Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2009, Vol. 45, No. 3, 802–817
DOI: 10.1214/08-AIHP187
© Association des Publications de l'Institut Henri Poincaré, 2009

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Upper bounds for minimal distances in the central limit theorem

Emmanuel Rio

UMR 8100 CNRS, Université de Versailles Saint Quentin en Y, Bât. Fermat, 45 Avenue des Etats Unis, 78035 Versailles Cedex, France
E-mail: rio@math.uvsq.fr

Received 23 May 2007; revised 18 June 2008; accepted 22 August 2008

Abstract. We obtain upper bounds for minimal metrics in the central limit theorem for sequences of independent real-valued random variables.

Résumé. Nous obtenons des majorations des distances minimales dans le théorème limite central pour les suites de variables aléatoires réelles indépendantes.

MSC: 60F05

Keywords: Fréchet–Dall'Aglio minimal metric; Wasserstein distance; Rates of convergence; Esseen's mean central limit theorem; Global central limit theorem

1. Introduction

Let Ω be a probability space, rich enough to generate the set of probability laws on $\mathbb{R} \times \mathbb{R}$. Let d be a pseudodistance on the set of real-valued random variables, such that $d(X,Y)$ depends only on the law of (X,Y). Then, according to Paul Lévy (see Note B in Fréchet (1950) for this fact) the minimal distance $\hat{d}$ associated with d is defined by

$$\hat{d}(\mu,\nu) = \inf\{d(X,Y):\ X \sim \mu, Y \sim \nu\},$$

where the infimum is taken over all random vectors (X,Y) with respective marginal laws μ and ν. We refer to Zolotarev (1976) for the properties of minimal distances. When $E = \mathbb{R}$, $r \geq 1$ and $d(X,Y) = \|X - Y\|_r$, we denote by W_r the so defined minimal distance on the space $\mathcal{M}_r$ of probability laws with a finite absolute moment of order r. This distance is often called Wasserstein distance of order r. More generally, if ψ is some Orlicz function and $\|\cdot\|_\psi$ is the Orlicz norm associated with this function, we will denote by W_ψ the minimal distance associated with $d(X,Y) = \|X - Y\|_\psi$. The distances W_r and W_ψ are homogeneous of degree 1.

Our aim in this paper is to provide upper bounds for the minimal distances between the normalized sum and the limiting Gaussian distribution in the independent case. We now recall the known results on this subject.

Throughout the paper, $X_1, X_2, \ldots$ is a sequence of independent real-valued random variables with mean zero and finite positive variance. We set $S_n = X_1 + X_2 + \cdots + X_n$ and $v_n = \operatorname{Var} S_n$. We denote by μ_n the law of $v_n^{-1/2} S_n$ and by γ_v the normal law with mean 0 and variance v. For independent and identically distributed (i.i.d.) random variables with finite absolute third moment, Esseen (1958) proved that

$$\lim_{n\to\infty} \sqrt{n}\, W_1(\mu_n, \gamma_1) = A_1(\mu_1), \tag{1.1}$$

where $A_1(\mu_1)$ is some non-negative explicit constant depending only on the law of X_1. Consequently the rate in the central limit for the distance W_1 is $\mathrm{O}(n^{-1/2})$. Next Zolotarev (1964) obtained

$$A_1(\mu_1) \le \frac{1}{2}\frac{\mathbb{E}(|X_1|^3)}{(\mathbb{E}(X_1^2))^{3/2}}. \tag{1.2}$$

He also proved that the constant $(1/2)$ in this inequality cannot be improved. For p in $]2,3]$ and i.i.d. random variables in $\mathbb{L}^p$, Ibragimov (1966) proved that

$$W_1(\mu_n,\gamma_1) = \mathrm{O}\big(n^{1-p/2}\big) \tag{1.3a}$$

and that this rate cannot be improved. For independent and non-identically distributed random variables, it follows from the non-uniform estimates of Bikelis (1966) that

$$W_1(\mu_n,\gamma_1) \le C v_n^{-p/2} \sum_{i=1}^{n} \mathbb{E}\big(|X_i|^p\big) \tag{1.3b}$$

for some constant C depending only on p, which generalizes the results of Ibragimov (1966).

Later Bártfai (1970) proved that $W_r(\mu_n,\gamma_1) = \mathrm{O}(n^{\varepsilon-1/2})$ for i.i.d. random variables with a finite moment generating function, for any $r \ge 1$. He also conjectured that

$$W_r(\mu_n,\gamma_1) = \mathrm{O}\big(n^{-1/2}\big) \quad \text{as } n \to \infty. \tag{1.4}$$

For $r > 2$ and i.i.d. random variables with a finite absolute moment of order r,

$$W_r(\mu_n,\gamma_1) = \mathrm{O}\big(n^{(1/r)-(1/2)}\big) \quad \text{as } n \to \infty, \tag{1.5}$$

as proved by Sakhanenko (1985). Furthermore this rate cannot be improved. Consequently the optimal rate needs a more stringent moment condition.

We now give an outline of our results. In Section 2, we give an answer to Bártfai's question. More precisely, we prove that, for i.i.d. random variables with a finite moment generating function in a neighborhood of the origin,

$$W_\psi(\mu_n,\gamma_1) = \mathrm{O}\big(n^{-1/2}\big) \quad \text{as } n \to \infty, \tag{1.6}$$

where ψ is the Orlicz function $\psi(x) = \exp(|x|) - 1$. This result implies that (1.4) holds for i.i.d. random variables with a finite moment generating function, for any $r \ge 1$.

In Section 4, we give more precise results for r in $[1,2]$ and sequences of independent random variables. In particular, our results imply that there exists some positive constant C depending only on r such that, for i.i.d. random variables in $\mathbb{L}^{r+2}$,

$$W_r(\mu_n,\gamma_1) \le C\big(\|X_1\|_{r+2}/\|X_1\|_2\big)^{1+2/r} n^{-1/2} \quad \text{for any positive } n. \tag{1.7}$$

The proof of (1.7) is mainly based on some functional inequality linking the Wasserstein distance of order r and Zolotarev's ideal distance of order r. This result was announced in Rio (1998). However, the proof was not correct (the constant appearing in his Theorem 1 should be modified). This is the reason why we give a new proof of this functional inequality in Section 3. In Section 5, we prove that inequality (1.7) cannot be improved. In Section 6, we extend the results of Section 4 to more general transportation costs. For example, for i.i.d. random variables with finite fourth moments, the results of Section 6 ensure that

$$W_2(P_{S_n^2/v_n}, P_{Y^2}) \le C\big(\|X_1\|_4/\|X_1\|_2\big)^2 n^{-1/2}, \tag{1.8}$$

where Y is a standard normal, which cannot be derived from the results of Section 4.

2. Minimal distances in the CLT for r.v.'s with finite Laplace transform

Throughout Section 2, $X_1, X_2, \ldots$ is a sequence of i.i.d. real-valued random variables satisfying the conditions

$$\mathbb{E}(X_1) = 0 \quad \text{and} \quad \mathbb{E}\bigl(\exp\bigl(t|X_1|\bigr)\bigr) < \infty \quad \text{for some } t > 0. \tag{2.1}$$

Let ψ be the convex function defined by $\psi(x) = \exp(|x|) - 1$. The Orlicz norm associated with ψ is defined by

$$\|X\|_\psi = \inf\bigl\{a > 0 \text{ such that } \mathbb{E}\bigl(\psi(X/a)\bigr) \le 1\bigr\}, \tag{2.2}$$

with the convention $\|X\|_\psi = +\infty$ if $\mathbb{E}(\psi(X/a)) > 1$ for any positive a.

Theorem 2.1 below proves that the Wasserstein distances $W_\psi(\mu_n, \gamma_1)$ converge to 0 at the rate $n^{-1/2}$ for sums of i.i.d. r.v.'s with finite Laplace transforms. Since $\lim_\infty x^{-r}\psi(x) = +\infty$ for any $r \ge 1$, it follows that (1.4) holds true for any $r \ge 1$ under condition (2.1). For random variables with a finite moment generating function, this result was stated by Rio (1993), pp. 785–786, in the case of random vectors with smooth density.

Theorem 2.1. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s satisfying* (2.1). *Then there exists some positive A such that $\sqrt{n}W_\psi(\mu_n, \gamma_{v_n}) \le A$ for any positive integer n.*

Proof. Let F_n denote the distribution function of S_n, F_n^{-1} denote the generalized inverse function of F_n and Φ denote the distribution function of a standard normal r.v. Y. Then $U_n = F_n^{-1}(\Phi(Y))$ has the same distribution as S_n and, by Theorem 8.1 in Major (1978),

$$\sqrt{v_n}W_\psi(\mu_n, \gamma_1) = \bigl\|U_n - \sqrt{v_n}Y\bigr\|_\psi.$$

It remains to prove that the sequence $(U_n - \sqrt{v_n}Y)_n$ is bounded in the Orlicz space $\mathbb{L}^\psi$ associated to ψ.

The idea of such a result goes back to Komlós, Major and Tusnády (1975), pp. 114–115. From a normal approximation theorem in Petrov (1975), which is Theorem A, Section 6, in Komlós, Major and Tusnády (1975), and using exactly the same arguments as in the proof of inequality (2.6), p. 118 of the same paper (this inequality does not need a Cramér type condition on X_1), one obtains that there are positive constants C and c such that

$$\bigl|U_n - \sqrt{v_n}Y\bigr| \le C\bigl(1 + n^{-1}U_n^2\bigr) \quad \text{if } |U_n| \le cn. \tag{2.3}$$

It follows that, for any positive constant $c' \le c$,

$$\bigl|U_n - \sqrt{v_n}Y\bigr| \le C + CU_n^2\mathbb{1}_{|U_n|\le c'n} + |U_n|\mathbb{1}_{|U_n|>c'n} + \sqrt{v_n}|Y|\mathbb{1}_{|U_n|>c'n}.$$

Hence

$$\bigl\|U_n - \sqrt{v_n}Y\bigr\|_\psi \le (C/\log 2) + C\bigl\|n^{-1}U_n^2\mathbb{1}_{|U_n|\le cn}\bigr\|_\psi + \|U_n\mathbb{1}_{|U_n|>c'n}\|_\psi + \sqrt{v_n}\|Y\mathbb{1}_{|U_n|>c'n}\|_\psi.$$

Now, from Lemma 4(ii), p. 276, in Massart (1989), applied to the random variables X_i with $w_i = 1$, there exist some positive constants c' and C' such that

$$\bigl\|n^{-1}U_n^2\mathbb{1}_{|U_n|\le c'n}\bigr\|_\psi \le C'. \tag{2.4}$$

Clearly, we may assume that $c' \le c$, reducing c' if necessary. To bound up the last terms on right-hand side in (2.3), it will be convenient to apply the elementary lemma below.

Lemma 2.1. *For any real-valued r.v. Z with mean* 0 *and any measurable set Γ*,

$$\mathbb{E}\bigl(\exp\bigl(a|Z|\mathbb{1}_\Gamma\bigr) - 1\bigr) \le \bigl(\mathbb{P}(\Gamma)\mathbb{E}\bigl(\mathrm{e}^{2aZ}\bigr)\mathbb{E}\bigl(\mathrm{e}^{-2aZ}\bigr)\bigr)^{1/2}.$$

Proof. From the elementary equality $\exp(a|Z|\mathbb{1}_\Gamma) - 1 = \mathbb{1}_\Gamma(\mathrm{e}^{a|Z|} - 1)$ together with the Schwarz inequality, we get that

$$\mathbb{E}\bigl(\exp\bigl(a|Z|\mathbb{1}_\Gamma\bigr) - 1\bigr) \le \bigl(\mathbb{P}(\Gamma)\mathbb{E}\bigl(\bigl(\mathrm{e}^{a|Z|} - 1\bigr)^2\bigr)\bigr)^{1/2}.$$

Now $(\mathrm{e}^{|u|} - 1)^2 \le (\mathrm{e}^u - \mathrm{e}^{-u})^2$, which ensures that

$$\mathbb{E}\bigl(\bigl(\mathrm{e}^{a|Z|} - 1\bigr)^2\bigr) \le \mathbb{E}\bigl(\mathrm{e}^{2aZ}\bigr) + \mathbb{E}\bigl(\mathrm{e}^{-2aZ}\bigr) - 2.$$

To complete the proof, apply the inequality $x + y - 2 \le xy - 1$ (valid for $x \ge 1$ and $y \ge 1$) to $x = \mathbb{E}(\mathrm{e}^{2aZ})$ and $y = \mathbb{E}(\mathrm{e}^{-2aZ})$ (observe that $x \ge 1$ and $y \ge 1$ since $\mathbb{E}(Z) = 0$). □

We now apply Lemma 2.1 to $\Gamma = (|U_n| \ge c'n)$. From the Chernoff bound for sums of i.i.d. random variables with finite Laplace transform, $\mathbb{P}(\Gamma) \le 2\exp(-nb)$ for some positive constant b. Now there exists some positive constant a such that

$$\mathbb{E}\bigl(\exp(2aU_n)\bigr)\mathbb{E}\bigl(\exp(-2aU_n)\bigr) \le \exp(nb) \quad \text{and} \quad \exp\bigl(4a^2 v_n\bigr) \le \exp(nb).$$

For this choice of a, and either $Z = U_n$ or $Z = \sqrt{v_n}Y$, Lemma 2.1 ensures that

$$\mathbb{E}\bigl(\exp\bigl(a|Z|\mathbb{1}_\Gamma\bigr) - 1\bigr) \le \sqrt{2} \le 3/2,$$

which implies that

$$\|U_n \mathbb{1}_{|U_n|>c'n}\|_\psi + \sqrt{v_n}\|Y\mathbb{1}_{|U_n|>c'n}\|_\psi \le 3a.$$

Hence Theorem 2.1 holds. □

We now give an application of Theorem 2.1 to the minimal distances between the Poisson distribution and the Gaussian distribution with the same expectation and the same variance. This corollary will be used to prove Theorem 4.1 in Section 4.

Corollary 2.2. *Let $\mathcal{P}(\lambda)$ denote the Poisson law with mean λ and $\mathcal{N}(m, v)$ denote the Gaussian law with mean m and variance v. Then* $\sup_{\lambda\ge 0} W_\psi(\mathcal{P}(\lambda), \mathcal{N}(\lambda, \lambda)) < \infty$.

Proof. Let $n = [\lambda]$. By the triangle inequality,

$$\begin{aligned} W_\psi\bigl(\mathcal{P}(\lambda), \mathcal{N}(\lambda, \lambda)\bigr) &\le W_\psi\bigl(\mathcal{P}(n), \mathcal{N}(n, n)\bigr) + W_\psi\bigl(\mathcal{P}(n), \mathcal{P}(\lambda)\bigr) + W_\psi\bigl(\mathcal{N}(n, n), \mathcal{N}(\lambda, \lambda)\bigr) \\ &\le A + W_\psi\bigl(\mathcal{P}(n), \mathcal{P}(\lambda)\bigr) + W_\psi\bigl(\mathcal{N}(n, n), \mathcal{N}(\lambda, \lambda)\bigr), \end{aligned} \tag{2.5}$$

by Theorem 2.1 applied to i.i.d. random variables with common law $\mathcal{P}(1)$.

Set $x = \lambda - n$. Then $\mathcal{P}(\lambda) = \mathcal{P}(n) * \mathcal{P}(x)$ and $\mathcal{N}(\lambda, \lambda) = \mathcal{N}(n, n) * \mathcal{N}(x, x)$. Hence, by Lemma 3, p. 382, in Zolotarev (1976),

$$W_\psi\bigl(\mathcal{P}(n), \mathcal{P}(\lambda)\bigr) \le W_\psi\bigl(\delta_0, \mathcal{P}(x)\bigr) \quad \text{and} \quad W_\psi\bigl(\mathcal{N}(n, n), \mathcal{N}(\lambda, \lambda)\bigr) \le W_\psi\bigl(\delta_0, \mathcal{N}(x, x)\bigr).$$

Let N_x be an r.v. with distribution $\mathcal{P}(x)$ and B_x be an r.v. with distribution $\mathcal{N}(0, x)$:

$$W_\psi\bigl(\delta_0, \mathcal{P}(x)\bigr) = \|N_x\|_\psi \quad \text{and} \quad W_\psi\bigl(\delta_0, \mathcal{N}(x, x)\bigr) = \|x + B_x\|_\psi \le x\|1\|_\psi + \sqrt{x}\|B_1\|_\psi.$$

Now

$$\mathbb{E}\bigl(\psi(N_x/a)\bigr) = \exp\bigl(x\bigl(\exp(1/a) - 1\bigr)\bigr) - 1 \le \mathbb{E}\bigl(\psi(N_1/a)\bigr),$$

which implies that $\|N_x\|_\psi \le \|N_1\|_\psi$. Hence, from the above facts,

$$W_\psi\big(\mathcal{P}(n),\mathcal{P}(\lambda)\big)+W_\psi\big(\mathcal{N}(n,n),\mathcal{N}(\lambda,\lambda)\big)\le \|N_1\|_\psi+(1/\log 2)+\|B_1\|_\psi. \tag{2.6}$$

Both (2.5) and (2.6) imply Corollary 2.2. □

3. A functional inequality for Wasserstein distances and Zolotarev distances

The main result of this section provides a comparison between the Wasserstein distance W_r and Zolotarev's ideal distance of order r.

Definition 3.1. *For $r > 1$, let $l = \sup \mathbb{N} \cap [0, r[$ and Λ_r be the class of l-times continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that $|f^{(l)}(x) - f^{(l)}(y)| \le |x - y|^{r-l}$ for any (x, y) in $\mathbb{R}^2$. The ideal distance Z_r of Zolotarev is defined by*

$$Z_r(\mu,\nu)=\sup\left\{\int_{\mathbb{R}} f\,\mathrm{d}\mu-\int_{\mathbb{R}} f\,\mathrm{d}\nu:\ f\in\Lambda_r\right\}.$$

Theorem 3.1 below is Theorem 1 in Rio (1998) with different constants.

Theorem 3.1. *For any $r > 1$ there exists a positive constant c_r such that, for any pair (μ, ν) of laws on the real line with finite absolute moments of order r,*

$$\big(W_r(\mu,\nu)\big)^r\le c_r Z_r(\mu,\nu).$$

Furthermore $c_r = 2r$ for r in $]1, 2]$ and $c_r = (r - 1)2^{r+1}$ for r in $]2, 3]$.

For large and entire values of r, Proposition 3.1 below provides estimates of c_r.

Proposition 3.1. *For $r \ge 4$ integer and $l = r - 1$, Theorem 3.1 holds true with*

$$c_r=\sqrt{\pi l}\big(\pi^2/\mathrm{e}\big)^{l/2}l^l\big(1+(1/l)\big)\quad \text{for } r \text{ even}\quad\text{and}\quad c_r=\sqrt{\pi l}\big(\pi^2/\mathrm{e}\big)^{l/2}l^l\exp\big(1/(6l)\big)\quad\text{for } r \text{ odd}.$$

Remark 3.1. *From Proposition 3.1, for any integer $r \ge 4$,*

$$W_r(\mu,\nu)\le \pi\mathrm{e}^{-1/2}r\big(Z_r(\mu,\nu)\big)^{1/r}.$$

Proof of Theorem 3.1. Let ϕ_σ denote the density of the normal law $N(0, \sigma^2)$ and $\mu_\sigma = \mu * \phi_\sigma.\lambda$. Clearly

$$W_r(\mu,\nu)=\lim_{\sigma\searrow 0}W_r(\mu_\sigma,\nu_\sigma)\quad\text{and}\quad Z_r(\mu,\nu)=\lim_{\sigma\searrow 0}Z_r(\mu_\sigma,\nu_\sigma).$$

Consequently it is enough to prove Theorem 3.1 for probability laws with strictly positive and smooth densities. So, let μ and ν be probability laws with distributions function F and G, respectively, and assume that F and G are C^∞ diffeomorphisms from $\mathbb{R}$ on $]0, 1[$. Recall that, for U r.v. with the uniform distribution over $[0, 1]$, the random vector $(F^{-1}(U), G^{-1}(U))$ has respective marginal laws μ and ν. Consequently, in order to prove Theorem 3.1, it is enough to prove that there exists some function h in Λ_r such that, for any u in $]0, 1[$,

$$c_r\big(h\big(G^{-1}(u)\big)-h\big(F^{-1}(u)\big)\big)\ge\big|G^{-1}(u)-F^{-1}(u)\big|^r. \tag{3.1}$$

Let A be the closed set of reals x such that $F(x) = G(x)$. By definition, $F(A) = G(A)$ and $F(A)$ is the set of reals u in $]0, 1[$ such that $F^{-1}(u) = G^{-1}(u)$. Throughout the proof, S is the sign function: $S(t) = \mathbb{1}_{t>0} - \mathbb{1}_{t<0}$.

If A is empty, then $F(x) > G(x)$ for any real x or $F(x) < G(x)$ for any real x. In that case, we choose

$$h(x) = 2^{r-l-1} \frac{S(F-G)x|x|^{r-1}}{r(r-1)\cdots(r-l+1)}. \tag{3.2}$$

Then $h^{(l)}(x) = 2^{r-l-1}(S(x))^{l-1}|x|^{r-l}S(F-G)$ and consequently h belongs to Λ_r. By definition of h,

$$\big(h\big(G^{-1}(u)\big) - h\big(F^{-1}(u)\big)\big) = 2^{r-1-l} \frac{|G^{-1}(u)|G^{-1}(u)|^{r-1} - F^{-1}(u)|F^{-1}(u)|^{r-1}|}{r(r-1)\cdots(r-l+1)}. \tag{3.3}$$

Now, for any $r > 1$ and any positive a, the function $(a+u)|a+u|^{r-1} - u|u|^{r-1}$ achieves its infimum for $u = -a/2$. Hence

$$\big|G^{-1}(u)\big|G^{-1}(u)\big|^{r-1} - F^{-1}(u)\big|F^{-1}(u)\big|^{r-1}\big| \geq 2^{1-r}\big|G^{-1}(u) - F^{-1}(u)\big|^{r},$$

which, together with (3.3), implies (3.1) with $c_r = 2^l r(r-1)\cdots(r-l+1)$.

Suppose now that A is not empty. Let $m = \inf A$ and $M = \sup A$. If M is finite, we define h on $[M, \infty[$ by

$$r(r-1)\cdots(r-l+1)\big(h(x) - h(M)\big) = 2^{r-l-1}S\big(F(x) - G(x)\big)(x-M)^r. \tag{3.4}$$

In the same way, if m is finite, we define h on $]-\infty, m]$ by

$$r(r-1)\cdots(r-l+1)\big(h(x) - h(m)\big) = -2^{r-l-1}S\big(F(x) - G(x)\big)(m-x)^r. \tag{3.5}$$

If $u > F(M)$, then $F^{-1}(u) > M$ and $G^{-1}(u) > M$. Now the sign of $F - G$ is constant over $]M, \infty[$. Hence, for any $u > F(M)$, from the superadditivity of x^r, (3.1) holds with $c_r = r(r-1)\cdots(r-l+1)2^{l+1-r}$. In the same way, if m is finite, from the definition (3.5), inequality (3.1) holds for $u < F(m)$ with the same constant.

We now define $h(x)$ on $[m, M] \cap \mathbb{R}$. Let the function ψ_r be defined on $[0, 1]$ in the following way. For $r > 3$, we set

$$\psi_r(x) = a_r\big(\sin(\pi x)\big)^{r-1}, \tag{3.6a}$$

where a_r is the largest positive real such that $\psi_r \in \Lambda_{r-1}$. Then $\psi_r^{(k)}(0) = \psi_r^{(k)}(1) = 0$ for any integer k in $[0, l[$.

For r in $]2, 3]$, we define ψ_r by

$$\psi_r(x) = (r-1)^{-1}x^{r-1} \quad \text{for } x \in [0, 1/4], \tag{3.6b}$$

$\psi_r(x) = 2\psi_r(1/4) - \psi_r((1/2) - x)$ for x in $[1/4, 1/2]$ and $\psi_r(x) = \psi_r(1-x)$ for x in $[1/2, 1]$. The function ψ_r satisfies $\psi_r(0) = \psi_r(1) = 0$ and $\psi_r'(0) = \psi_r'(1) = 0$.

For r in $]1, 2]$, we set

$$\psi_r(x) = \min\big(x^{r-1}, (1-x)^{r-1}\big). \tag{3.6c}$$

It is easy to check that the so defined functions ψ_r belong to Λ_{r-1}.

For x real in $[m, M]$, we define the function h by

$$h(x) = \int_c^x f(t)\,\mathrm{d}t, \tag{3.7a}$$

where c is some fixed number in $]m, M[$, $f(t) = 0$ for t in A and

$$f(t) = 2^{r-l-1}(b-a)^{r-1}\psi_r\big((t-a)/(b-a)\big)S\big(F(t) - G(t)\big) \tag{3.7b}$$

for t in $A^c \cap]m, M[$, with $a = \sup A \cap]-\infty, t]$ and $b = \inf A \cap [t, \infty[$.

We now prove that h belongs to Λ_r. By definition h is absolutely continuous on $\mathbb{R}$ l-times continuously differentiable on A^c and, for k in $[1, l]$ and t in $A^c \cap]m, M[$,

$$h^{(k)}(t) = 2^{r-l-1}(b-a)^{r-k}\psi_r^{(k-1)}\big((t-a)/(b-a)\big)S\big(F(t)-G(t)\big) \tag{3.8}$$

with $a = \sup A \cap]-\infty, t]$ and $b = \inf A \cap [t, \infty[$. Since ψ_r and the derivatives of ψ_r up to order $l-1$ vanish for $x = 0$ and $x = 1$, we get, by induction on k, that

$$\big|\psi_r^{(k-1)}(x)\big| \le \frac{\min(x^{r-k}, (1-x)^{r-k})}{(r-k)\cdots(r-l+1)}$$

for any integer k in $[1, l]$. It follows that

$$\big|h^{(k)}(t)\big| \le 2^{r-l-1}\frac{\min((t-a)^{r-k}, (b-t)^{r-k})}{(r-k)\cdots(r-l+1)}.$$

Hence, for any integer k in $[1, l]$, any t in $A^c \cap]m, M[$ and any x in A,

$$\big|h^{(k)}(t)\big| \le 2^{r-l-1}\frac{|t-x|^{r-k}}{(r-k)\cdots(r-l+1)}. \tag{3.9}$$

Furthermore, from (3.4) and (3.5), inequality (3.9) still holds for $t > M$ or $t < m$. Applying (3.9) to $k = 1$, we get that f is continuous, which ensures that h is C^1 and $h' = f$. Next, from (3.9), proceeding by induction on k, we get that, for any $k \le l$, h is k times differentiable and $h^{(k)}(x) = 0$ for x in A. From (3.9) again, $h^{(l)}$ is continuous. Now let x and y be two reals with $x < y$. If $[x, y] \cap A \ne \varnothing$, from (3.9),

$$\big|h^{(l)}(y) - h^{(l)}(x)\big| \le 2^{r-l-1} \sup_{t\in[x,y]} \big((t-x)^r + (y-t)^r\big) \le |y-x|^r.$$

If $[x, y]$ does not intersect A and $x > M$, from (3.4),

$$\big|h^{(l)}(y) - h^{(l)}(x)\big| \le |y-x|^r. \tag{3.10}$$

In the same way, if $y < m$, (3.10) holds. If $[x, y] \subset]m, M[$, x and y are in the same connex component of A^c, and consequently (3.8) holds true with $a < x < y < b$. Then (3.10) follows from (3.8) and the fact that ψ_r belongs to Λ_{r-1}. Hence h belongs to Λ_r.

It remains to prove (3.1) for u in $]F(m), F(M)[$. Since $F^{-1}(u) = G^{-1}(u)$ for any u in $F(A)$, it is sufficient to prove (3.1) for $u \notin F(A)$. Let $x = F^{-1}(u)$ and $y = G^{-1}(u)$. Clearly x does not belong to A and $m < x < M$. Setting $a = \sup A \cap]-\infty, x]$ and $b = \inf A \cap [x, \infty[$, we have: $a < F^{-1}(u) < b$ and $F(a) < u < F(b)$. Since $F(a) = G(a)$ and $F(b) = G(b)$, it implies that $G(a) < u < G(b)$ and $a < G^{-1}(u) < b$. Hence x and y are in the same connex component of A. Suppose, for example, that $G^{-1}(u) > F^{-1}(u)$. Then $F(t) > G(t)$ for any t in $]a, b[$ and

$$h(y) - h(x) = 2^{r-l-1}\int_x^y (b-a)^{r-1}\psi_r\big((t-a)/(b-a)\big)\,\mathrm{d}t.$$

Now, from the definition of ψ_r, it can be proven that, for t in $[x, y]$,

$$(b-a)^{r-1}\psi_r\big((t-a)/(b-a)\big) > (y-x)^{r-1}\psi_r\big((t-x)/(y-x)\big).$$

Hence, for $x = F^{-1}(u)$ and $y = G^{-1}(u)$,

$$h(y) - h(x) \ge 2^{r-l-1}(y-x)^r \int_0^1 \psi_r(t)\,\mathrm{d}t. \tag{3.11}$$

For r in $]1, 2]$,

$$I_r = \int_0^1 \psi_r(t)\,\mathrm{d}t = 2\int_0^{1/2} t^{r-1}\,\mathrm{d}t = \frac{1}{r}2^{1-r}.$$

Hence, from (3.4), (3.5) and (3.11), inequality (3.1) holds with $c_r = 2r$.

For r in $]2, 3]$, from the symmetry properties of ψ_r,

$$I_r = \int_0^1 \psi_r(t)\,\mathrm{d}t = \psi_r(1/4) = \frac{1}{r-1} 4^{1-r}.$$

Hence, from (3.11), inequality (3.1) holds for u in $]F(m), F(M)[$ with $c_r = (r-1)2^{r+1}$. For $u \notin]F(m), F(M)[$, from the definitions (3.4) and (3.5), inequality (3.1) holds with the constant $4r(r-1)$. Since $4r \leq 2^{r+1}$ for $r > 2$, we infer that (3.1) holds for any u in $]0, 1[$ with $c_r = (r-1)2^{r+1}$. Finally, for $r > 3$, inequality (3.1) holds with $c_r = (a_r I_r)^{-1}$.

Proof of Proposition 3.1. Recall that $l = r - 1$. Consequently

$$a_r^{-1} = \sup\bigl\{\bigl|\psi_r^{(l)}(t)\bigr| \colon t \in [0, 1]\bigr\}.$$

Now, by the De Moivre formula,

$$\psi_r(t) = (2\mathrm{i})^{-l} a_r \sum_{k=0}^{l} \frac{l!}{(l-k)!k!} (-1)^k \mathrm{e}^{\mathrm{i}\pi(l-2k)t}.$$

Hence

$$\psi_r^{(l)}(t) = \pi^l a_r \sum_{k=0}^{l} \frac{l!}{(l-k)!k!} 2^{-l} (-1)^k (l-2k)^l \mathrm{e}^{\mathrm{i}\pi(l-2k)t},$$

which implies that

$$a_r^{-1} \leq \pi^l \sum_{k=0}^{l} \frac{l!}{(l-k)!k!} 2^{-l} |l-2k|^l = \pi^l \mathbb{E}\bigl(|\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_l|^l\bigr), \tag{3.12}$$

if $(\varepsilon_k)_{k>0}$ is a sequence of i.i.d. random variables with $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$.

Let $T_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$. Suppose now that r is odd: $r = 2m + 1$ for some integer $m > 1$. If Y is a standard normal, it is known that

$$\mathbb{E}\bigl(T_n^{2m}\bigr) \leq n^m \mathbb{E}\bigl(Y^{2m}\bigr) = n^m \frac{(2m)!}{2^m m!}. \tag{3.13}$$

Next

$$I_r = \int_0^1 \bigl(\sin(\pi t)\bigr)^{2m} \mathrm{d}t = \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2 \cdot 4 \cdot 6 \cdots 2m} = 2^{-2m} \frac{(2m)!}{m!m!}. \tag{3.14}$$

Equations (3.12), (3.13) with $n = 2m$ and (3.14) ensure that

$$(a_r I_r)^{-1} \leq \bigl(4\pi^2 m\bigr)^m m! \quad \text{with } m = (r-1)/2 = l/2.$$

Now, recall that the sequence $(n^{-1/2}(\mathrm{e}/n)^n \exp(-1/(12n)) n!)_n$ is non-decreasing and converges to $\sqrt{2\pi}$. Consequently $m! \leq \sqrt{2\pi m}(m/\mathrm{e})^m \exp(1/12m)$ and (3.1) holds with

$$c_r = \sqrt{\pi l}\bigl(\pi^2/\mathrm{e}\bigr)^{l/2} l^l \exp\bigl(1/(6l)\bigr). \tag{3.15}$$

If r is even, $r = 2m$, then $l = 2m - 1$. By the Cauchy–Schwarz inequality,

$$\mathbb{E}\bigl(|T_l|^l\bigr) \leq \bigl(\mathbb{E}\bigl(T_l^{2m}\bigr)\mathbb{E}\bigl(T_l^{2m-2}\bigr)\bigr)^{1/2} = \left(\frac{l^l}{2^l} \frac{2m!(2m-2)!}{m!(m-1)!}\right)^{1/2}. \tag{3.16}$$

Next

$$I_r = \int_0^1 \big(\sin(\pi t)\big)^{2m-1}\,\mathrm{d}t = \frac{2\cdot 4\cdots(2m-2)}{3\cdot 5\cdots 2m-1}\frac{2}{\pi} = 2^{2m-1}\frac{(m-1)!(m-1)!}{(2m-1)!\pi}. \tag{3.17}$$

Equations (3.12), (3.16) and (3.17) ensure that

$$(a_r I_r)^{-2} \le \pi^2(2m-1)^2\left(\frac{\pi^2 l}{8}\right)^l \frac{(2m)!(2m-2)!}{m!(m-1)!}\left(\frac{(2m-2)!}{(m-1)!(m-1)!}\right)^2. \tag{3.18}$$

Now, recall that the sequence $(n^{-1/2}(\mathrm{e}/n)^n n!)_n$ is non-increasing and converges to $\sqrt{2\pi}$. Hence

$$\frac{(2m)!(2m-2)!}{m!(m-1)!} \le \frac{2^{4m-1}}{\mathrm{e}^{2m-1}} m^m (m-1)^{m-1} \quad \text{and} \quad \left(\frac{(2m-2)!}{(m-1)!(m-1)!}\right)^2 \le \frac{2^{4m-4}}{\pi(m-1)}.$$

Putting these inequalities in (3.18), one can prove that

$$(a_r I_r)^{-1} \le \sqrt{\pi l}\big(\pi^2/\mathrm{e}\big)^{l/2}(l+1)l^{l-1},$$

which completes the proof of Theorem 3.1. □

4. Upper bounds for the Wasserstein distances under moment assumptions

In this section we extend (1.3) to Wasserstein distances of order r for r in $]1,2]$. The main result is Theorem 4.1 below, which provides optimal rates of convergence for independent random variables under minimal moment conditions.

Theorem 4.1. *Let r in $]1,2]$. There exists some positive constant C_0 depending only on r such that, for any positive n and any sequence $(X_i)_{i>0}$ of independent real-valued random variables in $\mathbb{L}^{r+2}$ with mean* 0

$$\big(W_r(\mu_n,\gamma_1)\big)^r \le C_0 L_{r+2,n}, \quad \text{where } L_{r+2,n} = v_n^{-(r+2)/2}\sum_{i=1}^n \mathbb{E}\big(|X_i|^{r+2}\big).$$

Remark 4.1. *Sakhanenko* (1985) *proved that for any positive r there exists some constant c_0 depending only on r such that*

$$\big(W_{r+2}(\mu_n,\gamma_1)\big)^{r+2} \le c_0 L_{r+2,n}. \tag{4.1}$$

Suppose that the random variables X_i are identically distributed. From (4.1),

$$W_{r+2}(\mu_n,\gamma_1) \le c_0^{1/(r+2)}\big(\|X_1\|_{r+2}/\|X_1\|_2\big)n^{-r/2(r+2)}. \tag{4.2}$$

Hence (4.1) *provides some rate of convergence in the CLT for the metric W_{r+2}. However, in order to reach the optimal rate* $\mathrm{O}(n^{-1/2})$, *one needs to consider the weaker metric W_r, as shown by Theorem* 4.1, *which gives*

$$W_r(\mu_n,\gamma_1) \le C_0^{1/r}\big(\|X_1\|_{r+2}/\|X_1\|_2\big)^{(r+2)/r} n^{-1/2}. \tag{4.3}$$

Now, from Theorem 4.1 and Sakhanenko's estimate (4.1), we get upper bounds for the Wasserstein distances of order p, for p in $[r, r+2]$ in Corollary 4.2 below.

Corollary 4.2. *Under the assumptions of Theorem* 4.1, *for any p in* $[r, r+2]$,

(a) $\big(W_p(\mu_n, \gamma_1)\big)^p \le \max(C_0, c_0) L_{r+2,n}$.

Consequently, for i.i.d. random variables,

(b) $W_p(\mu_n, \gamma_1) \le \big(\max(C_0, c_0)\big)^{1/p} \big(\|X_1\|_{r+2}/\|X_1\|_2\big)^{(r+2)/p} n^{-r/2p}$.

Proof of Theorem 4.1. Our proof is mainly based on Theorem 3.1 and on asymptotic expansions of smooth functions in the CLT. By Theorem 1, p. 294 in Barbour (1986), for any function f in Λ_r,

$$\mathbb{E}\big(f\big(v_n^{-1/2} S_n\big)\big) = \mathbb{E}\big(f(Y)\big) - \frac{1}{2}\beta\mathbb{E}\big(D_2\Theta f(Y)\big) + \eta, \tag{4.4a}$$

where $\beta = v_n^{-3/2}\mathbb{E}(S_n^3)$, Y is a standard normal r.v. and $D_2\Theta f$ is a function depending only on f, and

$$|\eta| \le C_r L_{r+2,n}. \tag{4.4b}$$

If $\mathbb{E}(S_n^3) = 0$, then (4.4) ensures that $Z_r(\mu_n, \gamma_1) \le C_r L_{r+2,n}$, where Z_r is the Zolotarev distance as defined in Definition 3.1. From Theorem 3.1, we get that

$$\big(W_r(\mu_n, \gamma_1)\big)^r \le 2r C_r L_{r+2,n},$$

which implies Theorem 4.1 in that case.

Suppose now that $\mathbb{E}(S_n^3) \ne 0$. Let $(\Pi_t)_{t>0}$ denote the homogeneous Poisson process on $\mathbb{R}^+$ with Lebesgue measure as intensity. Take

$$N = \beta\Pi_{\beta^{-2}} - \beta^{-1}. \tag{4.5}$$

Then $\mathbb{E}(N) = 0$, $\operatorname{Var} N = 1$ and $\mathbb{E}(N^3) = \beta$. From the infinite divisibility of the Poisson process, for any positive integer m, N is the sum of m independent distributed random variables with the common law of $\beta\Pi_{1/m\beta^2} - (m\beta)^{-1}$. Hence, by (4.4) applied with $n = m$ and $v_n = 1$

$$\mathbb{E}\big(f(N)\big) = \mathbb{E}\big(f(Y)\big) - \frac{1}{2}\beta\mathbb{E}\big(D_2\Theta f(Y)\big) + \eta_m, \tag{4.6a}$$

with

$$\eta_m \le C_r m|\beta|^{r+2}\mathbb{E}\big(\big|\Pi_{1/m\beta^2} - \big(1/m\beta^2\big)\big|^{r+2}\big). \tag{4.6b}$$

As m tends to infinity, the upper bound in (4.6b) converges to $C_r|\beta|^r$. Hence, by (4.4) and (4.6),

$$Z_r(\mu_n, P_N) \le C_r\big(L_{r+2,n} + |\beta|^r\big). \tag{4.7}$$

In order to bound up $|\beta|^r$, define $\bar\mu = n^{-1}\sum_{k=1}^n P_{v_n^{-1/2}X_k}$ and let $\bar X$ be a random variable with law $\bar\mu$. Then $\mathbb{E}(\bar X^3) = \beta/n$ and

$$\big|\mathbb{E}\big(\bar X^3\big)\big|^r = \big(\mathbb{E}\big(|\bar X|^{2(r-1)/r + (r+2)/r}\big)\big)^r \le \big(\mathbb{E}\big(\bar X^2\big)\big)^{r-1}\mathbb{E}\big(|\bar X|^{r+2}\big)$$

by the Hölder inequality. Since $\mathbb{E}(\bar X^2) = 1/n$ and $\mathbb{E}(|\bar X|^{r+2}) = n^{-1}L_{r+2,n}$, we have:

$$|\beta|^r \le L_{r+2,n}. \tag{4.8}$$

By (4.7), (4.8) and Theorem 3.1, we get that

$$W_r(\mu_n, P_N) \le (4rC_r)^{1/r} L_{r+2,n}^{1/r}. \tag{4.9}$$

It remains to bound the Wasserstein distance between P_N and γ_1. From Corollary 2.2 together with the symmetry of the Gaussian law,

$$W_\psi(P_N, \gamma_1) \le |\beta| B \quad \text{with } B = \sup_{\lambda \ge 0} W_\psi\big(\mathcal{P}(\lambda), \mathcal{N}(\lambda, \lambda)\big). \tag{4.10}$$

Next, note that $\exp(x) - 1 \ge \max(x, x^2)$ for any positive x, which ensures that $\psi(x) \ge x^r$ for any r in $[1, 2]$. Hence, for any r in $[1, 2]$,

$$W_r(P_N, \gamma_1) \le W_\psi(P_N, \gamma_1) \le B|\beta| \le B L_{r+2,n}^{1/r}$$

by (4.8). Both the above inequality and (4.9) imply Theorem 4.1. □

Proof of Corollary 4.2. From the results of Dall'Aglio (1956) and Fréchet's remark (1957), for any $s \ge 1$,

$$W_s(\mu_n, \gamma_1) = \|Z_n\|_s, \quad \text{where } Z_n = v_n^{-1/2} F_n^{-1}\big(\Phi(Y)\big) - Y. \tag{4.11}$$

Now from the convexity properties of Hölder norms, for any p in $[r, r+2]$,

$$\mathbb{E}\big(|Z_n|^p\big) \le \big(\mathbb{E}|Z_n|^r\big)^{(r+2-p)/2}\big(\mathbb{E}|Z_n|^{r+2}\big)^{(p-r)/2} \le \max(c_0, C_0) L_{r+2,n} v_n^{(p-r-2)/2} \tag{4.12}$$

by (4.1) and Theorem 4.1. Hence Corollary 4.2(a) holds true; (b) follows immediately from (a). □

5. Lower bounds for minimal distances in the CLT for i.i.d. random variables

In this section we give a converse to inequality (1.7). Theorem 5.1 below proves that the estimate (1.7) cannot be improved.

Theorem 5.1. *For any $a \ge 1$, there exists a sequence $(X_i)_{i>0}$ of i.i.d. random variables with mean zero, satisfying $\mathbb{E}(|X|^{r+2}) = a^r$ for any $r \ge 0$, such that, for any $r \ge 1$,*

$$\liminf_{n \to \infty} \big(\sqrt{n} W_r(\mu_n, \gamma_1)\big)^r \ge b_r a^r \quad \textit{with } b_r = 2^{-r}/(r+1).$$

Remark 5.1. *For $r = 1$ the constant appearing in Theorem 5.1 is $1/4$, when Zolotarev's constant is $1/2$. In the case $a = 1$, Theorem 5.1 holds with the better constant $b_r = 1/(r+1)$ (see the proof below).*

Proof of Theorem 5.1. Define the law of X_1 by

$$\mathbb{P}(X_1 = a) = \mathbb{P}(X_1 = -a) = 1/\big(2a^2\big) \quad \text{and} \quad \mathbb{P}(X_1 = 0) = 1 - 1/a^2.$$

Then the distribution of S_n is a lattice distribution supported by $a\mathbb{Z}$ (for $a = 1$, the support is $1 + 2\mathbb{Z}$). Hence, if Y is an $N(0, 1)$-distributed random variable,

$$\sqrt{n} W_r(\mu_n, \gamma_1) \ge \big\|d\big(\sqrt{n}Y, a\mathbb{Z}\big)\big\|_r,$$

where $d(x, a\mathbb{Z})$ denotes the distance from x to the set $a\mathbb{Z}$. It follows that

$$\sqrt{n} W_r(\mu_n, \gamma_1) \ge a\big\|d\big(\sqrt{n}(Y/a), \mathbb{Z}\big)\big\|_r$$

(for $a = 1$, one can take $1 + 2\mathbb{Z}$ instead of $\mathbb{Z}$ in this inequality). Now the random variable $d(\sqrt{n}(Y/a), \mathbb{Z})$ has the density

$$f_n(x) = \frac{a}{\sqrt{2\pi n}} \sum_{k \in \mathbb{Z}} \exp\bigg(-\frac{a^2}{2n}(x-k)^2\bigg)$$

over $[-1/2, 1/2]$. Hence $f_n(x)$ converges to 1 as n tends to ∞, for any x in $[-1/2, 1/2]$, and the sequence of random variables $(d(\sqrt{n}(Y/a), \mathbb{Z}))_n$ converges in law to the uniform distribution over $[-1/2, 1/2]$ as n tends to ∞ (for $a = 1$, the sequence converges to the uniform distribution over $[-1, 1]$). Since these random variables take their values in $[-1/2, 1/2]$, the convergence of the $\mathbb{L}^r$ norms also holds. Consequently

$$\lim_{n\to\infty} \big\| d\big(\sqrt{n}(Y/a), \mathbb{Z}\big)\big\|_r^r = \int_{-1/2}^{1/2} |x|^r \, \mathrm{d}x,$$

which implies Theorem 5.1. □

6. Transportation costs in the CLT under moment assumptions

In this section, we extend the results of Section 4 to more general transportation costs. The tool is some extension of Theorem 3.1 to weighted Zolotarev type metrics. We first introduce some cost function between laws.

Definition 6.1. *Let $r \geq 1$ and $p \geq 0$. For μ and ν laws in the set $\mathcal{M}_{r+p}$ of probability laws with finite absolute moment of order $r + p$, with respective distribution functions F and G, we set*

$$\kappa_{r,p}(\mu, \nu) = \int_0^1 \big(1 + 2^{-p}\big(\big|F^{-1}(u)\big| + \big|G^{-1}(u)\big|\big)^p\big)\big|F^{-1}(u) - G^{-1}(u)\big|^r \, \mathrm{d}u.$$

In oder to state the extension of Theorem 3.1, we need to introduce weighted Hölder spaces and the corresponding Zolotarev type distances.

Definition 6.2. *For r in $]1, 2]$ and $p > 0$, let $\Lambda_{r,p}$ be the class of continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = f'(0) = 0$ and*

$$\big|f'(x) - f'(y)\big| \leq |x - y|^{r-1}\big(1 + |x|^p + |y|^p\big)$$

for any (x, y) in $\mathbb{R}^2$. For $r = 1$, let $\Lambda_{1,p}$ denote the class of absolutely continuous functions such that $f(0) = 0$ and $|f'(x)| \leq 1 + |x|^p$ almost everywhere. For μ and ν in $\mathcal{M}_{r+p}$, we set

$$Z_{r,p}(\mu, \nu) = \sup\left\{\int f \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu \colon \ f \in \Lambda_{r,p}\right\}.$$

Proposition 6.1 below is a modified version of Theorem 3.1.

Proposition 6.1. *Let r in $]1, 2]$ and $p \geq 1$. For any laws μ and ν in $\mathcal{M}_{r+p}$,*

$$\kappa_{r,p}(\mu, \nu) \leq 2^{r+p} rp Z_{r,p}(\mu, \nu).$$

The main result of the section is Theorem 6.1 below, which gives estimates of the rate of convergence of $\kappa_{r,p}(\mu_n, \gamma_1)$ to 0. The proof of this result is based on Proposition 6.1.

Theorem 6.1. *Let r in $]1, 2]$ and $p \geq 1$. Set $p^* = \max(p, 2)$. There exists some positive constant C depending only on r and p such that, for any positive n and any sequence $(X_i)_{i>0}$ of independent real-valued random variables in $\mathbb{L}^{r+p^*}$ with mean* 0,

$$\kappa_{r,p}(\mu_n, \gamma_1) \leq C(L_{r+2,n} + L_{r+p^*,n}),$$

provided that $L_{r+2,n} \leq 1$.

From Theorem 6.1, we can derive the result below, which gives estimates of the Wasserstein distances between the distributions of smooth functions of the random variables $S_n/\sqrt{v_n}$ and the standard normal r.v. Y.

Corollary 6.2. *Let r in $]1,2]$ and $p \geq 1$. Set $p^* = \max(p,2)$. There exists some positive constant C depending only on r and p such that, for any positive n and any sequence $(X_i)_{i>0}$ of independent real-valued random variables in $\mathbb{L}^{r+p^*}$ with mean* 0,

$$\sup_{f \in \Lambda_{1,p/r}} \big(W_r(P_{f(S_n/\sqrt{v_n})}, P_{f(Y)})\big)^r \leq C(L_{r+2,n} + L_{r+p^*,n}),$$

provided that $L_{r+2,n} \leq 1$.

Remark 6.1. *The additional condition $L_{r+2,n} \leq 1$ is not too restrictive, since the estimates in Theorem* 6.1 *and Corollary* 6.2 *are not efficient if $L_{r+2,n} > 1$. Let f be defined by $f(x) = x(1 + r(p+2r)^{-1}|x|^{p/r})$. Then the function f is in $\Lambda_{1,p/r}$ and furthermore f is an expanding map. Consequently Corollary* 6.2 *gives sharper estimates than Theorem* 4.1.

Proof of Proposition 6.1. As in Section 3, we may assume that the probability laws have strictly positive and smooth densities. To prove Proposition 6.1, it is enough to prove that there exists some function g in $\Lambda_{r,p}$ such that, for any u in $]0,1[$ with $F^{-1}(u) \neq G^{-1}(u)$,

$$2^{r+p} rp \frac{g(G^{-1}(u)) - g(F^{-1}(u))}{|G^{-1}(u) - F^{-1}(u)|^r} \geq 1 + \left(\frac{|F^{-1}(u)| + |G^{-1}(u)|}{2}\right)^p. \tag{6.1}$$

Let $A_0 = A \cup \{0\}$, where A is the set of reals x such that $F(x) = G(x)$. The function g is defined by

$$g(x) = \frac{2^r}{4p} \int_0^x \big(1 + |t|^p\big)\big(d(t, A_0)\big)^{r-1} S\big(F(t) - G(t)\big)\,\mathrm{d}t. \tag{6.2}$$

By definition g is continuously differentiable and

$$g'(t) = \frac{2^r}{4p}\big(1 + |t|^p\big)\big(d(t, A_0)\big)^{r-1} S\big(F(t) - G(t)\big) = \frac{1}{p}\big(1 + |t|^p\big)h'(t),$$

with h in Λ_r. The fact that g belongs to $\Lambda_{r,p}$ follows from Lemma 6.1 below.

Lemma 6.1. *Let h be a function in Λ_r satisfying $h'(0) = 0$ and g be defined by $g(0) = 0$ and $pg'(x) = h'(x)(1 + |x|^p)$ for any real x. Then g belongs to $\Lambda_{r,p}$.*

Proof. Let x and y reals with $|y| \geq |x|$. By definition of g,

$$p\big(g'(y) - g'(x)\big) = \big(h'(y) - h'(x)\big)\big(1 + |x|^p\big) + h'(y)\big(|y|^p - |x|^p\big).$$

Now $|y|^p - |x|^p \leq p|y|^{p-1}(|y| - |x|) \leq p|y|^{p+1-r}|y - x|^{r-1}$ and $|h'(y)| \leq |y|^{r-1}$. It follows that

$$p\big|g'(y) - g'(x)\big| \leq |y - x|^{r-1}\big(1 + |x|^p + p|y|^p\big),$$

which implies Lemma 6.1. □

We now prove (6.1). Let u be any real in $]0,1[$. Set $x = F^{-1}(u)$ and $y = G^{-1}(u)$. If $x = y$ then (6.1) holds true. Suppose that $x \neq y$. For example, take $x < y$. Then $F - G$ is positive over $[x, y]$ which ensures that A does not intersect $[x, y]$.

Let us distinguish two cases. If $0 \notin]x, y[$, then A_0 does not intersect $]x, y[$ and consequently

$$g(y) - g(x) = \int_x^y g'(t)\,\mathrm{d}t \geq \frac{2^r}{4p} \int_x^y \big(1 + |t|^p\big) \inf\big((t - x)^{r-1}, (y - t)^{r-1}\big)\,\mathrm{d}t. \tag{6.3}$$

Now the function $t \to 1 + |t|^p$ is convex, whence

$$1 + |t|^p \geq 1 + 2^{-p}|x+y|^p + p2^{-p}|x+y|^{p-1}S(x+y)(2t-x-y).$$

Putting this inequality in (6.3), integrating and noting that

$$\int_x^y (2t-x-y)\inf\bigl((t-x)^{r-1}, (y-t)^{r-1}\bigr)\,\mathrm{d}t = 0,$$

we get that

$$g(y) - g(x) \geq \frac{2^r}{4p}\bigl(1 + 2^{-p}|x+y|^p\bigr)\int_x^y \inf\bigl((t-x)^{r-1}, (y-t)^{r-1}\bigr)\,\mathrm{d}t,$$

which ensures that

$$g(y) - g(x) \geq (y-x)^r\bigl(1 + 2^{-p}\bigl(|x|+|y|\bigr)^p\bigr)/(2rp). \tag{6.4}$$

If $0 \in]x, y[$, then $g(y) - g(x) = g(y) - g(0) + g(0) - g(x)$. Arguing as in the proof of (6.4)

$$g(y) - g(0) \geq \frac{2^r}{4p}\int_0^y \bigl(1 + |t|^p\bigr)\inf\bigl(t^{r-1}, (y-t)^{r-1}\bigr)\,\mathrm{d}t \geq \frac{1}{2rp}y^r\bigl(1 + 2^{-p}y^p\bigr)$$

and

$$g(0) - g(x) \geq |x|^r\bigl(1 + 2^{-p}|x|^p\bigr)/(2rp).$$

Hence

$$rp\bigl(g(y) - g(x)\bigr) \geq \frac{1}{2}\bigl(y^r + |x|^r + 2^{-p}\bigl(y^{r+p} + |x|^{r+p}\bigr)\bigr) \geq \left(\frac{y-x}{2}\right)^r + 2^{-p}\left(\frac{y-x}{2}\right)^{r+p},$$

which ensures that

$$2^{r+p}rp\bigl(g(y) - g(x)\bigr) \geq (y-x)^r\bigl(1 + 2^{-p}\bigl(|x|+|y|\bigr)^p\bigr). \tag{6.5}$$

From (6.4) and (6.5), we get (6.1), which implies Proposition 6.1. □

Proof of Theorem 6.1. By Theorem 3 in Borisov, Panchenko and Skilyagina (1998), the expansion (4.4a) holds for f in $\Lambda_{r,p}$ with

$$|\eta| \leq C(L_{r+2,n} + L_{r+p^*,n}) \tag{6.6}$$

for some constant C depending only on r and p.

If $\mathbb{E}(S_n^3) = 0$, then both Proposition 6.1 and (6.6) ensure that

$$\kappa_{r,p}(\mu_n, \gamma_1) \leq rp2^{r+p}Z_{r,p}(\mu_n, \gamma_1) \leq Crp2^{r+p}(L_{r+2,n} + L_{r+p^*,n}).$$

Hence Theorem 6.1 holds true in that case.

If $\mathbb{E}(S_n^3) \neq 0$, then, repeating the arguments of Section 4, we get that, for the random variable N defined in (4.5),

$$\kappa_{r,p}(\mu_n, P_N) \leq C'(L_{r+2,n} + L_{r+p^*,n}) \tag{6.7}$$

for some constant C' depending only on r and p. Since the cost function $\kappa_{r,p}$ satisfies the triangle type inequality $\kappa_{r,p}(\mu, \nu) \leq C(\kappa_{r,p}(\mu, \pi) + \kappa_{r,p}(\pi, \nu))$ for some constant C depending only on r and p, the proof of Theorem 6.1 will be achieved if we prove that

$$\kappa_{r,p}(P_N, \gamma_1) \leq cL_{r+2,n} \tag{6.8}$$

for some constant c depending only on r and p. Now, for any laws μ and ν with respective d.f. F and G, from the definition of $\kappa_{r,p}$, Dall'Aglio's result and elementary convexity inequalities,

$$\begin{aligned}\kappa_{r,p}(\mu,\nu) &\le W_r^r(\mu,\nu)+\frac{1}{2}W_{r+p}^{r+p}(\mu,\nu)+2^{p-1}\int_0^1\left(\left|G^{-1}(u)\right|\right)^p\left|F^{-1}(u)-G^{-1}(u)\right|^r\,\mathrm{d}u\\ &\le W_r^r(\mu,\nu)+\frac{1}{2}W_{r+p}^{r+p}(\mu,\nu)+2^{p-1}W_{r+p}^r(\mu,\nu)W_{r+p}^p(\delta_0,\nu)\end{aligned}\tag{6.9}$$

by the Hölder inequality. We now apply this inequality to $\mu=P_N$ and $\nu=\gamma_1$. By (4.10), $W_\psi(P_N,\gamma_1)\le B|\beta|$, which implies that $W_q(P_N,\gamma_1)\le c_qB|\beta|$ for some constant c_q depending only on q, for any $q\ge 1$. Hence, by (6.9), $\kappa_{r,p}(P_N,\gamma_1)\le c'|\beta|^r(1+|\beta|^p)$ for some constant c' depending only on r and p. Finally, if $L_{r+2,n}\le 1$, then, by (4.8), $|\beta|^r\le L_{r+2,n}\le 1$. Hence, from the above inequality, (6.8) holds with $c=2c'$, which completes the proof of Theorem 6.1. □

Proof of Corollary 6.2. Let μ and ν be laws in $\mathcal{M}_{r+p}$ with respective d.f. F and G, and U be an r.v. with the uniform distribution over [0, 1]. Set $T=F^{-1}(U)$ and $Z=G^{-1}(U)$. Then T and Z have respective laws μ and ν, and

$$\sup_{f\in\Lambda_{1,p/r}}\left(W_r(P_{f(T)},P_{f(Z)})\right)^r\le\mathbb{E}\Big(\sup_{f\in\Lambda_{1,p/r}}|f(T)-f(Z)|^r\Big).$$

Now, for any f in $\Lambda_{1,p/r}$,

$$\left|f(T)-f(Z)\right|^r\le|T-Z|^r\left(1+\left(|T|\vee|Z|\right)^{p/r}\right)^r.$$

It follows that

$$\sup_{f\in\Lambda_{1,p/r}}\left(W_r(P_{f(T)},P_{f(Z)})\right)^r\le c\kappa_{r,p}(\mu,\nu)\tag{6.10}$$

for some constant c depending only on r and p. Corollary 6.2 follows immediately from Theorem 6.1 via (6.10). □

Acknowledgment

I would like to thank the referee for a thorough reading of the initial version and comments that helped me to improve the presentation and the composition of the paper.

References

A. D. Barbour. Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Related Fields* **72** (1986) 289–303. MR0836279

P. Bártfai. Über die entfernung der irrfahrtswege. *Studia Sci. Math. Hungar.* **5** (1970) 41–49. MR0275499

A. Bikelis. Estimates of the remainder term in the central limit theorem. *Litovsk. Mat. Sb.* **6** (1966) 323–346. MR0210173

I. S. Borisov, D. A. Panchenko and G. I. Skilyagina. On minimal smoothness conditions for asymptotic expansions of moments in the CLT. *Siberian Adv. Math.* **8** (1998) 80–95. MR1651906

G. Dall'Aglio. Sugli estremi deli momenti delle funzioni di ripartizione doppia. *Ann. Sc. Norm. Super Pisa Cl. Sci.* **10** (1956) 35–74. MR0081577

C. G. Esseen. On mean central limit theorems. *Kungl. Tekn. Högsk. Handl. Stockholm* **121** (1958) 31. MR0097111

M. Fréchet. *Recherches théoriques modernes sur le calcul des probabilités, premier livre. Généralités sur les probabilités, éléments aléatoires.* Paris, Gauthier-Villars, 1950. MR0038581

M. Fréchet. Sur la distance de deux lois de probabilité. *C. R. Acad. Sci. Paris* **244** (1957) 689–692. MR0083210

I. Ibragimov. On the accuracy of Gaussian approximation to the distribution functions of sums of independent random variables. *Theory Probab. Appl.* **11** (1966) 559–579. MR0212853

J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent rv's and the sample df. I. *Z. Wahrsch. Verw. Gebiete* **32** (1975) 111–131. MR0375412

P. Major. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8** (1978) 487–517. MR0520959

P. Massart. Strong approximation for multivariate empirical and related processes, via K.M.T. constructions. *Ann. Probab.* **17** (1989) 266–291. MR0972785

V. V. Petrov. *Sums of Independent Random Variables.* Berlin, Springer, 1975. MR0388499

E. Rio. Strong approximation for set-indexed partial-sum processes, via KMT constructions I. *Ann. Probab.* **21** (1993) 759–790. MR1217564

E. Rio. Distances minimales et distances idéales. *C. R. Acad. Sci. Paris* **326** (1998) 1127–1130. MR1647215

A. I. Sakhanenko. Estimates in the invariance principle. *Proc. Inst. Math. Novosibirsk* **5** (1985) 27–44. MR0821751

V. M. Zolotarev. On asymptotically best constants in refinements of mean central limit theorems. *Theory Probab. Appl.* **9** (1964) 268–276. MR0171315

V. M. Zolotarev. Metric distances in spaces of random variables and their distributions. *Math. USSR Sbornik* **30** (1976) 373–401. MR0467869