S. G. Bobkov and A. Marsiglietti (2020) "Entropic CLT for Smoothed Convolutions and Associated Entropy Bounds,"
International Mathematics Research Notices, Vol. 2020, No. 21, pp. 8057–8080
Advance Access Publication February 3, 2020
doi:10.1093/imrn/rnaa005

Entropic CLT for Smoothed Convolutions and Associated Entropy Bounds

Sergey G. Bobkov[1] **and Arnaud Marsiglietti**[2,*]

[1]School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA and [2]Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

**Correspondence to be sent to: e-mail: a.marsiglietti@ufl.edu*

We explore an asymptotic behavior of entropies for sums of independent random variables that are convolved with a small continuous noise.

1 Introduction

Let $(X_n)_{n\geq 1}$ be independent, identically distributed (i.i.d.) random vectors in $\mathbb{R}^d$ with an isotropic distribution, that is, with mean zero and an identity covariance matrix. By the central limit theorem (CLT), given a random vector X in $\mathbb{R}^d$, independent of all X_n's, the normalized sums

$$Z_n = \frac{1}{\sqrt{n}}\,(X + X_1 + \cdots + X_n) \tag{1}$$

are convergent weakly in distribution as $n \to \infty$ to the standard normal random vector Z with density

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

Suppose that X has a finite 2nd moment and an absolutely continuous distribution, so that Z_n have some densities p_n. A natural question of interest is whether or not this

Received February 21, 2019; Revised September 20, 2019; Accepted December 30, 2019

property (i.e., the weak CLT) may be strengthened as convergence of entropies

$$h(Z_n) = -\int_{\mathbb{R}^d} p_n(x) \, \log p_n(x) \, \mathrm{d}x$$

to the entropy of the Gaussian limit Z. The usual entropic CLT corresponds to the i.i.d. case with $X = 0$. Then, this CLT is known to hold, if and only if Z_n have densities p_n with finite $h(Z_n)$ for some or equivalently all n large enough [2] (see also [1, 5, 14, 17–19, 27]). What also seems remarkable, the presence of a small non-zero noise $X/\sqrt{n}$ in (1) may potentially enlarge the range of applicability of the entropic CLT. Here is one observation in this direction in terms of the characteristic function

$$f(t) = \mathbb{E}\, e^{i\langle t,X\rangle}, \quad t \in \mathbb{R}^d.$$

Theorem 1.1. If f is compactly supported, and X_1 has a non-lattice distribution, then

$$h(Z_n) \to h(Z) \quad \text{as } n \to \infty. \tag{3}$$

This convergence also holds for lattice distributions, if f is supported on the ball $|t| \leq T$ for some $T > 0$ depending on the distribution of X_1. One may take $T = 1/\beta_3$, assuming that the 3rd absolute moment

$$\beta_3 = \sup_{|\theta|=1} \mathbb{E}\, |\langle X_1, \theta\rangle|^3$$

is finite.

The assumption of compactness on the support of the characteristic function of X requires its density p to be the restriction to $\mathbb{R}^d$ of an entire function on $\mathbb{C}^d$ of exponential type by Paley–Wiener theorems (cf., e.g., [29]).

The entropic CLT (3) may equivalently be stated as the convergence

$$D(Z_n||Z) = \int_{\mathbb{R}^d} p_n(x) \, \log \frac{p_n(x)}{\varphi(x)} \, \mathrm{d}x \;\to\; 0 \quad (n \to \infty)$$

for the Kullback–Leibler distance (also called relative entropy or an informational divergence). It belongs to the family of so-called strong (informational) distances, which dominate many other metrics that are used in usual CLT's about the weak convergence of probability distributions. As was mentioned to us by one of the referees, one immediate consequence from (3) is the CLT for non-smoothed normalized sums with respect to the Kantorovich transport distance W_2 (cf. Remark 4.4 for details).

In general, the hypothesis on the support of f in Theorem 1.1 cannot be removed, but may be weakened by involving more delicate properties related to the location of zeros of the characteristic function. This may be seen from the following characterization in one important example under mild regularity assumptions on f.

Theorem 1.2. Suppose that X_1 has a uniform distribution on the discrete cube $\{-1, 1\}^d$, that is, with independent Bernoulli coordinates. Let the characteristic function f of X satisfy

$$\int_{\mathbb{R}^d} |f(t)|\,\mathrm{d}t < \infty, \quad \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\,\mathrm{d}t < \infty, \tag{4}$$

where $\|t\|$ denotes the distance from the point t to the lattice $\pi\mathbb{Z}^d$. Then, the entropic CLT (3) holds true, if and only if

$$f(\pi k) = 0 \quad \text{for all } \ k \in \mathbb{Z}^d, \ k \neq 0. \tag{5}$$

The 2nd moment assumption on X guarantees that f has a bounded continuous derivative $f'(t) = \nabla f(t)$ with its Euclidean length $|f'(t)|$. The assumption of integrability of f in (4) requires the density of X to be continuous on $\mathbb{R}^d$. In dimension $d = 1$, the condition (4) is fulfilled, as long as both f and f' are in L^1. If $d \geq 2$, (4) is more complicated, but is fulfilled, for example, under decay assumptions such as

$$|f(t)| \leq \frac{c}{((1+|t_1|)\dots(1+|t_d|))^{\alpha}}, \qquad |f'(t)| \leq \frac{c}{((1+|t_1|)\dots(1+|t_d|))^{\alpha}}, \tag{6}$$

holding for all $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ with some constants $\alpha > 1$ and $c > 0$.

Although an information-theoretic meaning of the property (5) is not clear, it is indeed connected with the entropy functional $h(X)$. Namely, under the conditions (4)-(5), it turns out that the entropy has to be non-negative. This is emphasized in the next statement, where we drop the isotropy condition and extend the Bernoulli case to arbitrary integer valued random vectors. As before, we assume that X is a continuous random vector in $\mathbb{R}^d$ with finite 2nd moment, which is independent of all X_n's.

Theorem 1.3. Let $(X_n)_{n\geq 1}$ be a sequence of independent, integer valued random vectors, whose components have variance one. Then

$$\limsup_{n\to\infty} h(Z_n) \leq h(X) + h(Z).$$

In particular, if $h(Z_n) \to h(Z)$ as $n \to \infty$, then necessarily $h(X) \geq 0$.

Actually, the independence assumption may further be weakened to the uncorrelatedness (as explained in Theorem 5.3 at the end of these notes).

We do not discuss here possible applications of the last conclusion in Theorem 1.3. Let us, however, stress that obtaining lower and upper bounds for the differential entropy, under various hypotheses or for different classes of probability distributions on the Euclidean space $\mathbb{R}^d$, is in itself an important and self-sufficient direction in information theory, which is motivated by many problems and is connected with other areas. For example, applications of lower bounds to rate-distortion theory and channel capacity were put forward in [23] (see also [12, 16, 22]). Let us also mention Bourgain's slicing problem in asymptotic geometric analysis, cf. [9]. As a main conjecture, it states that for any convex body K in $\mathbb{R}^d$ there is a hyperplane H such that the $(d-1)$-dimensional volume of the slice $H \cap K$ is bounded away from zero by a universal positive constant. It was shown in [6] that the latter may equivalently be formulated as the property that if X is a random vector in $\mathbb{R}^d$ with an isotropic log-concave distribution, then

$$h(X) \geq -cd$$

with some universal constant $c > 0$. Besides this conjecture, the past few years have seen a growing interest in the study of entropic inequalities as they shed new lights on fundamental problems in convex geometry (cf., e.g., [7, 10, 11]). We refer to the survey paper [21] for further details on the connections between entropic inequalities and geometric and functional inequalities.

The paper is organized as follows. We start in Section 2 with general upper and lower bounds on the Kullback–Leibler distance

$$D(X||Z) = \int_{\mathbb{R}^d} p(x) \, \log \frac{p(x)}{\varphi(x)} \, \mathrm{d}x \tag{7}$$

from the distribution of X to the standard normal law in terms of the L^2-distance

$$\Delta = \|p - \varphi\|_2 = \left(\int_{\mathbb{R}^d} (p(x) - \varphi(x))^2 \, dx \right)^{1/2}. \tag{8}$$

Throughout, Z denotes a standard normal random vector in $\mathbb{R}^d$, thus with density φ as in (2) and with characteristic function

$$g(t) = \mathbb{E}\, e^{i\langle t,Z\rangle} = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\, \varphi(x)\, dx = e^{-|t|^2/2}, \qquad t \in \mathbb{R}^d.$$

As usual, the Euclidean space $\mathbb{R}^d$ is endowed with the canonical inner product $\langle\cdot,\cdot\rangle$ and the norm $|\cdot|$. These bounds are applied in Section 3 to express the entropic CLT as convergence of densities in L^2. Theorems 1.1 and 1.2 (in a somewhat refined form) are proved in Section 4. Using Proposition 3.1, the proofs employ recent results obtained in [8] on local limit theorems with respect to the L^2 and L^∞-norms. Theorem 1.3 is proved in Section 5, where we also discuss the connection between entropy bounds and the entropic CLT.

2 General Bounds on Relative Entropy

Throughout this section, let X be a random vector in $\mathbb{R}^d$ with density p, and let Δ be defined according to (8).

Proposition 2.1. Suppose that $\mathbb{E}\,|X|^2 = d$. If $\Delta \le 1/e$, then

$$D(X||Z) \;\le\; c_d\, \Delta\, \log^{\frac{d+4}{4}}(1/\Delta) \tag{9}$$

with some constant $c_d > 0$ depending on d only. Moreover, if $\sup_x p(x) \le M$ for some constant $M \ge (2\pi)^{-d/2}$, then

$$D(X||Z) \;\ge\; \frac{1}{2M}\, \Delta^2. \tag{10}$$

First we collect a few elementary large deviation bounds.

Lemma 2.2. For any $T \ge 1$,

(a) $\int_{|x|\ge T} \varphi(x)\, dx \;\le\; 2d\, T^{d-2}\, e^{-T^2/2}$;

(b) $\int_{|x|\ge T} |x|^2\, \varphi(x)\, dx \;\le\; 2d\, T^d\, e^{-T^2/2}$.

Proof. Clearly, (a) follows from (b). To derive the 2nd bound, write

$$\mathbb{E}\,|Z|^2\, 1_{\{|Z|\geq T\}} \;=\; \int_{|x|\geq T} |x|^2\, \varphi(x)\, \mathrm{d}x \;=\; \frac{d\omega_d}{(2\pi)^{d/2}} \int_T^\infty r^{d+1}\, e^{-r^2/2}\, \mathrm{d}r, \tag{11}$$

where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$. Given $c > 1$, consider the function

$$u(T) = \int_T^\infty r^{d+1}\, e^{-r^2/2}\, \mathrm{d}r - cT^d\, e^{-T^2/2}.$$

We have $u(\infty) = 0$ and

$$u'(T) = \big((c-1)\, T^2 - cd\big)\, T^{d-1}\, e^{-T^2/2}.$$

Thus, $u(T)$ is decreasing in some interval $0 \leq T < T_0$ and is increasing in $T \geq T_0$. Therefore, $u(T) \leq 0$ for all $T \geq 1$, if $u(1) = 0$, that is, for

$$c = \sqrt{e} \int_1^\infty r^{d+1}\, e^{-r^2/2}\, \mathrm{d}r.$$

Using (11), we obtain

$$c = \sqrt{e}\, \frac{(2\pi)^{d/2}}{d\omega_d} \mathbb{E}\,|Z|^2\, 1_{\{|Z|\geq 1\}} \leq \sqrt{e}\, \frac{(2\pi)^{d/2}}{\omega_d},$$

so

$$\int_{|x|\geq T} |x|^2\, \varphi(x)\, \mathrm{d}x \;=\; \frac{d\omega_d}{(2\pi)^{d/2}} \left(u(T) + cT^d\, e^{-T^2/2}\right) \;\leq\; \sqrt{e}\, d\, T^d\, e^{-T^2/2}.$$

■

To get the upper bound (9), we also need to control the weighted quadratic tails in terms of the L^2-distance Δ.

Lemma 2.3. If $\mathbb{E}\,|X|^2 = d$, then for all $T \geq 1$,

$$\int_{|x|\geq T} |x|^2\, p(x)\, \mathrm{d}x \;\leq\; 2T^{\frac{d+4}{2}}\, \Delta + 2d\, T^d\, e^{-T^2/2}.$$

Proof. We have

$$\begin{aligned}\int_{|x|\ge T}|x|^2\,p(x)\,\mathrm{d}x &= d-\int_{|x|\le T}|x|^2p(x)\,\mathrm{d}x\\ &= \int_{|x|\le T}|x|^2\,(\varphi(x)-p(x))\,\mathrm{d}x+\int_{|x|\ge T}|x|^2\,\varphi(x)\,\mathrm{d}x\\ &\le \int_{|x|\le T}|x|^2\,|p(x)-\varphi(x)|\,\mathrm{d}x+\int_{|x|\ge T}|x|^2\,\varphi(x)\,\mathrm{d}x.\end{aligned}$$

The last integral is bounded by $2d\,T^d\,e^{-T^2/2}$. Also, by the Cauchy inequality,

$$\left(\int_{|x|\le T}|x|^2\,|p(x)-\varphi(x)|\,\mathrm{d}x\right)^2\le\int_{|x|\le T}|x|^4\,\mathrm{d}x\int_{\mathbb{R}^d}(p(x)-\varphi(x))^2\,\mathrm{d}x=\frac{d\omega_d}{d+4}\,T^{d+4}\,\Delta^2,$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$. Here, $\frac{d\omega_d}{d+4}<4$. ■

Lemma 2.4. For all $T\ge 1$,

$$\begin{aligned}D(X||Z) \le\ & 2d\,T^{d-2}\,e^{-T^2/2}+(2\pi)^{d/2}\int_{|x|\le T}(p(x)-\varphi(x))^2\,e^{|x|^2/2}\,\mathrm{d}x\\ &+\frac{2d-1}{2}\int_{|x|\ge T}|x|^2\,p(x)\,\mathrm{d}x+\int_{|x|\ge T}p\log\,p\,\mathrm{d}x.\end{aligned}\tag{12}$$

Proof. In definition (8), we split the integration into the two regions. Using the inequality $t\log t\le(t-1)+(t-1)^2$, $t\ge 0$, and applying the 1st bound of Lemma 2.2, we have

$$\begin{aligned}\int_{|x|\le T}\frac{p}{\varphi}\log\frac{p}{\varphi}\,\varphi\,\mathrm{d}x &\le \int_{|x|\le T}\left(\frac{p}{\varphi}-1\right)\varphi\,\mathrm{d}x+\int_{|x|\le T}\left(\frac{p}{\varphi}-1\right)^2\varphi\,\mathrm{d}x\\ &= \int_{|x|\ge T}(\varphi-p)\,\mathrm{d}x+\int_{|x|\le T}\frac{(p-\varphi)^2}{\varphi}\,\mathrm{d}x\\ &\le 2d\,T^{d-2}\,e^{-T^2/2}-\int_{|x|\ge T}p\,\mathrm{d}x+(2\pi)^{d/2}\int_{|x|\le T}(p(x)-\varphi(x))^2\,e^{|x|^2/2}\,\mathrm{d}x.\end{aligned}$$

For the 2nd region $|x|\ge T$, just write

$$\int_{|x|\ge T}p\log\frac{p}{\varphi}\,\mathrm{d}x = \int_{|x|\ge T}p\log p\,\mathrm{d}x+\frac{d}{2}\,\log(2\pi)\int_{|x|\ge T}p\,\mathrm{d}x+\frac{1}{2}\int_{|x|\ge T}|x|^2\,p(x)\,\mathrm{d}x.$$

Combining these relations and noting that $\log(2\pi) < 2$, we thus get

$$\begin{aligned} D(X||Z) \;\; &\leq \;\; 2d\,T^{d-2}\,e^{-T^2/2} + (2\pi)^{d/2} \int_{|x|\leq T} (p(x) - \varphi(x))^2\, e^{|x|^2/2}\, \mathrm{d}x \\ &\quad + (d-1) \int_{|x|\geq T} p(x)\,\mathrm{d}x + \frac{1}{2} \int_{|x|\geq T} |x|^2\, p(x)\,\mathrm{d}x + \int_{|x|\geq T} p \log p\, \mathrm{d}x. \end{aligned}$$

■

As a consequence, we obtain the following:

Lemma 2.5. For all $T \geq 1$,

$$D(X||Z) \;\leq\; (2d+1)\,T^{d-1}\,e^{-T^2/2} + \left((2\pi)^{d/2} + 1\right) e^{T^2/2}\,\Delta^2 + d \int_{|x|\geq T} |x|^2\, p(x)\,\mathrm{d}x.$$

Proof. We use the notation $a^+ = \max(a, 0)$. Subtracting $\varphi(x)$ from $p(x)$ and then adding, one can write

$$\begin{aligned} \int_{|x|\geq T} p \log p\, \mathrm{d}x \;\; &\leq \;\; \int_{|x|\geq T} p(x) \log^+(p(x))\, \mathrm{d}x \\ &\leq \;\; \int_{|x|\geq T} |p(x) - \varphi(x)|\, \log^+(p(x))\, \mathrm{d}x + \int_{|x|\geq T} \varphi(x)\, \log^+(p(x))\, \mathrm{d}x. \end{aligned}$$

Next, let us apply Cauchy's inequality together with the bound $(\log^+(t))^2 \leq 4e^{-2}\, t$ so that to estimate the last integral from above by

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

Here, using the 1st bound of Lemma 2.2, we have

$$\begin{aligned} \int_{|x|\geq T} \varphi(x)^2\, \mathrm{d}x \;\; &= \;\; \frac{1}{(4\pi)^{d/2}} \int_{|y|\geq T\sqrt{2}} \varphi(y)\, \mathrm{d}y \\ &\leq \;\; \frac{2d}{(4\pi)^{d/2}}\, (T\sqrt{2})^{d-2}\, e^{-T^2} \;<\; T^{d-1} e^{-T^2}. \end{aligned}$$

Therefore,

$$\int_{|x|\geq T} p \log p\, \mathrm{d}x \;\leq\; \int_{|x|\geq T} |p(x) - \varphi(x)|\, \log^+(p(x))\, \mathrm{d}x + T^{\frac{d-1}{2}}\, e^{-T^2/2}.$$

To simplify, the last integrand may be bounded by

$$\frac{1}{2}\,(p(x)-\varphi(x))^2+\frac{1}{2}\,\left(\log^+(p(x))\right)^2\leq\frac{1}{2}\,(p(x)-\varphi(x))^2+\frac{1}{2}\,p(x),$$

so,

$$\int_{|x|\geq T}p\log p\,\mathrm{d}x\;\leq\;\frac{1}{2}\,\Delta^2+\frac{1}{2}\int_{|x|\geq T}p(x)\,\mathrm{d}x+T^{\frac{d-1}{2}}\,e^{-T^2/2}.$$

Using this estimate in (12) together with $e^{|x|^2/2}\leq e^{T^2/2}$ for $|x|\leq T$, we get

$$\begin{aligned}D(X||Z)\quad&\leq\quad 2d\,T^{d-1}\,e^{-T^2/2}+(2\pi)^{d/2}\,e^{T^2/2}\int_{|x|\leq T}(p(x)-\varphi(x))^2\,\mathrm{d}x\\&\quad+\frac{2d-1}{2}\int_{|x|\geq T}|x|^2p(x)\,\mathrm{d}x+\frac{1}{2}\,\Delta^2+\frac{1}{2}\int_{|x|\geq T}p(x)\,\mathrm{d}x+T^{\frac{d-1}{2}}\,e^{-T^2/2}.\end{aligned}$$

■

Proof. of Proposition 2.1 Combining Lemma 2.5 with Lemma 2.3, we immediately get

$$D(X||Z)\;\leq\;(2d^2+2d+1)\,T^d\,e^{-T^2/2}+\left((2\pi)^{d/2}+1\right)e^{T^2/2}\,\Delta^2+2d\,T^{\frac{d+4}{2}}\,\Delta.$$

To get (9), it remains to take here

$$T=\sqrt{2\log(1/\Delta)+\frac{d}{2}\,\log\log(1/\Delta)}.$$

For the lower bound (10), let us recall that $D(X||Z)=h(Z)-h(X)$. By Taylor's expansion, for all $t\geq 0$ and $t_0>0$, there is a point t_1 between t and t_0 such that

$$t\log t=t_0\log t_0+(\log t_0+1)(t-t_0)+\frac{(t-t_0)^2}{2t_1}.$$

Inserting $t=p(x)$, $t_0=\varphi(x)$, we obtain a measurable function $t_1(x)$ with values between $p(x)$ and $\varphi(x)$, satisfying

$$p(x)\log p(x)\;=\;\varphi(x)\log\varphi(x)+(\log\varphi(x)+1)\,(p(x)-\varphi(x))+\frac{(p(x)-\varphi(x))^2}{2t_1(x)}.$$

Let us integrate this equality over x and use $\mathbb{E}\,|X|^2=d$ to get

$$-h(X)=-h(Z)+\frac{1}{2}\int_{\mathbb{R}^d}\frac{(p(x)-\varphi(x))^2}{t_1(x)}\,\mathrm{d}x.$$

Hence,

$$D(X||Z) = \frac{1}{2} \int_{\mathbb{R}^d} \frac{(p(x) - \varphi(x))^2}{t_1(x)} \, \mathrm{d}x.$$

It remains to use the assumptions $p(x) \le M$ and $\varphi(x) \le M$, so that $t_1(x) \le M$ as well. ■

3 Topological Properties of Relative Entropy

Applying Proposition 2.1 to a sequence of random vectors, we arrive at necessary and sufficient conditions for the convergence in the Kullback–Leibler distance D in terms of the L^2-distances

$$\Delta_n = \|p_n - \varphi\|_2 = \left(\int_{\mathbb{R}^d} (p_n(x) - \varphi(x))^2 \, \mathrm{d}x \right)^{1/2}.$$

More precisely, we have the following:

Proposition 3.1. Let $(Z_n)_{n \ge 1}$ be a sequence of random vectors in $\mathbb{R}^d$ with densities p_n. Suppose that as $n \to \infty$

(a) $\mathbb{E}\,|Z_n|^2 \to d$;
(b) $\Delta_n \to 0$.

Then $D(Z_n||Z) \to 0$ or equivalently $h(Z_n) \to h(Z)$ as $n \to \infty$. Conversely, if p_n are uniformly bounded, then the conditions (a)–(b) are also necessary for the convergence in D.

Before turning to the proof, let us recall a basic abstract definition of the Kullback–Leibler distance (i.e., relative entropy). Let X and Y be random elements in a measurable space Ω with distributions μ and ν, respectively. If μ is absolutely continuous with respect to ν and has density $h = d\mu/d\nu$, the relative entropy of μ with respect to ν is defined as

$$D(X||Y) = D(\mu||\nu) = \int_{\Omega} h \log h \, \mathrm{d}\nu = \int p \, \log \frac{p}{q} \, \mathrm{d}\lambda,$$

where in the last equality we assume that μ and ν have densities p and q with respect to the dominating measure λ on Ω, so that $h = p/q$ (which is well-defined λ-almost everywhere). This definition does not depend on the choice of λ, and one may always take $\lambda = \mu + \nu$, for example. If μ is not absolutely continuous with respect to ν, one puts $D(X||Y) = D(\mu||\nu) = \infty$. For basic properties of this functional, we refer an interested reader to [15], and here only mention one well-known relation:

$$\int_\Omega g\,\mathrm{d}\mu \ \le\ D(\mu||\nu) + \log \int_\Omega e^g\,\mathrm{d}\nu.$$

It holds for any measurable function g on Ω for which the right-hand side is finite (this relation easily follows from the elementary inequality $xy \le x\log x - x + e^y$, $x \ge 0$, $y \in \mathbb{R}$).

In the case where $\Omega = \mathbb{R}^d$ with Lebesgue measure λ, and choosing $g(x) = \varepsilon\,|x|^2$, $\varepsilon > 0$, we have in particular

$$\varepsilon\,\mathbb{E}\,|X|^2 \ \le\ D(X||Y) + \log\,\mathbb{E}\,e^{\varepsilon\,|Y|^2}.$$

If Y has a normal distribution, the last expectation is finite for some $\varepsilon > 0$. Therefore, finiteness of $D(X||Y)$ forces the random vector X in $\mathbb{R}^d$ to have a finite 2nd moment. One can now introduce an affine invariant functional

$$D(X) \ = \ \inf_Y \ \text{normal}\, D(X||Y),$$

where the infimum is running over all absolutely continuous normal distributions on $\mathbb{R}^d$. Thus, $D(X)$ represents the Kullback–Leibler distance from the distribution of X to the class of all non-degenerate Gaussian measures on $\mathbb{R}^d$. It is finite, only if the distribution of X is absolutely continuous and has a finite 2nd moment, and then this infimum is attained on the normal distribution with the same mean $a = \mathbb{E}X$ and covariance matrix V as for X (cf., e.g., [3, Section 10.7]).

Our next step is to quantify the properties (a)–(b) from Proposition 3.1 in terms of $D(X||Z)$, where Z is a standard normal random vector in $\mathbb{R}^d$. Denote by $\varphi_{a,V}$ the density of the normal law with these parameters, that is, let Y have density

$$\varphi_{a,V}(x) \ = \ \frac{1}{(2\pi)^{d/2}\sqrt{\det(V)}}\,\exp\left\{-\frac{1}{2}\left\langle V^{-1}(x-a), x-a\right\rangle\right\}, \qquad x \in \mathbb{R}^d,$$

so that $D(X) = D(X||Y)$. By the definition, if X has density p, we have

$$\begin{aligned} D(X||Z) \ &= \ \int_{\mathbb{R}^d} p(x)\log\frac{p(x)}{\varphi(x)}\,\mathrm{d}x \ = \ \int_{\mathbb{R}^d} p(x)\log\frac{p(x)}{\varphi_{a,V}(x)}\,\mathrm{d}x \\ &\quad - \frac{1}{2}\,\log\,\det(V) - \frac{1}{2}\,\mathbb{E}\left\langle V^{-1}(X-a), X-a\right\rangle + \frac{1}{2}\,\mathbb{E}\,|X|^2. \end{aligned}$$

Simplifying, we obtain an explicit formula

$$\begin{aligned} D(X||Z) &= D(X) + \frac{1}{2}\,|a|^2 + \frac{1}{2}\,\Big(\log \frac{1}{\det}(V) + \mathrm{Tr}(V) - d\Big) \\ &= D(X) + \frac{1}{2}\,|a|^2 + \frac{1}{2}\,\sum_{i=1}^{d}\bigg(\log \frac{1}{\sigma_i^2} + \sigma_i^2 - 1\bigg), \end{aligned} \tag{13}$$

where σ_i^2 are eigenvalues of the matrix V ($\sigma_i > 0$). Note that all the terms on the right-hand side are non-negative. This allows us to control the 1st two moments of X in terms of $D(X||Z)$. In particular, $|a|^2 \leq 2\,D(X||Z)$, so that the closeness of X to Z in relative entropy implies the closeness of the means. To come to a similar conclusion about the covariance matrices, consider the non-negative convex function

$$\psi(t) = \log \frac{1}{t} + t - 1, \qquad t > 0.$$

We have $\psi(1) = \psi'(1) = 0$ and $\psi''(t) = \frac{1}{t^2}$. If $|t-1| \leq 1$, by Taylor's formula about the point $t_0 = 1$ with some point t_1 between t and 1,

$$\psi(t) = \psi(1) + \psi'(1)(t-1) + \psi''(t_1)\,\frac{(t-1)^2}{2} \geq \frac{(t-1)^2}{8}.$$

For the values $t \geq 2$, we have a linear bound $\log\frac{1}{t} + t - 1 \geq c(t-1)$ with some constant $0 < c < 1$. Namely, write the latter inequality as $\log t \leq (1-c)(t-1)$, that is, $u(s) = \frac{\log(1+s)}{s} \leq 1 - c$ for $s \geq 1$. As easy to check, the function $u(s)$ is decreasing on the whole positive axis, so $u(s) \leq \log 2$ in $s \geq 1$. Hence, one may take $c = 1 - \log 2 > \frac{1}{8}$, and thus $\psi(t) \geq \frac{t-1}{8}$. The two bounds yield

$$\psi(t) \geq \frac{1}{8}\,\min\{|t-1|, |t-1|^2\}, \quad t > 0.$$

Let us summarize.

Lemma 3.2. Given a random vector X with mean a and covariance matrix V with eigenvalues σ_i^2, we have

$$D(X||Z) \geq D(X) + \frac{1}{2}\,|a|^2 + \frac{1}{16}\,\sum_{i=1}^{d}\min\left\{|\sigma_i^2 - 1|, (\sigma_i^2 - 1)^2\right\}.$$

In particular, putting $D = D(X||Z)$, we have

(a) $|a|^2 \le 2D$;

(b) $|\sigma_i^2 - 1| \le 4\sqrt{D} + 16D$ for all $i \le d$;

(c) $|\mathbb{E}\,|X|^2 - d\,| \le 4d\sqrt{D} + 16d\,D$.

Here, the closeness of all σ_i^2 to 1 may also be stated as closeness of V to the identity $d\times d$ matrix I_d in the (squared) Hilbert–Schmidt norm $\|V-I_d\|_{\rm HS}^2 = \sum_{i=1}^d (\sigma_i^2-1)^2$. These bounds have an application in the problem where one needs to determine whether or not there is convergence in relative entropy for a sequence of random vectors.

Corollary 3.3. Given a sequence of random vectors Z_n in $\mathbb{R}^d$ with means a_n and covariance matrices V_n, the property $D(Z_n||Z) \to 0$ as $n \to \infty$ is equivalent to the next three conditions:

$$D(Z_n) \to 0; \quad a_n \to 0; \quad V_n \to I_d.$$

Proof. of Proposition 3.1 First recall that

$$D(Z_n||Z) \ = \ -h(Z_n) + \frac{d}{2}\,\log(2\pi) + \frac{1}{2}\,\mathbb{E}\,|Z_n|^2, \qquad h(Z) = \frac{d}{2}\,\log(2\pi) + \frac{d}{2}.$$

Hence, if $\mathbb{E}\,|Z_n|^2 \to d$ like in (a), then $D(Z_n||Z) \to 0 \Leftrightarrow h(Z_n) \to h(Z)$. To show that the conditions (a)–(b) are sufficient for the convergence in D, denote by f_n the characteristic functions of Z_n. By the assumption and applying the Plancherel theorem,

$$\Delta_n \ = \ (2\pi)^{-d/2}\,\|f_n - g\|_2 \ \to \ 0$$

as $n \to \infty$. Define the random vectors $\tilde Z_n = b_n Z_n$, where $b_n^2 = d/\mathbb{E}\,|Z_n|^2$ ($b_n > 0$), so that $\mathbb{E}\,|\tilde Z_n|^2 = d$. They have densities $\tilde p_n(x) = \frac{1}{b_n^d}\,p_n(\frac{x}{b_n})$ with characteristic functions

$$\tilde f_n(t) = \mathbb{E}\,e^{i\langle t,\tilde Z_n\rangle} = f_n(b_n t), \qquad t \in \mathbb{R}^d.$$

Using $b_n \to 1$ and applying the Plancherel theorem once more together with the triangle inequality in L^2, we then get

$$\begin{aligned}
\tilde\Delta_n = \ (2\pi)^{-d/2}\,\|\tilde f_n - g\|_2 \ &= \ \frac{1}{(2\pi b_n)^{d/2}}\,\|f_n(t) - g(t/b_n)\|_2 \\
&\le \ \frac{1}{(2\pi b_n)^{d/2}}\,\|f_n(t) - g(t)\|_2 + \frac{1}{(2\pi b_n)^{d/2}}\,\|g(t/b_n) - g(t)\|_2 \\
&= \ \frac{1}{b_n^{d/2}}\,\Delta_n + \frac{1}{(2\pi b_n)^{d/2}}\,\|g(t/b_n) - g(t)\|_2.
\end{aligned}$$

Here, the last norm tends to zero, so, $\tilde{\Delta}_n \to 0$. We are in position to apply the upper bound (9) of Proposition 2.1 to $X = \tilde{Z}_n$, which yields $D(\tilde{Z}_n||Z) \to 0$ and thus

$$D(Z_n||Z) = D(\tilde{Z}_n||Z) - d\log b_n + \frac{d}{2}(b_n^2 - 1) \to 0. \tag{14}$$

Conversely, assuming that $D(Z_n||Z) \to 0$ and applying Corollary 3.3, we get the property (a). Hence, $b_n^2 = d/\mathbb{E}\,|Z_n|^2 \to 1$, and $D(\tilde{Z}_n||Z) \to 0$ according to the formula (14). By the assumption, $\tilde{p}_n$ are uniformly bounded, that is, $\tilde{p}_n(x) \le M$ with some constant M. We are in position to apply the lower bound (10), which yields $\tilde{\Delta}_n \to 0$ and therefore

$$\Delta_n = b_n^{d/2}\,(2\pi)^{-d/2}\,\|\tilde{f}_n(t) - g(b_n t)\|_2 \le b_n^{d/2}\,\tilde{\Delta}_n + b_n^{d/2}\,(2\pi)^{-d/2}\,\|g(t) - g(b_n t)\|_2 \to 0. \qquad \blacksquare$$

4 Proof of Theorems 1.1-1.2

From now on, let the random vectors Z_n be defined as the normalized sums according to (1). The proof of Theorem 1.1 is based on the following statement obtained in [8].

Lemma 4.1. ([8, Theorem 1.3]) There exists $T > 0$ depending on the distribution of X_1 with the following property. If f is supported on the ball $|t| \le T$, then the random vectors Z_n have continuous densities p_n such that

$$\|p_n - \varphi\|_\infty = \sup_x |p_n(x) - \varphi(x)| \to 0 \quad \text{as } n \to \infty. \tag{15}$$

If β_3 is finite, one may take $T = 1/\beta_3$. If X_1 has a non-lattice distribution, T may be arbitrary.

Recall that, in Theorems 1.1-1.2 we assume that $\mathbb{E}\,|X|^2 < \infty$, which implies $\mathbb{E}\,|Z_n|^2 = \frac{1}{n}\,\mathbb{E}\,|X|^2 + d \to d$ as $n \to \infty$. In addition, the uniform convergence (15) is stronger than

$$\|p_n - \varphi\|_2 \to 0 \quad \text{as } n \to \infty, \tag{16}$$

since

$$\begin{aligned}\|p_n - \varphi\|_2^2 &= \int_{\mathbb{R}^d} (p_n(x) - \varphi(x))^2\,dx \\ &\le \|p_n - \varphi\|_\infty \int_{\mathbb{R}^d} |p_n(x) - \varphi(x)|\,dx \le 2\,\|p_n - \varphi\|_\infty.\end{aligned}$$

By Proposition 3.1, both properties ensure that $D(Z_n||Z) \to 0$, and we obtain Theorem 1.1.

Now, let us turn to the Bernoulli case, that is, when X_1 has a uniform distribution on the discrete cube $\{-1,1\}^d$. Theorem 1.2 may slightly be refined in one direction by weakening the condition (4). As before, $\|t\|$ denotes the distance from the point $t \in \mathbb{R}^d$ to the lattice $\pi\mathbb{Z}^d$.

Theorem 4.2. Suppose that the characteristic function of X satisfies

$$f(\pi k) = 0 \quad \text{for all } \ k \in \mathbb{Z}^d, \ k \neq 0, \tag{17}$$

together with

$$\int_{\mathbb{R}^d} \frac{|f(t)|\,|f'(t)|}{\|t\|^{d-1}}\,\mathrm{d}t < \infty. \tag{18}$$

Then we have the entropic CLT, that is, $D(Z_n||Z) \to 0$ as $n \to \infty$. Conversely, if the entropic CLT holds together with

$$\int_{\mathbb{R}^d} |f(t)|\,\mathrm{d}t < \infty, \quad \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\,\mathrm{d}t < \infty, \tag{19}$$

then f satisfies (17). In this case the uniform local limit theorem (15) takes place.

The point of the refinement is that (18) is weaker than (19), which is exactly the condition (4) in Theorem 1.2. In dimension $d = 1$, (18) is fulfilled whenever f and f' are in L^2 (by Cauchy's inequality), that is, when the density p of the random variable X satisfies

$$\int_{-\infty}^{\infty} (1 + x^2)\, p(x)^2\,\mathrm{d}x < \infty$$

(which holds automatically, if p is bounded). If $d \geq 2$, (18) is fulfilled under the decay assumptions (6) with a weaker parameter constraint $\alpha > \frac{1}{2}$. This is the case, for example, where X is uniformly distributed in the (solid) cube $[-1,1]^d$, while (19) does not hold. In [8], it was shown that the properties (17)-(18) imply the L^2-convergence of densities (16), while (17) together with a stronger assumption (19) leads to the uniform convergence (15). Hence, we can apply Proposition 3.1 to conclude that $D(Z_n||Z) \to 0$. It was also shown there that the property (17) is fulfilled under the L^2-convergence (16). In order to arrive at a similar conclusion under an apriori weaker entropic CLT, we involve the assumption (19) and prove here:

Lemma 4.3. Suppose that X_1 has a uniform distribution on the discrete cube $\{-1,1\}^d$. If the condition (19) is fulfilled, then Z_n have uniformly bounded densities p_n.

Having this assertion, we therefore complete the proof of Theorem 4.2 and of Theorem 1.2 by appealing to Proposition 3.1 once more.

Proof. of Lemma 4.3 Put $v(t) = \cos(t_1)\dots\cos(t_d)$ for $t = (t_1,\dots,t_d) \in \mathbb{R}^d$. By the assumption (19), the characteristic functions

$$f_n(t) = f\left(\frac{t}{\sqrt{n}}\right) v^n\left(\frac{t}{\sqrt{n}}\right)$$

are integrable. Hence, Z_n have continuous densities given by the Fourier inversion formula

$$p_n(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f_n(t)\,dt. \tag{20}$$

Let us partition $\mathbb{R}^d$ into the cubes $Q_k = Q+\pi k$, $Q = [-\frac{\pi}{2},\frac{\pi}{2}]^d$, $k \in \mathbb{Z}^d$, so that $\|t\| = |t-\pi k|$ for $t \in Q_k$. Splitting the integration in (20), we can write

$$p_n(x) = \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d} I_{n,k}(x), \qquad I_{n,k}(x) = n^{d/2}\int_{Q_k} e^{-i\langle t,x\rangle\sqrt{n}} f(t) v^n(t)\,dt.$$

Putting $w_k(t) = f(\pi k + t)$ and using the periodicity of the cosine function together with the bound $0 \le \cos(u) \le e^{-u^2/2}$ for $|u| \le \frac{\pi}{2}$, we have

$$|I_{n,k}(x)| \le n^{d/2} J_{n,k}, \quad J_{n,k} = \int_Q |w_k(t)|\, e^{-n|t|^2/2}\,dt.$$

By Taylor's formula,

$$|f(\pi k + t) - f(\pi k)| \le |t|\int_0^1 |f'(\pi k + \xi t)|\,d\xi, \qquad t \in \mathbb{R}^d. \tag{21}$$

Hence, changing the variable $\xi t = s$, we get

$$\begin{aligned}\int_Q |f(\pi k + t) - f(\pi k)|\,dt &\le \int_0^1\int_Q |f'(\pi k + \xi t)|\,|t|\,d\xi\,dt \\ &= \int_Q \left[|f'(\pi k + s)|\,|s|\int_{\frac{2}{\pi}\|s\|_\infty}^1 \frac{d\xi}{\xi^{d+1}}\right] ds \le c_d\int_Q \frac{|f'(\pi k + s)|}{|s|^{d-1}}\,ds\end{aligned}$$

with some constant c_d depending on d only, where $\|s\|_\infty = \max_k |s_k|$ for $s = (s_1,\dots,s_d) \in \mathbb{R}^d$. Hence,

$$\pi^d\,|w_k(0)| = \pi^d\,|f(\pi k)| \le \int_{Q_k} |f(t)|\,dt + c_d\int_{Q_k} \frac{|f'(t)|}{\|t\|^{d-1}}\,dt.$$

The next summation over all k leads to

$$\sum_{k\in\mathbb{Z}^d} |w_k(0)| = \sum_{k\in\mathbb{Z}^d} |f(\pi k)| \le \frac{1}{\pi^d}\int_{\mathbb{R}^d} |f(t)|\,dt + \frac{c_d}{\pi^d}\int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\,dt < \infty, \tag{22}$$

where we applied the assumption (19). Put

$$\widetilde{J}_{n,k} = \int_Q (|w_k(t)| - |w_k(0)|)\, e^{-n|t|^2/2}\,dt.$$

By (21),

$$|w_k(t)| \le |w_k(0)| + |t| \int_0^1 |w_k'(\xi t)|\,d\xi.$$

Hence, again changing the variable $\xi t = s$, and then $\xi = \sqrt{n}\,|s|\,\frac{1}{u}$, we get

$$\begin{aligned}
\widetilde{J}_{n,k} &\le \int_Q \int_0^1 |t|\,|w_k'(\xi t)|\, e^{-n|t|^2/2}\,dt\,d\xi \\
&= \int_Q |w_k'(s)|\,|s| \left[\int_{\frac{2}{\pi}\|s\|_\infty}^1 \xi^{-d-1}\, e^{-n|s|^2/2\xi^2}\,d\xi\right] ds \\
&\le n^{-d/2} \int_Q |w_k'(s)|\,|s|^{-(d-1)} \left[\int_{|s|\sqrt{n}}^\infty u^{d-1}\, e^{-u^2/2}\,du\right] ds \\
&\le c_d\, n^{-d/2} \int_Q \frac{|w_k'(s)|}{|s|^{d-1}}\, e^{-n|s|^2/2}\,ds
\end{aligned}$$

with some constant c_d depending on the dimension, only. Performing summation over all k, we get

$$n^{d/2} \sum_{k\in\mathbb{Z}^d} \widetilde{J}_{n,k} \le c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\, e^{-n\|t\|^2/2}\,dt \le c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\,dt.$$

Due to (22), with some other d-dependent constants

$$n^{d/2} \sum_{k\in\mathbb{Z}^d} J_{n,k} \le c_d \int_{\mathbb{R}^d} |f(t)|\,dt + c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}}\,dt < \infty,$$

and thus $\sum_{k\in\mathbb{Z}^d} |I_{n,k}(x)|$ is bounded by a constant that does not depend on x. ■

Remark 4.4. To better realize the meaning of Theorem 1.1, let us also comment on the relationship between the entropic and transport CLTs. Given two random vectors X and

Y in $\mathbb{R}^d$ with distributions μ and ν, respectively, the (quadratic) Kantorovich distance is defined as

$$W_2(\mu,\nu) \,=\, W_2(X,Y) \,=\, \inf_\lambda \bigg(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^2 \, \mathrm{d}\lambda(x,y)\bigg)^{1/2},$$

where the infimum is running over all (Borel) probability measures λ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. It represents a metric in the space $M_2(\mathbb{R}^d)$ of all probability measures on $\mathbb{R}^d$ with finite 2nd moment, which is closely related to the weak topology. More precisely, given a sequence μ_n and a "point" ν in $M_2(\mathbb{R}^d)$, the convergence $W_2(\mu_n,\nu) \to 0$ holds true as $n \to \infty$ if and only if μ_n are weakly convergent to ν, that is,

$$\int_{\mathbb{R}^d} u(x)\, \mathrm{d}\mu_n(x) \to \int_{\mathbb{R}^d} u(x)\, \mathrm{d}\nu(x)$$

for any bounded continuous function u on $\mathbb{R}^d$, and $\int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\mu_n(x) \to \int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\nu(x)$ (cf., e.g., [31, p. 212]).

When ν is the standard Gaussian measure on $\mathbb{R}^d$, the relationship of W_2 with relative entropy was emphasized by Talagrand [30] who showed that

$$W_2^2(X,Z) \le 2D(X||Z)$$

holding for any random vector X in $\mathbb{R}^d$ with Z a standard normal random vector. Returning to the setting of Theorem 1.1, define the normalized sums

$$Z_n' = Z_n - \frac{1}{\sqrt{n}}\, X = \frac{1}{\sqrt{n}}\, (X_1 + \cdots + X_n).$$

By the classical CLT, the distributions μ_n' of Z_n' are weakly convergent to the Gaussian limit ν. Since also $\mathbb{E}\,|Z_n|^2 \,=\, \mathbb{E}\,|Z|^2 \,=\, d$, the above characterization of the convergence in the space $M_2(\mathbb{R}^d)$ ensures that $W_2(\mu_n',\nu) \to 0$, which is a transport CLT. A similar conclusion can also be made on the basis of Theorem 1.1. Indeed, choose for f a characteristic function supported on a suitable small ball $|t| \le T$, so that $D(Z_n||Z) \to 0$, by (3). Applying the Talagrand transport-entropy inequality, we get

$$W_2^2(Z_n',Z) \,\le\, 2W_2^2(Z_n,Z) + \frac{2}{n}\,\mathbb{E}\,|X|^2 \,\le\, 4D(Z_n||Z) + \frac{2}{n}\,\mathbb{E}\,|X|^2 \,\to\, 0.$$

A similar approach was used in [4] to study the rate of convergence in the one-dimensional transport CLT under the 4th moment assumption.

5 Entropy Bounds

Let $(X_n)_{n\geq 1}$ be a sequence of integer valued random vectors in $\mathbb{R}^d$, and let X be a continuous random vector in $\mathbb{R}^d$ with finite 2nd moment, independent of this sequence. As before, we define the normalized sums

$$Z_n = \frac{1}{\sqrt{n}}\,(X + X_1 + \cdots + X_n).$$

As is well known, when the 2nd moment $\mathbb{E}\,|U|^2$ of a continuous random vector U in $\mathbb{R}^d$ is fixed, its entropy is maximized on the normal distribution with the same 2nd moment (cf., e.g., [13]). In the case of independent and isotropic X_n's, we have $\mathbb{E}\,|Z_n|^2 = \frac{1}{n}\,\mathbb{E}\,|X|^2 + d \to d$ as $n \to \infty$. Hence, $\limsup_{n\to\infty} h(Z_n) \leq h(Z)$, where Z is a standard normal random vector in $\mathbb{R}^d$. The argument to derive a similar bound $\limsup_{n\to\infty} h(Z_n) \leq h(Z) + h(X)$ is based on two elementary lemmas, which involve the discrete Shannon entropy

$$H(Y) = -\sum_k p_k \log p_k.$$

Here, Y is a discrete random vector taking at most countably many values, say y_k, with probabilities p_k, respectively.

Lemma 5.1. Let X be a continuous random vector, and let Y be a discrete random vector independent of X, both with values in the Euclidean space $\mathbb{R}^d$. Then

$$h(X+Y) \leq h(X) + H(Y).$$

Lemma 5.1 can be derived implicitly from the ideas of [28] about the entropy of mixtures of discrete and continuous random variables. An explicit statement appears in [32, Lemma 11.2] (see also [26]). We include a proof for completeness:

Proof. Denote by p the density of X and let $p_k = P\{Y = y_k\}$ for some finite or infinite sequence y_k. Since X and Y are independent, $X+Y$ has density

$$q(z) = \sum_k p_k p(z - y_k).$$

We use the convention $u\log(u) = 0$ if $u = 0$. Note that, if $p(z - y_k) = 0$, then

$$p_k p(z - y_k) \log \sum_i p_i p(z - y_i) = 0 = p_k p(z - y_k) \log(p_k p(z - y_k)),$$

while in the case $p(z - y_k) > 0$, we have

$$\begin{aligned} & p_k p(z - y_k) \log \sum_i p_i p(z - y_i) \\ & \quad = p_k p(z - y_k) \log \left(p_k p(z - y_k) + \sum_{i \neq k} p_i p(z - y_i) \right) \\ & \quad = \; p_k p(z - y_k) \left[\log(p_k p(z - y_k)) + \log \left(1 + \frac{\sum_{i \neq k} p_i p(z - y_i)}{p_k p(z - y_k)} \right) \right] \\ & \quad \geq \; p_k p(z - y_k) \; \log(p_k p(z - y_k)). \end{aligned}$$

Hence, for all z,

$$p_k p(z - y_k) \; \log \sum_i p_i p(z - y_i) \; \geq \; p_k p(z - y_k) \; \log(p_k p(z - y_k)).$$

We may therefore conclude that

$$\begin{aligned} h(X+Y) \;\; & = \;\; - \int_{\mathbb{R}^d} q(z) \log q(z) \, \mathrm{d}z \\ & = \;\; - \sum_k \int_{\mathbb{R}^d} p_k p(z - y_k) \; \log \sum_i p_i p(z - y_i) \, \mathrm{d}z \\ & \leq \;\; - \sum_k \int_{\mathbb{R}^d} p_k p(z - y_k) \log(p_k p(z - y_k)) \, \mathrm{d}z \\ & = \;\; - \sum_k p_k \left(\int_{\mathbb{R}^d} p(z - y_k) \log p_k \, \mathrm{d}z + \int_{\mathbb{R}^d} p(z - y_k) \log p(z - y_k) \, \mathrm{d}z \right) \\ & = \;\; h(X) + H(Y). \end{aligned}$$

■

Let us note that a recent sharpening of Lemma 5.1 appears in [25, Theorem III.1], where it is shown that

$$h(X+Y) \leq h(X|Y) + TH(Y),$$

where $h(X|Y)$ is the conditional entropy, reducing to $h(X)$ on independence, and T is the supremum of the total variation of the conditional densities from their “mixture complements”, necessarily $T \leq 1$.

The following lemma is standard and has been used in several applications (see [24]):

Lemma 5.2. For any integer valued random variable Y with finite 2nd moment,

$$H(Y) \le \frac{1}{2}\log\Big(2\pi e\Big(\mathrm{Var}(Y) + \frac{1}{12}\Big)\Big). \tag{23}$$

The proof of Lemma 5.2, that we include for completeness, also combines both discrete and differential entropy:

Proof. Put $p_k = \mathbb{P}\{Y = k\}$, $k \in \mathbb{Z}$. Consider a continuous random variable $\widetilde{Y}$ with density q defined to be

$$q(x) = p_k \quad \text{if } x \in \big(k - \tfrac{1}{2}, k + \tfrac{1}{2}\big).$$

In other words,

$$q(x) = \sum_k p_k 1_{(k-\frac{1}{2},k+\frac{1}{2})}(x), \quad x \in \mathbb{R}.$$

Note that

$$\mathbb{E}\widetilde{Y} = \sum_k p_k \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} x\,\mathrm{d}x = \sum_k \frac{p_k}{2}\Big(\Big(k+\frac{1}{2}\Big)^2 - \Big(k-\frac{1}{2}\Big)^2\Big) = \sum_k k p_k = \mathbb{E}Y$$

and similarly

$$\mathbb{E}\widetilde{Y}^2 = \sum_k p_k \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} x^2\,\mathrm{d}x = \mathbb{E}Y^2 + \frac{1}{12}.$$

Hence, $\mathrm{Var}(\widetilde{Y}) = \mathrm{Var}(Y) + \frac{1}{12}$. Also,

$$\begin{aligned} h(\widetilde{Y}) &= -\int_{-\infty}^{\infty} \sum_k p_k 1_{(k-\frac{1}{2},k+\frac{1}{2})}(x)\, \log \sum_j p_j 1_{(j-\frac{1}{2},j+\frac{1}{2})}(x)\,\mathrm{d}x \\ &= -\sum_k p_k \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \log p_k\,\mathrm{d}x \;=\; H(Y). \end{aligned}$$

Now, since Gaussian distributions maximize the differential entropy for a fixed variance, we conclude that

$$H(Y) = h(\widetilde{Y}) \le \frac{1}{2}\log\big(2\pi e\,\mathrm{Var}(\widetilde{Y})\big) = \frac{1}{2}\log\Big(2\pi e\Big(\mathrm{Var}(Y) + \frac{1}{12}\Big)\Big).$$

■

We are now prepared to establish Theorem 1.3, in fact under somewhat weaker assumptions.

Theorem 5.3. Given a sequence $X_n = (X_{n,1}, \dots, X_{n,d})$ of random vectors with values in $\mathbb{Z}^d$, independent of X, assume that for each $k \le d$, the components $X_{n,k}$, $n \ge 1$, are uncorrelated and have variance one. Then,

$$\limsup_{n\to\infty} h(Z_n) \le h(X) + h(Z).$$

Proof. Putting $S_n = X_1 + \cdots + X_n$ and applying Lemma 5.1, we get

$$\begin{aligned} h(Z_n) = h\Big(\frac{X+S_n}{\sqrt{n}}\Big) &= h(X+S_n) - \frac{d}{2}\log n \\ &\le h(X) + H(S_n) - \frac{d}{2}\log n. \end{aligned}$$

Note that

$$S_n = (S_{n,1}, \dots, S_{n,d}), \qquad S_{n,k} = X_{1,k} + \cdots + X_{n,k} \quad (1 \le k \le d).$$

By the well-known subadditivity of entropy along components of a random vector (an abstract property on product spaces that is irrelevant to the independence assumption, cf., e.g., [20]), we have

$$H(S_n) \le H(S_{n,1}) + \cdots + H(S_{n,d}).$$

Here, the entropy functional on the left is applied to the d-dimensional random vector, while on the right-hand side of this inequality we deal with one-dimensional entropies. For each $k \le d$, the k-th component $S_{n,k}$ of the random vector S_n represents the sum of n uncorrelated integer valued random variables with variance one, so that $\mathrm{Var}(S_{n,k}) = n$. Hence, by (23) applied to $Y = S_{n,k}$, we have

$$H(S_{n,k}) \le \frac{1}{2}\log\Big(2\pi e\Big(n + \frac{1}{12}\Big)\Big) = \frac{1}{2}\log(2\pi e n) + O(1/n),$$

and therefore

$$H(S_n) \le \frac{d}{2}\log(2\pi e n) + O(1/n).$$

We conclude that

$$\limsup_{n\to\infty} h(Z_n) \le h(X) + \frac{d}{2}\log(2\pi e) = h(X) + h(Z).$$

■

Acknowledgments

The authors are grateful to both referees for useful comments.

Funding

The work was partially supported by Simons Foundation [to S.B.] and National Science Foundation [grant DMS-1855575 to S.B.].

[1] Artstein, S., K. Ball, F. Barthe, and A. Naor. "On the rate of convergence in the entropic central limit theorem." *Probab. Theory Related Fields* 129, no. 3 (2004): 381–90.

[2] Barron, A. R. "Entropy and the central limit theorem." *Ann. Probab.* 14, no. 1 (1986): 336–42.

[3] Bishop, C. M. *Pattern Recognition and Machine Learning.* Information Science and Statistics. New York: Springer, 2006. xx+738 pp.

[4] Bobkov, S. G. "Entropic approach to E. Rio's central limit theorem for W_2 transport distance." *Statist. Probab. Lett* 83, no. 7 (2013): 1644–8.

[5] Bobkov, S. G., G. P. Chistyakov, and F. Götze. "Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem." *Ann. Probab.* 41, no. 4 (2013): 2479–512.

[6] Bobkov, S. G. and M. Madiman. "The entropy per coordinate of a random vector is highly constrained under convexity conditions." *IEEE Trans. Inform. Theory* 57, no. 8 (2011): 4940–54.

[7] Bobkov, S. G. and M. Madiman. "Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures." *J. Funct. Anal.* 262, no. 7 (2012): 3309–39.

[8] Bobkov, S. G. and A. Marsiglietti. "Local limit theorems for smoothed Bernoulli and other convolutions." *Theory Probab. Appl.*64,4 (2019) Preprint:arXiv:1901.02984 [math.PR].

[9] Bourgain, J. "On high-dimensional maximal functions associated to convex bodies." *Amer. J. Math.* 108, no. 6 (1986): 1467–76.

[10] Caglar, U., M. Fradelizi, O. Guédon, J. Lehec, C. Schütt, and E Werner. "Functional versions of L^p-affine surface area and entropy inequalities." *Int. Math. Res. Not.* 2016 no. 4(2016), 1223–1250.

[11] Caglar, U. and E. Werner. "Divergence for s-concave and log concave functions." *Adv. Math.* 257 (2014): 219–47.

[12] Chung, H. W., B. M. Sadler, and A. O. Hero. "Bounds on variance for unimodal distributions." *IEEE Trans. Inform. Theory* 63, no. 11 (2017): 6936–49.

[13] Cover, T. M. and J. A. Thomas. *Elements of Information Theory*., 2nd ed. Hoboken,NJ: *Wiley-Interscience [John Wiley & Sons]*, 2006, xxiv+748 pp.

[14] Eldan, R, D. Mikulincer, and A Zhai. "The CLT in high dimensions: quantitative bounds via martingale embedding." Preprint: arXiv:1806.09087 [math.PR].

[15] van Erven, T. and P. Harremoës. "Rényi divergence and Kullback-Leibler divergence." *IEEE Trans. Inform. Theory* 60, no. 7 (2014): 3797–820.

[16] Hao, J. and V. Jog. "An entropy inequality for symmetric random variables." Preprint: arXiv:1801.03868 [cs.IT].

[17] Harremoës, P. and C Vignat. "A short information theoretic proof of CLT." (2004): preprint. Available at http://www-syscom.univ-mlv.fr/vignat/Signal/CLT.pdf.

[18] Johnson, O Information Theory and the Central Limit Theorem. London: Imperial College Press, 2004. xiv+209 pp.

[19] Johnson, O. and A. Barron. "Fisher information inequalities and the central limit theorem." *Probab. Theory Related Fields* 129, no. 3 (2004): 391–409.

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

[21] Madiman, M., J. Melbourne, and P Xu. "Forward and Reverse Entropy Power Inequalities in Convex Geometry." In *Convexity and Concentration*, pp. 427–85. *IMA Volumes in Mathematics and its Applications* 161. New York: Springer, 2017.

[22] Madiman, M., P. Nayar, and T. Tkocz. "Entropy versus variance for symmetric log-concave random variables and related problems." *Preprint: arXiv:1811.00345 [cs.IT]*.

[23] Marsiglietti, A. and V. Kostina. "A lower bound on the differential entropy of log-concave random vectors with applications." *Entropy* 20, no. 3 (2018): 24.

[24] Massey, J. L. "On the Entropy of Integer-Valued Random Variables." In *Proceedings of 1988 Beijing International Workshop on Information Theory, pp. C1.1–C1.4*, 1988.

[25] Melbourne, J., M. Madiman, and M. V. Salapaka. "Relationships between certain f-divergences." *57th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*2019 (2019): 1068–73.

[26] Melbourne, J., S. Talukdar, S. Bhaban, and M. V. Salapaka. "Error bounds on a mixed entropy Inequality." Preprint: arXiv:1805.11260 [cs.IT].

[27] Miclo, L. "Notes on the Speed of Entropic Convergence in the Central Limit Theorem." In *Stochastic Inequalities and Applications*. Progress in Probability 56. Basel: Birkhäuser, 129–56, 2003.

[28] Nair, C, B. Prabhakar, and D Shah. "On entropy for mixtures of discrete and continuous variables." arXiv:cs/0607075 [cs.IT].

[29] Rudin, W Functional Analysis, 2nd ed. *International Series in Pure and Applied Mathematics*. New York: McGraw-Hill, Inc., 1991. xviii+424 pp.

[30] Talagrand, M. "Transportation cost for Gaussian and other product measures." *Geom. Funct. Anal.* 6, no. 3 (1996): 587–600.

[31] Villani, C. *Topics in Optimal Transportation*. *Graduate Studies in Mathematics 58*. Providence, RI: American Mathematical Society, 2003, xvi+370 pp.

[32] Wang, L. and M. Madiman. "Beyond the entropy power inequality, via rearrangements." *IEEE Trans. Inform. Theory* 60, no. 9 (2014): 5116–37.