LOCAL LIMIT THEOREMS FOR DENSITIES IN ORLICZ SPACES

SERGEY G. BOBKOV

Abstract. Necessary and sufficient conditions for the validity of the central limit theorem for densities are considered with respect to the norms in Orlicz spaces. The obtained characterization unites several results due to Gnedenko and Kolmogorov (uniform local limit theorem), Prokhorov (convergence in total variation) and Barron (entropic CLT).

1. Introduction

Let $(X_n)_{n\geq 1}$ be independent copies of a random vector X in $\mathbb{R}^d$ with mean zero and an identity covariance matrix. By the central limit theorem (CLT), the normalized sums

$$Z_n = \frac{1}{\sqrt{n}}\,(X_1 + \cdots + X_n) \tag{1.1}$$

are weakly convergent to the standard Gaussian measure γ on $\mathbb{R}^d$ with density

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

Suppose that Z_n has an absolutely continuous distribution for some $n = n_0$, so that all $(Z_n)_{n\geq n_0}$ have densities p_n. The weak convergence then means that

$$\int_{\mathbb{R}^d} (p_n(x) - \varphi(x))\, u(x)\, \mathrm{d}x \to 0 \quad \text{as } n \to \infty$$

for any bounded continuous function u on $\mathbb{R}^d$. Local limit theorems deal with convergence of p_n to φ in a stronger sense. In particular, employing an approach by Prokhorov, it was proved by Ranga Rao and Varadarajan that

$$p_n(x) \to \varphi(x) \quad \text{as} \;\; n \to \infty \tag{1.2}$$

for almost all points $x \in \mathbb{R}^d$ (in the sense of the Lebesgue measure, cf. [12]). What is also natural, one may consider the convergence of densities in Orlicz spaces.

Given a Young function $\Psi : \mathbb{R} \to [0,\infty)$, that is, an even, convex function such that $\Psi(0) = 0$, $\Psi(t) > 0$ for $t > 0$, the Orlicz norm of a measurable function u on $\mathbb{R}^d$ is defined as

$$\|u\| = \|u\|_\Psi = \inf\Big\{\lambda > 0 : \int_{\mathbb{R}^d} \Psi(u(x)/\lambda)\, \mathrm{d}x \leq 1\Big\}.$$

2010 *Mathematics Subject Classification.* Primary 60E, 60F.

Key words and phrases. CLT for densities, Orlicz spaces, Entropy, Convolution.

School of Mathematics, University of Minnesota, Minneapolis, MN 55455 USA.

Research was partially supported by NSF grant DMS-1612961.

The associated Orlicz space $L^\Psi = L^\Psi(\mathbb{R}, \mathrm{d}x)$ contains all u with $\|u\|_\Psi < \infty$. For example, the choice $\Psi(t) = |t|^\alpha$ ($\alpha \geq 1$) leads to the usual L^α-norm $\|u\|_\alpha$. Let us include in this family the L^∞-norm $\|u\|_\infty = \mathrm{ess\,sup}_x\, |u(x)|$ as a (limit) Orlicz norm. Being specialized to probability densities, the convergence in any Orlicz norm occupies an intermediate position between the convergence in L^∞-norm (which is the strongest one) and the convergence in L^1-norm (the weakest one). Here, we prove the following characterization.

Theorem 1.1. *Suppose that Z_n have densities p_n for large enough n. For a given Orlicz norm, we have $\|p_n - \varphi\| \to 0$ as $n \to \infty$, if and only if $\|p_n\| < \infty$ for some $n = n_0$.*

For a large class of Orlicz norms, this statement may be given in a slightly different form. Recall that the Young function Ψ is said to satisfy the Δ_2-condition, if $\Psi(2t) \leq c\Psi(t)$ with some constant $c > 0$ independent of $t \geq 0$.

Corollary 1.2. *Suppose that Z_n have densities p_n for large enough n, and let the Young function Ψ satisfy the Δ_2-condition. Then*

$$\int \Psi\big(p_n(x) - \varphi(x)\big)\, \mathrm{d}x \to 0 \quad \text{as} \quad n \to \infty,$$

if and only if $\int \Psi(p_n(x))\, \mathrm{d}x < \infty$ for some $n = n_0$.

In the case of the L^∞-norm, Theorem 1.1 is essentially due to Gnedenko and Kolmogorov. Originally, sufficient conditions for the uniform local limit theorem

$$\sup_x |p_n(x) - \varphi(x)| \to 0 \qquad (n \to \infty) \tag{1.3}$$

were stated in [7] for the dimension $d = 1$ in the following way. It was assumed that, for some n, p_n belongs to L^α, $1 < \alpha \leq 2$, and satisfies an integrable Lipschitz condition (which was later removed in [8]). Here, the first assumption may formally be weakened to $\|p_n\|_\infty < \infty$ (for some n), which is not only sufficient, but is also necessary for (1.3) to hold, cf. Petrov [10]. But, once p_n is bounded, we have $\|p_n\|_\alpha < \infty$ for all $\alpha > 1$. Hence, Gnedenko-Kolmogorov's condition is necessary as well and may be formulated with arbitrary $\alpha > 1$. Bhattacharya and Ranga Rao [3] gave another description in terms of the characteristic function $f(t) = \mathbb{E}\, e^{i\langle t, X\rangle}$. Namely, (1.3) is equivalent to the so-called smoothing condition

$$\int_{\mathbb{R}^d} |f(t)|^\nu\, dt < \infty \quad \text{for some} \quad \nu \geq 1. \tag{1.4}$$

Let us also add that the property (1.4) implies not only boundedness, but also continuity of the densities p_n for sufficiently large n.

In the case of the L^α-norm ($\alpha > 1$) in Theorem 1.1, the requirement that $\|p_n\|_\alpha < \infty$ for some n returns us to the setting of Gnedenko-Kolmogorov's theorem and is therefore reduced to the smoothing condition. That is, the assertion $\|p_n - \varphi\|_\alpha \to 0$ does not depend on α in the range $1 < \alpha \leq \infty$ and is equivalent to (1.3)-(1.4).

In the case of the L^1-norm, Theorem 1.1 is due to Prokhorov [11]. It may be stated in terms of the total variation distance between the distribution μ_n of Z_n and the Gaussian measure γ as the assertion

$$\|\mu_n - \gamma\|_{\mathrm{TV}} = \|p_n - \varphi\|_1 \to 0 \qquad (n \to \infty).$$

Thus, it holds true without any condition as long as the densities p_n exist. This variant of the local limit theorem may also be viewed as a direct consequence from (1.2): By the well-known Scheffe's lemma, the pointwise convergence of probability densities (holding almost everywhere) implies the convergence in L^1-norm.

These particular cases show that the property $\|p_n - \varphi\| \to 0$ can involve a larger class of underlying probability distributions in comparison with the class described in (1.4), only when the norm $\|\cdot\|$ is weaker than all L^α-norms ($\alpha > 1$). In order to turn to an interesting example, let us remind the notion of the Kullback-Leibler distance

$$D(\mu||\nu) = D(p||q) = \int_\Omega p\, \log(p/q)\, \mathrm{d}\lambda,$$

also called the relative entropy or an information divergence. (Note that it is not a metric in the usual sense.) This quantity is well-defined in the setting of an abstract measurable space Ω for arbitrary probability measures ν and μ with densities p and q over a dominating σ-finite measure λ, assuming that μ is absolutely continuous with respect to ν (the definition does not depend on the choice of λ). In general, $0 \le D(\mu||\nu) \le \infty$, and $D(\mu||\nu) = 0$ if and only if $\mu = \nu$. This separation property may be quantified by means of the Pinsker-type inequality

$$D(\mu||\nu) \ge \frac{1}{2}\, \|\mu - \nu\|_{\mathrm{TV}}^2 = \frac{1}{2}\, \Big(\int_\Omega |p - q|\, \mathrm{d}\lambda\Big)^2,$$

cf. e.g. [6] and references therein. Returning to the normalized sums Z_n as in (1.1) with densities p_n on $\Omega = \mathbb{R}^d$ with respect to the Lebesgue measure λ, the Kullback-Leibler distance

$$D(\mu_n||\gamma) = D(p_n||\varphi) = \int_{\mathbb{R}^d} p_n \log(p_n/\varphi)\, \mathrm{d}x$$

thus dominates the L^1-distance, and we have the Pinsker inequality $D(p_n||\varphi) \ge \frac{1}{2}\, \|p_n - \varphi\|_1^2$. A corresponding description of the entropic CLT was obtained by Barron [2] (originally, in dimension one), and we give it below; see also [9], [4].

Theorem 1.3. *Suppose that Z_n have densities p_n for large enough n. Then $D(p_n||\varphi) \to 0$ as $n \to \infty$, if and only if $D(p_n||\varphi) < \infty$ for some n.*

Here, the last property may be stated as the finiteness of the differential entropy

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

(which does not exceed $h(\varphi)$ due to the second moment assumption, but in general may take the value $-\infty$). This is also equivalent to

$$\int_{\mathbb{R}^d} p_n(x)\, \log(1 + p_n(x))\, \mathrm{d}x < \infty. \tag{1.5}$$

As a next step, we show that Barron's theorem may be included in Theorem 1.1 as a particular case, by applying the next characterization of the convergence in D to the standard normal law. Introduce the Young function

$$\psi(t) = |t|\, \log(1 + |t|), \quad t \in \mathbb{R}.$$

Clearly, it satisfies the Δ_2-condition.

Theorem 1.4. *Given a sequence $(p_n)_{n \geq 1}$ of probability densities on $\mathbb{R}^d$, the property $D(p_n||\varphi) \to 0$ as $n \to \infty$ is equivalent to the following two conditions:*

a) $\int_{\mathbb{R}^d} |x|^2 \, p_n(x) \, \mathrm{d}x \to d$; *b)* $\|p_n - \varphi\|_\psi \to 0$ as $n \to \infty$.

Here, the last condition may be replaced with

$$\int_{\mathbb{R}^d} |p_n - \varphi| \, \log(1 + |p_n - \varphi|) \, \mathrm{d}x \, \to \, 0. \tag{1.6}$$

In the setting of Theorem 1.1, the integral in *a*) is just $\mathbb{E}\,|Z_n|^2 = d$ due to the basic assumption on the covariance matrix of the random vector X, so, the condition *a*) is fulfilled. Thus, the convergence in D implies the convergence in the Orlicz norm $\| \cdot \|_\psi$, which may also be formulated as (1.6). In turn, *b*) yields $\|p_n\|_\psi \leq \lambda$ for all sufficiently large n with some constant λ, implying that (1.5) is fulfilled. Hence $D(p_n||\varphi) < \infty$ as well, and we see that Theorem 1.3 is a consequence of Theorem 1.1.

The paper is organized as follows. For simplicity (mostly of notations), in the proof of Theorem 1.1 we will assume that the random vector X has an absolutely continuous distribution, so that $n_0 = 1$ (minor modifications should be done in order to involve the general case $n_0 \geq 1$, cf. e.g. [4]). As a preliminary step, first we recall a general scheme of decomposition of convolutions into the major and small parts (Section 2), and then a uniform local limit theorem is proved for the major part (Section 3). The material of these two sections is not new, and we include it here to make the proof to be self-contained. Final steps in the proof of Theorem 1.1 are made in Section 4. Before turning to the proof of Theorem 1.4, in Sections 5-6 we consider preliminary general bounds on the distance $D(p||\varphi)$, which relate them to the Orlicz norm, as well as to the mean and the covariance matrix associated to a given density p. In the last Section 7, we discuss topological properties of the convergence in relative entropy and prove Theorem 1.4.

2. **Decomposition of densities**

Assume that a random vector X has an absolutely continuous distribution with density w. Here, we describe a general scheme of decomposition of the convolution powers $w_n = w^{*n}$ into the two parts, one of which is a bounded density approximating w_n in a sufficiently sharp way, while the other one is small and can be controlled in terms of the Orlicz norm of w. This approach to local limit theorems goes back to the work by Prokhorov [11]. Let us write $M(q) = \|q\|_\infty$.

Given $0 < \delta_1 \leq \frac{1}{4}$, one may decompose $\mathbb{R}^d$ into two measurable sets of the form $\Omega_0 \subset \{w(x) \leq M\}$ and $\Omega_1 \subset \{w(x) \geq M\}$ such that

$$\int_{\Omega_0} w(x) \, \mathrm{d}x \, = \, \delta_0 \equiv 1 - \delta_1, \qquad \int_{\Omega_1} w(x) \, \mathrm{d}x \, = \, \delta_1.$$

As a result, we obtain the decomposition

$$w(x) = \delta_0 w_0(x) + \delta_1 w_1(x),$$

in which w_0 and w_1 are defined as the normalized restrictions of w to the sets Ω_0 and Ω_1, respectively. By the construction, $M(w_0) \le M/\delta_0 \le 2M$. Hence, putting

$$q_l = w_0^{*l} * w_1^{*(n-l)}, \qquad l = 0, 1, \dots, n,$$

we get the representation

$$w^{*n} = \sum_{l=0}^{n} C_n^l \, \delta_0^l \, \delta_1^{n-l} \, q_l,$$

where $C_n^l = \frac{n!}{l!(n-l)!}$ are usual binomial coefficients. Assuming that $n \ge 2$ and removing from this representation the first two terms, define

$$\tilde{w}_n \ = \ \frac{1}{1-\kappa_n} \sum_{l=2}^{n} C_n^l \, \delta_0^l \, \delta_1^{n-l} \, q_l, \quad \kappa_n = \delta_1^n + n\delta_0\delta_1^{n-1}, \tag{2.1}$$

where the normalizing constant is chosen to meet the requirement $\int \tilde{w}_n(x)\,\mathrm{d}x = 1$.

Definition 2.1. Let us call $\tilde{w}_n$ a canonical approximation for w_n with weight δ_0.

Lemma 2.2. *For $n \ge 2$, the probability density $\tilde{w}_n$ is bounded, continuous, and satisfies*

$$\int_{\mathbb{R}^d} |\tilde{w}_n(x) - w^{*n}(x)|\,\mathrm{d}x \ < \ \frac{1}{2^{n-1}}. \tag{2.2}$$

Moreover, the Fourier transform h_n of $\tilde{w}_n$ is an integrable function, satisfying

$$\int_{|t|\ge r} |h_n(t)|\,\mathrm{d}t \ < \ Ac^n \tag{2.3}$$

for any $r > 0$ with some constants $A > 0$ and $0 < c < 1$ which do not dependent on n (here, the constant c may depend on r).

Proof. By the definition (2.1),

$$\kappa_n = \delta_1^{n-1}\,(1 + n\delta_0\delta_1) \le 4^{-(n-1)}\left(1 + \frac{n}{4}\right) < 2^{-n}.$$

Therefore,

$$\int_{\mathbb{R}^d} |\tilde{w}_n(x) - w^{*n}(x)|\,\mathrm{d}x \ \le \ 2\kappa_n \ < \ 2^{-(n-1)},$$

proving the inequality (2.2).

Now, let

$$\hat{w}_j(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\, w_j(x)\,\mathrm{d}x, \qquad t \in \mathbb{R}^d \ \ (j = 0, 1)$$

denote the Fourier transforms of the densities w_0 and w_1. By the Riemann-Lebesgue lemma, $\hat{w}_0(t) \to 0$ as $|t| \to \infty$. In addition, $|\hat{w}_0(t)| < 1$ for all $t \neq 0$ (since otherwise, the distribution with density w_0 must be discrete, cf. [3]). Hence, for any fixed $r > 0$,

$$\beta = \sup_{|t|\ge r} |\hat{w}_0(t)| < 1.$$

Applying the Plancherel theorem, we get, for any integer $l \geq 2$,

$$\begin{aligned}\int_{|t|\geq r} |\hat{w}_0(t)|^l \, \mathrm{d}t &\leq \beta^{l-2} \int_{\mathbb{R}^d} |\hat{w}_0(t)|^2 \, \mathrm{d}t \\ &= (2\pi)^d \beta^{l-2} \int_{\mathbb{R}^d} w_0(x)^2 \, \mathrm{d}x \\ &\leq (2\pi)^d \beta^{l-2} \, M(w_0) \, \leq \, 2\,(2\pi)^d \beta^{l-2} \, M.\end{aligned}$$

Hence, the Fourier transform $\hat{q}_l$ of the density $q_l = w_0^{*l} * w_1^{*(n-l)}$ admits a similar bound

$$\int_{|t|\geq r} |\hat{q}_l(t)| \, \mathrm{d}t \leq \int_{|t|\geq r} |\hat{w}_0(t)|^l \, \mathrm{d}t \leq Ac^l, \qquad l = 2, \dots, n,$$

with some constants $A > 0$ and $0 < c < 1$ which do not depend on l. Since, by (2.1),

$$h_n(t) \, = \, \frac{1}{1-\kappa_n} \sum_{l=2}^n C_n^l \, \delta_0^l \, \delta_1^{n-l} \, \hat{q}_l(t),$$

we conclude that

$$\begin{aligned}\int_{|t|\geq r} |\hat{h}_n(t)| \, \mathrm{d}t &\leq \frac{A}{1-\kappa_n} \sum_{l=2}^n C_n^l \, (c\delta_0)^l \, \delta_1^{n-l} \\ &< \frac{A}{1-\kappa_n} \, (1 - (1-c)\,\delta_0)^n.\end{aligned}$$

It remains to recall that $\kappa < 1/4$, and then we arrive at (2.3). The latter inequality also guarantees that $\hat{w}_n$ are bounded and continuous, according to the inverse Fourier formula. $\square$

3. CLT for approximating densities

Let $X_1, X_2, \dots$ be independent copies of a random vector X in $\mathbb{R}^d$ with mean zero, an identity covariance matrix, and with density w. Denote by p_n the densities of the normalized sums

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

which are thus described by

$$p_n(x) = n^{d/2} \, w^{*n}(n^{1/2}x), \qquad x \in \mathbb{R}^d. \tag{3.1}$$

As we know from Definition 2.1 and Lemma 2.2, w^{*n} are well approximated by the functions $\tilde{w}_n$ which can be constructed and used with an arbitrary parameter $\delta_1 \in (0, 1/4]$. Hence, as a canonical approximation for p_n, one may use

$$\tilde{p}_n(x) = n^{d/2} \, \tilde{w}_n(n^{1/2}x). \tag{3.2}$$

Let us reformulate Lemma 2.2 in terms of the rescaled densities.

Lemma 3.1. *For $n \geq 2$, the probability density $\tilde{p}_n$ is bounded, continuous, and satisfies*

$$\int_{\mathbb{R}^d} |\tilde{p}_n(x) - p_n(x)| \, \mathrm{d}x \, < \, \frac{1}{2^{n-1}}. \tag{3.3}$$

Moreover, the Fourier transform $\tilde{f}_n$ of $\tilde{p}_n$ is an integrable function, satisfying

$$\int_{|t|\geq r\sqrt{n}} |\tilde{f}_n(t)|\,\mathrm{d}t \,<\, Ac^n \tag{3.4}$$

for any $r > 0$ with some constants $A > 0$ and $0 < c < 1$ which do not dependent on n (the constant c may depend on r).

We can now prove the uniform local limit theorem for the approximating densities $\tilde{p}_n$.

Lemma 3.2. *As $n \to \infty$, we have*

$$\sup_x |\tilde{p}_n(x) - \varphi(x)| \to 0. \tag{3.5}$$

Proof. Using the inversion formula, for all $x \in \mathbb{R}^d$, we have the representation

$$\tilde{p}_n(x) - \varphi(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\,(\tilde{f}_n(t) - g(t))\,\mathrm{d}t,$$

where $g(t) = e^{-|t|^2/2}$ is the Fourier transform of the standard normal density φ. Applying (3.4) with a certain number $r > 0$ which will be specified later on, we therefore obtain that

$$\|\tilde{p}_n - \varphi\|_\infty \leq Ac^n + (2\pi)^{-d} \int_{|t|\leq r\sqrt{n}} |\tilde{f}_n(t) - g(t)|\,\mathrm{d}t \tag{3.6}$$

with some constants $A > 0$ and $0 < c < 1$ which do not dependent on n.

Now, the distribution μ_n of Z_n has characteristic function

$$f_n(t) = \mathbb{E}\,e^{it\langle t,Z_n\rangle} = f\Big(\frac{t}{\sqrt{n}}\Big)^n, \qquad t \in \mathbb{R}^d,$$

where f is the characteristic function of X. Applying the property (3.3), we get

$$\sup_t |\tilde{f}_n(t) - f_n(t)| \,<\, \frac{1}{2^{n-1}}.$$

This means that one may replace $\tilde{f}_n$ with f_n in (3.6) by increasing the constants, so that

$$\|\tilde{p}_n - \varphi\|_\infty \leq Ac^n + (2\pi)^{-d} \int_{|t|\leq r\sqrt{n}} |f_n(t) - g(t)|\,\mathrm{d}t \tag{3.7}$$

with some $A > 0$ and $c \in (0,1)$ independent of n.

Here, the region of integration may further be decreased using the property that $f_n(t)$ is small for large $|t|$. Indeed, since the random vector X has mean zero and an identity covariance matrix, the characteristic function f admits a Taylor expansion up to the quadratic term in the form of Peano as $f(t) = 1 - \frac{1}{2}\,|t|^2 + o(|t|^2)$ as $t \to 0$. Hence, there exists $0 < r_0 < 1$ such that

$$|f(t)| \leq 1 - \frac{1}{4}\,|t|^2$$

in the ball $|t| \leq r_0$. This gives

$$|f_n(t)| \leq \Big(1 - \frac{1}{4n}\,|t|^2\Big)^n \leq e^{-|t|^2/4}, \qquad |t| \leq r_0\sqrt{n}.$$

It follows that, for any $T > 0$,

$$\int_{T\le|t|\le r_0\sqrt{n}} |f_n(t)|\,\mathrm{d}t \ \le\ \int_{T\le|t|\le r_0\sqrt{n}} e^{-|t|^2/4}\,\mathrm{d}t \ \le\ d\omega_d \int_T^\infty z^{d-1}\,e^{-z^2/4}\,\mathrm{d}z \ \le\ B\,e^{-T^2/8}$$

with some constant B which does not depend on n and T (where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$). Using a similar bound

$$\int_{|t|\ge T} |g(t)|\,\mathrm{d}t \ \le\ B\,e^{-T^2/8}$$

and putting $r = r_0$, from (3.7) we get

$$\|\tilde{p}_n - \varphi\|_\infty \le Ac^n + 2B\,e^{-T^2/8} + (2\pi)^{-d}\int_{|t|\le T} |f_n(t) - g(t)|\,\mathrm{d}t. \tag{3.8}$$

Finally, by the weak CLT, $f_n(t) \to g(t)$ for any $t \in \mathbb{R}^d$, and moreover, this convergence is uniform on every ball $|t| \le T$. One can therefore choose a sequence $T_n \uparrow \infty$ such that

$$\int_{|t|\le T_n} |f_n(t) - g(t)|\,\mathrm{d}t \to 0 \qquad (n \to \infty).$$

It remains to apply (3.8) with $T = T_n$, which leads to (3.5). □

4. Proof of Theorem 1.1

The proof of Theorem 1.1 is only needed in one direction. As was mentioned, we assume that the random vector X has an absolutely continuous distribution with density, say w. If the Orlicz norm $\|\cdot\|$ is generated by the Young function Ψ, without loss of generality we may also assume that $\Psi(1) = 1$. With this convention, let us start with general remarks.

Lemma 4.1. *For any measurable function u on $\mathbb{R}^d$,*

$$\|u\| \le \max\big\{\|u\|_1, \|u\|_\infty\big\}. \tag{4.1}$$

Proof. If $\|u\| = \|u\|_\infty$, (4.1) is immediate. Otherwise, let the norm be generated by the Young function Ψ such that $\Psi(1) = 1$. In view of the homogeneity of the inequality (4.1), we may assume that its right-hand hand side does not exceed 1, so that $\|u\|_1 \le 1$ and $\|u\|_\infty \le 1$. In this case, by the convexity of Ψ, we have $\Psi(t) \le |t|$ whenever $-1 \le t \le 1$. Hence

$$\int_{\mathbb{R}^d} \Psi(u(x))\,\mathrm{d}x \le \int_{\mathbb{R}^d} |u(x)|\,\mathrm{d}x \le 1,$$

which means that $\|u\|_\Psi \le 1$. □

The next elementary relation immediately follows from the definition of the Orlicz norm.

Lemma 4.2. *For any measurable function u on $\mathbb{R}^d$ and $\lambda \geq 1$, we have*

$$\|u(\lambda x)\| \leq \|u(x)\|.$$

Lemma 4.3. *For all non-negative measurable functions $u_1, \dots, u_N$ on $\mathbb{R}^d$ $(N \geq 2)$,*

$$\|u_1 * u_2 * \dots * u_N\| \,\leq\, \|u_1\|\, \|u_2\|_1\, \dots\, \|u_N\|_1. \tag{4.2}$$

Proof. One may argue by induction on N, and then it is sufficient to consider the case $N = 2$. If $\|\cdot\| = \|\cdot\|_\infty$, the inequality (4.2) is obvious. If $\|\cdot\| = \|\cdot\|_\Psi$, one may assume, by the homogeneity, that $\|u_1\|_\Psi = 1$ and $\|u_2\|_1 = 1$. By Jensen's inequality,

$$\Psi\big((u_1 * u_2)(x)\big) \,=\, \Psi\Big(\int u_1(x-y)\, u_2(y)\, \mathrm{d}y\Big) \,\leq\, \int \Psi(u_1(x-y))\, u_2(y)\, \mathrm{d}y,$$

so,

$$\int \Psi\big((u_1 * u_2)(x)\big)\, \mathrm{d}x \,\leq\, \iint \Psi(u_1(x-y))\, u_2(y)\, \mathrm{d}y\, \mathrm{d}x = 1.$$

□

Proof of Theorem 1.1. Let $n \geq 2$. We use the approximating functions $\tilde{p}_n$ for the densities p_n of Z_n, described in (3.1)-(3.2). By Lemma 3.2, $\|\tilde{p}_n - \varphi\|_\infty \to 0$ as $n \to \infty$, which implies $\|\tilde{p}_n - \varphi\|_1 \to 0$, by Scheffe's lemma (since all $\tilde{p}_n$ are probability densities). Applying Lemma 4.1, we may conclude that $\|\tilde{p}_n - \varphi\| \to 0$ as well.

In view of the triangle inequality in the Orlicz space, it remains to show that $\|\tilde{p}_n - p_n\| \to 0$. From (3.1)-(3.2) it follows that

$$\|\tilde{p}_n - p_n\| \,\leq\, n^{d/2}\, \big\|\tilde{w}_n(n^{1/2}x) - w^{*n}(n^{1/2}x)\big\|.$$

To simplify, one may apply Lemma 4.2, which gives

$$\|\tilde{p}_n - p_n\| \,\leq\, n^{d/2}\, \|\tilde{w}_n - w^{*n}\|. \tag{4.3}$$

Now, we need to return to the definition (2.1), which by the triangle inequality, yields

$$(1 - \kappa_n)\, \|\tilde{w}_n\| \,\leq\, \sum_{l=2}^{n} C_n^l\, \delta_0^l\, \delta_1^{n-l}\, \|q_l\|. \tag{4.4}$$

Let us recall that $q_l = w_0^{*l} * w_1^{*(n-l)}$, so that, by Lemma 4.3, $\|q_l\| \leq \|w_0\|$. Hence, from (4.4),

$$(1 - \kappa_n)\, \|\tilde{w}_n\| \,\leq\, \|w_0\|$$

and thus

$$\|\tilde{w}_n - (1 - \kappa_n)\, \tilde{w}_n\| \,=\, \kappa_n\, \|\tilde{w}_n\| \,\leq\, \frac{\kappa_n}{1 - \kappa}\, \|w_0\| \,<\, \frac{1}{2^{n-1}}\, \|w_0\|,$$

where we used $\kappa_n < 2^{-n}$ on the last step. Applying the latter estimate in (4.3), we get

$$\|\tilde{p}_n - p_n\| \,\leq\, n^{d/2}\, \|(1 - \kappa_n)\, \tilde{w}_n - w^{*n}\| + \frac{n^{d/2}}{2^{n-1}}\, \|w_0\|. \tag{4.5}$$

But, according to (2.1),

$$\begin{aligned} w^{*n} - (1-\kappa_n)\,\tilde{w}_n &= \delta_1^n q_0 + n\,\delta_0\delta_1^{n-1} q_1 \\ &= \delta_1^n w_1^{*n} + n\,\delta_0\delta_1^{n-1}\, w_0 * w_1^{*\,(n-1)}. \end{aligned}$$

Again by Lemma 4.3, the norm of this expression does not exceed

$$\delta_1^n\,\|w_1\| + n\,\delta_0\delta_1^{n-1}\,\|w_1\| \,<\, 2^{-n}\,\|w_1\|.$$

Inserting this in (4.5), we arrive at

$$\|\tilde{p}_n - p_n\| \,\leq\, n^{d/2}\,2^{-n}\,\big(\|w_1\| + 2\,\|w_0\|\big).$$

The last expression tends to zero exponentially fast as $n \to 0$, once we see that w_0 and w_1 have finite norms. But this follows from the decomposition $w = \delta_0 w_0 + \delta_1 w_1$ and the main hypothesis that $\|w\| < \infty$. □

Remark 4.4. Let us comment on the case where the Orlicz norm $\|\cdot\| = \|\cdot\|_\Psi$ corresponds to the Young function satisfying the Δ_2-condition $\Psi(2t) \leq c\Psi(t)$. This property may also be written as

$$\Psi(\lambda t) \leq c_\lambda \Psi(t), \qquad t \in \mathbb{R}, \tag{4.6}$$

where $\lambda > 1$ is an arbitrary fixed number, and the constant c_λ depends on λ only. It ensures that, for any measurable function u on $\mathbb{R}^d$, $\|u\|_\Psi < \infty$, if and only if $\int \Psi(u(x))\,\mathrm{d}x < \infty$. Indeed, by the convexity, $\Psi(\alpha t) \leq \alpha\Psi(t)$ for all $\alpha \in [0,1]$. Hence, to argue in one direction, if $\lambda = \int \Psi(u(x))\,\mathrm{d}x$ is finite, $\lambda > 1$, then

$$\int \Psi(u(x)/\lambda)\,\mathrm{d}x \leq \frac{1}{\lambda}\int \Psi(u(x))\,\mathrm{d}x = 1,$$

which means $\|u\|_\Psi \leq \lambda$. In the case $\lambda \leq 1$, necessarily $\|u\|_\Psi \leq 1$, by the definition of the Orlicz norm. Thus, $\|u\|_\Psi \leq \max(\lambda, 1) < \infty$. In the other direction, if $\lambda = \|u\|_\Psi < \infty$, then, by (4.6),

$$\int \Psi(u(x))\,\mathrm{d}x \,=\, \int \Psi(\lambda u(x)/\lambda)\,\mathrm{d}x \,\leq\, c_\lambda \int \Psi(u(x)/\lambda)\,\mathrm{d}x \,=\, c_\lambda \,<\, \infty.$$

By similar arguments, given a sequence of measurable functions $(u_n)_{n\geq 1}$ on $\mathbb{R}^d$, we have $\|u_n\|_\Psi \to 0$, if and only if $\int \Psi(u_n(x))\,\mathrm{d}x \to 0$ as $n \to 0$. This explains why Corollary 1.2 follows from Theorem 1.1.

Remark 4.5. The Δ_2-condition implies in particular that $\Psi(t) = O(t^\alpha)$ as $t \to \infty$ with some $\alpha \geq 1$. A necessary and sufficient condition for the property (4.6) to hold is that

$$\sup_{t\geq t_0} \frac{t\Psi'(t+)}{\Psi(t)} \,\leq\, C$$

for some $t_0 > 0$ and $C < \infty$, where $\Psi'(t+)$ denotes the right derivative of the function Ψ at the point t. See e.g. the notes [1].

5. Two-sided estimates on relative entropy

Before turning to the proof of Theorem 1.4, let us first derive one general two-sided bound on the relative entropy

$$D(p||q) = \int_\Omega p\,\log(p/q)\,\mathrm{d}\lambda, \tag{5.1}$$

which might be of independent interest. Here, p and q are probability densities on the abstract measure space (Ω,λ). We will assume that the probability measure $\mathrm{d}\mu = p\,\mathrm{d}\lambda$ is absolutely continuous with respect to $\mathrm{d}\nu = q\,\mathrm{d}\lambda$, that is, $q(x)=0 \Rightarrow p(x)=0$ for λ-almost all $x\in\Omega$.

Theorem 5.1. *With some absolute constants $c_1 > c_0 > 0$, we have*

$$\begin{aligned}\int_\Omega |p-q|\,\log\Big(1+c_0\,\frac{|p-q|}{q}\Big)\,\mathrm{d}\lambda \ &\le \ D(p||q)\\ &\le \ \int_\Omega |p-q|\,\log\Big(1+c_1\,\frac{|p-q|}{q}\Big)\,\mathrm{d}\lambda. \end{aligned}\tag{5.2}$$

The optimal values are $c_0 = 1/e$ and $c_1 = e-1$.

The point of (5.2) is that, in contrast with the integrand in (5.1), the integrands in (5.2) are non-negative. The integration in (5.2) may be restricted to the set $\{x\in\Omega : q(x)>0\}$.

Proof. Consider the function

$$H(u) = (1+u)\log(1+u) - u, \qquad u \ge -1,$$

so that

$$D(p||q) \ = \ \int_\Omega \frac{p}{q}\,\log\frac{p}{q}\,\mathrm{d}\nu \ = \ \int_\Omega H\Big(\frac{p-q}{q}\Big)\,\mathrm{d}\nu.$$

Hence, (5.2) would follow from the two-sided bound

$$|u|\,\log(1+c_0|u|) \ \le \ H(u) \ \le \ |u|\,\log(1+c_1|u|), \tag{5.3}$$

where we need to show that the same values $c_0 = 1/e$ and $c_1 = e-1$ are optimal.

Case 1: First, consider the region $u\ge 0$. Given a parameter $c>0$, the function

$$G(u) = H(u) - u\,\log(1+cu)$$

satisfies $G(u) = G'(0) = 0$, and

$$G'(u) = \log(1+u) - \log(1+cu) - 1 + \frac{1}{1+cu}, \qquad G'(\infty) = \log\frac{1}{c} - 1.$$

As easy to see, $G(\infty)=\infty$ if $c\le 1/e$, and $G(\infty) = -\infty$ if $c>1/e$. Moreover,

$$G''(u) \ = \ \frac{1}{1+u} - \frac{c}{1+cu} - \frac{c}{(1+cu)^2} \ = \ \frac{1-2c-c^2u}{(1+u)(1+cu)^2}.$$

Hence, if $c\le 1/2$, then G is convex in $u\le(1-2c)/c^2$ and concave in $u\ge(1-2c)/c^2$. In this case, $G(u)\ge 0$ for all $u\ge 0$, if and only if $G(\infty)\ge 0$, that is, $c\le 1/e$. Thus, the left inequality in (5.3) is fulfilled on the positive half-axis with an optimal value $c_0 = 1/e$.

The expression for the second derivative also shows that, in order that $G(u)$ be non-positive for all $u\ge 0$, it is necessary that $c\ge 1/2$. And if $c=1/2$, we get $G''(u)\le 0$, so, G is concave and thus non-positive. Hence, the right inequality in (5.3) is fulfilled on the positive half-axis with an optimal value $c_1 = 1/2$.

Case 2: Turning to the interval $[-1,0]$, let us make the substitution and consider the function

$$\begin{aligned} G(u) &= H(-u) - u \log(1+cu) \\ &= (1-u)\log(1-u) + u - u\log(1+cu), \qquad 0 \le u \le 1, \end{aligned}$$

with a parameter $c > 0$. We have $G(0) = G'(0) = 0$, $g(1) = 1 - \log(1+c)$,

$$G'(u) = -\log(1-u) - \log(1+cu) - 1 + \frac{1}{1+cu}.$$

Therefore, for G to be non-negative on $[0,1]$, it is necessary that $c \le e-1$, and for g to be non-positive on that interval, it is necessary that $c \ge e-1$. We also find

$$G''(u) = \frac{1}{1-u} - \frac{c}{1+cu} - \frac{c}{(1+cu)^2} = \frac{1 - 2c + c(4-c)u + 2c^2 u}{(1-u)(1+cu)^2}.$$

If moreover $c = 1/e$ as in Case 1 (when we considered the property $G \ge 0$), we have $G''(u) \ge 0$, so that, G is convex and thus non-negative. Thus, the left inequality in (5.3) is fulfilled on $[-1,0]$ with the same value $c_0 = 1/e$.

To get a reverse inequality, assume now that $c \ge e-1$ (which is necessary) and define $Q(u) = 1-2c+c(4-c)u+2c^2u$. We have $Q(0) = 1-2c < 0$ and $Q(1) = 1+2c+c^2 > 0$. Hence, there is a unique point $u_0 \in (0,1)$ such that $Q \le 0$ on $[0,u_0]$ and $Q \ge 0$ on $[u_0,1]$. Thus, G is concave on the first interval and is convex on the second one. Since $G(0) = G'(0) = 0$, the property $G \le 0$ on $[0,1]$ is therefore equivalent to $G(1) \le 0$, which is the case. Hence, the right inequality in (5.3) is fulfilled on $[0,1]$ with an optimal value $c_1 = e-1$. $\square$

Let us now specialize Theorem 5.1 to the case $\Omega = \mathbb{R}^d$ with the Lebesgue measure λ and the normal density $q = \varphi$, in which case, for any probability density p on $\mathbb{R}^d$,

$$\begin{aligned} \int_{\mathbb{R}^d} |p-\varphi| \log\Big(1 + c_0\,\frac{|p-\varphi|}{\varphi}\Big)\,dx &\le D(p||\varphi) \\ &\le \int_{\mathbb{R}^d} |p-\varphi| \log\Big(1 + c_1\,\frac{|p-\varphi|}{\varphi}\Big)\,dx. \end{aligned} \tag{5.4}$$

Since $\frac{c_0}{\varphi} \ge \frac{1}{e}\sqrt{2\pi} > 0.9$, and using the elementary inequality $\log(1+ct) \ge \min\{c,1\}\,\log(1+t)$, we see that the left integral in (5.4) is greater than or equal to

$$0.9 \int_{\mathbb{R}^d} |p(x)-\varphi(x)| \log\big(1 + |p(x)-\varphi(x)|\big)\,dx = 0.9 \int_{\mathbb{R}^d} \psi\big(p(x)-\varphi(x)\big)\,dx.$$

For an opposite inequality, one may use $\log(1+ab) \le \log a + \log(1+b)$ $(a \ge 1,\ b \ge 0)$, which allows us to bound the right-hand side of (5.4) from above by

$$\log\Big(c_1\,(2\pi)^{d/2}\Big) \int_{\mathbb{R}^d} |p(x)-\varphi(x)|\,dx + \frac{1}{2}\int_{\mathbb{R}^d} |x|^2\,|p(x)-\varphi(x)|\,dx + \int_{\mathbb{R}^d} \psi\big(p(x)-\varphi(x)\big)\,dx.$$

Here, the first factor may further be bounded by $d+1$. One may conclude.

Corollary 5.2. *For any probability density p on $\mathbb{R}^d$,*

$$0.9 \int_{\mathbb{R}^d} \psi\big(p(x) - \varphi(x)\big)\,{\rm d}x \ \ \leq \ \ D(p||\varphi) \ \leq \ \int_{\mathbb{R}^d} \psi\big(p(x) - \varphi(x)\big)\,{\rm d}x$$
$$+ \int_{\mathbb{R}^d} W_d(|x|)\,|p(x) - \varphi(x)|\,{\rm d}x, \qquad (5.5)$$

where $\psi(t) = |t|\,\log(1+|t|)$ and $W_d(t) = d + 1 + \frac{1}{2}\,t^2$.

The last integral in (5.5) represents the weighted total variation distance, with weight $W_d(|x|)$, between the standard Gaussian measure γ and the probability measure μ on $\mathbb{R}^d$ with density p.

6. **Bounds on moments in terms of relative entropy**

Let ξ be a random vector in $\mathbb{R}^d$ with an absolutely continuous distribution with density p. The finiteness of the relative entropy $D(p||\varphi)$ forces ξ to have a finite second moment, that is, $\mathbb{E}\,|\xi|^2 < \infty$. In that case, one may define the mean

$$a = \mathbb{E}\xi = \int_{\mathbb{R}^d} xp(x)\,{\rm d}x$$

(which is a vector in $\mathbb{R}^d$) and the covariance matrix R, which is an invertible, symmetric $d \times d$ matrix such that

$$\mathbb{E}\,\langle \xi - a, v\rangle^2 = \int_{\mathbb{R}^d} \langle x - a, v\rangle^2\, p(x)\,{\rm d}x = \langle Rv, v\rangle$$

for all $v \in \mathbb{R}^d$. Moreover, the smallness of $D(p||\varphi)$ insures that a is close to zero (which is the mean of a standard normal random vector Z in $\mathbb{R}^d$), while R should be close to the identity matrix I_d (which is the covariance matrix of Z). The following statement quantifying this property is recently proved in [5].

Lemma 6.1. *Putting $D = D(p||\varphi)$, we have*

$$D \ \geq \ \frac{1}{2}\,|a|^2 + \frac{1}{16}\,\sum_{i=1}^d \min\big\{|\sigma_i^2 - 1|, (\sigma_i^2 - 1)^2\big\}, \qquad (6.1)$$

where σ_i^2 are eigenvalues of the covariance matrix R. In particular,

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

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

c) $\big|\mathbb{E}\,|\xi|^2 - d\,\big| \leq 4d\sqrt{D} + 16d\,D.$

For completeness, let us include a short argument. Denote by q the density of the Gaussian measure on $\mathbb{R}^d$ with mean a and covariance matrix R, that is,

$$q(x) \ = \ \frac{1}{(2\pi)^{d/2}\sqrt{\det(R)}}\,\exp\Big\{-\frac{1}{2}\,\big\langle R^{-1}(x-a), x-a\big\rangle\Big\}, \qquad x \in \mathbb{R}^d.$$

Proof. By the definition,

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

Simplifying, we obtain an explicit formula

$$\begin{aligned} D &= D(p||q) + \frac{1}{2}\, |a|^2 + \frac{1}{2} \Big(\log \frac{1}{\det(R)} + \mathrm{Tr}(R) - d\Big) \\ &= D(p||q) + \frac{1}{2}\, |a|^2 + \frac{1}{2} \sum_{i=1}^d U(\sigma_i^2), \qquad U(t) = \log \frac{1}{t} + t - 1. \end{aligned} \tag{6.2}$$

All the terms on the right-hand side are non-negative, and we thus obtain (6.1), which in turn implies a).

For the next claim, note that the function $U(t)$ is convex in $t > 0$, and satisfies $U(1) = U'(1) = 0$, $U''(t) = 1/t^2$. If $|t-1| \leq 1$, by Taylor's formula about the point $t_0 = 1$ with some point t_1 between t and 1,

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

For the values $t \geq 2$, we have a linear bound $U(t) \geq \frac{1}{8}\,(t-1)$, so that the two bounds yield

$$U(t) \geq \frac{1}{8}\, \min\{|t-1|, |t-1|^2\}, \qquad t > 0,$$

which implies b). Finally, since $\mathbb{E}\,|\xi|^2 = \sigma_1^2 + \dots + \sigma_d^2$, the claim c) readily follows from b). $\square$

Note that the closeness of all eigenvalues to 1 may also be stated as closeness of R to the identity matrix. For example, in terms of the Hilbert-Schmidt norm, we have, by b),

$$\|R - I_d\|_{\mathrm{HS}}^2 \,=\, \sum_{i=1}^d (\sigma_i^2 - 1)^2 \,\leq\, Cd\, \max\big\{D(p||\varphi), D(p||\varphi)^2\big\}$$

with some absolute constant C.

7. **Proof of Theorem 1.4**

In one direction, we apply Corollary 5.2 and Lemma 6.1. Assuming that $D_n(p||\varphi) \to 0$ as $n \to \infty$, the first inequality in (5.5) shows that

$$\int \psi(p_n(x) - \varphi(x))\, \mathrm{d}x \to 0, \tag{7.1}$$

which is the required convergence (1.6). Since the Young function ψ satisfies the Δ_2-condition, the latter is equivalent to $\|p_n - \varphi\|_\psi \to 0$ (as explained in Remark 4.4). Moreover, by the inequality c) in Lemma 6.1 applied to the random vectors ξ_n in $\mathbb{R}^d$ with densities p_n, we also have

$$\big|\mathbb{E}\,|\xi_n|^2 - d\,\big| \,\leq\, 4d\sqrt{D(p_n||\varphi)} + 16d\,D(p_n||\varphi) \,\to\, 0.$$

This proves the property a) in Theorem 1.4.

Now, suppose that (7.1) holds true together with $\mathbb{E}\,|\xi_n|^2 \to d$. Using the second inequality in (5.5), it remains to show that

$$I_n \,=\, \int_{\mathbb{R}^d} W_d(|x|)\,|p_n(x) - \varphi(x)|\,\mathrm{d}x \,\to\, 0,$$

where $W_d(t) = d+1+\frac{1}{2}\,t^2$. Using the notation $z^+ = \max(z,0)$ and the identity $|z| = 2z^+ - z$ ($z \in \mathbb{R}$), the above integral may be rewritten (like in Scheffe's lemma) as

$$\begin{aligned} I_n \,&=\, 2\int_{\mathbb{R}^d} W_d(|x|)\,(\varphi(x) - p_n(x))^+\,\mathrm{d}x + \int_{\mathbb{R}^d} W_d(|x|)\,(p_n(x) - \varphi(x))\,\mathrm{d}x \\ &=\, 2\int_{\mathbb{R}^d} W_d(|x|)\,(\varphi(x) - p_n(x))^+\,\mathrm{d}x + \frac{1}{2}\,\big(\mathbb{E}\,|\xi_n|^2 - d\,\big). \end{aligned}$$

Here, the last integral tends to zero as $n \to \infty$. Splitting the integration over the ball $|x| \leq T_n$ and its complement, the second last integral may be bounded from above by

$$W_d(T_n)\,\|p_n - \varphi\|_1 + \int_{|x|\geq T_n} W_d(|x|)\,\varphi(x)\,\mathrm{d}x. \tag{7.2}$$

By the assumption (7.1), we have $\|p_n - \varphi\|_1 \to 0$ (since the $\|\cdot\|_\psi$-norm is stronger than the L^1-norm). Hence, one may choose a sequence T_n which grows to infinity sufficiently slow, so that the first term in (7.2) tends to zero as well. In that case, the whole expression in (7.2) tends to zero, and as a result, $I_n \to 0$. This finishes the proof of Theorem 1.4.

□

Remarks. Let us return to the normalized sums Z_n in (1.1) for independent identically distributed random variables $(X_n)_{n\geq 1}$ with common density w. To illustrate the range of applicability of the uniform local limit theorem (1.3), Gnedenko and Kolmogorov considered in [7]-[8] the example of the symmetric, compactly supported density

$$p(x) = \begin{cases} 0, & \text{if } |x| > 1/e, \\ \frac{\alpha}{2\,|x|\,\log^{\alpha+1}(1/|x|)}, & \text{if } |x| < 1/e, \end{cases}$$

with $\alpha = 1$. Define $w(x) = \frac{1}{\lambda}\,p(x/\lambda)$, where the constant $\lambda > 0$ is chosen so that $\mathbb{E}X_1^2 = 1$. As was shown, near the origin $x = 0$ the n-th convolution power $p^{*n}(x)$ admits a lower bound

$$p^{*n}(x) \geq \frac{c_n}{|x|\,\log^{\alpha n+1}(1/|x|)}$$

with some constant $c_n > 0$. Hence, all densities p_n of Z_n are unbounded in any neigbourhood of zero and therefore do not satisfy (1.3).

To illustrate the entropic central limit theorem, Barron [2] returned to this example, assuming that α is an arbitrary positive parameter. Although the densities p_n are still unbounded, it was noticed that the entropies $h(p_n)$ are finite as long as $n > 1/\alpha$. Hence, Z_n do satisfy the entropic CLT (by Theorem 1.3).

References

[1] Alexopoulos, J. A brief introduction to N–functions and Orlicz function spaces. Kent State University, 2004.

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

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

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

[5] Bobkov, S. G; Marsiglietti, A. Entropic CLT for smoothed convolutions, and associated entropy bounds. Preprint (2019).

[6] Fedotov, A.; Harremoës, P.; Topsoe, F. Refinements of Pinsker's inequality. IEEE Transactions on Information Theory 49 (2003), 1491–1498.

[7] Gnedenko, B. V.; Kolmogorov, A. N. Predelnye raspredeleniya dlya summ nezavisimyh sluchainyh velichin. (Russian) [Limit Distributions for Sums of Independent Random Variables] Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad], 1949. 264 pp.

[8] Gnedenko, B. V.; Kolmogorov, A. N. Limit distributions for sums of independent random variables. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954. ix+264 pp.

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

[10] Petrov, V. V. Sums of independent random variables. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer-Verlag, New York-Heidelberg, 1975. x+346 pp. Russian ed.: Moscow, Nauka, 1972, 414 pp.

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

[12] Ranga Rao, R.; Varadarajan, V. S. A limit theorem for densities. Sankhya 22 (1960), 261–266.