1/2}(x - K'(x)) \;=\; R_x^{-1/2}A'(x), \quad x\in\mathbb{R}^d,\end{aligned} \tag{7.1}$$

and recall that

$$\sigma_x = (\det R_x)^{\frac{1}{2d}} = (\det K''(x))^{\frac{1}{2d}}.$$

Proposition 7.1. *If the Laplace transform of a sub-Gaussian random vector X in $\mathbb{R}^d$ with finite $c = 1 + T_\infty(p||\varphi)$ is such that $A(h)\geq 0$ for all $h\in\mathbb{R}^d$, then for all $x\in\mathbb{R}^d$, $n\geq 6$, we have*

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \frac{1}{\sigma_x^d} \exp\left\{ -nA(x) - n\,|v_x|^2/2\right\} + \frac{B^d c^5}{\sqrt{n}}, \tag{7.2}$$

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

One should note that the term $|v_x|^2$ appearing on the right-hand side of (7.2) must be small for small values of $A(x)$. Indeed, according to Corollary 5.3, the minimal eigenvalue $\lambda = \min_{1\le j\le d} \lambda_j(x)$ of the matrix R_x admits a lower bound

$$\lambda \ge \frac{1}{c^2\,(8A(x)+2)^{d-1}}\, e^{-2A(x)-d}. \tag{7.3}$$

As a consequence, applying Proposition 2.4 in (7.1), we have

$$\begin{aligned} |v_x|^2 &\le \frac{1}{\lambda}\,|A'(x)|^2 \\ &\le 2c^2\,(8A(x)+2)^{d-1}\,A(x)\,e^{2A(x)+d} \le C^d c^2 A(x) \end{aligned} \tag{7.4}$$

for some absolute constant $C > 0$, where we assumed that $A(x) \le 1$ in the last step.

For the derivation of (7.2), we employ a general local limit theorem for densities on $\mathbb{R}^d$ with a quantitative error term, which was recently derived in [9].

Lemma 7.2. *Let $(\xi_k)_{k\ge1}$ be independent copies of a random vector ξ in $\mathbb{R}^d$ with mean zero, identity covariance matrix and finite third absolute moment. Assuming that ξ has a bounded density with maximum $M(\xi)$, the densities q_n of the normalized sums $Z_n = (\xi_1 + \dots + \xi_n)/\sqrt{n}$ satisfy*

$$\sup_x |q_n(x) - \varphi(x)| \le C^d\,\frac{1}{\sqrt{n}}\,M(\xi)^2\,\mathbb{E}\,|\xi|^3, \quad x \in \mathbb{R}^d,$$

with some absolute constant $C > 0$.

Proof of Proposition 7.1. Consider the term $Q_{h_n} p_n$ in (6.1) with $h_n = h\sqrt{n}$. By Proposition 2.2, this density has a convolution structure. It was also emphasized in (2.2) that, for any random vector X with density $p = p_X$,

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

Using this notation, we have $p_n = p_{S_n/\sqrt{n}}$ for the sum $S_n = X_1 + \dots + X_n$. Hence with $\lambda = 1/\sqrt{n}$,

$$Q_{h_n} p_n(x) = n^{d/2}\,(Q_h p_{S_n})(x\sqrt{n}) = n^{d/2}\,(Q_h p) * \dots * (Q_h p)(x\sqrt{n}),$$

where we applied Proposition 2.2 in the last step. Since $Q_h p$ serves as a density of the random vector $X(h)$, $Q_{h_n} p_n(x)$ represents the density for the normalized sum

$$Z_{n,h} = \frac{X_1(h) + \dots + X_n(h)}{\sqrt{n}},$$

assuming that $X_k(h)$ are independent.

Now, introduce the normalized sums

$$\widehat{Z}_{n,h} = \frac{\widehat{X}_1(h) + \dots + \widehat{X}_n(h)}{\sqrt{n}}$$

for the shifted distributions, i.e. with $X_k(h) = m_h + R_h^{1/2} \widehat{X}_k(h)$. Thus,

$$Z_{n,h} = m_h \sqrt{n} + R_h^{1/2} \widehat{Z}_{n,h}.$$

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

$$p_{n,h}(x) \ = \ \frac{1}{\sigma_h^d}\, \widehat{p}_{n,h}\big(R_h^{-1/2}(x - m_h\sqrt{n})\big).$$

At the points $x_n = x\sqrt{n}$ as in (6.1), we therefore obtain that

$$Q_{h_n} p_n(x_n) = p_{n,h}(x_n) = \frac{1}{\sigma_h^d}\, \widehat{p}_{n,h}\big(\sqrt{n}\, R_h^{-1/2}(x - m_h)\big).$$

Consequently, the equality (6.1) may be equivalently stated as

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

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

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = (2\pi)^{d/2}\, e^{-nA(x)}\, \frac{1}{\sigma_x^d}\, \widehat{p}_{n,x}\big(v_x\sqrt{n}\big). \tag{7.5}$$

We are in a position to apply Lemma 7.2 to the sequence $\xi_k = \widehat{X}_k(x)$ and write

$$\widehat{p}_{n,x}(z) = \varphi(z) + B^d\, \frac{\beta_3(x) M_x^2}{\sqrt{n}}, \quad z \in \mathbb{R}^d, \tag{7.6}$$

where

$$\beta_3(x) = \mathbb{E}\, \big|\widehat{X}(x)\big|^3, \quad M_x = M\big(\widehat{X}(x)\big),$$

and where the quantity $B = B_n(z)$ is bounded by an absolute constant. The latter maximum can be bounded by virtue of the upper bound (5.2):

$$\begin{aligned} M_x &= M\big(R_h^{-1/2}(X(h) - m_h)\big) \\ &= \sigma_x^d \, M(X(x)) \, = \, \sigma_x^d \, M(Q_x p) \, \le \, \frac{c\sigma_x^d}{(2\pi)^{d/2}} \, e^{A(x)}. \end{aligned}$$

In this case, (7.6) may be simplified with a new quantity $B = B_n(z)$ to

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

Inserting this in (7.5) with $z = v_x \sqrt{n}$, we arrive at

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \frac{1}{\sigma_x^d} \, e^{-nA(x) - n|v_x|^2/2} + B^d c^2 \, \frac{\beta_3(x) \, \sigma_x^d}{\sqrt{n}} \, e^{-(n-2)A(x)}, \tag{7.7}$$

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

In order to estimate $\beta_3(x)$, let us recall the bound (3.6) with $q = 3$ which implies

$$\mathbb{E} \, |X(x) - m_x|^3 \le C d^{3/2} \, (1 + A(x))^{3/2} \tag{7.8}$$

with some absolute constant C. Also, by (7.3), for any $w \in \mathbb{R}^d$,

$$\big|R_x^{-1/2} w\big| \le c \, (8A(x) + 2)^{\frac{d-1}{2}} \, e^{A(x) + d/2} \, |w|.$$

Applying this with $w = X(x) - m_x$ together with (7.8), we get

$$\begin{aligned} \beta_3(x) &\le c^3 \, (8A(x) + 2)^{\frac{3(d-1)}{2}} \, e^{3A(x) + 3d/2} \cdot C d^{3/2} \, (1 + A(x))^{3/2} \\ &\le C_1^d c^3 \, e^{4A(x)} \end{aligned}$$

for some absolute constant $C_1 > 0$. It remains to insert this bound in (7.7) leading to

$$\frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} = \frac{1}{\sigma_x^d} \, e^{-nA(x) - n|v_x|^2/2} + B^d c^5 \, \frac{\sigma_x^d}{\sqrt{n}} \, e^{-(n-6)A(x)}. \quad \square$$

8. Proof of Theorem 1.2

Recall that the assumptions 1)-2) stated before Theorem 1.2 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 vector with mean zero, identity covariance matrix, and finite constant $c = 1 + T_\infty(p||\varphi)$. In particular, the function

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

is non-negative on the whole space $\mathbb{R}^d$.

According to Corollary 4.2, the function $\sigma_x = (\det K''(x))^{\frac{1}{2d}}$ satisfies

$$\limsup_{A(x)\to 0} \sigma_x^{2d} \le 1. \tag{8.1}$$

First we show that the convergence $T_\infty(p_n||\varphi) \to 0$ is equivalent to

$$\lim_{A(x)\to 0} \sigma_x^{2d} = \lim_{A(x)\to 0} \det K''(x) = 1, \tag{8.2}$$

which is a compact version of the conditions $a') - b')$ mentioned after Theorem 1.2. Introduce the critical zones

$$A_n(a) = \Big\{x \in \mathbb{R}^d : A(x) \le \frac{a}{n-1}\Big\}, \quad a > 0, \ n \ge 2.$$

Sufficiency part. Putting $a = \log(1/\varepsilon)$, $\varepsilon \in (0,1)$, the upper bound (6.3) yields

$$\sup_{x \notin A_n(a)} \frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le ce^{d/2}\,\varepsilon.$$

As for the critical zone, the equality (7.2) is applicable for $n \ge 6$ and implies

$$\sup_{x \in A_n(a)} \frac{p_n(x\sqrt{n})}{\varphi(x\sqrt{n})} \le \sup_{x \in A_n(a)} \frac{1}{\sigma_x^d} + \frac{C}{\sqrt{n}},$$

where the constant C does not depend on n. Both estimates can be combined to give

$$1 + T_\infty(p_n||\varphi) \le \sup_{x \in A_n(a)} \frac{1}{\sigma_x^d} + ce^{d/2}\,\varepsilon + \frac{C}{\sqrt{n}}.$$

Choosing here $\varepsilon = \frac{1}{\sqrt{n}}$, one may conclude that a sufficient condition for the convergence $T_\infty(p_n||\varphi) \to 0$ as $n \to \infty$ is that

$$\liminf_{A(x)\to 0} \sigma_x^{2d} \ge 1, \tag{8.3}$$

which is equivalent to (8.2) in view of (8.1).

Necessity part. To see that the condition (8.2) is also necessary for the convergence in T_∞, let us return to the representation (7.2). Assuming that $T_\infty(p_n||\varphi) \to 0$, it implies that, for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{x \in A_n(a)} \frac{1}{\sigma_x^d} \exp\Big\{-n\Big(A(x) + \frac{1}{2}\,|v_x|^2\Big)\Big\} \le 1. \tag{8.4}$$

As explained in (7.4), $|v_x|^2 \le C^d c^2 A(x)$ for some absolute constant $C > 0$ whenever $A(x) \le 1$. In particular, this holds for all $x \in A_n(a)$ with $n \ge a+1$. Since $nA(x) \le 2a$ on the set $A_n(a)$, it follows that

$$A(x) + \frac{1}{2}\,|v_x|^2 \le \beta A(x) \le \frac{2\beta}{n}\,a,$$

where β depends on c and d only. Thus, (8.4) implies that

$$\limsup_{n\to\infty}\ \sup_{x\in A_n(a)} \frac{1}{\sigma_x^{2d}} \le e^{4\beta a}, \quad 0 < a \le 1.$$

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

$$\inf_{x\in A_n(a)} \sigma_x^{2d} \ge e^{-5\beta a}.$$

Since a may be as small as we wish, we conclude that necessarily

$$\forall \varepsilon > 0\ \exists\, \delta > 0\ \big[A(x) \le \delta\ \Rightarrow\ \det K''(x) \ge 1 - \varepsilon\big]. \tag{8.5}$$

But this is the same as (8.3), which is equivalent to (8.2).

Let us now explain why

$$\lim_{A(x)\to 0} K''(x) = I_d\ \Longleftrightarrow\ \lim_{A(x)\to 0} \det K''(x) = 1. \tag{8.6}$$

The implication "$\Rightarrow$" is obvious. For the opposite direction, we may assume that the necessary condition (8.5) is fulfilled. Since $\det K''(x) = \lambda_1(x)\dots\lambda_d(x)$ in terms of the eigenvalues $\lambda_j(x)$ of R_x, this condition may be stated in a weaker form as

$$\forall \varepsilon > 0\ \exists\, \delta > 0\ \big[A(x) \le \delta\ \Rightarrow\ \min_{1\le j\le d}\ \lambda_j(x) \ge (1-\varepsilon)^{1/d}\,\big]. \tag{8.7}$$

In order to reverse the conclusion, let us return to Corollary 4.2 and recall that the eigenvalues satisfy

$$\limsup_{A(x)\to 0} \lambda_j(x) \le 1, \quad 1 \le j \le d,$$

for any $x \in \mathbb{R}^d$, which is a stronger property compared to (8.1). Thus, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$A(x) \le \delta\ \Rightarrow\ \max_{1\le j\le d} \lambda_j(x) \le 1 + \varepsilon.$$

Being combined with (8.7) with $0 < \varepsilon < 1$, this gives $(1-\varepsilon)^{1/d} \le \lambda_j(x) \le 1 + \varepsilon$ and hence $|\lambda_j(x) - 1| \le \varepsilon$ for all j, as long as $A(x) \le \delta$. In this case it follows that

$$\|K''(x) - I_d\|_{\mathrm{HS}}^2 = \sum_{j=1}^d |\lambda_j(x) - 1|^2 \le d\varepsilon.$$

As a result, we obtain the opposite implication in (8.6).

It remains to see that the property

$$\lim_{A(x) \to 0} A''(x) = 0 \tag{8.8}$$

may be restated as the conditions $a) - b)$ in Theorem 1.2:

$a)$ $A''(x) = 0$ for every point $x \in \mathbb{R}^d$ such that $A(x) = 0$;
$b)$ $\lim_{k\to\infty} A''(x_k) = 0$ for every sequence $|x_k| \to \infty$ such that $A(x_k) \to 0$ as $k \to \infty$.

Obviously, the conditions $a) - b)$ follow from (8.8). For the converse direction, assume that (8.8) is not true. Then there would exist $\varepsilon > 0$ such that, for any $\delta > 0$, one can pick up a point $x \in \mathbb{R}^d$ with the property that

$$A(x) \le \delta \quad \text{and} \quad \|A''(x)\|_{\mathrm{HS}} > \varepsilon.$$

Choosing $\delta = \delta_k \downarrow 0$, we would obtain a sequence $x_k \in \mathbb{R}^d$ such that $A(x_k) \le \delta_k$ and $\|A''(x_k)\|_{\mathrm{HS}} > \varepsilon$. If this sequence is bounded, it would contain a convergent sub-sequence $x_{k_l} \to x$ with $A(x) = 0$ and $\|A''(x)\|_{\mathrm{HS}} \ge \varepsilon$, by continuity of the functions A and A''. But this contradicts $a)$. In the other case, one can subtract a sub-sequence such that $|x_{k_l}| \to \infty$, $A(x_{k_l}) \to 0$ as $l \to \infty$, while $\|A''(x_{k_l})\|_{\mathrm{HS}} > \varepsilon$. But this contradicts $b)$. $\square$

9. Proof of Corollaries 1.3-1.4

As in Theorem 1.2, suppose that the random vector X has mean zero and identity covariance matrix. In addition, 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: $L(t) \le e^{|t|^2/2}$ or equivalently $\Psi(t) \le 1$ for all $t \in \mathbb{R}^d$.

Recall that, by the separation property, we mean the relation

$$L(t) \le (1 - \delta)\, e^{|t|^2/2}, \tag{9.1}$$

which holds for all $t_0 > 0$ and $|t| \ge t_0$ with some $\delta = \delta(t_0)$, $\delta \in (0, 1)$.

Proof of Corollary 1.3. From (9.1) it follows that the log-Laplace transform and the function A 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 (8.8) is fulfilled automatically. $\square$

Next, let us apply Theorem 1.2 to the Laplace transforms $L(t) = \mathbb{E}\, e^{\langle t,X\rangle}$ with

$$\Psi(t) = L(t)\, e^{-|t|^2/2}, \quad t \in \mathbb{R}^d,$$

being periodic, that is, satisfying

$$\Psi(t+h) = \Psi(t), \quad t \in \mathbb{R}^d, \tag{9.2}$$

for some $h = (h_1, \dots, h_d) \in \mathbb{R}^d$, $h \neq 0$. Without loss of generality, let $h_i \geq 0$ and not all of them be zero. Put $[0,h] = [0,h_1] \times \cdots \times [0,h_d]$.

Proof of Corollary 1.4. The function $\Psi(t)$ is positive, and by the assumption (9.2), the function

$$A(t) = -\log \Psi(t) = \frac{1}{2}\,|t|^2 - K(t)$$

is h-periodic as well. To express the condition a) in terms of Ψ, let us differentiate the equality $\Psi(t) = e^{-A(t)}$ to get

$$\partial_{t_i} \Psi(t) = -\partial_{t_i} A(t)\, e^{-A(t)}$$

and

$$\partial^2_{t_i t_j} \Psi(t) = -\partial^2_{t_i t_j} A(t)\, e^{-A(t)} + \partial_{t_i} A(t)\, \partial_{t_j} A(t)\, e^{-A(t)}.$$

Thus, $A''(t) = -\Psi''(t)$ for every point $t \in \mathbb{R}^d$ such $A(t) = 0$. Recall that in this case, necessarily $A'(t) = 0$ and therefore $\Psi'(t) = 0$.

As for the condition b) in Theorem 1.2, it may be reduced to a). Indeed, assume that $A(x_k) \to 0$ for some sequence $x_k \in \mathbb{R}^d$ such that $|x_k| \to \infty$. Using the condition a), we need to show that $A''(x_k) \to 0$. The latter is equivalent to the assertion that from any sub-sequence x_{k_l} one may subtract a further sub-sequence $x_{k_l'}$ such that $A''(x_{k_l'}) \to 0$. For simplicity, let a given sub-sequence be the whole sequence x_k. By the periodicity (9.2), $A(x_k) = A(y_k)$ and $A''(x_k) = A''(y_k)$ for some $y_k \in [0,h]$. By compactness, there is a convergence sub-sequence $y_{k_l} \to y \in [0,h]$ as $l \to \infty$. But then, by continuity, $A(y) = 0$ and hence $A''(y) = 0$, by a). As a consequence, $A''(x_{k_l}) = A''(y_{k_l}) \to A''(y) = 0$. $\square$

10. Laplace transforms with separation property

In dimension $d = 1$, Corollary 1.3 can be illustrated by different examples. Let us recall two results from [7], assuming that X is a sub-Gaussian random variable with mean zero. In this case the characteristic function

$$f(z) = \mathbb{E}\, e^{izX}, \quad z \in \mathbb{C},$$

represents an entire function in the complex plane of order at most 2. If $f(z)$ does not have any real or complex zeros, a well-known theorem due to Marcinkiewicz [16] implies that the distribution of X is already Gaussian. Thus, non-normal sub-Gaussian distributions have characteristic functions that need to have zeros.

The strict sub-Gaussianity is defined by the relation

$$\mathbb{E}\, e^{tX} \leq e^{\sigma^2 t^2/2}, \quad t \in \mathbb{R}, \tag{10.1}$$

where $\sigma^2 = \mathrm{Var}(X)$ is the variance of X.

Proposition 10.1. *If the distribution of X is symmetric, and all zeros of $f(z)$ with $\mathrm{Re}(z) \geq 0$ lie in the cone centered on the real axis defined by*

$$|\mathrm{Arg}(z)| \leq \frac{\pi}{8},$$

then X is strictly sub-Gaussian. Moreover, if X is non-normal, then for any $t_0 > 0$, there exists $c = c(t_0)$ in the interval $0 < c < \sigma^2$ such that

$$\mathbb{E}\, e^{tX} \leq e^{ct^2/2}, \quad |t| \geq t_0. \tag{10.2}$$

The inequality (10.2) strengthens not only (10.1), but also the separation relation (9.1) (for $\sigma^2 = 1$). The claim about the strict sub-Gaussianity in Proposition 10.1 refines a theorem due to Newman [17], who considered the case where $f(z)$ has only real zeros (cf. also [11]). It was also shown in [7] that the condition $|\mathrm{Arg}(z)| \leq \frac{\pi}{8}$ is also necessary for the strict sub-Gaussianity of X, when it has a symmetric distribution, and its characteristic function $f(z)$ has exactly one zero z in the quadrant $\mathrm{Re}(z) > 0$, $\mathrm{Im}(z) > 0$.

The next assertion provides a sufficient condition for the property (10.2).

Proposition 10.2. *If X is non-normal, and the function $K(\sqrt{|t|})$ is concave on the half-axis $t > 0$ and is concave on the half-axis $t < 0$, then (10.2) holds true.*

In [7] one can find various examples illustrating these propositions. In particular, the symmetric Bernoulli and the uniform distribution on a symmetric interval are strictly sub-Gaussian, as well as convergent infinite convolutions of such distributions. Moreover, they satisfy the separation property (10.2).

Turning to the multidimensional situation, let us only mention two examples. For a sub-Gaussian random vector X in $\mathbb{R}^d$ with mean zero and covariance matrix $\sigma^2 I_d$, the notion of the strict sub-Gaussianity is defined by

$$\mathbb{E}\, e^{\langle t,X\rangle} \leq e^{\sigma^2 |t|^2/2}, \quad t \in \mathbb{R}^d. \tag{10.3}$$

This class of probability distributions is invariant under convolutions and weak limits.

It should be clear that, a product measure on $\mathbb{R}^d$ satisfies (10.3), if and only if all marginals are strictly sub-Gaussian. For spherically invariant distributions, (10.3) is also reduced to dimension one.

Proposition 10.3. *Suppose that the distribution of a sub-Gaussian random vector $X = (X_1, \dots, X_d)$ is spherically invariant. Then X is strictly sub-Gaussian, if and only if X_1 is strictly sub-Gaussian. In this case, if X is non-normal, and X_1 satisfies (10.2), then for any $t_0 > 0$, there exists $c = c(t_0)$, $0 < c < \sigma^2$, such that*

$$\mathbb{E}\, e^{\langle t,X\rangle} \leq e^{c|t|^2/2}, \quad t \in \mathbb{R}^d, \ |t| \geq t_0. \tag{10.4}$$

Proof. By the assumption, the random vectors X and UX are equidistributed for any linear orthogonal transformation of the space $\mathbb{R}^d$. Given $t \in \mathbb{R}^d$, choose U such that $U't = |t|e_1 = |t|\,(1, 0, \dots, 0)$. Then

$$\mathbb{E}\, e^{\langle t,X\rangle} = \mathbb{E}\, e^{\langle t,UX\rangle} = \mathbb{E}\, e^{\langle U't,X\rangle} = \mathbb{E}\, e^{|t|X_1}.$$

Since the distribution of X_1 is symmetric about the origin, (10.3) is equivalent to (10.1) with X_1 in place of X. The same is true about the equivalence of (10.2) and (10.4). $\square$

Here are two basic examples, where as before, p_n denotes the density of the normalized sum of n independent copies of X.

Corollary 10.4. *The uniform distribution on the Euclidean ball $B(r)$ in $\mathbb{R}^d$ with center at the origin and radius $r > 0$ satisfies (10.3). As a consequence, if $r^2 = d + 2$, then $T_\infty(p_n||\varphi) \to 0$ as $n \to \infty$.*

The assumption $r^2 = d+2$ corresponds to the requirement that $\mathbb{E}\,|X|^2 = d$ or $\mathbb{E} X_1^2 = 1$ for the random vector $X = (X_1, \dots, X_d)$ uniformly distributed in $B(r)$. This is equivalent to the property that X has identity covariance matrix.

Corollary 10.5. *The same assertion holds for the uniform distribution on the Euclidean sphere in $\mathbb{R}^d$ with center at the origin and radius $r = \sqrt{d}$.*

Note that the uniform distribution on the Euclidean sphere is not absolutely continuous. But its n-th convolution power for large n has a bounded, compactly supported density, so that $T_\infty(p_n||\varphi) < \infty$ for some $n = n_0$.

Introduce one dimensional probability densities

$$q_\alpha(x) = \frac{1}{c_\alpha}\,(1-x^2)^{\alpha-1}, \quad |x| < 1,$$

with parameter $\alpha > 0$, where (using the usual gamma-function)

$$\begin{aligned} c_\alpha &= \int_{-1}^{1} (1-x^2)^{\alpha-1}\,dx \\ &= \int_0^1 y^{-1/2}\,(1-y)^{\alpha-1}\,dy \;=\; \frac{\Gamma(\frac{1}{2})\Gamma(\alpha)}{\Gamma(\alpha+\frac{1}{2})} \end{aligned} \tag{10.5}$$

is the normalizing constant. For the proof of Corollaries 10.4-10.5, we need:

Lemma 10.6. *For any $\alpha > 0$, the distribution μ_α with density q_α is strictly sub-Gaussian. Moreover, the separation property (10.2) holds true.*

Proof. Expanding the cosh-function in power series, for the random variable ξ_α with density q_α the Laplace transform is given by

$$\begin{aligned} L_\alpha(t) &= \frac{1}{c_\alpha}\int_{-1}^{1} e^{tx}\,(1-x^2)^{\alpha-1}\,dx \;=\; \frac{2}{c_\alpha}\int_0^1 \cosh(tx)\,(1-x^2)^{\alpha-1}\,dx \\ &= \frac{2}{c_\alpha}\sum_{n=0}^\infty \frac{t^{2n}}{(2n)!}\int_0^1 x^{2n}\,(1-x^2)^{\alpha-1}\,dx \\ &= \frac{1}{c_\alpha}\sum_{n=0}^\infty \frac{t^{2n}}{(2n)!}\int_0^1 y^{n-\frac{1}{2}}\,(1-y)^{\alpha-1}\,dx \\ &= \frac{1}{\sqrt{\pi}}\sum_{n=0}^\infty \frac{\Gamma(s)}{\Gamma(n+s)}\,\frac{\Gamma(n+\frac{1}{2})}{(2n)!}\,t^{2n}, \end{aligned}$$

where we used (10.5) in the last step together with notation $s = \alpha + \frac{1}{2}$. From this expansion we find that

$$\sigma_\alpha^2 = \mathrm{Var}(\xi_\alpha) = \frac{1}{2\alpha+1} = \frac{1}{2s} \quad (\sigma_\alpha > 0).$$

Hence

$$L_\alpha(t/\sigma_\alpha) = \frac{1}{\sqrt{\pi}}\sum_{n=0}^\infty \frac{\Gamma(s)s^n}{\Gamma(n+s)}\,\frac{\Gamma(n+\frac{1}{2})}{(2n)!}\,(2t^2)^n$$

$$= 1 + \frac{1}{2}\,t^2 + \frac{1}{\sqrt{\pi}} \sum_{n=2}^\infty \frac{s^{n-1}}{(s+1)\dots(s+n-1)}\, \frac{\Gamma(n+\frac{1}{2})}{(2n)!}\,(2t^2)^n. \quad (10.6)$$

The fraction inside the sum represents an increasing function in s, and letting $s \to \infty$, we get

$$L_\alpha(t/\sigma_\alpha) \le \frac{1}{\sqrt{\pi}} \sum_{n=0}^\infty \frac{\Gamma(n+\frac{1}{2})}{(2n)!}\,(2t^2)^n = \sum_{n=0}^\infty \frac{1}{n!\,2^n}\,t^{2n} = e^{t^2/2}.$$

Thus, the Laplace transform of $X = \xi_\alpha/\sigma_\alpha$ is bounded by the Laplace transform of the standard normal law μ, which proves the strict sub-Gaussianity (this is consistent with the property that $\mu_\alpha \to \mu$ weakly as $\alpha \to \infty$).

To prove the refining property (10.2) for the random variable X, let us return to (10.6) and use the bound

$$\frac{s^{n-1}}{(s+1)\dots(s+n-1)} \le \beta^{n-1}, \quad \beta = \frac{s}{s+1}.$$

Then we get

$$L_\alpha(t/\sigma_\alpha) \le 1 + \frac{1}{\beta\sqrt{\pi}} \sum_{n=1}^\infty \frac{\Gamma(n+\frac{1}{2})}{(2n)!}\,(2\beta t^2)^n$$
$$= 1 + \frac{1}{\beta} \sum_{n=1}^\infty \frac{\beta^n}{n!\,2^n}\,t^{2n} \ = \ 1 + \frac{e^{\beta t^2/2} - 1}{\beta}.$$

Since $\beta < 1$, it is now easy to check that the last expression can be bounded by $e^{ct^2/2}$ for all $|t| \ge t_0$ with some constant $c \in (0,1)$ depending on β and $t_0 > 0$. $\square$

Proof of Corollaries 10.4-10.5. Without loss of generality, let $r = 1$. If the random vector $X = (X_1, \dots, X_d)$ is uniformly distributed in the unit ball $B(1)$, the distribution of X_1 has density q_α with $\alpha = (d+1)/2$. Similarly, if X is uniformly distributed in the unit sphere, the distribution of X_1 has density q_α with $\alpha = (d-1)/2$. It remains to apply Lemma 10.6 and Proposition 10.3. $\square$

11. Laplace transforms with periodic components

In order to describe examples illustrating Corollary 1.4, let us start with the following definition. We write $h = (h_1, \dots, h_d) \ge 0$, if all $h_j \ge 0$.

Definition. We say that the distribution μ of a random vector X in $\mathbb{R}^d$ is periodic with respect to the standard normal law, with period $h \ge 0$ ($h \ne 0$), if it has a density $p(x)$ such that the function

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

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

Here, q represents the density of μ with respect to the standard Gaussian measure γ on $\mathbb{R}^d$. We denote the class of all such distributions by $\mathfrak{F}_h$, and say that X belongs to $\mathfrak{F}_h$. In dimension $d = 1$, this class was studied in [7], and here we extend a number of one dimensional observations to higher dimensions. For this aim, introduce the componentwise multiplication of vectors $xy = (x_1y_1, \dots, x_dy_d)$, where $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d)$ are points in $\mathbb{R}^d$.

Proposition 11.1. *If X belongs to the class $\mathfrak{F}_h$, then for all $m \in \mathbb{Z}^d$,*

$$\mathbb{E}\, e^{\langle mh, X\rangle} = e^{|mh|^2/2}. \tag{11.1}$$

In particular, the random vector X is sub-Gaussian.

Proof. By the periodicity, $q(x - mh) = q(x)$ for all $x \in \mathbb{R}^d$ and $m \in \mathbb{Z}^d$. Hence, the random vector $X + mh$ has density

$$\begin{aligned} p(x - mh) &= q(x - mh)\varphi(x - mh) \\ &= q(x)\varphi(x)\, e^{\langle mh, x\rangle - \frac{1}{2}\,|mh|^2} \;=\; p(x)\, e^{\langle mh, x\rangle - \frac{1}{2}\,|mh|^2}. \end{aligned}$$

It remains to integrate this equality over the variable x, which leads to (11.1).

Next, starting from (11.1), it is easy to see that $\mathbb{E}\, e^{c|X|^2} < \infty$ for some $c > 0$. □

As a consequence, the Laplace transform $L(t) = \mathbb{E}\, e^{\langle t, X\rangle}$, $t \in \mathbb{R}^d$, is finite and may be extended to the d-dimensional complex space $\mathbb{C}^d$ as an entire function $L(z) = L(z_1, \dots, z_d)$. By saying “entire”, it is meant that a given function on $\mathbb{C}^d$ is entire with respect to every complex coordinate $z_j = t_j + iy_j$ ($t_j, y_j \in \mathbb{R}$) and is C^∞-smooth as a function of $2d$ real variables $t_1, y_1, \dots, t_d, y_d$.

This property of $L(z)$ may be further refined.

Proposition 11.2. *If X belongs to $\mathfrak{F}_h$, then its Laplace transform is an entire function of order 2 (with respect to every complex coordinate). Moreover,*

$$|L(z)| \le e^{|t|^2 + |h|^2}, \quad z = t + iy \in \mathbb{C}^d. \tag{11.2}$$

Proof. Since $\mathrm{Re}\,\langle z, X\rangle = \langle t, X\rangle$ for $z = t + iy$, we have $|L(z)| \le L(t)$. Hence, one may assume that $y = 0$ in (11.2).

We employ the convexity of the function $K(t) = \log L(t)$. For simplicity, suppose that $t = (t_1, \dots, t_d) \in \mathbb{R}^d_+$. Take an integral vector $m = (m_1, \dots, m_d)$ with positive components such that $(m_j - 1)h_j \le t_j < m_j h_j$ for all $j \le d$. Since t lies in the cube with sides $[0, m_j h_j]$, this point may be written as a convex mixture of vertices of the cube

$$t = \sum_{\varepsilon} a_\varepsilon v_\varepsilon, \quad a_\varepsilon \ge 0, \quad \sum_{\varepsilon} a_\varepsilon = 1.$$

Here the vertex $v_\varepsilon = (\varepsilon_1 m_1 h_1, \dots, \varepsilon_d m_d h_d)$ is parametrized by the tuple $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$ with $\varepsilon_j = 0$ or $\varepsilon_j = 1$ for each $j \le d$. By Jensen's inequality and (11.1),

$$\begin{aligned} K(t) \le \sum_{\varepsilon} a_\varepsilon K(v_\varepsilon) \ &= \ \frac{1}{2} \sum_{\varepsilon} a_\varepsilon \Big[\sum_{j=1}^d \varepsilon_j^2 m_j^2 h_j^2 \Big] \\ &\le \frac{1}{2} \sum_{j=1}^d m_j^2 h_j^2 \ \le \ \frac{1}{2} \sum_{j=1}^d (t_j + h_j)^2. \end{aligned}$$

Dropping the condition on the sign of t_j, more generally we obtain that

$$K(t) \le \frac{1}{2} \sum_{j=1}^d (|t_j| + h_j)^2 \le \sum_{j=1}^d (t_j^2 + h_j^2) = |t|^2 + |h|^2.$$

Thus, we obtain (11.2) which shows that $L(z)$ is an entire function of order at most 2. On the other hand, by (11.1), it is an entire function of order at least 2. $\square$

Let us also mention the periodicity property for convolutions.

Proposition 11.3. *If X belongs to $\mathfrak{F}_h$, then Z_n belongs to $\mathfrak{F}_{h\sqrt{n}}$.*

Here as before, $Z_n = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$ denotes the normalized sum of independent copies X_k of the random vector X. The proof is similar to the proof of Proposition 10.4 from [7] for the one dimensional case, so we omit it.

12. Characterizations of periodicity in terms of Laplace transform

Fix $h = (h_1, \dots, h_d) \in \mathbb{R}^d$, $h \neq 0$, for simplicity with $h_j \ge 0$. Here we prove:

Proposition 12.1. *If X belongs to $\mathfrak{F}_h$, then the function*

$$\Psi(t) = L(t)\, e^{-|t|^2/2}, \quad t \in \mathbb{R}^d, \tag{12.1}$$

is periodic with period h. It can be extended to the complex space $\mathbb{C}^d$ as an entire function of order at most 2. Conversely, if the function $\Psi(t)$ for a sub-Gaussian random vector X is h-periodic, then X belongs to $\mathfrak{F}_h$, as long as the characteristic function $f(t)$ of X is integrable on $\mathbb{R}^d$.

The last claim is based on the following general observation.

Lemma 12.2. *Let $f(z) = \mathbb{E}\, e^{i\langle z,X\rangle}$ $(z \in \mathbb{C}^d)$ be the characteristic function of a sub-Gaussian random vector X. If $f(t)$ and $f(t+ih)$ are integrable in $t \in \mathbb{R}^d$, then*

$$\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f(t)\, dt = \int_{\mathbb{R}^d} e^{-i\langle t+ih,x\rangle} f(t+ih)\, dt. \tag{12.2}$$

Proof. Write $z = (z_1, \dots, z_d)$ with $z_j \in \mathbb{C}$. By assumption, $f(z) = f(z_1, \dots, z_d)$ is well-defined and represents an entire function. In addition,

$$|f(z)| \le \mathbb{E}\, |e^{i\langle z,X\rangle}| = \mathbb{E}\, e^{-\langle y,X\rangle} \le \mathbb{E}\, e^{|y|\,|X|}, \quad z = t + iy,\ t, y \in \mathbb{R}^d. \tag{12.3}$$

One may rewrite the first integral in (12.2) in a different way using contour integration. For this aim, let $X_\varepsilon = X + \varepsilon Z$, where $\varepsilon > 0$ and Z is a standard normal random vector in $\mathbb{R}^d$ independent of X. The random vector X_ε is also sub-Gaussian and has an entire characteristic function

$$f_\varepsilon(z) = f(z)\, e^{-\varepsilon^2 z^2/2}, \quad z \in \mathbb{C}^d,$$

where $z^2 = z_1^2 + \dots + z_d^2$. Moreover, by (12.3), for all $t \in \mathbb{R}^d$,

$$|f_\varepsilon(t + iy)| \le K e^{-\varepsilon^2 |t|^2/2}, \quad y = (y_1, \dots, y_j),\ |y_j| \le h_j, \tag{12.4}$$

with some constant K which does not depend on t.

Putting $x = (x_1, \dots, x_d)$, $t = (t_1, \dots, t_d)$, let us integrate the left integrand in (12.2) with respect to t_1, keeping the remaining variables $t_2, \dots, t_d$ fixed. Given $T > 0$, consider the rectangle contour with sides

$$\begin{aligned} &C_1 = [-T, T], && C_2 = [T, T + ih_1], \\ &C_3 = [T + ih_1, -T + ih_1], && C_4 = [-T + ih_1, -T], \end{aligned}$$

so that to apply Cauchy's theorem which yields

$$\int_{C_1} e^{-iz_1x_1} f_\varepsilon(z)\, dz_1 + \int_{C_2} e^{-iz_1x_1} f(z)\, dz_1 + \int_{C_3} e^{-iz_1x_1} f_\varepsilon(z)\, dz_1 + \int_{C_4} e^{-iz_1x_1} f_\varepsilon(z)\, dz_1 = 0.$$

For points $z_1 = t_1 + iy_1$ on the contour, $|e^{-iz_1x_1}| = e^{x_1y_1} \le e^{|x|h_1}$. Hence, by the decay property (12.4), the integrals over C_2 and C_4 are vanishing as $T \to \infty$, and

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{-it_1x_1} f_\varepsilon(t)\,dt_1 &= \lim_{T\to\infty}\int_{C_1} e^{-iz_1x_1} f_\varepsilon(z)\,dz_1 \\
&= -\lim_{T\to\infty}\int_{C_3} e^{-iz_1x_1} f_\varepsilon(z)\,dz_1 \\
&= \int_{-\infty}^{\infty} e^{-i(t_1+ih_1)x_1} f_\varepsilon(t_1+ih_1,t_2,\dots,t_d)\,dt_1.
\end{aligned}
$$

Thus,

$$
\int_{-\infty}^{\infty} e^{-it_1x_1} f_\varepsilon(t)\,dt_1 = \int_{-\infty}^{\infty} e^{-i(t_1+ih_1)x_1} f_\varepsilon(t_1+ih_1,t_2,\dots,t_d)\,dt_1. \tag{12.5}
$$

Turning to the next variable t_2, first note that the function

$$
g(z_2,\dots,z_d) = \int_{-\infty}^{\infty} e^{-i(t_1+ih_1)x_1} f_\varepsilon(t_1+ih_1,z_2,z_3,\dots,z_d)\,dt_1, \quad z_j = t_j + iy_j \in \mathbb{C},
$$

is entire and admits a similar bound as (12.4)

$$
|g(z_2,\dots,z_d)| \le K e^{-\varepsilon^2(t_2^2+\dots+t_d^2)/2}, \quad |y_j| \le h_j.
$$

Hence, one may perform the contour integration like in the previous step leading to

$$
\int_{-\infty}^{\infty} e^{-it_2x_2}\, g(t_2,t_3,\dots,t_d)\,dt_2 = \int_{-\infty}^{\infty} e^{-i(t_2+ih_2)x_2}\, g(t_2+ih_2,t_3,\dots,t_d)\,dt_2.
$$

By the Fubini theorem and using (12.5), we then get

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it_1x_1-it_2x_2} f_\varepsilon(t)\,dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(t_1+ih_1)x_1-i(t_2+ih_2)x_2} f_\varepsilon(t_1+ih_1,t_2+ih_2,t_3,\dots,t_d)\,dt_1dt_2.
\end{aligned}
$$

Repeating the process, we will arrive at the equality for the d-dimensional integrals

$$
\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f_\varepsilon(t)\,dt = \int_{\mathbb{R}^d} e^{-i\langle t+ih,x\rangle} f_\varepsilon(t+ih)\,dt.
$$

It remains to let $\varepsilon \to 0$ and make use of the Lebesgue dominated convergence theorem, which provides the desired equality in (12.2). $\square$

Proof of Proposition 12.1. By periodicity of q, changing the variable $x = y + h$, we have, for any $t \in \mathbb{R}^d$,

$$\begin{aligned} L(t+h) &= \int_{\mathbb{R}^d} e^{\langle t+h,x\rangle}\, q(x)\,\varphi(x)\,dx \\ &= \int_{\mathbb{R}^d} e^{\langle t+h,y+h\rangle}\, q(y+h)\,\varphi(y+h)\,dy \\ &= \int_{\mathbb{R}^d} e^{\langle t+h,y+h\rangle}\, q(y)\,\varphi(y)\,e^{-\langle y,h\rangle - |h|^2/2}\,dy \;=\; L(t)\,e^{\langle t,h\rangle + |h|^2/2}. \end{aligned}$$

Hence

$$L(t+h)\,e^{-|t+h|^2/2} = L(t)\,e^{-|t|^2/2},$$

which is the first claim of the proposition. Since $L(z)$ is an entire function of order 2, the formula (12.1) admits a natural extension to $\mathbb{C}^d$

$$\Psi(z) = L(z)\,e^{-\frac{1}{2}\,(z_1^2+\dots+z_d^2)}, \quad z = (z_1,\dots,z_d) \in \mathbb{C}^d,$$

which is an entire function with respect to every component z_j of order at most 2. By analyticity and periodicity on $\mathbb{R}^d$,

$$\Psi(z+h) = \Psi(z) \quad \text{for all } z \in \mathbb{C}^d. \tag{12.6}$$

Turning to the last claim of the proposition, we need to prove the periodicity of the function $q(x) = p(x)/\varphi(x)$. The characteristic function of X admits an entire extension using the formula

$$f(z) = \mathbb{E}\,e^{i\langle z,X\rangle} = L(iz) = \Psi(iz)\,e^{-z^2/2}, \quad z \in \mathbb{C}^d,$$

where again $z^2 = z_1^2 + \dots + z_d^2$. Hence, by (12.6), $f(t+ih)\,e^{(t+ih)^2/2} = f(t)\,e^{t^2/2}$, i.e.

$$f(t+ih) = f(t)\,e^{-i\langle t,h\rangle + |h|^2/2} \quad \text{for all } t \in \mathbb{R}^d. \tag{12.7}$$

This identity also shows that, due to the integrability of $f(t)$, the function $f(t+ih)$ is integrable as well.

By the integrability of $f(t)$, the random vector X has a continuous density given by the Fourier inversion formula

$$p(x) = \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f(t)\,dt, \quad x \in \mathbb{R}^d.$$

This yields

$$q(x) = \frac{p(x)}{\varphi(x)} = \frac{1}{(2\pi)^{d/2}}\, e^{|x|^2/2} \int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f(t)\,dt$$

and

$$q(x+h) = \frac{1}{(2\pi)^{d/2}}\, e^{|x|^2/2}\, e^{\langle x,h\rangle + |h|^2/2} \int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle - i\langle t,h\rangle} f(t)\,dt.$$

Hence, we need to show that

$$\int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f(t)\,dt = e^{\langle x,h\rangle + |h|^2/2} \int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle - i\langle t,h\rangle} f(t)\,dt.$$

Here, the left integral may be rewritten according to Lemma 12.1, so that the above equality is restated as

$$\int\limits_{\mathbb{R}^d} e^{-i\langle t+ih,x\rangle} f(t+ih)\,dt = e^{\langle x,h\rangle + |h|^2/2} \int\limits_{\mathbb{R}^d} e^{-i\langle t,x\rangle - i\langle t,h\rangle} f(t)\,dt. \tag{12.8}$$

Moreover, by (12.7), the first integrand is equal to

$$e^{-i\langle t+ih,x\rangle}\ e^{-i\langle t,h\rangle + |h|^2/2} f(t).$$

This proves (12.8). $\square$

Remark. Since $f(t) = L(it) = \Psi(it)\, e^{-t^2/2}$, $t \in \mathbb{R}^d$, the integrability assumption in Proposition 12.1 is fulfilled, if $\Psi(z)$ has order smaller than 2.

13. Periodic components via trigonometric series

Proposition 12.1 is applicable to a variety of interesting examples including the underlying distributions whose Laplace transform has the form

$$L(t) = \Psi(t)\, e^{|t|^2/2}, \quad t \in \mathbb{R}^d, \tag{13.1}$$

where Ψ is a 2π-periodic function of the form

$$\Psi(t) = 1 - cP(t), \quad P(t) = \sum_{k \in \mathbb{Z}^d} c_k e^{i\langle k,t\rangle} \tag{13.2}$$

(for simplicity, in the sequel we say "2π-periodic" instead of "$(2\pi, \dots, 2\pi)$-periodic"). Here $c_k = a_k - ib_k$ are complex coefficients which are supposed to satisfy

$$\sum_{k\in\mathbb{Z}^d}^{\infty} e^{|k|^2/2}\,|c_k| < \infty, \tag{13.3}$$

and $c \in \mathbb{R}$ is a non-zero parameter. To ensure that $P(t)$ is real-valued, we assume that $a_{-k} = a_k$ and $b_{-k} = -b_k$ for all $k \in \mathbb{Z}^d$, in which case

$$P(t) = \sum_{k\in\mathbb{Z}^d} \big(a_k \cos\langle k, t\rangle + b_k \sin\langle k, t\rangle\big).$$

Any such function with real coefficients a_k and b_k can be written in the form (13.2).

Let us note that the function $\Psi(t)$ defined in the equality (13.1) for a sub-Gaussian random vector X is smooth with $\Psi(0) = L(0) = 1$. Differentiating (13.1), we see that X has mean zero and identity covariance matrix if and only if $\Psi'(0) = 0$ in vector sense and $\Psi''(0) = 0$ in matrix sense. For the Ψ-functions as in (13.2), this is equivalent to $P(0) = P'(0) = P''(0) = 0$, that is,

$$\sum_{k\in\mathbb{Z}^d} a_k = 0, \quad \sum_{k\in\mathbb{Z}^d} b_k k = 0, \quad \sum_{k=1}^{\infty} a_k\, k\otimes k = 0,$$

where $k \otimes k$ denotes the $d \times d$ matrix with entries $k_i k_j$, $1 \le i, j \le d$, $k = (k_1, \dots, k_d)$. All the series are well convergent due to the condition (13.3).

Proposition 13.1. *If $P(0) = P'(0) = P''(0) = 0$ and $|c|$ is small enough, then $L(t)$ represents the Laplace transform of a sub-Gaussian random vector X with mean zero, identity covariance matrix, and with density $p = q\varphi$, where q is a bounded, 2π-periodic function. This random vector is strictly sub-Gaussian, if $P(t) \ge 0$ for all $t \in \mathbb{R}^d$ and if $c > 0$ is small enough.*

Proof. The functions $u_\lambda(x) = \varphi(x)\cos\langle\lambda, x\rangle$ and $v_\lambda(x) = \varphi(x)\sin\langle\lambda, x\rangle$ with parameter $\lambda \in \mathbb{R}^d$ have respectively the Laplace transforms

$$\int_{\mathbb{R}^d} e^{\langle t,x\rangle} u_\lambda(x)\,dx = e^{-|\lambda|^2/2}\,\cos\langle\lambda, t\rangle\, e^{|t|^2/2},$$

$$\int_{\mathbb{R}^d} e^{\langle t,x\rangle} v_\lambda(x)\,dx = e^{-|\lambda|^2/2}\,\sin\langle\lambda, t\rangle\, e^{|t|^2/2}.$$

Define

$$q(x) = 1 - c\sum_{k\in\mathbb{Z}^d} e^{|k|^2/2}\big(a_k \cos\langle k, x\rangle + b_k \sin\langle k, x\rangle\big). \tag{13.4}$$

It is bounded due to the condition (13.3) and is non-negative for sufficiently small $|c|$, more precisely, if

$$\sum_{k\in\mathbb{Z}^d}^{\infty} e^{|k|^2/2}\,(|a_k| + |b_k|) \leq \frac{1}{|c|}.$$

Moreover, the Laplace transform of the function $p(x) = q(x)\varphi(x)$ is exactly

$$\int_{\mathbb{R}^d} e^{\langle t,x\rangle} p(x)\,dx = (1 - cP(t))\,e^{-|t|^2/2}, \quad t \in \mathbb{R}^d.$$

Recall that the requirement $P(0) = 0$ guarantees that $\int_{\mathbb{R}^d} p(x)\,dx = 1$, while the property that X has mean zero and identity covariance matrix is equivalent to $P'(0) = P''(0) = 0$.

Finally, if $P(t) \geq 0$ for all $t \in \mathbb{R}^d$ and $c > 0$ is small enough, then $0 < \Psi(t) \leq 1$, which means that X is strictly sub-Gaussian. $\square$

14. Examples involving trigonometric polynomials

Some specific examples in Proposition 13.1 are based on trigonometric polynomials. More precisely, suppose that the random vector X has the Laplace transform

$$L(t) = \Psi(t)\,e^{|t|^2/2}, \quad t \in \mathbb{R}^d, \tag{14.1}$$

with

$$\Psi(t) = 1 - cQ(t)^2, \quad Q(t) = \sum_{k\in\mathbb{Z}^d} \big(a_k \cos\langle k,t\rangle + b_k \sin\langle k,t\rangle\big), \tag{14.2}$$

where $c > 0$, assuming that the sum contains only finitely many real coefficients.

Corollary 14.1. *If $Q(0) = Q'(0) = 0$ and $c > 0$ is small enough, $L(t)$ represents the Laplace transform of a strictly sub-Gaussian random vector X with mean zero, identity covariance matrix, and with density $p = q\varphi$, where q is a bounded, 2π-periodic function.*

Let us complement this statement with the assertion about the central limit theorem for the distance

$$T_\infty(p_n||\varphi) = \sup_x \frac{p_n(x) - \varphi(x)}{\varphi(x)},$$

where p_n denotes the density of the normalized sum Z_n constructed for n independent copies of X.

Corollary 14.2. *Under the assumptions of Corollary 14.1, $T_\infty(p_n||\varphi) \to 0$ as $n \to \infty$, if and only if*

$$\forall\, t \in [0,2\pi]^d \quad \big[Q(t) = 0 \;\Rightarrow\; Q'(t) = 0\big]. \tag{14.3}$$

This is the case where $Q(t) \geq 0$ for all $t \in [0,2\pi]^d$.

Proof. We apply Proposition 13.1 with $P(t) = Q(t)^2$. Then $P(0) = Q(0)^2 = 0$ and $P'(0) = 2Q(0)\,Q'(0) = 0$. In addition,

$$P''(t) = 2\,Q(t)Q''(t) + 2\,Q'(t) \otimes Q'(t), \tag{14.4}$$

so that $P''(0) = 0$ as well. This proves Corollary 14.1.

For the assertion in Corollary 14.2, we appeal to Corollary 1.4 and simplify the implication $\Psi(t) = 0 \Rightarrow \Psi''(t) = 0$. Here, the hypothesis means that $Q(t) = 0$, in which case $\Psi''(t) = -cP''(t) = -2c\,Q'(t) \otimes Q'(t)$, according to (14.4). The latter matrix is zero, if and only if $Q'(t) = 0$. $\square$

Example 14.3. Consider the Laplace transform

$$L(t) = \big(1 - c\sin^{2m}(t_1 + \dots + t_d)\big)\, e^{|t|^2/2}, \quad t = (t_1, \dots, t_d) \in \mathbb{R}^d,$$

with a fixed integer $m \geq 2$, where $c > 0$ is small enough (depending on m and d). In this case, the conditions of Corollary 14.1 are fulfilled for

$$Q(t) = \sin^m(t_1 + \dots + t_d).$$

Since the condition (14.3) in Corollary 14.2 is fulfilled as well, we obtain the assertion about the CLT (although $Q(t)$ does not need be non-negative for odd values of m).

An interesting feature of this example is that $L(t) = e^{|t|^2/2}$ on countably many hyperplanes $t_1 + \dots + t_d = \pi l$, $l \in \mathbb{Z}$.

Example 14.4. Modifying the previous example, put

$$P(t) = Q(t)^2, \quad Q(t) = \big(1 - 4\sin^2(t_1 + \dots + t_d)\big)\sin^2(t_1 + \dots + t_d).$$

In this case, $Q(0) = Q'(0) = 0$, so that the conditions of Corollary 14.1 are fulfilled.

We have $Q(t) = 0$ for $t_1 = \pi/6$, $t_j = 0$ for $j \geq 2$. At this point $\partial_{t_1} Q(t) \neq 0$, so that the condition (14.3) is not fulfilled. Hence, by Corollary 14.2, the CLT with respect to T_∞ does not hold in this example.

Acknowledgment

The research has been supported by the NSF grant DMS-2154001 and the GRF – SFB 1283/2 2021 – 317210226.

Data availability

No data was used for the research described in the article.

References

[1] S. Artstein, K.M. Ball, F. Barthe, A. Naor, On the rate of convergence in the entropic central limit theorem, Probab. Theory Relat. Fields 129 (3) (2004) 381–390.
[2] A.R. Barron, Entropy and the central limit theorem, Ann. Probab. 14 (1) (1986) 336–342.
[3] S.G. Bobkov, G.P. Chistyakov, Bounds for the maximum of the density of the sum of independent random variables (Russian) Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 408 Veroyatn. Stat. 18 (2012) 62–73, 324; transl. in J. Math. Sci. (N.Y.) 199 (2014), no. 2, 100–106.
[4] S.G. Bobkov, G.P. Chistyakov, F. Götze, Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem, Ann. Probab. 41 (4) (2013) 2479–2512.
[5] S.G. Bobkov, G.P. Chistyakov, F. Götze, Berry-Esseen bounds in the entropic central limit theorem, Probab. Theory Relat. Fields 159 (3–4) (2014) 435–478.
[6] S.G. Bobkov, G.P. Chistyakov, F. Götze, Rényi divergence and the central limit theorem, Ann. Probab. 47 (1) (2019) 270–323.
[7] S.G. Bobkov, G.P. Chistyakov, F. Götze, Strictly subgaussian distributions, Electron. J. Probab. 29 (62) (2024) 1–28.
[8] S.G. Bobkov, F. Götze, Central limit theorem for Rényi divergence of infinite order, Ann. Probab. 53 (2) (2025) 453–477.
[9] S.G. Bobkov, F. Götze, Berry-Esseen bounds in the local limit theorems, Lith. Math. J. 65 (1) (2025) 50–66.
[10] S.G. Bobkov, M. Madiman, The entropy per coordinate of a random vector is highly constrained under convexity conditions, IEEE Trans. Inf. Theory 57 (8) (2011) 4940–4954.
[11] V.V. Buldygin, Yu.V. Kozachenko, Metric Characterization of Random Variables and Random Processes, Translated from the 1998 Russian original by V. Zaiats Transl. Math. Monogr., vol. 188, American Mathematical Society, Providence, RI, 2000, xii+257 pp.
[12] H.E. Daniels, Saddlepoint approximations in statistics, Ann. Math. Stat. 25 (1954) 631–650.
[13] F. Esscher, On the probability function in the collective theory of risk, Skand. Aktuarietidskr. 15 (3) (1932) 175–195.
[14] A.I. Khinchin, Mathematical Foundations of Statistical Mechanics, Translated by G. Gamow. Dover Publications, Inc., New York, N.Y., 1949, viii+179 pp.
[15] J.V. Linnik, An information-theoretic proof of the central limit theorem with Lindeberg conditions, Theory Probab. Appl. 4 (1959) 288–299.
[16] J.J. Marcinkiewicz, Sur une propriété de la loi de Gauss (French), Math. Z. 44 (1) (1939) 612–618.
[17] C.M. Newman, Inequalities for Ising models and field theories which obey the Lee-Yang theorem, Commun. Math. Phys. 41 (1975) 1–9.
[18] Yu.V. Prokhorov, A local theorem for densities (Russian) Dokl. Akad. Nauk SSSR 83 (1952) 797–800.
[19] A.Yu. Zaitsev, Estimates for the Lévy-Prokhorov distance in the multidimensional central limit theorem for random vectors with finite exponential moments (Russian) Teor. Veroâtn. Primen. 31 (2) (1986) 246–265, English translation: Theory Probab. Appl. 31:2 (1986), 203–220.