RÉNYI DIVERGENCE AND THE CENTRAL LIMIT THEOREM

S. G. BOBKOV[1,4], G. P. CHISTYAKOV[2,4], AND F. GÖTZE[3,4]

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

1. **Introduction**

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

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

represents an important measure of deviation of the distribution P of X from the distribution Q of Z, which has been frequently used especially in Statistics and Information Theory (cf. e.g. [Le], [L-V], [V]). This strong distance-like quantity may be included in the hierarchy of Rényi divergences (relative α-entropies)

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

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

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

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

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

1991 *Mathematics Subject Classification.* Primary 60E, 60F.

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

1) School of Mathematics, University of Minnesota, USA.

2) Faculty of Mathematics, University of Bielefeld, Germany.

3) Faculty of Mathematics, University of Bielefeld, Germany.

4) Research was supported by SFB 1283, Humboldt Foundation, and NSF grant DMS-1612961.

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

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

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

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

Different aspects of such strong CLT's, including the non-i.i.d. situation and the problem of rates or Berry-Esseen bounds, were studied by many authors, and we refer to [Li], [S-M], [A-B-B-N1], [B-J], [J], [B-C-G2-3]. There is also an increasing interest to other limit theorems in other strong distances such as the relative Fisher information, cf. [B-C-G4], [B-C-K], [To1-2].

As for convergence in the CLT with respect to D_α with $\alpha > 1$, not much is known so far. This case seems to be quite different in nature, and here the distance restricts the range of applicability of the CLT quite substantially. When focusing on the particular value $\alpha = 2$, we are concerned with the behavior of the quantity

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

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

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

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

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

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

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

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

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

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

For $s = 3$ (1.2) becomes

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 1.3. *Suppose that $D_\alpha(Z_n||Z)$ is finite for some $n = n_0$, and let the property* (1.4) *be fulfilled. Then, for all n large enough and for all $x \in \mathbb{R}^d$,*

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

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

Thus, (1.5) is implied by the convergence $D_\alpha(Z_n||Z) \rightarrow 0$. Non-uniform bounds in the normal approximation have been intensively studied in the literature, cf. [Pe1-2], [I-L], [A1-2]. However, existing results start with weaker hypotheses (e.g. moment assumptions) and either provide a polynomial error of approximation with respect to x (such as $\frac{1}{1+|x|^3}$), or deal with narrow zones contained in regions $|x| = o(\sqrt{n})$.

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

2. **Background on Rényi Divergence**

First let us briefly review some general properties of the Rényi divergences. More details can be found in van Erven and Harremoës [E-H]; cf. also [Le], [S], [G-S].

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

Definition 2.1. Let $\alpha \in (0, \infty)$, $\alpha \neq 1$. The Rényi divergence of P from Q and the divergence power or relative Tsallis entropy of index α are the quantities

$$\begin{aligned} D_\alpha(X||Z) &= D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \int \Big(\frac{p}{q}\Big)^\alpha q \, d\mu, \\ T_\alpha(X||Z) &= T_\alpha(P||Q) = \frac{1}{\alpha - 1} \bigg[\int \Big(\frac{p}{q}\Big)^\alpha q \, d\mu - 1 \bigg]. \end{aligned}$$

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

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

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

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

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

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

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

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

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

This extends the classical Pinsker inequality for the Kulback-Leibler distance (when $\alpha = 1$), with best constant due to Csiszár, cf. [Pi], [Cs].

The following general property is important for comparing the Rényi divergences.

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

The monotonicity of D_α is discussed in [E-H], Theorem 3. As for T_α, let $0 < \alpha < \beta$, $\alpha, \beta \neq 1$. The functions $c \rightarrow e^{ct_0} - 1$ $(t_0 \geq 0)$ and $t \rightarrow \frac{e^{ct}-1}{t}$ are non-decreasing in $c \geq 0$ and $t > 1$, resp. Hence, in case $\alpha > 1$, we get, using monotonicity of D_α,

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

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

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

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

$$\begin{aligned} D_0(P||Q) &= \lim_{\alpha \downarrow 0} D_\alpha(P||Q), \qquad D_\infty(P||Q) = \lim_{\alpha \rightarrow \infty} D_\alpha(P||Q), \\ D_1(P||Q) &= \lim_{\alpha \uparrow 1} D_\alpha(P||Q). \end{aligned}$$

As easy to check, $D_0(P||Q) = -\log\, Q\{p(x) > 0\}$ and $D_\infty(P||Q) = \log\, {\rm ess\,sup}_P\, \frac{p(x)}{q(x)}$.

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

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

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

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

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

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

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

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

Another important property of these distances is the contractivity under mappings.

Proposition 2.3. *For any measurable map S from Ω to any measurable space Ω',*

$$D_\alpha(S(X)||S(Z)) \leq D_\alpha(X||Z) \qquad (\alpha \geq 1). \tag{2.1}$$

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

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

that is,

$$1 + (\alpha-1)\, T_\alpha(X||Z) = \sup_{\mathbb{E}\, \eta(Z)^\beta \leq 1} \big(\mathbb{E}\, \eta(X)\big)^\alpha, \tag{2.2}$$

where the supremum is taken over all measurable $\eta : \Omega \rightarrow \mathbb{R}_+$ such that $\mathbb{E}\, \eta(Z)^\beta \leq 1$. Similarly,

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

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

The property (2.1) is closely related to the so called data processing inequality in Information Theory, namely

$$D_\alpha(P_{\mathfrak{A}}||Q_{\mathfrak{A}}) \leq D_\alpha(P||Q),$$

where $P_{\mathfrak{A}}$ and $Q_{\mathfrak{A}}$ denote restrictions of the measures P and Q to an arbitrary σ-subalgebra $\mathfrak{A}$ in Ω (cf. [E-H], Theorem 1).

3. Pearson-Vajda Distances

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

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

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

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

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

For our further purpose, it will be useful to relate the Rényi divergence power T_α to χ_α. Both quantities are metrically equivalent, as seen by the following elementary observation.

Proposition 3.2. *Put $T_\alpha = T_\alpha(P||Q)$ and $\chi_\alpha = \chi_\alpha(P||Q)$ for $\alpha > 1$. We always have*

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

and conversely,

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

Proof. To derive (3.1), we apply the triangle inequality in $L^\alpha(q^{1-\alpha}\,d\mu)$, which yields

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

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

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

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

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

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

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

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

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

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

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

Therefore,

$$\psi(1) \ge \frac{1}{2}\,\alpha(\alpha-1)\,\mathbb{E}\,\eta^2\Big(1+\frac{1}{3}\,\eta\Big)^{\alpha-2}.$$

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

$$\psi(1) \ge \frac{1}{2}\,\alpha(\alpha-1)\,\Big(\frac{3}{4}\Big)^{2-\alpha}\,\mathbb{E}\,\big(\eta^2\,1_A + \eta^\alpha 1_B\big).$$

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

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

By Hölder's inequality,

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

so $\mathbb{E}\,\eta^2\,1_A \ge (\mathbb{E}\,|\eta|^\alpha\,1_A)^{2/\alpha}$ and thus

$$\mathbb{E}\,\big(\eta^2\,1_A + \eta^\alpha\,1_B\big) \ge U = u_0^{2/\alpha} + u_1, \quad \text{where } \ u_0 = \mathbb{E}\,|\eta|^\alpha\,1_A, \quad u_1 = \mathbb{E}\,|\eta|^\alpha\,1_B.$$

If $u = u_0 + u_1 \le 1$, then $u^{\frac{2}{\alpha}} \le 2^{\frac{2}{\alpha}-1}\,(u_0^{\frac{2}{\alpha}} + u_1^{\frac{2}{\alpha}}) \le 2U$, by Jensen's inequality and since $u_1 \le 1$. In the case $u \ge 1$, we have $2U - u = 2u_0^{\frac{2}{\alpha}} - 2u_0 + u \ge 2u_0^{\frac{2}{\alpha}} - 2u_0 + 1$ which is positive for all $u_0 \ge 0$. So, $2U \ge \min(u, u^{2/\alpha})$ in both cases, that is,

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

As a result,

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

which yields the first inequality in (3.2).

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

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

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

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

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

4. **Basic Exponential Inequalities**

We now focus on the case, where $\Omega = \mathbb{R}$ is the real line with Lebesgue measure μ, and where $Z \sim N(0,1)$ is a standard normal random variable, i.e., with density

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Given a random variable X with density p, the Rényi divergence and the Tsallis distance of index $\alpha > 1$ with respect to Z are then given by the formulas

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

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

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

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

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

Proof. Let X have density p such that the integral $C = \big(\int_{-\infty}^{\infty} p(x)^\alpha\, \varphi(x)^{1-\alpha}\, dx\big)^{1/\alpha}$ is finite. By the Hölder inequality with dual exponents (β, α),

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

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

An alternative (although almost equivalent) variant of Proposition 4.1 is:

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

$$\mathbb{E}\, e^{tX} \leq C e^{\beta t^2/2} \tag{4.1}$$

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

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

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

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

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

Proof. Write $\mathbb{E}\, e^{tX} = \int_{-\infty}^{\infty} e^{tx}\, p(x)\, dx$ in terms of density p of X. Here, we split integration over $(0,\infty)$ into the two regions. Given $t > 0$, by the Hölder inequality,

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

where we used the monotonicity of $\beta tx - \frac{1}{2}\, x^2$ in the interval $0 \leq x \leq \beta t$. Similarly,

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

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

$$\mathbb{E}\, e^{tX}\, 1_{\{X>0\}}\, e^{-\frac{\beta t^2}{2}} \leq (2\pi)^{-1/(2\beta)} \big(1 + (\alpha-1) T_\alpha(X||Z)\big)^{1/\alpha} \Big(\frac{\beta t}{2}\Big)^{1/\beta} e^{-\frac{\beta t^2}{8}} + \delta(t) \to 0.$$

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

5. Laplace and Weierstrass Transforms

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence

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

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

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

Proof of Proposition 5.1. Let p be the density of X. The Weierstrass transform can be applied to the function

$$u(x) = \varphi(x)^{-\frac{1}{\beta}}\, p(x),$$

which has finite norm $\|u\|_\alpha = (1 + (\alpha - 1)\, T_\alpha)^{1/\alpha}$. Putting $\bar{x} = \frac{1}{n}\,(x_1 + \dots + x_n)$ and using dx for $dx_1 \dots dx_n$, the expectation we have to estimate is

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

where $Q_i = \sum_{j=1}^n (x_i - x_j)^2$. First, we apply Hölder's inequality and put $t = 2\beta$, to get

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

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

$$\Big(\int_{-\infty}^{\infty} (W_t u(x_1))^\gamma\, dx_1\Big)^{\frac{1}{\beta}}\, \|u\|_\alpha \,=\, \|W_t u\|_\gamma^{n-1}\, \|u\|_\alpha \,\leq\, \big((2\pi t)^{\frac{\alpha-\gamma}{2\gamma\alpha}}\, \beta^{\frac{\alpha-\gamma}{2\gamma\beta}}\, \|u\|_\alpha\big)^{n-1}\, \|u\|_\alpha.$$

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

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

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

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

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

we similarly obtain

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

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

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

This leads to

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

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

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

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

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

$$\Big|\mathbb{E}\, e^{\frac{1}{2\beta} Z_n^2} - \mathbb{E}\, e^{\frac{1}{2\beta} Z^2}\Big| \ \leq \ c_{n,\alpha} \sum_{k=0}^{n-1} \frac{n!}{k!\,(n-k)!}\, \|\varphi^{1/\alpha}\|_\alpha^k\, \|v\|_\alpha^{n-k} = c_{n,\alpha} \Big((1 + \|v\|_\alpha)^n - 1\Big).$$

□

6. Connections with Fourier Transform

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

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

In this case, necessary and sufficient conditions for finiteness of this divergence may be given in terms of the characteristic function

$$f(t) = \mathbb{E}\, e^{itX}, \qquad t \in \mathbb{R}.$$

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

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

Proof. By the very definition,

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

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

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

It remains to apply Plancherel's theorem. □

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

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

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

Proposition 6.2. *For any random variable X,*

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

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

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

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

are connected by the identity

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

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

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

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

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

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

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

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

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

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

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

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

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

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

thus proving (6.2).

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

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

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

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

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

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

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

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

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

7. Exponential Series

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

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

In particular, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$. Each H_k is a polynomial of degree k with integer coefficients. Depending on k being even or odd, H_k contains even resp. odd powers only. These polynomials may be defined explicitly via

$$H_k(x) = \mathbb{E}\,(x + iZ)^k, \quad Z \sim N(0,1).$$

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

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

Equivalently, the Hermite functions $\varphi_k = H_k\varphi$ form a complete orthogonal system in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$ with squares of the L^2-norms in this space given by $\int_{-\infty}^{\infty} \frac{\varphi_k(x)^2}{\varphi(x)}\,dx = k!$ Summarizing we have:

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

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

which converges in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$. The coefficients are given by $c_k = \int_{-\infty}^{\infty} u(x)\,H_k(x)\,dx$, and we have Parseval's identity

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

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

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

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

The integral condition of Proposition 7.3 is illustrated in [Cr] on the example of the Gaussian functions $u(x) = e^{-\lambda x^2}$ ($\lambda > 0$). In this case, the corresponding exponential series can be explicitly computed, and at $x = 0$ it is given by the series

$$\frac{1}{\sqrt{2\lambda}} \sum_{k=0}^{\infty} \frac{(2k)!}{(k!)^2\, 4^k} \Big(1 - \frac{1}{2\lambda}\Big)^k.$$

It is absolutely convergent for $\lambda > \frac{1}{4}$, simply convergent for $\lambda = \frac{1}{4}$ and divergent for $\lambda < \frac{1}{4}$.

8. Normal Moments

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

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

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

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

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

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

which we call the normal moments of X. In particular, $c_0 = 1$, $c_1 = \mathbb{E}X$, $c_2 = \mathbb{E}X^2 - 1$.

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

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

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

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

with $k \geq 3$ (where we use the branch of the logarithm determined by $\log f(0) = 0$).

Let us emphasize one simple algebraic property of normal moments.

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

a) $\gamma_j(X) = 0$ *for all* $j = 3, \dots, k-1$;

b) $\mathbb{E}H_j(X) = 0$ *for all* $j = 3, \dots, k-1$;

c) $\mathbb{E}X^j = \mathbb{E}Z^j$ *for all* $j = 3, \dots, k-1$.

In this case

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

Proof. Let us first describe the structure of the coefficients in (8.3) used for $N = k$. Repeated differentiation of $f(t)\, e^{t^2/2} = \mathbb{E}\, e^{it(X+iZ)}$ yields $\frac{d^j}{i^j\, dt^j}\, \big[f(t)\, e^{t^2/2}\big]\big|_{t=0} = \mathbb{E}\,(X+iZ)^j$. Hence, we get indeed $c_j = \mathbb{E}\, H_j(X)$ for all $j \leq k$.

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

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

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

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

In general, the moments of X may be expressed easily in terms of the normal moments. Indeed, the Chebyshev-Hermite polynomials have the generating function

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

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

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

Expanding e^{xz} into the power series and comparing the coefficients in front of z^k, we get

$$x^k = k! \sum_{j=0}^{[k/2]} \frac{1}{(k-2j)!\, j!\, 2^j}\, H_{k-2j}(x).$$

Hence, if $\mathbb{E}\,|X|^k < \infty$, then

$$\mathbb{E}X^k = k! \sum_{j=0}^{[k/2]} \frac{1}{(k-2j)!\, j!\, 2^j}\, \mathbb{E}\, H_{k-2j}(X). \tag{8.6}$$

Now, let us describe the connection between the normal moments and the χ^2-distance. The series in (8.3) is absolutely convergent as $N \to \infty$, when f is analytic in the complex plane. Hence we have the expansion

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

assuming condition (8.1). Moreover, using the Parseval identity in Proposition 7.1, we have

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

and hence arrive at the following relation:

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

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

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

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

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

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

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

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

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

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

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

N-th partial sum of that series,

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

It represents the Fourier transform of the function $p_N(x) = \varphi(x) \sum_{k=0}^N c_k\, \frac{H_k(x)}{k!}$ which is the N-th partial sum of the exponential series in (8.2). Since, by the assumption,

$$\sum_{k=0}^{\infty} \frac{c_k^2}{k!} < \infty,$$

p_N converge to some p in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$, by Proposition 7.1. In particular, p_N converge in $L^2(\mathbb{R}, dx)$, and by Plancherel's theorem, f_N also converge in $L^2(\mathbb{R}, dx)$ to the Fourier transform $\hat{p}$ of p. But $f_N(t) \to f(t)$ for all t, so $f(t) = \hat{p}(t)$ almost everywhere. Thus we conclude that f belongs to $L^2(\mathbb{R}, dx)$ and is equal to the Fourier transform of p. Hence, X has an absolutely continuous distribution, and p is density of X.

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

There is a natural generalization of the identity (8.9) for the random variables

$$X_t = \sqrt{t}\, X + \sqrt{1-t}\, Z, \qquad 0 \leq t \leq 1,$$

where $Z \sim N(0,1)$ is independent of X.

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

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

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

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

Proof. It is known, e.g., as a direct consequence of the identity $H_k(x) = \mathbb{E}\,(x + iZ)^k$, that the Hermite polynomials satisfy the binomial formula

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

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

9. **Behavior of Rényi divergence under Convolutions**

The obvious question, when describing convergence in the central limit theorem in the D_α-distance is, does it remain finite for sums of independent summands with finite D_α-distances? The answer is affirmative and is made precise in the following:

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

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

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

$$1 + (\alpha - 1)\,T_\alpha(aX + bY||Z) \ \leq\ \big(1 + (\alpha - 1)\,T_\alpha(X||Z)\big)\,\big(1 + (\alpha - 1)\,T_\alpha(Y||Z)\big). \quad (9.1)$$

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

$$D_\alpha(a_1 X_1 + \dots + a_n X_n||Z) \ \leq\ D_\alpha(X_1||Z) + \dots + D_\alpha(X_n||Z),$$

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

$$D(a_1 X_1 + \dots + a_n X_n||Z) \ \leq\ \max\{D(X_1||Z), \dots, D(X_n||Z)\},$$

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

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

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

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

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

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

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

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

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

By Cauchy's inequality,

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

This gives

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

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

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

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

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

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

One may also ask whether or not $\chi^2(aX+bY,Z)$ remains finite, when $\chi^2(X,Z)$ is finite, and Y is "small" enough. If p is density of X, the density of $aX+bY$ is given by

$$q(x) = \frac{1}{|a|}\,\mathbb{E}\,p\Big(\frac{x-bY}{a}\Big), \qquad x \in \mathbb{R},$$

which is a convex mixture of densities on the line. Applying Cauchy's inequality, we have

$$\frac{q(x)^2}{\varphi(x)} \ \le \ \frac{1}{a^2}\,\mathbb{E}\,\frac{p(\frac{x-bY}{a})^2}{\varphi(x)},$$

and using $(ax+by)^2 \le x^2+y^2$, we get an elementary bound

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

That is, we arrive at:

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

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

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

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

Example 9.3. Suppose that X has density of the form

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

where π is a probability measure on the positive half-axis. The existence of $\chi^2(X,Z)$ implies that $\sigma^2 < 2$ for π-almost all σ^2, i.e., π should be supported on the interval

$(0,2)$. Squaring (9.4) and integrating over x, we find that

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

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

$$\int_0^1\!\!\int_0^1 \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\,d\pi(\sigma_1^2)\,d\pi(\sigma_2^2)<\infty \ \text{ and } \ \int_1^2\!\!\int_1^2 \frac{1}{\sqrt{4-(\sigma_1^2+\sigma_2^2)}}\,d\pi(\sigma_1^2)\,d\pi(\sigma_2^2)<\infty.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 9.4. Consider a density of the form

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

where a_k is a normalizing constant, $k = n_0 - 1$, and let f_1 denote its Fourier transform (i.e., the characteristic function). Define the distribution of X via its characteristic function

$$f(t) = \alpha f_1(t) + (1-\alpha)\, \frac{\sin(\gamma t)}{\gamma t}$$

with a sufficiently small parameter $\alpha > 0$ and $\gamma = \sqrt{3\, \frac{1+\alpha f_1''(0)}{1-\alpha}}$. It is easy to check that $f''(0) = -1$, which guarantees that $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$. Furthermore, it is not difficult to show that the densities p_n of Z_n admit the two-sided bounds

$$\frac{b_n'}{1+|x|^{n/2k}}\, e^{-x^2/4} \le p_n(x) \le \frac{b_n''}{1+|x|^{n/2k}}\, e^{-x^2/4} \qquad (x \in \mathbb{R}),$$

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

10. Superadditivity of χ^2 with Respect to Marginals

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

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

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

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

where X is a random vector in $\mathbb{R}^d$ with density p (assuming that the above integrals exits). These functionals are known to be subadditive and superadditive with respect to the components: Writing $X = (X', X'')$ with $X' \in \mathbb{R}^{d_1}$, $X'' \in \mathbb{R}^{d_2}$ $(d_1 + d_2 = d)$, one always has

$$h(X) \le h(X') + h(X''), \qquad I(X) \ge I(X') + I(X'') \tag{10.1}$$

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

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

Hence, by (10.1), these information-theoretic distances are both superadditive, i.e.,

$$D(X||Z) \ge D(X'||Z') + D(X''||Z''), \qquad I(X||Z) \ge I(X'||Z') + I(X''||Z''),$$

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

We now establish a similar property for the χ^2-distance, which can be more conveniently stated in the setting of a Euclidean space H, say of dimension d, with norm $|\cdot|$ and inner product $\langle \cdot, \cdot \rangle$. If X is a random vector in H with density p, and $Z \sim N(0, \mathrm{I}_d)$ is a normal random vector with mean zero and an identity covariance operator I_d, then (according to the abstract definition),

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

where $\varphi(x) = \frac{1}{(2\pi)^{d/2}}\, e^{-|x|^2/2}$ $(x \in H)$ is the density of Z.

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

$$\chi^2(X,Z) \ \geq\ \chi^2(X',Z') + \chi^2(X'',Z''), \tag{10.2}$$

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

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

Proof. Let $H = \mathbb{R}^d$ and $X = (\xi_1, \dots, \xi_d)$. Note that $\chi^2(X,Z)$ is invariant under orthogonal transformations U of the space, i.e., $\chi^2(U(X), Z) = \chi^2(X,Z)$. Hence, without loss of generality, one may assume that $X' = (\xi_1, \dots, \xi_{d_1})$ and $X'' = (\xi_{d_1+1}, \dots, \xi_d)$. Moreover, to simplify the argument (notationally), let $d_1 = d_2 = 1$.

The finiteness of $\chi^2(X,Z)$ means that the random vector $X = (\xi_1, \xi_2)$ has density $p = p(x_1,x_2)$ such that

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

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

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

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

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

Moreover, we have Parseval's equality

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

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

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

Hence, by Proposition 8.1,

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

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

11. **Asymptotic Expansions and Lower Bounds**

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

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

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

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

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

with

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

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

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

with a fixed integer $s \geq 2$, inequality (11.1) can be reversed up to an error term of order $o(n^{-(s-1)})$. This reduces our task to the study of the asymptotic behavior of the integrals $I(M_n(s))$, using the following result due to Petrov (cf. [Pe1-2], [B-C-G1]).

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

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

In this formula

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

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

For example, for $k = 3$ (11.3) yields

$$p_n(x) = \varphi(x) + \frac{\gamma_3}{3!\sqrt{n}}\, H_3(x)\varphi(x) + o\Big(\frac{1}{\sqrt{n}}\Big)\frac{1}{1+|x|^3},$$

where $\gamma_3 = \mathbb{E}X^3$ and $H_3(x) = x^3 - 3x$. More generally, if $\gamma_3 = \dots = \gamma_{k-1} = 0$, i.e., the first $k-1$ moments of X are the same as for a standard normal law, then (11.3) simplifies to

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

This local limit theorem may be used to derive the following expansion. Here and in the sequel, we use the standard notation $(\alpha)_m = \alpha(\alpha-1)\dots(\alpha-m+1)$.

Lemma 11.2. *Under the assumptions of Proposition* 11.1 *with* $k = 2s$ $(s \geq 2)$,

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

with

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

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

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

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

Proof. The representation (11.3) with $k = 2s$ may be written as

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

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

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

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

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

so that

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

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

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

with $\sup_x |C_n(x)| \leq C$ (where C is a constant). Thus, integration in (11.9) leads to

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

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

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

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

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

Now, using a multinomial formula, write

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

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

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

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

$$\int_{-\infty}^{\infty} H_{\nu_1}(x) \dots H_{\nu_k}(x) \, \varphi(x) \, dx = 0 \qquad (\nu_1 + \dots + \nu_k \text{ is odd}).$$

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

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

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

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

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

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

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

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

with coefficients given in (11.6). *In particular, if $\gamma_j = 0$ for $j = 3, \dots, s-1$, then*

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

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

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

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

Combining the lower bound (11.2) with the upper bound (9.2) yields:

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

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

12. Necessity Part in Theorem 1.2 ($d = 1$)

Again, let $X, X_1, X_2, \dots$ be i.i.d. random variables with characteristic function $f(t) = \mathbb{E}\, e^{itX}$, and let $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$. The necessity part in Theorem 1.2 does not require any moment assumptions on the mean and variance. As a preliminary step, the next lemma provides a subgaussian bound on the Laplace transform $f(iy) =$

$\mathbb{E}\, e^{-yX}$ subject to the sublinear growth of $D_\alpha(Z_n||Z)$. Recall that $\alpha > 1$, and we denote its conjugate value by $\beta = \alpha/(\alpha-1)$.

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

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

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

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

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

$$f(iy) \le \exp\Big\{\frac{1}{\alpha n}\, \log\big(1+(\alpha-1)\, T_\alpha\big)\Big\}\, e^{\beta y^2/2}.$$

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

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

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

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

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

Proof. Since $f(\frac{t}{\sqrt{n}})^n$ is the characteristic function of Z_n, the integral in (12.2) is

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

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

$$\big|\mathbb{E}\, e^{\frac{1}{2\beta}\, Z_{2nk}^2} - \mathbb{E}\, e^{\frac{1}{2\beta}\, Z^2}\big| \ \le\ 4^k\, \Big(\big(1+\chi_\alpha(Z_n||Z)^{1/\alpha}\big)^{2k} - 1\Big), \qquad Z \sim N(0,1).$$

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

Proof of the necessity part in Theorem 1.2 for $d = 1$. Assume that $D_\alpha(Z_n||Z) \to 0$ as $n \to \infty$ and fix an integer $k \ge \alpha/2$. Given a fixed number

$\delta > 0$, let us decompose

$$\int_{-\infty}^{\infty} f\big(iy/\sqrt{nk}\big)^{2nk} e^{-\beta y^2}\,dy \ = \ I_1 + I_2$$
$$= \ \bigg(\int_{|y|\le\delta\sqrt{nk}} + \int_{|y|>\delta\sqrt{nk}}\bigg) f\big(iy/\sqrt{nk}\big)^{2nk} e^{-\beta y^2}\,dy. \qquad (12.3)$$

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

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

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

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

where $\theta = \theta(y)$ is a quantity such that $|\theta| \le C$. Assuming that δ is small enough, e.g. $\delta \le \min\{r, (\beta-1)/(2C\sqrt{k})\}$, this relation allows us to rewrite the integral I_1 as

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

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

$$\int_{|y|\le\delta\sqrt{nk}} \big|e^{-(\beta-1)\,y^2+\theta y^3/\sqrt{n}} - e^{-(\beta-1)\,y^2}\big|\,dy \ \le \int_{|y|\le\delta\sqrt{nk}} \frac{C|y|^3}{\sqrt{n}}\, e^{-(\beta-1)\,y^2 + \frac{C|y|^3}{\sqrt{n}}}\,dy$$
$$\le \frac{C}{\sqrt{n}} \int_{|y|\le\delta\sqrt{nk}} |y|^3\, e^{-(\beta-1)\,y^2/2}\,dy \ = \ O\Big(\frac{1}{\sqrt{n}}\Big).$$

Hence

$$I_1 = \int_{|y|\le\delta\sqrt{nk}} e^{-(\beta-1)\,y^2} dy + O\Big(\frac{1}{\sqrt{n}}\Big) \ = \ \sqrt{\pi(\alpha-1)} + O\Big(\frac{1}{\sqrt{n}}\Big), \qquad n \to \infty.$$

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

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

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

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

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

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

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

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

13. **Pointwise Upper Bounds for Convolutions of Densities**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

By Proposition 4.3, $\mathbb{E}\,e^{cZ_{n_0}^2} < \infty$ and hence $\mathbb{E}\,e^{cX^2} < \infty$ for some $c > 0$. In particular, $f(t)$ and all characteristic functions $f_n(t) = \mathbb{E}\,e^{itZ_n} = f(t/\sqrt{n})^n$ are extended as entire functions to the complex plane. Moreover, since the map $h \to e^{hx}p_n(x)$ for a fixed $n \geq \beta n_0$ is continuous from $\mathbb{R}$ to $L^1(\mathbb{R})$, the family $\{e^{hx}p_n(x)\}_{0\leq h\leq y}$ is compact in $L^1(\mathbb{R})$. Hence, $f_n(t)$ tends to zero at infinity uniformly in every strip $|\mathrm{Im}\,t| \leq y < \infty$ (by the Riemann-Lebesgue lemma). Applying Cauchy's theorem to the rectangle contour $[-T,T] \cup [T,T+iy] \cup [T+iy,-T+iy] \cup [-T+iy,-T]$, the inversion formula may therefore be written as

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

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

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

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

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

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

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

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

Therefore,

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

and (13.4) results in the upper bound

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

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

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

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

Now, since $\beta\geq 2$, we are allowed to apply the Hausdorff-Young inequality to get

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

To estimate the last integral, factorize its integrand as $\big(e^{-\alpha yu\sqrt{n_0}/\sqrt{n}}\,\varphi(u)^{\alpha-1}\big)\,\frac{p_{n_0}(u)^\alpha}{\varphi(u)^{\alpha-1}}$ and bound the expression in the brackets by $(2\pi)^{-\frac{\alpha-1}{2}}\,e^{\frac{\alpha\beta y^2 n_0}{2n}}$. This gives

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

Hence, (13.5) results in the upper bound

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

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

14. Sufficiency Part in Theorem 1.2 ($d = 1$).

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

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

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

According to Lemma 11.2, the integrals of the form

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

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

$$\int_{|x| > M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx \, = \, O\Big(\frac{1}{n^{\kappa s}}\Big), \qquad n \to \infty. \tag{14.1}$$

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

$$I_1 = \int_{-\infty}^{-x_0\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx, \quad I_2 = \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx, \quad I_3 = \int_{-x_1\sqrt{n}}^{-M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx$$

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

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

$$\begin{aligned} I_1 \ &\leq \ (2\pi)^{\frac{\alpha-1}{2}}\, \delta^{\alpha n} \int_{-\infty}^{-x_0\sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{\alpha n/2}\, dx \\ &\leq \ (2\pi)^{\frac{\alpha-1}{2}}\, \delta^{\alpha n} \beta\sqrt{n} \int_{-\infty}^{\infty} \psi(u)^m\, du, \qquad m \leq \frac{\alpha n}{2}, \end{aligned}$$

where on the last step we used $\psi \leq 1$. By Corollary 5.2, the last integral is convergent whenever $m = kn_0$, $k \geq \alpha$. One may take, for example, $k = [\alpha] + 1$, which ensures the condition $m \leq \frac{\alpha n}{2}$ for all sufficiently large n. Hence

$$I_1 \, \leq \, C\delta_1^n \qquad (n \geq n_1)$$

with some constants $C > 0$, $x_0 > 0$ and $\delta < \delta_1 < 1$, depending on the density p only.

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

$$\begin{aligned} I_2 &\le A_\alpha n^{\alpha/2} \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta}\,dx \\ &= A_\alpha^\alpha\,\beta\,n^{(\alpha+1)/2} \int_{-x_0/2}^{-x_1/2} \psi(u)^{n-n_\beta}\,du \ \le\ A_\alpha^\alpha\,\beta\,n^{(\alpha+1)/2}\,(x_0-x_1)\,\delta_2^{n-n_\beta} \end{aligned}$$

which again decays exponentially fast like I_1.

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

$$\psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta} \le \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n/2} \le \exp\Big\{-\frac{\beta-1}{4}\,\frac{x^2}{2\beta^2}\Big\} = e^{-x^2/(8\alpha\beta)}$$

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

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

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

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

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

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

Since $T_\alpha(Z_n||Z) = \frac{1}{\alpha-1}\,(I_0 + I_1 + I_2 + I_3 - 1)$, and using the expansion (11.6), we conclude that, for any sufficiently large s and hence for any $s = 3, 4, \dots$,

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

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

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

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

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

Corollary 14.1. *Given a random variable X with $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 2$, there exists $n_1 \ge 1$ such that the sequence $D_\alpha(Z_n||Z)$ is non-increasing for $n \ge n_1$.*

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

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

Hence, the increments

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

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

It is interesting to know, whether or not Corollary 14.1 is true with $n_1 = 1$, which would generalize the monotonicity of the relative entropy along normalized convolutions, cf. [A-B-B-N2], [M-B], [T-V].

15. **Non-uniform Local Limit Theorem**

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

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

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

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

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

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

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

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

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

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

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

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

Now substitute t with $t\sqrt{n}$ and raise the above inequality to the power $1/n$. Letting $n \to \infty$ along a suitable subsequence, we arrive in the limit at

$$\mathbb{E}\, e^{tX} \, \leq \, e^{\beta t^2/2}, \qquad t \in \mathbb{R}.$$

Thus, this subgaussian property is indeed implied by the local limit theorem (15.2).

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

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

Then, it will be sufficient to restrict the supremum in (15.1) to Δ_0 and to bound

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

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

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

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

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

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

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

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

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

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

Finally, in order to study the asymptotic behavior of

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

we invoke the Edgeworth expansion (11.3) of Proposition 11.1. After division by $\varphi(x)^{1/\beta}$, the remainder term there will be $O(n^{-\frac{s-1}{2}})$ uniformly on Δ_0, as soon as $\frac{k-2}{2} - \frac{l-1}{\beta} \geq s-1$. Pick up such an integer k (necessarily $k \geq 3$). As a result, we may replace $p_n(x)$ in the definition of J_0 by the Edgeworth approximation described on the right-hand side of (11.3). The first (potentially) non-zero term in it has the form $\frac{1}{s!}\, \gamma_s\, H_s(x)\varphi(x)\, n^{-\frac{s-2}{2}}$, while all remaining terms are $\varphi(x) q_\nu(x)\, n^{-\nu/2}$ with $\nu \geq s-1$. After division by $\varphi(x)^{1/\beta}$, such terms may therefore contribute in the supremum a quantity which is $O(n^{-\frac{s-1}{2}})$. This means that we are left with the leading term, resulting in (15.1). □

16. The Multidimensional Case

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

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

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

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

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

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

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

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

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

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

Like in dimension one, the finiteness of $D_\alpha(Z_{n_0}||Z)$ guarantees the existence of bounded continuous densities p_n for Z_n for all $n \geq n_\beta = \max(\beta, 2) n_0$, cf. Proposition 13.1. In addition, $\mathbb{E}\, e^{c|X|^2} < \infty$ for some $c > 0$. In particular, the characteristic function $f(t) = \mathbb{E}\, e^{i\langle X, t\rangle}$ extends as an entire function to the d-dimensional complex space $\mathbb{C}^d$. Most important properties of the densities p_n rely upon the function

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

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

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

The first assertion is a multidimensional analog of Proposition 4.3; it can be proved with very similar arguments as in dimension one. The second assertion generalizing

Corollary 5.2 can be proved by using the contractivity properties of the d-dimensional Weierstrass transform

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

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

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

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

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

$$p_n(x) \ \leq\ A_{\alpha,d}\, n^{d/2}\, e^{-|x|^2/(2\beta)}\, \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta}, \tag{16.3}$$

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

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

Proof of Theorem 1.2 (Sufficiency part) **and Theorem 1.3**. We need to explore the asymptotic behavior of

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

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

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

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

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

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

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

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

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

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

$$I_1 \ \leq \ C_1\,\delta^{\alpha n} \int_{|x| > x_0 \sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{\alpha n/2} dx \ \leq \ C_1\,\delta^{\alpha n} \int_{\mathbb{R}^d} \psi(u)^{k n_0}\,du \ \leq \ C_2\,\delta_1^n$$

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

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

$$\begin{aligned} I_2 \ &\leq \ C_1\,n^{d\alpha/2} \int_{x_1\sqrt{n} < |x| < x_0 \sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n_1} dx \\ &= \ C_2\,n^{d(\alpha+1)/2} \int_{x_1\sqrt{n} < |x| < x_0 \sqrt{n}} \psi(u)^{n_1}\,du \ \leq \ C_3\,n^{d(\alpha+2)/2}\,x_0^d\,\delta_2^{n_1} \end{aligned}$$

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

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

Therefore, by (16.2), in this ball $w_n(x) \ \leq \ A_{\alpha,d}^\alpha\,n^{d\alpha/2}\,e^{-|x|^2/(8\alpha\beta)}$, which gives with $x_1 = \beta r$

$$\begin{aligned} I_3 \ &\leq \ C_1\,n^{d\alpha/2} \int_{M_n < |x| < x_1 \sqrt{n}} e^{-|x|^2/(8\alpha\beta)}\,dx \\ &< \ C_2\,n^{d\alpha/2}\,\mathbb{P}\{|Z|^2 > M_n^2/(8d\alpha\beta)\} \ \leq \ C_3\,n^{d\alpha/2}\,e^{-M_n^2/(8d\alpha\beta)} \ = \ C_3\,n^{-(\frac{l-1}{4d\alpha\beta} - \frac{\alpha}{2})}. \end{aligned}$$

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

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

$$J_0 = \sup_{|x| \leq M_n} \frac{|p_n(x) - \varphi_s(x)|}{\varphi(x)^{1/\beta}}, \quad J_1 = \sup_{|x| \geq x_1\sqrt{n}} \frac{p_n(x)}{\varphi(x)^{1/\beta}}, \quad J_2 = \sup_{M_n \leq |x| \leq x_1\sqrt{n}} \frac{p_n(x)}{\varphi(x)^{1/\beta}}$$

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

17. Some Examples and Counter-Examples

Given a random variable X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, consider the function $\psi(t) = e^{-t^2}\,\mathbb{E}\,e^{tX}$ ($t \in \mathbb{R}$). As before, put $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$, where X_j's are independent copies of X. One immediate consequence of Theorem 1.1 (with $n_0 = 1$) is the following characterization. As usual, Z denotes a standard normal random variable.

Theorem 17.1. *Assume that the random variable X has a density p such that*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Distributions with Gaussian component. Consider random variables of type

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

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

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

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

$$\mathbb{E}\, e^{t\xi} \leq e^{\sigma^2 t^2/2}, \qquad t \in \mathbb{R}, \tag{17.4}$$

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

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

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

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

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

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

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

under the same constraint $a < a_\sigma$. Thus, by Theorem 17.1, $\chi^2(Z_n,Z) \to 0$ as $n \to \infty$, if $a < \frac{1}{\sqrt{\sigma^2-1}}$. In case $\sigma^2 \leq 2$, this convergence holds for all admissible (a,b).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

18. **Convolution of Bernoulli with Gaussian**

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

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

$$\mathbb{E}\, e^{tX} < e^{\beta t^2/2} \tag{18.1}$$

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

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

Let us describe explicitly one family of distributions satisfying the assertion of this proposition. Returning to one of the previous examples, consider random variables of the form

$$X_p = a\xi + bZ \qquad (a, b > 0),$$

assuming that ξ takes two values q and $-p$ with probabilities p and q, respectively ($p, q > 0$, $p + q = 1$), while $Z \sim N(0,1)$ is independent of ξ. Then $\mathbb{E}X_p = 0$, and we have the constraint

$$\mathbb{E}X_p^2 = pq\, a^2 + b^2 = 1. \tag{18.2}$$

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

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

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

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

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

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

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

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

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

which makes sense provided that $\sigma^2 > \beta pq$. According to [B-H-T], Proposition 2.3, the subgaussian constant for the Bernoulli distribution is known to be

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

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

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

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

References

[A1] Amosova, N. N. Narrow zones of local normal attraction. (Russian) Teor. Veroyatn. Primenen. 35 (1990), no. 1, 138–143. Translation in: Theory Probab. Appl. 35 (1990), no. 1, 140–145 (1991).

[A2] Amosova, N. N. A remark on a local limit theorem for large deviations. (Russian) Teor. Veroyatn. Primenen. 35 (1990), no. 4, 754–756. Translation in: Theory Probab. Appl. 35 (1990), no. 4, 758–760 (1991).

[A-B-B-N1] Artstein, S.; Ball, K. M.; Barthe, F.; Naor, A. On the rate of convergence in the entropic central limit theorem. Probab. Theory Related Fields 129 (2004), no. 3, 381–390.

[A-B-B-N2] Artstein, S.; Ball, K. M.; Barthe, F.; Naor, A. Solution of Shannon's problem on the monotonicity of entropy. J. Amer. Math. Soc. 17 (2004), no. 4, 975–982.

[B-C] Bally, V.; Caramellino, L. Asymptotic development in the CLT in total variation distance. Bernoulli 22 (2016), 2442–2485.

[B] Barron, A. R. Entropy and the central limit theorem. Ann. Probab. 14 (1986), no. 1, 336–342.

[B-J] Barron, A. R.; Johnson, O. Fisher information inequalities and the central limit theorem. Probab. Theory Related Fields 129 (2004), no. 3, 391–409.

[B-RR] Bhattacharya, R. N.; Ranga Rao, R. Normal approximation and asymptotic expansions. John Wiley & Sons, Inc. 1976. Also: Soc. for Industrial and Appl. Math., Philadelphia, 2010.

[B-C-G1] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Non-uniform bounds in local limit theorem in case of fractional moments. I. Math. Methods of Statistics, 20 (2011), no. 3, 171–191; II. Math. Methods of Statistics, 20 (2011), no. 4, 269–287.

[B-C-G2] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. Ann. Probab. 41 (2013), no. 4, 2479–2512.

[B-C-G3] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Berry-Esseen bounds in the entropic central limit theorem. Probab. Theory Related Fields 2014 (159), 435–478.

[B-C-G4] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Fisher information and the central limit theorem. Probab. Theory Related Fields 159 (2014), issue 1-2, 1–59.

[B-C-K] Bobkov, S. G.; Chistyakov, G. P.; Kösters, H. The entropic Erdös-Kac limit theorem. J. Theor. Probab. 28 (2015), no. 4, 1520–1555.

[B-G] Bobkov, S. G.; Götze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. 163 (1999), no. 1, 1–28.

[B-H-T] Bobkov, S. G.; Houdré, C.; Tetali, P. The subgaussian constant and concentration inequalities. Israel J. Math. 156 (2006), 255–283.

[B-P-T] Borland, L.; Plastino, A. R.; Tsallis, C. Information gain within nonextensive thermostatistics. J. Math. Phys. 39 (1998), no. 12, 6490–6501.

[C] Carlen, E. A. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. 101 (1991), no. 1, 194–211.

[Cr] Cramér, H. On some classes of series used in mathematical statistics. Proc. 6th Scand. Math. Congr. Copenhagen, 1925, 399–425. See also: Harald Cramér. Collected works. Volume I, Ed. A. Martin-Löf, Springer-Verlag, 438–464.

[Cs] Csiszár, I. Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. 2 (1967), 299–318.

[D-C-T] Dembo, A., Cover, T. M., Thomas, J. A. Information-theoretic inequalities. IEEE Trans. Inform. Theory, 37 (1991), no. 6, 1501–1518.

[E-H] van Erven, T., Harremoës, P. Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inform. Theory 60 (2014), no. 7, 3797–3820.

[F] Fomin, S. V. The central limit theorem: convergence in the norm $\|u\| = (\int_{-\infty}^{\infty} u^2(x)\, e^{x^2/2}\, dx)^{1/2}$. (Russian) Problems of the theory of probability distributions, VII. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 119 (1982), 218–229.

[G-S] Gibbs, A. L.; Su, F. E. On choosing and bounding probability metrics. Intern. Statist. Review 70 (2002), issue 3, 419–435.

[G] Gilardoni, G. L. On Pinsker's and Vajda's type inequalities for Csiszar's f-divergences. IEEE Trans. Inform. Theory 56 (2010), no. 11, 5377–5386.

[J] Johnson, O. Information theory and the central limit theorem. Imperial College Press, London, 2004, xiv+209 pp.

[M-B] Madiman, M.; Barron, A. Generalized entropy power inequalities and monotonicity properties of information. IEEE Trans. Inform. Theory 53 (2007), no. 7, 2317–2329.

[H-W] Hirschman, I. I.; Widder, D. V. The convolution transform. Princeton University Press, Princeton, N. J., 1955. x+268 pp.

[I-L] Ibragimov, I. A., Linnik, Yu. V. Independent and stationary sequences of random variables. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov. Translation from the Russian edited by J. F. C. Kingman. Wolters-Noordhoff Publishing, Groningen, 1971, 443 pp.

[K] Kullback, S. A lower bound for discrimination in terms of variation. IEEE Trans. Inform. Theory, T-13, 1967, 126–127.

[K-L] Kullback, S., Leibler, R. A. On information and sufficiency. Ann. Math. Stat. 22 (1951), 79–86.

[Le] LeCam, L. M. Asymptotic Methods in Statistical Decision Theory. Springer Series in Statistics. Springer-Verlag, New York, 1986. xxvi+742 pp.

[L] Lieb, E. H. Some convexity and subadditivity properties of entropy. Bull. Amer. Math. Soc. 81 (1975), 1–13.

[L-V] Liese, F.; Vajda, I. Convex statistical distances. With German, French and Russian summaries. Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987. 224 pp.

[Li] Linnik, Ju. V. An information-theoretic proof of the central limit theorem with the Lindeberg condition. Theory Probab. Appl. 4 (1959), 288–299.

[N] Nielsen, F. On the Chi square and higher-order Chi distances for approximating f-divergences. IEEE Signal Processing Letters. Vol. 21, issue 1, Jan. 2014.

[O-V] Otto, F.; Villani, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. 173 (2000), no. 2, 361–400.

[Pe1] Petrov, V. V. Local limit theorems for sums of independent random variables. (Russian) Teor. Verojatnost. i Primenen. 9 (1964), 343–352.

[Pe2] Petrov, V. V. Sums of Independent Random Variables. Berlin Heidelberg New York (1975).

[Pi] Pinsker, M. S. Information and information stability of random variables and processes. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964, xii+243 pp.

[Pr] Prohorov, Yu. V. A local theorem for densities. (Russian) Doklady Akad. Nauk SSSR (N.S.) 83, (1952), 797–800.

[R] Rényi, A. On measures of entropy and information. Proc. 4th Berkeley Sympos. Math. Statist. and Prob. Vol. I, pp. 547–561, 1961, Univ. California Press, Berkeley, Calif.

[S] Shiryaev, A. N. Probability. Springer-Verlag, 1996.

[S-M] Sirazhdinov, S. H.; Mamatov, M. On mean convergence for densities. Theory Probab. Appl., 7 (1962), No. 4, 424–428.

[Sz] Szegö, G. Orhtogonal polynomials. 3rd ed., American Math. Soc. Publications 23. Amer. Math. Soc., Providence, RI.

[To1] Toscani, G. Entropy inequalities for stable densities and strengthened central limit theorems. J. Stat. Phys. 165 (2016), no. 2, 371–389.

[To2] Toscani, G. The fractional Fisher information and the central limit theorem for stable laws. Ric. Mat. 65 (2016), no. 1, 71–91.

[Ts] Tsallis, C. Generalized entropy-biased criterion for consistent testing. Phys. Rev. E 58 (1998), 1442–1445.

[T-V] Tulino, A. M., Verdú, S. Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof. IEEE Trans. Inform. Theory 52, 4295–4297 (2006)

[V] Vajda, I. Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht-Borston-London, 1989, 432 pp.