Multidimensional local limit theorem for densities in Orlicz spaces

Sergey Bobkov
University of Minnesota

High Dimensional Probability 14-19 June 2020
Bedlewo, Poland, online 19 June 2020

Central limit theorem

Setting: $\{X_k\}_{k\geq 1}$ iid in $\mathbb{R}^d$, $\mathbb{E}X_1 = 0$, $\operatorname{cov}(X_1) = \mathrm{Id}$,

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

$$Z \sim N(0, \mathrm{Id}), \quad \varphi(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}, \quad x \in \mathbb{R}^d.$$

CLT: $Z_n \Rightarrow Z$ weakly in distribution, $\mathbb{E}\, u(Z_n) \to \mathbb{E}\, u(Z)$, i.e.

$$\int u(x)\, d\mu_n(x) \to \int u(x)\, \varphi(x)\, dx \qquad (n \to \infty)$$

for any bounded continuous $u : \mathbb{R}^d \to \mathbb{R}$, where $\mu_n = \mathcal{L}(Z_n)$.

Local limit theorems: If Z_n have densities φ_n for large/some n, then φ_n approximate φ.

Theorem (Ranga Rao-Varadarajan 1960): $\varphi_n(x) \to \varphi(x)$ a.e. Hence CLT.

Local limit theorems in L^1 and L^∞

Theorem (Prokhorov 1952): If Z_n have densities φ_n for large n, then

$$\|\varphi_n - \varphi\|_1 = \int |\varphi_n(x) - \varphi(x)|\, dx \to 0.$$

Proof: Decomposition of densities. Consequence of RR-V (by applying Scheffe's lemma).

Theorem (Gnedenko-Kolmogorov 1949, 1954): If Z_n has density $\varphi_n \in L^p$ for some n and $1 < p \le 2$, then φ_n exist and are bounded for all large n, and

$$\|\varphi_n - \varphi\|_\infty = \operatorname{ess\,sup}_x |\varphi_n(x) - \varphi(x)| \to 0.$$

Equivalent necessary and sufficient condition: $\varphi_n \in L^\infty$ for some n.

Notes

1) $\|\varphi_n\|_\infty < \infty \Rightarrow \varphi_{2n}$ is continuous, so $\operatorname{ess\,sup} = \sup$.

2) φ_n has Fourier transform $f_n(t) = f(\frac{t}{\sqrt{n}})^n$. If $\|\varphi_n\|_p < \infty$, $1 < p \le 2$, then by the Hausdorff-Young inequality, $\|f_n\|_q < \infty$, $q = \frac{p}{p-1}$. Hence $\|f\|_{nq} < \infty$ and $\|\varphi_m\|_\infty < \infty$ for all $m \ge nq$.

Local limit theorems in L^p

What about

$$\|\varphi_n - \varphi\|_p = \left(\int |\varphi_n(x) - \varphi(x)|^p \, dx \right)^{1/p} \quad (p > 1) \; ?$$

Corollary. Suppose that Z_n have densities φ_n for all large n. The following properties are equivalent:

$a)$ $\|\varphi_n - \varphi\|_p \to 0$;
$b)$ $\|\varphi_n - \varphi\|_\infty \to 0$.

Proof. In general $\|u\|_p \le \|u\|_1 + \|u\|_\infty$, so $b) \Rightarrow a)$.
Assuming $a)$, we get $\|\varphi_n - \varphi\|_q \to 0$ for $1 < q \le p$. Hence $b)$, by Gnedenko-Kolmogorov.

Bhachattarya-Ranga Rao (1976):

In terms of the characteristic function $f(t) = \mathbb{E}\, e^{it \cdot X_1}$, Z_n have bounded densities φ_n for large (some) n, if and only

$$f \in L^q(\mathbb{R}^d) \;\; \text{for some } q \;\; \text{(smoothing condition)}.$$

Entropic CLT

Kullback–Leibler distance (relative entropy): Given μ, ν probability measures on a measurable space Ω with densities $p = \frac{d\mu}{d\lambda}$, $q = \frac{d\nu}{d\lambda}$,

$$D(\mu||\nu) = D(p||q) = \int p \log(p/q)\, d\lambda.$$

Pinsker-type inequality

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

Theorem (Barron 1986, $d = 1$). If Z_n has density φ_n such that $D(\varphi_n||\varphi) < \infty$ for some n, then

$$D(\varphi_n||\varphi) \to 0.$$

Proof: de Bruijn identity (simplified Harremoës-Vignat 2005)
Rates: B-Chistyakov-Goetze 2013 (decomposition of densities)
Note: In terms of entropies $h(p) = -\int p \log p\, dx$,

$$D(\varphi_n||\varphi) = h(\varphi) - h(\varphi_n).$$

Open: How to express $D(\varphi_n||\varphi) < \infty$ in terms of $f(t)$?

Local limit theorem in Orlicz spaces

Young function $\Psi : \mathbb{R} \to [0, \infty)$, even, convex, $\Psi(0) = 0$, $\Psi(t) > 0$ for $t > 0$.
Orlicz norm for u on $\mathbb{R}^d$

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

The Orlicz space L^Ψ contains all u with $\|u\|_\Psi < \infty$.
Limit Orlicz norm: $\|u\|_\infty = \operatorname{ess\,sup}_x |u(x)|$.

Theorem 1. Suppose that Z_n have densities φ_n for large n. For a given Orlicz norm, $\|\varphi_n - \varphi\| \to 0$ if and only if $\|\varphi_n\| < \infty$ for some n.

Corollary. Suppose that Z_n have densities φ_n for large n, and let $\Psi(2t) \le c\Psi(t)$ with some $c > 0$ independent of $t \ge 0$ (Δ_2-condition). Then

$$\int \Psi\big(\varphi_n(x) - \varphi(x)\big)\, dx \to 0$$

if and only if $\int \Psi(\varphi_n(x))\, dx < \infty$ for some n.

Convergence in D via Orlicz norms

Introduce the Young function

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

Theorem 2. 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:

$a)$ $\int |x|^2 \, p_n(x) \, dx \to d$

$b)$ $\|p_n - \varphi\|_\psi \to 0$ as $n \to \infty$.

The last condition may be replaced with

$$\int |p_n - \varphi| \, \log(1 + |p_n - \varphi|) \, dx \to 0.$$

Two-sided estimate on relative entropy

Lemma 1. Given probability densities p, q on a measure space (Ω, λ), we have

$$\int |p-q| \, \log\left(1 + c_0 \frac{|p-q|}{q}\right) d\lambda \;\le\; D(p||q) \;\le\; \int |p-q| \, \log\left(1 + c_1 \frac{|p-q|}{q}\right) d\lambda.$$

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

Consider

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

so that

$$D(p||q) \;=\; \int \frac{p}{q} \log\frac{p}{q} \, d\nu \;=\; \int H\Big(\frac{p-q}{q}\Big) \, q \, d\lambda.$$

Hence, Lemma 1 would follow from the two-sided bound

$$|t| \, \log(1 + c_0|t|) \;\le\; H(t) \;\le\; |t| \, \log(1 + c_1|t|).$$

Relative entropy with respect to normal

Lemma 2. For any probability density p on $\mathbb{R}^d$,

$$0.9 \int \psi\big(p(x) - \varphi(x)\big)\, dx \;\le\; D(p||\varphi)$$

$$\le\; \int \psi\big(p(x) - \varphi(x)\big)\, dx + \int W_d(|x|)\, |p(x) - \varphi(x)|\, dx,$$

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

Proof. By Lemma 1, on $\Omega = \mathbb{R}^d$ with $\lambda = \mathrm{mes}$ and $q = \varphi$,

$$\int |p - \varphi|\, \log\Big(1 + c_0\, \frac{|p - \varphi|}{\varphi}\Big)\, dx \;\le\; D(p||\varphi)$$

$$\le\; \int |p - \varphi|\, \log\Big(1 + c_1\, \frac{|p - \varphi|}{\varphi}\Big)\, dx.$$

Using $\frac{c_0}{\varphi} > 0.9$ and $\log(1 + c_0 t) \ge c_0 \log(1 + t)$, the lower bound on D follows. Since $\log(1 + ab) \le \log a + \log(1 + b)$ ($a \ge 1$, $b \ge 0$), the last integral $\le$

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

Bounds on moments in terms of D

Lemma 3 (B-Marsliglietti). Given a r.v. X in $\mathbb{R}^d$ with finite $D = D(p||\varphi)$,

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

where $a = \mathbb{E}X$ and σ_i^2 are eigenvalues of the covariance matrix R of X. In particular,

$a)$ $|a|^2 \leq 2D$;
$b)$ $|\sigma_i^2 - 1| \leq 4\sqrt{D} + 16D$ for all $i \leq d$;
$c)$ $\left|\mathbb{E}\,|X|^2 - d\,\right| \leq 4d\sqrt{D} + 16d\,D$.

Finiteness of $D(p||\varphi)$ forces X to have a finite second moment. So, one may define $a = \mathbb{E}X$ and the covariance matrix R, which is an invertible, symmetric $d \times d$ matrix such that

$$\mathbb{E}\,\langle X - a, v\rangle^2 = \langle Rv, v\rangle\,, \quad v \in \mathbb{R}^d.$$

Thus, the smallness of D insures that a is close to zero, while R is close to the identity matrix I_d.

Proof of Lemma 3

Denote by q the density of $N(a,R)$, that is,

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

By definition,

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

By Taylor,

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

which implies $b)$. Since $\mathbb{E}\,|X|^2 = \sigma_1^2 + \cdots + \sigma_d^2$, $c)$ follows from $b)$.

Proof of Theorem 2

($\Rightarrow$) Assume $D(p_n||\varphi) \to 0$. By Lemma 2 (first inequality), $\int \psi(p_n - \varphi)\,dx \to 0$.
By Lemma 3, if $\xi_n \sim p_n$,

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

($\Leftarrow$) By Lemma 2 (second inequality), it remains to show that

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

where $W_d(t) = d + 1 + \frac{1}{2}\,t^2$. Since $|z| = 2z^+ - z$,

$$\begin{aligned} I_n \;&=\; 2\int W_d(|x|)\,(\varphi(x) - p_n(x))^+\,dx + \frac{1}{2}\left(\mathbb{E}\,|\xi_n|^2 - d\right) \\ &\le\; 2\,W_d(T_n)\,\|p_n - \varphi\|_1 + 2\int_{|x|\ge T_n} W_d(|x|)\,\varphi(x)\,dx + o(1). \end{aligned}$$

But, $\|p_n - \varphi\|_1 \to 0$, since $\|\cdot\|_\psi$-norm is stronger than L^1-norm. Choose T_n growing to infinity sufficiently slow.

Decomposition of convolution powers

Let w be a probability density on $\mathbb{R}^d$, $w_n = w^{*n}$ convolution power.

Tool (Prokhorov 1952): Decompose $w_n = \mathrm{Part}_1 + \mathrm{Part}_2$ into two parts, one of which is bounded and approximates 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.

Given $0 < \delta_1 \le \frac{1}{4}$, decompose $\mathbb{R}^d$ into two sets $\Omega_0 \subset \{w \le M\}$ and $\Omega_1 \subset \{w \ge M\}$ such that

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

Decomposition: $w = \delta_0 p_0 + \delta_1 p_1$ in which p_0 and p_1 are the normalized restrictions of w to Ω_0 and Ω_1. As a result,

$$w^{*n} = \sum_{l=0}^{n} C_n^l\,\delta_0^l\,\delta_1^{n-l}\,q_l, \qquad q_l = p_0^{*l} * p_1^{*(n-l)}.$$

Removing the first two terms, define a canonical approximation for w_n

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

CLT for approximating densities

Let $X_1, X_2, \dots$ be iid in $\mathbb{R}^d$ with mean zero, identity covariance matrix, and with density w. Denote by φ_n the densities of

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

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

As a canonical approximation for p_n, one may use

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

Lemma 4. $\tilde{\varphi}_n$ is bounded, continuous, and satisfies

$$\int |\tilde{\varphi}_n(x) - \varphi_n(x)|\, dx \, < \, \frac{1}{2^{n-1}}.$$

Moreover,

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

Proof: Fourier inverse formula.

Properties of Orlicz norms

If $\|\cdot\| = \|\cdot\|_\Psi$, assume that $\Psi(1) = 1$.

Lemma 5. For any measurable u on $\mathbb{R}^d$,

$a)$ $\|u\| \leq \max\big\{\|u\|_1, \|u\|_\infty\big\}$.

$b)$ $\|u(\lambda x)\| \leq \|u(x)\|$ for all $\lambda \geq 1$.

$c)$ For all measurable functions $u_1, \ldots, u_N$ on $\mathbb{R}^d$ $(N \geq 2)$,

$$\|u_1 * u_2 * \cdots * u_N\| \ \leq\ \|u_1\|\,\|u_2\|_1 \ \ldots\ \|u_N\|_1.$$

Proof. $a)$ follows from $\Psi(t) \leq |t|$ for $|t| \leq 1$. In $c)$ consider the case $N = 2$ (using induction) and assume that $\|u_1\| = 1$ and $\|u_2\|_1 = 1$. If $\|\cdot\| = \|\cdot\|_\Psi$, then by Jensen,

$$\Psi\big((u_1 * u_2)(x)\big) \ = \ \Psi\Big(\int u_1(x-y)\,u_2(y)\,dy\Big) \ \leq \ \int \Psi(|u_1(x-y)|)\,|u_2(y)|\,dy,$$

so

$$\int \Psi\big((u_1 * u_2)(x)\big)\,dx \ \leq \iint \Psi(|u_1(x-y)|)\,|u_2(y)|\,dy\,dx = 1.$$

Proof of Theorem 1

Assume that X_1 has density w with $\|w\| < \infty$. Need to show: $\|\varphi_n - \varphi\| \to 0$.

From the decomposition $w = \delta_0 p_0 + \delta_1 p_1$, it follows that $\|p_i\| < \infty$. Indeed, $p_0 \le M/\delta_0$, and with $\lambda = \|w\|_\Psi$

$$1 = \int \Psi(w/\lambda)\,dx = \int_{\Omega_0} \Psi(\delta_0 p_0/\lambda)\,dx + \int_{\Omega_1} \Psi(\delta_1 p_1/\lambda)\,dx,$$

so $\|p_1\|_\Psi \le \lambda/\delta_1$.

By Lemma 4, $\|\tilde{\varphi}_n - \varphi\|_\infty \to 0$ which implies $\|\tilde{\varphi}_n - \varphi\|_1 \to 0$ (Scheffe's lemma). By Lemma 5a, $\|\tilde{\varphi}_n - \varphi\| \to 0$ as well. In view of the triangle inequality in the Orlicz space, it remains to see that

$$\|\tilde{\varphi}_n - \varphi_n\| \to 0.$$

Recall

$$\tilde{\varphi}_n(x) = n^{d/2}\,\tilde{w}_n(n^{1/2}x), \quad \varphi_n(x) = n^{d/2}\,w_n(n^{1/2}x),$$

where $w_n = w^{*n}$. Applying Lemma 5b, it follows that

$$\|\tilde{\varphi}_n - \varphi_n\| \le n^{d/2}\,\|\tilde{w}_n - w_n\|.$$

Proof of Theorem 1 (cont.)

Return to the definition

$$\tilde{w}_n \;=\; \frac{1}{1-\varepsilon_n} \sum_{l=2}^{n} C_n^l \, \delta_0^l \, \delta_1^{n-l} \, q_l, \qquad q_l = p_0^{*l} * p_1^{*(n-l)},$$

with $\varepsilon_n = \delta_1^n + n \, \delta_0 \delta_1^{n-1} \leq 2^{-(n-1)}$ which yields

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

By Lemma 5c, $\|q_l\| \leq \|p_0\|$, so $(1-\varepsilon_n) \, \|\tilde{w}_n\| \;\leq\; \|p_0\|$ and thus

$$\|\tilde{w}_n - (1-\varepsilon_n) \, \tilde{w}_n\| \;=\; \varepsilon_n \, \|\tilde{w}_n\| \;\leq\; \frac{\varepsilon_n}{1-\varepsilon_n} \, \|p_0\|.$$

This gives

$$\|\tilde{\varphi}_n - \varphi_n\| \;\leq\; n^{d/2} \, \|(1-\varepsilon_n) \, \tilde{w}_n - w_n\| + \frac{\varepsilon_n}{1-\varepsilon_n} \, n^{d/2} \, \|p_0\|.$$

Here

$$w_n - (1-\varepsilon_n) \, \tilde{w}_n \;=\; \delta_1^n q_0 + n \, \delta_0 \delta_1^{n-1} q_1 \;=\; \delta_1^n p_1^{*n} + n \, \delta_0 \delta_1^{n-1} \, p_0 * p_1^{*(n-1)}.$$

By Lemma 5c, the norm of this expression does not exceed $\varepsilon_n \, \|p_1\|$, and we arrive at

$$\|\tilde{\varphi}_n - \varphi_n\| \;\leq\; c n^{d/2} \, 2^{-n} \left(\|p_0\| + \|p_1\|\right) \to 0.$$

Δ_2-condition

The Δ_2-condition $\Psi(2t) \le c\Psi(t)$ implies that $\Psi(t) = O(t^p)$ as $t \to \infty$ with some $p \ge 1$. A necessary and sufficient condition for this property is that

$$\sup_{t \ge t_0} \frac{t\Psi'(t+)}{\Psi(t)} \le C$$

for some $t_0 > 0$ and $C < \infty$.

The Δ_2-condition may be written more generally as

$$\Psi(\lambda t) \le c_\lambda \Psi(t), \qquad t \in \mathbb{R},$$

where the constant c_λ depends on the fixed number $\lambda > 1$ only.

Claim.

$a)$ For any measurable function u on $\mathbb{R}^d$, $\|u\|_\Psi < \infty$ if and only if $\int \Psi(u(x))\, dx < \infty$.

$b)$ Given a sequence of measurable functions $(u_n)_{n\ge1}$ on $\mathbb{R}^d$, we have $\|u_n\|_\Psi \to 0$ if and only if $\int \Psi(u_n(x))\, dx \to 0$ as $n \to 0$.