Probab. Theory Relat. Fields (2014) 159:435–478
DOI 10.1007/s00440-013-0510-3

Berry–Esseen bounds in the entropic central limit theorem

Sergey G. Bobkov · Gennadiy P. Chistyakov · Friedrich Götze

Received: 22 August 2011 / Revised: 15 May 2013 / Published online: 12 June 2013

Abstract Berry–Esseen-type bounds for total variation and relative entropy distances to the normal law are established for the sums of non-i.i.d. random variables.

Keywords Entropy · Entropic distance · Central limit theorem · Berry–Esseen bounds

Mathematics Subject Classification (1991) Primary 60E

Contents

This research was partially supported by NSF grant DMS-1106530 and SFB 701.

S. G. Bobkov
School of Mathematics, University of Minnesota,
127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA
e-mail: bobkov@math.umn.edu

G. P. Chistyakov · F. Götze (✉)
Fakultät für Mathematik, Universität Bielefeld,
Postfach 100131, 33501 Bielefeld, Germany
e-mail: goetze@math.uni-bielefeld.de

G. P. Chistyakov
e-mail: chistyak@math.uni-bielefeld.de

1 Introduction

Let $X_1, \dots, X_n$ be independent (not necessarily identically distributed) random variables with mean $\mathbf{E}X_k = 0$ and finite variances $\sigma_k^2 = \mathbf{E}X_k^2 (\sigma_k > 0)$. Put $B_n = \sum_{k=1}^n \sigma_k^2$. Under additional moment assumptions, the normalized sum

$$S_n = \frac{X_1 + \dots + X_n}{\sqrt{B_n}}$$

has approximately a standard normal distribution in a weak sense. More precisely (see [19]), the closeness of the distribution function $F_n(x) = \mathbf{P}\{S_n \leq x\}$ to the standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

has been studied intensively in terms of the Lyapunov ratios

$$L_s = \frac{\sum_{k=1}^n \mathbf{E}\,|X_k|^s}{B_n^{s/2}}.$$

In particular, if all X_k have finite third absolute moments, the classical Berry–Esseen theorem indicates that

$$\sup_x |F_n(x) - \Phi(x)| \leq C L_3, \tag{1.1}$$

where C is an absolute constant (cf. e.g. [12, 14, 19]).

One of the most remarkable features of (1.1) is that the number of summands does not *explicitly* appear in it, while in the i.i.d. case, that is, when X_k have identical distributions, L_3 is of order $\frac{1}{\sqrt{n}}$, which is best possible for the Kolmogorov distance under the 3-rd moment condition (see, for example [19, p.169]).

In this paper we consider the closeness of F_n to Φ in terms of generally stronger distances, such as total variation and relative entropy. Given two distribution functions F and G, introduce the notation

$$\|F - G\|_{\mathrm{TV}} = 2 \sup_A \left| \int_A dF(x) - \int_A dG(x) \right|$$

for the total variation distance between F and G (where the supremum is running over all Borel subsets A of the real line). If F is absolutely continuous with respect to G (as measures) and has density $u = dF/dG$, one defines the Kullback–Leibler distance or the relative entropy of F with respect to G by

$$D(F||G) = \int_{-\infty}^{+\infty} u \, \log u \, dG.$$

If F is not absolutely continuous with respect to G, one puts $D(F||G) = +\infty$.

Our aim is to establish bounds for $\|F_n - \Phi\|_{\mathrm{TV}}$ and $D(F_n||\Phi)$ by using the Lyapunov ratios similarly as in (1.1). Note, however, that these distances are not informative, for example, when all summands have discrete distributions, in which case $\|F_n - \Phi\|_{\mathrm{TV}} = 2$, $D(F_n||\Phi) = +\infty$. Therefore, some assumptions are needed or desirable, such as absolute continuity of distributions F_{X_k} of X_k. But even with this assumption we cannot exclude the case that our distances from S_n to the normal law may be growing when the F_{X_k} are close to discrete distributions. To prevent such behaviour, one may require that the densities of X_k should be bounded on a reasonably large part of the real line. This can be guaranteed quite naturally, for instance, by using the entropy functional, defined for a random variable X with density p by

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

Once X has a finite second moment, the entropy is well-defined as a Lebesgue integral, although the value $h(X) = -\infty$ is possible. Introduce a related functional

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

where Z is a normal random variable with density $q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-a)^2}{2\sigma^2}\}$ having the same mean a and variance σ^2 as X. Note that this functional is affine invariant, that is, $D(c_0 + c_1 X) = D(X)$, for all $c_0 \in \mathbf{R}$, $c_1 \neq 0$, and in this sense it depends neither on the mean, nor the variance of X.

The quantity $D(X)$ may also be regarded as the relative entropy $D(F_X||F_Z)$, where F_X and F_Z are the corresponding distributions of X and Z. It represents the Kullback–Leibler distance from F_X to the class of all normal laws on the real line and is often referred to as the "entropic distance to normality". In general, $0 \leq D(X) \leq +\infty$, and the equality $D(X) = 0$ is possible, when X is normal, only. Moreover, by the Pinsker-Csiszár-Kullback inequality [11,13,17,21], the entropic distance dominates the total variation in the sense that

$$D(X) \geq \frac{1}{2} \|F_X - F_Z\|_{\mathrm{TV}}^2.$$

Thus, finiteness of $D(X)$ guarantees that F_X is separated from the class of discrete probability distributions, and if it is small, one may speak about the closeness of F_X to normality in a rather strong sense. Using D for both purposes, one can obtain refinements of Berry–Esseen's inequality (1.1) in terms of the total variation and the entropic distances to normality for the distributions F_n. The fact that the convergence in the central limit theorem can be studied in terms of the entropy was first noticed by Linnik [18], see also Brown [8], Barron [2], Carlen and Soffer [9].

We start with a quantitative bound for the total variation distance.

Theorem 1.1 *Let D be a non-negative number. Assume that the independent random variables $X_1, \dots, X_n$ have finite third absolute moments, and that $D(X_k) \le D (1 \le k \le n)$. Then*

$$\|F_n - \Phi\|_{\mathrm{TV}} \le C L_3, \tag{1.2}$$

where the constant $C = C_D$ depends on D, only.

In particular, if all X_k are identically distributed with $\mathbf{E} X_1^2 = 1$, we get

$$\|F_n - \Phi\|_{\mathrm{TV}} \ \le \ \frac{C}{\sqrt{n}}\, \mathbf{E}\, |X_1|^3 \tag{1.3}$$

with a constant C depending on $D(X_1)$, only. Although (1.2)–(1.3) seem to be new, related estimates in the i.i.d.-case were studied by many authors. For example, in the early 1960s Mamatov and Sirazhdinov [27] found an exact asymptotic $\|F_n - \Phi\|_{\mathrm{TV}} = \frac{c}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})$, where the constant c is proportional to $|\mathbf{E} X_1^3|$, and which holds under the assumption that the distribution of X_1 has a non-trivial absolutely continuous component (cf. also [22,25]).

Now, let us turn to the entropic distance to normality.

Theorem 1.2 *Assume that the independent random variables $X_1, \dots, X_n$ have finite fourth moments, and that $D(X_k) \le D (1 \le k \le n)$. Then*

$$D(S_n) \le C L_4, \tag{1.4}$$

where $C = C_D$ depends on D, only.

In (1.2) and (1.4) one may take $C_D = e^{c(D+1)}$, where c is an absolute constant. Moreover, as we will see in Theorems 11.2 and 12.3 below, C_D can be chosen to be independent of D (i.e., to be just a numerical constant), provided that the respective Lyapunov ratios are smaller than a certain numerical value, while D is not too large, namely, if

$$D \le c \log \frac{1}{L_3} \quad \text{and} \quad D \le c \log \frac{1}{L_4}$$

with some absolute constant $c > 0$.

These Berry–Esseen-type estimates are consistent in view of the Pinsker inequality. In some sense, one may consider (1.4) as a stronger assertion than (1.2), which is indeed the case, when L_4 is of order L_3^2. (In general $L_3^2 \le L_4$.)

In the i.i.d. case as in (1.3), the inequality (1.4) becomes

$$D(S_n) \le \frac{C}{n}\, \mathbf{E}X_1^4,$$

where C depends on $D(X_1)$ only. Thus, we obtain an error bound of order $O(1/n)$ under the 4th moment assumption. Note that the property $D(S_n) \to 0$ always holds under the second moment assumption (with finite entropy of X_1). This is the statement of the entropic central limit theorem, which is due to Barron [2]. Here, the convergence may have an arbitrarily slow rate. Nevertheless, the expected typical rate $D(S_n) = O(\frac{1}{n})$ was known to hold in some cases, for example, when X_1 has a distribution satisfying an integro-differential inequality of Poincaré-type. These results are due to Artstein et al. [1], and Barron and Johnson [3]; cf. also [16]. Recently, an exact asymptotic for $D(S_n)$ has been studied in [5]. If the entropy and the 4th moment of X_1 are finite, it was shown that

$$D(S_n) = \frac{c}{n} + o\left(\frac{1}{n \log n}\right), \quad c = \frac{1}{12}\,(\mathbf{E}X_1^3)^2.$$

Moreover, with finite 3rd absolute moment (and infinite 4th moment) such a relation may not hold, and it may happen that $D(S_n) \ge n^{-(1/2+\varepsilon)}$ for all n large enough with a given prescribed $\varepsilon > 0$. This holds, for example, when X_1 has density

$$p(x) = \int_{1/e}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}\, dP(\sigma),$$

where P is a probability measure on $(\frac{1}{e}, +\infty)$ with density $\frac{dP(\sigma)}{d\sigma} = (\sigma \log \sigma)^{-4}$ for $\sigma \ge e$ and with an arbitrary extension to the interval $\frac{1}{e} < \sigma < e$ satisfying $\int_{1/e}^{+\infty} \sigma^2\, dP(\sigma) = 1$.

Therefore, in the general non-i.i.d.-case, the Lyapunov coefficient L_3 cannot be taken as an appropriate quantity for bounding the error in Theorem 1.2, and L_4 seems more relevant. This is also suggested by the result of [1] for the weighted sums

$$S_n = a_1 X_1 + \cdots + a_n X_n \quad (a_1^2 + \cdots + a_n^2 = 1)$$

of i.i.d. random variables X_k such that $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 = 1$. Namely, it is proved there that

$$D(S_n) \le \frac{L(a)}{c/2 + (1 - c/2)L(a)}\, D(X_1), \tag{1.5}$$

where $L(a) = a_1^4 + \cdots + a_n^4$ and $c \geq 0$ is an optimal constant in the Poincaré-type inequality $c\,\mathrm{Var}(u(X_1)) \leq \mathbf{E}\,[u'(X_1)]^2$. But for the sequence $a_k X_k$ and $s = 4$, the corresponding Lyapunov coefficient is exactly $L_4 = L(a)\,\mathbf{E}X_1^4$. Therefore, when $c = c(X_1)$ is positive, (1.5) yields the estimate

$$D(S_n) \leq \frac{2D(X_1)}{c\,\mathbf{E}X_1^4}\,L_4,$$

which is of similar nature as in (1.4).

Another interesting feature of (1.4) is that it may be connected with transportation cost inequalities for the distributions F_n of S_n in terms of the quadratic Kantorovich distance W_2 (also called the Wasserstein distance). For random variables X and Z with finite second moments and distributions F_X and F_Z, this distance is defined by

$$W_2^2(F_X, F_Z) = \inf_{\pi} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |x-y|^2\, d\pi(x,y),$$

where the infimum is taken over all probability measures π on the plane $\mathbf{R}^2$ with marginals F_X and F_Z. The value $W_2^2(F_X, F_Z)$ is interpreted as the minimal expenses needed to transport F_Z to F_X, provided that it costs $|x-y|^2$ to move any "particle" x to any "particle" y.

The metric W_2 is of weak type in the sense that it can be used to metrize the weak convergence of probability distributions ([29]). Moreover, if $Z \sim N(0,1)$ is standard normal, this distance, i.e., $W_2(F_X, F_Z) = W_2(F_X, \Phi)$, may be bounded in terms of the relative entropy by virtue of Talagrand's transportation inequality

$$W_2^2(F_X, \Phi) \leq 2D(F_X||\Phi) \tag{1.6}$$

(cf. [28], or [7] for a different approach). If additionally X has mean zero and unit variance, then $D(F_X||\Phi) = D(X)$. Hence, applying (1.6) with $X = S_n$, we get, by Theorem 1.2,

$$W_2(F_n, \Phi) \leq C\sqrt{L_4}, \tag{1.7}$$

where C depends on D. In fact, this inequality holds true with C being an absolute constant. This result is due to Rio [23], who also studied more general Wasserstein distances W_r, by relating them to Zolotarev's "ideal" metrics. It has also been noticed in [23] that the 4th moment condition is essential, so the Laypunov's ratio L_4 in (1.7) cannot be replaced with L_3 including the i.i.d.-case (like in Theorem 1.2).

The paper starts with general bounds on the total variation and the Kullback–Leibler distance to the standard normal law in terms of characteristic functions. In the proof of Theorems 1.1–1.2, these bounds will be applied to special probability distributions $\widetilde{F}_n$ that approximate F_n sufficiently well. These distributions are constructed according to the so-called quantile density decomposition whose general properties are discussed separately. Several sections are devoted to the construction and the study of basic properties of $\widetilde{F}_n$ and their characteristic functions.

2 General bounds on total variation and entropic distance

Assume that a random variable X has an absolutely continuous distribution F with density p and finite first absolute moment. We do not require that it has mean zero and/or unit variance.

First, we recall an elementary bound for the total variation distance $\|F - \Phi\|_{\rm TV}$ in terms of the characteristic function

$$f(t) = \mathbf{E}\, e^{itX} = \int_{-\infty}^{+\infty} e^{itx}\, p(x)\, dx \quad (t \in \mathbf{R}).$$

Introduce the characteristic function $g(t) = e^{-t^2/2}$ of the standard normal law.

In the sequel, we use the notation

$$\|u\|_2 = \bigg(\int_{-\infty}^{+\infty} |u(t)|^2\, dt\bigg)^{1/2}$$

to denote the L^2-norm of a measurable complex-valued function u on the real line (with respect to Lebesgue measure).

Proposition 2.1 *We have*

$$\|F - \Phi\|_{\rm TV}^2 \ \le\ \frac{1}{2}\, \|f - g\|_2^2 + \frac{1}{2}\, \|f' - g'\|_2^2. \tag{2.1}$$

This bound is standard (cf. e.g. [15, Lemma 1.3.1]). In fact, the inequality (2.1) remains to hold for an arbitrary probability distribution in place of Φ with finite first absolute moment and characteristic function g. However, the general case will not be needed in the sequel. Note that the assumption $\mathbf{E}\,|X| < +\infty$ guarantees that f is continuously differentiable, so that the last L^2-norm in (2.1) makes sense.

Let Z denote a standard normal random variable, with density $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$. Consider the relative entropy

$$D(X||Z) = D(F||\Phi) = \int_{-\infty}^{+\infty} p(x)\, \log \frac{p(x)}{\varphi(x)}\, dx. \tag{2.2}$$

As a preliminary bound for this quantity, we first derive:

Lemma 2.2 *For all $T \geq 0$,*

$$D(X||Z) \leq e^{-T^2/2} + \sqrt{2\pi} \int_{-T}^{T} (p(x) - \varphi(x))^2 \, e^{x^2/2} \, dx$$
$$+ \frac{1}{2} \int_{|x| \geq T} x^2 \, p(x) \, dx + \int_{|x| \geq T} p(x) \log p(x) \, dx. \tag{2.3}$$

Proof We split the integral in (2.2) into the two regions. For the interval $|x| \leq T$, using the elementary inequality $t \log t \leq (t-1) + (t-1)^2$, $t \geq 0$, we have

$$\int_{-T}^{T} \frac{p}{\varphi} \log \frac{p}{\varphi} \, \varphi \, dx \leq \int_{-T}^{T} \left(\frac{p}{\varphi} - 1\right) \varphi \, dx + \int_{-T}^{T} \left(\frac{p}{\varphi} - 1\right)^2 \varphi \, dx$$
$$= \int_{|x| \geq T} (\varphi - p) \, dx + \int_{-T}^{T} \frac{(p - \varphi)^2}{\varphi} \, dx$$
$$= 2\,(1 - \Phi(T)) - \int_{|x| \geq T} p(x) \, dx + \sqrt{2\pi} \int_{-T}^{T} (p(x) - \varphi(x))^2 \, e^{x^2/2} \, dx.$$

For the second region, just write

$$\int_{|x| \geq T} p(x) \log \frac{p(x)}{\varphi(x)} \, dx = \int_{|x| \geq T} p(x) \log p(x) \, dx$$
$$+ \log \sqrt{2\pi} \int_{|x| \geq T} p(x) \, dx + \frac{1}{2} \int_{|x| \geq T} x^2 \, p(x) \, dx.$$

It remains to collect these relations and use $\log \sqrt{2\pi} < 1$ together with a well-known elementary inequality $1 - \Phi(T) \leq \frac{1}{2} \, e^{-T^2/2}$. Thus, Lemma 2.2 is proved. □

Remark If p is bounded by a constant M, the estimate (2.3) yields

$$D(X||Z) \leq e^{-T^2/2} + \sqrt{2\pi} \int_{-T}^{T} (p(x) - \varphi(x))^2 \, e^{x^2/2} \, dx$$
$$+ \frac{1}{2} \int_{|x| \geq T} x^2 \, p(x) \, dx + \log M \int_{|x| \geq T} p(x) \, dx.$$

This bound might be of interest in other applications, although it involves the maximum of the density. For our purposes, the important integral in (2.3), $\int_{|x| \geq T} p(x) \log p(x) \, dx$, will be bounded in a different way and in terms of the characteristic functions, without involving the parameter M.

3 Entropic distance and Edgeworth-type approximation

To estimate the integrals in (2.3) in terms of the characteristic functions like in Proposition 2.1, define

$$\varphi_\alpha(x) = \varphi(x)\left(1 + \alpha\,\frac{x^3 - 3x}{3!}\right),$$

where α is a parameter. These functions appear with α proportional to $n^{-1/2}$ in the Edgeworth-type expansions up to order 3 for densities of the normalized sums $S_n = \frac{X_1+\dots+X_n}{\sqrt{B_n}}$ of i.i.d. summands, cf. e.g. [19]. In the non-i.i.d. case such expansions hold as well with

$$\alpha = \frac{1}{B_n^{3/2}} \sum_{k=1}^n \mathbf{E} X_k^3.$$

Note that every φ_α has the Fourier transform

$$g_\alpha(t) = \int_{-\infty}^{+\infty} e^{itx} \varphi_\alpha(x)\,dx = g(t)\left(1 + \alpha\,\frac{(it)^3}{3!}\right),$$

where $g(t) = e^{-t^2/2}$.

Proposition 3.1 *Let X be a random variable with $\mathbf{E}\,|X|^3 < +\infty$. For all $\alpha \in \mathbf{R}$,*

$$D(X||Z) \le \alpha^2 + 4\,(\|f - g_\alpha\|_2 + \|f''' - g_\alpha'''\|_2), \tag{3.1}$$

where Z is a standard normal random variable and f is the characteristic function of X.

The assumption on the 3rd absolute moment is needed to ensure that f has first three continuous derivatives.

As a particular case, the inequality (3.1) is valid for $\alpha = 0$, as well. Then it becomes

$$D(X||Z) \le 4\,(\|f - g\|_2 + \|f''' - g'''\|_2),$$

which may be viewed as a full analog of Proposition 2.1. However, with properly chosen values of α, (3.1) may provide a much better asymptotic approximation (especially when applying it to the sums of independent random variables).

Proof We may assume that the characteristic function f and its first three derivatives are square integrable, so that the right-hand side of (3.1) is finite. Note that in this case, X has an absolutely continuous distribution with some density p.

We apply Lemma 2.2. Given $T \geq 0$ to be specified later on, let us start with the estimation of the last integral in (2.3). Define the even function $\widetilde{p}(x) = p(x) + p(-x)$, so that $p \log p \leq p \log^+ \widetilde{p}$ (where we use the notation $a^+ = \max\{a, 0\}$). Subtracting $\varphi_\alpha(x)$ from $p(x)$ and then adding, one can write

$$\int_{|x|\geq T} p(x) \log p(x)\, dx \leq \int_{|x|\geq T} p(x) \log^+ \widetilde{p}(x)\, dx$$

$$\leq \int_{-\infty}^{+\infty} |p(x) - \varphi_\alpha(x)| \log^+ \widetilde{p}(x)\, dx$$

$$+ \int_{|x|\geq T} \varphi_\alpha(x) \log^+ \widetilde{p}(x)\, dx.$$

But the function $\varphi_\alpha - \varphi$ is odd, so the last integral does not depend on α and is equal to

$$\int_{|x|\geq T} \varphi(x) \log^+ \widetilde{p}(x)\, dx. \tag{3.2}$$

To estimate it from above, one may use Cauchy's inequality together with the elementary bound $(\log^+ t)^2 \leq Ct$, where the optimal constant C is equal to $4e^{-2}$. Since $\int_{-\infty}^{+\infty} \widetilde{p}(x)\, dx = 2$, (3.2) does not exceed

$$\left(\int_{|x|\geq T} \varphi(x)^2\, dx\right)^{1/2} \left(\int_{|x|\geq T} \left(\log^+ \widetilde{p}(x)\right)^2 dx\right)^{1/2} \leq \left(\int_{|x|\geq T} \varphi(x)^2\, dx\right)^{1/2} \frac{2\sqrt{2}}{e}.$$

On the other hand,

$$\left(\int_{|x|\geq T} \varphi(x)^2\, dx\right)^{1/2} = \left(\frac{1}{\sqrt{\pi}} \left(1 - \Phi(T\sqrt{2})\right)\right)^{1/2} \leq \frac{1}{\pi^{1/4}\sqrt{2}}\, e^{-T^2/2},$$

where we applied the inequality $1 - \Phi(x) \leq \frac{1}{2}\, e^{-x^2/2}$ $(x \geq 0)$. Thus, using $\frac{2\sqrt{2}}{e} \cdot \frac{1}{\pi^{1/4}\sqrt{2}} < 1$ to simplify the constant, we get

$$\int_{|x|\geq T} p(x) \log p(x)\, dx \leq \int_{-\infty}^{+\infty} |p(x) - \varphi_\alpha(x)| \log^+ \widetilde{p}(x)\, dx + e^{-T^2/2}.$$

Here, again by the Cauchy inequality, the last integral does not exceed

$$\frac{2\sqrt{2}}{e}\left(\int_{-\infty}^{+\infty}(p(x)-\varphi_\alpha(x))^2\,dx\right)^{1/2} = \frac{2\sqrt{2}}{e}\cdot\frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{+\infty}|f(t)-g_\alpha(t)|^2\,dt\right)^{1/2},$$

where we applied Plancherel's formula. The constant in front of the last integral is smaller than $\frac{1}{2}$, so we arrive at the estimate

$$\int_{|x|\geq T} p(x)\log p(x)\,dx \;\leq\; \frac{1}{2}\,\|f-g_\alpha\|_2 + e^{-T^2/2}. \tag{3.3}$$

Now, let us turn to the next to the last integral in (2.3). Once more, subtracting $\varphi_\alpha(x)$ from $p(x)$ and then adding, one can write

$$\int_{|x|\geq T} x^2 p(x)\,dx \leq \int_{-\infty}^{\infty} x^2\,|p(x)-\varphi_\alpha(x)|\,dx + \int_{|x|\geq T} x^2\,\varphi_\alpha(x)\,dx.$$

Since the function $\varphi_\alpha - \varphi$ is odd, the last integral is equal to

$$\int_{|x|\geq T} x^2\varphi(x)\,dx \;=\; \frac{2}{\sqrt{2\pi}}\int_T^{+\infty} x^2\,e^{-x^2/2}\,dx \;=\; 2(1-\Phi(T)) + \frac{2}{\sqrt{2\pi}}\,T e^{-T^2/2}$$

(by direct integration by parts). Hence, using $2(1-\Phi(T)) \leq e^{-T^2/2}$ once more, we get

$$\begin{aligned}\frac{1}{2}\int_{|x|\geq T} x^2 p(x)\,dx &\leq \frac{1}{2}\int_{-\infty}^{+\infty} x^2\,|p(x)-\varphi_\alpha(x)|\,dx \\ &\quad + \frac{1}{2}\,e^{-T^2/2} + \frac{1}{\sqrt{2\pi}}\,T e^{-T^2/2}.\end{aligned} \tag{3.4}$$

In addition, by Cauchy's inequality,

$$\begin{aligned}\left(\int_{-\infty}^{+\infty} x^2\,|p(x)-\varphi_\alpha(x)|\,dx\right)^2 &\leq \int_{-\infty}^{+\infty}\frac{dx}{1+x^2}\int_{-\infty}^{+\infty}(1+x^2)\,x^4\,(p(x)-\varphi_\alpha(x))^2\,dx \\ &= \pi\int_{-\infty}^{+\infty}(x^4+x^6)\,(p(x)-\varphi_\alpha(x))^2\,dx\end{aligned}$$

$$\leq \pi \int_{-\infty}^{+\infty} (1+2x^6)\,(p(x)-\varphi_\alpha(x))^2\,dx.$$

But, by Plancherel's formula,

$$\int_{-\infty}^{+\infty} (p(x)-\varphi_\alpha(x))^2\,dx = \frac{1}{2\pi}\,\|f-g_\alpha\|_2^2 \tag{3.5}$$

$$\int_{-\infty}^{+\infty} x^6\,(p(x)-\varphi_\alpha(x))^2\,dx = \frac{1}{2\pi}\,\|f'''-g_\alpha'''\|_2^2. \tag{3.6}$$

Hence,

$$\int_{-\infty}^{+\infty} x^2\,|p(x)-\varphi_\alpha(x)|\,dx \leq \Big(\frac{1}{2}\,\|f-g_\alpha\|_2^2 + \|f'''-g_\alpha'''\|_2^2\Big)^{1/2}$$
$$\leq \|f-g_\alpha\|_2 + \|f'''-g_\alpha'''\|_2,$$

and from (3.4),

$$\frac{1}{2}\int_{|x|\geq T} x^2 p(x)\,dx \leq \frac{1}{2}\,e^{-T^2/2} + \frac{1}{\sqrt{2\pi}}\,T e^{-T^2/2}$$
$$+\frac{1}{2}\,\|f-g_\alpha\|_2 + \frac{1}{2}\,\|f'''-g_\alpha'''\|_2. \tag{3.7}$$

Using the bounds (3.3) and (3.7) in the inequality (2.3), we therefore obtain that

$$D(X||Z) \leq \frac{5}{2}\,e^{-T^2/2} + \frac{1}{\sqrt{2\pi}}\,T e^{-T^2/2}$$
$$+\sqrt{2\pi}\int_{-T}^{T} (p(x)-\varphi(x))^2\,e^{x^2/2}\,dx$$
$$+\|f-g_\alpha\|_2 + \|f'''-g_\alpha'''\|_2. \tag{3.8}$$

Next, let us consider the integral in (3.8). First, writing

$$p(x)-\varphi(x) = \big(p(x)-\varphi_\alpha(x)\big) + \alpha\,\frac{x^3-3x}{3!}\,\varphi(x)$$

and applying an elementary inequality $(a+b)^2 \leq \frac{a^2}{1-t} + \frac{b^2}{t}$ $(a, b \in \mathbf{R}, 0 < t < 1)$ with $t = 1/6$, we get

$$(p(x) - \varphi(x))^2 \le \frac{6}{5}\,(p(x) - \varphi_\alpha(x))^2 + \alpha^2\,\frac{(x^3-3x)^2}{6}\,\varphi(x)^2,$$

or equivalently,

$$(p(x) - \varphi(x))^2\,e^{x^2/2} \le \frac{6}{5}\,(p(x) - \varphi_\alpha(x))^2\,e^{x^2/2} + \frac{1}{\sqrt{2\pi}}\,\alpha^2\,\frac{(x^3-3x)^2}{6}\,\varphi(x).$$

Integrating this inequality over the interval $[-T, T]$ and using $\mathbf{E}\,(Z^3 - 3Z)^2 = 6$, where $Z \sim N(0, 1)$, we obtain

$$\sqrt{2\pi}\int_{-T}^{T}(p(x)-\varphi(x))^2\,e^{x^2/2}\,dx \;\le\; \frac{6}{5}\,\sqrt{2\pi}\int_{-T}^{T}(p(x)-\varphi_\alpha(x))^2\,e^{x^2/2}\,dx + \alpha^2. \tag{3.9}$$

To estimate the last integral, first note that the function $t \to e^{t/2}/(2+t)$ is increasing for $t \ge 0$. Hence, for all $|x| \le T$,

$$e^{x^2/2} = \frac{e^{x^2/2}}{2+x^2}\,(2+x^2) \le \frac{e^{T^2/2}}{2+T^2}\,(3+x^6),$$

and thus, using (3.5)–(3.6),

$$\begin{aligned}\int_{-T}^{T}(p(x)-\varphi_\alpha(x))^2\,e^{x^2/2}\,dx &\le \frac{e^{T^2/2}}{2+T^2}\int_{-T}^{T}(3+x^6)\,(p(x)-\varphi_\alpha(x))^2\,dx\\ &\le \frac{3}{2\pi}\,\frac{e^{T^2/2}}{2+T^2}\,(\|f-g_\alpha\|_2^2 + \|f'''-g_\alpha'''\|_2^2).\end{aligned}$$

Putting $\varepsilon = \|f - g_\alpha\|_2 + \|f''' - g_\alpha'''\|_2$, we therefore get from (3.9)

$$\sqrt{2\pi}\int_{-T}^{T}(p(x)-\varphi(x))^2\,e^{x^2/2}\,dx \le \frac{18}{5\sqrt{2\pi}}\,\frac{e^{T^2/2}}{2+T^2}\,\varepsilon^2 + \alpha^2.$$

Inserting this inequality in (3.8) leads to

$$D(X||Z) \;\le\; \frac{5}{2}\,e^{-T^2/2} + \frac{1}{\sqrt{2\pi}}\,Te^{-T^2/2} + \frac{18}{5\sqrt{2\pi}}\,\frac{e^{T^2/2}}{2+T^2}\,\varepsilon^2 + \varepsilon + \alpha^2. \tag{3.10}$$

It remains to optimize this bound over all $T \ge 0$. As before, consider the function $\psi(t) = e^{t/2}/(2+t)$. It is increasing for $t \ge 0$ with $\psi(0) = \frac{1}{2}$. If $0 \le \varepsilon \le 2$, define $T = T_\varepsilon$ to be the (unique) solution to the equation

$$\psi(T^2) = \frac{1}{\varepsilon}.$$

In this case,

$$Te^{-T^2/2} \cdot \frac{1}{\varepsilon} = Te^{-T^2/2} \cdot \frac{e^{T^2/2}}{2+T^2} \le \frac{1}{2},$$

so $Te^{-T^2/2} \le \frac{\varepsilon}{2}$. Furthermore, note that

$$e^{-T^2/2} \cdot \frac{1}{\varepsilon} = e^{-T^2/2} \cdot \frac{e^{T^2/2}}{2+T^2} \le \frac{1}{2},$$

so $e^{-T^2/2} \le \frac{\varepsilon}{2}$. Applying these bounds in (3.10), we arrive at

$$D(X||Z) \le \frac{5\varepsilon}{4} + \frac{1}{\sqrt{2\pi}}\,\frac{\varepsilon}{2} + \frac{18}{5\sqrt{2\pi}}\,\varepsilon + \varepsilon + \alpha^2 \le 4\varepsilon + \alpha^2,$$

which is exactly the desired inequality (3.1).

In case $\varepsilon \ge 2$, let us return to (3.8) and apply it with $T = 0$. This yields

$$D(X||Z) \le \frac{5}{2} + \varepsilon < 4\varepsilon,$$

which is even better than (3.1). Thus, Proposition 3.1 is proved. □

4 Quantile density decomposition

In order to effectively apply Propositions 2.1 and 3.1, one has to manage two different tasks. The first one is to estimate integrals such as

$$\int_{-T}^{T} |f(t) - g_\alpha(t)|^2\, dt, \qquad \int_{-T}^{T} |f'''(t) - g_\alpha'''(t)|^2\, dt$$

over sufficiently large t-intervals with properly chosen values of the parameter α. When the characteristic function f has a multiplicative structure, i.e., corresponds to the sum of a large number of small independent summands, this task can be attacked by using classical Edgeworth-type expansions (for characteristic functions). Such expansions are well-known for the non-i.i.d. case, as well, and we consider one of them in Sect. 12.

The second task concerns an estimation of integrals such as

$$\int_{|x|\ge T} |f(t)|^2\, dt, \qquad \int_{|x|\ge T} |f'''(t)|^2\, dt,$$

which in general do not need to be small or even finite. The finiteness is guaranteed, for example, when f is the Fourier transform of a bounded density p. For some purposes such as obtaining local limit theorems, it is therefore natural to restrict oneself to the case of bounded densities. For other purposes, such as an estimation of the total variation or relative entropy, the density p may be slightly modified, so that the new density, say $\widetilde{p}$, will be bounded, and at the same time will only slightly change the total variation distance or relative entropy with respect to the standard normal law.

To this aim, we shall use the so-called quantile density decomposition, based on the following elementary observation. (This decomposition will be needed regardless of whether the densities are bounded or not.)

Proposition 4.1 *Let X be a random variable with density p. Given $0 < \kappa < 1$, the real line can be partitioned into two Borel sets A_0, A_1 such that $p(x) \leq p(y)$, for all $x \in A_0$, $y \in A_1$, and*

$$\int_{A_0} p(x)\,dx = \kappa, \quad \int_{A_1} p(x)\,dx = 1 - \kappa.$$

The argument is based on the continuity of the measure $p(x)\,dx$ and is omitted.

Clearly, for some real number m_κ we get

$$A_0 \subset \{x \in \mathbf{R} : p(x) \leq m_\kappa\}, \quad A_1 \subset \{x \in \mathbf{R} : p(x) \geq m_\kappa\}.$$

Here, m_κ represents a quantile (or one of the quantiles) for the function p viewed as a random variable on the probability space $(\mathbf{R}, p(x)\,dx)$. In other words, $m_\kappa = m_\kappa(p(X))$ is a quantile of order κ for the random variable $p(X)$. If $\kappa = \frac{1}{2}$, the index is usually omitted, and then $m = m(p(X))$ denotes a median of $p(X)$.

Definition 4.2 Define the densities p_0 and p_1 to be the normalized restrictions of p to the sets A_0 and A_1, respectively. As a result, we have an equality

$$p(x) = \kappa p_0(x) + (1 - \kappa)\, p_1(x), \tag{4.1}$$

which we call the quantile density decomposition for p (respectively—the median density decomposition, when $\kappa = \frac{1}{2}$).

Let us mention one obvious, but important property of the functionals $m_\kappa(p(X))$, assuming that X has a finite second moment.

Proposition 4.3 *The functionals*

$$Q_\kappa(X) = m_\kappa(p(X))\sqrt{\mathrm{Var}(X)}$$

are affine invariant. That is, for all $a \in \mathbf{R}$ and $b \neq 0$, $Q_\kappa(a + bX) = Q_\kappa(X)$.

More precisely, let p and q denote the densities of the random variables X and $a + bX$, respectively. If $m_\kappa(p(X))$ is a specific quantile participating in the definition of $Q_\kappa(X)$, we have the relation $m_\kappa(q(a + bX)) = |b|^{-1}\, m_\kappa(p(X))$ which should be used in order to define $Q_\kappa(a + bX)$. With this agreement, $Q_\kappa(a + bX) = Q_\kappa(X)$.

5 Properties of the quantile decomposition

In this section we establish basic properties of the quantile density decomposition. Although for purposes of Theorems 1.1–1.2 the median decomposition is sufficient, the general case is no more difficult (but may be used to provide more freedom especially for improving D-dependent constants).

First, let us bound from above the quantiles $m_\kappa = m_\kappa(p(X))$ in terms of the entropic distance to normality.

Proposition 5.1 *Let X be a random variable with finite variance σ^2 $(\sigma > 0)$, having an absolutely continuous distribution, and let $0 < \kappa < 1$. Then*

$$m_\kappa \leq \frac{1}{\sigma\sqrt{2\pi}}\, e^{(D(X)+1)/(1-\kappa)}.$$

In particular,

$$m \leq \frac{1}{\sigma\sqrt{2\pi}}\, e^{2D(X)+2}.$$

Proof By Proposition 4.3, we may assume that X has mean zero and variance one. Let $A = \{x \in \mathbf{R} : p(x) \geq m_\kappa\}$. By the definition of the quantiles,

$$\int_A p(x)\,dx \geq 1 - \kappa.$$

Since $p(x) \geq m_\kappa$ on the set A, we have

$$\begin{aligned}
\int_{-\infty}^{+\infty} p(x)\log\left(1 + \frac{p(x)}{\varphi(x)}\right) dx &\geq \int_A p(x)\log\left(1 + \frac{m_\kappa}{\varphi(x)}\right) dx \\
&\geq \int_A p(x)\log\frac{m_\kappa}{\varphi(x)}\,dx \\
&= \log(m_\kappa\sqrt{2\pi})\int_A p(x)\,dx + \frac{1}{2}\int_A x^2 p(x)\,dx \\
&\geq (1-\kappa)\,\log(m_\kappa\sqrt{2\pi}).
\end{aligned}$$

On the other hand, using an elementary inequality $t\log(1+t) - t\log t \leq 1$ $(t \geq 0)$, we get

$$\int_{-\infty}^{+\infty} p(x)\log\left(1 + \frac{p(x)}{\varphi(x)}\right) dx = \int_{-\infty}^{+\infty} \frac{p(x)}{\varphi(x)}\log\left(1 + \frac{p(x)}{\varphi(x)}\right)\varphi(x)\,dx$$

$$\leq \int_{-\infty}^{+\infty} \frac{p(x)}{\varphi(x)} \log \frac{p(x)}{\varphi(x)}\, \varphi(x)\, dx + 1 \;=\; D(X) + 1.$$

Hence, $(1-\kappa)\log(m_\kappa\sqrt{2\pi}) \leq D(X) + 1$, and the proposition follows. □

Now, let V_0 and V_1 be random variables with densities p_0 and p_1 from the quantile decomposition (4.1). They have means $a_j = \mathbf{E}\,V_j$ and variances $\sigma_j^2 = \mathrm{Var}(V_j)$, connected by

$$\kappa a_0 + (1-\kappa)\, a_1 = \mathbf{E}X,$$

and

$$(\kappa a_0^2 + (1-\kappa)\, a_1^2) + (\kappa\sigma_0^2 + (1-\kappa)\,\sigma_1^2) = \mathbf{E}X^2, \tag{5.1}$$

provided that X has a finite second moment.

The next step is to prove upper bounds for the entropies of V_0 and V_1.

Proposition 5.2 *If X has mean zero and finite second moment, then*

$$\kappa D(V_0) + (1-\kappa)\, D(V_1) \leq D(X) - \kappa\log\kappa - (1-\kappa)\log(1-\kappa).$$

In particular, in case of the median decomposition,

$$D(V_0) + D(V_1) \leq 2D(X) + 2\log 2.$$

Proof Let $\mathrm{Var}(X) = \sigma^2 (\sigma > 0)$. We may assume that $D(X)$ is finite. By Definition 4.2,

$$\begin{aligned} -h(V_0) &= \int_{-\infty}^{+\infty} p_0(x)\log p_0(x)\, dx \\ &= \int_{A_0} (p(x)/\kappa)\log(p(x)/\kappa)\, dx \;=\; -\log\kappa + \frac{1}{\kappa}\int_{A_0} p(x)\log p(x)\, dx, \end{aligned}$$

and similarly, $-h(V_1) = -\log(1-\kappa) + \frac{1}{1-\kappa}\int_{A_1} p(x)\log p(x)\, dx$. Adding the two equalities with weights, we get

$$-\kappa h(V_0) - (1-\kappa)\, h(V_1) = -\kappa\log\kappa - (1-\kappa)\log(1-\kappa) - h(X). \tag{5.2}$$

Recall that

$$\begin{aligned} D(V_0) &= h(Z_0) - h(V_0), \quad \text{where } Z_0 \sim N(a_0, \sigma_0^2), \\ D(V_1) &= h(Z_1) - h(V_1), \quad \text{where } Z_1 \sim N(a_1, \sigma_1^2), \\ D(X) &= h(Z) \,-\, h(X), \quad \text{where } Z \sim N(0, \sigma^2). \end{aligned}$$

Hence, from (5.2),

$$\begin{aligned}\kappa D(V_0) + (1-\kappa)\, D(V_1) &= \kappa h(Z_0) + (1-\kappa)\, h(Z_1) \\ &\quad - \kappa \log \kappa - (1-\kappa) \log(1-\kappa) + (D(X) - h(Z)) \\ &= \kappa \log(\sigma_0 \sqrt{2\pi e}\,) + (1-\kappa)\, \log(\sigma_1 \sqrt{2\pi e}\,) \\ &\quad -\kappa \log \kappa - (1-\kappa) \log(1-\kappa) + (D(X) - \log(\sigma \sqrt{2\pi e}\,)) \\ &= -\kappa \log \kappa - (1-\kappa) \log(1-\kappa) + D(X) + \log \frac{\sigma_0^{\kappa} \sigma_1^{1-\kappa}}{\sigma}.\end{aligned}$$

Finally, by (5.1), and the arithmetic-geometric inequality,

$$\sigma_0^{2\kappa} \sigma_1^{2(1-\kappa)} \le \kappa \sigma_0^2 + (1-\kappa)\, \sigma_1^2 \le \sigma^2,$$

so, $\frac{\sigma_0^{\kappa} \sigma_1^{1-\kappa}}{\sigma} \le 1$. Proposition 5.2 is proved. □

The following bounds provide a quantitative measure in terms of $D(X)$ of non-degeneracy of the distributions of V_j via positivity of their variances σ_j^2.

Proposition 5.3 *Let X be a random variable with mean zero and variance* σ^2 ($\sigma > 0$), *having finite entropy. Then*

$$\sigma_0 > \sigma\, e^{-(D(X)+4)/\kappa}, \quad \sigma_1 > \sigma\, e^{-(D(X)+4)/(1-\kappa)}.$$

Proof By homogeneity with respect to σ, one may assume that $\sigma = 1$.

We modify the argument from the proof of Proposition 5.1. First note that

$$\begin{aligned}\log(\sigma_0 \sqrt{2\pi e}\,) &= D(V_0) - \int_{-\infty}^{+\infty} p_0(x) \log p_0(x)\, dx \\ &\ge - \int_{-\infty}^{+\infty} p_0(x) \log p_0(x)\, dx \\ &= - \int_{A_0} (p(x)/\kappa)\, \log(p(x)/\kappa)\, dx \\ &= \log \kappa - \frac{1}{\kappa} \int_{A_0} p(x)\, \log p(x)\, dx,\end{aligned} \tag{5.3}$$

where A_0 is a set from Definition 4.2.

In order to estimate the last integral, put $r(x) = e^{-a^2 x^2/2}$ with parameter $a > 0$. Using the property $r(x) \le 1$ and once more the inequality $t \log(1+t) \le t \log t + 1 (t \ge 0)$, we get

$$\begin{aligned}
\int_{A_0} p(x)\,\log p(x)\,dx &\le \int_{-\infty}^{+\infty} p(x)\,\log\left(1+\frac{p(x)}{r(x)}\right)\,dx \\
&= \int_{-\infty}^{+\infty} \frac{p(x)}{r(x)}\,\log\left(1+\frac{p(x)}{r(x)}\right)\,r(x)\,dx \\
&\le \int_{-\infty}^{+\infty} \left[\frac{p(x)}{r(x)}\,\log\frac{p(x)}{r(x)}+1\right]\,r(x)\,dx \\
&= \int_{-\infty}^{+\infty} p(x)\,\log p(x)\,dx + \frac{a^2}{2}\int_{-\infty}^{+\infty} p(x)\,x^2\,dx + \int_{-\infty}^{+\infty} r(x)\,dx \\
&= D(X) - \log(\sqrt{2\pi e}) + \left(\frac{a^2}{2} + \frac{1}{a}\,\sqrt{2\pi}\right).
\end{aligned}$$

The right-hand side is minimized for $a = (2\pi)^{1/6}$ in which case we obtain that

$$\int_{A_0} p(x)\,\log p(x)\,dx \le D(X) - \log(\sqrt{2\pi e}\,) + \frac{3}{2}\,(2\pi)^{1/3} < D(X) + 1.35.$$

Together with (5.3), the above estimate yields

$$\log(\sigma_0\sqrt{2\pi e}\,) > \log\kappa - \frac{1}{\kappa}\,(D(X) + 1.35).$$

But $\log(\sqrt{2\pi e}\,) \sim 1.42 < \frac{1.42}{\kappa}$, so $\log\sigma_0 > \log\kappa - \frac{1}{\kappa}\,(D(X)+2.77)$, or equivalently,

$$\sigma_0 > \kappa\, e^{-(D(X)+2.77)/\kappa}.$$

Finally, using $\kappa > e^{-1/\kappa}$, the above estimate may be simplified to

$$\sigma_0 > e^{-(D(X)+3.77)/\kappa},$$

which gives the first estimate on σ_0. The second estimate for σ_1 is similar. □

Note that in case of the median decomposition, Proposition 5.3 becomes

$$\sigma_0 > c\sigma\, e^{-2D(X)}, \quad \sigma_1 > c\sigma\, e^{-2D(X)},$$

where c is a positive absolute constant. One may take $c = e^{-8}$, for example.

6 Entropic bounds for cramer constants of characteristic functions

If a random variable X has an absolutely continuous distribution with density, say p, then, by the Riemann–Lebesgue theorem, its characteristic function

$$f(t) = \mathbf{E}\, e^{itX} = \int_{-\infty}^{+\infty} e^{itx}\, p(x)\, dx \quad (t \in \mathbf{R})$$

satisfies $f(t) \to 0$, as $t \to \infty$. Hence, for all $T > 0$,

$$\delta_X(T) = \sup_{|t| \geq T} |f(t)| < 1.$$

An important problem is how to quantify this separation property (that is, separation from 1) by giving explicit upper bounds on the quantity $\delta_X(T)$, sometimes called Cramer's constant. (At least $\delta_X(T) < 1$ is referred to as Cramer's condition (C).) This problem arises naturally in local limit theorems for densities of the sums of non-identically distributed independent summands (cf. e.g. [26]). Furthermore, it appears in the study of bounds and rates of convergence in the central limit theorem for strong metrics including the total variation and relative entropy. For our purposes, it is desirable to bound $\delta_X(T)$ explicitly in terms of the entropy of X or, what is more relevant, in terms of the entropic distance to normality $D(X)$. A preliminary answer may be given in terms of the variance $\sigma^2 = \mathrm{Var}(X)$, when it is finite, and where the density p is uniformly bounded.

Proposition 6.1 *Assume $p(x) \leq M$ a.e. Then, for all t real,*

$$|f(t)| \leq 1 - c\, \frac{\min\{1, \sigma^2 t^2\}}{M^2 \sigma^2}, \tag{6.1}$$

where $c > 0$ is an absolute constant.

In a slightly different form, this bound was obtained in the mid 1960s by Statulevičius [26]. He also considered more complicated quantities reflecting the behavior of the density p on non-overlapping intervals of the real line.

The inequality (6.1) can be generalized by involving non-bounded densities, but then M should be replaced by other quantities such as quantiles $m_\kappa = m_\kappa(p(X))$ of the random variable $p(X)$. One can also remove any assumption on the moments of X by replacing the standard deviation by the quantiles of the random variable $X - X'$, where X' is an independent copy of X. We refer to [6] for details, where the following bound is derived.

Proposition 6.2 *Let X be a random variable with finite variance σ^2 and finite entropy. Then, for all t real,*

$$|f(t)| \leq 1 - c\, \min\{1, \sigma^2 t^2\}\, e^{-4D(X)}, \tag{6.2}$$

where $c > 0$ is an absolute constant.

At the expense of a worse constant in the exponent, this bound can be derived directly from (6.1) by combining it with Propositions 5.1 and 5.3.

Indeed, we may assume that $\mathbf{E}X = 0$. Let V_0 and V_1 be random variables with densities p_0 and p_1 from the median decomposition (4.1), that is, for $\kappa = \frac{1}{2}$, and denote by f_0 and f_1 the corresponding characteristic functions, so that $f = \frac{1}{2} f_0 + \frac{1}{2} f_1$. Hence, for all t,

$$|f(t)| \leq \frac{1}{2} |f_0(t)| + \frac{1}{2}. \tag{6.3}$$

Since p_0 is bounded—more precisely, $p_0(x) \leq m = m(p(X))$, one can apply Proposition 6.1 to the random variable V_0 with $M = m$. Then (6.1) and (6.3) give

$$|f(t)| \leq 1 - c\, \frac{\min\{1, \sigma_0^2 t^2\}}{2m^2 \sigma_0^2},$$

where $\sigma_0^2 = \mathrm{Var}(V_0)$.

Note that $\sigma_0^2 \leq 2\sigma^2$, according to (5.1). Hence, by Proposition 5.1,

$$m^2 \sigma_0^2 \leq 2m^2 \sigma^2 \leq \frac{1}{\pi}\, e^{4D(X)+4}.$$

This gives

$$|f(t)| \leq 1 - c_1 \min\{1, \sigma_0^2 t^2\}\, e^{-4D(X)}.$$

But, by Proposition 5.3, $\sigma_0^2 > c_2 \sigma^2\, e^{-4D(X)}$, hence,

$$|f(t)| \leq 1 - c_3 \min\{1, \sigma^2 t^2\}\, e^{-8\,D(X)}$$

with some absolute constants $c_j > 0$ $(j = 1, 2, 3)$.

7 Repacking of summands

We now consider a sequence of independent (not necessarily identically distributed) random variables $X_1, \dots, X_n$ and their sum $S_n = X_1 + \dots + X_n$. Let $\mathbf{E}X_k = 0$, $\mathbf{E}X_k^2 = \sigma_k^2$ $(\sigma_k > 0)$. One may always assume without loss of generality that $\sigma_1^2 + \dots + \sigma_n^2 = 1$, so that $\mathrm{Var}(S_n) = 1$.

In addition, all X_k are assumed to have absolutely continuous distributions, having finite entropies in each place where the functional D is used.

To study integrability properties of the characteristic function f_n of S_n (more precisely—of its slightly modified variants $\widetilde{f_n}$), it will be more convenient to work with a different representation,

$$S_n = V_1 + \dots + V_N,$$

where the new independent summands represent appropriate partial sums of the X_l resulting in almost equal variances, such that at the same time the number of blocks, N, is still reasonably large. Such a representation may be introduced just by taking

$$V_k = \sum_{n_{k-1}<l\le n_k} X_l, \tag{7.1}$$

where $n_0 = 0$ and $n_k = \max\{l \le n : \sigma_1^2 + \cdots + \sigma_l^2 \le \frac{k}{N}\}$. In order that V_k have almost equal variances, the number of new summands should be restricted in terms of the parameter

$$\sigma = \max_l \sigma_l$$

which in general may be an arbitrary real number between $\frac{1}{\sqrt{n}}$ and 1.

Lemma 7.1 *If $N \le \frac{1}{2\sigma^2}$, then for each $k = 1, \dots, N$,*

$$\frac{1}{2N} < \mathrm{Var}(V_k) < \frac{2}{N}. \tag{7.2}$$

Proof If $n_1 = n$, then necessarily $N = 1$ and $V_1 = S_n$, so (7.2) holds immediately.

If $n_1 < n$, then, by the definition, $\mathrm{Var}(V_1) \le \frac{1}{N}$ and $\mathrm{Var}(V_1 + X_{n_1+1}) > \frac{1}{N}$. The latter implies $\mathrm{Var}(V_1) > \frac{1}{N} - \sigma^2 \ge \frac{1}{2N}$, thus proving (7.2) for $k = 1$.

Now, let $2 \le k \le N$. Again by the definition, $\mathrm{Var}(S_{n_k}) \le \frac{k}{N}$ and $\mathrm{Var}(S_{n_{k-1}+1}) > \frac{k-1}{N}$. The latter implies $\mathrm{Var}(S_{n_{k-1}}) > \frac{k-1}{N} - \sigma^2$. Combining the two bounds, we get

$$\mathrm{Var}(V_k) = \mathrm{Var}(S_{n_k}) - \mathrm{Var}(S_{n_{k-1}}) \le \frac{k}{N} - \left(\frac{k-1}{N} - \sigma^2\right) = \frac{1}{N} + \sigma^2 < \frac{2}{N}.$$

On the other hand,

$$\mathrm{Var}(V_k) > \left(\frac{k}{N} - \sigma^2\right) - \frac{k-1}{N} = \frac{1}{N} - \sigma^2 \ge \frac{1}{2N}.$$

Lemma 7.1 is proved. □

Thus, to obtain the property (7.2), it seems suggestive to take $N = [\frac{1}{2\sigma^2}]$ (the integer part). However, this choice is not used in the Proof of Theorems 1.1–1.2, since we need to express N as a suitable function of Lyapunov's coefficients.

As another useful property of the representation (7.1), let us mention the following.

Lemma 7.2 *If $\max_{l\le n} D(X_l) \le D$, then $\max_{k\le N} D(V_k) \le D$, as well.*

This is due to the general bound $D(X+Y) \le \max\{D(X), D(Y)\}$, which holds for arbitrary independent random variables with finite second moments and absolutely continuous distributions. It can easily be derived, for example, from the entropy power inequality

$$e^{2h(X+Y)} \geq e^{2h(X)} + e^{2h(Y)},$$

cf. [10].

Now, let ρ_k denote density of the random variable V_k. For each ρ_k, one may consider a median density decomposition

$$\rho_k(x) = \frac{1}{2}\,\rho_{k0}(x) + \frac{1}{2}\,\rho_{k1}(x) \tag{7.3}$$

in accordance with Definition 4.2 for the parameter $\kappa = \frac{1}{2}$.

In particular, $\rho_{k0}(x) \leq m$, where $m = m(\rho_k(V_k))$ is a median of the random variable $\rho_k(V_k)$. Note that by Proposition 5.1 with $X = V_k$ and Lemmas 7.1–7.2, if $\max_{j \leq n} D(X_j) \leq D$, we immediately obtain that

$$m(\rho_k(V_k)) \,\leq\, \frac{1}{v_k\sqrt{2\pi}}\, e^{2D+2} \,\leq\, \sqrt{N}\, e^{2D+2}, \tag{7.4}$$

where $v_k = \sqrt{\mathrm{Var}(V_k)}$.

Let V_{kj} be random variables with densities ρ_{kj} and characteristic functions

$$\hat{\rho}_{kj}(t) = \mathbf{E}\, e^{itV_{kj}} = \int_{-\infty}^{+\infty} e^{itx}\, \rho_{kj}(x)\, dx, \quad j = 0, 1.$$

We collect their basic properties in the following lemma.

Lemma 7.3 *Assume that* $N \leq \frac{1}{2\sigma^2}$ *and* $\max_{l \leq n} D(X_l) \leq D$. *For all* $k \leq N$ *and* $j = 0, 1$,

a) $D(V_{kj}) \leq 2D + 2$,
b) $\mathrm{Var}(V_{kj}) > \frac{1}{2N}\, e^{-4(D+4)}$,
c) $|\hat{\rho}_{kj}(t)| \leq 1 - c\, e^{-12\,D}$ *for all* $|t| \geq \sqrt{N}$ *with an absolute constant* $c > 0$.

Proof The first assertion follows from Lemma 7.2 and Proposition 5.2 applied with $X = V_k$. For the second one, combine Proposition 5.3 with $X = V_k$ and Lemmas 7.1–7.2 to get

$$v_{kj} > v_k\, e^{-2(D(V_k)+4)} \geq v_k\, e^{-2(D+4)} \geq \frac{1}{\sqrt{2N}}\, e^{-2(D+4)},$$

where $v_{kj}^2 = \mathrm{Var}(V_{kj})$ ($v_{kj} > 0$). For the assertion in c), combine Proposition 6.2 for $X = V_{kj}$ and the previous steps, which give

$$\begin{aligned} |\hat{\rho}_{kj}(t)| &\le 1 - c\, \min\{1, v_{kj}^2 t^2\}\, e^{-4D(V_{kj})} \\ &\le 1 - c\, \min\{1, e^{-4(D+4)}\, t^2/(2N)\}\, e^{-4(2D+2)} \\ &\le 1 - c'\, \min\{1, t^2/N\}\, e^{-12\,D} \end{aligned}$$

with some absolute constants $c, c' > 0$. □

8 Decomposition of convolutions

Starting from the representation $S_n = V_1 + \cdots + V_N$ with the summands defined in (7.1), one can write the density of S_N as the convolution

$$p_n = \rho_1 * \cdots * \rho_N,$$

where ρ_k denotes the density of V_k. Moreover, a direct application of the median decomposition (7.3) leads to the representation

$$p_n = 2^{-N} \sum (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N}),$$

where the summation is carried out over all 2^N sequences δ_k with values 0 and 1, and with the convention that

$$\rho_{k0}^{\delta_k} * \rho_{k1}^{1-\delta_k} = \begin{cases} \rho_{k0}, & \text{if } \delta_k = 1, \\ \rho_{k1}, & \text{if } \delta_k = 0. \end{cases}$$

Let an integer number $m_0 \ge 0$ be given (For our purposes, since we will need to control 3 derivatives in Proposition 3.1, one may take $m_0 = 3$). For $N \ge m_0 + 1$, we split the above sum into the two parts, so that

$$p_n = q_{n0} + q_{n1},$$

where

$$q_{n0} = 2^{-N} \sum_{\delta_1 + \cdots + \delta_N > m_0} (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N}),$$

$$q_{n1} = 2^{-N} \sum_{\delta_1 + \cdots + \delta_N \le m_0} (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N}).$$

Put

$$\varepsilon_n = \int_{-\infty}^{+\infty} q_{n1}(x)\,dx = 2^{-N} \sum_{k=0}^{m_0} \frac{N!}{k!\,(N-k)!}.$$

One can easily see that

$$\varepsilon_n \leq 2^{-(N-1)}\, N^{m_0}. \tag{8.1}$$

Definition 8.1 Put

$$\widetilde{p}_n(x) = p_{n0}(x) = \frac{1}{1-\varepsilon_n}\, q_{n0}(x), \tag{8.2}$$

and similarly $p_{n1}(x) = \frac{1}{\varepsilon_n}\, q_{n1}(x)$. Thus, we get the decomposition

$$p_n(x) = (1-\varepsilon_n)\, p_{n0}(x) + \varepsilon_n\, p_{n1}(x). \tag{8.3}$$

Accordingly, introduce the associated characteristic functions

$$\widetilde{f}_n(t) = f_{n0}(t) = \int_{-\infty}^{+\infty} e^{itx}\, \widetilde{p}_n(x)\,dx, \quad f_{n1}(t) = \int_{-\infty}^{+\infty} e^{itx}\, p_{n1}(x)\,dx.$$

The probability densities $\widetilde{p}_n(x) = p_{n0}(x)$ are bounded and provide a strong approximation for $p_n(x)$. Indeed, from (8.3) it follows that

$$|\widetilde{p}_n(x) - p_n(x)| = \varepsilon_n |p_{n0}(x) - p_{n1}(x)| \tag{8.4}$$

which together with the bound (8.1) immediately implies:

Proposition 8.2 *For all $n \geq N \geq m_0 + 1$,*

$$\int_{-\infty}^{+\infty} |\widetilde{p}_n(x) - p_n(x)|\,dx \leq 2^{-(N-2)}\, N^{m_0}.$$

In particular, the corresponding characteristic functions satisfy, for all $t \in \mathbf{R}$,

$$|\widetilde{f}_n(t) - f_n(t)| \leq 2^{-(N-2)}\, N^{m_0}. \tag{8.5}$$

Note that that Proposition 8.2 uses an absolute continuity of distributions of X_k, only (for the construction of $\widetilde{p}_n$ and $\widetilde{f}_n$), and does not need any moment assumption.

To obtain a bound for the derivatives of characteristic functions similar to (8.5), we involve basic hypotheses $\mathbf{E}X_k = 0, \mathbf{E}X_k^2 < +\infty$, assuming that the sum $S_n =$

$X_1+\cdots+X_n$ has the second moment $\mathbf{E}S_n^2 = 1$. We shall use the associated Lyapunov ratios, thus given by

$$L_s = \sum_{k=1}^n \mathbf{E}\,|X_k|^s.$$

Our basic tool will be Rosenthal's inequality

$$\mathbf{E}\,|S_n|^s \le C_s \left(1 + \sum_{j=1}^n \mathbf{E}\,|X_j|^s\right) = C_s\,(1+L_s), \quad s \ge 2, \tag{8.6}$$

which holds true with some constants C_s, depending on s, only (cf. e.g. [20,24]). Note that in case $1 \le s \le 2$, there is also an obvious bound $\mathbf{E}\,|S_n|^s \le 1$.

Proposition 8.3 *Assume that L_s is finite $(s \ge 2)$. For all $n \ge N \ge m_0 + 1$,*

$$\int_{-\infty}^{+\infty} |x|^s\,|\widetilde{p}_n(x) - p_n(x)|\,dx \;\le\; C_s(1+L_s)\,2^{-(N-3)}\,N^{m_0+s}.$$

In particular, if s is an integer, the sth derivative of the corresponding characteristic functions satisfies, for all t real,

$$|\widetilde{f}_n^{(s)}(t) - f_n^{(s)}(t)| \;\le\; C_s(1+L_s)\,2^{-(N-3)}\,N^{m_0+s}.$$

Here, the constant C_s is the same as in (8.6). For the values $s = 1$ and $s = 2$, it is better to use $\mathbf{E}\,|S_n| \le 1$ and $\mathbf{E}S_n^2 = 1$ instead of (8.6). For $s = 3$, Rosenthal's inequality can be shown to hold with constant $C_3 = 2$. Hence, we obtain:

Corollary 8.4 *Let $n \ge N \ge m_0 + 1$ and $t \in \mathbf{R}$. Then, for $s = 1, 2$, we have*

$$|\widetilde{f}_n^{(s)}(t) - f_n^{(s)}(t)| \;\le\; 2^{-(N-3)}\,N^{m_0+s}.$$

Moreover, if L_3 is finite,

$$|\widetilde{f}_n'''(t) - f_n'''(t)| \;\le\; (1+L_3)\,2^{-(N-4)}\,N^{m_0+3}.$$

Proof of Proposition 8.3 Let V_{kj} $(1 \le k \le N,\, j = 0, 1)$ be independent random variables with respective densities ρ_{kj} from the median decomposition (7.3) for the random variables V_k. For each sequence $\delta = (\delta_k)_{1\le k\le N}$ with values 0 and 1, the convolution

$$\rho^{(\delta)} = (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N})$$

represents the density of the sum

$$S(\delta) = \sum_{k=1}^{N} (\delta_k V_{k0} + (1 - \delta_k) V_{k1}).$$

By the assumption, all moments $\mathbf{E}\,|X_k|^s$ are finite, and (7.3) yields

$$\mathbf{E}\,|V_k|^s = \frac{1}{2}\,\mathbf{E}\,|V_{k0}|^s + \frac{1}{2}\,\mathbf{E}\,|V_{k1}|^s. \tag{8.7}$$

Hence, for the L^s-norm $\|S(\delta)\|_s = (\mathbf{E}\,|S(\delta)|^s)^{1/s}$, using the Minkowski inequality, we have

$$\begin{aligned} \|S(\delta)\|_s &\leq \sum_{k=1}^{N} \|\delta_k V_{k0} + (1 - \delta_k) V_{k1}\|_s \\ &\leq \sum_{k=1}^{N} (\delta_k\, \|V_{k0}\|_s + (1 - \delta_k)\, \|V_{k1}\|_s) \\ &\leq 2^{1/s} \sum_{k=1}^{N} \|V_k\|_s, \end{aligned} \tag{8.8}$$

where (8.7) was used in the last step. But

$$\frac{1}{N} \sum_{k=1}^{N} \|V_k\|_s = \frac{1}{N} \sum_{k=1}^{N} (\mathbf{E}\,|V_k|^s)^{1/s} \leq \left(\frac{1}{N} \sum_{k=1}^{N} \mathbf{E}\,|V_k|^s\right)^{1/s}, \tag{8.9}$$

so

$$\mathbf{E}\,|S(\delta)|^s \leq 2N^{s-1} \sum_{k=1}^{N} \mathbf{E}\,|V_k|^s \leq 2N^s\,\mathbf{E}\,|S_n|^s,$$

where we used $\mathbf{E}\,|V_k|^s \leq \mathbf{E}\,|S_n|^s$ (due to Jensen's inequality).

Write $\mathbf{E}\,|S(\delta)|^s = \int_{-\infty}^{+\infty} |x|^s\,\rho^{(\delta)}(x)\,dx$. Recalling the definition of q_{nj} and ε_n, we get

$$\begin{aligned} \int_{-\infty}^{+\infty} |x|^s\, q_{n0}(x)\,dx &= 2^{-N} \sum_{\delta_1 + \dots + \delta_N > m_0} \mathbf{E}\,|S(\delta)|^s \;\leq\; 2\,\mathbf{E}\,|S_n|^s\,(1 - \varepsilon_n)\,N^s, \\ \int_{-\infty}^{+\infty} |x|^s\, q_{n1}(x)\,dx &= 2^{-N} \sum_{\delta_1 + \dots + \delta_N \leq m_0} \mathbf{E}\,|S(\delta)|^s \;\leq\; 2\,\mathbf{E}\,|S_n|^s\,\varepsilon_n N^s. \end{aligned}$$

Hence, by the definition of p_{n0},

$$\int_{-\infty}^{+\infty} |x|^s \, p_{n0}(x)\,dx \ \le\ 2\,\mathbf{E}\,|S_n|^s\,N^s,$$

and similarly for p_{n1}. But, from (8.4),

$$|x|^s\,|\widetilde{p}_n(x) - p_n(x)| \le \varepsilon_n |x|^s\,(p_{n0}(x) + p_{n1}(x)),$$

so, applying (8.1),

$$\int_{-\infty}^{+\infty} |x|^s\,|\widetilde{p}_n(x) - p_n(x)|\,dx \ \le\ \mathbf{E}\,|S_n|^s\,2^{-(N-3)}N^{m_0+s}.$$

It remains to apply (8.6). □

9 Entropic approximation of p_n by $\tilde{p}_n$

As before, let $X_1, \dots, X_n$ be independent random variables with $\mathbf{E}X_k = 0$, $\mathbf{E}X_k^2 = \sigma_k^2$ ($\sigma_k > 0$), such that $\sigma_1^2 + \dots + \sigma_n^2 = 1$. Moreover, let X_k have absolutely continuous distributions with finite entropies, and let p_n denote the density of the sum

$$S_n = X_1 + \dots + X_n.$$

Put $\sigma^2 = \max_k \sigma_k^2$.

The next step is to extend the assertion of Propositions 8.2–8.3 to relative entropies with respect to the standard normal distribution on the real line with density

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

Thus put

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

Recall that the modified densities $\widetilde{p}_n$ are constructed in Definition 8.1 with arbitrary integers $0 \le m_0 < N \le n$ on the basis of the representation (7.1), based on the independent random variables V_k and the median decomposition (7.3) for the densities ρ_k of V_k.

Proposition 9.1 *Let $D = \max_k D(X_k)$. Given that $m_0 + 1 \le N \le \frac{1}{2\sigma^2}$, we have*

$$|\widetilde{D}_n - D_n| < 2^{-(N-6)}\,N^{m_0+1}\,(D+1). \tag{9.1}$$

We shall use a few elementary properties of the convex function $L(u) = u \log u$ $(u \geq 0)$.

Lemma 9.2 *For all* $u, v \geq 0$ *and* $0 \leq \varepsilon \leq 1$,

a) $L((1-\varepsilon)\,u + \varepsilon v) \leq (1-\varepsilon)\,L(u) + \varepsilon L(v)$;
b) $L((1-\varepsilon)\,u + \varepsilon v) \geq (1-\varepsilon)\,L(u) + \varepsilon L(v) + uL(1-\varepsilon) + vL(\varepsilon)$.

Proof of Proposition 9.1 Define

$$D_{nj} = \int p_{nj}(x) \log \frac{p_{nj}(x)}{\varphi(x)}\,dx \quad (j = 0, 1),$$

so that $\widetilde{D}_n = D_{n0}$, where the densities p_{nj} have been defined in (8.2)–(8.3).

By Lemma 9.2 *a*), $D_n \leq (1-\varepsilon_n)D_{n0} + \varepsilon_n D_{n1}$. On the other hand, by Lemma 9.2 *b*),

$$D_n \geq ((1-\varepsilon_n)D_{n0} + \varepsilon_n D_{n1}) + \varepsilon_n \log \varepsilon_n + (1-\varepsilon_n)\log(1-\varepsilon_n).$$

The two estimates give

$$|\widetilde{D}_n - D_n| \leq \varepsilon_n(D_{n0} + D_{n1}) - \varepsilon_n \log \varepsilon_n - (1-\varepsilon_n)\log(1-\varepsilon_n). \tag{9.2}$$

Hence, we need to give appropriate bounds on both D_{n0} and D_{n1}.

To this aim, as before, let V_{kj} $(1 \leq k \leq N,\ j = 0, 1)$ be independent random variables with respective densities ρ_{kj} from the median decomposition (7.3) for V_k, and put $v_{kj}^2 = \mathrm{Var}(V_{kj})$. As in the previous section, for each sequence $\delta = (\delta_k)_{1\leq k\leq N}$ with values 0 and 1, consider the convolution

$$\rho^{(\delta)} = (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N}),$$

i.e., the densities of the random variables

$$S(\delta) = \sum_{k=1}^N (\delta_k V_{k0} + (1-\delta_k)V_{k1}).$$

By convexity of the function $u \log u$,

$$D_{n1} \ \leq \ \frac{1}{\varepsilon_n}\, 2^{-N} \sum_{\delta_1+\cdots+\delta_N \leq m_0} \int_{-\infty}^{+\infty} \rho^{(\delta)}(x)\ \log \frac{\rho^{(\delta)}(x)}{\varphi(x)}\,dx, \tag{9.3}$$

$$D_{n0} \ \leq \ \frac{1}{1-\varepsilon_n}\, 2^{-N} \sum_{\delta_1+\cdots+\delta_N > m_0} \int_{-\infty}^{+\infty} \rho^{(\delta)}(x)\ \log \frac{\rho^{(\delta)}(x)}{\varphi(x)}\,dx. \tag{9.4}$$

In general, if S denotes a random variable with variance v^2 ($v > 0$) having density ρ, and if Z is a standard normal random variable, the relative entropy of S with respect to Z is connected with the entropic distance to normality $D(S)$ by the simple formula

$$D(S||Z) = \int \rho(x) \log \frac{\rho(x)}{\varphi(x)}\, dx = D(S) + \log \frac{1}{v} + \frac{\mathbf{E}S^2 - 1}{2}. \tag{9.5}$$

In the case $S = S(\delta)$, applying Lemma 7.3 *b*), we have

$$v^2 = \sum_{k=1}^{N} [\,\delta_k v_{k0}^2 + (1-\delta_k)\, v_{k1}^2] \geq \frac{1}{2}\, e^{-4(D+4)},$$

hence

$$\log \frac{1}{v} \leq 2D + 9.$$

In addition, by (8.8)–(8.9) in the particular case $s = 2$, and using $\sum_{k=1}^N \mathrm{Var}(V_k) = \mathrm{Var}(S_n) = 1$, we have $\mathbf{E}S(\delta)^2 \leq 2N$. Therefore, for the random variable $S = S(\delta)$ we obtain from (9.5)

$$D(S(\delta)||Z) \leq D(S(\delta)) + (2D + 9) + N. \tag{9.6}$$

The remaining term, $D(S(\delta))$, can be estimated by virtue of the same general inequality $D(X+Y) \leq \max\{D(X), D(Y)\}$ mentioned after Lemma 7.2. This bound can be applied to all summands of $S(\delta)$, which together with Lemma 7.3 *a*) gives

$$D(S(\delta)) \leq \max_{1 \leq k \leq N} \max\{D(V_{k0}), D(V_{k1})\} \leq 2D + 2.$$

Applying this in (9.6), we arrive at

$$\int_{-\infty}^{+\infty} \rho^{(\delta)}(x) \log \frac{\rho^{(\delta)}(x)}{\varphi(x)}\, dx = D(S(\delta)||Z) \leq 4D + 11 + N.$$

Finally, by (9.3)–(9.4), we have similar bounds for D_{n0} and D_{n1}, namely,

$$D_{n0} \leq 4D + 11 + N, \quad D_{n1} \leq 4D + 11 + N.$$

Having obtained these estimates, we are prepared to return to (9.2), which thus gives

$$|\widetilde{D}_n - D_n| \leq 2\varepsilon_n (4D + 11 + N) + \varepsilon_n \log \frac{1}{\varepsilon_n} + (1-\varepsilon_n) \log \frac{1}{1-\varepsilon_n}. \tag{9.7}$$

To simplify this bound, consider the function $H(\varepsilon) = \varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon}$, which is defined for $0 \le \varepsilon \le 1$, is concave and symmetric about the point $\frac{1}{2}$, where it attains its maximum $H(\frac{1}{2}) = \log 2$. Recall (8.1), that is, $\varepsilon_n \le d_n = 2^{-(N-1)}\, N^{m_0}$.

If $d_n \ge \frac{1}{2}$, then

$$H(\varepsilon_n) \le \log 2 \le 2d_n = 2^{-(N-2)}\, N^{m_0}. \tag{9.8}$$

Note that

$$\log \frac{1}{d_n} = m_0 \log \frac{1}{N} + (N-1) \log 2 < N.$$

Hence, in the other case $d_n \le \frac{1}{2}$, we have

$$H(\varepsilon_n) \le H(d_n) \le 2d_n \log \frac{1}{d_n} \le 2^{-(N-2)}\, N^{m_0+1}. \tag{9.9}$$

Comparing (9.8) and (9.9), we see that they can be combined to the following estimate

$$H(\varepsilon_n) \le 2^{-(N-2)}\, N^{m_0+1},$$

which is valid regardless of whether d_n is greater or smaller than $\frac{1}{2}$.

Using this estimate in (9.7), we finally get

$$\begin{aligned} |\widetilde{D}_n - D_n| &\le 2^{-(N-2)}\, N^{m_0}\, (4D + 11 + N) + 2^{-(N-2)}\, N^{m_0+1} \\ &= 2^{-(N-2)}\, N^{m_0}\, (4D + 11 + 2N). \end{aligned}$$

Since $4D + 11 + 2N < 2^4\, N(D+1)$, we arrive at the desired inequality (9.1). □

10 Integrability of characteristic functions $\tilde{f}_n$ and their derivatives

Now we turn to the question of quantitative bounds for the modified characteristic functions $\widetilde{f}_n$ in terms of the maximal entropic distance to normality

$$D = \max_{k \le n} D(X_k).$$

Again, let $X_1, \dots, X_n$ be independent random variables with $\mathbf{E}X_k = 0$, $\mathbf{E}X_k^2 = \sigma_k^2$ ($\sigma_k > 0$), such that $\sigma_1^2 + \dots + \sigma_n^2 = 1$. Moreover, all X_k are assumed to have absolutely continuous distributions with finite entropies.

We assume that the modified density $\widetilde{p}_n$ and its characteristic function $\widetilde{f}_n$ have been constructed for arbitrary integers $m_0 + 1 \le N \le n$. Put $\sigma = \max_k \sigma_k$.

Proposition 10.1 *If* $m_0 \geq 1$ *and* $m_0 + 1 \leq N \leq \frac{1}{2\sigma^2}$, *then*

$$\int\limits_{|t| \geq \sqrt{N}} |\widetilde{f}_n(t)|^2\, dt \leq C\sqrt{N}\, e^{-cN} \tag{10.1}$$

with some positive constants C and c, depending on D, only.

In fact, one can choose the constants to be of the form $C = e^{2D+4}$ and $c = c_0 e^{-12\, D}$, where c_0 is a positive absolute factor.

Proof Consider any convolution

$$\rho = (\rho_{10}^{\delta_1} * \rho_{11}^{1-\delta_1}) * \cdots * (\rho_{N0}^{\delta_N} * \rho_{N1}^{1-\delta_N})$$

participating in the definition of q_{n0}, that is, with $\delta_1 + \cdots + \delta_N > m_0$. It has the characteristic function

$$\hat{\rho}(t) = \int_{-\infty}^{+\infty} e^{itx} \rho(x)\, dx = \prod_{k=1}^{N} \hat{\rho}_{k0}(t)^{\delta_k}\, \hat{\rho}_{k1}(t)^{1-\delta_k}, \tag{10.2}$$

where $\hat{\rho}_{kj}$ denote the characteristic functions of the random variables V_{kj} from the median decomposition (4.1) with $X = V_k$ $(1 \leq k \leq N,\ j = 0, 1)$. In every such convolution there are at least $m_0 + 1$ terms ρ_{k0} for which $\delta_k = 1$. Without loss of generality, let $k = N$ be one of them, so that $\delta_N = 1$. Then, we may write

$$\hat{\rho}(t) = \hat{\rho}_{N0}(t) \prod_{k=1}^{N-1} \hat{\rho}_{k0}(t)^{\delta_k}\, \hat{\rho}_{k1}(t)^{1-\delta_k}. \tag{10.3}$$

By Lemma 7.3 *c*), and using the inequality $1 - x \leq e^{-x}$ $(x \in \mathbf{R})$, we get for all $|t| \geq \sqrt{N}$,

$$|\hat{\rho}_{kj}(t)| \leq \exp\{-c_0 e^{-12\, D}\} \tag{10.4}$$

with some absolute constant $c_0 > 0$. Inserting this in (10.3) and using $N \geq 2$ leads to

$$|\hat{\rho}(t)|^2 \ \leq\ A\, |\hat{\rho}_{N0}(t)|^2, \quad A = \exp\{-c_0 e^{-12\, D} N\}, \tag{10.5}$$

where $c_0 > 0$ is a different absolute constant.

Now, integrate (10.5) over the region $|t| \geq \sqrt{N}$ and use Plancherel's formula. Applying the property $\rho_{N0}(x) \leq m = m(\rho_N(V_N))$, we get

$$\int\limits_{|t| \geq \sqrt{N}} |\hat{\rho}(t)|^2\, dt \leq A \int_{-\infty}^{+\infty} |\hat{\rho}_{N0}(t)|^2\, dt = 2\pi A \int_{-\infty}^{+\infty} \rho_{N0}(x)^2\, dx \leq 2\pi A\, m. \tag{10.6}$$

But, as noted in (7.4), we have $m \le e^{2D+2}\sqrt{N}$, so together with $2\pi < e^2$ (10.6) gives the desired bound

$$\int\limits_{|t|\ge\sqrt{N}} |\hat{\rho}(t)|^2\,dt \;\le\; e^{2D+4}\sqrt{N}\,e^{-cN} \quad (c = c_0 e^{-12\,D})$$

for $\hat{\rho}$. But $\widetilde{f}_n$ is a finite convex combination of such functions, and (10.1) immediately follows. □

Next, we shall extend Proposition 10.1 to the derivatives of $\widetilde{f}_n$, which are needed up to order $s = 3$ in case of finite 4th moments of X_k. Assume that $s \ge 1$ is an arbitrary integer.

Consider the characteristic functions $\hat{\rho}$ in (10.2). Recall that $\widetilde{f}_n$ represents a convex combination of such characteristic functions over all sequences $\delta = (\delta_1, \dots, \delta_N)$ such that $\delta_1 + \cdots + \delta_N \ge m_0 + 1$. Hence, it will be sufficient to derive an estimate, such as (10.1), for any admissible fixed sequence δ.

Put

$$u_k = \hat{\rho}_{k0}^{\delta_k}\,\hat{\rho}_{k1}^{1-\delta_k} \quad (1 \le k \le N),$$

which is the characteristic function of the random variable $\delta_k V_{k0} + (1-\delta_k)\,V_{k1}$.

Thus, $\hat{\rho} = \prod_{k=1}^N u_k$. For the sth derivative of the product we write a general polynomial formula

$$\hat{\rho}^{(s)} = \sum \binom{s}{s_1 \dots s_N} u_1^{(s_1)} \dots u_N^{(s_N)},$$

where the summation runs over all integer numbers $s_1, \dots, s_N \ge 0$, such that $s_1 + \cdots + s_N = s$.

Fix such a sequence $s_1, \dots, s_N$. Note that it contains at most s non-zero terms. The sequence $\delta = (\delta_1, \dots, \delta_N)$ defining ρ satisfies $\delta_1 + \cdots + \delta_N \ge m_0 + 1$. Hence, in the row $u_1^{(s_1)}, \dots, u_N^{(s_N)}$ there are at least m_0+1 terms corresponding to $\delta_k = 1$. Therefore, if $m_0 \ge s$, there is at least one index, say k, for which $\delta_k = 1$ and in addition $s_k = 0$. For simplicity, let $k = N$, so that

$$\psi \equiv u_1^{(s_1)} \dots u_N^{(s_N)} = \hat{\rho}_{N0}\,u_1^{(s_1)} \dots u_{N-1}^{(s_{N-1})}. \tag{10.7}$$

If $s_k > 0$, then

$$|u_k^{(s_k)}(t)| \le \mathbf{E}\,|\delta_k V_{k0} + (1-\delta_k)\,V_{k1}|^{s_k} \le \max\{\mathbf{E}\,|V_{k0}|^{s_k}, \mathbf{E}\,|V_{k1}|^{s_k}\}.$$

But, by the decomposition (7.3) and Jensen's inequality,

$$\frac{1}{2}\,\mathbf{E}\,|V_{k0}|^{s_k} + \frac{1}{2}\,\mathbf{E}\,|V_{k1}|^{s_k} = \mathbf{E}\,|V_k|^{s_k} \le \mathbf{E}\,|S_n|^{s_k},$$

so $|u_k^{(s_k)}(t)| \le 2\,\mathbf{E}\,|S_n|^{s_k}$. Hence,

$$\prod_{s_k>0} |u_k^{(s_k)}(t)| \ \le\ 2^s \prod_{s_k>0} \mathbf{E}\,|S_n|^{s_k} \ \le\ 2^s \prod_{s_k>0} (\mathbf{E}\,|S_n|^s)^{s_k/s} = 2^s\,\mathbf{E}\,|S_n|^s. \quad (10.8)$$

When $s_k = 0$, we apply the estimate (10.4) on Cramer's constants, which may be used in (10.7). Note that (10.4) is fulfilled for at least $(N-1)-(s-1) \ge N - m_0$ indices $k \le N-1$. Hence, using also (10.8), we get

$$|\psi(t)| \le C\,|\hat\rho_{N0}(t)|\,\exp\{-c_0(N-m_0)\,e^{-12\,D}\}, \quad C = 2^s\,\mathbf{E}\,|S_n|^s.$$

In case $N \ge 2m_0$, one may simplify this bound by writing $N - m_0 \ge \frac{N}{2}$. In addition, since the sum of the multinomial coefficients in the representation of $\hat\rho^{(s)}$ is equal to N^s, and using Jensen's inequality for the quadratic function, we arrive at

$$|\hat\rho^{(s)}(t)|^2 \le\ A\,|\hat\rho_{N0}(t)|^2, \quad A = CN^s\exp\{-c_0 e^{-12\,D} N\},$$

with some absolute constant $c_0 > 0$. It remains to integrate this inequality like in (10.6) over the region $|t| \ge \sqrt{N}$ and apply the estimate (7.4). As a result, we obtain

$$\int\limits_{|t|\ge\sqrt{N}} |\hat\rho^{(s)}(t)|^2\,dt \le Ae^{2D+4}\sqrt{N}.$$

Since $\widetilde{f}_n$ is a convex combination of the functions $\hat\rho^{(s)}$, a similar inequality holds for $\widetilde{f}_n(t)$, as well. That is,

$$\int\limits_{|t|\ge\sqrt{N}} |\widetilde{f}_n^{(s)}(t)|^2\,dt \ \le\ 2^s\,\mathbf{E}\,|S_n|^s\,e^{2D+4}\,\exp\{-c_0e^{-12\,D}N\}\,N^{s+1/2}.$$

For $s=1$ and $s=2$, we have $\mathbf{E}\,|S_n|^s \le 1$, while for $s \ge 3$, one may use Rosenthal's inequality (8.6). In particular, for $s=3$ it gives $\mathbf{E}\,|S_n|^3 \le 2(1+L_3)$.

Summarizing the results obtained so far, we have:

Proposition 10.2 *Let $m_0 \ge 3$ and $2m_0 \le N \le \frac{1}{2\sigma^2}$. Then*

$$\int\limits_{|t|\ge\sqrt{N}} |\widetilde{f}_n^{(s)}(t)|^2\,dt \le CN^{s+1/2}\,e^{-cN} \quad (s=1,2) \quad (10.9)$$

with positive constants C and c, depending on D, only. Moreover, if L_s is finite, $s \ge 3$ integer, and $m_0 \ge s$, then

$$\int\limits_{|t|\ge\sqrt{N}} |\widetilde{f}_n^{(s)}(t)|^2\,dt \le C \cdot C_s(1+L_s)\,N^{s+1/2}\,e^{-cN}.$$

Here, the constants $C = e^{2D+4}$ and $c = c_0 e^{-12\,D}$ are of the same form as in Proposition 10.1, and C_s is a constant in Rosenthal's inequality (8.6). In particular, for $s = 3$, we arrive at

$$\int\limits_{|t| \geq \sqrt{N}} |\widetilde{f}_n'''(t)|^2 \, dt \leq C(1 + L_3)\, N^{7/2}\, e^{-cN}. \tag{10.10}$$

Note also that, for $s = 0$, (10.9) is true, as well, and returns us to Proposition 10.1.

11 Proof of Theorem 1.1 and its refinement

We are now ready to complete the proof of Theorems 1.1–1.2 and develop some refinements. Thus, let $X_1, \dots, X_n$ be independent random variables with mean zero and finite third absolute moments, having finite entropies, and such that the sum $S_n = X_1 + \cdots + X_n$ has variance $\mathrm{Var}(S_n) = 1$. The relevant quantity in our bounds will be the Lyapunov coefficient

$$L_3 = \sum_{k=1}^{n} \mathbf{E}\,|X_k|^3$$

and the maximal entropic distance to normality $D = \max_k D(X_k)$.

To bound the total variation distance $\|F_n - \Phi\|_{\mathrm{TV}}$ from the distribution F_n of S_n to the standard normal law Φ, one may apply the general bound (2.1) of Proposition 2.1. However, it is only applicable when the characteristic function f_n of S_n and its derivative are square integrable. But even in the case that, for example, each density p_n of S_n is bounded individually, we still could not properly bound the maximum of the convolutions of these densities explicitly in terms of D and L_3. That is why, we are forced to consider modified forms of p_n.

Thus, consider these modifications $\widetilde{p}_n$ together with their Fourier transforms $\widetilde{f}_n$ described in Definition 8.1. By the triangle inequality,

$$\|F_n - \Phi\|_{\mathrm{TV}} \leq \|\widetilde{F}_n - \Phi\|_{\mathrm{TV}} + \|\widetilde{F}_n - F_n\|_{\mathrm{TV}}, \tag{11.1}$$

where $\widetilde{F}_n$ denotes the distribution with density $\widetilde{p}_n$.

In the construction of $\widetilde{p}_n$ it suffices to take the values $m_0 = 3$ and $6 \leq N \leq \frac{1}{2\sigma^2}$. Then, by Proposition 8.2,

$$\|\widetilde{F}_n - F_n\|_{\mathrm{TV}} \;=\; \int\limits_{-\infty}^{+\infty} |\widetilde{p}_n(x) - p_n(x)|\, dx \;\leq\; 2^{-(N-2)}\, N^3. \tag{11.2}$$

This gives a sufficiently good bound on the last term in (11.1), if N is sufficiently large.

The first term on the right-hand side of (11.1) can be bounded by virtue of (2.1), which gives

$$\|\widetilde{F}_n - \Phi\|_{\mathrm{TV}}^2 \le \frac{1}{2}\,\|\widetilde{f}_n - g\|_2^2 + \frac{1}{2}\,\|(\widetilde{f}_n)' - g'\|_2^2, \tag{11.3}$$

where $g(t) = e^{-t^2/2}$. To estimate the L^2-norms, first write

$$\frac{1}{2}\,\|\widetilde{f}_n - g\|_2^2 \le \frac{1}{2}\int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t) - g(t)|^2\,dt + \int\limits_{|t|>\sqrt{N}} |\widetilde{f}_n(t)|^2\,dt + \int\limits_{|t|>\sqrt{N}} g(t)^2\,dt.$$

Since $|\widetilde{f}_n(t) - f_n(t)| \le 2^{-(N-2)}\,N^3$, we have

$$\begin{aligned}\frac{1}{2}\int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t){-}g(t)|^2\,dt &\le \int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t) - f_n(t)|^2\,dt + \int\limits_{|t|\le\sqrt{N}} |f_n(t) - g(t)|^2\,dt \\ &\le \int\limits_{|t|\le\sqrt{N}} |f_n(t) - g(t)|^2\,dt + 2^{-(2N-5)}\,N^{7/2}.\end{aligned} \tag{11.4}$$

In addition, by Proposition 10.1,

$$\int\limits_{|t|\ge\sqrt{N}} |\widetilde{f}_n(t)|^2\,dt \le C\sqrt{N}\,e^{-cN} \tag{11.5}$$

with $C = e^{2D+4}$ and $c = c_0 e^{-12\,D}$, where c_0 is an absolute positive constant.

Using a well-known bound $1 - \Phi(x) \le \frac{1}{x}\,\varphi(x)$ $(x > 0)$, we easily get $\int_{|t|>\sqrt{N}} g(t)^2\,dt < e^{-N}$. Together with (11.4)–(11.5), and since one may always assume that $c_0 \le \frac{1}{2}$, the latter gives

$$\frac{1}{2}\,\|\widetilde{f}_n - g\|_2^2 \le \int\limits_{|t|\le\sqrt{N}} |f_n(t) - g(t)|^2\,dt + C\sqrt{N}\,e^{-cN} \tag{11.6}$$

with D-dependent constants $C = C_0 e^{2D}$ and $c = c_0 e^{-12\,D}$ (where C_0 and c_0 are numerical).

A similar analysis based on the application of Proposition 8.3 (cf. Corollary 8.4) and Proposition 10.2 with $s = 1$ leads to an analogous estimate

$$\frac{1}{2}\,\|(\widetilde{f}_n)' - g'\|_2^2 \le \int\limits_{|t|\le\sqrt{N}} |f_n'(t) - g'(t)|^2\,dt + CN^{3/2}\,e^{-cN}.$$

Together with (11.6) it may be applied in (11.3), and then we get

$$\|\widetilde{F}_n - \Phi\|_{\mathrm{TV}}^2 \le \int\limits_{|t|\le\sqrt{N}} |f_n(t) - g(t)|^2\,dt + \int\limits_{|t|\le\sqrt{N}} |f_n'(t) - g'(t)|^2\,dt + CN^{3/2}\,e^{-cN}.$$

It is time to appeal to the classical theorem on the approximation of f_n by the characteristic function of the standard normal law, cf. e.g. [4].

Lemma 11.1 *Assume $L_3 \le 1$. Up to an absolute constant A, in the interval $|t| \le L_3^{-1/3}$ we have*

$$|f_n(t) - g(t)| \le AL_3\,e^{-t^2/4},$$

and similarly for the first three derivatives of $f_n - g$.

In fact, the above inequality holds in the larger interval $|t| \le 1/(4L_3)$. But this will not be needed for the present formulation of Theorem 1.1.

Thus, if in addition to the original condition $6 \le N \le \frac{1}{2\sigma^2}$ we require that $\sqrt{N} \le L_3^{-1/3}$, Lemma 11.1 may be applied, and we get

$$\|\widetilde{F}_n - \Phi\|_{\mathrm{TV}} \le AL_3 + CN^{3/2}\,e^{-cN}.$$

Using this together with (11.2) in (11.1), we arrive at

$$\|F_n - \Phi\|_{\mathrm{TV}} \le AL_3 + CN^{3/2}\,e^{-cN}, \tag{11.7}$$

where A is some positive absolute constant, while $C = C_0 e^{2D}$ and $c = c_0 e^{-12\,D}$, as before.

Proof of Theorem 1.1 To finish the argument, we may take $N = [\frac{1}{2}\,L_3^{-2/3}]$, so that $\sqrt{N} \le L_3^{-1/3}$. In view of the elementary bound $\sigma \le L_3^{1/3}$, the condition $N \le \frac{1}{2\sigma^2}$ is fulfilled, as well. Finally, the condition $N \ge 6$ just restricts us to smaller values of L_3, and, for example, $L_3 \le \frac{1}{64}$ would work. Indeed, in this case, $\frac{1}{2}\,L_3^{-2/3} \ge 8$, so $N \ge 8$.

Thus, if $L_3 \le \frac{1}{64}$, then (11.7) holds true. But since $N \ge \frac{1}{4}\,L_3^{-2/3}$, the last term in (11.7) is dominated by any power of L_3 (up to constants). For example, using $e^x \ge \frac{1}{2}\,x^3$ $(x \ge 0)$, we get

$$N^{3/2}\,e^{-cN} \le \frac{2}{c^3}\,N^{-3/2} \le \frac{16}{c^3}\,L_3 = \frac{16}{c_0^3}\,e^{36D}\,L_3.$$

Hence, (11.7) implies

$$\|F_n - \Phi\|_{\mathrm{TV}} \le CL_3, \tag{11.8}$$

with $C = C_0 e^{36D}$, where C_0 is a positive numerical constant.

Finally, if $L_3 > \frac{1}{64}$, (11.8) automatically holds with $C = 128$, and Theorem 1.1 is proved. □

Note, however, that the inequality (11.7) contains more information in comparison with Theorem 1.1. Again assume, as above, that $L_3 \leq \frac{1}{64}$ and take $N = [\frac{1}{2} L_3^{-2/3}]$. If $D \leq \frac{1}{24} \log \frac{1}{L_3}$, then

$$cN \geq c_0 e^{-12\,D} \cdot \frac{1}{4} L_3^{-2/3} \geq c_0 L_3^{1/2} \cdot \frac{1}{4} L_3^{-2/3} = \frac{c_0}{4} L_3^{-1/6}$$

and $C = C_0 e^{2D} \leq C_0 L_3^{-1/12}$. Hence,

$$C N^{3/2} e^{-cN} \leq C_0 L_3^{-1/12} \cdot L_3^{-1} \cdot e^{-\frac{c_0}{4} L_3^{-1/6}} \leq C_0' L_3$$

with some absolute constant C_0'. As a result, (11.7) yields $\|F_n - \Phi\|_{\mathrm{TV}} \leq (A + C_0')\, L_3$, and we arrive at:

Theorem 11.2 *Assume that the independent random variables X_k have mean zero and finite third absolute moments. If $L_3 \leq \frac{1}{64}$ and $D(X_k) \leq \frac{1}{24} \log \frac{1}{L_3} (1 \leq k \leq n)$, then*

$$\|F_n - \Phi\|_{\mathrm{TV}} \leq C L_3, \tag{11.9}$$

where C is an absolute constant.

One should note that in the range $L_3 > \frac{1}{64}$ the inequality (11.9) holds, as well, namely, with $C = 128$ and without any constraint on $D(X_k)$.

12 Proof of Theorem 1.2 and its refinement

In the proof of Theorem 1.2, we apply the general bound (3.1) of Proposition 3.1 to the modified densities $\widetilde{p}_n$ constructed under the same constraints $m_0 = 3$ and $6 \leq N \leq \frac{1}{2\sigma^2}$, as in the proof of Theorem 1.1. It then gives

$$\widetilde{D}_n \leq \alpha^2 + 4(\|\widetilde{f}_n - g_\alpha\|_2 + \|(\widetilde{f}_n)''' - g_\alpha'''\|_2),$$

where $\widetilde{D}_n$ is the relative entropy of $\widetilde{F}_n$ with respect to Φ and

$$g_\alpha(t) = g(t) \left(1 + \alpha \, \frac{(it)^3}{3!}\right), \quad \alpha = \sum_{k=1}^n \mathbf{E} X_k^3.$$

As we know from Proposition 9.1, $\widetilde{D}_n$ provides a good approximation for the entropic distance $D_n = D(S_n)$, namely

$$|\widetilde{D}_n - D_n| < 2^{-(N-6)}\, N^4\, (D + 1).$$

Hence,

$$D_n \le \alpha^2 + 4(\|\widetilde{f}_n - g_\alpha\|_2 + \|(\widetilde{f}_n)''' - g_\alpha'''\|_2) + 2^{-(N-6)} N^4 (D+1). \quad (12.1)$$

On the other hand, the closeness of f_n and g_α on relatively large intervals is provided by:

Lemma 12.1 *Assume $L_4 \le 1$. Up to an absolute constant A, in the interval $|t| \le L_4^{-1/6}$ we have*

$$|f_n(t) - g_\alpha(t)| \le A L_4\, e^{-t^2/4}, \quad (12.2)$$

and similarly for the first four derivatives of $f_n - g_\alpha$.

Again, we refer to [4], where one can find several variants of such bounds.

We also use the following elementary relations, cf. e.g. [19, p. 139, Lemma 2].

Lemma 12.2 $\alpha^2 \le L_3^2 \le L_4$.

Now, assume that $L_4 \le 1$. To estimate the L^2-norms in (12.1), again write

$$\|\widetilde{f}_n - g_\alpha\|_2^2 \le \int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t) - g_\alpha(t)|^2\, dt + 2\int\limits_{|t|>\sqrt{N}} |\widetilde{f}_n(t)|^2\, dt + 2\int\limits_{|t|>\sqrt{N}} |g_\alpha(t)|^2\, dt. \quad (12.3)$$

Using $|\widetilde{f}_n(t) - f_n(t)| \le 2^{-(N-2)} N^3$ and the inequality (12.2) with $|t| \le \sqrt{N} \le L_4^{-1/6}$, we have

$$\begin{aligned}\int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t) - g_\alpha(t)|^2\, dt &\le 2\int\limits_{|t|\le\sqrt{N}} |\widetilde{f}_n(t) - f_n(t)|^2\, dt \\ &\quad + 2\int\limits_{|t|\le\sqrt{N}} |f_n(t) - g_\alpha(t)|^2\, dt \\ &\le A L_4^2 + 2^{-(2N-5)} N^{7/2} \end{aligned} \quad (12.4)$$

with some absolute constant A.

The middle integral on the right-hand side of (12.3) has been already estimated in (11.5).

In addition, using $t^6 g(t) \le 6^3/e^3$, we have

$$|g_\alpha(t)|^2 = g(t)^2 \left(1 + \alpha^2 \frac{t^6}{36}\right) < (1+\alpha^2)\, g(t) \le 2\, g(t),$$

where we applied Lemma 12.2 together with the assumption $L_4 \le 1$ (so that $|\alpha| \le 1$). Hence,

$$\int\limits_{|t|>\sqrt{N}} |g_\alpha(t)|^2\,dt < 2 \int\limits_{|t|>\sqrt{N}} e^{-t^2/2}\,dt < 2\,e^{-N/2}.$$

One may combine this bound with (11.5) and (12.4), and then (12.3) gives

$$\|\widetilde{f}_n - g_\alpha\|_2^2 \le AL_4^2 + 2^{-(2N-5)}\,N^{7/2} + C\sqrt{N}\,e^{-cN} + 4\,e^{-N/2}$$

with $C = e^{2D+4}$ and $c = c_0 e^{-12\,D}$ as in (11.5), where c_0 is an absolute positive constant. Since one may always choose $c_0 \le \frac{1}{2}$, the above inequality may be simplified as

$$\|\widetilde{f}_n - g_\alpha\|_2 \le AL_4 + CN^{1/4}\,e^{-cN}$$

with some absolute constant A and D-dependent constants $C = C_0 e^{2D}$ and $c = c_0 e^{-12\,D}$.

By a similar analysis based on the application of Corollary 8.4 and Proposition 10.2 with $s = 3$ (cf. (10.10)), we also have an analogous estimate

$$\|\widetilde{f}_n''' - g_\alpha'''\|_2 \le AL_4 + CN^{7/4}\,e^{-cN}.$$

Hence, (12.1) together with Lemma 12.2 yields

$$D_n \le AL_4 + CN^{7/4}\,e^{-cN}, \tag{12.5}$$

where A is absolute, and $C = C_0 e^{2D}$ and $c = c_0 e^{-12\,D}$, as before. The obtained estimate holds true, as long as $6 \le N \le \frac{1}{2\sigma^2}$ and $\sqrt{N} \le L_4^{-1/6}$ with $L_4 \le 1$.

Proof of Theorem 1.2 The last condition, $\sqrt{N} \le L_4^{-1/6}$, is satisfied for $N = [\frac{1}{2}\,L_4^{-1/3}]$. Then, by the elementary bound $\sigma \le L_4^{1/4}$, we also have $N \le \frac{1}{2\sigma^2}$. The condition $N \ge 6$ restricts us to smaller values of L_4. If, for example, $L_4 \le 4^{-6}$, we have $\frac{1}{2}\,L_4^{-1/3} \ge 8$ and hence $N \ge 8$.

Thus, if $L_4 \le 4^{-6}$, then (12.5) holds true. But, since $N \ge \frac{1}{4}\,L_4^{-1/3}$, the last term in (12.5) is dominated by any power of L_4. In particular, using $e^x \ge \frac{1}{25}\,x^5$ ($x \ge 0$), we get

$$N^2\,e^{-cN} \le \frac{25}{c^5}\,N^{-3} \le \frac{25 \cdot 4^5}{c^5}\,L_4 = \frac{25 \cdot 4^5}{c_0^5}\,e^{60\,D}\,L_4.$$

Hence, (12.5) yields

$$D_n \le C L_4 \tag{12.6}$$

with $C = C_1 e^{2D}\, e^{60\,D} = C_1\, e^{62\,D}$, where C_1 is an absolute constant.

Finally, for $L_4 > 4^{-6}$, one may use the relation $D_n \le D$ (according to the entropy power inequality), which shows that (12.6) holds with $C = 4^6 D$. Theorem 1.2 is proved. □

Now, again assume, as above, that $L_4 \le 4^{-6}$ and take $N = [\frac{1}{2}\, L_4^{-1/3}]$. If $D \le \frac{1}{48}\, \log \frac{1}{L_4}$, then $cN \ge c_0 L_4^{1/4} \cdot \frac{1}{4}\, L_4^{-1/3} = \frac{c_0}{4}\, L_4^{-1/12}$ and $C = C_0 e^{2D} \le C_0 L_4^{-1/24}$. Hence,

$$C N^{7/4}\, e^{-cN} \le C_0 L_4^{-1/24} \cdot L_4^{-7/12} \exp\Big\{-\frac{c_0}{4}\, L_4^{-1/12}\Big\} \le C_0' L_4$$

with some absolute constant C_0'. As a result, (12.5) yields $D_n \le (A + C_0')\, L_4$, and we arrive at another variant of Theorem 1.2.

Theorem 12.3 *Assume that the independent random variables X_k have mean zero and finite fourth absolute moments. If $L_4 \le 2^{-12}$ and $D(X_k) \le \frac{1}{48}\, \log \frac{1}{L_4} (1 \le k \le n)$, then*

$$D(S_n) \le C L_4,$$

where C is an absolute constant.

Here, the two assumptions about L_4 and $D = \max_k D(X_k)$ may be united by just one relation $L_4 \le \min\{2^{-12}, e^{-48D}\}$. When not paying attention to the value of numerical constants, this relation may be written in a more compact form as

$$L_4 \le c\, e^{-D/c},$$

where $c > 0$ is an absolute constant.

Let us illustrate this result in the scheme of weighted sums

$$S_n = a_1 X_1 + \cdots + a_n X_n$$

of independent identically distributed random variables X_k, such that $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 = 1$, and with coefficients such that $a_1^2 + \cdots + a_n^2 = 1$. In this case $L_4 = \mathbf{E}X_1^4 \sum_{k=1}^n a_k^4$, so Theorem 12.3 is applicable, when the last sum is sufficiently small.

Corollary 12.4 *Assume that X_1 has density with finite entropy, and let $\mathbf{E}X_1^4 < +\infty$. If the coefficients satisfy*

$$\sum_{k=1}^n a_k^4 \le \frac{c}{\mathbf{E}X_1^4}\, e^{-D(X_1)/c},$$

then

$$D(S_n) \le C\,\mathbf{E}X_1^4 \sum_{k=1}^n a_k^4,$$

where C and c are positive absolute constants.

For example, in case of equal coefficients, so that $S_n = \frac{X_1+\dots+X_n}{\sqrt{n}}$, the conclusion becomes

$$D(S_n) \le \frac{C}{n}\,\mathbf{E}X_1^4, \quad \text{for all } n \ge n_1,$$

which holds true with an absolute constant C and $n_1 = 2^{12}e^{48D(X_1)}\,\mathbf{E}X_1^4$.

13 The case of bounded densities

In this Section we add a few remarks about Theorems 1.1–1.2 for the case where the densities of the summands X_k are bounded.

First, let us note that, if a random variable X has an absolutely continuous distribution with a bounded density $p(x) \le M$, where M is a constant, and if the variance $\sigma^2 = \mathrm{Var}(X)$ is finite ($\sigma > 0$), then X has finite entropy, and moreover,

$$D(X) \le \log(M\sigma\sqrt{2\pi e}). \tag{13.1}$$

Indeed, if Z is a standard normal random variable, and assuming (without loss of generality) that $\sigma = 1$, we have

$$D(X) = h(Z) - h(X) = \log(\sqrt{2\pi e}) + \int_{-\infty}^{+\infty} p(x)\log p(x)\,dx,$$

which immediately implies (13.1).

It is worthwile to note that, similarly to D, the functional $X \to M\sigma$ is affine invariant, where $M = \operatorname{ess\,sup}_x p(x)$. Therefore, $M\sigma$ does not depend neither on the mean, nor the variance of X. In addition, one always has $M\sigma \ge \frac{1}{\sqrt{12}}$, and the equality is achieved only for X which is uniformly distributed in a finite interval of the real line. (Without proof this lower bound is already mentioned in [26].)

Using (13.1), Theorems 1.1 and 1.2 admit formulations involving the maximum of the densities. In the statement below, let $(X_k)_{1\le k\le n}$ be independent random variables with mean zero and variances $\sigma_k^2 = \mathbf{E}X_k^2 (\sigma_k > 0)$, such that $\sum_{k=1}^n \sigma_k^2 = 1$. Let F_n be the distribution function of the sum $S_n = X_1 + \dots + X_n$.

Corollary 13.1 *Assume that every X_k has density bounded by M_k. If* $\max_k M_k\sigma_k \le M$*, then*

$$\|F_n - \Phi\|_{\mathrm{TV}} \le CL_3, \tag{13.2}$$

where the constant C depends on M, only. Moreover,

$$D(S_n) \le CL_4. \tag{13.3}$$

Here, one may take $C = C_0 M^c$ with some positive absolute constants C_0 and c.
In particular, consider the weighted sums

$$S_n = a_1 X_1 + \cdots + a_n X_n$$

of independent identically distributed random variables X_k, such that $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 = 1$, and with coefficients satisfying $a_1^2 + \cdots + a_n^2 = 1$. If X_1 has density, bounded by M, (13.2)–(13.3) yield respectively

$$\|F_n - \Phi\|_{\mathrm{TV}} \le C_M \, \mathbf{E}\,|X_1|^3 \sum_{k=1}^n |a_k|^3, \quad D(S_n) \le C_M \, \mathbf{E}X_1^4 \sum_{k=1}^n a_k^4,$$

where C_M depends on M, only. (One may take $C_M = C_0 M^c$.)

Moreover, if $\sum_{k=1}^n |a_k|^3$ or, respectively, $\sum_{k=1}^n a_k^4$ are sufficiently small, the constant C_M may be chosen to be independent of M. In particular, in the i.i.d. case, where $S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$, the last bound may also be written with an absolute constant C, i.e.,

$$D(S_n) \le \frac{C}{n} \mathbf{E}X_1^4, \quad \text{for all } \ n \ge n_1.$$

One may take, e.g., $n_1 = 2^{12} (M\sqrt{2\pi e})^{48}\, \mathbf{E}X_1^4$.

Acknowledgments We would like to thank M. Ledoux for pointing us to the relationship between Theorem 1.2 and the transportation inequality of E. Rio. We also thank the referees for careful reading of the manuscript and very helpful remarks.

References

1. Artstein, S., Ball, K.M., Barthe, F., Naor, A.: On the rate of convergence in the entropic central limit theorem. Probab. Theory Relat. Fields **129**(3), 381–390 (2004)
2. Barron, A.R.: Entropy and the central limit theorem. Ann. Probab. **14**(1), 336–342 (1986)
3. Barron, A.R., Johnson, O.: Fisher information inequalities and the central limit theorem. Probab. Theory Relat. Fields **129**(3), 391–409 (2004)
4. Bhattacharya, R.N., Ranga Rao, R.: Normal Approximation and Asymptotic Expansions. Wiley, New York (1976). Also: Soc. for Industrial and Appl. Math., Philadelphia (2010)
5. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. Ann. Probab. ArXiv:1104.3994 v1 [math.PR] (2011)
6. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Bounds for characteristic functions in terms of quantiles and entropy. Electron. Commun. Probab. 17 (2012), paper no. 21, electronic
7. Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**(1), 1–28 (1999)
8. Brown, L.D.: A Proof of the Central Limit Theorem Motivated by the Cramer-Rao Inequality. Statistics And Probability: Essays in Honor of C. R. Rao, pp. 141–148. North-Holland, Amsterdam (1982)
9. Carlen, E.A., Soffer, A.: Entropy production by block variable summation and central limit theorems. Comm. Math. Phys. **140**(2), 339–371 (1991)

10. Cover, T.M., Dembo, A., Thomas, J.A.: Information-theoretic inequalities. IEEE Trans. Inf. Theory **37**(6), 1501–1518 (1991)
11. Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hung. **2**, 299–318 (1967)
12. Esseen, C.-G.: Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math. **77**, 1–125 (1945)
13. Fedotov, A.A., Harremoës, P., Topsøe, F.: Refinements of Pinsker's inequality. IEEE Trans. Inf. Theory **49**(6), 1491–1498 (2003)
14. Feller, W.: An Introduction to Probability Theory and its Applications, vol. II, 2nd edn. Wiley, New York (1971)
15. Ibragimov, I.A., Linnik, J.V.: Independent and Stationarily Connected Variables. Izdat. "Nauka", Moscow (1965)
16. Johnson, O.: Information Theory and the Central Limit Theorem. Imperial College Press, London (2004)
17. Kullback, S.: A lower bound for discrimination in terms of variation. IEEE Trans. Inf. Theory **T–13**, 126–127 (1967)
18. Linnik, J.V.: An information-theoretic proof of the central limit theorem with the Lindeberg condition. Theory Probab. Appl. **4**, 288–299 (1959)
19. Petrov, V.V.: Sums of independent random variables. Springer, Berlin (1975)
20. Pinelis, I.F., Utev, S.A.: Estimates of moments of sums of independent random variables. Theory Probab. Appl. **29**(3), 574–577 (1984)
21. Pinsker, M.S.: Information and information stability of random variables and processes. Translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco (1964)
22. Prohorov, Y.V.: A local theorem for densities (Russian). Doklady Akad. Nauk SSSR (N.S.) **83**, 797–800 (1952)
23. Rio, E.: Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. **45**(3), 802–817 (2009)
24. Rosenthal, H.P.: On the subspaces of L^p $(p\ 2)$ spanned by sequences of independent random variables. Isr. J. Math. **8**, 273–303 (1970)
25. Senatov, V.V.: Central Limit Theorem. Exactness of Approximation and Asymptotic Expansions (Russian). TVP Science Publishers, Moscow (2009)
26. Statulevičius, V.A.: Limit theorems for densities and the asymptotic expansions for distributions of sums of independent random variables. Theory Probab. Appl. **10**(4), 682–695 (1965)
27. Sirazhdinov, S.H., Mamatov, M.: On mean convergence for densities. Theory Probab. Appl. **7**(4), 424–428 (1962)
28. Talagrand, M.: Transportation cost for Gaussian and other product measures. Geom. Funct. Anal. **6**(3), 587–600 (1996)
29. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)