Rényi divergence and the central limit theorem

Sergey Bobkov (University of Minnesota)

based on joint work with Gennadiy Chistyakov
and Friedrich Götze (Bielefeld University)

MSRI workshop "Geometric functional analysis and applications" 13–17 November 2017, Berkeley

---

Rényi divergence

X, Z random elements in a measure space (Ω, μ)
P, Q distributions with densities $p = \frac{dP}{d\mu}$, $q = \frac{dQ}{d\mu}$
$0 < \alpha < \infty$

Definition (Rényi 1961, Tsallis 1998)
The Rényi divergence of P from Q of index α is

$$D_\alpha(X||Z) \;=\; D_\alpha(P||Q) \;=\; \frac{1}{\alpha - 1} \log \int \left(\frac{p}{q}\right)^\alpha q \, d\mu.$$

The Rényi divergence power or relative Tsallis entropy

$$T_\alpha(X||Z) \;=\; T_\alpha(P||Q) \;=\; \frac{1}{\alpha - 1} \left[\int \left(\frac{p}{q}\right)^\alpha q \, d\mu - 1 \right].$$

Monotone transformations:

$$D_\alpha \;=\; \frac{1}{\alpha - 1} \log \big(1 + (\alpha - 1)\, T_\alpha\big),$$
$$T_\alpha \;=\; \frac{1}{\alpha - 1} \big[\, e^{(\alpha-1) D_\alpha} - 1\big].$$

Equivalence: $D_\alpha \sim T_\alpha$ (when small)

Properties

- Independence of the dominating measure μ

- Separation: $D_\alpha(P||Q) \geq 0$, and $D_\alpha(P||Q) = 0$ if and only if $P = Q$.

- Monotonicity: The functions $\alpha \to D_\alpha(P||Q)$ and $\alpha \to T_\alpha(P||Q)$ are non-decreasing.

- Contractivity under mappings:

$$D_\alpha(S(X)||S(Z)) \leq D_\alpha(X||Z) \qquad (\alpha \geq 1).$$

- Range $0 < \alpha < 1$: All D_α are comparable to each other and are metrically equivalent to the total variation $\|P - Q\|_{\mathrm{TV}}$. Gilardoni's inequality (2010):

$$D_\alpha(P||Q) \geq \frac{\alpha}{2} \|P - Q\|_{\mathrm{TV}}^2.$$

This extends the Pinsker inequality for the Kullback-Leibler distance ($\alpha = 1$).

- Range $\alpha \geq 1$: $D_\alpha(P||Q) < \infty \Rightarrow P \ll Q$.

Particular cases

- $\alpha = 1/2$ (Hellinger distance)
- $\alpha = 1$ (Kullback-Leibler distance, relative entropy):

$$D(X||Z) = D(P||Q) = \int p \log \frac{p}{q} \, d\mu.$$

- $\alpha = 2$ (χ^2-distance, quadratic Renyi divergence):

$$\begin{aligned} D_2(X||Z) &= \log \int \frac{p^2}{q} \, d\mu, \\ \chi^2(X,Z) = T_2(X||Z) &= \int \frac{(p-q)^2}{q} \, d\mu. \end{aligned}$$

In all cases

$$\begin{aligned} \frac{1}{2} \|P - Q\|_{\mathrm{TV}}^2 &\le D(X||Z) \\ &\le D_2(X||Z) \le \chi^2(X,Z). \end{aligned}$$

Goodness of fit test (Karl Pearson 1900): If Q is unknown distribution with k atoms, $P = P_n$ empirical, then

$$n\chi^2(P_n, Q) \Rightarrow \chi^2_{k-1} = \mathcal{L}(Z_1^2 + \cdots + Z_{k-1}^2)$$

Rényi divergence from Gaussian

$\Omega = \mathbb{R}^d$ with Lebesgue measure $d\mu(x) = dx$
X random vector with density $p(x)$
$Z \sim N(0, \mathrm{I})$ standard normal random vector with density

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

Rényi divergence–Tsallis distance of index α are given by

$$D_\alpha(X||Z) \;=\; \frac{1}{\alpha - 1}\,\log \int \frac{p^\alpha}{\varphi^{\alpha-1}}\, dx,$$
$$T_\alpha(X||Z) \;=\; \frac{1}{\alpha - 1}\int \frac{p^\alpha}{\varphi^{\alpha-1}}\, dx - 1.$$

Relative entropy ($\alpha = 1$), $\mathbb{E}X = 0$, $\mathrm{cov}(X) = \mathrm{I}$

$$D(X||Z) = h(Z) - h(X)$$

in terms of Shannon entropy $h(X) = -\int p\, \log p\; dx$.
Pearson ($\alpha = 2$)

$$\chi^2(X, Z) = \int \frac{(p - \varphi)^2}{\varphi}\, dx.$$

Exponential integrability

Let $d = 1$, $Z \sim N(0,1)$, $\beta = \frac{\alpha}{\alpha-1}$.

Note: If $D(X||Z) < \infty$, then $\mathbb{E}X^2 < \infty$.

Proposition 1. If $T_\alpha = T_\alpha(X||Z) < \infty$ for $\alpha > 1$, then X has an absolutely continuous distribution and finite moments of any order. Moreover,

$$\mathbb{E}\, e^{cX^2} \ < \ \infty \quad \text{for all} \ \ c < 1/(2\beta).$$

For all $t \in \mathbb{R}$,

$$\mathbb{E}\, e^{tX} \leq Ce^{\beta t^2/2} \quad \text{with } C = \Big(1 + (\alpha - 1)\, T_\alpha\Big)^{1/\alpha}.$$

It is possible that $T_\alpha < \infty$, while $\mathbb{E}\, e^{\frac{1}{2\beta}X^2} = \infty$.
If p is density of X,

$$\begin{aligned}
\mathbb{E}\, e^{tX} \ &= \ \int_{-\infty}^{\infty} p(x)\, e^{tx}\, dx \\
&= \ \int_{-\infty}^{\infty} \frac{p(x)}{\varphi(x)^{1/\beta}} \cdot e^{tx} \varphi(x)^{1/\beta}\, dx \\
&\leq \ C \left(\int_{-\infty}^{\infty} e^{\beta tx}\, \varphi(x)\, dx \right)^{1/\beta} = \ Ce^{\beta t^2/2}.
\end{aligned}$$

Improved integrability for convolutions

Case $\alpha = 2$: If $\chi^2 = \chi^2(X, Z) < \infty$, then

$$\mathbb{E}\, e^{cX^2} \, < \, \infty \quad \text{for all} \;\; c < 1/4.$$

For all $t \in \mathbb{R}$,

$$\mathbb{E}\, e^{tX} \le C e^{t^2/4} \quad \text{with} \;\; C = (1+\chi^2)^{1/2}.$$

Proposition 2. If X_1, X_2 are independent copies of X,

$$\mathbb{E}\, e^{\frac{1}{4}\,(\frac{X_1+X_2}{\sqrt{2}})^2} \; \le \; 2\,(1+\chi^2).$$

For general $\alpha > 1$ with conjugate $\beta = \frac{\alpha}{\alpha-1}$, we need $k \ge \alpha$ normalized convolutions to include the critical coefficient $c = 1/(2\beta)$: If $T_\alpha = T_\alpha(X||Z)$ is finite, then

$$\mathbb{E}\, e^{\frac{1}{2\beta}\,Z_k^2} \; \le \; 2^k \Big(1 + (\alpha-1)\,T_\alpha\Big)^{\frac{k}{\alpha}}.$$

Moreover, if $T_\alpha(Z_k||Z) \to 0$, then

$$\mathbb{E}\, e^{\frac{1}{2\beta}\,Z_k^2} \, \to \, \mathbb{E}\, e^{\frac{1}{2\beta}\,Z^2} \; = \; \sqrt{\pi\,(\alpha-1)}.$$

Proof. Use of Plancherel theorem in case $\alpha = 2$ (Weierstrass transform for $\alpha > 1$).

Exponential series and normal moments

Question: How to connect $\chi^2(X,Z)$ to the moments of X ?
Exponential orthogonal series (Cramér):

$$p(x) = \varphi(x) \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x)$$

converges in $L^2(\mathbb{R}, \frac{dx}{\varphi(x)})$ if and only if $\sum_{k=0}^{\infty} c_k^2 < \infty$.
Fourier coefficients (normal moments of X):

$$c_k = \int_{-\infty}^{\infty} H_k(x)\, p(x)\, dx = \mathbb{E}\, H_k(X) = \mathbb{E}\,(X+iZ)^k.$$

In particular, $c_0 = 1$, $c_1 = \mathbb{E}X$, $c_2 = \mathbb{E}X^2 - 1$.
Taylor series around zero for the characteristic function:

$$f(t) = \mathbb{E}\, e^{itX} = e^{-t^2/2} \sum_{k=0}^{n} \frac{c_k}{k!} (it)^k + o(|t|^n).$$

Proposition 3. If $\chi^2(X,Z) < \infty$, then

$$\chi^2(X,Z) \ = \sum_{k=1}^{\infty} \frac{1}{k!} \big(\mathbb{E}\, H_k(X)\big)^2.$$

Conversely, if X has all moments and the series is convergent, then $\chi^2(X,Z) < \infty$. In particular, X has density.

CLT for strong metrics

$X, X_1, X_2, \dots$ i.i.d. random variables, $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$,

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

CLT: as $n \to \infty$

$$F_n(x) = \mathbb{P}\{Z_n \le x\} \to \Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy.$$

Total variation distance. Prokhorov (1952):

$$\|F_n - \Phi\|_{\mathrm{TV}} \to 0 \iff \|F_{n_0} - \Phi\|_{\mathrm{TV}} < 2,$$

F_n has an absolutely continuous component for some $n = n_0$ (in particular, if X has density).

Kullback-Leibler distance (relative entropy). Barron (1986):

$$D(Z_n||Z) \to 0 \iff D(Z_{n_0}||Z) < \infty$$

for some $n = n_0$. In particular, when X has density p such that $\int_{-\infty}^{\infty} p(x) \log p(x)\, dx < \infty$.

Rates, Berry-Esseen bounds, the non-i.i.d. case: Linnik (1959), Sirazhdinov, Mamatov (1962), Artstein, Ball, Barthe, Naor (2004), Barron, Johnson (2004), B-C-G (2013-2016), Toscani (2016), Bally, Caramellino (2016).

CLT for χ^2 distance

Fomin (1982): Suppose that X has a compactly supported, symmetric, piecewise differentiable density p such that the coefficients in

$$p(x) \,=\, \varphi(x) \sum_{k=1}^{\infty} \frac{\sigma_k}{2^k k!} \, H_{2k}(x)$$

satisfy $\sup_{k \geq 2} \sigma_k < 1$. Then $\chi^2(Z_n, Z) = O(\frac{1}{n})$ as $n \rightarrow \infty$.

Example: Uniform distribution on $(-\sqrt{3}, \sqrt{3})$.

Theorem 1. $\chi^2(Z_n, Z) \rightarrow 0$, if and only if $\chi^2(Z_n, Z) < \infty$ for some $n = n_0$, and

$$\mathbb{E}\, e^{tX} < e^{t^2} \qquad \text{for all } t \neq 0.$$

Remark. It is possible that $\mathbb{E}\, e^{tX} < e^{t^2}$ for all $t \neq 0$ except for one $t_0 > 0$. Consider $X = a\xi + bZ$ assuming that ξ takes values q and $-p$ with probabilities p and q such that

$$\frac{p - q}{\log p - \log q} > pq.$$

Edgeworth-type expansion

If $\chi^2(Z_n, Z) \to 0$, then as $n \to \infty$

$$\chi^2(Z_n, Z) \;=\; \sum_{j=1}^{s-2} \frac{c_j}{n^j} + O\Big(\frac{1}{n^{s-1}}\Big)$$

for every fixed $s = 3, 4, \dots$ with c_j certain polynomials in the moments $\alpha_k = \mathbb{E}X^k$, $k = 3, \dots, j+2$.

Case $s = 3$:

$$\chi^2(Z_n, Z) \;=\; \frac{\alpha_3^2}{6n} + O\Big(\frac{1}{n^2}\Big),$$

Case $\alpha_3 = 0$, $s = 4$:

$$\chi^2(Z_n, Z) \;=\; \frac{(\alpha_4 - 3)^2}{24\,n^2} + O\Big(\frac{1}{n^3}\Big).$$

CLT for Renyi divergence

$X, X_1, X_2, \dots$ i.i.d. random vectors in $\mathbb{R}^d$, with $\mathbb{E}X = 0$ and identity covariance.

Denote by $\alpha^* = \frac{\alpha}{\alpha-1}$ the conjugate index for $\alpha > 1$, and by Z a standard normal random vector in $\mathbb{R}^d$.

Theorem 2. $D_\alpha(Z_n||Z) \to 0$ if and only if $D_\alpha(Z_n||Z) < \infty$ for some $n = n_0$, and

$$\mathbb{E}\, e^{\langle t,X\rangle} < e^{\alpha^*|t|^2/2} \quad \text{for all } \; t \in \mathbb{R}^d, \; t \neq 0.$$

In this case,

$$D_\alpha(Z_n||Z) = O(1/n),$$

and even $D_\alpha(Z_n||Z) = O(1/n^2)$, if the distribution of X is symmetric.

In fact, all distances have similar rates

$$D_\alpha(Z_n||Z) \sim T_\alpha(Z_n||Z) \sim \frac{\alpha}{2}\,\chi^2(Z_n, Z)$$

once they tend to zero.

Examples

• Uniform distribution.
Let $X \sim (-\sqrt{3}, \sqrt{3})$. It has Laplace transform

$$\mathbb{E}\, e^{tX} = \frac{\sinh(t\sqrt{3})}{t\sqrt{3}} < e^{t^2/2}, \qquad t \in \mathbb{R} \ \ (t \neq 0),$$

with first moments $\alpha_2 = 1$, $\alpha_3 = 0$, $\alpha_4 = \frac{9}{5}$. Therefore,

$$\chi^2(Z_n, Z) \ = \ \frac{3}{50\, n^2} + O\Big(\frac{1}{n^3}\Big),$$
$$D_\alpha(Z_n||Z) \ = \ \frac{\alpha}{2}\, \chi^2(Z_n, Z) + O\Big(\frac{1}{n^3}\Big).$$

• Log-concave probability distributions on $\mathbb{R}^d$.
Let X have density $p(x) = e^{-V(x)}$ with mean zero, identity covariance and such that $V''(x) \geq c\, \mathrm{I}$ for some $c > 0$ (Bakry-Emery criterion, necessarily $c \leq 1$). Then

$$\mathbb{E}\, e^{tg(X)} \leq e^{t^2/(2c)}, \qquad t \in \mathbb{R}.$$

for any g on $\mathbb{R}^d$ such that $\|g\|_{\mathrm{Lip}} \leq 1$ and $\mathbb{E}\, g(X) = 0$. Hence

$$D_\alpha(Z_n||Z) \to 0 \text{ as } n \to \infty, \quad \text{whenever} \ \ \alpha < \frac{1}{1-c}.$$

.

Necessity part in Theorem 2 (preparation)

The characteristic function $f(t) = \mathbb{E}\, e^{itX}$ is entire on $\mathbb{C}$, and $f(iy) = \mathbb{E}\, e^{-yX}$.

Lemma 1. If $\lim D_\alpha(Z_n||Z)\big] = 0$, then for all $y \in \mathbb{R}$

$$f(iy) \le e^{\beta y^2/2}$$

and for any integer $k \ge \alpha/2$,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f\big(iy/\sqrt{kn}\big)^{2kn}\, e^{-\beta y^2}\, dy \;=\; \sqrt{\pi(\alpha-1)}.$$

Proof. By Proposition 1 applied to Z_n in place of X,

$$f(iy/\sqrt{n})^n \;=\; \mathbb{E}\, e^{-yZ_n} \;\le\; C_n\, e^{\beta y^2/2}$$

with $C_n = (1+(\alpha-1)\, T_\alpha(Z_n||Z))^{1/\alpha}$. After change of variable

$$f(iy) \;\le\; C_n^{1/n}\, e^{\beta y^2/2} \;\to\; e^{\beta y^2/2}.$$

Since $f(\frac{t}{\sqrt{n}})^n$ is the ch.f. of Z_n, the integral in Lemma is

$$\int_{-\infty}^{\infty} \big(\mathbb{E}\, e^{-yZ_{nk}}\big)^2\, e^{-\beta y^2}\, dy \;=\; \int_{-\infty}^{\infty} \mathbb{E}\, e^{-\sqrt{2}\, y\, Z_{2nk}}\, e^{-\beta y^2}\, dy$$

$$= \sqrt{\frac{\pi}{\beta}}\, \mathbb{E}\, e^{\frac{1}{2\beta}\, Z_{2nk}^2} \;\to\; \sqrt{\frac{\pi}{\beta}}\, \mathbb{E}\, e^{\frac{1}{2\beta}\, Z^2},$$

by Proposition 2 on last step.

Necessity part in Theorem 2

If $D_\alpha(Z_n||Z) \to 0$, then, by Lemma 1,

$$\psi(y) = f(iy)\, e^{-\beta y^2/2} \le 1 \quad \text{for all} \;\; y \in \mathbb{R}.$$

Need to show:

$$\psi(y) < 1 \quad \text{for all} \;\; y \neq 0.$$

Fix $k \ge \alpha/2$ and $\delta > 0$ (small), and decompose

$$\int_{-\infty}^{\infty} f\big(iy/\sqrt{nk}\big)^{2nk} e^{-\beta y^2}\, dy \;=\; I_1 + I_2$$

$$= \left(\int_{|y| \le \delta\sqrt{nk}} + \int_{|y| > \delta\sqrt{nk}} \right) f\big(iy/\sqrt{nk}\big)^{2nk} e^{-\beta y^2}\, dy. \quad (1)$$

Write

$$g(t) = \log f(t) = -\frac{1}{2}\, t^2 + \sum_{m=3}^{\infty} a_m t^m.$$

Since $\sum_{m=3}^{\infty} |a_m t^m| \le c|t|^3$ for $|t| \le r$, $r > 0$ small, so,

$$f\big(iy/\sqrt{nk}\big)^{2nk} = \exp\{y^2 + \theta y^3/\sqrt{n}\}, \quad y \in \big[-r\sqrt{nk}, r\sqrt{nk}\,\big],$$

where $|\theta| \le c$. Assuming $\delta \le \min\{r, (\beta - 1)/(2c\sqrt{k})\}$,

$$I_1 = \int_{|y| \le \delta\sqrt{nk}} e^{-(\beta-1)\, y^2 + \theta y^3/\sqrt{n}}\, dy.$$

Here $\theta y^3/\sqrt{n}$ may be removed at the expense of $O(\frac{1}{\sqrt{n}})$. Hence

$$I_1 = \int_{|y|\le\delta\sqrt{nk}} e^{-(\beta-1)\,y^2} dy + O\Big(\frac{1}{\sqrt{n}}\Big) = \sqrt{\pi(\alpha-1)} + O\Big(\frac{1}{\sqrt{n}}\Big).$$

Applying this in (1), we have $I_2 \to 0$, or equivalently

$$\int_{|u|>\delta} \psi(u)^{2nk}\, du = \int_{|u|>\delta} \big(f(iu)\, e^{-\beta u^2/2}\big)^{2nk}\, du = o\Big(\frac{1}{\sqrt{n}}\Big)$$

for any sufficiently small $\delta > 0$ and hence for any $\delta > 0$.

Assume $\psi(u_0) = 1$ for some $u_0 > 0$, which implies $\psi'(u_0) = 0$. Hence the power series representation at this point

$$\psi(u) - 1 = c_l(u-u_0)^l + \sum_{j=l+1}^{\infty} c_j(u-u_0)^j$$

starts with $c_l \neq 0$ for some $l \geq 2$. Since $\psi(u) \leq 1$ for all $u \in \mathbb{R}$, necessarily $l = 2m$ and $c_l < 0$. Hence,

$$\psi(u) \geq 1 - b_1(u-u_0)^{2m} \geq e^{-b_0(u-u_0)^{2m}}.$$

for $|u - u_0| \leq r_0 < u_0$ with some constants $b_1, b_0 > 0$. Choosing $\delta = u_0 - r_0$, this neighborhood is contained in (δ, ∞), and

$$\int_{|u|>\delta} \psi(u)^{2nk}\, du \ \geq \ \int_{|u-u_0|<\delta} \exp\big\{-2nk\cdot c_0(u-u_0)^{2m}\big\}\, du$$

$$= \ 2\int_0^{\delta} \exp\big\{-2nk\cdot c_0 x^{2m}\big\}\, dx \ \geq \ \frac{c}{n^{1/(2m)}}.$$

Pointwise upper bounds on densities

The following observation holds without assuming that X has mean zero and variance one. Let f be the characteristic function of X, and define

$$\psi(u) \,=\, f(iu)\, e^{-\beta u^2/2} \,=\, \mathbb{E}\, e^{-uX}\, e^{-\beta u^2/2}, \qquad u \in \mathbb{R},$$

where $\beta = \frac{\alpha}{\alpha-1}$.

Question: If $T_\alpha(X||Z) < \infty$, can we bound the density $p(x)$ of X pointwise? Answer: No. However, assume $T_\alpha(Z_{n_0}||Z) < \infty$.

Lemma 2. For any $n \geq n_\beta = \max(\beta, 2)\, n_0$, the normalized sum Z_n has a continuous bounded density p_n satisfying

$$p_n(x) \,\leq\, \frac{A_\alpha \sqrt{n}}{\sqrt{2\pi n_0}}\, e^{-x^2/(2\beta)}\, \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta}, \qquad x \in \mathbb{R}.$$

In particular, there exist $x_0 > 0$ and $\delta \in (0,1)$ such that, for all n large enough,

$$p_n(x) \,\leq\, \delta^n e^{-x^2/(2\beta)}\, \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n/2}, \qquad |x| \geq x_0\sqrt{n}.$$

Proof. Use contour integration.

Proof of Lemma 2

Let $\alpha = 2$, $n_0 = 1$, so that f_n integrable for $n \geq 2$ and

$$p_n(x) = e^{yx}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f((t+iy)/\sqrt{n})^n\,dt$$

for any fixed $y > 0$. Let $x < 0$. Using $|f(t+iy)| \leq f(iy)$ and changing variable,

$$p_n(x) \;\leq\; e^{yx} f\Big(\frac{iy}{\sqrt{n}}\Big)^{n-2}\,\frac{\sqrt{n}}{2\pi}\int_{-\infty}^{\infty}\Big|f\Big(t+\frac{iy}{\sqrt{n}}\Big)\Big|^2\,dt.$$

The function $t \to f(t+iy/\sqrt{n}) = \mathbb{E}\,e^{itX - yX/\sqrt{n}}$ is the Fourier transform of $e^{-yu/\sqrt{n}}\,p(u)$ and

$$e^{-2yu/\sqrt{n}}\,p(u)^2 \;=\; \big(e^{-2yu/\sqrt{n}}\,\varphi(u)\big)\,\frac{p(u)^2}{\varphi(u)} \;\leq\; \frac{1}{\sqrt{2\pi}}\,e^{\frac{2y^2}{n}}\,\frac{p(u)^2}{\varphi(u)}.$$

By Plancherel,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big|f\Big(t+\frac{iy}{\sqrt{n}}\Big)\Big|^2\,dt \;\leq\; \frac{1}{\sqrt{2\pi}}\,e^{2y^2/n}\,(1+\chi^2).$$

Hence

$$\begin{aligned} p_n(x) \;&\leq\; \sqrt{\frac{n}{2\pi}}\,(1+\chi^2)\;e^{yx+2y^2/n}\,f(iy/\sqrt{n})^{n-2} \\ &=\; \sqrt{\frac{n}{2\pi}}\,(1+\chi^2)\;e^{yx+y^2}\,\psi(y/\sqrt{n})^{n-2}. \end{aligned}$$

Choose $y = -x/2$.

Sufficiency part in Theorem 2

Let $X, X_1, X_2, \dots$ be i.i.d., $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, with ch.f. $f(t) = \mathbb{E}\, e^{itX}$. As before, put

$$\psi(u) = f(iu)\, e^{-\beta u^2/2}, \qquad \beta = \frac{\alpha}{\alpha - 1}, \quad Z \sim N(0,1).$$

Assuming that $\psi(u) < 1$ for all $u \neq 0$, we need to show that

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

satisfy $T_\alpha(Z_n||Z) \to 0$ as long as $T_\alpha(Z_{n_0}||Z) < \infty$ for some n_0. By Lemma 2, Z_n have densities p_n which are continuous and bounded whenever $n \geq n_\beta$.

Using Edgeworth expansions, the integrals

$$I_0 = \int_{|x| \leq M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx \quad \text{with} \quad M_n = \sqrt{2(s-1)\log n}$$

admit an asymptotic expansion in powers of $1/n$ up to $1/n^{s-1}$. So, it remains to bound the integral of $p_n^\alpha/\varphi^{\alpha-1}$ over $|x| > M_n$ by a polynomially small quantity. In fact, for any large enough $s \geq 3$ and some constant $\kappa > 0$,

$$\int_{|x| > M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\, dx \,=\, O\Big(\frac{1}{n^{\kappa s}}\Big), \qquad n \to \infty.$$

For definiteness, let $x < -M_n$, and define

$$I_1 = \int_{-\infty}^{-x_0\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\,dx, \quad I_2 = \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\,dx,$$

$$I_3 = \int_{-x_1\sqrt{n}}^{-M_n} \frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}}\,dx$$

with parameters $0 < x_1 < x_0$ and assuming that $M_n < x_1\sqrt{n}$ (otherwise, $I_3 = 0$).

By Lemma 2, for all large n, with some $\delta \in (0,1)$, $x_0 > 0$,

$$\begin{aligned} I_1 \ &\leq\ (2\pi)^{\frac{\alpha-1}{2}}\,\delta^{\alpha n} \int_{-\infty}^{-x_0\sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{\alpha n/2}\,dx \\ &\leq\ (2\pi)^{\frac{\alpha-1}{2}}\,\delta^{\alpha n}\beta\sqrt{n} \int_{-\infty}^{\infty} \psi(u)^m\,du, \qquad m \leq \frac{\alpha n}{2}, \end{aligned}$$

where on the last step we used $\psi \leq 1$. The last integral is convergent whenever $m = kn_0$, $k \geq \alpha$. Hence

$$I_1 \ \leq\ C\delta_1^n \qquad (n \geq n_1)$$

with some constants $C > 0$, $x_0 > 0$ and $\delta < \delta_1 < 1$, depending on the density p only.

By assumption, $\delta_2 = \max_{-x_0 \leq u \leq -x_1} \psi(u) < 1$, and Lemma 2 yields

$$\begin{aligned} I_2 \ &\le \ A_\alpha n^{\alpha/2} \int_{-x_0\sqrt{n}}^{-x_1\sqrt{n}} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta}\, dx \\ &= \ A_\alpha^\alpha\, \beta\, n^{(\alpha+1)/2} \int_{-x_0/2}^{-x_1/2} \psi(u)^{n-n_\beta}\, du \\ &\le \ A_\alpha^\alpha\, \beta\, n^{(\alpha+1)/2}\, (x_0 - x_1)\, \delta_2^{n-n_\beta} \end{aligned}$$

which again decays exponentially fast like I_1.

Near zero, $h(u) = \log f(iu) \sim \frac{1}{2}\, u^2$, hence $|h(u)| \le \frac{1+\beta}{4}\, |u|^2$ in some disc $|u| \le r$, when r is sufficiently small, implying $|f(iu)| \le e^{(1+\beta)|u|^2/4}$. Hence

$$\psi(u) \le e^{-\frac{1}{4}\,(\beta-1)\,|u|^2}, \qquad |u| \le r,$$

and

$$\begin{aligned} \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n-n_\beta} \ &\le \ \psi\Big(-\frac{x}{\beta\sqrt{n}}\Big)^{n/2} \\ &\le \ \exp\Big\{-\frac{\beta-1}{4}\,\frac{x^2}{2\beta^2}\Big\} = e^{-x^2/(8\alpha\beta)} \end{aligned}$$

for all $n \ge 2n_\beta$ and $-\beta r\sqrt{n} < x < 0$. Therefore, By Lemma 2, in this interval

$$\frac{p_n(x)^\alpha}{\varphi(x)^{\alpha-1}} \le A_\alpha^\alpha\, n^{\alpha/2}\, e^{-x^2/(8\beta)},$$

which results with $x_1 = \beta r$ in the bound

$$\begin{aligned} I_3 &\le A_\alpha^\alpha \, n^{\alpha/2} \int_{-x_1\sqrt{n}}^{-M_n} e^{-x^2/(8\beta)} \, dx \\ &\le \sqrt{2\pi\beta} \, A_\alpha^\alpha \, n^{\alpha/2} \, e^{-M_n^2/(8\beta)} = \sqrt{2\pi\beta} \, A_\alpha^\alpha \, n^{-(\frac{s-1}{4\beta} - \frac{\alpha}{2})}. \end{aligned}$$

Collecting these bounds, we obtain that $I_1 + I_2 + I_3 = o(n^{-s/8\beta})$ for a sufficiently large s.