AN OPTIMAL BERRY–ESSEEN TYPE THEOREM FOR INTEGRALS OF SMOOTH FUNCTIONS

LUTZ MATTNER AND IRINA SHEVTSOVA

Abstract. We prove a Berry–Esseen type inequality for approximating expectations of sufficiently smooth functions f, like $f = |\cdot|^3$, with respect to standardized convolutions of laws $P_1, \ldots, P_n$ on the real line by corresponding expectations based on symmetric two-point laws $Q_1, \ldots, Q_n$ isoscedastic to the P_i. Equality is attained for every possible constellation of the Lipschitz constant $\|f''\|_{\mathrm{L}}$ and the variances and the third centred absolute moments of the P_i. The error bound is strictly smaller than $\frac{1}{6}$ times the Lyapunov ratio times $\|f''\|_{\mathrm{L}}$, and tends to zero also if n is fixed and the third standardized absolute moments of the P_i tend to one.

In the homoscedastic case of equal variances of the P_i, and hence in particular in the i.i.d. case, the approximating law is a standardized symmetric binomial one.

The inequality is strong enough to yield for some constellations, in particular in the i.i.d. case with n large enough given the standardized third absolute moment of P_1, an improvement of a more classical and already optimal Berry–Esseen type inequality of Tyurin (2009).

Auxiliary results presented include some inequalities either purely analytical or concerning Zolotarev's ζ-metrics, and some binomial moment calculations.

Contents

2010 *Mathematics Subject Classification.* Primary 60E15; Secondary 60F05.

Key words and phrases. Approximation by Rademacher averages, binomial and normal approximation, extreme point methods, moment and characteristic function inequalities, osculatory inequalities, penultimate approximation, Zolotarev's ζ-metrics.

The work was partially supported by the Russian Foundation for Basic Research (projects 15-07-02984-a and 16-31-60110-mol_a_dk) and by the Ministry for Education and Science of Russia (grant No. MD-2116.2017.1).

1. Introduction and main results

1.1. **Introduction.** In statistics and various other applications of probability theory, inconvenient or even intractable distributions are often approximated by relying on some limit theorem. The most popular among such approximations is the *normal* approximation to distributions of sums of a large number n of independent or weakly dependent random variables with appropriate mean and variance which is based on the central limit theorem. However, to use effectively any approximation in practice one needs an explicit and convenient estimate of its accuracy, and such an estimate may be not as sharp as one might wish. For the purpose of improving the error-bounds one can introduce further terms into the approximating law (leading to the so-called asymptotic expansions) and reach arbitrarily high accuracy, but this requires some additional assumptions on the original distribution. For example, in case of approximating the sums of independent random variables these conditions are: (i) finiteness of the higher-order moments of the random summands and (ii) some kind of smoothness either of the distributions of random summands or of the metric under consideration.

On the other hand, from the general theory of summation of independent random variables it follows that approximation by infinitely divisible distributions may be more effective even without any moment conditions due to the better error-bound, which is, in the i.i.d. case, for the Kolmogorov metric, of the order $O(n^{-2/3})$ [2, 3, 4], rather than $O(n^{-1/2})$ as usual in the CLT, but such an approximation may be inconvenient, because the sequence of *penultimate* approximating infinitely divisible distributions that guarantees the rate $O(n^{-2/3})$ may be very complicated and usually is not given in an explicit form. Recall that an approximation, depending on the sample size n not only through location-scale parameters and, in the present context, usually being merely asymptotically normal itself, is sometimes called a *penultimate* approximation, a terminology apparently first introduced in extreme value theory [22]. A recent example of an explicit and convenient penultimate approximation even in the total variation metric, but only for distributions with an absolutely continuous part and finite fourth-order moments, can be found in [9], where an infinitely divisible shifted-gamma approximation with matching first three moments was proved to have the rate $O(n^{-1})$.

In this paper, as an alternative to the normal approximation, we propose and evaluate another penultimate approximation only assuming finiteness of the third-order moments. Our approximation is in the i.i.d. case of the same rate $O(n^{-1/2})$ as the normal approximation, but its error bound depends more favourably on the standardized third absolute moments of the convolved distributions, and can in fact tend to zero even for n fixed. As the approximating distribution we take the n-fold convolution of the symmetric two-point laws with the same variances as the original laws, which is asymptotically normal itself. Thus, in a terminology used for example in [35, chapter 4], our approximations are laws of Rademacher averages rather than Gaussian laws.

As a corollary, for the approximation of a standardized characteristic function by its Taylor polynomial of degree 2, a new explicit and asymptotically exact error-bound given the absolute third-order moment is obtained in (12) below.

Moreover, trivially using the triangle inequality together with the asymptotic normality of the penultimate distribution, which is valid to a higher order due to vanishing third cumulants and due to the smoothness of the metric under consideration, we obtain a sharp upper bound for the accuracy of the normal approximation which improves an already optimal estimate due to Tyurin [61, 62, 63] for some constellations (see Theorems 1.2, 1.14

below). This improvement is possible due to a more favourable dependence of our estimate on the moments of the convolved distributions.

First attempts of a more effective use of the information on the first three moments of the convolved distributions in the estimates of the accuracy of the normal approximation for the Kolmogorov metric were undertaken by Ikeda [30] and Zahl [67] who used some additional conditions on the values of the first three moments and by Prawitz [46] and Bentkus [5] who made those conditions more explicit (for the detailed review see [59, Sections 2.1.1 and 2.4]). The problem of optimization of use of moment-type information in the estimates of the accuracy of the normal approximation was posed in [55, 56, 54, 57] where it was called the problem of *optimization of the structure* of convergence rate estimates and where this problem was partially solved for estimates of the Kolmogorov and the weighted uniform metrics.

More precisely, let $X, X_1, \dots, X_n$ denote independent and, for simplicity in this introduction, identically distributed (i.i.d.) random variables (r.v.'s) on a certain probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, $\mathbb{E}|X|^3 < \infty$,

$$\widetilde{S}_n \;=\; (X_1 + \ldots + X_n)/\sqrt{n}\,, \quad \Delta_n(X) \;:=\; \sup_{x\in\mathbb{R}} |\mathbb{P}(\widetilde{S}_n \le x) - \Phi(x)|, \tag{1}$$

where Φ is the standard normal distribution function. The problem of optimization of the structure of *asymptotic* convergence rate estimates stated in [55, 56, 54, 57] may be formulated as follows: For every $\varrho \in [1, \infty[$ find the greatest lower bound of all possible numbers $g(\varrho)$ such that for all sufficiently large n

$$\sup_{X\,:\,\mathbb{E}X=0,\,\mathbb{E}X^2=1,\,\mathbb{E}|X|^3=\varrho} \Delta_n(X) \;\le\; \frac{g(\varrho)}{\sqrt{n}} + \varepsilon\left(\frac{\varrho}{\sqrt{n}}\right) \quad \text{with } \lim_{\ell\to 0+} \frac{\varepsilon(\ell)}{\ell} = 0. \tag{2}$$

It is easy to see that this greatest lower bound has the form

$$g_*(\varrho) = \lim_{\ell\to 0+} \sup_{n\in\mathbb{N},\,X} \left\{\sqrt{n}\,\Delta_n(X)\colon \mathbb{E}X = 0,\ \mathbb{E}X^2 = 1,\ \mathbb{E}|X|^3 = \varrho \le \ell\sqrt{n}\right\},$$

moreover, inequality (2) with $g(\varrho) = g_*(\varrho)$ still holds true, hence, the greatest lower bound is attained and equals to $g_*(\varrho)$ for every $\varrho \ge 1$.

The problem of evaluation of the optimal function $g_*(\varrho)$, $\varrho \ge 1$, is very complicated. Historically the first investigations were done in the class of linear functions of the form $g(\varrho) = c\varrho$, where Chistyakov [11] finally managed to find an optimal one. Namely, from Esseen's result [20] it follows that $c \ge C_E$, where

$$C_E \;:=\; \frac{\sqrt{10} + 3}{6\sqrt{2\pi}} \;=\; 0.4097\ldots,$$

and Chistyakov [11] proved that (2) holds true with

$$g(\varrho) \;=\; g_1(\varrho) \;:=\; C_E \cdot \varrho \quad \text{for } \varrho \ge 1.$$

In particular, we have an exact linear upper bound $g_*(\varrho) \le C_E \cdot \varrho$.

In fact, Esseen's results from [20] allow to construct tighter lower bounds for g_* in the class of nonlinear functions. More explicitly, from Esseen's [19] asymptotic expansion

$$\mathbb{P}(\widetilde{S}_n \le x) \;=\; \Phi(x) + (1 - x^2)\mathrm{e}^{-x^2/2} \cdot \frac{\mathbb{E}X^3}{6\sqrt{2\pi n}} + \frac{h\psi_n(x)\mathrm{e}^{-x^2/2}}{\sqrt{2\pi n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty,$$

uniformly in $x \in \mathbb{R}$, where h is the span in case of a lattice distribution of X and $h = 0$ otherwise, ψ_n is a certain $h/(\sigma\sqrt{n})$-periodic $[-\frac{1}{2}, \frac{1}{2}]$-valued function, in [20] Esseen, first, deduced that

$$\lim_{n\to\infty} \Delta_n(X)\sqrt{n} \quad = \quad \frac{|\mathbb{E}X^3| + 3h}{6\sqrt{2\pi}}. \tag{3}$$

Second, he considered and solved an extremal problem yielding an exact upper bound of the R.H.S. of (3) in terms of $\mathbb{E}|X|^3$ only, namely, he proved that

$$\sup_{X\colon \mathbb{E}X=0, \mathbb{E}X^2=1, \mathbb{E}|X|^3<\infty} \frac{|\mathbb{E}X^3| + 3h}{\mathbb{E}|X|^3} \quad = \quad \sqrt{10} + 3,$$

with equality attained iff $X \sim \pm X_{\varrho_E}$, that is, iff X is distributed as either X_ϱ or $-X_\varrho$ with $\varrho = \varrho_E$, where here and below

$$\varrho_E \quad := \quad \sqrt{20(\sqrt{10} - 3)/3} \quad = \quad 1.0401\ldots,$$

and X_ϱ with $\varrho \geq 1$ denotes a r.v. with the two-point distribution uniquely defined by the conditions $\mathbb{E}X_\varrho = 0$, $\mathbb{E}X_\varrho^2 = 1$, $\mathbb{E}X_\varrho^3 \geq 0$, and $\mathbb{E}|X_\varrho|^3 = \varrho$, namely

$$\mathbb{P}\Big(X_\varrho = -\sqrt{\frac{p}{q}}\Big) = q := 1 - p, \quad \mathbb{P}\Big(X_\varrho = \sqrt{\frac{q}{p}}\Big) = p := \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\varrho}{2}\sqrt{\varrho^2 + 8} - \frac{\varrho^2}{2} - 1}\,.$$

In particular, $p = p_E := (4 - \sqrt{10})/2 = 0.4188\ldots$ for $\varrho = \varrho_E$, the span $h_\varrho := 1/\sqrt{pq} = 2\sqrt{2}\Big/\sqrt{\varrho^2 - \varrho\sqrt{\varrho^2 + 8} + 4}$ and

$$B(\varrho) \quad := \quad \mathbb{E}X_\varrho^3 \quad = \quad \sqrt{\varrho^2/2 + \varrho\sqrt{\varrho^2 + 8}/2 - 2} \qquad \text{for } \varrho \in [1, \infty[. \tag{4}$$

This, finally, made it possible to evaluate in the same paper [20] the value of the asymptotically the best constant in the Berry–Esseen inequality, which was, in fact, already introduced above as C_E:

$$\sup_{X\colon \mathbb{E}X=0, \mathbb{E}X^2=1, \mathbb{E}|X|^3<\infty} \lim_{n\to\infty} \frac{\Delta_n(X)\sqrt{n}}{\mathbb{E}|X|^3} \quad = \quad C_E.$$

The expression on the L.H.S. here trivially serves as a lower bound for the factor c in (2) with $g(\varrho) = c\varrho$. We refer to [17] for the corresponding result where arbitrary intervals replace the unbounded ones $]-\infty, x]$ implicit in the definition (1) of $\Delta_n(X)$.

Turning now to not necessarily linear functions, Esseen's result (3) immediately implies that the optimal function g_* satisfies

$$g_*(\varrho) \quad \geq \quad \frac{|\mathbb{E}X_\varrho^3| + 3h_\varrho}{6\sqrt{2\pi}} \quad = \quad \frac{1}{6\sqrt{2\pi}} \cdot \frac{2\sqrt{\varrho\sqrt{\varrho^2 + 8} - \varrho^2 - 2} + 6\sqrt{2}}{\sqrt{\varrho^2 - \varrho\sqrt{\varrho^2 + 8} + 4}} \quad =: \quad g_0(\varrho)$$

for every $\varrho \geq 1$. Since $g_*(\varrho) \leq g_1(\varrho)$ and $g_0(\varrho) = g_1(\varrho) := C_E \cdot \varrho$ iff $\varrho = \varrho_E$, we conclude that $g_*(\varrho) = g_1(\varrho)$ iff $\varrho = \varrho_E$, i.e. Chistyakov's optimal linear upper bound is indeed optimal only in one point $\varrho = \varrho_E$.

In recent papers [55, 54] the optimal function g_* was found on a nondegenerate, but still bounded interval, containing ϱ_E, namely, it has been proved that (2) holds for every $\varrho \geq 1$ with

$$g(\varrho) \quad = \quad g_2(\varrho) \quad := \quad \inf_{c \geq 2/(3\sqrt{2\pi})} \{c\varrho + K(c)\} \quad \leq \quad \frac{2\varrho}{3\sqrt{2\pi}} + \sqrt{\frac{2\sqrt{3} - 3}{6\pi}},$$

with equality throughout for $\varrho \geq \sqrt{6\sqrt{3}}(4-\sqrt{3})/6 = 1.2185\ldots$, where

$$K(c) \;:=\; \frac{M(p(\lambda),\lambda)}{6\sqrt{2\pi}}\bigg|_{\lambda=6\sqrt{2\pi}c-3}, \quad M(p,\lambda) \;:=\; \frac{1-\lambda+2(\lambda+2)p-2(\lambda+3)p^2}{\sqrt{p(1-p)}},$$

$$p(\lambda) \;:=\; \frac{1}{2}-\sqrt{\frac{\lambda+1}{\lambda+3}}\sin\left(\frac{\pi}{6}-\frac{1}{3}\arctan\sqrt{\lambda^2+2\frac{\lambda-1}{\lambda+3}}\right),$$

and it has been demonstrated that $g_2(\varrho) = g_0(\varrho)$ for $1 \leq \varrho \leq 3^{1/4}(4-\sqrt{3})/\sqrt{6} = 1.2185\ldots$, yielding the equality $g_*(\varrho) = g_0(\varrho)$ on the above interval. In particular, $g_*(1) = g_2(1) = 1/\sqrt{2\pi} = 0.1989\ldots$. Also $g_2(\varrho)$ is asymptotically optimal for $\varrho \to \infty$ in the sense of

$$\lim_{\varrho\to\infty}\frac{g_2(\varrho)}{g_*(\varrho)} \;=\; 1.$$

Moreover, in [55, 54] there have been obtained explicit estimates of the remainder term ε in (2) with $g = g_2$ yielding the universal bound $\varepsilon(\ell) \leq 2\ell^{3/2}$ for all $\ell > 0$.

An extension of (2) to the non-i.i.d. case has also been obtained in [55, 54] in the form

$$\sup_{n,\,X_1,\ldots,X_n}\sup_x |\mathbb{P}(\widetilde{S}_n \leq x)-\Phi(x)| \;\leq\; \tau\cdot g_2(\ell/\tau)+3\ell^{7/6}$$

with the same function g_2, where the supremum is taken over all n and all centred distributions of $X_1,\ldots,X_n$ such that $\sum_{i=1}^n \mathbb{E}|X_i|^3/(\sum_{i=1}^n\sigma_i^2)^{3/2} = \ell$, $\sum_{i=1}^n\sigma_i^3/(\sum_{i=1}^n\sigma_i^2)^{3/2} = \tau$, $\sigma_i^2 := \mathbb{E}X_i^2$, $i = 1,\ldots,n$. These bounds improve the earlier results of Bentkus [5] and Prawitz [46].

In case of i.i.d. two-point random summands distributed as X_ϱ recently Schulz [50] proved that the remainder term ε in (2) with $g = g_2$ can be omitted for $1 \leq \varrho \leq 5\sqrt{2}/6 = 1.1785\ldots$ (which corresponds to $p \in [1/3, 2/3]$), generalizing the earlier result by Hipp and Mattner [26] originally obtained for $\varrho = 1$. Also, in [50] it is proved that

$$\Delta_n(X_\varrho) \;\leq\; \frac{\sqrt{10}+3}{6\sqrt{2\pi}}\varrho \quad \text{for every} \quad \varrho \geq 1 \text{ and } n \in \mathbb{N}.$$

The present paper can in its main parts be regarded as a transfer and then improvement of some of the above results from the Kolmogorov to the appropriate Zolotarev metric, namely ζ_3.

For the related topic of asymptotic expansions of expectations of smooth functions in the CLT, where rigorous results go back at least to Cramér [12, p. 45, (41a)] in the case of characteristic functions, and to von Bahr [64] in the case of moments and absolute moments, we may refer in chronological order to the surveys in [6, section 25, that section apparently unchanged from its earlier 1986 edition], [24, Chapter 2], and [42, pp. 196–197], and to the more recent papers [8, 7, 31]. From the vast literature on asymptotic expansions of distribution functions, and thus expectations of certain non-smooth functions, for which one may also consult the monographs just cited, let us just mention the recent paper [1].

This paper is organized as follows. Subsections 1.2, 1.3, 1.4, and 1.5 present exact formulations of the main results with discussion. Sections 5 and 6 contain the proofs of the main results. The latter are based on Hoeffding's [28] and Tyurin's [61, 62, 63] results for extremal values of linear and quasi-convex functionals under given moment conditions treated in a novel way in section 3, the previously obtained bound on the third-order moment given the absolute third-order moment [58] as well as a new exact absolute third moment recentering inequality presented in Lemma 2.5, various properties of ζ-metrics, in particular in connection with the s-convex ordering [13] as treated in section 4, and the properties of the Krawtchouk polynomials [36] associated to the symmetric binomial law used in section 6.

The main results of this paper have been anounced without proofs in [38].

1.2. **Notation.** Let $\mathrm{Prob}(\mathbb{R})$ stand for the set of all probability distributions on the real line, $\mathrm{Prob}_s(\mathbb{R}) := \{P \in \mathrm{Prob}(\mathbb{R}) : \nu_s(P) := \int |x|^s \,\mathrm{d}P(x) < \infty\}$ for $s > 0$, $\sigma^2(P) := \inf\{\int (x-a)^2 \,\mathrm{d}P(x) : a \in \mathbb{R}\}$ for $P \in \mathrm{Prob}(\mathbb{R})$, $\mathcal{P}_3 := \{P \in \mathrm{Prob}_3(\mathbb{R}) : \sigma(P) > 0\}$, $\mu_k(P) := \int x^k \,\mathrm{d}P(x)$ for $P \in \mathrm{Prob}_k(\mathbb{R})$ with $k \in \mathbb{N}$, $\mu(\cdot) := \mu_1(\cdot)$. We write N_σ for the centred normal law on $\mathbb{R}$ with standard deviation $\sigma \in [0,\infty[$, and $\mathrm{N} := \mathrm{N}_1$ for the standard normal law with distribution function Φ. The one-point law concentrated at $a \in \mathbb{R}$ is denoted by δ_a. For $n \in \mathbb{N} = \{1,2,\ldots\}$, let $\mathrm{B}_{n,\frac12} := (\frac12(\delta_0+\delta_1))^{*n}$, where the asterisk $*$ indicates convolution, denote the binomial law with $\mathrm{B}_{n,\frac12}(\{k\}) = \mathrm{b}_{n,\frac12}(k) := \binom{n}{k}2^{-n}$ for $k \in \mathbb{N}_0 := \mathbb{N}\cup\{0\}$. If $P \in \mathcal{P}_3$, then we let $\widetilde{P}$ denote its standardization, that is, the image of P under the map $x \mapsto (x-\mu(P))/\sigma(P)$, and

$$\varrho(P) \;:=\; \nu_3\left(\widetilde{P}\right) \;=\; \int \left|\frac{x-\mu(P)}{\sigma(P)}\right|^3 \mathrm{d}P(x) \;=\; \widetilde{P}|\cdot|^3 \;=\; P\left|\frac{\cdot-\mu(P)}{\sigma(P)}\right|^3$$

its standardized third absolute moment; of course then $\varrho(P) \geq 1$, and $\varrho(P) = 1$ iff $\widetilde{P} = \frac12(\delta_{-1}+\delta_1)$. Further, let $\widetilde{\mathcal{P}_3} := \{P \in \mathcal{P}_3 \colon \mu(P) = 0,\ \sigma(P) = 1\} = \{\widetilde{P} \colon P \in \mathcal{P}_3\}$. The tilde notation just introduced should not lead to confusion with a more standard one, used also here, for indicating equality of laws of random variables, as in $X \sim Y$, or for specifying the law of a random variable, as in $X \sim P$.

We use the terms like “positive”, “increasing”, and “convex” in the wide sense, adding “strictly” when appropriate. Also, “interval” may refer to any convex subset of $\mathbb{R}$, possibly degenerated to one point or even to the empty set. Finally, we use the de Finetti indicator notation, (statement) $:= 1$ if “statement” is true, (statement) $:= 0$ otherwise, for example in (47) below.

1.3. **The auxiliary functions A and B.** Recalling the definition of $B(\varrho)$ given in (4) above let us put

$$A(\varrho) \;:=\; \varrho^{-1}B(\varrho) \;=\; \sqrt{\tfrac12\sqrt{1+8\varrho^{-2}}+\tfrac12-2\varrho^{-2}} \quad \text{for } \varrho \in [1,\infty[. \tag{5}$$

The notation A here is as used in [58, pp. 194, 208], so let us note that there is an inconsequential typo in the formula for $A'(\varrho)$ in [58, p. 208], where $\varrho^{3/2}$ should be $\varrho^3/2$.

Lemma 1.1. *The functions A and B are strictly concave and increasing, with $A(1) = B(1) = 0$, $\lim_{\varrho\to1} A(\varrho)/\sqrt{\varrho-1} = \sqrt{8/3}$, and $\lim_{\varrho\to\infty} A(\varrho) = 1$. In particular, we have $0 < A(\varrho) < 1$ and $\varrho - 1 < B(\varrho) < \varrho$ for $\varrho \in]1,\infty[$.*

Proof. We have $(A^2(\varrho))' = 4\varrho^{-3}(1-(1+8\varrho^{-2})^{-1/2})$ strictly decreasing and positive for $\varrho \in [1,\infty[$, hence A^2 strictly concave and increasing, thus $A = \sqrt{A^2}$ strictly concave and increasing as well, and also $\lim_{\varrho\to1} A^2(\varrho)/(\varrho-1) = (A^2)'(1) = 8/3$. B is obviously strictly increasing and, by [58, p. 209], satisfies $B'' < 0$ and is hence strictly concave; hence $B(\varrho)/(\varrho-1) = (B(\varrho)-B(1))/(\varrho-1)$ is strictly decreasing and hence > 1. ☐

1.4. **The main result (Rademacher average approximation) and some consequences.** If $I \subseteq \mathbb{R}$ is an interval and E is a Banach space over $\mathbb{R}$ or $\mathbb{C}$, and with its norm denoted by $|\cdot|$ since the most interesting cases here are $E = \mathbb{R}$ and $E = \mathbb{C}$, then we use the standard notation $\mathcal{C}(I,E)$ for the continuous E-valued functions on I, $\mathcal{C}^m(I,E)$ for the ones $m \in \mathbb{N}_0$ times continuously differentiable, and $\mathcal{C}^{m,\alpha}(I,E)$ for those $f \in \mathcal{C}^m(I,E)$ whose m-th derivative $f^{(m)}$ has a finite Hölder constant

$$\|f^{(m)}\|_{\mathrm{L},\alpha} \;\coloneqq\; \sup_{x,\,y\in I,\,x\neq y} \frac{|f^{(m)}(x)-f^{(m)}(y)|}{|x-y|^\alpha} \tag{6}$$

of order $\alpha \in \left]0,1\right]$. It is well known that for E finite-dimensional, and also more generally as discussed in [16], the condition $f \in \mathcal{C}^{m,1}(I,E)$ is equivalent to $f^{(m)}$ being absolutely continuous with its then Lebesgue-almost everywhere existing derivative $f^{(m+1)}$ satisfying

$$\begin{aligned}\|f^{(m)}\|_{\mathrm{L}} \;&\coloneqq\; \|f^{(m)}\|_{\mathrm{L},1} \;=\; \|f^{(m+1)}\|_\infty \\ &\coloneqq\; \inf\{M\in\mathbb{R} : |f^{m+1}| \le M \text{ Lebesgue-almost everywhere on } I\}.\end{aligned}$$

Our main result is:

Theorem 1.2. *Let $n \in \mathbb{N}$, $P_1,\ldots,P_n \in \mathcal{P}_3$, E be a Banach space, and $f \in \mathcal{C}^{2,1}(\mathbb{R},E)$. Then we have*

$$\left|\widetilde{\mathop{\ast}_{i=1}^{n} P_i}\, f - \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\, f\right| \;\le\; \frac{\|f''\|_{\mathrm{L}}}{6}\sum_{i=1}^n \frac{\sigma_i^3}{\sigma^3} B(\varrho_i) \tag{7}$$

with $\sigma_i \coloneqq \sigma(P_i)$, $Q_i \coloneqq \frac{1}{2}(\delta_{-\sigma_i}+\delta_{\sigma_i})$, $\sigma \coloneqq \left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}$, and $\varrho_i \coloneqq \varrho(P_i)$. If each P_i is a two-point law and if the centred third moments of the P_i are all ≥ 0 or all ≤ 0, and if also $f(x) = cx^3$ for $x \in \mathbb{R}$, with a constant $c \in E$, then equality holds in (7).

Clearly, in the homoscedastic case of $\sigma_1 = \ldots = \sigma_n$, the approximating law $\widetilde{\ast_{i=1}^n Q_i}$ in Theorem 1.2 is just the standardized symmetric binomial law $\widetilde{\mathrm{B}_{n,\frac{1}{2}}}$. And in the i.i.d. case of $P_1 = \ldots = P_n =: P$, inequality (7) further simplifies to

$$\left|\widetilde{P^{*n}}f - \widetilde{\mathrm{B}_{n,\frac{1}{2}}}f\right| \;\le\; \frac{B(\varrho(P))}{6\sqrt{n}}\|f''\|_{\mathrm{L}}, \tag{8}$$

with equality whenever P is a two-point law and $f(x) = cx^3$.

Here are three examples of applications of Theorem 1.2, of which the first one, however, is a mock one.

Example 1.3. Theorem 1.2 formally yields [58, Theorem 6], namely

$$\max_{P\in\mathcal{P}_3,\varrho(P)=\varrho}\left|\int x^3\,\mathrm{d}\widetilde{P}(x)\right| \;=\; B(\varrho) \quad \text{for } \varrho \in [1,\infty[\tag{9}$$

with equality attained for two-point laws, by applying (8) with $E = \mathbb{R}$, $n = 1$, and $f(x) \coloneqq x^3$, since for $P \in \mathcal{P}_3$, we have

$$\left|\int x^3\,\mathrm{d}\widetilde{P}(x)\right| \;=\; \left|\widetilde{P}f - \widetilde{\mathrm{B}_{1,\frac{1}{2}}}f\right|$$

and $\|f''\|_{\mathrm{L}} = 6$. However, (9) is used in Step 6 of our proof of Theorem 1.2 given in section 5.

Example 1.4. In Theorem 1.2, let $E = \mathbb{C}$ and $f(x) = \mathrm{e}^{\mathrm{i}tx}$ for some $t \in \mathbb{R}$. Then, writing φ for the characteristic function of $\widetilde{\ast_{i=1}^n P_i}$, we get

$$\left|\varphi(t) - \prod_{i=1}^n \cos\left(\frac{\sigma_i t}{\sigma}\right)\right| \;\le\; \frac{|t|^3}{6}\sum_{i=1}^n \frac{\sigma_i^3 B(\varrho_i)}{\sigma^3}, \tag{10}$$

since here $\|f''\|_{\mathrm{L}} = \sup_{x\in\mathbb{R}} |f'''(x)| = |t|^3$. In (10), we have asymptotic equality for $t \to 0$ if all the P_i are two-point laws with equi-signed third centred moments, by equality in (7) for $f = (\cdot)^3$ and by a Taylor expansion inside the modulus on the left hand side of (10).

Moreover, using

$$0 \;\leq\; \prod_{i=1}^n \cos(t_i) - 1 + \frac{1}{2}\sum_{i=1}^n t_i^2 \;\leq\; \frac{1}{24}\sum_{i=1}^n t_i^4 + \frac{1}{4}\sum_{i<j} t_i^2 t_j^2 \quad \text{for } t \in \mathbb{R}^n, \tag{11}$$

which follows by rewriting the central term in (11) with the help of i.i.d. Rademacher variables $\xi_1, \ldots, \xi_n$ as

$$\prod_{i=1}^n \mathbb{E}\mathrm{e}^{\mathrm{i}t_i\xi_i} - 1 + \frac{1}{2}\sum_{i=1}^n t_i^2 \mathbb{E}\xi_i^2 \;=\; \mathbb{E}\Big(\cos\Big(\sum_{i=1}^n t_i\xi_i\Big) - 1 + \frac{1}{2}\Big(\sum_{i=1}^n t_i\xi_i\Big)^2\Big)$$

and applying $0 \leq \cos(x) - 1 + \frac{1}{2}x^2 \leq \frac{1}{24}x^4$ inside the last expectation above, we obtain from (10) the following estimate for the accuracy of the approximation of φ by the first terms of its Taylor expansion:

$$\left|\varphi(t) - 1 + \frac{t^2}{2}\right| \;\leq\; \frac{|t|^3}{6}\sum_{i=1}^n \frac{\sigma_i^3 B(\varrho_i)}{\sigma^3} + \frac{t^4}{24}\sum_{i=1}^n \frac{\sigma_i^4}{\sigma^4} + \frac{t^4}{4}\sum_{i<j}\frac{\sigma_i^2\sigma_j^2}{\sigma^4} \quad \text{for } t \in \mathbb{R}.$$

In particular, with $n = 1$ we have

$$\left|\mathbb{E}\mathrm{e}^{\mathrm{i}tX} - 1 + \frac{t^2}{2}\right| \;\leq\; A(\varrho)\frac{\varrho|t|^3}{6} + \frac{t^4}{24} \tag{12}$$

for all $t \in \mathbb{R}$ and an arbitrary r.v. X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$, $\varrho := \mathbb{E}|X|^3 < \infty$, where the inequality turns into the asymptotic equality as $t \to 0$ whenever X is a two-point r.v.

Inequality (12) for small t improves the bound

$$\left|\mathbb{E}\mathrm{e}^{\mathrm{i}tX} - 1 + \frac{t^2}{2}\right| \;\leq\; \frac{\varrho|t|^3}{6}\inf_{0<\lambda<1/2}\{\lambda A(\varrho) + q_3(\lambda)\}$$

obtained in [58, Corollary 4], where

$$q_3(\lambda) \;:=\; \sup_{x>0}\frac{6}{x^3}\left|\mathrm{e}^{\mathrm{i}x} - 1 - \mathrm{i}x - \frac{(\mathrm{i}x)^2}{2} - \lambda\frac{(\mathrm{i}x)^3}{6}\right| \;\geq\; 1 - \lambda \quad \text{for } 0 \leq \lambda \leq 1/2,$$

with the final inequality following from considering $x \downarrow 0$. Indeed, for every $\varrho \geq 1$, we have $A(\varrho) < 1$ by Lemma 1.1 and hence get

$$\inf_{0<\lambda<1/2}\{\lambda A(\varrho) + q_3(\lambda)\} \;\geq\; \inf_{0<\lambda<1/2}\{\lambda A(\varrho) + 1 - \lambda\} \;=\; \frac{A(\varrho)+1}{2} \;>\; A(\varrho).$$

Example 1.5. Applying Theorem 1.2 to $E = \mathbb{R}$ and $f = |\cdot|^3$ in the i.i.d. case yields: For i.i.d. $X_i \sim P \in \mathcal{P}_3$, we have

$$\left|\mathbb{E}\left|\widetilde{\textstyle\sum_{i=1}^n X_i}\right|^3 - \widetilde{\mathrm{B}_{n,\frac{1}{2}}}|\cdot|^3\right| \;\leq\; \frac{B(\varrho(P))}{\sqrt{n}}, \tag{13}$$

by $\|f''\|_{\mathrm{L}} = 6$, where by formula (86) stated and proved below, we have explicitly

$$\widetilde{\mathrm{B}_{n,\frac{1}{2}}}|\cdot|^3 \;=\; \begin{cases} \left(2n^{\frac{1}{2}} + n^{-\frac{1}{2}} - n^{-\frac{3}{2}}\right)\mathrm{b}_{n,\frac{1}{2}}(\lfloor\frac{n}{2}\rfloor) & \text{if } n \text{ is odd,} \\ 2n^{\frac{1}{2}}\mathrm{b}_{n,\frac{1}{2}}(\frac{n}{2}) & \text{if } n \text{ is even.} \end{cases}$$

Let us note that $\widetilde{\mathrm{B}_{n,\frac{1}{2}}|\cdot|^3}$ can not be replaced by any other function of n without invalidating (13), since the R.H.S. of (13) is zero if the X_i are symmetrically Bernoulli-distributed; an analogous remark applies to every application of Theorem 1.2 in the i.i.d. case.

In Theorem 1.8 below, we rewrite Theorem 1.2 in terms of Zolotarev's distance ζ_3. On the one hand this actually prepares for the proof of Theorem 1.2. On the other hand it allows, by simply using the triangle inequality combined with Theorem 1.9 below, to obtain the quite sharp normal approximation result in Theorem 1.10. Since in turn the proof of Theorem 1.9 uses ζ_4, let us recall here the definition and some basic and well-known properties of ζ_s in general. For more properties of Zolotarev distances needed in the present paper, including new results as well as apparently previously unpublished detailed proofs of some "well-known" results, we refer to section 4 below. Standard references on ζ-distances include the monographs [68], [47], and [52].

We will use the notation introduced around (6), here with $I = E = \mathbb{R}$.

Definition 1.6 (ζ-distances)**.** Let $s > 0$. With $m := \lceil s-1 \rceil \in \mathbb{N}_0$ and $\alpha := s - m \in\,]0,1]$, we put

$$\mathcal{F}_s \;:=\; \{f \in \mathcal{C}^{m,\alpha}(\mathbb{R},\mathbb{R}) : \|f^{(m)}\|_{\mathrm{L},\alpha} \le 1\}, \qquad \mathcal{F}_s^\infty \;:=\; \{f \in \mathcal{F}_s : f \text{ bounded}\}.$$

For $P, Q \in \mathrm{Prob}(\mathbb{R})$ then

$$\zeta_s(P,Q) \;:=\; \sup_{f\in\mathcal{F}_s^\infty} |Pf - Qf| \tag{14}$$

is called the *Zolotarev distance of order s* from P to Q, and one further defines a weighted variation distance as

$$\nu_s(P,Q) \;:=\; \int |x|^s \,\mathrm{d}|P-Q|(x).$$

Let us note that in [68, p. 44] and [52, p. 100], our $\mathcal{F}_s^\infty$ is denoted by $\mathcal{F}_s$, and that in these books our $\mathcal{F}_s$ is implicitly used without any convenient notation. The latter may have led to some of the clearly existing confusion in the literature. For example, one finds in several publications, usually obscured by employing random variable notation, in effect the definition (14) with $\mathcal{F}_s$ in place of $\mathcal{F}_s^\infty$, which makes sense, and then no difference by the apparently not completely trivial Theorem 1.7(d) below, iff $P, Q \in \mathrm{Prob}_s(\mathbb{R})$. As a recent example of such an unclear "definition" without assuming $P, Q \in \mathrm{Prob}_s(\mathbb{R})$, we can mention [40, (8), the case of $s = 1$, $\mu = \nu$ the standard Cauchy law, once $Y = X$ and once $Y = -X$, f the identity] where, however, the error is immediately admitted.

Theorem 1.7 (Well-known facts about ζ_s)**.** *Let $s = m + \alpha$ be as in Definition 1.6.*

(a) *For $P, Q \in \mathrm{Prob}(\mathbb{R})$, the value of $\zeta_s(P,Q)$ does not change if in the definition of $\mathcal{F}_s$ the functions f are assumed to be E-valued rather than $\mathbb{R}$-valued, with E any Banach space not degenerated to one point.*

(b) *On $\mathrm{Prob}(\mathbb{R})$, ζ_s is an extended metric, that is, a metric except that it may also assume the value ∞.*

(c) *For $P \in \mathrm{Prob}(\mathbb{R})$ and $Q \in \mathrm{Prob}_s(\mathbb{R})$, we have the equivalence chain*

$$\begin{aligned} \zeta_s(P,Q) < \infty \;&\Leftrightarrow\; P \in \mathrm{Prob}_s(\mathbb{R}) \text{ and } \mu_j(P) = \mu_j(Q) \text{ for } j \in \{1,\ldots,m\} \\ &\Leftrightarrow\; P \in \mathrm{Prob}_s(\mathbb{R}) \text{ and } \zeta_s(P,Q) \le \frac{\Gamma(1+\alpha)}{\Gamma(1+s)}\nu_s(P,Q). \end{aligned} \tag{15}$$

Hence, if $c_1, \ldots, c_m \in \mathbb{R}$ *are given, then* ζ_s *is a metric on the (possibly empty) set* $\{P \in \mathrm{Prob}_s(\mathbb{R}) : \mu_j(P) = c_j$ *for* $j \in \{1, \ldots, m\}\}$. *In particular,* ζ_3 *is a metric on* $\widetilde{\mathcal{P}_3}$.

(d) *Let* $P, Q \in \mathrm{Prob}_s(\mathbb{R})$. *Then we may omit the boundedness condition on* f *in the definition* (14)*, that is, we have*

$$\zeta_s(P,Q) \;=\; \sup_{f\in\mathcal{F}_s} |Pf - Qf|, \tag{16}$$

and we further have

$$|Pf - Qf| \;\leq\; \|f^{(m)}\|_{\mathrm{L},\alpha}\, \zeta_s(P,Q) \qquad \text{for } f \in \mathcal{C}^{m,\alpha}(\mathbb{R},\mathbb{R}). \tag{17}$$

References or proofs for Theorem 1.7 are given in section 4, together with further facts about ζ_s. With the above preparations, we can state:

Theorem 1.8 (essentially Theorem 1.2 rewritten)**.** *Let* $n \in \mathbb{N}$ *and* $\sigma, P_i, Q_i, \varrho_i$ *for* $i \in \{1, \ldots, n\}$ *be as in Theorem 1.2. Then we have*

$$\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}\;,\; \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\right) \;\leq\; \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3 B(\varrho_i), \tag{18}$$

with equality whenever each P_i *is a two-point law and also the centred third moments of the* P_i *are all* ≥ 0 *or all* ≤ 0.

Indeed, if Theorem 1.2 is assumed to be true, then applying the definition of ζ_3 immediately yields inequality (18), and using also (16) from Theorem 1.7(d) yields the accompanying equality statement. Conversely, if (18) is proved, then, using (17), we get Theorem 1.2 in the case of $E = \mathbb{R}$ and except for the equality statement.

1.5. **Normal approximation.** Coming now to the normal approximation results following from Theorem 1.8, let us first consider in Theorem 1.10 below the i.i.d. case. There

$$\varepsilon_n := \zeta_3\left(\widetilde{\mathrm{B}_{n,\frac{1}{2}}}, \mathrm{N}\right) \qquad \text{for } n \in \mathbb{N} \tag{19}$$

plays the role of a higher order error term, as is made explicit by the following auxiliary result.

Theorem 1.9. *For* $n \in \mathbb{N}$*, we have, with the first equality to be read from right to left due to the* $O(n^{-2})$,

$$\begin{aligned}
\frac{1}{6\sqrt{2\pi}\,n} + O\left(\frac{1}{n^2}\right) \;&=\; \frac{1}{6}\left\{\begin{array}{ll} \left|\left(2n^{\frac{1}{2}} + n^{-\frac{1}{2}} - n^{-\frac{3}{2}}\right)\mathrm{b}_{n,\frac{1}{2}}(\lfloor\frac{n}{2}\rfloor) - \frac{4}{\sqrt{2\pi}}\right| & \text{if } n \text{ is odd,} \\ \left|2n^{\frac{1}{2}}\mathrm{b}_{n,\frac{1}{2}}(\frac{n}{2}) - \frac{4}{\sqrt{2\pi}}\right| & \text{if } n \text{ is even} \end{array}\right\} \\
&=\; \left|\left(\mathrm{N} - \widetilde{\mathrm{B}_{n,\frac{1}{2}}}\right)\frac{|\cdot|^3}{6}\right| \;\leq\; \varepsilon_n \\
&<\; \frac{1}{3\sqrt{2\pi}\,n} + \left(\frac{4+\zeta(\frac{1}{2})}{\sqrt{2\pi}} - 1\right)\frac{1}{6n^{3/2}} \\
&<\; \frac{0.1330}{n} + \frac{0.0022}{n^{3/2}} \;\leq\; \frac{0.1352}{n},
\end{aligned} \tag{20}$$

where $\zeta(\cdot)$ *is the Riemann zeta-function, in particular* $\zeta(\frac{1}{2}) = -1.4603\ldots$.

The above lower bound for ε_n holds even with equality in case of $n = 1$, by Example 4.3 below, and we conjecture that, in the general case, it is at least asymptotically exact.

Theorem 1.10. *For $P \in \mathcal{P}_3$ and $n \in \mathbb{N}$, we have*

$$\zeta_3\left(\widetilde{P^{*n}}, \mathrm{N}\right) \quad \leq \quad \frac{B(\varrho(P))}{6\sqrt{n}} + \varepsilon_n, \tag{21}$$

where, on the right, the leading term for $n \to \infty$ is optimal in the sense of

$$\frac{B(\varrho)}{6} \quad = \quad \sup_{P \in \mathcal{P}_3, \varrho(P)=\varrho} \ \sup_{f \in \mathcal{F}_3} \lim_{n\to\infty} \sqrt{n}\left|\widetilde{P^{*n}} f - \mathrm{N} f\right| \qquad \text{for } \varrho \in [1,\infty[, \tag{22}$$

and the leading term for $\varrho \to 1+$ is asymptotically exact in the sense of

$$\varepsilon_n \quad = \quad \lim_{\varrho \to 1+} \ \inf_{P \in \mathcal{P}_3, \varrho(P)=\varrho} \zeta_3\left(\widetilde{P^{*n}}, \mathrm{N}\right) \qquad \text{for } n \in \mathbb{N}. \tag{23}$$

Remark 1.11. In (22) we can also interchange the limit with the supremum over $f \in \mathcal{F}_3$ producing the definition of ζ_3, namely, we have

$$\frac{B(\varrho)}{6} \quad = \quad \sup_{P \in \mathcal{P}_3, \varrho(P)=\varrho} \lim_{n\to\infty} \sqrt{n}\, \zeta_3\left(\widetilde{P^{*n}}, \mathrm{N}\right) \qquad \text{for } \varrho \in [1,\infty[.$$

In fact, every permutation of the two suprema and the limit in (22) can be used.

Remark 1.12. Inequality (21) often improves Tyurin's estimate [61, 62, 63] (with [61] actually being the final one among the three papers)

$$\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}, \mathrm{N}\right) \quad \leq \quad \frac{1}{6\sigma^3}\sum_{i=1}^{n} \sigma_i^3 \varrho_i \quad \text{for } P_1 \ldots, P_n \in \mathcal{P}_3 \tag{24}$$

in the i.i.d. case, where the latter takes the form

$$\zeta_3(\widetilde{P^{*n}}, \mathrm{N}) \quad \leq \quad \frac{\varrho(P)}{6\sqrt{n}} \qquad \text{for } P \in \mathcal{P}_3 \text{ and } n \in \mathbb{N} \tag{25}$$

and is optimal in the sense that the factor $1/6$ cannot be made less if $\varrho(P)$ is allowed to be arbitrarily large. Indeed, in view of $B(\varrho) < \varrho$ and $\varepsilon_n = O(n^{-1})$, inequality (21) improves (25) for *every* value of $\varrho \geq 1$ and every sufficiently large $n \in \mathbb{N}$, namely, iff

$$6\sqrt{n}\,\varepsilon_n \quad < \quad \varrho - B(\varrho), \tag{26}$$

which is surely true for

$$n \quad \geq \quad \left(\frac{6 \cdot 0.1352}{\varrho - B(\varrho)}\right)^2 \quad = \quad \frac{0.65804\ldots}{(\varrho - B(\varrho))^2}.$$

Here is a table of the values of ϱ and n satisfying condition (26), where, for convenience, we also provide values of $B(\varrho)$ rounded up:

$\varrho \leq$	1.01	1.10	1.18	1.25	1.31	1.52	1.78	2.18	2.41	2.52
$B(\varrho) \leq$	0.17	0.53	0.72	0.85	0.95	1.27	1.60	2.07	2.32	2.44
$n \geq$	1	2	3	4	5	10	20	50	80	100

Example 1.13. Let P be an exponential distribution. Then $\varrho = 12e^{-1} - 2 = 2.4145\ldots$, $B(\varrho) = 2.3248\ldots$, and condition (26) holds surely for $n \geq 80$.

If P is a uniform distribution, then $\varrho = 3\sqrt{3}/4 = 1.2990\ldots$, $B(\varrho) = 0.9302\ldots$, and condition (26) holds surely for $n \geq 5$.

If P is the Bernoulli distribution with parameter $p \in]0, \frac{1}{2}]$, then, denoting $q := 1-p$, we have $\varrho(P) = (p^2+q^2)/\sqrt{pq}$, $B(\varrho) = (q-p)/\sqrt{pq}$, $\varrho - B(\varrho) = 2p\sqrt{p/q}$, and condition (26) holds for $n \geq 1$ if $p \geq 0.45$, $n \geq 2$ if $p \geq 0.38$, $n \geq 3$ if $p \geq 0.34$, $n \geq 4$ if $p \geq 0.31$, $n \geq 17$

if $p \geq 0.2$, and $n \geq 144$ if $p \geq 0.1$. In particular, in the symmetric case ($p = 1/2$) our bound (21) is sharper than (25) for every $n \in \mathbb{N}$.

If P is the Poisson distribution with parameter $\lambda > 0$, then:
for $\lambda = 1$ we have $\varrho = 1.7357\ldots$, $B(\varrho) = 1.5448\ldots$, and (26) holds for $n \geq 18$;
for $\lambda = 2$ we have $\varrho = 1.6640\ldots$, $B(\varrho) = 1.4543\ldots$, and (26) holds for $n \geq 15$;
for $\lambda = 3$ we have $\varrho = 1.6408\ldots$, $B(\varrho) = 1.4244\ldots$, and (26) holds for $n \geq 14$;
for $\lambda = 6$ we have $\varrho = 1.6181\ldots$, $B(\varrho) = 1.3948\ldots$, and (26) holds for $n \geq 13$.

If P is the geometric distribution with $P_i(\{k\}) = p(1-p)^k$ for $k = 0, 1, 2, \ldots$, then, with $p = 0.1$, we have $\varrho = 2.4158\ldots$, $B(\varrho) = 2.3262\ldots$, and (26) holds for $n \geq 80$.

Now we present extensions of some of the above results to the non-i.i.d. case.

Theorem 1.14. *For $P_i, Q_i, \varrho_i, \sigma_i, \sigma$ as in Theorem 1.2 we have*

$$\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} P_i},\, \mathrm{N}\right) \;\leq\; \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3 B(\varrho_i) + \zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} Q_i},\, \mathrm{N}\right). \tag{27}$$

Further, if $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$, then

$$\begin{aligned}\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} Q_i},\, \mathrm{N}\right) &\leq \frac{1}{6}\left(2\sqrt{\frac{2}{\pi}} - 1\right)\frac{\sigma_1^3}{\sigma^3} + \frac{1}{6\sqrt{2\pi}}\sum_{k=1}^{n-1}\frac{\sigma_{k+1}^3 \min\{1, \sqrt{n}\,\sigma_{k+1}/\sigma\}}{\sigma^3\sqrt{k}} \\ &\leq 0.0993 \cdot \frac{\sigma_1^3}{\sigma^3} + 0.0665\sum_{k=1}^{n-1}\frac{\sigma_{k+1}^3}{\sigma^3\sqrt{k}}.\end{aligned}$$

Remark 1.15. Inequality (27) improves Tyurin's already optimal bound (24) iff

$$\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} Q_i},\, \mathrm{N}\right) \;<\; \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3\left(\varrho_i - B(\varrho_i)\right).$$

Thus, as already indicated at the end of subsection 1.1, Theorems 1.2 and 1.14 can be regarded as extensions of the results previously obtained in [55, 56, 54, 57] for the uniform metric to ζ_3-metric, so that the bounds (21) and (27) can be called *asymptotic estimates with an optimal structure.*

2. Auxiliary analytic results

2.1. Two-point Hermite interpolation, and approximation in $\mathcal{F}_s$.

The purpose of Lemma 2.1 is to prepare through its parts (c) and (d) for a proof of Lemma 2.2, which in turn is used in section 4 below in our proof of Theorem 1.7.

Lemma 2.1 (On two-point Hermite interpolation polynomials)**.** *Let $m_0, m_1 \in \mathbb{N}_0$, $V_i := \mathbb{R}^{\{0,\ldots,m_i\}}$ for $i \in \{0,1\}$, and $V := V_0 \times V_1$. For distinct $x_0, x_1 \in \mathbb{R}$ and for $y = (y_0, y_1) = \left((y_{0,j})_{j=0}^{m_0}, (y_{1,j})_{j=0}^{m_1}\right) \in V$, let $p = p_{x_0,x_1,y} = p_{x_0,x_1,y_0,y_1}$ denote the Hermite interpolation polynomial defined by being a polynomial of degree at most m_0+m_1+1 and satisfying the condition*

$$p^{(j)}(x_i) \;=\; y_{i,j} \quad \text{for } i \in \{0,1\} \text{ and } j \in \{0,\ldots,m_i\}. \tag{28}$$

(a) Linearity. *Given distinct $x_0, x_1 \in \mathbb{R}$, the map $V \ni y \mapsto p_{x_0,x_1,y}$ is linear with respect to the obvious vector space structures; in particular we have $p_{x_0,x_1,y_0,y_1} = p_{x_0,x_1,y_0,0} + p_{x_0,x_1,0,y_1} = p_{x_0,x_1,y_0,0} + p_{x_1,x_0,y_1,0}$ for $y_0 \in V_0$ and $y_1 \in V_1$.*

(b) Change of variables. *For $y\in V$ and distinct $x_0,x_1\in\mathbb{R}$, we have*

$$p_{x_0,x_1,y}(x) \;=\; p_{0,1,z}\Big(\frac{x-x_0}{x_1-x_0}\Big) \qquad \text{for } x\in\mathbb{R}$$

with $z\in V$ defined by $z_{i,j}:=(x_1-x_0)^j y_{i,j}$ for $i\in\{0,1\}$ and $j\in\{0,\ldots,m_i\}$.

(c) Positivity. *Let $-\infty<x_0<x_1<\infty$ and let $(y_0,y_1)\in V$ satisfy*

$$y_{0,j} \;\geq\; 0 \text{ for } j\in\{0,\ldots,m_0\}, \qquad (-1)^j y_{1,j} \;\geq\; 0 \text{ for } j\in\{0,\ldots,m_1\}. \tag{29}$$

Then either $p>0$ on $]x_0,x_1[$, or $y_0=0$, $y_1=0$, $p=0$.

(d) Bounds. *Let $\|\cdot\|$ be a norm on the vector space V. Then there exists a constant $c=c_{\|\cdot\|}\in\,]0,\infty[$ such that the following holds: If $y\in V$ and if $-\infty<x_0<x_1<\infty$, then*

$$\sup_{x\in[x_0,x_1]}\Big|p^{(k)}_{x_0,x_1,y}(x)\Big| \;\leq\; c\,\|y\|\,\frac{1\vee|x_1-x_0|^{m_0\vee m_1}}{|x_1-x_0|^k} \qquad \text{for } k\in\mathbb{N}_0. \tag{30}$$

Proof. The existence and uniqueness of p are well-known, and easily imply (a) and (b).

(c) By (a) and (b), the latter applied to $p_{x_0,x_1,y_0,0}$ and also to $p_{x_1,x_0,y_1,0}$, we may assume that we have $x_0=0$, $x_1=1$, $y_1=0$. Then the case of $y_0=0$ is trivial, and so we assume from now on that at least one coordinate of y_0 is even strictly positive, and we put

$$k \;:=\; \max\{j\in\{0,\ldots,m_0\}: y_{0,j}>0\}.$$

We then have

$$p^{(j)}(x) \;>\; 0 \qquad \text{for } x>0 \text{ sufficiently close to } 0 \tag{31}$$

for $j\in\{0,\ldots,k\}$.

Assume from now on, to get a contradiction, that we do not have $p>0$ on $]0,1[$. Then, by (31) with $j=0$ and by the intermediate value theorem, we have $p(\xi)=0$ for some $\xi\in\,]0,1[$. Hence, understanding "n zeros" to mean "at least n zeros, counting multiplicity" in this proof, $p=p^{(0)}$ has $1+(m_1+1)=m_1+2$ zeros in $]0,1]$, namely one zero at ξ and m_1+1 zeros at 1.

If now $k\geq 1$ and if $j\in\{0,\ldots,k-1\}$ is such that $p^{(j)}$ has m_1+2 zeros in $]0,1]$, then there is an $\eta=\eta_j\in\,]0,1]$ with $p^{(j)}(\eta)=0$ and such that $p^{(j)}$ has m_1+2 zeros in $[\eta,1]$, and then (31) with $j+1$ in place of j together with $p^{(j)}(0)\geq 0$ implies that the maximum of $p^{(j)}$ over $[0,\eta]$ is attained at a point in $]0,\eta[$, and hence, in addition applying Rolle's theorem on $[\eta,1]$, we conclude that $p^{(j+1)}$ has $1+(m_1+2-1)=m_1+2$ zeros in $]0,1]$.

The preceding two paragraphs yield that $p^{(k)}$ has m_1+2 zeros in $]0,1]$, and we have $p^{(k+1)}(0)=\ldots=p^{(m_0)}(0)=0$, with the latter condition of course being empty if $k=m_0$. Hence $p^{(k+1)}$ has $(m_0-k)+(m_1+2-1)=m_0+m_1+1-k$ zeros in [0,1] and is of degree at most $m_0+m_1+1-(k+1)=m_0+m_1-k$, so we have $p^{(k+1)}=0$ and hence p of degree at most $k\leq m_0$, yielding $p(\xi)=\sum_{j=0}^{m_0}y_{0,j}\xi^j/j!>0$, a contradiction.

(d) If $k\geq m_0+m_1+2$, then $p^{(k)}=0$, and then (30) is trivially true even with $c=0$; hence we may assume that $k\in\{0,\ldots,m_0+m_1+1\}$ is fixed in this proof. Using finite-dimensionality of V, we may further assume that $\|\cdot\|=\|\cdot\|_\infty$, that is, $\|y\|=\max_{i,j}|y_{i,j}|$ for $y\in V$, see e.g. [51, pp. 192, 175]. Given now y and x_0,x_1 as in the claim, we apply (b) with z as defined there to get

$$\sup_{x\in[x_0,x_1]}\Big|p^{(k)}_{x_0,x_1,y}(x)\Big| \;=\; \sup_{x\in[x_0,x_1]}\Big|\tfrac{1}{(x_1-x_0)^k}p^{(k)}_{0,1,z}\Big(\tfrac{x-x_0}{x_1-x_0}\Big)\Big| \;\leq\; \tfrac{c}{(x_1-x_0)^k}\|z\|_\infty \;\leq\; \text{R.H.S.(30)},$$

where c denotes the norm of the linear map $V \ni z \mapsto p^{(k)}_{0,1,z}|_{[0,1]} \in \mathcal{C}([0,1],\mathbb{R})$, with respect to the supremum norms on the two vector spaces, and $c < \infty$ by finite-dimensionality of V again, see e.g. [51, p. 279]. □

We recall the definitions of $\mathcal{F}_s^\infty$ and $\mathcal{F}_s$ from Definition 1.6.

Lemma 2.2 (Denseness of $\mathcal{F}_s^\infty$ in $\mathcal{F}_s$)**.** *Let $s \in]0,\infty[$ and $f \in \mathcal{F}_s$. Then there exist a sequence (f_n) in $\mathcal{F}_s^\infty$ and constants $a,b \in [0,\infty[$ with $f_n \to f$ pointwise and $|f_n| \le a+b|\cdot|^s$ for $n \in \mathbb{N}$. If $f = c|\cdot|^s$ with $c \ge 0$, then (f_n) can be chosen to satisfy also $f_n \ge 0$ for $n \in \mathbb{N}$.*

Proof. Let $m \in \mathbb{N}_0$ and $\alpha \in]0,1]$ with $s = m+\alpha$. We will use the notation of Lemma 2.1 with $m_1 := m_2 := m$.

Let $n \in \mathbb{N}$. We define $y \in V = \mathbb{R}^{\{0,\ldots,m\}} \times \mathbb{R}^{\{0,\ldots,m\}}$ by $y_{0,j} := \frac{n-1}{n} f^{(j)}(n)$ and $y_{1,j} := 0$ for $j \in \{0,\ldots,m\}$, and we then apply Lemma 2.1(d) with $k := m+1$, $x_0 := n$, and $x_1 := b_n$ with $b_n \ge n+1$ chosen so large that we have $c\|y\|(b_n - n)^{-\alpha} \le \frac{1}{2n}$ and hence, by (30), so that $p_n := p_{n,b_n,y}$ satisfies

$$\left|p_n^{(m+1)}(x)\right| \;\le\; \tfrac{1}{2n}(b_n - n)^{\alpha-1} \qquad \text{for } x \in [n,b_n]. \tag{32}$$

We analogously choose $a_n \le -n-1$ with $|a_n|$ so large that the polynomial q_n of degree at most $2m+1$ and with $q_n^{(j)}(a_n) = 0$ and $q_n^{(j)}(-n) = \frac{n-1}{n} f^{(j)}(-n)$ for $j \in \{0,\ldots,m\}$ satisfies

$$\left|q_n^{(m+1)}(x)\right| \;\le\; \tfrac{1}{2n}(-n - a_n)^{\alpha-1} \qquad \text{for } x \in [a_n,-n]. \tag{33}$$

We finally put, using the de Finetti notation introduced in subsection 1.2,

$$f_n(x) \;:=\; (a_n \le x \le -n)q_n(x) + (|x| < n)\tfrac{n-1}{n} f(x) + (n \le x \le b_n)p_n(x) \qquad \text{for } x \in \mathbb{R}.$$

Then $f_n \in \mathcal{C}^m(\mathbb{R},\mathbb{R})$ and f_n is bounded. Thus to get $f_n \in \mathcal{F}_s^\infty$, it remains to prove that

$$\sup_{u,\,v\in\mathbb{R},\,u<v} \frac{|f_n^{(m)}(v) - f_n^{(m)}(u)|}{|v-u|^\alpha} \;\le\; 1. \tag{34}$$

So let $-\infty < u < v < \infty$, and let us abbreviate $g := f_n^{(m)}$. Then $g(u) = g(u \vee a_n)$ and $g(v) = g(v \wedge b_n)$ and hence $|g(v)-g(u)|/|v-u|^\alpha \le |g(v\wedge b_n) - g(u \vee a_n)|/|v\wedge b_n - u \vee a_n|^\alpha$, and so we may assume $a_n \le u$ and $v \le b_n$. In the case of $a_n \le u \le -n$ and $n \le v \le b_n$, we use in the second step below (32) and (33), and also (34) with f in place of f_n, to get

$$\begin{aligned}
|g(v) - g(u)| \;&\le\; |g(v) - g(n)| + |g(n) - g(-n)| + |g(-n) - g(u)| \\
&\le\; \tfrac{1}{2n}(b_n - n)^{\alpha-1}|v-n| + \tfrac{n-1}{n}|n-(-n)|^\alpha + \tfrac{1}{2n}(-n-a_n)^{\alpha-1}|-n-u| \\
&\le\; \tfrac{1}{2n}|v-n|^\alpha + \tfrac{n-1}{n}|n-(-n)|^\alpha + \tfrac{1}{2n}|-n-u|^\alpha \\
&\le\; |v-n|^\alpha \vee |n-(-n)|^\alpha \vee |-n-u|^\alpha \\
&\le\; |v-u|^\alpha.
\end{aligned}$$

The remaining cases needed to prove (34) are similar or simpler.

Obviously, $f_n \to f$ pointwise. Further, by Lemma 2.1(c), we have $f_n \ge 0$ in case of $f = c|\cdot|^s$ with $c \ge 0$.

Let $g \in \mathcal{F}_s$. If $s \le 1$, then we have $|g(x) - g(0)| \le |x|^s$ and hence $|g| \le a + b|\cdot|^s$ for $a := g(0)$ and $b := 1$. If $s > 1$, then we have for $x \in \mathbb{R}$ the Taylor formula

$$g(x) \;=\; \sum_{j=0}^{m-1} \frac{g^{(j)}(0)}{j!} x^j + \int_0^1 \frac{(1-\lambda)^{m-1}}{(m-1)!} g^{(m)}(\lambda x) x^m \,\mathrm{d}\lambda \tag{35}$$

and get $|g(x)| \leq \sum_{j=0}^{m-1} c_j |x|^j + \int_0^1 \frac{(1-\lambda)^{m-1}}{(m-1)!} \left(|g^{(m)}(0)| + |x|^\alpha\right) |x|^m \,\mathrm{d}\lambda \leq a + b|x|^s$ for certain constants c_j and a, b depending only on the availability of bounds for the derivatives up to the order m of g at zero. Hence, by the construction of the sequence (f_n), we have constants a, b with $|f_n| \leq a + b|\cdot|^s$ for each n. □

2.2. **On some special osculatory interpolations and a moment inequality.** Here our goal is the elementary Lemma 2.4, whose trivial consequence Lemma 2.5 is used in the final Step 7 of the proof of Theorem 1.2 in section 5. As for the title of the present subsection, recall that a function f is called first order osculatory at a point x_0 to a function g if we have $f(x_0) = g(x_0)$ and $f'(x_0) = g'(x_0)$.

Let $I \subseteq \mathbb{R}$ be a nondegenerate interval and $s \in \mathbb{N}_0$. Then, following here closely [44], a function $f\colon I \to \mathbb{R}$ is said to be s-convex on I iff for every choice of $s+1$ pairwise distinct points $x_0, \ldots, x_s \in I$ the $(s+1)$-st divided difference $[x_0, x_1, \ldots, x_s; f]$ is positive (recall that "positive" means ≥ 0, see subsection 1.2). This divided difference may be defined as

$$[x_0, x_1, \ldots, x_s; f] := \frac{U(x_0, \ldots, x_s; f)}{V(x_0, \ldots, x_s)},$$

where

$$U(x_0, \ldots, x_s; f) := \begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_0 & x_1 & \ldots & x_s \\ \vdots & \vdots & \ddots & \vdots \\ x_0^{s-1} & x_1^{s-1} & \ldots & x_s^{s-1} \\ f(x_0) & f(x_1) & \ldots & f(x_s) \end{vmatrix},$$

$V(x_0, \ldots, x_s) := U(x_0, \ldots, x_s; (\cdot)^s) = \prod_{i<j}(x_j - x_i)$ is the Vandermonde determinant. Alternatively one can set [34, Chapter 15]

$$[x; f] = f(x), \quad [x_0, x_1, \ldots, x_k; f] = \frac{[x_1, \ldots, x_k; f] - [x_0, \ldots, x_{k-1}; f]}{x_k - x_0} \quad \text{for } k \in \{1, \ldots, s\}.$$

As $V(x_0, \ldots, x_s) > 0$ for $x_0 < x_1 < \ldots < x_s$, a function f is s-convex on I iff we have $U(x_0, \ldots, x_s; f) \geq 0$ for all $x_0 < x_1 < \ldots < x_s \in I$. Thus, from the definition it immediately follows that a function is 0-convex iff it is nonnegative, 1-convex iff it is nondecreasing, and 2-convex iff it is convex in the usual sense. Higher order convexity was first considered by Hopf in his dissertation [29] and was further extensively developed by Popoviciu in his thesis [45].

If $P(x_1, \ldots, x_s; f|\cdot)$ is the unique Lagrange polynomial of degree at most $s-1$ that interpolates f at the points $x_1 < x_2 < \ldots < x_s$, then [45], [34, Chapter 15]

$$f(x) - P(x_1, \ldots, x_s; f|x) = \frac{U(x_1, \ldots, x_s, x; f)}{V(x_1, \ldots, x_s)} = [x_1, \ldots, x_s, x; f] \prod_{i=1}^{s}(x - x_i),$$

and thus f is s-convex on I iff for every choice of $-\infty =: x_0 < x_1 < \ldots < x_s < x_{s+1} := +\infty$ we have

$$(-1)^{i+s}(f(x) - P(x_1, \ldots, x_s; f|x)) \geq 0 \quad \text{for } i \in \{0, \ldots, s\},\ x \in \,]x_i, x_{i+1}[\, \cap I.$$

If $s \geq 2$, then a continuous function f is s-convex on I iff on the interior of I the derivative $f^{(s-2)}$ exists and is convex [29, 45, 34]. If f is s times differentiable on I, then f is s-convex iff $f^{(s)} \geq 0$ on I [45, 34].

Lemma 2.3. *Let $I \subseteq \mathbb{R}$ be an interval, $s, t \in I$ with $s \neq t$, and $f : I \to \mathbb{R}$ twice differentiable with*

$$f(s) \;=\; f'(s) \;=\; f(t) \;=\; f'(t) \;=\; 0 \tag{36}$$

and f'' convex on I. Then we have $f \geq 0$ on I. If further $u \in I \setminus \{s,t\}$ satisfies $f(u) = 0$, then we have $f = 0$ on the convex hull of $\{s,t,u\}$.

Proof. The existence of f'' and its convexity yield the 4-convexity of f; hence for every choice of $t_1, t_2, t_3, t_4 \in I$ with $t_1 < t_2 < t_3 < t_4$, the Lagrange interpolation polynomial p of degree ≤ 3 with $p(t_j) = f(t_j)$ for each j satisfies for $x \in I$ respectively $f(x) \geq p(x)$ if $t_4 \leq x$ or $t_2 \leq x \leq t_3$ or $x \leq t_1$, and $f(x) \leq p(x)$ if $t_3 \leq x \leq t_4$ or $t_1 \leq x \leq t_2$. This continues to hold if some, but not all, of the t_j coincide and p is accordingly the corresponding Hermite interpolation polynomial, in view of the continuous dependence of the latter on (t_1, t_2, t_3, t_4) due to the continuity of f'', compare [14, p. 119, Theorem 6.3].

To prove now the lemma, we may assume $s < t$. Assumption (36) says that $p := 0$ is the Hermite interpolation polynomial of degree ≤ 3 for f and the nodes $t_1 := t_2 := s$ and $t_3 := t_4 := t$, and hence we get $f \geq 0$ on I. If further u is as stated, then we prove also $f \leq 0$ on the convex hull of $\{s,t,u\}$, by applying the previous paragraph to $p := 0$, but now with $(t_1, t_2, t_3, t_4) := (u, s, s, t)$ if $u < s$, $:= (s, u, u, t)$ if $s < u < t$, using that then also $f'(u) = 0$ due to $f \geq 0$ and $f(u) = 0$, and finally $:= (s, t, t, u)$ if $t < u$. □

Lemma 2.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable and let $s, t \in \mathbb{R}$ with $|s| \neq |t|$.*

(a) *There are unique $a, b, c, d \in \mathbb{R}$ such that*

$$g(x) \;:=\; a + bx + cx^2 + d|x|^3 \qquad \text{for } x \in \mathbb{R}$$

satisfies

$$g(s) = f(s),\; g'(s) = f'(s),\; g(t) = f(t),\; g'(t) = f'(t). \tag{37}$$

(b) *If f is a polynomial of degree at most 3, then g is a global upper or lower bound for f. More precisely, if $f(x) = A + Bx + Cx^2 + Dx^3$ for $x \in \mathbb{R}$, then we have the equivalence chains*

$$f \leq g \text{ on } \mathbb{R} \;\Leftrightarrow\; d \geq 0 \;\Leftrightarrow\; D \cdot (s+t) \geq 0, \tag{38}$$

$$f \geq g \text{ on } \mathbb{R} \;\Leftrightarrow\; d \leq 0 \;\Leftrightarrow\; D \cdot (s+t) \leq 0, \tag{39}$$

and the inequality between f and g in (38) or (39) is strict on all of $\mathbb{R} \setminus \{s,t\}$ iff $D \neq 0$ and $st < 0$. In any case, we have $a = A + Da_0$, $b = B + Db_0$, $c = C + Dc_0$, $d = Dd_0$, where

$$a_0 = \frac{4|st|^3}{(s+t)(s^2 + 4|st| + t^2)}, \qquad b_0 = \frac{6s^2t^2}{s^2 + 4|st| + t^2},$$

$$c_0 = -\frac{12s^2t^2}{(s+t)(s^2 + 4|st| + t^2)}, \qquad d_0 = \frac{(|s| + |t|)^3}{(s+t)(s^2 + 4|st| + t^2)}$$

in case of $st \leq 0$, and $a_0 = b_0 = c_0 = 0$ and $d_0 = \operatorname{sgn}(s) = \operatorname{sgn}(t)$ in case of $st > 0$.

(c) *If $f(x) = |x - r|^3$ for $x \in \mathbb{R}$, with some $r \in \mathbb{R} \setminus \{0\}$, and if $s = v \cdot \operatorname{sgn}(r)$ and $t = -u \cdot \operatorname{sgn}(r)$ for some u, v with $u > v \geq 0$, then we have $f \leq g$ on $\mathbb{R}$, and this inequality is strict on $\mathbb{R} \setminus \{s,t\}$ unless $v = 0$. More explicitly,*

$$|x - r|^3 \;\leq\; a + bx + cx^2 + d|x|^3, \tag{40}$$

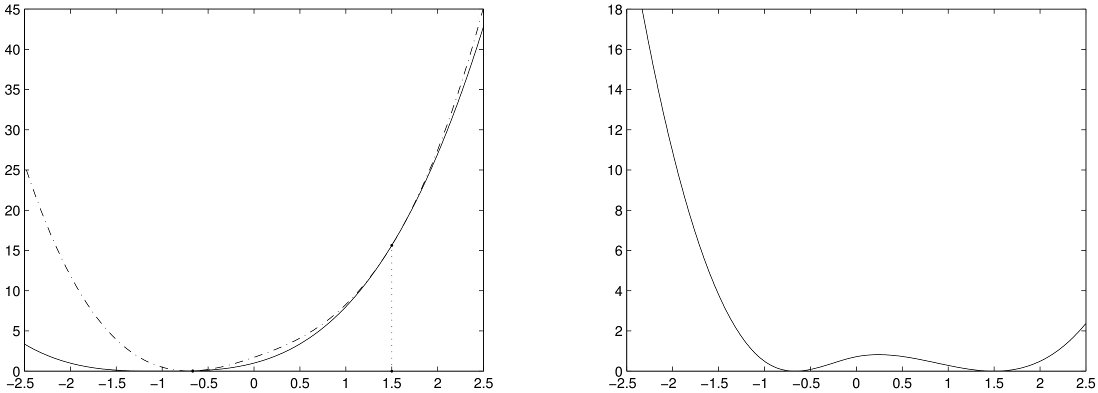

FIGURE 1. Left: plots of the functions $f(x) = |x+1|^3$ (solid line) and $g(x) = a + bx + cx^2 + d|x|^3$ (dashdot line) from Lemma 2.4(c) with $u = 3/2$, $v = 2/3$. Right: plot of the difference $g(x) - f(x)$.

where $a = a_r(u,v)$, $b = b_r(u,v)$, $c = c_r(u,v)$, $d = d_r(u,v)$, *with*

$$a_r(u,v) = |r|^3 + \frac{4u^3v^3}{(u-v)(u^2+4uv+v^2)}, \tag{41}$$

$$b_r(u,v) = -\mathrm{sgn}(r)\left(3r^2 + \frac{6u^2v^2}{u^2+4uv+v^2}\right), \tag{42}$$

$$c_r(u,v) = 3|r| - \frac{12u^2v^2}{(u-v)(u^2+4uv+v^2)}, \tag{43}$$

$$d_r(u,v) = \frac{(u+v)^3}{(u-v)(u^2+4uv+v^2)} \tag{44}$$

for $v \le |r|$ *and*

$$a_r(u,v) = |r|\frac{6u^4v^2+6u^2v^4+12u^3v^2|r|-12u^2v^3|r|-4u^3vr^2-4uv^3r^2-u^4r^2-v^4r^2+6u^2v^2r^2}{(u-v)\,(u+v)\,(u^2+4uv+v^2)},$$

$$b_r(u,v) = 3r\frac{-4u^2v^2-4u^3v-4uv^3-3u^2v|r|+3uv^2|r|+u^3|r|-v^3|r|-4uvr^2}{(u+v)\,(u^2+4uv+v^2)},$$

$$c_r(u,v) = 3|r|\frac{u^4+v^4-6u^2v^2-4u^3v-4uv^3+4u^3|r|-4v^3|r|+2u^2r^2+2v^2r^2}{(u-v)\,(u+v)\,(u^2+4uv+v^2)},$$

$$d_r(u,v) = \frac{(u-v+2|r|)\,(u^2+v^2+4uv-2u|r|+2v|r|-2r^2)}{(u-v)\,(u^2+4uv+v^2)}$$

for $v > |r|$. *Equality in* (40) *is attained at least* (*and at most as well if* $v > 0$) *at the two points* $x = -u \cdot \mathrm{sgn}(r)$ *and* $x = v \cdot \mathrm{sgn}(r)$.

Using monotonicity of the expectation, Lemma 2.4(c) trivially yields the following

Lemma 2.5. *For every* $r \in \mathbb{R} \setminus \{0\}$, $u > v \ge 0$ *and every* $P \in \mathrm{Prob}_3(\mathbb{R})$, *we have*

$$\int |x-r|^3 \,\mathrm{d}P(x) \;\le\; a + b\int x\,\mathrm{d}P(x) + c\int x^2\,\mathrm{d}P(x) + d\int |x|^3\,\mathrm{d}P(x),$$

where the coefficients $a = a_r(u,v), b = b_r(u,v), c = c_r(u,v)$*, and* $d = d_r(u,v)$ *are defined in Lemma 2.4(c), with equality iff the distribution* P *is concentrated in the two points* $v\cdot\mathrm{sgn}(r),\ -u\cdot\mathrm{sgn}(r)$.

Remark 2.6. Lemma 2.5 generalizes [60, Lemma 2], where the stated inequality was proved only in the case of $v > |r|$.

Proof of Lemma 2.4. (a) Condition (37) is a system of linear equations for a,b,c,d with the determinant

$$\begin{aligned}
\begin{vmatrix} 1 & s & s^2 & |s|^3 \\ 0 & 1 & 2s & 3s|s| \\ 1 & t & t^2 & |t|^3 \\ 0 & 1 & 2t & 3t|t| \end{vmatrix}
&= \begin{vmatrix} 1 & 2s & 3s|s| \\ t-s & t^2-s^2 & |t|^3-|s|^3 \\ 1 & 2t & 3t|t| \end{vmatrix} \\
&= \begin{vmatrix} t^2-s^2-2s(t-s) & |t|^3-|s|^3-3(t-s)s|s| \\ 2t-2s & 3t|t|-3s|s| \end{vmatrix} \\
&= (t-s)\begin{vmatrix} t-s & |t|^3+2|s|^3-3ts|s| \\ 2 & 3t|t|-3s|s| \end{vmatrix} \\
&= (t-s)\Big((t-s)(3t|t|-3s|s|)-2|t|^3-4|s|^3+6ts|s|\Big) \\
&= (t-s)\Big(|t|^3-|s|^3+3ts|s|-3ts|t|\Big) \\
&= (t-s)\,(|t|-|s|)\Big(t^2+s^2+|ts|-3ts\Big) \;\neq\; 0.
\end{aligned}$$

(b) Lemma 2.3 applied to $g-f$ or to $f-g$ yields the first equivalences in (38) and (39), even without knowing d explicitly. One next easily checks in case of $A = B = C = 0$ and $D = 1$ that the stated formulae for a,b,c,d solve the interpolation problem (37). The case of arbitray A,B,C,D then follows by the linearity of the interpolation operator mapping f to g according to part (a). Using now the explicit formula for $d = Dd_0$, one obviously gets the second equivalences in (38) and (39).

In case of $D = 0$ or $st \geq 0$, we have g identical to f at least on a half-line. In case of $D \neq 0$ and $st < 0$, the existence of any $u \in \mathbb{R}\setminus\{s,t\}$ with $f(u) = g(u)$ would imply by Lemma 2.3 that $f = g$ holds in some neighbourhood of zero, which implies $D = d = 0$, a contradiction to $D \neq 0$.

(c) The case $v > |r|$ is proved in [60, Lemma 1]. Let now $v \leq |r|$. By writing $f(x) = |r|^3\left|\frac{x}{-r}+1\right|^3$ and considering $\frac{x}{-r}$ as the new variable, we may assume that $r = -1$, that is, $f(x) = |x+1|^3$ for $x \in \mathbb{R}$, and

$$-1 \;\leq\; s \;=\; -v \;\leq\; 0 \;\leq\; v \;<\; t \;=\; u, \quad v \;\leq\; 1. \tag{45}$$

Let $\widetilde{f}(x) = (x+1)^3$ for $x \in \mathbb{R}$. Since $s,t \in [-1,\infty[$ and $f = \widetilde{f}$ on $[-1,\infty[$, our present g is also the osculatory interpolation to the polynomial $\widetilde{f}$. Hence the present formulae for the coefficients of g follow from part (b) with $A = D = 1$, $B = C = 3$, and in view of $s+t = u-v > 0$ we get from (38) that $f(x) = \widetilde{f}(x) \leq g(x)$ holds for $x \in [-1,\infty[$, and in view of $st = -uv \leq 0$ we have either equality iff $x \in \{s,t\}$, or $s = v = 0$. So, setting

$$h(x) \;\coloneqq\; g(x)-f(x) \;=\; a+1+(b+3)x+(c+3)x^2+(1-d)x^3 \quad \text{for } x \in\,]-\infty,-1],$$

it is enough to prove now $h < 0$ on $]-\infty,-1[$.

We have

$$h'(x) \;=\; b+3+2(c+3)x+3(1-d)x^2, \qquad h''(x) \;=\; 2(c+3)+6(1-d)x,$$

and, using $u > v \geq 0$ from (45) and also (44), we get

$$W \;:=\; (u-v)(u^2+4uv+v^2) \;>\; 0, \qquad d-1 \;=\; \frac{2v^2(3u+v)}{W} \;\geq\; 0$$

and hence, for $x \in \,]-\infty,-1[$, using (43) with $r=-1$ in the central step, and $v \in [0,1]$ from (45) in the last,

$$h''(x) \;\geq\; 2(c+3)+6(d-1) \;=\; \frac{12}{W}u^2\left(u+3v-2v^2\right) \;>\; 0$$

Thus h' is strictly increasing, and hence we get, for $x \in \,]-\infty,-1[$,

$$h'(x) \;<\; h'(-1) \;=\; b-2c-3d \;=\; -\frac{6u^2(1-v)(u+3v+v(u-v))}{W} \;\leq\; 0,$$

so that h is strictly decreasing, and we get, again for $x \in \,]-\infty,-1[$,

$$h(x) \;>\; h(-1) \;=\; a-b+c+d \;=\; \frac{2u^2(v-1)^2(2uv+u+3v)}{W} \;\geq\; 0$$

as desired. □

2.3. **Sign change counting.** The notation and facts of this subsection are used in the formulation and the proof of Theorem 4.2, which in turn is used in Steps 6 and 7 of the proof of Theorem 1.2 in section 5. Lemma 2.8 refines [13, Lemma 4.2].

For sets $A,B \subseteq \mathbb{R}$ and $n \in \mathbb{N}_0$, we put $A^n_< := \{x \in A^n : x_1 < x_2 < \ldots < x_n\}$, $A^n_\leq := \{x \in A^n : x_1 \leq x_2 \leq \ldots \leq x_n\}$, and $A \leq B :\Leftrightarrow x \leq y$ for every choice of $x \in A$ and $y \in B$, and we define $A < B$ similarly.

Let now $D \subseteq \mathbb{R}$ and let $f : D \to \mathbb{R}$ be a function. Then, with a notation as in [32, p. 20], one calls

$$\begin{aligned} S^-(f) \;&:=\; \sup\{n \in \mathbb{N}_0 : \exists\, x \in D^{n+1}_< \text{ with } f(x_i)f(x_{i+1}) < 0 \text{ for } i \in \{1,\ldots,n\}\} \\ &\in\; \mathbb{N}_0 \cup \{\infty\} \end{aligned}$$

the *(possibly infinite) number of (inequivalent) sign changes* of f, and the restrictions of f obey the rule

(46) $$S^-(f|_{A\cup B}) \;\leq\; S^-(f|_A) + S^-(f|_B) + 1 \qquad \text{for } A,B \subseteq D \text{ with } A \leq B.$$

Let us from now on assume for simplicity that $D = I$ is an interval. For $n \in \mathbb{N}_0$ then clearly $S^-(f) = n$ is equivalent to the following condition: There exist a $z = (z_1,\ldots,z_n) \in I^n_\leq$ and nonempty (but possibly one-point) intervals $I_0,\ldots,I_n$ with $\bigcup_{j=0}^n I_j = I$ and such that, for $j \in \{0,\ldots,n\}$, we have $f(x)f(y) \geq 0$ for $x,y \in I_j$, but in case of $j \geq 1$ also $\sup I_{j-1} = z_j = \inf I_j$ and $f(x)f(y) < 0$ for some $x \in I_{j-1}$ and $y \in I_j$. If this condition holds, let us call every z as above a *sign change tuple* of f, every entry z_i of such a z a *sign change* of f, and two different sign changes of f *inequivalent* if they both occur in one sign change tuple. If in addition f is left- or right-continuous, then obviously every such z belongs to $I^n_<$ and the corresponding intervals I_j are nondegenerate. Let us finally call $f\colon I \to \mathbb{R}$ *lastly positive* if we have $f \geq 0$ on I or there is an $x_0 \in I$ with $f(x_0) > 0$ and $f \geq 0$ on $]x_0,\infty[\, \cap I$, and *essentially lastly positive* if we have $f \geq 0$ Lebesgue-a.e. on I or there is an $x_0 \in I$ with $f \geq 0$ Lebesgue-a.e. on $[x_0,\infty[\, \cap I$ and not $f = 0$ Lebesgue-a.e. on $[x_0,\infty[\, \cap I$.

We will need the following variant of Rolle's theorem.

Lemma 2.7. *Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to [0, \infty[$ be absolutely continuous, not identically zero, and vanishing in the limit at the boundary points* $\inf I$ *and* $\sup I$*. Then there exist $\xi, \eta \in I$ with $\xi < \eta$ and $f'(\xi) > 0 > f'(\eta)$.*

Proof. We choose a maximizer x_0 for f. Then x_0 is not a boundary point of I, and we have $\int_x^{x_0} f'(t)\,\mathrm{d}t = f(x_0) - f(x) > 0$ for some $x < x_0$ sufficiently close to $\inf I$, and then $f'(\xi) > 0$ for some $\xi \in]x, x_0[$. Similarly, $f'(\eta) < 0$ for some $\eta \in]x_0, x[$ with some $x > x_0$ close to $\sup I$. □

Lemma 2.8. *Let I be a nondegenerate interval, $a = \inf I$, $b = \sup I$, $f : I \to \mathbb{R}$ be absolutely continuous, and let $f' : I \to \mathbb{R}$ be almost everywhere a derivative of f.*

(a) *If $\lim_{t\to b-} f(t) = 0$ and if f' is essentially lastly positive, then so is $-f$.*

(b) *We have*

$$S^-(f) \;\leq\; S^-(f') + 1 - \Big(\lim_{x\to a+} f(x) = 0\Big) - \Big(\lim_{x\to b-} f(x) = 0\Big) \tag{47}$$

except when $f = 0$ and $S^-(f') = 0$.

More precisely, if $S^-(f') = n \in \mathbb{N}_0$, then also $m := S^-(f)$ is finite, and, if f is not identically zero, with $y \in I^m_<$ and $z \in I^n_\leq$ denoting any sign change tuples of f and f' respectively, and with

$$J \;:=\; \Big\{ j \in \{0, \ldots, m\} : 1 \leq j \leq m-1, \text{ or } j = 0 \text{ and } \lim_{x\to a+} f(x) = 0, \\ \text{or } j = m \text{ and } \lim_{x\to b-} f(x) = 0 \Big\} \tag{48}$$

and $y_0 := a$ and $y_{m+1} := b$, for every $j \in J$, there is a $k \in \{1, \ldots, n\}$ with $z_k \in]y_j, y_{j+1}[$.

Proof. (a) Obvious from $-f(x) = \lim_{y\to b-}(f(y) - f(x)) = \lim_{y\to b-} \int_x^y f'(t)\,\mathrm{d}t$ for $x \in I$.

(b) It suffices to prove the second claim since, under the stated conditions, it yields the existence of an injective function $k(\cdot) : J \to \{1, \ldots, n\}$, hence $\#J \leq n$ and thus (47), and since the remaining cases of $S^-(f') = \infty$ or $f = 0$ are trivial.

So let $S^-(f') = n \in \mathbb{N}_0$, f not identically zero, and $z \in I^n_\leq$ a sign change tuple of f'. With corresponding intervals $I_0, \ldots, I_n$ as above, we have, for each $j \in \{0, \ldots, n\}$, either $f' \leq 0$ on I_j or $f' \geq 0$ on I_j, and hence $S^-(f|_{I_j}) \leq 1$, and hence $m := S^-(f) \leq 2n+1 < \infty$, by applying (46) n times. So let $y \in I^m_<$ be a sign change tuple of f, let J be defined by (48), $y_0 := a$, $y_{m+1} := b$, and let $j \in J$. Applying Lemma 2.7 to $f|_{]y_j, y_{j+1}[}$ or its negative yields a k as claimed. □

2.4. **Partial sums of reciprocals of square roots.** As usual, the symbol ζ without any subscript denotes the Riemann zeta-function. In particular, $\zeta(\frac{1}{2})$ is a negative number as indicated in (49) below, see `oeis.org/A059750` in [41].

Lemma 2.9. *For $n \in \mathbb{N} = \{1, 2, \ldots\}$, we have*

$$\sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} - 2\sqrt{n} \;<\; \zeta(\tfrac{1}{2}) \;=\; -1.46035\ldots, \tag{49}$$

with equality in the limit as $n \to \infty$.

Proof. Let a_n denote the left hand side of the inequality in (49). Then $a_n - a_{n+1} = 2\sqrt{n}\left(\sqrt{1+\frac{1}{n}} - \left(1 + \frac{1}{2n}\right)\right) < 0$ by the tangent bound at 1 for the concave function $\sqrt{\cdot}$. Hence the sequence $(a_n)_{n\geq 1}$ is strictly increasing. Since we have $\lim_{n\to\infty} a_n = \zeta(1/2)$ by

[25, p. 333, (13.10.7) with $s = \sigma = \frac{1}{2}$], or see [66, p. 192, (4.1)] for a more elementary proof, the inequality in (49) follows. □

3. On few-point reduction theorems

In this section, we recall some reduction theorems partially used below, with apparently some novelty in part (b) of the first one. For Tyurin's Theorem 3.3 we provide a proof perhaps more natural than the original one.

The term "component" below is meant in the usual topological sense of "maximal connected subset", here of a subset M of $\mathbb{R}^k$.

Theorem 3.1 (essentially Richter 1957, [48])**.** *Let P be a law on the measurable space $(\mathcal{X}, \mathcal{A})$, let $k \in \mathbb{N}$, and let $f_1, \ldots, f_k$ be real-valued and P-integrable functions on $\mathcal{X}$.*

(a) *There exists a law Q on $(\mathcal{X}, \mathcal{A})$ concentrated in $k + 1$ or fewer points such that $Pf_i = Qf_i$ holds for each $i \in \{1, \ldots, k\}$.*

(b) *Assume in addition that $M \coloneqq \{(f_1(x), \ldots, f_k(x)) : x \in \mathcal{X}\}$ has at most k components. Then conclusion (a) holds with "k or fewer" in place of "$k + 1$ or fewer".*

Proof. Let $F(x) \coloneqq (f_1(x), \ldots, f_k(x))$ for $x \in \mathcal{X}$, so that M as defined in **(b)** above is the image of the function F, and let C denote the convex hull of M. Then we have $y \coloneqq \int F \,\mathrm{d}P \in C$, by part of the multivariate Jensen inequality as in [21, p. 74, Lemma 3] or [18, p. 348, Theorem 10.2.6], noting that the measurability condition imposed on C in the second reference is not used anywhere in the proof.

(a) By the Carathéodory theorem [27, p. 29, Theorem 1.3.6], the point y is a convex combination of $k + 1$ or fewer points in M, that is, there exist not necessarily distinct $x_1, \ldots, x_{k+1} \in \mathcal{X}$ and $p_1, \ldots, p_{k+1} \in [0, 1]$ with $\sum_{j=1}^{k+1} p_j = 1$ and $y = \sum_{j=1}^{k+1} p_j F(x_j)$, that is, $Pf_i = Qf_i$ holds for $Q = \sum_{j=1}^{k+1} p_j \delta_{x_j}$ and each $i \in \{1, \ldots, k\}$.

(b) Under the additional hypothesis, the Fenchel-Bunt refinement [27, p. 30, Theorem 1.3.7, see also pp. 245–246] of the Carathéodory theorem yields that y is a convex combination of k or fewer points in C, and we conclude as before. □

For $\mathcal{X}$ a Borel subset of $\mathbb{R}$, Theorem 3.1(a) is contained in [48, p. 153, Satz 4]. For $\mathcal{X}$ an interval in $\mathbb{R}$ and for the special case of continuous f_i, in which case M is connected, Theorem 3.1(b) is [48, p. 153, Satz 5], whereas in our version and say in case of $k \geq 3$, one of the functions f_i could for example be an indicator of a subinterval of $\mathcal{X}$, since then, assuming the remaining functions to be continuous, M would have at most three components. For a general measurable space $(\mathcal{X}, \mathcal{A})$, Theorem 3.1(a) is stated in [33], where also further references are given.

In the course of the proof of our main result below, Theorem 3.1(a) allows us to restrict attention to 5-point laws, which are still rather complex objects. Using instead Theorem 3.1(b) would permit us to consider only 4-point laws. However, the following generalization of a result [28, p. 269, Theorem 2.1 with $n = 1$] of Hoeffding from 1955, combined with Theorem 3.1(a) and with the concavity of the function B from (4), allows a reduction to 3-point laws, which turn out to be sufficiently tractable analytically. Let us remark that using just Hoeffding's result would again only lead to a reduction to 4-point laws.

For the rest of this section all laws considered are finitely supported and are hence for notational simplicity regarded as defined on the power set of the basic set $\mathcal{X}$.

Theorem 3.2 (implicitly Hoeffding 1955, [28])**.** *Let $\mathcal{X}$ be a set, let $k \in \mathbb{N}$, and let $f_1, \ldots, f_k$ be real-valued functions on $\mathcal{X}$. Then every finitely supported law P on $\mathcal{X}$ is* *a finite convex combination* $\sum_{j=1}^{n}\lambda_j P_j$ *of laws* P_j *each concentrated on* $k+1$ *or fewer support points of* P *and satisfying* $P_j f_i = P f_i$ *for each* $i \in \{1,\ldots,k\}$.

Proof. Replacing $\mathcal{X}$ by $\{x\in\mathcal{X} : P(\{x\})>0\}$, we may assume that $\mathcal{X}$ is finite and is the set of all support points of P. Then

$$K \;\;:=\;\; \{Q\in \mathrm{Prob}(\mathcal{X}) : Qf_i = Pf_i \text{ for } i\in\{1,\ldots,k\}\}$$

is a convex and compact subset of the finite-dimensional vector space of all $\mathbb{R}$-valued measures on $\mathcal{X}$, with $P\in K$. Hence, by Minkowski's theorem [27, p. 42, Theorem 2.3.4], P is a finite convex combination $\sum_{j=1}^{n}\lambda_j P_j$ of extreme points P_j of K, and then each P_j is concentrated in at most $k+1$ points:

Indeed, suppose that $Q=\sum_{x\in\mathcal{X}} q_x\delta_x \in K$ is such that its set of support points $\mathcal{X}_0 := \{x\in\mathcal{X} : q_x>0\}$ contains at least $k+2$ elements. Then

$$\left\{ r\in\mathbb{R}^{\mathcal{X}_0} : \sum_{x\in\mathcal{X}_0} r_x = 0,\, \sum_{x\in\mathcal{X}_0} r_x f_i(x) = 0 \text{ for } i\in\{1,\ldots,k\}\right\}$$

is a subspace of dimension at least 1 of $\mathbb{R}^{\mathcal{X}_0}$, hence contains a nonzero r, so that we have

$$Q_\pm \;\;:=\;\; Q\pm\varepsilon\sum_{x\in\mathcal{X}_0} r_x\delta_x \;\;\in\;\; K\setminus\{Q\}$$

for some $\varepsilon>0$, and $Q=\frac{1}{2}(Q_+ + Q_-)$. Thus Q is not an extreme point of K. □

Theorem 3.3 (Tyurin 2009, [61, 62, 63])**.** *Let* $\mathcal{X}$ *be a set,* $k\in\mathbb{N}$*,* $f_1,\ldots,f_k$ *real-valued functions on* $\mathcal{X}$*,* $c_1,\ldots,c_k\in\mathbb{R}$*, and*

$$\mathcal{P} \;\;:=\;\; \{P\in\mathrm{Prob}(\mathcal{X}) : \#\mathrm{supp}\,P<\infty,\, Pf_i = c_i \text{ for } i\in\{1,\ldots,k\}\}\,.$$

Let $F:\mathcal{P}\to\overline{\mathbb{R}}$ *be quasi-convex, that is, satisfying* $F(\lambda P+(1-\lambda)Q)\le\max\{F(P),F(Q)\}$ *for* $P,Q\in\mathcal{P}$ *and* $\lambda\in[0,1]$*. Then*

$$\sup\{F(P) : P\in\mathcal{P}\} \;\;=\;\; \sup\{F(P) : P\in\mathcal{P},\, \#\mathrm{supp}\,P\le k+1\}.$$

Proof. Applying the representation $P=\sum_{j=1}^{n}\lambda_j P_j$ from Theorem 3.2, and the quasi-convexity condition on F extended by induction, immediately yields the claim. □

Let us finally mention [65, 43] as starting points for some more sophisticated results related to this section.

4. Auxiliary results for Zolotarev's ζ-metrics

Proof of Theorem 1.7. (a) An obvious Hahn-Banach argument, as in Step 2 of the proof of Theorem 1.2 in section 5 below.

(b) Definiteness of ζ_s, that is, the implication $\zeta_s(P,Q)=0\Rightarrow P=Q$, is of course very well-known, for example as a consequence of the uniqueness theorem for characteristic functions. The remaining claims are obvious.

(d) Relation (16) follows from Lemma 2.2 using dominated convergence. Inequality (17) follows from (16) using the linearity of expectations.

(c) If $\zeta_s(P,Q)<\infty$, then we apply Lemma 2.2 to $f:=\left(\prod_{j=0}^{m-1}(s-j)\right)^{-1}|\cdot|^s\in\mathcal{F}_s$ to get $\infty>\zeta_s(P,Q)\ge|Pf_n-Qf_n|\to|Pf-Qf|$ using dominated convergence for Qf_n, dominated convergence for Pf_n in case of $Pf<\infty$, and Fatou's Lemma for Pf_n in case of $Pf=\infty$, and we conclude that $Pf<\infty$, that is $P\in\mathrm{Prob}_s(\mathbb{R})$; and for $j\in\{1,\ldots,m\}$ and $n\in\mathbb{N}$ then (16) from part (d) applies to the monomial $n(\cdot)^j\in\mathcal{F}_s$, and letting $n\to\infty$

yields $\mu_j(P) = \mu_j(Q)$. If the second condition in (15) holds, then the third follows easily using (35), compare [52, pp. 102–103]. Finally, the third condition in (15) implies the first, in view of $\nu_s(P,Q) \le \nu_s(P) + \nu_s(Q)$. The remaining claims follow obviously. □

Let us next recall two further well-known properties of ζ_s, with $s \in \,]0,\infty[$ arbitrary, needed below. The first is its *regularity*

$$\zeta_s(P*R, Q*R) \;\le\; \zeta_s(P,Q) \quad \text{for } P,Q,R \in \mathrm{Prob}(\mathbb{R}) \tag{50}$$

proved e.g. in [52, p. 101], which, given Theorem 1.7(b), is equivalent to its *semiadditivity*

$$\zeta_s\Big(\mathop{\ast}_{i=1}^{n} P_i, \mathop{\ast}_{i=1}^{n} Q_i\Big) \;\le\; \sum_{i=1}^{n} \zeta_s(P_i,Q_i) \quad \text{for } n \in \mathbb{N} \text{ and } P_i, Q_i \in \mathrm{Prob}(\mathbb{R}), \tag{51}$$

compare [52, p. 48]. To formulate the second, we use here, as well as later in some proofs, the obvious random variable notation $\zeta_s(X,Y) := \zeta_s(P,Q)$ if X,Y are $\mathbb{R}$-valued r.v.'s with $X \sim P$ and $Y \sim Q$. Then we have the *homogeneity*

$$\zeta_s(aX, aY) \;=\; a^s \zeta_s(X,Y) \quad \text{for } a \in [0,\infty[\text{ and } \mathbb{R}\text{-valued r.v.'s } X \text{ and } Y, \tag{52}$$

the obvious proof of which being given in [52, p. 102].

The following Lemma, which is presented in [52, pp. 108-112] without explicit constants, allows us in the proof of Theorem 1.14, in a case where $aX \sim P$ and $aY \sim Q$ with small a, to use the homogeneity (52) with a better exponent than possible by just using (50). We recall that N_σ denotes the centred normal law on $\mathbb{R}$ with variance σ^2.

Lemma 4.1. *Let $P,Q \in \mathrm{Prob}(\mathbb{R})$ and $s,t,\sigma \in \,]0,\infty[$. Then we have*

$$\zeta_s(P*\mathrm{N}_\sigma, Q*\mathrm{N}_\sigma) \;\le\; C_{s,t}\frac{\zeta_{s+t}(P,Q)}{\sigma^t} \tag{53}$$

with the finite constant $C_{s,t}$ defined as follows: Writing

$$s = \ell + \alpha, \quad t = m + \beta \qquad \text{with } \ell, m \in \mathbb{N}_0 \text{ and } \alpha,\beta \in \,]0,1]$$

and letting φ denote the standard normal density, we put

$$D_k \;:=\; \int |\varphi^{(k)}(x)|\,\mathrm{d}x, \qquad D_{k,\alpha} \;:=\; \int |x|^\alpha\, |\varphi^{(k)}(x)|\,\mathrm{d}x \qquad \text{for } k \in \mathbb{N}_0,$$

$$C_{s,t} \;:=\; \begin{cases} D_m^{\frac{1-\alpha-\beta}{1-\alpha}} \cdot D_{m+1,\alpha}^{\frac{\beta}{1-\alpha}} & \text{if } \alpha+\beta \le 1, \\[2ex] D_{m+1}^{\frac{\alpha+\beta-1}{\alpha}} \cdot (2D_{m+1,\alpha})^{\frac{1-\beta}{\alpha}} & \text{if } \alpha+\beta > 1. \end{cases}$$

In particular, if $t \in \mathbb{N}$, hence $m = t-1$, $\beta = 1$, and $\alpha+\beta > 1$, then $C_{s,t} = D_{m+1} = D_t = \int |\varphi^{(t)}(x)|\,\mathrm{d}x$, and the first few of these constants can be explicitly computed, for example

$$C_{s,1} \;=\; \int |\varphi'(x)|\,\mathrm{d}x \;=\; \frac{2}{\sqrt{2\pi}}, \qquad C_{s,2} \;=\; \int |\varphi''(x)|\,\mathrm{d}x \;=\; \frac{4}{\sqrt{2\pi \mathrm{e}}}.$$

Proof. We shall follow the outline of the reasoning employed in [52, Lemma 2.10.1]. Let $\varphi_\sigma(x) := \sigma^{-1}\varphi(x/\sigma)$ for $x \in \mathbb{R}$. Given any $f \in \mathcal{F}_s^\infty$, and writing

$$g(x) \;:=\; \int f(x+z)\varphi_\sigma(z)\,\mathrm{d}z \quad \text{and} \quad h(x) \;:=\; \frac{\sigma^t g(x)}{C_{s,t}} \quad \text{for } x \in \mathbb{R}, \tag{54}$$

it is sufficient to prove that $h \in \mathcal{F}_{s+t}^\infty$, for then we would get

$$|(P*\mathrm{N}_\sigma)f - (Q*\mathrm{N}_\sigma)f| \;=\; |Pg - Qg| \;=\; \frac{C_{s,t}}{\sigma^t}|Ph - Qh| \;\le\; \text{R.H.S.(53)}$$

as desired. So let $f \in \mathcal{F}_s^\infty$ and let g and h be defined through (54). Then h is obviously bounded, and, with

$$n \;:=\; \lceil s+t-1\rceil \;=\; \begin{Bmatrix} \ell+m \\ \ell+m+1 \end{Bmatrix} \text{ if } \alpha+\beta \begin{Bmatrix} \le \\ > \end{Bmatrix} 1 \quad\text{and}\quad \gamma \;:=\; s+t-n \;\in\;]0,1],$$

it remains to prove that we have

$$\left|g^{(n)}(x)-g^{(n)}(y)\right| \;\le\; \frac{C_{s,t}}{\sigma^t}|x-y|^\gamma \quad \text{for } x,y\in\mathbb{R}. \tag{55}$$

If $k\in\mathbb{N}_0$ with $k\ge\ell$, then we obtain, for $x,y\in\mathbb{R}$,

$$g^{(\ell)}(x) \;=\; \int f^{(\ell)}(x+z)\varphi_\sigma(z)\,\mathrm{d}z \;=\; \int f^{(\ell)}(z)\varphi_\sigma(x-z)\,\mathrm{d}z, \tag{56}$$

$$g^{(k)}(x) \;=\; \int f^{(\ell)}(z)\varphi_\sigma^{(k-\ell)}(x-z)\,\mathrm{d}z \;=\; \int f^{(\ell)}(x-z)\varphi_\sigma^{(k-\ell)}(z)\,\mathrm{d}z, \tag{57}$$

$$|g^{(k)}(x)-g^{(k)}(y)| \;\le\; \int\left|f^{(\ell)}(x-z)-f^{(\ell)}(y-z)\right|\left|\varphi_\sigma^{(k-\ell)}(z)\right|\mathrm{d}z \;\le\; |x-y|^\alpha\frac{D_{k-\ell}}{\sigma^{k-\ell}} \tag{58}$$

where, to justify differentiation under the integral, we may in (56) apply the dominated convergence theorem successively using polynomial bounds on the derivatives $f',\ldots,f^{(\ell)}$, compare (35) and the ensuing line, and we may treat (57) similarly, or remember it as a well-known special case of the differentiability of Laplace transforms, see for example [37, Example]; at the last step in (58) we used $f\in\mathcal{F}_s$ and the change of variables $z\mapsto\sigma z$. Specializing (57) to $k:=\ell+m+1$ and using at the first step below $\int\varphi_\sigma^{(m+1)}(z)\,\mathrm{d}z=0$ yields

$$\begin{aligned} |g^{(\ell+m+1)}(x)| \;&=\; \left|\int\left(f^{(\ell)}(x-z)-f^{(\ell)}(x)\right)\varphi_\sigma^{(m+1)}(z)\,\mathrm{d}z\right| \\ &\le\; \int|z|^\alpha\left|\varphi_\sigma^{(m+1)}(z)\right|\mathrm{d}z \;=\; \frac{D_{m+1,\alpha}}{\sigma^{m+1-\alpha}} \quad \text{for } x\in\mathbb{R}. \end{aligned} \tag{59}$$

Let us now first assume that we have $\alpha+\beta\le1$, and hence $n=\ell+m$ and $\gamma=\alpha+\beta$. Then, using (59) at the second step below, we get

$$\text{L.H.S.(55)} \;\le\; \|g^{(n+1)}\|_\infty\cdot|x-y| \;\le\; \frac{D_{m+1,\alpha}}{\sigma^{m+1-\alpha}}|x-y| \quad \text{for } x,y\in\mathbb{R},$$

and taking a geometric mean of this bound and the one from (58) with $k:=n$, with the exponents $u:=\beta/(1-\alpha)\in\,]0,1]$ and $1-u$, yields (55) in the present case.

Let us finally assume that we have $\alpha+\beta>1$, and hence $n=\ell+m+1$ and $\gamma=\alpha+\beta-1$. Then, applying below (59) to x and to y, we get

$$\text{L.H.S.(55)} \;\le\; \frac{2D_{m+1,\alpha}}{\sigma^{m+1-\alpha}} \quad \text{for } x,y\in\mathbb{R},$$

and taking a geometric mean of this bound and the one from (58) with $k:=n$, with the exponents $v:=(1-\beta)/\alpha\in[0,1[$ and $1-v$, yields (55) again. □

In Steps 6 and 7 of our proof of Theorem 1.2, we will use Theorem 4.2 stated below, which collects or refines results known from [68], [13], and [10]. In particular, Theorem 4.2(b) contains [13, Theorems 3.3 and 4.3] and [10, p. 353, first part of Theorem 2], and adds a converse to the latter, while Theorem 4.2(c,d) seems to be new.

Let us first recall the definition of the s-convex order of laws on $\mathbb{R}$ in accordance with [13, p. 590], [10, p. 351], [39, p. 39, Definition 1.6.2 a)], and [53, p. 139], but being here somewhat more explicit with respect to the appropriate integrability assumptions: If $s \in \mathbb{N}$, then

$$P \;\; \leq_{s\text{-cx}} \;\; Q \tag{60}$$

is defined to mean that $P, Q \in \mathrm{Prob}_{s-1}(\mathbb{R})$ and that $Pf \leq Qf$ holds for every s-convex function $f : \mathbb{R} \to \mathbb{R}$ such that Pf and Qf are well-defined (possibly infinite). Thus $\leq_{1\text{-cx}}$ is just the usual stochastic order $\leq_{\mathrm{st}}$ on $\mathrm{Prob}(\mathbb{R})$, $\leq_{2\text{-cx}}$ is the usual convex order $\leq_{\mathrm{cx}}$ on $\mathrm{Prob}_1(\mathbb{R})$, and $\leq_{3\text{-cx}}$ is what we use below. By considering the s-convex function $\pm(\cdot)^k$ with $k \in \{1, \ldots, s-1\}$, it is clear that (60) necessitates

$$\mu_j(P) \;\; = \;\; \mu_j(Q) \;\; \in \;\; \mathbb{R} \quad \text{ for } j \in \{1, \ldots, s-1\}. \tag{61}$$

For $x \in \mathbb{R}$ and $\alpha \in [0, \infty[$, we agree to the standard notation $x_-^\alpha := (x_-)^\alpha$ and $x_+^\alpha := (x_+)^\alpha$ if $\alpha > 0$, and $x_-^0 := (x \leq 0)$ and $x_+^0 := (x \geq 0)$, which is not in general the same as $(x_-)^0$ and $(x_+)^0$ due to $0^0 := 1$. For a law $P \in \mathrm{Prob}(\mathbb{R})$, let F and $\overline{F}$ denote its ordinary and "upper" distribution functions, that is, $F(x) := P(]-\infty, x])$ and $\overline{F}(x) := P([x, \infty[)$ for $x \in \mathbb{R}$, and we then define $F_k(t)$ and $\overline{F}_k(t)$ for $k \in \mathbb{N}$ and $t \in \mathbb{R}$ inductively by $F_1 := F$, $\overline{F}_1 := \overline{F}$,

$$F_{k+1}(t) \;\; := \;\; \int_{-\infty}^t F_k(x)\,\mathrm{d}x, \qquad \overline{F}_{k+1}(t) \;\; := \;\; \int_t^\infty \overline{F}_k(x)\,\mathrm{d}x, \tag{62}$$

and hence get, as follows by inserting the right hand sides from (63) into the integrals in (62) and using Fubini,

$$F_k(t) \;\; = \;\; \int \frac{(x-t)_-^{k-1}}{(k-1)!}\,\mathrm{d}P(x), \qquad \overline{F}_k(t) \;\; = \;\; \int \frac{(x-t)_+^{k-1}}{(k-1)!}\,\mathrm{d}P(x). \tag{63}$$

By (63), the functions F_k and $\overline{F}_k$ are finite-valued in particular if $P \in \mathrm{Prob}_{k-1}(\mathbb{R})$, and then (62) with $k-1$ in place of k yields

$$\lim_{t \to -\infty} F_k(t) = 0, \qquad \lim_{t \to \infty} \overline{F}_k(t) = 0. \tag{64}$$

In Theorem 4.2(a,d) below, symmetry of $P - Q$ is to be understood in the usual sense of $(P-Q)(B) = (P-Q)(-B)$ for every Borel set $B \subseteq \mathbb{R}$.

Theorem 4.2 (ζ-distances, s-convex orderings, cut conditions)**.** *Let $s \in \mathbb{N}$ and let $P, Q \in \mathrm{Prob}_{s-1}(\mathbb{R})$ satisfy the moment condition (61). Let further $F, \overline{F}, G, \overline{G}$ denote the respective ordinary and complementary distribution functions of P, Q and, with $F_k, \overline{F}_k, G_k, \overline{G}_k$ as in (62) and (63), let $H_k := G_k - F_k$ and $\overline{H}_k := \overline{G}_k - \overline{F}_k$ for $k \in \{1, \ldots, s\}$.*

(a) *For $k \in \{1, \ldots, s\}$ and $t \in \mathbb{R}$, we have*

$$(-1)^{k-1} H_k(t) + \overline{H}_k(t+) \;\; = \;\; 0, \tag{65}$$

$$(-1)^{k-1} H_k(t-) + \overline{H}_k(t) \;\; = \;\; 0, \tag{66}$$

and, if $P - Q$ is symmetric, then also

$$\overline{H}_k(-t) \;\; = \;\; (-1)^k \overline{H}_k(t+)\,; \tag{67}$$

here the one-sided limit signs, namely "+" in the argument of $\overline{H}_k$ in (65) and (67), and "−" in the argument of H_k in (66), can be omitted if $k \geq 2$.

Let I denote the smallest interval satisfying $P(I) = Q(I) = 1$. Then, for each $k \in \{1, \ldots, s\}$, we have $\overline{H}_k = 0$ on $\mathbb{R} \setminus I$ and

$$\lim_{t \to -\infty} \overline{H}_k(t) \;\; = \;\; \lim_{t \to \infty} H_k(t) \;\; = \;\; 0. \tag{68}$$

If in addition $P,Q \in \mathrm{Prob}_s(\mathbb{R})$*, then we have*

$$\zeta_s(P,Q) \;=\; \int |\overline{H}_s(x)|\,\mathrm{d}x, \tag{69}$$

and a function $f \in \mathcal{F}_s$ *satisfies*

$$\zeta_s(P,Q) \;=\; Qf - Pf \tag{70}$$

iff its Lebesgue-a.e. existing derivative of order s *satisfies*

$$f^{(s)}(x) \;=\; \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \quad \text{if} \quad \overline{H}_s(x) \begin{Bmatrix} < \\ > \end{Bmatrix} 0, \qquad \text{for Lebesgue-a.e. } x \in I. \tag{71}$$

(b) *For* $k \in \{1,\ldots,s\}$*, let* (B_k) *denote the condition "*$\overline{H}_k$ *has at most* $s-k$ *sign changes and is lastly positive". Then we have the implications* $(B_1) \Rightarrow (B_2) \Rightarrow \ldots \Rightarrow (B_s) \Leftrightarrow \overline{H}_s \geq 0 \Leftrightarrow P \leq_{s-\mathrm{cx}} Q$*. If in addition* $P,Q \in \mathrm{Prob}_s(\mathbb{R})$*, then* $P \leq_{s-\mathrm{cx}} Q$ *is further equivalent to* $\zeta_s(P,Q) = \frac{1}{s!}(\mu_s(Q) - \mu_s(P))$*, that is, to* (70) *holding for the function* $f \in \mathcal{F}_s$ *given by*

$$f(x) \;\coloneqq\; \tfrac{1}{s!}x^s \qquad \text{for } x \in \mathbb{R}. \tag{72}$$

(c) *For* $k \in \{1,\ldots,s\}$*, let* (C_k) *denote the condition "*$\overline{H}_k$ *has exactly* $s-k+1$ *sign changes and is lastly positive". Then we have the implications* $(C_1) \Rightarrow (C_2) \Rightarrow \ldots \Rightarrow (C_s)$*. If in addition* $P,Q \in \mathrm{Prob}_s(\mathbb{R})$*, then* (C_s) *is further equivalent to* (70) *holding, with some sign change point* x_0 *of* $\overline{H}_s$*, for the function* $f \in \mathcal{F}_s$ *given by*

$$f(x) \;\coloneqq\; \tfrac{1}{s!}|x - x_0|^s \qquad \text{for } x \in \mathbb{R}, \tag{73}$$

and this remains true if "some" is replaced by "some and every". Further, if (C_k) *holds for some* $k \in \{1,\ldots,s-1\}$*, then each sign change point of* $\overline{H}_s$ *belongs to the interior of the convex hull of the entries of every sign change tuple of* $\overline{H}_k$*.*

(d) *Assume that we have* $P,Q \in \mathrm{Prob}_s(\mathbb{R})$*,* $P-Q$ *symmetric, and* $\overline{H}_s$ *with exactly one sign change. Then* s *is odd, and* (70) *holds with* $f(x) \coloneqq |x|^s/s!$ *for* $x \in \mathbb{R}$*.*

Proof. (a) For every $t \in \mathbb{R}$, (63) yields that

$$(-1)^{k-1}F_k(t) + \overline{F}_k(t+) \;=\; (-1)^{k-1}F_k(t-) + \overline{F}_k(t) \;=\; \int \frac{(x-t)^{k-1}}{(k-1)!}\,\mathrm{d}P(x)$$

is a function of $\mu_1(P),\ldots,\mu_{k-1}(P)$, and $(-1)^{k-1}G_k(t) + \overline{G}_k(t+)$ is the same function of $\mu_1(Q),\ldots,\mu_{k-1}(Q)$; hence (61) yields (65) and (66). If now $P-Q$ is assumed to be symmetric, then, using this at the second step below, and using (63) applied to Q and to P at the first and fourth steps, and (65) at the fifth, we get (67) through

$$\begin{aligned}
\overline{H}_k(-t) \;&=\; \int \frac{(x+t)_+^{k-1}}{(k-1)!}\,\mathrm{d}(Q-P)(x) \;=\; \int \frac{(-x+t)_+^{k-1}}{(k-1)!}\,\mathrm{d}(Q-P)(x) \\
&=\; \int \frac{(x-t)_-^{k-1}}{(k-1)!}\,\mathrm{d}(Q-P)(x) \;=\; H_k(t) \;=\; (-1)^k\overline{H}_k(t+).
\end{aligned}$$

Back in the general case, since $(\cdot - t)_+^{k-1}$ is $(P+Q)$-a.e. equal to a polynomial of degree $\leq k-1$ if $t \in \mathbb{R}\setminus I$, namely $(P+Q)$-a.e. $(\cdot - t)_+^{k-1} = (\cdot - t)^{k-1}$ if $\{t\} < I$ and $(\cdot - t)_+^{k-1} = 0$ if $\{t\} > I$, we get $\overline{H}_k = 0$ on $\mathbb{R}\setminus I$. Claim (68) follows using (64) and (65).

Assume now $P,Q \in \mathrm{Prob}_s(\mathbb{R})$. If $f \in \mathcal{F}_s$, then the representation $f(x) = \sum_{j=0}^{s-1}\frac{f^{(j)}(0)}{j!}x^j + \int_0^x \frac{(x-y)^{s-1}}{(s-1)!}f^{(s)}(y)\,\mathrm{d}y = \sum_{j=0}^{s-1}\frac{f^{(j)}(0)}{j!}x^j + \int_{\mathbb{R}}\Big((0 \leq y < x) - (x \leq y < 0)\Big)\frac{(x-y)^{s-1}}{(s-1)!}f^{(s)}(y)\,\mathrm{d}y$

and a Fubini calculation, valid due to $\|f^{(s)}\|_\infty \leq 1$ and the moment assumption just introduced, and using (65) with $k = s$, yield the formula

$$Qf - Pf \quad = \quad \int f^{(s)}(x)\overline{H}_s(x)\,\mathrm{d}x. \tag{74}$$

By applying (74) to $f \in \mathcal{F}_s^\infty$ and using $\|f^{(s)}\|_\infty \leq 1$ we get "$\leq$" in (69). By applying (74) to a function $f \in \mathcal{F}_s$ with $f^{(s)}(x) = \operatorname{sgn}(\overline{H}_s(x))$ for Lebesgue-a.e. x, and using Theorem 1.7(d), we get "$\geq$" in (69). Finally, (69) and (74) yield the claim involving (71).

(b) Using (68), the implications $(B_1) \Rightarrow (B_2) \Rightarrow \ldots \Rightarrow (B_s)$ follow from Lemma 2.8 up to the statement involving (47), since (62) yields $\overline{H}'_{k+1}(t) = -\overline{H}_k(t)$ for $k \in \{1,\ldots,s-1\}$ and $t \in \mathbb{R}$, except for at most countably many t in case of $k = 1$. The equivalence $(B_s) \Leftrightarrow \overline{H}_s \geq 0$ is trivial, and the equivalence $\overline{H}_s \geq 0 \Leftrightarrow P \leq_{s-\mathrm{cx}} Q$ is [13, Theorem 3.2], using (63). Since (71) holds for f from (72) iff $\overline{H}_s \geq 0$, using the left-continuity of $\overline{H}_s$ and also $\overline{H}_s = 0$ on $\mathbb{R} \setminus I$ for the "only if" part, the final equivalence follows from part (a).

(c) Let $k \in \{1,\ldots,s-1\}$ and assume (C_k). Then, as in the proof of part (b), we deduce that $\overline{H}_{k+1}$ has at most $s-k$ sign changes and is lastly positive. If $\overline{H}_{k+1}$ even had at most $s-k-2 = (s-1)-(k+1) \in \mathbb{N}_0$ sign changes, then $k+1 \leq s-1$, and hence part (b) applied with $s-1$ in place of s would yield $P \leq_{(s-1)-\mathrm{cx}} Q$ and hence $\zeta_{s-1}(P,Q) = \frac{1}{(s-1)!}(\mu_{s-1}(Q) - \mu_{s-1}(P)) = 0$ and thus $P = Q$ by Theorem 1.7(b), in contradiction to (C_k). If $\overline{H}_{k+1}$ had exactly $s-k-1$ sign changes, then, on the one hand, part (b) as it stands would yield $\overline{H}_s \geq 0$, but on the other hand, by (C_k), there would exist a $t_0 \in \mathbb{R}$ such that the left-continuous function $(-1)^{s-k+1}\overline{H}_k$ would be ≥ 0 on $]-\infty, t_0]$ and actually > 0 on some nondegenerate subinterval $]t_1, t_0]$, so that, in view of $\overline{H}_k(t+) = (-1)^k H_k(t)$ by (65), the expression $(-1)^{s+1}H_k(t) = (-1)^{s-k+1}\overline{H}_k(t+)$ would be ≥ 0 for $t \in]-\infty, t_0[$ and > 0 for $t \in [t_1, t_0[$, and hence $\overline{H}_s(t_0) = (-1)^s H_s(t_0-) < 0$ by (66) and the recursion (62), a contradiction. Thus indeed (C_{k+1}) holds.

Let $x_0 \in \mathbb{R}$ and f be as in (73). Then $f^{(s)}(x) = \operatorname{sgn}(x - x_0)$ for $x \in \mathbb{R} \setminus \{x_0\}$, and hence (71) holds iff (x_0) is a sign change tuple for $\overline{H}_s$ and $\overline{H}_s$ is lastly positive. Hence the stated equivalence involving "some" and "some and every" follows using part (a).

The final claim of part (c) follows using the "More precisely" statement of Lemma 2.8.

(d) Suppose that 0 were no sign change point of $\overline{H}_s =: h$. Then at least one of the following three conditions would be violated: (i) $h(x)h(y) \geq 0$ for $x, y \in]-\infty, 0[$, (ii) $h(x)h(y) \geq 0$ for $x, y \in]0, \infty[$, (iii) $h(x)h(y) < 0$ for some $x < 0 < y$. If (i) or (ii) were false, that is, $h(x)h(y) < 0$ for some $x, y \in I$ with $I =]-\infty, 0[$ or $I =]0, \infty[$, then (67) would yield $h(-x+)h(-y+) < 0$, and hence $h(u)h(v) < 0$ for some $u, v \in -I$, leading to $S^-(h) \geq 2$, a contradiction. If (i) and (ii) were true but (iii) not, then $S^-(h) = 0$, again a contradiction.

Thus 0 is a sign change point of $\overline{H}_s$, and hence part (c) yields, since condition (C_s) is fulfilled, that (70) holds with f from (73) with $x_0 = 0$.

Hence, if s were even, then (70) would hold with f from (72), but then by part (b) we would have (B_s), that is, $\overline{H}_s$ would have no sign changes, a contradiction. Therefore s is odd. □

From the following example, which in particular computes ε_1 from (19), the results (75) and (76) are used in the proofs of Theorems 1.14 and 1.9 in section 6 below.

Example 4.3. *Let* $Q := \frac{1}{2}(\delta_{-1}+\delta_1)$. *Then we have*

$$\varepsilon_1 \;=\; \zeta_3(Q,\mathrm{N}) \;=\; \frac{1}{6}\left(\frac{4}{\sqrt{2\pi}}-1\right) \;<\; 0.0993, \tag{75}$$

$$\zeta_4(Q,\mathrm{N}) \;=\; \frac{1}{12} \;<\; 0.0834, \tag{76}$$

$$\zeta_s(Q,\mathrm{N}) \;=\; \infty \quad \text{for } s\in\,]4,\infty[. \tag{77}$$

Proof. Claim (77) follows from Theorem 1.7(c) with $m \geq 4$, since $\mu_4(Q) = 1 \neq 3 = \mu_4(\mathrm{N})$.

For proving (75) and (76) using Theorem 4.2, let us change here the notation and put for the rest of this proof

$$P \;:=\; \tfrac{1}{2}(\delta_{-1}+\delta_1), \qquad Q := \mathrm{N}.$$

Then, using from now on the notation of Theorem 4.2 with these P, Q, and first with $s \in \{1,2,3,4\}$ arbitrary, we have (61), and the function $\overline{H}_1 = \overline{G} - \overline{F}$ obviously has the unique sign change tuple $(-1,0,1)$ and hence exactly three sign changes, and is lastly positive.

If now $s = 4$, then assumption (B_1) of Theorem 4.2(b) is fulfilled, and, with $f(x) := x^4/4!$ from (72), we accordingly get

$$\zeta_4(P,Q) \;=\; Qf - Pf \;=\; \frac{1}{4!}(3-1) \;=\; \frac{1}{12}\,.$$

If, finally, $s = 3$, then assumption (C_1) of Theorem 4.2(c) is fulfilled, hence so is (C_3), and, by symmetry of P and of Q, Theorem 4.2(d) now yields

$$\zeta_3(P,Q) \;=\; \frac{1}{3!}\left(Q|\cdot|^3 - P|\cdot|^3\right) \;=\; \frac{1}{6}\left(\frac{4}{\sqrt{2\pi}}-1\right).$$

□

5. Proofs of the main results

Proof of Theorem 1.2. We will use random variable notation whenever this appears to be more convenient. So, in addition to the assumptions of Theorem 1.2, let $X_i \sim P_i$ and $Y_i \sim Q_i$ be $2n$ independent random variables on some probability space with expectation operator $\mathbb{E}$. Without loss of generality, we assume the P_i to be centred, that is, $\mathbb{E}X_i = 0$ for each i.

Step 1. Equality in (7) occurs under the stated conditions. Indeed, we then have $\mathop{\widetilde{\ast}}_{i=1}^n Q_i\, f = 0$ by symmetry, and thus

$$\begin{aligned}\text{L.H.S.(7)} \;&=\; \left|\mathbb{E}\, c\left(\frac{1}{\sigma}\sum_{i=1}^n X_i\right)^3\right| \;=\; \frac{1}{6\sigma^3}\left|\sum_{i=1}^n \mathbb{E}X_i^3\right| 6|c| \\ &=\; \frac{1}{6\sigma^3}\sum_{i=1}^n \sigma_i^3 \left|\mathbb{E}\left(\frac{X_i}{\sigma_i}\right)^3\right| \|f''\|_{\mathrm{L}} \;=\; \text{R.H.S.(7)}\end{aligned}$$

by the equality statement in Example 1.3, that is, by a rather easy part of [58, Theorem 6].

Step 2. We may assume that the Banach space E is the real line $\mathbb{R}$, with the norm being the usual modulus. Indeed, assume Theorem 1.2 to be true in this special case.

Then, for the given general f, the Hahn-Banach theorem [49, Theorem 5.20] yields an $\mathbb{R}$-linear functional $\ell : E \to \mathbb{R}$ of norm 1 satisfying the first of the following equalities

$$\text{L.H.S.}(7) \;=\; \ell\left(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}\, f - \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\, f\right) \;=\; \widetilde{\mathop{\ast}_{i=1}^{n} P_i}\, \ell\circ f - \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\, \ell\circ f,$$

and thus an application of inequality (7) to $\ell\circ f$ in place of f and using $\|(\ell\circ f)''\|_{\mathrm{L}} = \|\ell\circ f''\|_{\mathrm{L}} \le \|f''\|_{\mathrm{L}}$ yields inequality (7) as stated (for example, in the particular case of $E = \mathbb{C}$ we may put $\ell(z) := \Re(cz)$, where $\Re$ stands for the real part and $c = c_f \in \mathbb{C}$ is such that $|c| = 1$ and $c \cdot \left(\widetilde{\ast_{i=1}^{n} P_i}\, f - \widetilde{\ast_{i=1}^{n} Q_i}\, f\right)$ is real and ≥ 0).

Step 3. It is enough to prove inequality (18), since we have $|Pf - Qf| \le \|f''\|_{\mathrm{L}}\, \zeta_3(P,Q)$ for $P,Q \in \mathrm{Prob}_3(\mathbb{R})$ and $f \in \mathcal{C}^{2,1}(\mathbb{R},\mathbb{R})$ by (17) with $s := 3$, and in view of Steps 1 and 2.

Step 4. It is enough to prove inequality (18) in case of $n = 1$, since assuming this special case to be true yields the penultimate step below in

$$\begin{aligned}\text{L.H.S.}(18) &= \zeta_3\left(\frac{1}{\sigma}\sum_{i=1}^{n} X_i\,,\,\frac{1}{\sigma}\sum_{i=1}^{n} Y_i\right) = \frac{1}{\sigma^3}\zeta_3\left(\sum_{i=1}^{n} X_i\,,\,\sum_{i=1}^{n} Y_i\right) \le \frac{1}{\sigma^3}\sum_{i=1}^{n}\zeta_3(X_i,Y_i)\\ &= \frac{1}{\sigma^3}\sum_{i=1}^{n}\sigma_i^3\zeta_3(\widetilde{X_i},\widetilde{Y_i}) \le \frac{1}{\sigma^3}\sum_{i=1}^{n}\sigma_i^3\frac{\beta_i A(\varrho_i)}{6\sigma_i^3} = \text{R.H.S.}(18),\end{aligned}$$

where we have used the homogeneity (52) at the second and fourth steps, and the semi-additivity (51) at the third.

Step 5. Let us write for the rest of this proof

$$Q \;:=\; \tfrac{1}{2}(\delta_{-1} + \delta_1). \tag{78}$$

By Step 4, it remains to prove that we have

$$\zeta_3(P,Q) - \frac{B(\varrho(P))}{6} \;\le\; 0 \tag{79}$$

for $P \in \widetilde{\mathcal{P}_3}$ or, equivalently in view of the alternative representation (16) of ζ_3, that

$$Pf - Qf - \frac{B(\varrho(P))}{6} \;\le\; 0 \tag{80}$$

holds for $P \in \widetilde{\mathcal{P}_3}$ and $f \in \mathcal{C}^{2,1}(\mathbb{R},\mathbb{R})$ with $\|f''\|_{\mathrm{L}} \le 1$. Let $f_1(x) := x$, $f_2(x) := x^2$, and $f_3(x) := |x|^3$ for $x \in \mathbb{R}$. Given now $P \in \widetilde{\mathcal{P}_3}$ and $f \in \mathcal{C}^{2,1}(\mathbb{R},\mathbb{R})$ with $\|f''\|_{\mathrm{L}} \le 1$, we can apply Theorem 3.1(a) to P and to the functions f_1, f_2, f_3, and $f_4 := f$ to conclude, since the left hand side of (80) is a function of Pf_3 and Pf_4, that it is enough to prove (80) under the additional assumption that P has at most 5 support points. (Using instead of Theorem 3.1(a) the a bit deeper Theorem 3.1(b), which applies by the continuity of the functions f_i and the connectedness of $\mathbb{R}$, we could reduce “5” above to “4”, but this does not appear to help in what follows.) Hence it is enough to prove (79) for $P \in \mathcal{P}$ where

$$\mathcal{P} \;:=\; \{P \in \mathrm{Prob}(\mathbb{R}) : \#\mathrm{supp}\, P < \infty, Pf_1 = 0, Pf_2 = 1\}.$$

Let $F(P)$ be the left hand side of (79) for $P \in \mathcal{P}$. Then F is a convex $\mathbb{R}$-valued functional on $\mathcal{P}$, since $P \mapsto \varrho(P) = Pf_3$ is linear on $\mathcal{P}$, B is concave by Lemma 1.1, and $P \mapsto \zeta_3(P,Q)$ is convex since it is the supremum of the affine functionals $P \mapsto Pf - Qf$ with $f \in \mathcal{C}^{2,1}(\mathbb{R},\mathbb{R})$. Hence Tyurin’s Theorem 3.3, with $k := 2$, shows that it is enough to prove (79) for P standardized and having at most three support points. So, for the remaining two steps, let

$$P \quad = \quad p\delta_\alpha + q\delta_\beta + (1-p-q)\delta_\gamma \tag{81}$$

with some $\alpha \le \beta \le \gamma$, $p, q > 0$, $p + q < 1$, $p\alpha + q\beta + (1-p-q)\gamma = 0$, and $p\alpha^2 + q\beta^2 + (1-p-q)\gamma^2 = 1$. Let us further apply the notation $\overline{H}_k$ of Theorem 4.2 with $s := 3$ to the present P from (81) and Q from (78). Then $\overline{H}_1$ has at most $5-2 = 3$ sign changes, since with $S := \{\alpha, -1, \beta, 1, \gamma\}$, only the elements of $S \setminus \{\min S, \max S\}$ can be sign changes.

Step 6. Assume in this step that $\overline{H}_1$ has at most two sign changes. Then, since $\overline{H}_1$ or $-\overline{H}_1$ is lastly positive, Theorem 4.2(b) applied to (P,Q) or to (Q,P) yields the first equality in

$$\zeta_3(P,Q) \quad = \quad \left|\int \frac{x^3}{6}\,\mathrm{d}(P-Q)(x)\right| \quad = \quad \frac{1}{6}\left|\int x^3\,\mathrm{d}P(x)\right| \quad \le \quad \frac{B(\varrho(P))}{6},$$

where the final inequality comes from [58, Theorem 6], that is, from (9) of Example 1.3.

Step 7. Assume finally that $\overline{H}_1$ has exactly three sign changes. Then we have $\alpha < -1 < \beta < 1 < \gamma$, and the (unique) sign change tuple of $\overline{H}_1$ is $(-1, \beta, 1)$, with the interior of the convex hull of its coordinates being $]-1, 1[$. Hence Theorem 4.2(c), with $s = 3$ and with the condition (C_1) being fulfilled, yields the existence of an $r \in\,]-1, 1[$ satisfying

$$\zeta_3(P,Q) \quad = \quad \frac{1}{6}\left(\int |x-r|^3\,\mathrm{d}P(x) - \int |x-r|^3\,\mathrm{d}Q(x)\right). \tag{82}$$

If $r = 0$, then R.H.S.(82) $= \frac{1}{6}(\varrho(P) - 1) \le \frac{1}{6}B(\varrho(P))$, using Lemma 1.1.

So let now $r \ne 0$. Then there is a (unique) two-point law $P' \in \widetilde{\mathcal{P}_3}$ with $\varrho(P') = \varrho(P)$ and concentrated in points $v \cdot \mathrm{sgn}(r)$ and $-u \cdot \mathrm{sgn}(r)$ with certain $u > v > 0$, compare the distribution of X_ϱ in subsection 1.1 above. Lemma 2.5 yields

$$\int |x-r|^3\,\mathrm{d}P(x) \quad < \quad a_r(u,v) + c_r(u,v) + d_r(u,v)\varrho(P) \quad = \quad \int |x-r|^3\,\mathrm{d}P'(x)$$

using also standardizedness of P, P' and $\varrho(P') = \varrho(P)$. Hence, using also (82) in the first step below, we get

$$\zeta_3(P,Q) \quad < \quad \int \frac{1}{6}|x-r|^3\,\mathrm{d}(P'-Q)(x) \quad \le \quad \zeta_3(P',Q).$$

Finally, Step 6 applied to P' in place of P, which is legitimate since the $\overline{H}_1$ corresponding to the two-point law P' has at most two sign changes, yields $\zeta_3(P',Q) \le \frac{1}{6}B(\varrho(P')) = \frac{1}{6}B(\varrho(P))$. □

6. Proofs concerning ζ_3-distances between normal and convolutions of symmetric two-point laws

Proof of Theorem 1.14. Without loss of generality we may assume that

$$\sum_{i=1}^n \sigma_i^2 \quad = \quad 1. \tag{83}$$

Let $Y, Y_1, \ldots, Y_n, Z, Z_1, \ldots, Z_n$ be independent r.v.'s with $Y \sim \frac{1}{2}(\delta_{-1} + \delta_1)$, $Y_i \sim Q_i$ and hence $Y_i \sim \sigma_i Y$, $Z \sim \mathrm{N}$, and $Z_i \sim \mathrm{N}_{\sigma_i}$ and hence $Z_i \sim \sigma_i Z$, for $i \in \{1, \ldots, n\}$. Let further $T_k := Z_1 + \ldots + Z_k + Y_{k+1} + \ldots + Y_n$ for $k \in \{0, \ldots, n\}$. Then, using (83), we get

$T_0 \sim \widetilde{\ast_{i=1}^n Q_i} = \ast_{i=1}^n \widetilde{Q_i}$ and $T_n \sim \mathrm{N}$ and hence, writing in this proof ε_n for a quantity more general than the one introduced in (19), we get

$$\varepsilon_n \;:=\; \zeta_3\Big(\widetilde{\mathop{\ast}_{i=1}^{n} Q_i},\,\mathrm{N}\Big) \;=\; \zeta_3(T_0,T_n) \;\le\; \zeta_3(T_0,T_1)+\sum_{k=1}^{n-1}\zeta_3(T_k,T_{k+1})$$

by using the triangle inequality at the last step.

The regularity (50) and the homogeneity (52) of ζ_3 yield

$$\zeta_3(T_0,T_1) \;\le\; \zeta_3(Y_1,Z_1) \;=\; \sigma_1^3\zeta_3(Y,Z).$$

Noting that the r.v. $Z_1+\ldots+Z_k$ occurring in T_k and in T_{k+1} has the centred normal distribution with variance $\sum_{i=1}^k \sigma_i^2$ and applying Lemma 4.1 with $s=3$ and $t=1$, we get

$$\zeta_3(T_k,T_{k+1}) \;\le\; \sqrt{\frac{2}{\pi}}\cdot\frac{\zeta_4(Y_{k+1},Z_{k+1})}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \;=\; \sqrt{\frac{2}{\pi}}\cdot\frac{\zeta_4(Y,Z)\sigma_{k+1}^4}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \qquad \text{for } k\in\{1,\ldots,n-1\},$$

so that

$$\varepsilon_n \;\le\; \zeta_3(Y,Z)\sigma_1^3+\sqrt{\frac{2}{\pi}}\,\zeta_4(Y,Z)\sum_{k=1}^{n-1}\frac{\sigma_{k+1}^4}{\sqrt{\sum_{i=1}^k \sigma_i^2}}.$$

Using now the assumptions $\sigma_1\ge\sigma_2\ge\ldots\ge\sigma_n$ and (83), we have $\sigma_1^2+\ldots+\sigma_k^2\ge k/n$ and also $\sigma_1^2+\ldots+\sigma_k^2\ge k\sigma_{k+1}^2$, which yields

$$\varepsilon_n \;\le\; \zeta_3(Y,Z)\sigma_1^3+\sqrt{\frac{2}{\pi}}\,\zeta_4(Y,Z)\sum_{k=1}^{n-1}\frac{\sigma_{k+1}^3\min\{1,\sqrt{n}\,\sigma_{k+1}\}}{\sqrt{k}}. \tag{84}$$

Inserting now the values for $\zeta_3(Y,Z)$ and $\zeta_4(Y,Z)$ from (75) and (76) in Example 4.3 yields the claim. □

Proof of Theorem 1.9. For the upper bound we observe that formula (84), specialized to the homoscedastic case $\sigma_1=\ldots=\sigma_n=1/\sqrt{n}$, yields

$$\varepsilon_n \;=\; \zeta_3\left(\widetilde{\mathrm{B}_{n,\frac12}},\mathrm{N}\right) \;\le\; \frac{\zeta_3(Y,Z)}{n^{3/2}}+\sqrt{\frac{2}{\pi}}\cdot\frac{\zeta_4(Y,Z)}{n^{3/2}}\sum_{k=1}^{n-1}\frac{1}{\sqrt{k}},$$

which can further be simplified by use of Lemma 2.9 to give

$$\varepsilon_n \;<\; 2\sqrt{\frac{2}{\pi}}\cdot\frac{\zeta_4(Y,Z)}{n}+\frac{\zeta_3(Y,Z)+\zeta(\frac12)\sqrt{\frac{2}{\pi}}\,\zeta_4(Y,Z)}{n^{3/2}},$$

and now the claimed upper bound for ε_n follows if we substitute the explicit values of $\zeta_3(Y,Z)$ and $\zeta_4(Y,Z)$ as in the preceding proof.

For the lower bound, let us recall for $n,k\in\mathbb{N}_0$ the kth Krawtchouk polynomial P_k^n associated to the symmetric binomial law $\mathrm{B}_{n,\frac12}$ as defined in [36, pp. 130, 151–154, the case of $q=2$ and hence $\gamma=1$] and also, with the unnecessary restriction $k\le n$, in [15, section 6.2 on p. 298, the special case of $p=\frac12$ and hence $\gamma=1$], that is,

$$P_k^n(x) \;:=\; \sum_{j=0}^{k}(-1)^j\binom{x}{j}\binom{n-x}{k-j} \qquad \text{for } x\in\mathbb{R},$$

so that we have in particular

$$P_0^n(x) \;=\; 1, \qquad P_1^n(x) \;=\; -2\left(x-\tfrac{n}{2}\right)$$

and the recursion

$$(k+1)P^n_{k+1}(x) \;=\; (n-2x)P^n_k(x) - (n-k+1)P^n_{k-1}(x) \quad \text{for } k\in\{1,\ldots,n-1\}$$

and hence further

$$\begin{aligned} P^n_2(x) &= 2\left(\left(x-\tfrac{n}{2}\right)^2 - \tfrac{n}{4}\right), \\ P^n_3(x) &= -\tfrac{4}{3}\left(x-\tfrac{n}{2}\right)^3 + \left(n-\tfrac{2}{3}\right)\left(x-\tfrac{n}{2}\right). \end{aligned}$$

If now $n,k\in\mathbb{N}$, then, from the cited sources, we have for $a\in\mathbb{N}_0$

$$\sum_{x=0}^{a} P^n_k(x)\mathrm{b}_{n,\frac12}(x) \;=\; \tfrac{n-a}{k}P^{n-1}_{k-1}(a)\mathrm{b}_{n,\frac12}(a), \tag{85}$$

and hence in particular

$$\begin{aligned} \sum_{x=0}^{a} P^n_1(x)\mathrm{b}_{n,\frac12}(x) &= (n-a)\mathrm{b}_{n,\frac12}(a), \\ \sum_{x=0}^{a} P^n_3(x)\mathrm{b}_{n,\frac12}(a) &= \tfrac{2}{3}(n-a)\left(\left(a-\tfrac{n-1}{2}\right)^2 - \tfrac{n-1}{4}\right)\mathrm{b}_{n,\frac12}(a) \end{aligned}$$

and thus

$$\begin{aligned} \sum_{x=0}^{a}\left(x-\tfrac{n}{2}\right)\mathrm{b}_{n,\frac12}(x) &= \tfrac{a-n}{2}\mathrm{b}_{n,\frac12}(a), \\ \sum_{x=0}^{a}\left(x-\tfrac{n}{2}\right)^3\mathrm{b}_{n,\frac12}(x) &= \sum_{x=0}^{a}\left(-\tfrac{3}{4}P^n_3(x) - \left(\tfrac{3}{8}n-\tfrac{1}{4}\right)P^n_1(x)\right)\mathrm{b}_{n,\frac12}(x) \\ &= \tfrac{a-n}{2}\left(\left(a-\tfrac{n-1}{2}\right)^2 + \tfrac{1}{2}n - \tfrac{1}{4}\right)\mathrm{b}_{n,\frac12}(a), \end{aligned}$$

and finally

$$\begin{aligned} \sum_{x=0}^{n}\left|x-\tfrac{n}{2}\right|^3\mathrm{b}_{n,\frac12}(x) &= -2\sum_{x=0}^{\lfloor\frac{n}{2}\rfloor}\left(x-\tfrac{n}{2}\right)^3\mathrm{b}_{n,\frac12}(x) \\ &= \begin{cases} \tfrac14 n^2\mathrm{b}_{n,\frac12}(\tfrac{n}{2}) & \text{if } n \text{ is even,} \\ \left(\tfrac14 n^2+\tfrac18 n-\tfrac18\right)\mathrm{b}_{n,\frac12}(\lfloor\tfrac{n}{2}\rfloor) & \text{if } n \text{ is odd.} \end{cases} \end{aligned} \tag{86}$$

Recalling the local Edgeworth expansion for binomial laws (see, e.g., [23, § 51, Theorem 1])

$$\sqrt{\tfrac{n}{4}}\,\mathrm{b}_{n,\frac12}(k) \;=\; \Phi'(z)\left(1-\frac{z^4-6z^2+3}{12n}\right) + O(n^{-2})$$

uniformly in $z := (k-\frac{n}{2})/\sqrt{\frac{n}{4}}$ with $k\in\mathbb{Z}$, we thus get, writing $\alpha_n := \frac{n}{2}-\lfloor\frac{n}{2}\rfloor$, and using at the last step below $2\alpha_n^2=\alpha_n$,

$$\begin{aligned} \varepsilon_n &\ge \left|\int \frac{|\cdot|^3}{6}\,\mathrm{d}\left(\mathrm{N}-\widetilde{\mathrm{B}_{n,\frac12}}\right)\right| \\ &= \frac16\left|\frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}}\sum_{x=0}^{n}\left|x-\tfrac{n}{2}\right|^3\mathrm{b}_{n,\frac12}(x)\right| \;=\; \frac16\left|\frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}}\mathrm{R.H.S.}(86)\right| \\ &= \frac16\left|\frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}}\left(\tfrac{n^2}{4}+\alpha_n\tfrac{n}{4}+O(1)\right)\sqrt{\tfrac{4}{n}}\left(\Phi'\left(-\alpha_n/\sqrt{\tfrac{n}{4}}\right)\left(1-\tfrac{3+O(n^{-1})}{12n}\right)+O(n^{-2})\right)\right| \end{aligned}$$

$$= \frac{4}{6\sqrt{2\pi}} \left|1 - \left(1 + \tfrac{\alpha_n}{n}\right)\left(1 - \tfrac{2\alpha_n^2}{n}\right)\left(1 - \tfrac{1}{4n}\right) + O(n^{-2})\right| = \frac{1}{6\sqrt{2\pi}\,n} + O\left(\tfrac{1}{n^2}\right). \quad \square$$

Acknowledgements

We thank Jürgen Müller for showing us the limit result in [25] and [66] used in the proof of Lemma 2.9, and Bero Roos for helpful remarks on an earlier version of the present paper.

References

[1] Angst, J. and Poly, G. (2017). A weak Cramér condition and application to Edgeworth expansions. *Electron. J. Probab.* **22** (59), 1–24.

[2] Arak, T.V. (1981). On the convergence rate in Kolmogorov's uniform limit theorem. I, II. *Teor. Veroyatn. Primen.* **26**, 235–245; **26**, 449–463 (Russian). English transl. *Theory Probab. Appl.* **26**, 219–239; **26**, 437–451.

[3] Arak, T.V. (1982). An improvement of the lower bound for the rate of convergence in Kolmogorov's uniform limit theorem. *Teor. Veroyatn. Primen.* **27**, 767–772 (Russian). English transl. *Theory Probab. Appl.* **27**, 826–832.

[4] Arak, T.V. and Zaitsev, A.Yu. (1986). Uniform Limit Theorems for Sums of Independent Random Variables. *Tr. MIAN USSR.* **174**, 3–214 (Russian). English transl. *Proc. Steklov Instit. Math.* **174**, 1–222 (1988).

[5] Bentkus, V. (1994). On the asymptotical behavior of the constant in the Berry–Esseen inequality. *J. Theoret. Probab.* **7**, 211–224.

[6] Bhattacharya, R.N. and Ranga Rao, R. (2010). *Normal Approximation and Asymptotic Expansion.* SIAM. [In effect a third edition, extending the ones from 1986 and 1976.]

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

[8] Borisov, I.S. and Skilyagina, G.I. (1996). On asymptotic expansion of the moments of smooth functions in the central limit theorem. *Sibirskii mat. zhurnal* **37**, 519–525 (Russian). English transl. *Siberian Math. J.* **37**, 447–453.

[9] Boutsikas, M. (2015) Penultimate gamma approximation in the CLT for skewed distributions. *ESAIM: Probability and Statistics* **19**, 590–604.

[10] Boutsikas, M.V. and Vaggelatou, E. (2002). On the distance between convex-ordered random variables, with applications. *Advances in Applied Probability* **34**, 349–374.

[11] Chistyakov, G.P. (2001, 2001, 2002). A new asymptotic expansion and asymptotically best constants in Lyapunov's theorem. I, II, III. *Teor. Veroyatn. Primen.* **46**, 326–344; **46**, 573–579; **47**, 475–497 (Russian). English transl. *Theory Probab. Appl.* **46**, 226–242 (2002); **46**, 516–522 (2002); **47**, 395–414 (2003).

[12] Cramér, H. (1928). On the composition of elementary errors. First paper: Mathematical deductions. *Skand. Aktuarietidskr.* **11**, 13–74.

[13] Denuit, M., Lefèvre, C., and Shaked, M. (1998). The s-convex orders among real random variables, with applications. *Math. Inequal. Appl.* **1**, 585–613.

[14] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation.* Springer.

[15] Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions: Variations on a theme of De Moivre. *Statistical Science* **6**, 284–302.

[16] Diestel, J. and Uhl. J.J. (1977). *Vector Measures.* American Mathematical Society.

[17] Dinev, T. and Mattner, L. (2012). The asymptotic Berry–Esseen constant for intervals. *Teor. Veroyatn. Primen.* **57**, 381–325 (English, with Russian summary). English transl. *Theory Probab. Appl.* **57**, 323–325 (2013).

[18] Dudley, R.M. (2003). *Real Analysis and Probability.* Second Edition. Cambridge University Press.

[19] Esseen, C.-G. (1945) Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77**, 1–125.

[20] Esseen, C.-G. (1956) A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39**, 160–170.

[21] Ferguson, T.S. (1967). *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press.

[22] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philosphical Soc.* **24**, 180–190.
[23] Gnedenko, B.V. and Kolmogorov, A.N. (1968). *Limit Distributions for Sums of Independent Random Variables.* Addison–Wesley.
[24] Ghosh, J.K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics. Vol. 4. Institute of Mathematical Statistics.
[25] Hardy, G.H. (1949). *Divergent Series.* Oxford University Press, Oxford.
[26] Hipp, C. and Mattner, L. (2007). On the normal approximation to symmetric binomial distributions. *Teor. Veroyatn. Primen.* **52**, 610–617 (English, with Russian summary). English transl. *Theory Probab. Appl.* **52**, 516–523 (2008).
[27] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2001). *Fundamentals of Convex Analysis.* Springer.
[28] Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.* **26**, 268–275.
[29] Hopf, E. (1926). *Über die Zusammenhänge zwischen gewissen höheren Differenzenquotienten reeller Funktionen einer reellen Variablen und deren Differenzierbarkeitseigenschaften.* Norddeutsche Buchdruckerei und Verlagsanstalt A.-G., Berlin.
[30] Ikeda, S. (1959) A note on the normal approximation to the sum of independent random variables. *Ann. Inst. Stat. Math.* **11**, 121–130.
[31] Jiao, Y. (2012). Zero bias transformation and asymptotic expansions. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques.* **48**, 258–281.
[32] Karlin, S. (1968). *Total Positivity.* Stanford University Press, Stanford, California.
[33] Kemperman, J.H.B. (1968). The general moment problem, a geometric approach. *Ann. Math. Statist.* **39**, 93–122.
[34] Kuczma, M. (2009). *An Introduction to the Theory of Functional Equations and Inequalities. Cauchy's Equation and Jensen's Inequality.* 2nd Ed., Birkhäuser.
[35] Ledouox, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* First Reprint 2002, Springer.
[36] MacWilliams, F.J. and Sloane, N.J.A. (1977). *The Theory of Error-Correcting Codes.* Third printing 1981, North-Holland.
[37] Mattner, L. (2001). Complex differentiation under the integral. *Nieuw Arch. Wiskd. (5)* **2**, 32–35.
[38] Mattner, L. and Shevtsova, I.G. (2017). An optimal Berry–Esseen type inequality for expectations of smooth functions. *Dokl. Akad. Nauk* **474**, 535–539 (Russian). English transl. *Dokl. Math.* **95**, 250–253 (2017).
[39] Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Wiley.
[40] Neininger, R and Sulzbach, H. (2015). On a functional contraction method. *Ann. Probab.* **43**, 1777–1822.
[41] *The On-Line Encyclopedia of Integer Sequences* (N. J. A. Sloane, editor, 2010). `https://oeis.org`
[42] Petrov, V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables.* Oxford University Press.
[43] Pinelis, I. (2016). On the extreme points of moments sets. *Math. Meth. Oper. Res.* **83**, 325–349.
[44] Pinkus, A. and Wulbert, D. (2005). Extending n-convex functions. *Studia Math.* **171**, 125–152.
[45] Popoviciu, T. (1933). *Sur quelques propriétés des fonctions d'une ou de deux variables réelles.* Théses. Faculté des Sciences de Paris. See also Mathematica (Cluj) **8**, 1–85 (1934).
[46] Prawitz, H. (1975). On the remainder in the central limit theorem. I: One-dimensional independent variables with finite absolute moments of third order. *Scand. Actuar. J.* **3**, 145–156.
[47] Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley.
[48] Richter, H. (1957). Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Blätter der deutschen Gesellschaft für Versicherungsmathematik* **3**, 147–161.
[49] Rudin, W. (1987). *Real and Complex Analysis.* Third Edition, McGraw-Hill.
[50] Schulz, J. (2016). *The Optimal Berry–Esseen Constant in the Binomial Case.* Dissertation, Universität Trier. Available at `http://ubt.opus.hbz-nrw.de/volltexte/2016/1007/`
[51] Schwartz, L. (1991). *Analyse I. Théorie des ensembles et topologie.* Nouveau tirage février 1995, Hermann.
[52] Senatov, V.V. (1998). *Normal Approximation: New Results, Methods and Problems.* VSP, Utrecht, Netherlands.
[53] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders.* Springer.

[54] Shevtsova, I. (2012). Moment–type estimates with asymptotically optimal structure for the accuracy of the normal approximation. *Annales Mathematicae et Informaticae* **39**, 241–307.

[55] Shevtsova, I.G. (2012). On the accuracy of the normal approximation for sums of independent random variables. *Dokl. Akad. Nauk* **443**, 555–560 (Russian). English transl. *Dokl. Math.* **85**, 274–278.

[56] Shevtsova, I.G. (2012). On the accuracy of the normal approximation for sums of independent symmetric random variables. *Dokl. Akad. Nauk* **443**, 671–676 (Russian). English transl. *Dokl. Math.* **85**, 292–296.

[57] Shevtsova, I.G. (2012). Moment–type estimates with an improved structure for the accuracy of the normal approximation to distributions of sums of independent symmetric random variables. *Teor. Veroyatn. Primen.* **57**, 499–532 (Russian). English transl. *Theory Probab. Appl.* **57**, 468–496 (2013).

[58] Shevtsova, I. (2014). On the accuracy of the approximation of the complex exponent by the first terms of its Taylor expansion with applications. *J. Math. Anal. Appl.* **418**, 185–210.

[59] Shevtsova, I.G. (2016). *Accuracy of the Normal Approximation: Methods of Estimation and New Results*. Argamak–Media (Russian).

[60] Shevtsova, I.G. (2017). A moment inequality with application to convergence rate estimates in the global CLT for Poisson–binomial random sums. *Teor. Veroyatn. Primen.* **62**, 345–364 (Russian).

[61] Tyurin, I. (2009). New estimates of the convergence rate in the Lyapunov theorem. *arXiv:0912.0726*.

[62] Tyurin, I.S. (2009). On the accuracy of the Gaussian approximation. *Dokl. Akad. Nauk* **429**, 312–316 (Russian). English transl. *Dokl. Math.* **80**, 840–843.

[63] Tyurin, I.S. (2010). On the convergence rate in Lyapunov's theorem. *Teor. Veroyatn. Primen.* **55**, 250–270 (Russian). English transl. *Theory Probab. Appl.* **55**, 253–270 (2011).

[64] von Bahr, B. (1965). On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36**, 808–818.

[65] Winkler, G. (1988). Extreme points of moment sets. *Math. Oper. Res.* **13**, 581–587.

[66] Wirths, K.-J. (2015). Telescoping estimates for smooth series. *Canadian Math. Bull.* **58**, 188–195.

[67] Zahl, S. (1963) *Bounds for the Central Limit Theorem Error*. Ph.D. thesis. Harward University, Cambridge, MA. See also SIAM J. Appl. Math. **14** (6), 1225–1245 (1966).

[68] Zolotarev, V.M. (1997). *Modern Theory of Summation of Random Variables*. VSP, Utrecht, The Netherlands.

Universität Trier, Fachbereich IV – Mathematik, 54286 Trier, Germany
E-mail address: mattner@uni-trier.de

Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, and Russian Academy of Sciences, Federal Research Scientific Center "Computer Science and Control," Institute for Informatics Problems, Moscow, Russia
E-mail address: ishevtsova@cs.msu.ru