Abstract. We explore an asymptotic behavior of entropies for sums of independent random variables that are convoluted 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 mean zero and an identity covariance matrix. By the central limit theorem (CLT), given a random vector \(X\) in \(\mathbb{R}^d\), the normalized sums

\[ Z_n = \frac{1}{\sqrt{n}} (X + X_1 + \cdots + X_n) \]

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

\[ \varphi(x) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2}, \quad x \in \mathbb{R}^d. \]

Suppose that \(X\) has a finite second 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 property (that is, the weak CLT) may be strengthened as convergence of entropies

\[ h(Z_n) = -\int_{\mathbb{R}^d} p_n(x) \log p_n(x) \, dx \]

to the entropy of the Gaussian limit \(Z\). The usual entropic CLT corresponds to the purely 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, cf. \([1, 7, 8, 2]\). What also seems remarkable, the presence of a small non-zero noise \(X/\sqrt{n}\) in (1.1) may potentially enlarge the range of applicability of the entropic CLT. Here is one observation in this direction in terms of the characteristic function

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

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

\[ h(Z_n) \to h(Z) \quad \text{as} \quad n \to \infty. \]
This convergence also holds for lattice distributions, if $f$ is supported on the ball $|t| \leq 1/\beta_3$, assuming that the 3-rd absolute moment

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

is finite.

The entropic CLT (1.3) may equivalently be stated as the convergence

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

for the Kullback-Leibler distance (also called relative entropy or an informational divergence).

In general, the hypothesis on the support of $f$ 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

$$\int_{\mathbb{R}^d} |f(t)| \, dt < \infty, \quad \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}} \, dt < \infty,$$  \hspace{1cm} (1.4)

where $\|t\|$ denotes the distance from the point $t$ to the lattice $\pi \mathbb{Z}^d$. Then, the entropic CLT (1.3) holds true, if and only if

$$f(\pi k) = 0 \quad \text{for all} \quad k \in \mathbb{Z}^d, \ k \neq 0.$$  \hspace{1cm} (1.5)

The second moment assumption on $X$ guarantees that $f$ has a bounded continuous derivative $f'(t) = \nabla f(t)$ with its Euclidean length $|f'(t)|$. In dimension $d = 1$, the condition (1.4) is fulfilled, as long as both $f$ and $f'$ are in $L^1$. If $d \geq 2$, (1.4) is more complicated, but is fulfilled, for example, under decay assumptions such as

$$|f(t)| \leq \frac{c}{((1 + |t_1|) \ldots (1 + |t_d|))^\alpha}, \quad |f'(t)| \leq \frac{c}{((1 + |t_1|) \ldots (1 + |t_d|))^\alpha},$$  \hspace{1cm} (1.6)

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

Although an information-theoretic meaning of the property (1.5) is not clear, it is indeed connected with the entropy functional $h(X)$. The following observation in the same setting as in Theorem 1.2 is of an independent interest. As before, we assume that $X$ is a continuous random vector in $\mathbb{R}^d$ with finite second moment.

**Theorem 1.3.** Assume that $X_1$ has a uniform distribution on $\{-1,1\}^d$. If $h(Z_n) \to h(Z)$ as $n \to \infty$, then necessarily $h(X) \geq 0$.

Lower bounds on differential entropy are of interest in the study of certain fundamental limits of information theory. For example, applications to rate-distortion theory and channel capacity were put forward in [10] (see also [4], [6], [9], [14]).
The paper is organized as follows. We start in Section 2 with general upper and lower bounds on the Kullback-Leibler distance
\[ D(X || Z) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{\varphi(x)} \, dx \] (1.7)
from the distribution of \( X \) to the standard normal law in terms of the \( L^2 \)-distance
\[ \Delta = \| p - \varphi \|_2 = \left( \int_{\mathbb{R}^d} (p(x) - \varphi(x))^2 \, dx \right)^{1/2}. \] (1.8)
Throughout, \( Z \) denotes a standard normal random vector in \( \mathbb{R}^d \), thus with density \( \varphi \) as in (1.2) and with characteristic function
\[ g(t) = \mathbb{E} e^{i(t,Z)} = \int_{\mathbb{R}^d} e^{i(t,x)} \varphi(x) \, dx = e^{-|t|^2/2}, \quad t \in \mathbb{R}^d. \]
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 \( L^2 \). Theorem 1.1 and Theorem 1.2 (in a somewhat refined form) are proved in Section 4; the proofs employ recent results obtained in [3] on local limit theorems with respect to the \( L^2 \) and \( L^\infty \)-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 \( \Delta \) be defined according to (1.8).

**Proposition 2.1.** Suppose that \( \mathbb{E} |X|^2 = d \). If \( \Delta \leq 1/e \), then
\[ D(X || Z) \leq c_d \Delta \log \frac{d+4}{1} (1/\Delta) \] (2.1)
with some constant \( c_d > 0 \) depending on \( d \) only. Moreover, if \( \sup_x p(x) \leq M \) for some constant \( M \geq (2\pi)^{-d/2} \), then
\[ D(X || Z) \geq \frac{1}{2M} \Delta^2. \] (2.2)

First we collect a few elementary large deviation bounds.

**Lemma 2.2.** For any \( T \geq 1 \),
\[ a) \int_{|x| \geq T} \varphi(x) \, dx \leq 2d T^{d-2} e^{-T^2/2}; \]
\[ b) \int_{|x| \geq T} |x|^2 \varphi(x) \, dx \leq 2d T^d e^{-T^2/2}. \]

**Proof.** Clearly, \( a) \) follows from \( b) \). To derive the second bound, write
\[ \mathbb{E} |Z|^2 1_{\{|Z| \geq T\}} = \int_{|x| \geq T} |x|^2 \varphi(x) \, dx = \frac{d\omega_d}{(2\pi)^{d/2}} \int_T^{\infty} r^{d+1} e^{-r^2/2} \, dr, \]
where $\omega_d$ denotes the volume of the unit ball in $\mathbb{R}^d$. Given $c > 0$, consider the function

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

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

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

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

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

Integration by parts gives

$$c \leq \sqrt{e} \int_0^\infty r^{d+1} e^{-r^2/2} dr = \sqrt{e} d \int_0^\infty r^{d-1} e^{-r^2/2} dr,$$

so

$$\int_{|x|\geq T} |x|^2 \varphi(x) dx = \frac{\int_{T}^{\infty} r^{d+1} e^{-r^2/2} dr}{\int_0^{\infty} r^{d-1} e^{-r^2/2} dr} \leq \sqrt{e} d T^d e^{-T^2/2}.$$  

□

To get the upper bound (2.1), we also need to control the weighted quadratic tails in terms of the $L^2$-distance $\Delta$.

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

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

**Proof.** We have

$$\int_{|x|\geq T} |x|^2 p(x) dx = d - \int_{|x|\leq T} |x|^2 p(x) dx$$

$$= \int_{|x|\leq T} |x|^2 (\varphi(x) - p(x)) dx + \int_{|x|\geq T} |x|^2 \varphi(x) dx$$

$$\leq \int_{|x|\leq T} |x|^2 |p(x) - \varphi(x)| dx + \int_{|x|\geq T} |x|^2 \varphi(x) dx.$$  

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

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

where $\omega_d$ is the volume of the unit ball in $\mathbb{R}^d$. Here, $\frac{\omega_d}{d+4} < 2$. □
Lemma 2.4. For all $T \geq 1$,
\[
D(X||Z) \leq 2d T^{d-2} e^{-T^2/2} + (2\pi)^{d/2} \int_{|x| \leq T} (p(x) - \varphi(x))^2 e^{|x|^2/2} \, dx
\]
\[
+ \frac{2d-1}{2} \int_{|x| \geq T} |x|^2 p(x) \, dx + \int_{|x| \geq T} p \log p \, dx.
\]
\[(2.3)\]

Proof. In definition (1.8), we split the integration into the two regions. Using the inequality $t \log t \leq (t-1) + (t-1)^2$, $t \geq 0$, and applying the first bound of Lemma 2.2, we have
\[
\int_{|x| \leq T} \frac{p}{\varphi} \log \frac{p}{\varphi} \varphi \, dx \leq \int_{|x| \leq T} \left( \frac{p}{\varphi} - 1 \right) \varphi \, dx + \int_{|x| \leq T} \left( \frac{p}{\varphi} - 1 \right)^2 \varphi \, dx
\]
\[
= \int_{|x| \geq T} (\varphi - p) \, dx + \int_{|x| \leq T} \frac{(p - \varphi)^2}{\varphi} \, dx
\]
\[
\leq 2d T^{d-2} e^{-T^2/2} - \int_{|x| \geq T} p \, dx + (2\pi)^{d/2} \int_{|x| \leq T} (p(x) - \varphi(x))^2 e^{|x|^2/2} \, dx.
\]

For the second region $|x| \geq T$, just write
\[
\int_{|x| \geq T} p \log \frac{p}{\varphi} \, dx = \int_{|x| \geq T} p \log p \, dx
\]
\[
+ \frac{d}{2} \log(2\pi) \int_{|x| \geq T} p \, dx + \frac{1}{2} \int_{|x| \geq T} |x|^2 p(x) \, dx.
\]
Combining these relations and noting that $\log(2\pi) < 2$, we thus get
\[
D(X||Z) \leq 2d T^{d-2} e^{-T^2/2} + (2\pi)^{d/2} \int_{|x| \leq T} (p(x) - \varphi(x))^2 e^{|x|^2/2} \, dx
\]
\[
+ (d-1) \int_{|x| \geq T} p(x) \, dx + \frac{1}{2} \int_{|x| \geq T} |x|^2 p(x) \, dx + \int_{|x| \geq T} p \log p \, dx.
\]
As a consequence, we obtain:

Lemma 2.5. For all $T \geq 1$,
\[
D(X||Z) \leq (2d + 1) T^{d-2} e^{-T^2/2} + ((2\pi)^{d/2} + 1) e^{T^2/2} \Delta^2 + d \int_{|x| \geq T} |x|^2 p(x) \, dx.
\]

Proof. We use the notation $a^+ = \max(a, 0)$. Subtracting $\varphi(x)$ from $p(x)$ and then adding, one can write
\[
\int_{|x| \geq T} p \log p \, dx \leq \int_{|x| \geq T} p(x) \log^+(p(x)) \, dx
\]
\[
\leq \int_{|x| \geq T} |p(x) - \varphi(x)| \log^+(p(x)) \, dx + \int_{|x| \geq T} \varphi(x) \log^+(p(x)) \, dx.
\]
Next, let us apply Cauchy’s inequality together with the bound \((\log^+(t))^2 \leq 4e^{-2t}\) so that to estimate the last integral from above by
\[
\left( \int_{|x| \geq T} \varphi(x)^2 \, dx \right)^{1/2} \left( \int_{|x| \geq T} \left( \log^+(p(x))^2 \right) \, dx \right)^{1/2} \leq \frac{2}{e} \left( \int_{|x| \geq T} \varphi(x)^2 \, dx \right)^{1/2}.
\]
Here, using the first bound of Lemma 2.2, we have
\[
\int_{|x| \geq T} \varphi(x)^2 \, dx = \frac{1}{(4\pi)^{d/2}} \int_{|y| \geq T\sqrt{2}} \varphi(y) \, dy \leq \frac{2d}{(4\pi)^{d/2}} (T\sqrt{2})^{d-2} e^{-T^2} < T^{d-2} e^{-T^2}.
\]
Therefore,
\[
\int_{|x| \geq T} p \log p \, dx \leq \int_{|x| \geq T} |p(x) - \varphi(x)| \log^+(p(x)) \, dx + T^{d-2} e^{-T^2/2}.
\]
To simplify, the last integrand may be bounded by
\[
\frac{1}{2} (p(x) - \varphi(x))^2 + \frac{1}{2} \left( \log^+(p(x))^2 \right) \leq \frac{1}{2} (p(x) - \varphi(x))^2 + \frac{1}{2} p(x),
\]
so,
\[
\int_{|x| \geq T} p \log p \, dx \leq \frac{1}{2} \Delta^2 + \frac{1}{2} \int_{|x| \geq T} p(x) \, dx + T^{d-2} e^{-T^2/2}.
\]
Using this estimate in (2.3) together with \(e^{\frac{1}{2}x^2} \leq e^{T^2/2}\) for \(|x| \leq T\), we get
\[
D(X||Z) \leq 2d T^{d-2} e^{-T^2/2} + (2\pi)^{d/2} e^{T^2/2} \int_{|x| \leq T} \left( p(x) - \varphi(x) \right)^2 \, dx
\]
\[
+ \frac{2d - 1}{2} \int_{|x| \geq T} |x|^2 p(x) \, dx + \frac{1}{2} \Delta^2 + \frac{1}{2} \int_{|x| \geq T} p(x) \, dx + T^{d-2} e^{-T^2/2}.
\]

**Proof of Proposition 2.1.** Combining Lemma 2.5 with Lemma 2.3, we immediately get
\[
D(X||Z) \leq (2d^2 + 2d + 1) T^d e^{-T^2/2} + ((2\pi)^{d/2} + 1) e^{T^2/2} \Delta^2 + 2d T^{d+4} \Delta.
\]
To get (2.1), it remains to take here
\[
T = \sqrt{2 \log(1/\Delta) + \frac{d}{2} \log \log(1/\Delta)}.
\]

For the lower bound (2.2), 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 \(E|X|^2 = d\) to get

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

Hence

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

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

\[\Box\]

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

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

More precisely, we have:

**Proposition 3.1.** Let \((Z_n)_{n \geq 1}\) be a sequence of random vectors in \(\mathbb{R}^d\) with densities \(p_n\).

Suppose that as \(n \to \infty\)

a) \(E|Z_n|^2 \to d; \quad b) \Delta_n \to 0.\)

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 quantify the properties a) – b) in terms of the relative entropy. It makes also sense to consider an affine invariant functional

\[D(X) = \inf_{Z \text{ normal}} D(X||Z),\]

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, if and only if the distribution of \(X\) is absolutely continuous and has a finite second moment, and then this infimum is attained on the normal distribution with the same mean \(a = EX\) and covariance matrix \(V\) as for \(X\). Denote by \(\varphi_{a,V}\) the density of the normal law with these parameters, that is,

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

By the definition, if \(X\) has density \(p\), we have

\[D(X||Z) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{\varphi(x)} \, dx = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{\varphi_{a,V}(x)} \, dx\]

\[-\frac{1}{2} \log \det(V) - \frac{1}{2} \mathbb{E} \langle V^{-1}(X-a),X-a \rangle + \frac{1}{2} \mathbb{E}|X|^2.\]
Simplifying, we obtain an explicit formula

\[ D(X||Z) = D(X) + \frac{1}{2} |a|^2 + \frac{1}{2} \left( \log \frac{1}{\det(V)} + \text{Tr}(V) - d \right) \]

\[ = D(X) + \frac{1}{2} |a|^2 + \frac{1}{2} \sum_{i=1}^{d} \left( \log \frac{1}{\sigma_i^2} + \sigma_i^2 - 1 \right), \quad (3.1) \]

where \( \sigma_i^2 \) 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 first 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

\[ \psi(t) = \log \frac{1}{t} + t - 1, \quad t > 0. \]

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 \( \log t \leq (1-c)(t-1) \), i.e., \( u(s) = \log(1+s) \leq 1-c \) 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 \( \psi(t) \geq \frac{(t-1)^2}{8} \). 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 \( \sigma_i^2 \), we have

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

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

a) \(|a|^2 \leq 2D; \)

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

c) \(|E|X|^2 - d| \leq 4d\sqrt{D} + 16dD. \)

Here, the closeness of all \( \sigma_i^2 \) 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 \( \|V - I_d\|^2_{\text{HS}} = \sum_{i=1}^{d} (\sigma_i^2 - 1)^2 \). 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

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

Hence, if $\mathbb{E}|Z_n|^2 \to d$ like in a), then $D(Z_n||Z) \to 0 \iff h(Z_n) \to h(Z)$.

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 $\tilde{p}_n(x) = \frac{1}{b_n} p_n(t/b_n)$ with characteristic functions

\[ \tilde{f}_n(t) = \mathbb{E} e^{i(t;\tilde{Z}_n)} = f_n(b_n t), \quad t \in \mathbb{R}^d. \]

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

\[ \tilde{\Delta}_n = \|\tilde{p}_n - \varphi\|_2 = (2\pi)^{-d/2} \|\tilde{f}_n - g\|_2 \]

\[ = \frac{1}{(2\pi b_n)^{d/2}} \|f_n(t) - g(t/b_n)\|_2 \]

\[ \leq \frac{1}{(2\pi b_n)^{d/2}} \|f_n(t) - g(t)\|_2 + \frac{1}{(2\pi b_n)^{d/2}} \|g(t/b_n) - g(t)\|_2 \]

\[ = \frac{1}{b_n^{d/2}} \Delta_n + \frac{1}{(2\pi b_n)^{d/2}} \|g(t/b_n) - g(t)\|_2. \]

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

\[ D(Z_n||Z) = D(\tilde{Z}_n||Z) - d \log b_n + \frac{d}{2} (b_n^2 - 1) \to 0. \] (3.2)

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 (3.2). By the assumption, $\tilde{p}_n$ are uniformly bounded, that is, $\tilde{p}_n(x) \leq M$ with some constant $M$. We are in position to apply the lower bound (2.2) which yields $\tilde{\Delta}_n \to 0$ and therefore

\[ \Delta_n = b_n^{d/2} (2\pi)^{-d/2} \|\tilde{f}_n(t) - g(b_n t)\|_2 \]

\[ \leq b_n^{d/2} (2\pi)^{-d/2} \|\tilde{f}_n(t) - g(t)\|_2 + b_n^{d/2} (2\pi)^{-d/2} \|g(t) - g(b_n t)\|_2 \]

\[ = b_n^{d/2} \Delta_n + b_n^{d/2} (2\pi)^{-d/2} \|g(t) - g(b_n t)\|_2 \to 0. \]

□
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 (1.1). The proof of Theorem 1.1 is based on the following statement obtained in [3].

**Lemma 4.1.** There exists $T > 0$ depending on the distribution of $X_1$ with the following property. If $f$ is supported on the ball $|t| \leq T$, then the random vectors $Z_n$ have continuous densities $p_n$ such that

$$\sup_x |p_n(x) - \varphi(x)| \to 0 \quad \text{as} \quad n \to \infty. \tag{4.1}$$

If $\beta_3$ 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 $E|X|^2 < \infty$, which implies $E|Z_n|^2 = \frac{1}{n} E|X|^2 + d \to d$ as $n \to \infty$. In addition, the uniform convergence (4.1) is stronger than

$$\|p_n - \varphi\|_2 \to 0 \quad \text{as} \quad n \to \infty. \tag{4.2}$$

By Proposition 3.1, both properties being valid simultaneously 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 (1.4). 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

$$f(\pi k) = 0 \quad \text{for all} \quad k \in \mathbb{Z}^d, \; k \neq 0, \tag{4.3}$$

together with

$$\int_{\mathbb{R}^d} \frac{|f(t)| |f'(t)|}{\|t\|^{d-1}} \, dt < \infty. \tag{4.4}$$

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

$$\int_{\mathbb{R}^d} |f(t)| \, dt < \infty, \quad \int_{\mathbb{R}^d} \frac{|f'(t)|}{\|t\|^{d-1}} \, dt < \infty, \tag{4.5}$$

then $f$ satisfies (4.3). In this case the uniform local limit theorem (4.1) takes place.

The point of the refinement is that (4.4) is weaker than (4.5), which is exactly the condition (1.4) in Theorem 1.2. In dimension $d = 1$, (4.4) 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

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

(which holds automatically, if $p$ is bounded). If $d \geq 2$, (4.4) is fulfilled under the decay assumptions (1.6) 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 (4.5) does not hold.

In [3], it was shown that the properties (4.3)-(4.4) imply the $L^2$-convergence of densities (4.2), while (4.3) together with a stronger assumption (4.5) leads to the uniform convergence (4.1). Hence, we can apply Proposition 3.1 to conclude that $D(Z_n || Z) \to 0$. 

It was also shown there that the property (4.3) is fulfilled under the $L^2$-convergence (4.2). In order to arrive at a similar conclusion under an apriori weaker entropic CLT, we involve the assumption (4.5) and prove here:

**Lemma 4.3.** Suppose that $X_1$ has a uniform distribution on the discrete cube $\{-1,1\}^d$. If the condition (4.5) 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 $v(t) = \cos(t_1)\ldots\cos(t_d)$ for $t = (t_1,\ldots,t_d) \in \mathbb{R}^d$. By the assumption (4.5), 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

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(t,x)} f_n(t) \, dt. \quad (4.6)$$

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 (4.6), we can write

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

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

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

By Taylor’s formula,

$$|f(\pi k + t) - f(\pi k)| \leq |t| \int_0^1 |f'(\pi k + \xi t)| \, d\xi, \quad t \in \mathbb{R}^d. \quad (4.7)$$

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

$$\int_Q |f(\pi k + t) - f(\pi k)| \, dt \leq \int_0^1 \int_Q |f'(\pi k + \xi t)||t| \, d\xi \, dt = \int_Q \left[|f'(\pi k + s)||s| \int_0^1 \frac{d\xi}{\xi^{d+1}}\right] ds \leq c_d \int_Q \frac{|f'(\pi k + s)|}{|s|^{d-1}} \, ds$$

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

Hence

$$\pi^d |w_k(0)| = \pi^d |f(\pi k)| \leq \int_{Q_k} |f(t)| \, dt + c_d \int_{Q_k} \frac{|f'(t)|}{|t|^{d-1}} \, dt.$$
The next summation over all \( k \) leads to
\[
\sum_{k \in \mathbb{Z}^d} |w_k(0)| = \sum_{k \in \mathbb{Z}^d} |f(\pi k)| \leq \frac{1}{\pi^d} \int_{\mathbb{R}^d} |f(t)| \, dt + \frac{c_d}{\pi^d} \int_{\mathbb{R}^d} \frac{|f'(t)|}{||t||^{d-1}} \, dt < \infty, \tag{4.8}
\]
where we applied the assumption (4.5).

Put
\[
\tilde{J}_{n,k} = \int_Q (|w_k(t)| - |w_k(0)|) e^{-n|t|^2/2} \, dt.
\]
By (4.7),
\[
|w_k(t)| \leq |w_k(0)| + |t| \int_0^1 |w_k'(\xi t)| \, d\xi.
\]
Hence, again changing the variable \( \xi t = s \), and then \( \xi = \sqrt{n} |s| \frac{1}{u} \), we get
\[
\tilde{J}_{n,k} \leq \int_Q \int_0^1 |t| |w_k'(\xi t)| e^{-n|t|^2} \, dt \, d\xi
\]
\[
= \int_Q \left| w_k'(s) \right| |s| \left[ \int_0^1 \xi^{-d-1} e^{-n|s|^2/\xi^2} \, d\xi \right] ds
\]
\[
\leq n^{-d/2} \int_Q \left| w_k'(s) \right| |s|^{-(d-1)} \left[ \int_{|s|\sqrt{n}}^\infty u^{d-1} e^{-u^2} \, du \right] ds
\]
\[
\leq c_d n^{-d/2} \int_Q \frac{|w_k'(s)|}{|s|^{d-1}} e^{-n|s|^2/2} \, ds
\]
with some constant \( c_d \) depending on the dimension, only. Performing summation over all \( k \), we get
\[
n^{d/2} \sum_{k \in \mathbb{Z}^d} \tilde{J}_{n,k} \leq c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{||t||^{d-1}} e^{-n||t||^2/2} \, dt \leq c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{||t||^{d-1}} \, dt.
\]
Due to (4.8), with some other \( d \)-dependent constants
\[
n^{d/2} \sum_{k \in \mathbb{Z}^d} J_{n,k} \leq c_d \int_{\mathbb{R}^d} |f(t)| \, dt + c_d \int_{\mathbb{R}^d} \frac{|f'(t)|}{||t||^{d-1}} \, dt < \infty,
\]
and thus \( \sum_{k \in \mathbb{Z}^d} |I_{n,k}(x)| \) is bounded by a constant which does not depend on \( x \).

\[\square\]

5. Entropy bounds and Theorem 1.3

Let \((X_n)_{n \geq 1}\) be i.i.d. random vectors in \(\mathbb{R}^d\) uniformly distributed on the discrete cube \((-1, 1)^d\), and let \(X\) be a continuous random vector in \(\mathbb{R}^d\) with finite second moment. 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 second moment \(\mathbb{E} |Y|^2\) of a continuous random vector \(Y\) in \(\mathbb{R}^d\) is fixed, its entropy is maximized on the normal distribution with the same second moment (see, e.g., [5]). In our case, \(\mathbb{E} |Z_n|^2 = \frac{1}{n} \mathbb{E} |X|^2 + d \to d\) as \(n \to \infty\). Hence \(\lim_{n \to \infty} h(Z_n) \leq h(Z)\), where \(Z\) is a standard normal random vector in \(\mathbb{R}^d\). Assuming the entropic CLT holds true, for
the proof of Theorem 1.3, it is sufficient to derive the bound \( \limsup_{n \to \infty} h(Z_n) \leq h(Z) + h(X) \).

The argument 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 [13] about the entropy of mixtures of discrete and continuous random variables. An explicit statement appears in [15, Lemma 11.2] (see also [12]). 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

\[
p_k p(z - y_k) \log \sum_i p_i p(z - y_i) = p_k p(z - y_k) \log \left( p_k p(z - y_k) + \sum_{i \neq k} p_i p(z - y_i) \right)
\]

\[
= p_k p(z - y_k) \left[ \log(p_k p(z - y_k)) + \log \left( 1 + \frac{\sum_{i \neq k} p_i p(z - y_i)}{p_k p(z - y_k)} \right) \right]
\]

\[
\geq p_k p(z - y_k) \log(p_k p(z - y_k)).
\]

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
\[
    h(X + Y) = -\int_{\mathbb{R}^d} q(z) \log q(z) \, dz
\]
\[
    = -\sum_k \int_{\mathbb{R}^d} p_k p(z - y_k) \log \sum_i p_i p(z - y_i) \, dz
\]
\[
    \leq -\sum_k \int_{\mathbb{R}^d} p_k p(z - y_k) \log(p_k p(z - y_k)) \, dz
\]
\[
    = -\sum_k p_k \left( \int_{\mathbb{R}^d} p(z - y_k) \log p_k \, dz + \int_{\mathbb{R}^d} p(z - y_k) \log p(z - y_k) \, dz \right)
\]
\[
    = h(X) + H(Y).
\]

\[\square\]

Let \((Y_n)_{n \geq 1}\) be a sequence of i.i.d. Bernoulli random variables. A simple analysis of the binomial probabilities shows that as \(n \to \infty\),
\[
    H(Y_1 + \cdots + Y_n) = \frac{1}{2} \log(2\pi en) + o(1).
\]

We only need an upper bound in this asymptotic relation, which in turn can be stated in the class of all lattice distributions. The following lemma is standard and has been used in several applications (see \[11\]):

**Lemma 5.2.** For any integer valued random variable \(Y\) with finite second moment,
\[
    H(Y) \leq \frac{1}{2} \log \left( 2\pi e \left( \text{Var}(Y) + \frac{1}{12} \right) \right).
\]

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 \(\tilde{Y}\) with density \(q\) defined to be
\[
    q(x) = p_k \quad \text{if} \quad x \in \left( k - \frac{1}{2}, k + \frac{1}{2} \right).
\]
In other words,
\[
    q(x) = \sum_k p_k 1_{\left( k - \frac{1}{2}, k + \frac{1}{2} \right)}(x), \quad x \in \mathbb{R}.
\]
Note that
\[
    \mathbb{E}\tilde{Y} = \sum_k p_k \int_{k - \frac{1}{2}}^{k + \frac{1}{2}} x \, dx = \sum_k \frac{p_k}{2} \left( \left( k + \frac{1}{2} \right)^2 - \left( k - \frac{1}{2} \right)^2 \right) = \sum_k k p_k = \mathbb{E}Y
\]
and similarly
\[
    \mathbb{E}\tilde{Y}^2 = \sum_k p_k \int_{k - \frac{1}{2}}^{k + \frac{1}{2}} x^2 \, dx = \mathbb{E}Y^2 + \frac{1}{12}.
\]
Hence \( \text{Var}(\widetilde{Y}) = \text{Var}(Y) + \frac{1}{12} \). Also,

\[
\begin{align*}
\mathbb{H}(\widetilde{Y}) &= -\int_{-\infty}^{\infty} \sum_{k} p_k 1_{(k-\frac{1}{2}, k+\frac{1}{2})}(x) \log \sum_{j} p_j 1_{(j-\frac{1}{2}, j+\frac{1}{2})}(x) \, dx \\
&= -\sum_{k} p_k \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \log p_k \, dx = H(Y).
\end{align*}
\]

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

\[
H(Y) = \mathbb{H}(\widetilde{Y}) \leq \frac{1}{2} \log \left( 2\pi e \text{Var}(\widetilde{Y}) \right) = \frac{1}{2} \log \left( 2\pi e \left( \text{Var}(Y) + \frac{1}{12} \right) \right).
\]

Proof of Theorem 1.3. By (5.1)-(5.2) applied to \( Y = Y_1 + \cdots + Y_n \), we have

\[
H(Y) \leq \frac{1}{2} \log \left( 2\pi en + \frac{1}{12} \right) = \frac{1}{2} \log(2\pi en) + O(1/n).
\]

Let \( (Y_{i,j})_{1 \leq i \leq d, 1 \leq j \leq n} \) be a collection of independent Bernoulli random variables, independent of \( X \). The random vectors \( X_j = (Y_{1,j}, \ldots, Y_{d,j}) \) are independent and uniformly distributed on the discrete cube \( \{-1, 1\}^d \), \( j = 1, \ldots, n \). Hence, applying Lemma 5.1 together with (5.2), we get

\[
\begin{align*}
\mathbb{H}\left( \frac{X + X_1 + \cdots + X_n}{\sqrt{n}} \right) &= \mathbb{H}(X + X_1 + \cdots + X_n) - \frac{d}{2} \log n \\
&\leq \mathbb{H}(X) + \frac{d}{2} \log(2\pi e n) - \frac{d}{2} \log n \\
&\leq \mathbb{H}(X) + \frac{d}{2} \log(2\pi e) + o(1), \quad \text{as } n \to \infty.
\end{align*}
\]

We conclude that

\[
\limsup_{n \to \infty} \mathbb{H}\left( \frac{X + Y_1 + \cdots + Y_n}{\sqrt{n}} \right) \leq \mathbb{H}(X) + \frac{d}{2} \log(2\pi e) = \mathbb{H}(X) + \mathbb{H}(Z),
\]

implying that

\[
\mathbb{H}(Z) = \lim_{n \to \infty} \mathbb{H}(Z_n) \leq \mathbb{H}(X) + \mathbb{H}(Z).
\]

References


