Statistics and Probability Letters 83 (2013) 1644–1648

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Entropic approach to E. Rio's central limit theorem for W_2 transport distance[☆]

Sergey G. Bobkov

School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form 24 March 2013
Accepted 24 March 2013
Available online 30 March 2013

Keywords:
Central limit theorem
Berry–Esseen-type bounds
Transport metrics

ABSTRACT

The central limit theorem is considered with respect to the transport distance W_2. We discuss an alternative approach to a result of E. Rio, based on a Berry–Esseen-type bound for the entropic distance to the normal distribution.

Let X and Z be random variables with distributions F and G, having finite second moments. The Kantorovich distance $W_2(F,G)$ between F and G, also called the quadratic Wasserstein distance, is defined by

$$W_2^2(F,G) = W_2^2(X,Z) = \inf\left\{ \mathbf{E}\,(X'-Z')^2 : X' \sim F,\ Z' \sim G\right\}, \tag{1}$$

where the infimum is taken over all random variables X' and Z' with distributions F and G, respectively. More precisely,

$$W_2^2(F,G) = \inf_\pi \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-z)^2\, d\pi(x,z)$$

with infimum taken over all probability measures π on the plane $\mathbf{R}^2$ having F and G as marginal projections.

Under mild moment constraints, this distance metrizes the weak topology in the space of probability measures on the real line. So, it is of a certain interest to know how to bound W_2 in various limit theorems. Here we consider the usual central limit theorem and the related problem of the normal approximation with respect to W_2.

Thus, let $S_n = X_1 + \cdots + X_n$ be the sum of n independent random variables such that $\mathbf{E}X_k = 0$, $\sum_k \mathbf{E}X_k^2 = 1$. The closeness of the distribution of S_n to the standard normal law may be quantified under higher order moment assumptions in terms of the Lyapunov coefficients

$$L_s = \sum_{k=1}^n \mathbf{E}\,|X_k|^s, \quad s \geq 2.$$

For example, the Berry–Esseen theorem indicates that, up to a numerical constant C,

$$\sup_x |\mathbf{P}\{S_n \leq x\} - \mathbf{P}\{Z \leq x\}| \leq CL_3,$$

where Z is a standard normal random variable (in which case we write $Z \sim N(0,1)$). This inequality quantifies the Kolmogorov (uniform) distance between the corresponding distribution functions. In part concerning the W_2-distance, principal results in this direction are due to E. Rio, who made in particular the following remarkable observation (cf. also Rio, 2011).

☆ Partially supported by the NSF grant DMS-1106530 and Simons Fellowship.
E-mail address: bobkov@math.umn.edu.

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.03.020

Theorem 1 (Rio, 2009). *We have*

$$W_2^2(S_n, Z) \le CL_4, \tag{2}$$

where $Z \sim N(0,1)$ and C is an absolute constant.

Restricting (2) to the i.i.d. case and writing the sum in a more convenient way as $S_n = (\xi_1 + \cdots + \xi_n)/\sqrt{n}$ with $\mathbf{E}\xi_1 = 0$, $\mathbf{E}\xi_1^2 = 1$ and $\mathbf{E}\xi_1^4 < \infty$, Theorem 1 provides a typical expected rate of the normal approximation with respect to n, namely,

$$W_2^2(S_n, Z) \le \frac{C\mathbf{E}\xi_1^4}{n}. \tag{3}$$

In contrast with the Berry–Esseen theorem, this bound relies upon the finiteness of the 4-th moment. However, as was also shown in Rio (2009), the distances $W_2(S_n, Z)$ may decay to zero at a lower rate under the weaker moment assumption $\mathbf{E}\,|\xi_1|^s < \infty$, $2 < s < 4$.

Being somewhat non-trivial, the proof of Theorem 1 given in Rio (2009) is based on the study of the relationship between W_2 and Zolotarev's ideal metrics ζ_s, as well as on Poisson-type approximations. In this note we describe an alternative approach to this result which involves an entropic distance to normality and uses the property that $W_2(F, G)$ does not change considerably under slight smoothing operations.

First let us introduce basic notations. If a random variable X has density p, and a random variable Z has density q, the Kullback–Leibler distance from the distribution of X to the distribution of Z (also called an informational divergence or the relative entropy of X with respect to Z) is defined by

$$D(X \parallel Z) = \int_{-\infty}^{\infty} p(x)\, \log\frac{p(x)}{q(x)}\, dx.$$

This quantity is not symmetric in coordinates (X, Z), but is always non-negative and vanishes, if and only if $p = q$ a.e. Moreover, it majorizes (a function of) the total variation distance between the distributions. What will be more relevant is that, when Z is standard normal, we have M. Talagrand's entropy-transport inequality

$$W_2^2(X, Z) \le 2D(X \parallel Z), \tag{4}$$

cf. Talagrand (1996), or Bobkov and Götze (1999) for a different proof. Such a relation could be used in the proof of Theorem 1, once we are able to bound the relative entropy. And this turns out indeed possible under proper "smoothing" assumptions.

For the sum S_n as before, introduce its characteristic function $f_n(t) = \mathbf{E}\, e^{itS_n}$.

Theorem 2. *If $f_n(t)$ is vanishing outside the interval $|t| \le \frac{1}{4\sqrt{L_4}}$, then*

$$D(S_n \parallel Z) \le CL_4, \tag{5}$$

where $Z \sim N(0,1)$ and C is an absolute constant.

Combining (4) and (5), we are led to the desired estimate $W_2^2(S_n, Z) \le CL_4$, which holds however under an additional support hypothesis on $f_n(t)$. The latter may be removed when applying (5) to the smoothed distributions, namely — to the random variables of the form

$$S_n(\tau) = \sqrt{1-\tau^2}\, S_n + \tau\,\xi, \quad 0 \le \tau \le 1,$$

where ξ is independent of S_n, with $\mathbf{E}\xi = 0$ and $\mathbf{E}\xi^2 = 1$.

Proof of Theorem 1 (*On the Basis of* Theorem 2). By the definition (1),

$$W_2(S_n(\tau), S_n) \le \left(\mathbf{E}\,(S_n(\tau) - S_n)^2\right)^{1/2} \le 2\tau. \tag{6}$$

To meet the requirements of Theorem 2, let ξ have finite 4-th moment and a characteristic function $v(t) = \mathbf{E}\, e^{it\xi}$ vanishing outside an interval $[-T, T]$ of length of order 1. For example, one may take the 4-th convolution power of the characteristic function $h(t) = (1-|t|)^+$ and then normalize and rescale it, so that $v(t) = a\,(h*h*h*h)(bt)$ would satisfy $v(0) = 1$ and $\mathbf{E}\xi^2 = -v''(0) = 1$. By the construction, v is 4 times differentiable, so $\mathbf{E}\xi^4 < \infty$.

To derive (2), we may assume that L_4 is small enough. Let $L_4 \le 1/(16T^2)$ and take $\tau = 4\sqrt{L_4}\,T$. Then we may apply Theorem 2 to the sequence $\sqrt{1-\tau^2}\,X_1, \ldots, \sqrt{1-\tau^2}\,X_n, \tau\xi$, in which case for the corresponding Lyapunov coefficient we have

$$L_4(\tau) = (1-\tau^2)^2 L_4 + \tau^4 \mathbf{E}\xi^4 \le AL_4, \quad A = 1 + 16\,T^2\,\mathbf{E}\xi^4.$$

By (4)–(5), this gives

$$W_2^2(S_n(\tau), Z) \le 2D(S_n(\tau) \parallel Z) \le 2CL_4(\tau) \le 2CAL_4.$$

Hence, by (6), and using the triangle inequality for W_2, we get $W_2(S_n, Z) \le C_\xi\sqrt{L_4}$, where the constant C_ξ depends on the distribution of ξ, only. $\square$

As we see from the proof, a good point of this approach is that, avoiding the use of the special one dimensional formula

$$W_2^2(F,G) = \int_0^1 (F^{-1}(p) - G^{-1}(p))^2\,dp,$$

it can be adapted to cover the multidimensional case (of independent random vectors in $\mathbf{R}^d$). Such extension deserves a separate consideration, and we do not discuss it here. Let us only mention that for the dimensions d higher than 4, the rate $1/n$ like in (3) requires the finiteness of the 5-th absolute moment (as noticed in Bobkov et al. (in press-a) for the entropic central limit theorem in $\mathbf{R}^d$).

As a main argument leading to Theorem 2, we use an Edgeworth-type approximation of f_n by the corrected normal "characteristic function"

$$g_\alpha(t) = \left(1 + \alpha\,\frac{(it)^3}{3!}\right) e^{-t^2/2}$$

with $\alpha = \sum_{k=1}^n \mathbf{E}X_k^3$. First let us recall a routine but rather standard fact, which is based on Taylor's expansion for the characteristic functions of X_k near zero up to order 4, cf. e.g. Statulevičius (1965), Bhattacharya and Ranga Rao (1976).

Lemma 3. *If $L_4 < \infty$, then in the interval $|t|\ \max(L_4^{1/4}, L_4^{1/6}) \leq 1$,*

$$\left|\frac{d^p}{dt^p}\,(f_n(t) - g_\alpha(t))\right| \leq CL_4\left(|t|^{4-p} + |t|^{6+p}\right) e^{-t^2/2}, \quad p = 0,1,2,3,4,$$

where C is an absolute constant.

Lemma 4. *Let $L_s < \infty$, for some integer $s \geq 3$. Then, in the interval $|t| \leq \frac{1}{4L_3}$,*

$$\left|\frac{d^p}{dt^p}\,f_n(t)\right| \leq C_p\,(L_{p^*} + 1)\,e^{-t^2/4}, \quad p = 0,\dots,s,$$

where $p^* = \max(p,2)$, and where the constants C_p depend on p, only.

Here, in the cases $p = 0,1,2$ the term $L_{p^*} = 1$ may be removed. This bound is also known, but is usually stated for the particular value $p = 0$, only. We could not find a reference for the case $p \geq 1$ and therefore will include a standard argument at the end of this note.

In the next auxiliary assertion, we use the notation $\|u\|_2 = \left(\int_{-\infty}^{\infty} |u(t)|^2\,dt\right)^{1/2}$.

Lemma 5. *Let X be a random variable with $\mathbf{E}\,|X|^3 < \infty$. For any $\alpha \in \mathbf{R}$,*

$$D(X \,\|\, Z) \leq \alpha^2 + 4\left(\|f - g_\alpha\|_2 + \|f''' - g_\alpha'''\|_2\right),$$

where $Z \sim N(0,1)$, and f is the characteristic function of X.

Note that the finiteness of $\|f\|_2$ guarantees the existence of a density of X, while the finiteness of the 3-rd absolute moment of X insures that f has continuous derivatives up to order 3. The proof of Lemma 5 can be found in a recent paper by Bobkov et al. (in press-b), where Berry–Esseen-type bounds are treated under entropic assumptions.

Proof of Theorem 2. By Lemma 5,

$$D(S_n \,\|\, Z) \leq \alpha^2 + 4\left(\|f_n - g_\alpha\|_2 + \|f_n''' - g_\alpha'''\|_2\right), \quad \alpha = \sum_{k=1}^n \mathbf{E}X_k^3. \tag{7}$$

Using Lemmas 3 and 4 with $s = 4$, and the general relation $L_3^2 \leq L_4$, one may write down a more unified bound for $f_n - g_\alpha$ (at the expense of a non-essential constant in the exponent)

$$\left|\frac{d^p}{dt^p}\,(f_n(t) - g_\alpha(t))\right| \leq CL_4\,e^{-t^2/4}, \quad |t| \leq (16L_4)^{-1/2},\ p = 0,1,2,3,4, \tag{8}$$

where C is an absolute constant. Due to the assumption that $f_n(t) = 0$ in $|t| \geq (16L_4)^{-1/2}$, the inequalities (7) and (8) with $p = 0$ and $p = 3$ yield

$$\begin{aligned} D(S_n \,\|\, Z) \leq\ & \alpha^2 + 4\left(\int_{|t|\leq(16L_4)^{-1/2}} \left(CL_4\,e^{-t^2/4}\right)^2 dt + \int_{|t|\geq(16L_4)^{-1/2}} |g_\alpha(t)|^2\,dt\right)^{1/2} \\ & + 4\left(\int_{|t|\leq(16L_4)^{-1/2}} \left(CL_4\,e^{-t^2/4}\right)^2 dt + \int_{|t|\geq(16L_4)^{-1/2}} |g_\alpha'''(t)|^2\,dt\right)^{1/2}. \end{aligned}$$

But $|\alpha| \leq L_3 \leq \sqrt{L_4}$, so all these integrals do not exceed L_4^2, up to a numerical constant. $\square$

Proof of Lemma 4. Put $\beta_{p,k} = \mathbf{E}\,|X_k|^p$ and denote by v_k the characteristic function of X_k. First let us recall why one may bound $|f_n(t)|$ on a large interval (cf. e.g. Petrov, 1975, par. 2, Chapter V). Let X_k' be an independent copy of X_k. Then, $\mathbf{E}\,(X_k - X_k')^2 = 2\sigma_k^2$ and, by Jensen's inequality, $\mathbf{E}\,|X_k - X_k'|^3 \leq 4\beta_{3,k}$. Hence, by Taylor's expansion, for any t real,

$$|v_k(t)|^2 = \mathbf{E}\,e^{it(X_k - X_k')} = 1 - \sigma_k^2 t^2 + \frac{4\theta_k}{3!}\,\beta_{3,k}|t|^3 \leq \exp\left\{-\sigma_k^2 t^2 + \frac{4\theta_k}{3!}\,\beta_{3,k}|t|^3\right\}$$

with some $\theta_k = \theta_k(t)$ such that $|\theta_k| \leq 1$. Multiplying these inequalities, we get

$$|f_n(t)|^2 \leq \exp\left\{-t^2 + \frac{2}{3}\,L_3|t|^3\right\},$$

so,

$$|f_n(t)| \leq e^{-t^2/3}, \quad \text{for } |t| \leq \frac{1}{2L_3}. \tag{9}$$

Next, to estimate the absolute value of the p-th derivative of f_n, one may use the polynomial formula

$$f_n^{(p)}(t) = \sum \left(\frac{p!}{q_1! \dots q_n!}\right) v_1^{(q_1)}(t) \cdots v_n^{(q_n)}(t), \tag{10}$$

where the summation runs over all integers $q_k \geq 0$ such that $q_1 + \cdots + q_n = p$.

We first assume that

$$\sigma^2 = \max_k\,\sigma_k^2 \leq \sigma^2(p) = \frac{1}{p}\left(1 - \frac{1}{4^{1/3}}\right).$$

To bound the derivatives of v_k in (10), one may use $|v_k^{(q)}(t)| \leq \beta_{q,k}$ which is good in the case $q \geq 2$. For $q = 1$, we have $v_k'(0) = 0$ and $|v_k''(t)| \leq \sigma_k^2$, so, $|v_k'(t)| \leq \sigma_k^2 |t|$. Therefore, in all cases

$$|v_k^{(q_k)}(t)| \leq \beta_{q_k^*,k}\,(1 + |t|), \quad q_k \geq 1,\ q_k^* = \max(q_k, 2). \tag{11}$$

On the other hand, by the assumption on σ,

$$c^2 = \mathrm{Var}\left(\sum_{k:\,q_k=0} X_k\right) = 1 - \sum_{k:\,q_k\geq 1} \sigma_k^2 \geq 1 - p\sigma^2 \geq \frac{1}{4^{1/3}}.$$

Hence, the 3-rd Lyapunov coefficient for the collection $\left(\frac{1}{c}X_k\right)_{q_k=0}$ does not exceed $2L_3$. Applying the inequality (9) to the sum $\frac{1}{c}\sum_{k:\,q_k=0} X_k$, we obtain that

$$\prod_{k:\,q_k=0} |v_k(t)| \leq e^{-t^2/3}, \quad \text{for } |t| \leq \frac{1}{4L_3}. \tag{12}$$

Now, write $(q_1, \dots, q_n) = (1, \dots, k_1, \dots, k_l, \dots, 1)$, i.e., specifying the indices k for which $q_k \geq 1$, and let $l = \mathrm{card}\{k \leq n : q_k \geq 1\}$. Put $p_1 = q_{k_1}, \dots, p_l = q_{k_l}$ and combine (11) with (12), so as to write in the same interval

$$\prod_{k=0}^{n} |v_k^{(q_k)}(t)| \leq (1 + |t|)^l\,e^{-t^2/3}\,\beta_{p_1^*,k_1} \dots \beta_{p_l^*,k_l}.$$

Using this estimate in (10) and performing summation over all k_j's, we get

$$|f_n^{(p)}(t)| \leq p!\,\widetilde{L}_p\,(1 + |t|)^p\,e^{-t^2/3} \tag{13}$$

with constant

$$\widetilde{L}_p = \sum L_{p_1^*} \dots L_{p_l^*},$$

where the summation is running over all integers $l = 1, \dots, p$ and $p_1, \dots, p_l \geq 1$ such that $p_1 + \cdots + p_l = p$.

Clearly, $\widetilde{L}_1 = 1$ and $\widetilde{L}_2 = 2$. If $p \geq 3$, using the property that the function $q \to L_q^{1/(q-2)}$ is not decreasing in $q > 2$ (due to Markov's inequality), we get

$$L_{p_1^*} \dots L_{p_l^*} = \prod_{j:\,p_j\geq 2} L_{p_j} \leq \prod_{j:\,p_j\geq 2} L_p^{(p_j-2)/(p-2)} = L_p^{\nu}. \tag{14}$$

Here

$$(p-2)\nu = \sum_{j=1}^{l} (p_j - 2)\, 1_{\{p_j \geq 2\}} = p - 2l + \sum_{j:\, p_j = 1} 1 \leq p - 2$$

with the last inequality holding for $l \geq 2$. It also holds for $l = 1$, since then $\sum_{j:\, p_j=1} 1 = 0$ due to the assumption $p \geq 3$. Hence $\nu \leq 1$, so $L_p^{\nu} \leq L_p + 1$, which implies

$$\widetilde{L}_p \;\leq\; (L_p + 1) \sum_{l=1}^{p} \sum_{p_1+\cdots+p_l=p} 1 = A_p(L_p+1) \quad (p \geq 3).$$

Thus, $\widetilde{L}_p \leq A_p(L_{p^*}+1)$ in all cases with a constant A_p depending p, only. Using also $(1+|t|)^p\, e^{-t^2/3} \leq B_p\, e^{-t^2/4}$, we get from (13) the desired inequality

$$|f_n^{(p)}(t)| \leq C_p (L_{p^*}+1)\, e^{-t^2/4}, \quad |t| \leq \frac{1}{4L_3}, \tag{15}$$

with some p-depending constant C_p.

Finally, for the remaining values $\sigma > \sigma(p)$, necessarily $L_3 > \sigma^3(p)$ and $\frac{1}{4L_3} < \frac{1}{4\sigma^3(p)}$, so, it suffices to consider the interval $|t| < \frac{1}{4\sigma^3(p)}$. Applying Rosenthal's inequality $\mathbf{E}\,|S_n|^p \leq D_p(L_{p^*}+1)$, in which D_p depends only on p, we get

$$|f_n^{(p)}(t)| \leq \mathbf{E}\,|S_n|^p \leq D_p(L_{p^*}+1) \leq D_p'(L_{p^*}+1)\, e^{-t^2/4}, \quad |t| < \frac{1}{4\sigma^3(p)}.$$

Here one may take $D_p' = D_p\, e^{1/(64\,\sigma^6(p))}$. Thus, (15) holds without any constraint on σ. $\square$

Acknowledgments

We would like to thank M. Ledoux for pointing to the paper by E. Rio, and N. Gozlan, for reading the manuscript and valuable comments.

References

Bhattacharya, R.N., Ranga Rao, R., 1976. Normal Approximation and Asymptotic Expansions. John Wiley & Sons, Inc.

Bobkov, S.G., Chistyakov, G.P., Götze, F., 2011. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. Ann. Probab. Preprint (in press-a).

Bobkov, S.G., Chistyakov, G.P., Götze, F., 2012. Berry–Esseen bounds in the entropic central limit theorem. Probab. Theory Related Fields. Preprint (in press-b).

Bobkov, S.G., Götze, F., 1999. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. 163 (1), 1–28.

Petrov, V.V., 1975. Sums of Independent Random Variables. Springer-Verlag, New York, Heidelberg, Berlin. p. x+345.

Rio, E., 2009. Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. 45 (3), 802–817.

Rio, E., 2011. Asymptotic constants for minimal distance in the central limit theorem. Electron. Commun. Probab. 16, 96–103.

Statulevičius, V.A., 1965. Limit theorems for densities and the asymptotic expansions for distributions of sums of independent random variables. Theory Probab. Appl. 10 (4), 682–695.

Talagrand, M., 1996. Transportation cost for Gaussian and other product measures. Geom. Funct. Anal. 6 (3), 587–600.